Performance measurement: is it now more scientific?

The role of performance measurements and clinical indicators has changed from being a minor component in the management of health care to one where the demands for greater accountability, safety, efficiency, and quality have accentuated the role of measurement and reporting. For those who have been involved in the development of indicators this will be seen as progress. However, the increased demand also requires that the field is associated with greater scientific rigour [1].

In this issue, three papers provide a new approach to one component that is used to define indicators: how to develop a quality framework. The framework is the first of four stages that are required in the application of clinical indicators, namely: (i) defining the measures (derived from the framework); (ii) collecting the data; (iii) appropriate analyses and interpretation; (iv) acting on the results.

Below we look at some of the methods used during the last 20 years in this first stage of defining indicators and then consider the methods presented in this issue. The first of the following two quotes illustrates that there has always been concern with the lack of measurement. In the second, Smyth summarizes the situation rather well.

We must formulate some method of hospital report showing as nearly as possible what are the results of the treatment obtained at different institutions. This report must be made out and published by each hospital in a uniform manner, so that comparison will be possible. With such a report as a starting point, those interested can begin to ask questions as to management and efficiency.

E. A. Codman, Product of a Hospital, 1914

If the financial records of any hospital were audited as casually and as ineffectively as in the quality of its patient care, the Administrator and the Governing Board would probably land in gaol. Or barring such an unfortunate happening, the least they could hope for would be financial chaos, unpaid bills, and a richly deserved reputation for business incompetence.

Actually there is much more reason to do a continuing and adequate medical audit in a hospital than there is to keep meticulous and informative financial records. The medical audit deals with the life and health of people: the financial audit is concerned only with money.

John Smyth, Surgeon, Royal Newcastle Hospital, NSW Medical Journal of Australia 1 (10), 1959

By the early 1980s, the measurement of quality began to be focused on the taxonomy given by Donabedian [2] who classified measures as being related to structure, process, or outcome. In the 1990s the evolution of clinical performance in health care became more apparent. As an example, the Australian government and the Australian Council on Health-care Standards funded the development of clinical indicators that were to be integrated into the accreditation process [3]. The indicators were developed by members of the clinical colleges, through expert discussion and consensus followed by piloting the measures in a hospital and modifying the definitions as required. The indicator’s validity was based on the expert groups and relevant literature. There was no overall framework in which they were developed, and the results varied between the different colleges in terms of the number of indicators and whether process or outcome was measured. However, despite this rather slow process, by 2003 there were 245 individual indicators from 20 clinical specialty areas and 715 hospitals reporting one or more of these indicators [4].

Although this growth in the collection of indicators may look promising, there has also been concern expressed that more research is required in the definition of indicators [5]. Alternative frameworks have been developed such as the following which has six dimensions: safety, effectiveness, appropriateness, consumer participation, efficiency, and access. Indicators have been defined within each of these dimensions.

Modified Delphi techniques have been used to rank a battery of indicators in terms of importance, utility, and feasibility [6]. Other approaches have identified the key criteria for an indicator and then developed and evaluated indicators based on these criteria. Some of the criteria identified are that indicators be valid, reliable, and useful, have large denominators and a significant number of events for the numerator, show variation between institutions and allow comparisons, create no perverse incentives, are not subject to confounding or case-mix differences, and can be derived from existing data or are readily available.

In general, indicators have been developed by experts without a framework or more recently the framework is defined in advance and the indicators developed within the framework. The methods used by Nabitz [7] and Roeg [8] tackle the problem from a different perspective. A brainstorming approach is used to identify all aspects of the care required in as broad a manner as possible. In the examples by Nabitz, 73 concepts were identified. Next, concept mapping is carried out which requires each staff person to put these concepts into categories and score them in terms of importance. The results can be written as a 73 by 73 matrix with a zero if the two concepts are not in the same category and a one if they are. Summing the individual matrices gave a data file that principal components can analyse and identify the first two factor scores for each concept. These are used in the $x$–$y$ plots, and the factor scores are also used by the cluster analysis to group similar concepts. The result is a framework with eight dimensions or clusters each of which has an importance score. Using this framework, the definition of indicators can then be carried...
out by ensuring that those clusters with a high score are given more importance.

Some readers may have been involved in developing indicators through committees that seem to go on endlessly: a frustrating process. The concept mapping approach has several advantages: it can be carried out in one day, can be analysed using standard software (although the authors chose not to do this), the dimensions are likely to encompass more aspects than the three proposed by Donabedian or the six listed above. Note, however, that the task of defining the indicators and determining their feasibility still has to be carried out. Those who have been involved in developing indicators may like to ask whether using ‘concept mapping’ would have helped them. From my own experience of being on committees that developed indicators for both hospitals and primary care, the answer is probably ‘maybe’. The task of creating clinical indicators for primary care may have had the most to gain since it was the least developed area at the time. It could be that indicators for hospitals have been thought about for so long that the need to carry out concept mapping may be less useful. For new areas such as substance abuse dependency the benefits are likely to be greater.

If there are concerns with the technique, the one that Nabitz comments on is worth considering. The results are internally consistent for the 16 staff that they used, but how the results would generalize to all staff or another set of staff is not known: that is, is there external validity? Note that the two papers used the same method in very similar contexts and came up with a somewhat different framework. This may be evidence that internal validity does not guarantee reproducibility across other studies. The reasons for this may be many, and one could involve the experience and perspectives of the staff. For example, the attitudes of medical officers are driven by cost, clinical significance, and privacy, those of managers by minimizing the number of indicators while maximizing utility, and data managers and statisticians are concerned with the precision and bias of the measurements. Although the method requires a mixture of staff, giving equal weight to each category of staff may not be appropriate. Should those with more experience in measurement be given more weight when summing the matrices and determining the importance scores?

Considering the approach used to analyse the results, it would be interesting to see what happens when three or more principal component scores are found, and what impact this has when the cluster analysis is carried out. Although the results could no longer be plotted as an \( x-y \) plot, if there were three significant factors as found by Roeg [8], it would be possible to determine clusters based on these three factors. By not including all three factors, one is throwing away possible useful information. These points can be addressed by further research.

However, despite the issues that I have raised, it is worth remembering that the conditions for the development of a science are: meeting the demands from the outside users, providing evidence of some success in answering their requests, and an ongoing development of new techniques and ideas. These papers would appear to satisfy these three conditions and maybe John Smyth, quoted above, would agree that the development of indicators has become more scientific.

Robert W. Gibberd
Head, Discipline in Quality in Healthcare
Faculty of Health,
University of Newcastle NSW Australia
robert.gibberd@newcastle.edu.au