The King's Fund

Ideas that change health care

Consultation response

# The Nuffield Trust health and social care ratings review

**February 2013**

We welcome the opportunity to contribute to the independent review by the Nuffield Trust commissioned by the Secretary of State for Health to consider whether aggregate ratings of provider performance should be used in health and social care.

This short paper sets out the views of The King's Fund. Overall, while we support the government's commitment to make more information about the quality of care and services publicly available to patients and the public, it is not clear to us that a nationally defined aggregate rating for every health and social care provider would be useful for any of the intended audiences. We set out in more detail below the basis for this position.

Further, we are concerned that there should be coherence between different Department of Health policies relating to quality assurance, performance assessment, pay-for-performance, transparency, accountability, etc. If these are not well-aligned, they could, at best, be confusing for users and, at worst, contradictory. It is important to see how ratings would fit with the Care Quality Commission (CQC) inspection regime, the friends and family test, quality accounts and the NHS Commissioning Board announcement of consultant-level clinical indicators.

We draw on two of our reports that provide insights into how patients use information to make choices and the impact these have on driving improvements in providers: *Choosing a high-quality hospital: the role of nudges, scorecard design and information* (Dixon *et al* 2010b) and *Patient choice: how patients choose and how providers respond* (Dixon *et al* 2010a). We suggest that the review may wish to refer to the reports directly. We have responded to specific questions in the review in more detail below.

The King's Fund
11-13 Cavendish Square
London W1G 0AN
Tel 020 7307 2400

Registered charity: 1126980
www.kingsfund.org.uk

The King's Fund is an independent charity working to improve health and health care in England. We help to shape policy and practice through research and analysis; develop individuals, teams and organisations; promote understanding of the health and social care system; and bring people together to learn, share knowledge and debate. Our vision is that the best possible care is available to all.

© The King's Fund 2013

## Q 6. Do you think a summary rating for a provider is useful?

The usefulness of a summary rating depends on: by whom it is intended to be used; for what purpose (whether choice, performance assessment, quality improvement, pay-for-performance, etc); and how it is constructed. For example, patients and the public may want relatively simple but reliable markers of quality, whereas information for quality improvement needs to be more granular and to have the support of clinicians and others working in the NHS.

A summary rating may be useful for some people, but it is unlikely to suit everyone. Some patients have been shown to prefer a composite or summary measure of quality, others distrust them and want to be able to see the details themselves. In focus groups we found that in general, older people and people with lower levels of education preferred the aggregate measures of quality while younger people and those with higher levels of education preferred to see the individual indicators (Dixon A, *et al* 2010b). There appear to be two explanations for the findings. First, the group with higher levels of education appears to have a greater level of distrust about who had compiled the summary measure, assuming that this was done by the government and therefore they did not trust their judgement. Where the rating was derived from the overall rating of their care provided by previous patients, they were less concerned. The second reason relates to cognitive burden. The experimental stage of the research showed that, other than for those with high levels of numeracy, most people found it difficult to interpret and weigh up the information and come to a view about which was the highest- performing provider. It is likely that this group found the individual level data complex to interpret and the summary measure provided them with a simple shortcut to reach a decision. It is therefore important to give patients and the public an option as to whether or not they would prefer to see a summary measure.

However, there are a number of reasons why a summary rating generally is not very useful, particularly when applied at the organisational level. Most health care organisations provide a heterogeneous range of services. Any summary measure is likely to mask variation across the range of services and may be misleading, providing a false reassurance that all services are of a similarly high quality. For example, acute hospitals provide a vast range of services and there is evidence from external inspections and academic studies that the quality of care varies from one part of a hospital to another. Summary ratings, therefore, do not provide a potential user with useful information to help guide their choice of where to be treated. Nor do they provide useful information for those working in acute hospitals to know how their services compare with others in similar institutions. That requires service- or specialty-specific data.

General practices also provide a range of services, including prevention and health promotion; chronic disease management; and day-to-day care for minor conditions. A summary measure may mask variation and different standards both for the different types of services they provide and between GPs within a practice.

Residential care homes are perhaps the best examples of reasonably homogeneous services in the health and social care sector, in terms of both population and services provided and so summary ratings might be of more use here, although it is important that where an organisation owns several homes the unit to which the rating applies is the individual care home. This is an area in which, unlike the NHS, there is a dearth of data and a need for signposting and where more readily available information would be a great benefit. Arguably, the NHS and the social care sector need different approaches with regard to the provision of publicly available information.

There are a number of important methodological considerations when developing a composite measure of quality or performance. These include: which indicators to use; what weight to give the individual indicators; what adjustment to apply to either the individual indicators and/or the aggregate score; where to set the thresholds between different categories of performance; how to treat missing data or small numbers; and how to ensure stability in the measure while at the same time enabling the indicators to be updated.

There are also important practical considerations when considering how to present data on the performance of health and social care services. The first is the level at which to present data, ie, at an organisational, service/ unit or team/individual level. Second, it is important to define the intended audience and the purpose

for producing the information. Finally, consideration needs to be given to the source of the information (government, independent regulator, commercial provider, patient organisation), the medium through which it is presented to the audience (newspaper, website, consumer report) and presentation (ranking or side-by-side comparison, symbols or affective labels).

It is also important to be clear what the rating is based on. With hotels and accommodation it is possible to have a four-star hotel with great facilities, but for the people who have stayed there giving it a poor rating in terms of the quality of the service on, for example, TripAdvisor. This situation could apply to care homes where the range of facilities available as well as the standard of care may vary. A care home might have a limited range of facilities but still be delivering care to a high standard. Similarly a private sector hospital may be able to offer private rooms with television, but the clinical quality of care may be poor. A rating system could usefully grade establishments in the way that hotels are graded in major guides, ie, there can be a high-quality one-star organisation.

## Q 7 How should any rating be used, who should use it and why?

If we take the potential audiences for this information and consider why they might want information about the quality of services, it appears that an aggregate rating at the level of the organisation is unlikely to be useful. However, there may be some value in an aggregate rating at the unit/service level, though this has some disadvantages (see below) and would need to be presented alongside other disaggregated information.

For patients, users and carers, information about quality of care may be useful to support decision-making about where to seek care. This could particularly be the case for decisions that are not urgent, such as registering with a general practice, finding a residential or nursing home (unless this follows admission to hospital or a sudden crisis that requires rapid admission to a care home or hospital) or choosing where to be referred for an outpatient appointment, diagnostic testing or elective treatment. Critical to patients is that the information is relevant to their situation or condition. This means they want to know about the outcomes and quality of care for people like them, who had the same treatment.

We also know that patients' preferences (ie, what aspects of quality are important to them) differ and are not necessarily stable over time. In generating an aggregate rating, the information provider has to make a judgement about the relative importance of different aspects of care. Their judgement may not be in line with the preferences of a particular patient or user. For example, The King's Fund inquiry into the quality of general practice in England (The King's Fund 2011) showed that while some patients put great value on speed and convenience of access, others put greater value on continuity of care with the same doctor. A single rating would hide any differences between practices on their level of performance in these two areas.

It may be possible to reduce the cognitive burden, but still allow people to make a decision in line with their own preferences and weightings. In the study by Dixon *et al* (2010b) we asked patients before they were presented with information to consider which aspects of quality were important to them. We found that by 'coaching' patients before showing them information about the quality of care, patients were clearer about what was important to them, searched the information more systematically and made higher quality choices. These stated preferences could have been used to create a tailored summary rating online that would have made the comparison between services easier, but respected the values of the individual.

For those managing services, it is vital that they have information and data about the quality of the services they are responsible for. Many boards currently lack the depth of information they need to confidently assure the quality of services across all service lines within their organisation. A common feature of high-performing health care organisations in the USA is their use of data to inform day-to-day clinical practice, drive improvements and underpin the internal performance management systems (Bohmer 2011).

For those providing direct services, it is important that they have up-to-date information about the performance of their own service that helps them both to track the quality of care over time (in order to drive improvements) and to compare their performance with that of other similar units or services. For this purpose, clinical audit data currently provides the most detailed and clinically

relevant information. This also ensures the engagement and buy-in of clinicians and other professionals into the process, which is critically important if the information is to drive quality improvement. Real-time patient feedback collected at ward or service level is also increasingly available and can provide insights into areas where care may not be meeting the expectations of users. But this data is primarily for the purposes of improvement and should only be used for judgement with great caution. Revalidation will also require individual doctors and clinicians to collect information about their own performance.

For regulators and commissioners, information and data is needed to assure them that standards of quality are being met, outcomes are being achieved and that the providers have robust systems of clinical governance to assure quality of care. This data, unlike those used by providers themselves, is being used for judgement (ie, licensing, pay-for-performance, contracting and so on). Whether it comes from routine sources, direct from the provider or from inspections, the data needs to be specific to the service being commissioned/regulated, sensitive and specific enough to pick up if there are problems with the quality of care, and appropriately adjusted to take account of differences in the severity/case mix of patients. These considerations suggest that summary ratings are of very limited use to of any these user groups

If the purpose is to make providers pay more attention to quality issues then research suggests that the main mechanism for this is not through choice but rather through reputational effects. Publishing comparative or benchmarked performance data at both organisational and individual consultant level can have a positive impact particularly on the performance of the worst-performing or lowest-ranked individuals/organisations, although there are also risks of gaming and risk selection. This suggests that the main use of an aggregate or summary rating would be to create reputational effects particularly for poor-performing organisations through so-called naming and shaming, publishing league tables or rankings. The recent announcement by the NHS Commissioning Board that it will publish individual consultant level data for a number of surgical procedures is an example of this.

## Q 8. What might be the key advantages of having a rating?

Ratings reduce the requirement for patients and users to make sense of the complex information available on the quality of information. However, there are other ways of doing this. Symbols are particularly helpful for those less numerate, but need to be tested as some symbols, for example, traffic lights, can be more confusing than helpful for some. Affective labels, using words that convey meaning such as good or excellent rather than pass, met or average, can also be useful.

It is possible to create domains of quality that reduce the level of detail and provide summative measures on, for example, clinical quality, patient experience, safety. The results of the Care Quality Commission's inspections and quality assurance reports, and Monitor's oversight of governance and financial stability, could be helpful here and incorporated as they were with the star ratings and the annual health check.

Summary ratings can provide a simple reassurance for patients, users and the public that a provider is meeting basic standards. The CQC already publishes information about whether providers are meeting standards and presents this in a very clear way across five domains with green ticks and grey and red crosses to convey the degree to which standards are being met. It is not clear that further aggregation of these domains into a single rating would be useful to the public. Above all, a single rating, or domain ratings, should provide unambiguous markers of quality if used for the public. Aggregate measures inevitably have to be qualified, as they inherently depend on the choice of the component parts.

The main advantage of having a rating that we can see is to provide the basis for a overall ranking of organisations/service or individuals in order to highlight where organisations are not performing well and to create external pressure to improve. Evidence from the Wisconsin study (Hibbard *et al* 2005) showed the powerful effect of publishing comparative performance reports compared to just feeding the information back to providers without transparency.

## Q 9. What might be the key disadvantages of having a rating?

Composite ratings can mask important variation in the quality of care within organisations. Patients will be offered false assurance that services of 'high-performing' organisations are of high quality across the board. There is the risk, already apparent in the Mid Staffordshire case, that the regulator or commissioner may give a false positive, ie, giving a positive overall rating, and then subsequently care will be found to be of poor quality. If this risk is realised, the ratings will be undermined and there will be an impact on the reputation of the organisation responsible for them.

Summary ratings are based on the values and judgements of those compiling the ratings, and on choices between very wide-ranging sets of indicators. These may not be the aspects of performance that are most important to the end user, and inevitably there is a trade-off between comprehensiveness and the ensuing complexity on the one hand, and simplicity but inadequate coverage of salient issues on the other. The choice of indicators, and how they are weighted and aggregated, can also result in very different relative rankings of providers. It also leaves them open to manipulation and means that there can be contradictory information in the public domain if different reports/ sources use different methods, which can be confusing for patients and the public.

An example of this, at a much simpler level than ratings, is the divergence between the summary hospital-level mortality indicator (SHMI) and the hospital standardised mortality ratio (HSMR), and the many reports of how these are influenced by coding artefacts. Importantly, it cannot be assumed that summary ratings will have buy-in from clinicians, as they offer little potential for quality improvement. They could also prove to be detrimental to staff morale, and could deflect from the arguably more important task of focusing the minds of clinicians and managers on more granular-level information about quality. Clinical engagement in the development of summary measures will be all important if they are to have credibility.

With the NHS facing unprecedented financial pressures, it will be important to assess the costs (centrally and to providers) and benefits of producing aggregate ratings. There may be insights to be had by looking at the experience of the Healthcare Commission and of the Commission for Health Improvement in carrying out these studies. The production of star ratings and then the annual health check was reportedly very labour intensive – producing ratings across health and social care will be even more so. It is not clear if these ratings will represent value for money, especially given the limited potential for quality improvement.

Finally, it is possible that organisations will try to game the rankings. For example, if a particular service or aspect of quality is given significant weight in the ranking an organisation may focus on the services to the neglect of others. There is also the risk that if they are not properly adjusted there might be risk of selection to avoid more difficult patients whose outcomes may bring down the overall performance.

## Q 10. How should ratings best be presented and reported?

If ratings are used they should be published with clear 'health warnings' alongside more disaggregated data about clinical quality, safety and patient experience.

While most users will not easily understand the detail of the methodology, it is important that the information provider makes it clear which indicators have been used, what value has been placed on them, the quality and source of the underpinning data and how the information should be used/interpreted. There should be a clear 'health warning' alongside the summary measure, along the lines of 'this is a high-level summary of the overall quality of care, it does not guarantee that the quality of all services provided by this organisation is of the same standard'. Ideally, there should be the option for the user to access more disaggregated data about clinical quality, safety and patient experience or to state their own preferences and for summary information to be provided that is tailored to the aspects most important to them.

## Q 11. What can be learned from previous experience of using summary ratings in health and social care?

There have been a number of studies of the impact of the star rating system used by the Healthcare Commission as well as more broadly based analyses of the limitations of summary indicators more generally. These have revealed a number of technical problems with ratings systems of this kind.

a) There is no uniquely correct way of choosing which indicators to use – ie, which elements of an organisation's performance are taken into account and how performance is measured, or of combining the various elements to calculate a summary measure. The rank of any one provider can vary by a large amount as a result of small changes in the method of calculation or of small changes in performance on any one element. This has been demonstrated by Jacobs *et al* (2004), who found that using a different method of combining (weighting) indicators by the Commission for Health Improvement produced very marked changes in the rankings and hence the stars. However, there is no objective way of choosing between weighting methods in this context. The only area of consistency identified by Jacobs *et al* was low performance: the same performers were identified as poor using different methods of combination.

b) If overall ratings are divided into bands of different levels of performance, as they were with the Healthcare Commission's star rating system, there is no objective way of defining where the line should be drawn between the bands. In practice, organisations just above and below any dividing line are likely to have very similar scores, and any differences are likely to lie within the margins of error attached to the underlying data used to derive them.

c) Gravelle *et al* (2012) and studies cited by these authors found different quality measures are not closely correlated. As some of the measures they investigated were condition-specific it follows that any summary indicator relating to a whole hospital is likely to give misleading information to a patient considering where to choose to be treated for those conditions. For example, Rowan *et al* (2004) found that there was no relationship between star ratings and standards of adult critical care.

d) Summary indicators can be volatile or inconsistent with other methods of evaluating performance. The methodology underlying star ratings was complex and changes were made between years in the way the overall score was calculated. As a result, changes in the ratings could not be taken as a guide to changes in performance (Snelling 2003). Clearly, this weakness may in principle be overcome by using a wider range of information and maintaining a consistent methodology. But as the Healthcare Commission found, the desired information may not be available across the whole of a hospital's range of activities.

e) Indicators may be misleading if they do not take into account contextual information such as random variation and measurement error that might impact differentially on actual or recorded performance (Jacobs *et al* 2006).

f) Any publication of performance data risks stigmatising or de-motivating the poor performers and making it harder to attract staff. Conversely, they may provide a stimulus to improvement (Mannion *et al* 2005). It could be argued that the balance between these two effects is likely to be more unfavourable with summary indicators than with indicators that bear closely on particular aspects of performance. The latter points to where action needs to be taken, particularly if the indicators are sound, ie, accurately reflect poor/good performance. In contrast, summary indicators give no information to either providers or consumers as to where poor performance lies.

## Q 12. What can we learn from experience in other countries?

There is a substantial body of international work relating to the use of composite indicators in health as well as other fields. The literature in the United States relating to the use of summary indicators both for health care providers and insurers also confirms the importance of the issues listed above, in particular the lack of an objectively justifiable weighting method. One result of this is confusion on the part of users as to what is the 'right' index to use.

There are (at least) three national systems of hospital rating in use in the United States, as well as some operating at state level. The national systems give very different results. For example, one puts the big teaching and research hospitals at the top; another rates small community hospitals best. A study of websites providing hospital ratings confirms this conclusion (Rothberg *et al* 2008). This study also found variations between rating agencies in reported performance for treatment of specific diagnoses using a condition-specific composite.

On the other hand, some countries, including Sweden and Canada, opt for publishing robust information on quality without attempting to aggregate it into a rating. In Sweden data on evidence-based indicators of quality and efficiency, developed in partnership with health care commissioners and providers, is published at county council level and increasingly at provider level to support transparency and patient choice. See: www.socialstyrelsen.se/lists/artikelkatalog/attachments/18336/2011-5-18.pdf

## Q 13. What can we learn from experience in other sectors?

We have briefly looked at some of the literature on how ratings have impacted on schools and comment on whether these are likely to apply to health and social care.

a) Ofsted bases its ratings on a combination of the outcomes of visits and interviews with staff and pupils. Given resource constraints, visits can take place every few years (in schools judged satisfactory or better): hence there is a risk that any rating becomes out of date. A key assumption behind this approach is that performance typically does not decline rapidly once a given level is reached and that it can be checked between visits from published information.

The timeliness of the information on which a summary rating is based would need to be made clear and further research carried out to understand whether the performance of health and social care organisations is stable over time. Research could also try to identify which factors trigger changes in the performance of an organisation, eg, changes in senior staff (care home manager, clinical director, chief executive) or sudden deterioration in financial position, and this information be used to trigger an inspection.

b) The 'satisfactory' rating band does not distinguish between schools getting better, ie, moving from a lower band, and schools getting worse, ie, moving from a higher band. But this is critical to making an informed choice between schools.

The categories used to describe the performance of organisations are important and carry meaning. Health and social care has experience of star ratings but these may not have a clear meaning for the public and are somewhat tainted by their association with performance management. Categories such as 'below standard', 'meeting standards', 'exceeding standards' could convey the performance where there are defined standards. Relative performance such as 'below average'or 'above average' can be misleading and relies on people understanding a statistical concept. An organisation could be performing above average but the quality of the care may not be good in absolute terms. 'Improving' or 'deteriorating' convey the trajectory of performance, but do not given any sense of the relative quality of care as compared to other providers. The public and those who represent them, eg, governors, Healthwatch, may find such ratings a useful basis on which to challenge a local provider.

c) Judgements that a school is satisfactory appear to have led to a large number of schools 'coasting' as they are under no particular pressure to improve. In addition, the overall judgement can hide areas of poor performance.

It is important that there are external pressures not only on those who are judged to be poor-performing but also on those who are in the middle of the pack to improve and drive for excellence. It is important in designing any rating system that there are positive reputational benefits of being identified as a high-performing or excellent trust. Three-star ratings were previously associated with being able to become a foundation trust and having more autonomy over the running of the organisation. In the NW Premier Advancing Quality scheme there were financial rewards associated with the relative performance of the organisations who participated.

## Conclusion

While we support the government's commitment to make more information about the quality of care and services publicly available to patients and the public, it is not clear to us that a nationally defined aggregate rating for all health and social care providers would be very useful. It is vital that the information that is made available to patients and the public about the quality of care of different providers is clearly presented. Where possible it should be tailored to individual preferences, both in terms of what aspects of quality are important to them but also to their preferred way of viewing information and their cognitive and numeric ability. We think there is potential for some use of summary information at the level of service/unit/care home by domains of care but do not think that a single national summary rating for providers of health and social care is desirable.

## Sources


Bohmer RMJ (2011). 'The four habits of high-value health care organizations'. *New England Journal of Medicine*, vol 365, no 22, pp 2045–7.

Dixon A, Appleby J, Robertson R, Burge P, Devlin N, Magee H (2010a). *Patient Choice: How patients choose and how providers respond*. London: The King's Fund. Available at: www.kingsfund.org.uk/publications/patient-choice (accessed on 12 February 2013).

Dixon A, Boyce T, Fasolo B, Reutskaja E (2010b). *Choosing a High-Quality Hospital: The role of nudges, scorecard design and information*. London: The King's Fund. Available at: hwww.kingsfund.org.uk/publications/choosing-high-quality-hospital (accessed on 13 February 2013).

Dixon A, Boyce T, Fasolo B, Reutskaja E (2010c). 'Helping Patients Choose: How to improve the design of comparative scorecards of hospital quality'. *Patient Education and Counseling*, vol 78, no 3, pp 344–9. Available at: www.kingsfund.org.uk/publications/articles/helping-patients-choose-how-improve-design-comparative-scorecards-hospital (accessed on 13 February 2013).

Goodwin N, Dixon A, Poole T, Raleigh V (2011). *Improving the Quality of Care in General Practice: Report of an independent inquiry commissioned by The King's Fund*. London: The King's Fund. Available at: www.kingsfund.org.uk/projects/gp-inquiry (accessed on 12 February 2013).

Gravelle H, Santos R, Siciliani L, Goudie R (2012). *Hospital Quality Competition under Fixed Prices*. CHE Research Paper 80. Available at: www.york.ac.uk/media/che/documents/papers/researchpapers/CHERP80_hospital_quality_competition_fixedprices.pdf (accessed on 12 February 2013).

Hibbard J, Stockard J, Tusler M (2005). 'Hospital performance reports: Impact on quality, market share, and reputation'. *Health Affairs*, vol 4, no 24, pp 1150–60. Available at: http://content.healthaffairs.org/content/24/4/1150.full (accessed on 12 February 2013).

Jacobs R, Martin S, Goddard M, Gravelle H, Smith PC (2006). 'Exploring the determinants of NHS performance ratings: lessons for performance assessment systems'. *Journal of Health Services Research and Policy*, vol 11, no 4, pp 211–7.

Jacobs R, Smith P, Goddard M (2004). *Measuring performance: an examination of composite performance indicators technical paper 29*. York: Centre for Health Economics. Available at: www.york.ac.uk/che/pdf/tp29.pdf (accessed on 12 February 2013).

Mannion R, Davies, H, Marshall M(2005). 'Impact of star performance ratings in English acute hospital trusts'. *Journal of Health Services Research and Policy*, vol 10, no 1, pp 18–24. Available at: www.ncbi.nlm.nih.gov/pubmed/15667700 (accessed on 12 February 2013).

Rothberg M, Morsi E, Benjamin E, Pekow P, Lindanauer P (2008). 'Choosing the best hospitals; the limitation of public quality reporting'. *Health Affairs*, vol 27, no 6, pp 1680–7. Available at: http://content.healthaffairs.org/content/27/6/1680.full (accessed on 13 February 2013).

Rowan K, Harrison D, Brady A, Black N (2004). 'Hospitals' star ratings and clinical outcomes: ecological study'. *British Medical Journal*, vol 328, pp 924–5. Available at: www.bmj.com/content/328/7445/924 (accessed on 13 February 2013).

Snelling I (2003). 'Do star ratings really effect hospital performance?' *Journal of Health Organization and Management*, vol 117, no 3, pp 210–23.