



HealthStat: The introduction of the new, online, performance information tool to the health system

No. 2 2009

Contents	
Editorial	1
Summary results	2
Introduction	3
Measuring performance in health systems internationally	5
How HealthStat works	8
Synopsis of results	11
Key issues in performance information	13
Conclusion	17
Bibliography	19

Editorial

HealthStat is the new performance information tool the HSE uses to collect information from Irish public health services. Members should be aware of the importance of the introduction of this system as it will be an important database to be drawn upon by the HSE in reporting to the Oireachtas (Parliamentary Questions, Annual Output Statements etc) and as the monthly HealthStat reports attract a lot of media and public interest because of the overall traffic light rating (green, amber and red) given to hospitals. It is also of interest in the context of the challenges the HSE faced in monitoring its budget in 2008.

It is presently restricted to monthly results from 29 teaching, regional and general hospitals but will expand to cover additional general and speciality hospitals later this year and, from next year, all health and social care services provided in the community by Local Health Offices. The results are published online by the HSE.

*Committees Team, June 2009
Library & Research Service
Central Enquiry Desk, Tel. 618-4701*

No liability is accepted to any person arising out of any reliance on the contents of this paper. Nothing herein constitutes professional advice of any kind. This document contains a general summary of developments and is not complete or definitive. It has been prepared for distribution to Members to aid them in their Parliamentary duties. It is not for general circulation outside the Houses of the Oireachtas. Authors are available to discuss the contents of these papers with Members and their staff

Table 1: HealthStat Summary Results: March 2009

Hospitals	Overall	Access	Integration	Resources
Adelaide and Meath Hospital inc NCH Tallaght				
Beaumont Hospital				
Cavan General Hospital				
Connolly Hospital				
Cork University Hospital				
Kerry General Hospital				
Letterkenny General Hospital				
Louth County Hospital Dundalk				
Mallow General Hospital				
Mater Hospital				
Mayo General Hospital, Castlebar				
Mercy University Hospital, Cork				
Merlin Park Hospital, Galway		See UCHG		See UCHG
Midland Regional Hospital, Mullingar				
Midland Regional Hospital, Portlaoise				
Midland Regional Hospital, Tullamore				
Mid-Western Regional Hospital Limerick				
Our Lady of Lourdes Hospital, Drogheda				
Our Lady's Hospital, Navan				
Portiuncula Hospital, Ballinasloe				
Sligo General Hospital				
South Infirmary Victoria Hospital, Cork				
St. Colmcille's Hospital, Loughlinstown				
St. James's Hospital				
St. Luke's General Hospital, Kilkenny				
St. Vincent's University Hospital				
University College Hospital Galway				
Waterford Regional Hospital				
Wexford General Hospital				

Source: adapted from HSE website: <http://www.hse.ie/eng/Healthstat/> (Accessed 02/06/2009)

- : Green means very good performance
- : Amber means average performance, room for improvement
- : Red means unsatisfactory, requiring urgent attention

Legend - Source : <http://www.hse.ie/eng/Healthstat/overall/>

Introduction to HealthStat

The cost of the health service

The importance of measuring performance in the Irish health system cannot be overestimated as such measurement is an integral part of achieving value for money.

Despite a small reduction in the voted expenditure of the Department of Health & Children in the April 2009 budget, it remains (at €15.957 million) second only to the Department of Social & Family Affairs in terms of overall public spending. HealthStat currently covers 29 hospitals¹, but it is intended to expand its coverage during 2009 and 2010. The budget in respect of HealthStat of those 29 hospitals in 2009 is €3.4 billion.

If we make an international comparison in terms of health spending per capita Ireland does rank above the OECD average (in US dollars, \$3,082 in 2006 compared with \$2,824) though it remains lower than in countries such as Norway and Switzerland².

The introduction of HealthStat

The introduction of HealthStat complements and enhances the range of performance measurement tools available to the HSE. The role of such tools has been summarised as:

“Performance measurement has considerable potential in health service management in enabling national priorities for health reform to be translated into organisational and individual objectives, to provide a focus on results, and to enhance accountability” (Butler, 2002, p. 1).

Existing approaches which measure performance in the health sector include the Public Health Information System (a database managed by the Department of Health & Children) and the Hospital in-Patient Enquiry (HIPE) which is a health information system used to collect clinical, demographic and administrative data on discharges from

hospitals in Ireland. The HIPE scheme is managed by the Economic and Social Research Institute.

Results from the HSE’s new performance information system, HealthStat, were published for the first time in March. The system uses a ‘traffic light’ classification to award green, amber or red lights to hospitals, symbolising good, average or unsatisfactory performance (see Table 1, p.2). The traffic lights summarise each hospital’s performance on a diverse package of indicators, organised around three key themes of access, integration and resources.

HealthStat is to be used to improve performance in the provision of hospital services in Ireland, and to be used by a range of key stakeholders, including managers and clinicians as well as the HSE (HSE, 2009b). The HealthStat Chair, Maureen Lynott stated:

*“HealthStat is a powerful tool but it is complex—like the services it represents. It is designed for use by professionals working in the health system and also by other professional groups with an interest in the health system”.*³

Table 2 summarises the key features of HealthStat and Table 3 those features that it does not possess:

Table 2 – What HealthStat does

- Systematically collects data on a range of aspects of performance;
- Compares performance to agreed and shared targets;
- Publishes monthly information on how hospitals perform against the targets;
- Uses a ‘traffic light’ system to summarise a wide range of complex material;
- Follows up with suggested actions for improving performance.

¹ Ireland currently has 51 public hospitals. A list can be found at: http://www.hse.ie/eng/Find_a_Service/Hospitals/

² OECD Health Data Note (2008)

³ HSE press release ‘HSE’s HealthStat – a finger on the pulse of Hospital performance’ 23/04/2009.

Table 3 – What healthStat does not do

- It does not measure clinical outcomes or standards;
- It does not measure population health outcomes;
- It does not rank hospitals by performance;
- It does not provide interactive online material;
- It is not intended to guide patients in choosing hospitals.

This Spotlight locates HealthStat in the context of comparative performance measurement and management systems.

The following section briefly reviews some of the sources of variation in performance information systems in different countries, including short overviews of systems which operated in England, Finland and New Zealand at various periods. Later sections focus on the details of how HealthStat operates and identify some key issues from the comparative literature on performance measurement.

Global trends

The HealthStat system reflects Irish circumstances, and its aims are specific to those of its designers within the HSE. However, it also clearly reflects global trends in performance measurement in the public sector, in particular the use of public reporting of summary (or composite) measures as to how institutions are meeting (or failing to meet) targets (Papanicolas et al., 2008, Jacobs et al., 2007). As a result, some of the lessons learned from the international experience of performance measurement and management may be relevant to Ireland.

The growing use of performance measurement in health systems worldwide reflects pressures on health systems due to a combination of resource constraints, the growing complexity of medical issues, and an on-going demand for high quality care (Helfert et al., 2005). In the light of these pressures, policy makers often turn to performance measurement tools, drawing on a range of approaches used across the public and private sectors:

“In general, performance measurement seeks to monitor, evaluate and communicate the extent to which various aspects of the health system meet their key objectives” (Smith et al., 2008, p. 2).

Health systems vary substantially from country to country, and so too do the ways in which their performance is measured and managed. In overview, HealthStat is made up of three key elements:

- a data collection process;
- the production of performance statistics in graphical dashboard format;⁴
- a monthly accountability forum (Turner, 2009)⁵.

The way in which performance is measured in health care systems in different countries varies across all three of these dimensions. For example, countries vary in what aspects of the health system they choose to measure (or what data they collect), for whom the information is intended, and to what purposes the information will be put. Performance measurement may encompass very wide-ranging population health measures such as life expectancy. Alternatively, it may focus on specific clinical measures such as post-operative mortality rates in various hospitals, or may address labour productivity measures in individual organisations (Smith et al., 2008, p. 4).⁶

In addition data can be collected for various purposes and for use by different groups of stakeholders – it may be aimed narrowly at hospital managers or physicians, more broadly at regulators or government, or perhaps at patients, to enable them to make a choice of provider where this is appropriate.

Alongside these complex dimensions are choices about the best way of presenting and reporting information, in particular when this information is to be placed in the public domain. Quite often, public reporting goes hand in hand with the allocation of a ‘score’ to an institution, perhaps tied to a visual symbol like a star or a numbered grade (see case studies below), or as in HealthStat, by a system of traffic lights and dashboards (Jacobs et al., 2007, Smith et al., 2008). This overall score (or composite measure) can sometimes be used to rank performance or compare the performance of particular organisations in what is sometimes known as a league table (Smith et al., 2008).

⁴ The dashboard is a series of graphs intended to convey information about performance, and is likened to the way a car dashboard conveys information to the driver (HSE, 2009b). Section 2 discusses this in more detail.

⁵ Mark Turner is Assistant National Director of the Decision Support Unit of the HSE, which is responsible for the HealthStat

programme. This Spotlight draws extensively on his paper presented to the Statistical and Social Inquiry Society of Ireland in April 2009 (Turner, 2009), particularly in Sections 2 and 3.

⁶ Papanicolas et al (2008) suggest five broad areas: population health; individual health outcomes; clinical quality and appropriateness of care; responsiveness of health system; equity; and productivity. Only some of these areas are covered by HealthStat (see section 2).

This latter approach has been used extensively in the UK across the public sector (Hood, 2007).

On the one hand, publishing openly in the form of 'report cards' or other summaries can enhance accountability and stimulate provider improvement (Papanicolas et al., 2008). On the other hand, it has been suggested that publication in tandem with the use of composite measures can have unforeseen adverse consequences.

These pros and cons are discussed in some detail in the section entitled 'Synopsis of Results' (p. 11), as are some of the common pitfalls encountered in attempting to measure performance.

The following section provides brief summaries of performance measurement for hospitals in England, Finland and New Zealand at various times. New Zealand and England were chosen as they represent countries which have previously been the subjects of comparative reviews *vis-à-vis* the situation in Ireland (Butler, 200, p. 2) and Finland was chosen to give an example of another EU member state.

Hospital benchmarking case studies

England: From 1998 on, different performance measurement systems developed within the constituent countries of the UK (Bevan and Hamblin, 2009).

In England, star ratings were applied to acute trusts in 2001, to ambulance trusts in 2002, and to all types of NHS trusts between 2003 and 2005, after which star ratings were dropped (Ibid).

Organisations received a score ranging from 0 stars (worst performance) to 3 stars (best performance), based on a set of 'key targets' and also on a wider range of indicators in a 'balanced scorecard'.⁷

Key targets included such dimensions as ambulance response times, time from GP referral to outpatient appointment, hospital cleanliness, admission from A&E, financial management etc.

Waiting times of various kinds dominated these key targets. The regime "*rewarded success and penalised failure in a process of naming and shaming*" (Bevan, 2009).

From 2005, the regime was replaced with one which drew on a broader range of indicators, and used a four point scale of excellent, good, fair and weak (*Society Guardian*, 2005).

Subsequently there has been much debate about the merits and demerits of the target driven approach (Gubb, 2009; Bevan, 2009). Bevan and Hamblin (2009, p. 169) suggest that:

"All of the criticism of star ratings recognise its undeniable effect, but have also identified six significant general problems: in measuring what matters, selection of targets, nature of measures, aggregation for ranking, gaming⁸ and damaging morale".

Some of these criticisms are discussed later in *Spotlight* in the context of general issues around performance measurement (p. 13).

Finland: The benchmarking project in Finland began in 1997 with a pilot programme which expanded to cover nearly all publicly delivered specialised health care in Finland.

Since 2006, benchmarking data is integrated into national statistics (Linna, 2006). Data collected from hospitals includes inpatient and outpatient care, as well as information on diagnoses and procedures (Smith et al., 2008).

One of the aims of the project "*was to develop a new measure to describe the output of hospitals better than the traditional measures such as admissions or outpatient visits*" (Linna, 2006).

⁷ A 'balanced scorecard' is a particular form of performance measurement which groups metrics around a set of indicators which is intended to provide a balanced view of an organisation (Wyatt, 2004, p. 78).

⁸ i.e. Manipulating the system

The benchmarking data is put to a range of uses: examples include setting hospital budget target levels, reallocating resources and restructuring care processes, such as the balance between inpatient and outpatient care (Linna, 2006). Results indicate productivity varies considerably from hospital to hospital (Smith et al., 2008).

One account of the process suggests that 'reliable, timely and interesting data' is one of the most important prerequisites, and one of the greatest obstacles is trust in data quality (Linna, 2006).

New Zealand: The Ministry of Health in New Zealand has responsibility for monitoring the performance of public hospitals. In pursuit of this, Hospital Benchmark Reports are produced quarterly, and provide information on a range of key indicators. (Walton and Goodwin, 2007).

Data is provided on 15 measures, ranging across financial, resource, efficiency and patient satisfaction issues. The data is adapted from a balanced scorecard measure (introduced in 2001), which was adapted in 2004, and revised in 2005/6 (ibid). Targets are set in some cases and where there are no independent targets, the results of District Health Boards (DHBs) (who have responsibility for hospitals) are compared with either the average or median results for all DHBs combined.⁹

Reports are published on the Ministry of Health website. The most recent, for October – December 2008, notes:

“Comparative data is presented in this report at a highly aggregated level, and without further analysis or evaluation of performance other than the ranking of DHB results against each measure. A poor ranking in the HBI report is not the same as poor performance” (New Zealand Ministry of Health, 2009, p.2).

Again, the section of this *Spotlight* entitled 'Key Issues' (p. 13) further examines methodological issues of this kind in some detail.

⁹ Information from the Ministry of Health website: <http://www.moh.govt.nz/moh.nsf/indexmh/dhb-hospital-benchmark> (Accessed 26/05/2009).

What does HealthStat measure?

As suggested in the introduction, one of the key questions in performance measurement is what exactly to measure and how, or in other words, what indicators or metrics to use. It is very important to note that, at present, HealthStat does not measure clinical outcomes or standards.¹⁰

The issue of what to measure (and how) evolved gradually during the development of the HealthStat project (Turner, 2009). In their 2008 report on the 2006 *Report of the Comptroller and Auditor General*, the Committee of Public Accounts commented on the then absence of adequate performance indicators at hospital level in Ireland:

“The problem up to now has been that, while a lot of data is collected, it is treated in an unconnected fashion and key decision makers do not have access to analysed data that would make for effective decision making. The committee noted that, as part of the HSE’s Transformation Programme 2007-2010, an integrated statistical system (HealthStat) will be used to collect and connect such information. This should enable standards and benchmarks to be set and should be a major driver of efficiency in the system” (Committee of Public Accounts, 2008, p. 5).

Early in 2007, work began to examine and document sources of performance information in the HSE, and *“to consider how such information could be integrated and used in a way that would promote continuous improvement in the performance of health services delivery”* (Turner, 2009).¹¹

¹⁰ It is worth noting this because much of the comparative literature focuses on performance measurement in relation to clinical issues (see Bevan and Hamblin 2009 for an overview)

¹¹ Information on the HSE transformation programme is available here:

<http://www.hse.ie/eng/Publications/corporate/transformation.html> (Accessed 22/04/2009).

From its inception, metrics (measures) were organised around the three themes of Access, Integration and Resources:

- Access measures are focused on issues such as waiting times for various services;
- Integration is concerned with whether the services received are patient-centred;
- Resources measures how staff and financial resources are being used (HSE, 2009b).

The number of metrics (measures) used to collect information on these three themes were organised around 12 areas (Turner, 2009, HSE, 2009a). The initial choice of metrics was intuitive to some extent:

“Examination of performance measurement approaches in other countries was of limited assistance in the development of a system for Ireland as other health services had different perspectives due to their varying states of maturity and scope. So the 12 areas nominated by the CEO and his advisor were taken as the starting point. Faith in these 12 areas was confirmed during the first annual review of the HealthStat metrics in mid 2008 when a selected group of hospitals verified that HealthStat had hit on the right fundamental measures” (Turner, 2009, p. 5)

Around 600 or so data items per hospital were needed for these 12 areas, approximately 80% of which were available monthly, and the remainder of which (120 new data items) were not centrally collected. Responsibility for these additional measures was assigned to the Decision Support Unit of the HSE.

Table 4 (adapted from HSE material) provides examples of measures used to capture these three themes, which currently comprise 38 separate metrics.

Table 4: Access, Integration and Resources: examples of measures

Access	Waiting times for: planned procedures; emergency department admissions; diagnostics; therapies; outpatient clinics.
Integration	Day case rates; average length of stay; day-of procedure admission rates; delayed discharges; use of inpatient beds.
Resources	Staffing and absenteeism; management of social work; occupational therapy; radiology and consultant clinics; budget/spend; meeting activity targets.

Source: (HSE, 2009b)

Targets and traffic lights: good, satisfactory and bad performance

HealthStat uses targets to give an indication of how individual hospitals are performing.

There is some variation around how performance is measured across the 38 metrics. In some cases an absolute target is used, in other cases a relative target is used, while in some more, group averages are used.

An example of an absolute target is the percentage of staff hours lost due to absenteeism, which should not exceed 3.5%, while relative targets measure hospitals against what is known as ‘best in Ireland’. This is the average of the performance of the three best hospitals in the peer group for that measure (HSE, 2009a, p. 10). In addition, some measures use a group average.

It is important to note that in some cases, no ‘traffic light’ indicator is available.

Targets are a key element in how performance is displayed in what is referred to as a ‘dashboard’, which is a collection of graphics illustrating to what extent a hospital has met its target.

In effect traffic lights are used as shorthand for performance as measured by closeness to targets.

- **Green:** on target or within acceptable tolerance of the target for individual metric;
- **Amber:** outside target/tolerance and is of concern;
- **Red:** significantly outside target/acceptable tolerance and of major concern (HSE, 2009a, pp. 10-11).

In summary, hospitals receive a traffic light for most metrics, depending on how close or far away they are from an agreed target. These traffic lights are then used to generate a traffic light for each of the three themes, and finally, a traffic light for overall performance, as described in the introduction and summarised in Table 1.

The overall traffic light indicator is generated by a numerical score, which is the mean of the three theme scores combined (which are also numerical).

Within each theme, red is deemed to be equal to three, amber to two and green to one. These are added and then divided by the number of indicators to arrive at the traffic light score.

This methodology can, at times, lead to what initially appear to be anomalous results, as is discussed in more detail in the section entitled ‘Synopsis of Results’ (p. 11) in relation to the results for the Adelaide and Meath hospital.

Another important point to note is that traffic lights are aggregated without weighting: that is, all the measures are given equal value. There are plans, however, to add a weighting facility in the future (Turner, 2009, p. 9). While the issue of weighting can appear very technical, it involves a consideration of whether some aspects of hospital performance are more important than others, and it is discussed in more detail below (p. 13).

Following up results: the Forum

Behn (2008) examines what he terms ‘PerformanceStat’ systems, and suggests that meaningful if not ‘relentless’ follow-up is fundamental to the success of performance information systems of this kind.

In HealthStat, follow-up is organised around a monthly Forum, intended to be a key means of promoting best practice. The Forum was based on a model in use in Baltimore, USA, in which regular meetings to discuss performance are held (see Box 1).

Citing the Baltimore model among others, Behn (2008) suggests that: *“The principal mechanism for follow-up consists of the questions asked during one meeting about the problems examined, the data analyzed, the strategies debated, and the commitments made at previous meetings.”*

Indeed, without this kind of verbal and public follow-up, other mechanisms may be completely ineffective” (Behn, 2008, p. 6).

Box 1 describes the evolution and structure of the HealthStat forum.

The fact that the Chief Executive Officer of the HSE leads the meetings and that the Chief Executive Officers of a number of hospitals attend these meetings underlines the importance of the Forum.

Box 1: The HealthStat Forum

The HealthStat model was influenced by the CitiStat data-tracking and management tool developed in Baltimore, USA, in 2000. This was, in turn, based on an earlier policing system (CompStat) from New York Police Department (Perez and Rushing, 2007). The basic model of these two systems has been followed by many organisations, and one of the most important elements has been said to be the ‘follow-up’ element (Behn, 2008). In Baltimore, various city departments collect information on a variety of indicators, and these are discussed at weekly meetings. According to a Scottish evaluation of the CitiStat model: “CitiStat meetings are intensive and formal; service directors are asked to account for their Department’s performance and offered support to improve performance where necessary” (Sharp et al., 2006, p. i). They have also been described as “interrogative, high pressure and confrontational” Turner, 2009, p. 13).

One of the most important elements which HealthStat has taken from CitiStat is the monthly forum, which is an important part of the follow-up element. Meetings are led by the HSE CEO, and also attended by hospital CEOs and Clinical Directors. Approximately 4-6 hospitals attend, with other hospitals getting written actions. The aim is said to be to share best practice and to address problem areas in specific hospitals. Each hospital will attend at least once a year. The first HealthStat forum based on this model was held in February 2008, and monthly subsequently.

HealthStat: limitations for the user

The heart of the HealthStat system is the ‘dashboard’. Each hospital’s performance is presented with a series of graphs, each illustrating performance on a particular measure (or set of measures). Together these graphs are termed a ‘dashboard’.

As previously outlined, each hospital also receives a traffic light for each measure (where relevant), and for each theme, and finally an overall traffic light outcome, and these form part of the overall dashboard. This material is published online monthly, on a hospital-by-hospital basis. In terms of using the online system (to review monthly results, for example) and, in particular, to attempt to compare hospital performance, there are two issues which should be borne in mind.

The first is that results are presented on a hospital-by-hospital basis. What this means is that HealthStat does not currently compare hospital performance online. Hospitals are measured against targets rather than being ranked on comparative performance. However, from July 2009 comparative tables on some key metrics will be published online, making comparison on these measures easier.¹² The second is that the underlying data is not always published online. This can mean that a precise quantitative description of what a particular measure is showing is not always available alongside the graph, although the underlying data can be acquired on request from the HSE.¹³ This means that precise descriptions of performance, or detailed comparisons can take some time for the outside (non-HSE) user.

¹² Personal communication, Mark Turner, 29/05/2009.

¹³ Ibid.

Synopsis of results: issues in interpretation

No hospital received an overall satisfactory or 'green' rating in the first results for January 2009 (published in March 2009).¹⁴ The majority (25) received an 'amber' (average performance) result, while 4 received an overall 'red' result, implying that the performance was unsatisfactory and needed urgent attention. Nor did any hospital receive an overall green light in the subsequent results for February and March.

However, there has been variation in both the number of overall red traffic lights, and in which hospitals received red lights across the three months published to date.

the National Children's Hospital Tallaght), and the Mid Western Hospital Limerick.

There has also been variation across the three themes in terms of traffic light results, as Table 5 illustrates. In general, hospitals appear to be performing least well under the resources theme, but the number of red traffic lights awarded for resources declined from 20 to 13 between February and March.

Nor does this end the variation which can be seen in the three months of results to date, as is clear if we turn our focus to look in more detail at the level of individual hospitals.

Table 5: Traffic light outcomes (January – March 2009)

	Access			Integration			Resources		
	Jan	Feb	March	Jan	Feb	March	Jan	Feb	March
Green	3	4	4	6	5	5	0	0	1
Amber	20	18	17	22	23	24	10	8	13
Red	5	6	6	1	1	0	18	20	13

In the results for January, the four which received a red traffic light were Cork University Hospital, Mayo General Hospital, Our Lady of Lourdes Hospital, Drogheda, and University College Hospital Galway.

In the February results, two hospitals received an overall red light. These were Mayo General Hospital and Our Lady of Lourdes, Drogheda.

The March results looked different again, with three red lights overall. Although once again Mayo General Hospital received a red light, this time the other two hospitals were the Adelaide and Meath Hospital (incorporating

The example of the Adelaide and Meath including the National Childrens Hospital, Tallaght, is interesting:

- In the results for January, this hospital received an overall amber light, and below this level, two reds (for access and resources) and an amber (for integration).
- In the results for February, the hospital again received an overall amber, but this time it received one red (for access) and two ambers (for integration and resources).
- In the results for March, the hospital received a red overall, but at the same time still received one red (for access) and two ambers (for integration and resources).

¹⁴ 'Hospitals fail to get green light in performance rating' *Irish Examiner*, 24/04/2009, <http://www.irishexaminer.com/ireland/idqlojmhkf/>; 'State's Hospitals fail to make the grade' *The Irish Times*, 24/04/2009, <http://www.irishtimes.com/newspaper/ireland/2009/0324/1224243317913.html>. (Accessed 20/04/2009).

The explanation for this apparent anomaly is the procedures for calculating traffic lights summarised in the previous section. Traffic lights are based on a numerical score, and the sum of two low amber marks can, for example, generate a higher score than the sum of a low amber and a high red mark.

This example points to the necessity for bearing in mind that the apparent simplicity of the traffic light system conceals a great deal of complexity. There are many issues to be borne in mind when interpreting 'composite' indicators like traffic lights, such as whether improvements in performance indicators are due to genuine improvements, or to underlying issues with the data or with the methodology for compiling and reporting that data. The section entitled 'Key issues in performance information' (p. 13) examines these key issues in some detail.

Before turning to this, it is worth considering some of the outcomes which have emerged to date at the level of individual metrics.

As Table 5 indicates, many hospitals performed poorly on resources in January (18 red lights) and February (20 red lights), but fewer performed as badly in March (13 red lights).

At the individual metric basis this reflects outcomes such as the fact that in January and February 27 of the 29 hospitals did not operate within their January budget, but this fell to 25 hospitals in March.¹⁵

Likewise, in January 23 of the 29 received a red light on the absenteeism target (3.5% absenteeism or less), 20 received a red light on absenteeism in February, but this fell to 14 in March.¹⁶

Another resource-based measure is the number of patients who fail to attend outpatient clinics.¹⁷ The measure has an absolute target of 10%. In January and February only one hospital received a green light on this target, while three received greens in March. The level of 'did not attend' was considerably more than 10% in some cases.

Finally, of course, hospitals vary individually across the three months in terms of traffic lights for the three themes and on each individual metric, but this is not examined further here. Again, however, it underscores the need for caution in interpreting HealthStat outcomes, particularly at the aggregate traffic light level.

¹⁵ This refers to metric R1, which has a target based on the percentage variance of hospital expenditure from budget (HSE, 2009a, p. 40).

¹⁶ This refers to hospitals which received a red traffic light on metric R5 which has a target of 3.5% absenteeism or less (HSE, 2009a, p. 44).

¹⁷ This refers to metric R10c, and the target has been set by HealthStat based on current performance and the experience of other countries (see HSE, 2009a).

What should be measured?

One of the first steps in designing a performance information system is the choice of indicators or metrics. The comparative literature suggests that the choice of indicator is often opportunistic, based on whatever data is easily available. If this is the case, it is possible that important dimensions are neglected simply because it is difficult to measure them (Goddard and Jacobs, 2008).

Data shortages aside, the choice of indicator may reflect the purpose of the performance information system, i.e. whether it is primarily aimed at providers, consumers, clinicians, funders and so on (Smith et al., 2008).

We earlier (p. 8) clarified that HealthStat measures evolved from processes internal to the HSE, required some new data collection processes, and is a system designed for use by professionals working in the health system and other professional groups with an interest in the health system.¹⁸ All of these factors need to be considered in evaluating the system.

A related issue concerns what happens to the things which are not measured (for whatever reason) (Bevan and Hamblin, 2009). There is a possibility that performance in areas which are not measured may worsen as a side-effect of a shift in focus, and as untargeted areas are neglected (Smith et al., 2008, Papanicolas et al., 2008).

Are all the measures equally important?

In the HealthStat system, every metric is treated as being of equal importance for the purposes of calculating the traffic lights; individual metrics are simply aggregated.

However, it is intended that a facility for weighting will be added during 2009 (Turner, 2009). Adding weighting raises important issues. Perhaps the most obvious is that it involves making decisions on what the most important aspects of hospital performance are. Goddard and Jacobs (2008, p.9) point out that weights are essentially value judgements about the importance of different elements of performance. Likewise, Smith et al. suggest:

“All the evidence suggests that there exist great variations in the importance different people attach to different aspects of performance, so the specification of a single set of weights is fundamentally a political action” (Smith et al., 2008, p.9).

From a methodological perspective, one study suggests that changes to the weighting structure of performance indicators can have a profound impact on the rankings of organisations (Jacobs et al., 2007, p. 6). The study suggests that this can occur if an indicator in respect of which the organisation either excels or not is given more weight (Ibid). It points to a range of different methods for making choices on weights from getting expert opinions or public views, through to the use of statistical techniques.

What is the cause of poor performance?

HealthStat awards traffic lights which reflect the extent to which targets are or are not being met. An important question may be *“what has caused the observed performance and to which practitioners, organisations or agencies should variations in performance be attributed?”* (Smith et al., 2008, p. 7). Much of the literature on assigning the correct causality in health performance measures is concerned with clinical outcomes, or population health measures, neither of which are the focus of HealthStat.

¹⁸ HSE press release ‘HSE’s HealthStat – a finger on the pulse of Hospital performance’ 23/04/2009:

http://83.71.161.17/eng/News/Top_Story/HSE%20%80%99S_HEALTHSTAT_-_A_FINGER_ON_THE_PULSE_OF_HOSPITAL_PERFORMANCE.shortcut.html?showDoc=1&

¹⁹ Weighting involves assigning some indicators more importance than others, so that they contribute a higher proportion of the final score. It suggests that not everything we wish to measure is of the same order of importance.

However, the general point behind these studies, which is that metrics should only aim to capture performance which is directly attributable to the organisation concerned, remains relevant to the Irish case. The HealthStat forum may offer an opportunity for these issues to be raised by the organisation concerned.

A related issue concerns the design of targets, and the calculation of the 'traffic light' measures. How much variance from targets should be allowed, and how should decisions on this issue be made (Normand, 2009)?

At present, if an institution is within 15% of target it is given a green light, between 15% and 35% the light is amber, while over 35% gets a red (Turner, 2009). This issue, in effect, concerns the 'conceptual framework' of the designers of the system (Smith et al., 2009).

Is public reporting useful?

One of the key arguments in favour of public reporting is that it is a useful way of encouraging improvement in performance.

Information about performance can stimulate improvement in various ways: by identifying areas which should be changed; as a result of consumer power; or through 'naming and shaming' (Berwick et al., 2003, cited in Bevan and Hamblin, 2009; also Hibbard, 2008).

If performance information is not published, then it requires providers to act spontaneously when alerted to poor performance for change to occur (See the review of the literature on this in Bevan and Hamblin, 2009).

The second way improvement might occur is deemed to arise when information aimed at patients is published, enabling them to make choices about providers, especially, perhaps, when clinical care is being measured. However, HealthStat does not measure clinical care, and the information is aimed at health professionals and health managers rather than consumers. Nor does the available comparative evidence suggest that consumers do in fact draw on performance information to make choices about health care.

It has been argued though that the format of public reports may be important in facilitating consumer use (Hibbard, 2008).

From the Irish perspective, of more relevance may be arguments that public reporting helps to stimulate efforts to improve performance because of 'reputational' effects, or the impact of 'naming and shaming' (Hibbard, 2008; Bevan and Hamblin, 2009). This might operate through the HealthStat forum or through media reports of the monthly summaries. Incentives and/or penalties may also add to the impact of public reporting (Ibid.)

Is the data reliable?

Clearly, in order to be useful, performance information needs to be "defined and presented in such a way that it accurately and reliably measures the concepts being examined" (Papanicolas et al., 2008). There are many reasons why data might not be reliable.

Under some circumstances, public disclosure can result in unreliable data, due to what is known as 'gaming' (Jacobs et al., 2007; Smith et al. 2008; Bevan and Hamblin, 2009). Public reporting (especially if combined with incentives or penalties) can provide incentives to providers to ensure that the data they provide meets the targets.

For example, following the introduction of the star system in England (see p.6) British Medical Association surveys suggest 'dubious' practices such as "*moving patients to clinical decision units, making patients wait in ambulances, admitting patients unnecessarily, discharging people too early, and miscoding data*" were all employed to distort data (Gubb, 2009).

Bevan acknowledges that gaming can be an issue but suggests that:

"Gaming does not mean that we ought to reject targets but rather that they are being taken seriously: we should therefore make audit and random checks on gaming practices integral to an effective regime of targets" (Bevan, 2009).

At present, there are no specific audit procedures in the HealthStat process.

Do traffic lights simplify or distort?²⁰

The traffic light system in HealthStat is in line with a growing use of composite (or aggregate) performance measures in the public sector (and health systems) worldwide (Smith et al., 2008; Jacobs et al., 2007).

“Composite measures combine separate performance indicators into a single index or measure and are often used to rank or compare the performance of different practitioners, organizations or systems, by providing a bigger picture and offering a more rounded view of performance” (Smith et al., 2008).

Composite measures are attractive because they seem to provide a way of summarising complex information and avoiding ‘information overload’ (Smith et al., 2008). In the case of HealthStat, the overall traffic light summarises some 38 metrics underpinned by around 350 data items, provided by 5 different sources (HSE, 2009a).²¹ However, there are potential pitfalls with using composite measures.

As Normand (2009) commented with regard to HealthStat, a question concerns whether there is a trade-off between simplification and meaning (Normand, 2009). There are various methodological issues which can mean that composite indicators can be misleading (Goddard and Jacobs, 2008; Bevan and Hamblin, 2009).

Some of these have been described above: measures may be based on what data is available or can be measured rather than what is important; data may be unreliable; the weighting methods might distort the data; and the cause of poor performance may be erroneously attributed to organisations when it is in fact due to other factors.

In addition, there are important statistical and methodological concerns involved, particularly with measuring clinical outcomes, but also with other aspects of performance.

Hence, Smith et al. (2008), suggest that:

“Indicators should be presented with full acknowledgements of any data limitations, including uncertainty estimates and lack of timeliness”.

Table 6 below was compiled by Jacobs et al. (2007), and summarises the arguments for and against composite performance measures such as the HealthStat traffic lights.

²⁰ The wording of this heading is adapted from Normand (2009).

²¹ The full list of sources and metrics is available in HSE (2009a).

Table 6: Arguments for and against the use of composite performance measures

Arguments for	Arguments against
Places performance at the centre of the policy arena	By aggregating measures, composites may disguise failings in parts of the system
Offers policymakers a summary of complex multi-dimensional issues	Difficult to determine the source of poor performance or where to focus remedial action
Presents the 'big' picture and may be easier to interpret	A comprehensive composite may rely on poor quality data in some dimensions
Can offer a rounded assessment of performance	If certain performance measures are excluded, it may distort behaviour in undesirable ways
May promote accountability and facilitate communication with the public	Individual performance measures used in composites are often contentious
Easier to track progress of a single indicator over time rather than a whole package of indicators	Composite measures depend crucially on weights attached to each performance dimension, but the methodology for obtaining and using weights is not straightforward
May stimulate better analytic methods and data quality	
Indicates which organisations represent beacons of best performance	
Indicates priority organisations for improvement efforts	

Source: Jacob et al. (2007), Table 1, p. 2

Conclusion

There is a large literature on what is required in order to attain effective performance measurement in health care systems.

However, it is important to realise that much of this is based on situations where consumers have substantial choice of provider, and where performance measurement systems include information on clinical outcomes.

An example is the four requirements for effective performance measurement identified by Hibbard (2003, as cited in Bevan and Hamblin, 2009).

She suggested that a system must be:

- A ranking system;
- Published and widely disseminated;
- Easily understood by the public (so that they can see which providers are performing well and poorly); and
- Followed up by future reports (that show whether performance has improved or not) (Cited in Bevan and Hamblin, 2009, p. 165).

Although this clearly refers to the impact on consumers and is thus less relevant in Ireland, nevertheless the general points about publication, comprehensibility and follow-up seem likely to be equally relevant here.

Smith et al. (2008), also consider the question of effectiveness:

“Given the increasing demand for performance measurement and given the large set of actors and responsibilities, it is important that policy-makers consider what makes performance indicators effective in improving system performance and accountability” (Smith et al. 2008, p. 16).

Their suggested features are listed under 12 separate headings and are too extensive to be addressed fully here.

However, some of their points have been outlined in this *Spotlight* and include issues such as:

- Only measuring performance which is genuinely attributable to the organisation concerned;
- That a system should not just measure what is easy and/or available; and
- Should clearly point out any technical or statistical problems with the data, with estimates of the likely margin of error where appropriate.

Some of the additional suggestions made by Smith et al. (2009), are that there should be a clear conceptual framework, with clear and consistent definitions of statistically sound indicators.

If there are incentives for performance outcomes, then these need to be monitored to ensure that there are no adverse consequences.

Three general points made by Smith et al. (2008), which are worth drawing attention to are:

- Policy makers should pay attention to the broader health system, and to the organisational context within which performance data are collected and disseminated;
- Performance measurement systems should be monitored and evaluated frequently;
- The political aspects need to be managed; this involves ensuring specific interest groups do not capture the process, and encouraging debate (Smith et al, 2008, p. 17).

These wide-ranging points about performance measurement in health systems may become more relevant as the HealthStat project advances.

At present, HealthStat covers 29 hospitals only.

However, it is intended that it will eventually be extended to cover the remaining general and specialised hospitals.

It will also include the wide range of services that are provided through the 32 Local Health Offices including general practitioner services, public health nursing, child health services, community welfare, chiropody, ophthalmic, speech therapy, social work, addiction counselling and treatment, physiotherapy, occupational therapy, psychiatric services and home help.

The HealthStat system is currently aimed at facilitating debate amongst Health professionals. However, it is available online and as it is developed may become increasingly useful for other users and should increase transparency in debates over value for money in the delivery of primary health services.

The international case studies briefly reviewed in this *Spotlight*, the experience of the pilot phase of HealthStat, and the comparative literature generally indicate that performance measurement and management require an adaptable and reflective process to ensure positive outcomes over the long term.

The accessible nature of HealthStat and its future use by the Members of the Oireachtas may assist in that process.

Bibliography

- Behn, R. D. (2008) PerformanceStat as a leadership Strategy: it don't mean a Thing if it Ain't Got the Follow-Up. *Twelfth Annual Conference of the International Research Society for Public Management*. Brisbane Australia.
- Berwick, D. M., Brent, J. & Coye.M.J. (2003) Connections between quality measurement and improvement. *Medical Care*, 41, 130-138.
- Bevan, G. (2009) Head to head: Have targets done more harm than good in the English NHS? No. *BMJ*, 338:a3129.
- Bevan, G. & Hamblin, R. (2009) Hitting and missing targets by ambulance services for emergency calls: effects of different systems of performance measurement within the UK. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172, 161-190.
- Butler Michelle(2000) *Performance Measurement in the Health Sector*:CPMR Discussion Paper 14
- Committee of Public Accounts (2008) *Third Interim Report on the 2006 Report of the Comptroller and Auditor General: Expenditure on Health Services*.
- Goddard, M. & Jacobs, R. (2008) Using composite indicators to measure performance in health care. *Euro Observer: The Health Policy Bulletin of the European Observatory on Health Systems and Policies*, 10, pp. 8-9.
- Gubb, J. (2009) Head to Head: Have targets done more harm than good in the English NHS? Yes. *BMJ*, 338:a3130.
- Helfert, M., Henry, P., Leist, S. & Zellner, G. (2005) Healthcare performance indicators: Preview of frameworks and an approach for healthcare process-development. IN Soliman, K. S. (Ed.) *Information Management in Modern Enterprise: Issues & Solutions - Proceedings of The 2005 International Business Information Management Conference , July 5 - 7, 2005 Lisbon, Portugal*
- Hibbard, J. H. (2008) What Can We Say about the Impact of Public Reporting? Inconsistent Execution Yields Variable Results. *Annals of Internal Medicine*, 148, 160-161.
- Hood, C. (2007) Public Service Management by numbers: Why Does it Vary? Where has it Come From? What are the Gaps and Puzzles? *Public Money and Management*, April, 95-102.
- HSE (2009a) *HealthStat for Hospitals User Guide*.
http://www.hse.ie/eng/Healthstat/linksinfo/HealthStat_for_Hospitals_User_Guide.pdf.
- HSE (2009b) *HealthStat: Supporting high performance in the Irish Health System*. HealthStat Information Booklet available online at:
<http://www.hse.ie/eng/Healthstat/healthstatbooklet.pdf> (Accessed 07/04/2009).
- Jacobs, R., Goddard, M. & Smith, P. C. (2007) Composite performance measures in the public sector. *Policy Discussion Briefing*. Centre for Health Economics, University of York.

- Linna, M. (2006) Benchmarking hospital productivity. *Health Policy Monitor* 26/04/2006
Available at
http://www.hpm.org/en/Surveys/THL/07/Benchmarking_hospital_productivity.html.
- New Zealand Ministry of Health (2009) DHB Hospital benchmark Information: report for the Quarter October- December 2008. Wellington: Ministry of Health.
- Normand, C. (2009) HealthStat - supporting high performance in the health system: some comments. *Reply to Paper by Mark Turner to the Statistical and Social Inquiry Society of Ireland*, 30/04/2009
- Papanicolas, I., C.Smith, P. & Mossialos, E. (2008) Principles of performance measurement. *Euro Observer: The Health Policy Bulletin of the European Observatory on Health Systems and Policies*, 10, pp. 1-5.
- Perez, T. & Rushing, R. (2007) The CitiStat Model: How Data-Driven Government can Increase Efficiency & Effectiveness. Centre for American Progress.
- Sharp, C., Jones, J. & M.Smith, A. (2006) What do we measure and why? An Evaluation of the CitiStat Model of Performance Management and its Applicability to the Scottish Public Sector. *Scottish Executive Social Research*. Edinburgh.
- Smith, P. C., Mossialos, E. & Panaicolas, I. (2008) Background document: Performance measurement for health system improvement: experiences, challenges and prospects. *WHO European Ministerial Conference on Health Systems: "Health systems and Wealth"*. Tallinn, Estonia.
- Turner, M. (2009) HealthStat - supporting high performance in the health system. *Paper to the Statistical and Social Inquiry Society of Ireland*. 30/04/2009.
- Walton, L. & Goodwin, N. J. (2007) Public hospital benchmarking. *Health Policy Monitor*, October. Available online at
http://www.hpm.org/en/Surveys/The_University_of_Auckland/10/Public_hospital_benchmarking.html.