LSE Health and NHS Confederation Seminar Series 2010

‘A view from Europe: Transferring Health Policy Knowledge from Europe to the English Context’

Measuring & Managing Performance

**Introduction**

This seminar took place at the LSE on 27 April 2010. It was chaired by Peter Smith of Imperial and included presentations by Niek Klazinga (University of Amsterdam/OECD), Mark Exworthy (Royal Holloway), Chris Ham (King's Fund) and Gwyn Bevan (LSE).

Topics covered in the seminar included:
- How and why do we measure performance?
- How do we develop and use quality indicators and quality systems in health care systems?
- Have targets improved performance in the English NHS?
- How does measuring performance affect management within the NHS?

**Niek Klazinga, OECD**

Niek Klazinga of the OECD and University of Amsterdam began the seminar with a discussion of measuring and managing health systems performance, arguing that when utilised improperly, performance management can result in sub-optimisation. That is, if one starts to increase management in one area of health care because it is measurable, this does not always lead to overall improvement in the health system.

In the discussion, Peter Smith (Imperial) stated that international comparisons are probably the single biggest influence on policy-makers and Klazinga discussed the three main reasons for using international performance comparisons: accountability, strategic decision making as well as learning and improvement. In using these, however, it is necessary to be aware of common hazards and other considerations that are discussed below. Throughout his talk, Klazinga focussed on quality of care, namely effectiveness, safety and patient-centredness and referred to the work by the OECD in this areas.

Firstly, measuring is not universal. We need to keep in mind the reasons for which they will be used and ensure that the measures have validity, relevance and utility within this context. His example is that if one wishes to compare departments within
a hospital, the metrics used would be different than if one were comparing hospitals within a country or region. Similarly, data collected for overall health systems performance and international comparisons is different; Different levels of managers need have different information needs; it is important for users of the data to be involved in developing the indicators.

Measuring and governing with indicators is not stand-alone. Performance management linked to other agendas within the health system, such as sustainability, integrated care, patient-centred care, equity, level of regulation in the market and incentive structure – performance indicators/measurement is only one strategy of improving care and outcomes. Indicators used are also a reflection of what outcomes are seen as important and must take into account varying perspectives from public health, medicine, management science, economics, as well as societal and individual values.

To demonstrate the hazards of international comparisons, Klazinga used two specific examples. In terms of cancer, most countries collect large amounts of data on screening rates, survival and mortality, but there is a distinct lack of data on what stage the cancer was when diagnosed, which underlies much of the other outcome data. Additionally, during the discussion, the differences between using 1-year and 5-year survival rates (which can be markedly different) was raised. He also discussed how hospital mortality rates are very much affected by secondary diagnoses and co-morbidities, which are not always reported in a standard manner; this is an area in which improvements in information technology and linking patients records can benefit the quality of data collected.

Finally, simply looking at data is inadequate and one must also delve into the reasons underlying these differences, such as organisational or financial features of the health system or external factors. Overall, performance measures need to be examined in wider systems context and need to take into account guidelines, safety, patient experience, total quality management, as well.

**Mark Exworthy, RHUL**

Mark Exworthy started his talk with the premise that performance is a contested concept, open to challenge, different interpretations and measurements – and thus consequences. He focuses on the English context, in which performance has been the main frame for policy reform and research in recent years. He also spoke of the shift in focus: measuring performance was done at the organisational level (i.e. hospital or clinic), was anonymous, focussed on inputs and outputs, conducted by peers (i.e. surgeons assessing surgeons) and was to be a developmental and learning exercise. It is increasingly becoming focussed on individuals and outcomes; is judgemental and involved named rankings and is increasingly assessed externally.

Recent attention has been directed to the development and use of “performance products” such as indicators, targets, benchmarks and comparisons. Although there are a number of benefits to these approaches, they are not without their shortcomings, especially if one replies solely on quantitative measures: It is unclear how “performance products” improve performance and if they have unintended
consequences elsewhere in the system. Another limitation is “performance chum:” that current indicators are often quite different than previous indicators. Exworthy then focussed on how these products neglect the ‘informal’ and subjective aspects of performance.

Formal measure of performance (i.e. quantitative ‘hard’ metrics) can be complimented by formal measures (i.e. soft information, views and perceptions). Although these subjective measures are often incomplete, they can offer richer data and capture information on reputation and credibility. The point is not to focus on formal or informal, but to examine the interplay between the two.

He cited the case of Foundation Trusts, which have not performed as well as expected. These were introduced in 2004 and have been given much more financial and operational autonomy than other Trusts, in-line with the overall decentralisation of public services. Although they have the ability to exercise this autonomy, many have lacked the willingness to do so: they already felt the had enough autonomy before becoming Foundation Trusts, they are exposed to greater risk, are uncertain about just how far they can take this autonomy and some managers interviewed in the study also feared they could impact negatively on other services and providers in the local area.

A second example was that of the Mid-Staffordshire NHS Foundation Trust. It had performed very well by quantitative measures, but qualitatively patients and staff knew about its poor performance, even referring to one of the wards a “Beirut.” This poor performance was later revealed when 400 patients died unnecessarily in a three-year period and regulators found a number of other failings. A final example is a Primary Care Trust in which three different hospitals were managed in three very different ways, based on qualitative aspects, rather than ‘hard’ indicators: collegiate networks, informal knowledge and the responsiveness to management.

Exworthy’s case studies highlighted how the qualitative data can fill in the gaps in quantitative data. He also concluded by asking what we do with our indicators and the consequences of performance management and the public disclosure of this information. Here he looks at public reporting of surgery mortality and how this can affect patient choices as well as the surgeons whose performance is rated. For example, the performance of the surgeons in questions is affected by the junior surgeons who work with them and senior surgeons may be hesitant to have junior surgeons in the theatre during high-risk operations (anticipating a negative effect on rating), thus affecting the quality of training.

Overall, managing performance necessitates we decide what to measure and how to do so. Returning to the point that performance is contested, we must be aware also of what does not get measured and how this affects performance, as well.
Chris Ham, King's Fund

Chris Ham looked at the broader context of improving performance in the English NHS and the approaches taken by the Labour government over the past 13 years. He mentioned recent King’s Fund review that has reported significant progress in the NHS on in access, waiting times and other key areas. His talk continued by asking why performance has improved, which – as is generally agreed – is very much attributed to the “targets and terror” regime in the NHS. This phrase was coined by Christopher Hood (Oxford) in a joint paper with Gwyn Bevan and refers to the aggressive policy of targets together with sanctions for poorly performing managers and the publication of waiting times data at the hospital level. That which is measured gets managed because it gives managers an incentive to see their targets improve (and hence secure their jobs). A second reason for improvements has been regulation: the role of the Care Quality Commission (and its predecessors) and Monitor, which evaluates Foundation Trusts. Finally, market-based mechanisms have also impacted upon improved performances such as increased patient choice and provider competition; however the impact of these appears to be relatively limited in comparison to “targets and terror” and the role of regulators.

There are, of course, limitations to the use of targets and the command and control style of management. Ham cited the NHS Next Stage Review from 2008 as well as several other reports which argued that the government had created a “culture of compliance” in which the “opportunity for learning was crowded out by the fear of failure.” Here he brings up Sir Michael Barber’s point that targets and command and control management is fully justified when performance is weak, but as performance improves, management style needs to be adjusted to this.

Although in many ways performance management can be seen as a British preoccupation, there are lessons to be learnt from outside the UK. Ham discussed studies from high performing organisations, such as Kaiser Permanente, Jonkoping County Council in Sweden, the Veterans’ Health Administration and Intermountain Health Care in the US. The shared characteristics of high performing organisations include: sustained investment in building leadership capacity at all levels and investment in developing skills in improving quality and delivery of care; part of the latter involved giving front-line staff more involvement and agency to make improvements themselves. Finally, these organisations demonstrate a constancy of purpose and clarity of vision. They have stability in both their managers and the long-term strategies and visions. Ham contrasted Kaiser Permanente’s “culture of commitment” to the UK’s “culture of compliance.” Kaiser promotes the commitment amongst staff of wanting to be the best at what they do and in this sense are “living the dream” of being a self-improving organisation. A point raised in the discussion to this effect is that we should exploit the natural competitiveness of those in the health care profession and further rely on “peer pressure” to improve performance.

Although recent Cabinet Office Strategy Unit reports echo these themes, the UK faces the challenge that our system – and the involvement of the government in it, is set up such that we see reforms overlaid up reform and stability is in lacking.

Finally, Ham asked what should be the focus of performance management. Rather than looking at NHS organisations or individuals, he suggested we look at local
systems of care, or local health economics: that is, the way in which primary care trusts work with together with local authorities and other health provider and tackle the scale of challenges and argued that they should aim to improving outcomes for their populations, not just service users. To this end, regulators need to change how they work with organisations and what they measure, otherwise individual organisations will continue to look after their own self-interest rather than working with other public sector actors.

Gwyn Bevan, LSE
Gwyn Bevan of the LSE asked how do we move from information to improvement; that is, how do we use the information gathered to make changes in performance. He looked at three pathways for improving performance. In this framework, providers may use information to make changes without external pressure (change pathway). That is, if they see they are performing badly, they will act accordingly to improve. This is Julian LeGrand’s notion of knights: politicians and civil servants do the best job they can to respond to the needs of the population they serve. One issue with this pathway is that one needs to understand why one is performing badly (not just that one is). The other issue is that it does not seem to affect change.

Secondly, the selection pathway is when providers may make changes in response to market pressures in which patients choose good providers over poor ones. He looked briefly at data from the US that has a much longer history of public “report cards,” specifically citing the New York State Cardiac Surgery Reporting system. One consideration in this is whether or not patients really do have choice or the motivation to change. For example, if there is only one hospital within a 50-mile radius, to what extent can a patient choose? Secondly, there is simply little evidence that patients switch providers, even when given information on performance.

Finally, providers may respond to systems which report on performance in a way that may affect their own reputation (reputation), for example “naming and shaming,” as discussed by Chris Ham before. Bevan examines the relative impact of these three pathways, using the UK as a natural experiment. Since 1997, policies in England have been based on the change, reputation and then the selection and reputation pathways; Part of Labour manifesto in 1997 was to dramatically reduce waiting times. New Labour tried initially tried the change pathway, but when it did not produce change, the government introduced its system of “Star ratings” in which NHS organisations were given targets (such as waiting times) and publically “named and shamed” if these were not met (2000-2005).

Since devolution in 1999, however, Scotland, Wales and Northern Ireland have continued with policies on the change pathway (or the policy of “rewarding failure”) and Bevan described how evidence from the four countries demonstrates the effectiveness of the reputation pathway and the relative ineffectiveness of the selection and reputation pathways combined. Overall, waiting times decreased in England, whilst they increased or decreased only slightly in the other countries. The most striking example was that of ambulance response times for life-threatening emergencies in which there was a 20% difference between England and both Northern Ireland and Wales.
Bevan cited Judith Hibbard (University of Oregon)’s work on what is needed in order for this ‘knavish pathway’ to work: ranking, published and widely disseminated information, easily understood information and the promise of future reporting (to give incentive to change). The paradox is that this information is not used by patients to switch providers, yet managers of poor hospitals respond to repair the perceived damage to their reputation (rather than their market share).

The downside, of course, as Bevan pointed out, is that managers and other NHS staff may feel antagonised. He cited nurses who were concerned about even going outside in uniform at the height of the Star Ratings. Partly because of this and in the hope of further improving performance, the government moved towards the combined selection and reputation pathways. Sir Michael Barber had the idea that the reputation pathway would take a system from poor to adequate but not to good. LeGrand proposed that market pressure in the form of a quasi-market and consumer choice could take the system to the next level and make it great. Unfortunately, this does not seem to be the case. Rather, it appears the NHS has incurred the transaction costs of the market without the benefits.

Finally, Bevan concluded by suggestion we further examine why devolution has not worked. For instance, a country the size of Wales should be able to deliver services well. Secondly, he suggested the idea of local standards (in addition to national requirements) that may hold providers accountable to local goals, as well.
Discussion & Questions

Peter Smith, the chair, noted three main themes coming from the session. Firstly, the importance of consistency over time: consistency of vision, regulation and market incentives. Although Bevan had noted that the market incentives did not work in the UK experience, Smith argued that the market never been tested because there had been no long-term consistency and in order for this to happen, the market needs assurance that policy will not simply change in 6 months according to political whim. Secondly was the tension between self-improvement and self-regulations and external pressure and one can ask what is the best balance to find. Finally, there is the concern that the use of targets tends to concentrate activity on the areas measured, whilst neglecting others: what gets measured gets managed.