Quality: link with effectiveness

Trevor A Sheldon

There is no consensus over a definition of quality of health care. However, increasingly more user-oriented definitions have been applied, such as the ability of the elements of care to achieve legitimate medical and non-medical goals or to meet legitimate customer needs. In the management of the NHS it has been the level of activity itself which traditionally has been seen as the marker of performance. Doing more was equated with doing better. Such a philosophy is embodied in the Department of Health’s “efficiency index,” which measures change in the level of crude activity per unit of expenditure change. More recently, the NHS has introduced measures of quality which relate to targets in the patient’s charter and are based, primarily, on measuring not how much is done but how fast it is done – that is, waiting times.

Challenging the notion that “faster is better”

The issue of waiting times was analysed recently by Frankel and West, who showed that waiting times should not necessarily be seen as some expression of global deficiencies of the NHS. Firstly, they are made up of a rather small number of procedures; secondly, they are a relatively small percentage of the total throughput; and, lastly, a simple league table takes no account of the severity of the condition of people on waiting lists and the relative benefit they stand to gain from treatment.

More worrying than the actual aggregate time spent waiting for procedures is that some of the procedures people are waiting for are relatively cheaper and more cost effective at increasing quality of life (for example, hip replacements and cataract surgery) than many procedures for which there is little waiting. Improving this aspect of quality lies not in some blanket target for waiting time but redistributing resources away from those procedures which are less cost effective or whose impact on health is less. A sizable part of the waiting list may comprise people waiting for inappropriate treatments; it is even more worrying to think that the process of compelling patients even more rapidly towards unnecessary treatment will be interpreted as a quality improvement.

The targeting of waiting times as a measure of quality fails on at least three counts:

- Distinguishing between inefficiency and excess "legitimate" demand
- Considering the relative ability to benefit different client groups
- On a more practical level it will produce a range of unintended consequences resulting from agents attempting to play the system and manipulate data.

Thus waiting times are likely to be a rather poor proxy for quality in the NHS. Because something is easy to measure does not mean it is a good measure. This is not to argue against comparative waiting list data being available, since the results can usefully promote constructive discussion which generates important insights about the functioning of the service and ways that it can be improved.

The “bottom line” in health care

A simple, though not comprehensive, breakdown of the attributes of quality which can be used for assessment includes access to, continuity of, comprehensiveness of, effectiveness of, and (more controversially) the efficiency of health care. Efficiency is an important inclusion because if resources are not used efficiently, by implication some other allocation could produce more benefit with the same resources.

This paper concentrates on the issue of effectiveness and appropriateness, primarily because they may justifiably be regarded as the “bottom line” in health care. Fundamentally, patients seek health care in anticipation of a valued change in health status or quality of life, rather than any utility that may be derived from the actual process of care, though this does not deny the great importance of professionals showing respect, kindness, and consideration towards their patients. Broadly, effectiveness implies that a treatment is likely to do more good than harm. Appropriateness, defined in several ways, will here be taken to mean the use of health services that improve health outcomes valued by patients that are based on sound scientific evidence. Treatments that are not indicated and from which patients are unlikely to benefit are a waste of public money; welfare could be improved by treating other patients with more appropriate treatments.

In Britain there seems to be an increasing concentration on standards of amenity as a measure of quality; decor and other aspects of the physical environment may take precedence over other, more important aspects of quality. In the United States, despite the existence of comparative outcomes information competition between hospitals is still primarily based on the quality of amenity. Because of the asymmetry of information between patient and providers, patients can be unduly influenced by the wrong informational cues. As a result they may make health care decisions based on signals which are unrelated to the likelihood of their benefiting from care,
and for this reason providing understandable information on effectiveness for health service users is potentially so important in order to promote informed choice.

One of the reasons that quality initiatives in Britain have paid insufficient attention to the dimension of clinical effectiveness is that managers and other key decision makers in the NHS are rather reluctant to do so, and they lack the confidence to get involved in what are seen as clinical areas. A “glass wall” separates

A “glass wall” separates managers . . . from the clinical coalface.

managers (particularly in purchasing authorities) from the clinical coalface. Though medical audit is now more openly discussed between managers and clinicians, and purchasers and providers, many issues of quality have yet to reach down fully into the clinical processes. The way in which professional practice changes to incorporate new evidence is scrutinised and is generally “validated” is still a rather mysterious and mainly confidential process. Some of the inertia is due to the sentimental, though largely unethical, attachment to the notion of clinical autonomy. However, effectiveness and outcome information are not some optional extra, and managers need to be encouraged and supported to pay much more attention to activity at a fundamental level in trying to improve effectiveness.

Effectiveness and appropriateness of health care may be considered at three levels: the treatment decision for individual patients, the outcomes for providers, and population outcomes.

Effectiveness of health care for individual patients

There is international evidence that many health care interventions are used in cases where patients are unlikely to benefit and are therefore inappropriate. For example, the RAND Corporation found widespread inappropriate use of health services for several procedures. Using this approach in Britain, it found that over 20% of coronary angiography was inappropriate and 16% of patients underwent coronary artery bypass surgery for reasons judged to be inappropriate. Though the message is robust, interpretation of these actual estimates requires care, owing to the lack of adequate validation of the RAND Corporation’s approaches to determining levels of effectiveness.

Much inappropriate care reflects the inherent difficulties of practising medicine and practising in other clinical fields which have become so complex, where the outcomes are longer term, and where it is impossible for any one practitioner to keep up with published work and correctly interpret it. It is partly in acknowledgement of this impossible information task that specialised groups are asked to produce syntheses of the evidence and that clinical practice guidelines are being promoted.

The next few years will be critical for us in the development of aids to decision making such as clinical guidelines and their local application in the form of protocols. It has been proposed that purchasers, instead of concentrating on developing contracts for procedures, should purchase agreed protocols as a unit of appropriate treatment. The initiative by Graham Winyard, deputy chief medical officer, to promote clinical effectiveness by urging the adoption of clinical practice guidelines and quality standards for clinical services might have a significant impact if carried out in a scientific fashion, based soundly on evidence to ensure that the guidelines are valid, and in close collaboration with local managers and clinicians.

However, the market is already being flooded with poor quality guidelines which have not been developed according to scientific criteria and are not based on rigorous reviews of the literature. It often seems as though professional groups see the development of guidelines as good in itself – and even as a way to protect professional interests by simply validating existing practice. The work of Grimshaw and Russell points to the difficulty of guidelines to change clinical practice;12 we must ensure that it is changed in the right way.

Given the difficulty in changing professional behaviour, our energies should be devoted to those areas of clinical practice where good evidence exists that change will be worthwhile – in that there is good published evidence (from randomised controlled trials) for or against particular healthcare interventions and there is sufficient evidence that professional behaviour significantly diverges from best practice, taking into account the consistent need for professional judgements deviating from guidelines, owing to the heterogeneity of the patient population.

Utilisation review

In order to identify better where efforts should be concentrated and how this should be monitored, the practice of utilisation review should be introduced, as a complement to guidelines and as part of quality initiatives of providers and purchasers. Utilisation review is (usually) a review of the case notes of patients with reference to a set of defined criteria, indicating the degree to which (allowing for the inevitable uncertainties in health care procedures) were appropriate to the health care needs of the patient and the extent that guidelines, if in place, are being adhered to. In Britain the use of utilisation review where there is good scientific evidence of what constitutes appropriate care would allow purchasers to identify the extent of inappropriateness. There is little evidence that purchasers are aware of the distribution of reasons why patients are being treated and the extent to which this is appropriate.
Some recent evidence indicates that professionals are not availing themselves of easy access to the best information on effectiveness\textsuperscript{14} and are not basing their practice on the best evidence.\textsuperscript{15, 16} The studies provide indirect examples of how a utilisation review, by examining samples of case notes and using routine data, could be used to identify inappropriate activity and monitor attempts at improvement.

\textbf{. . . professionals . . . are not basing their practice on the best evidence.}

Utilisation review linked to clear criteria developed with national and local professional groups can become a useful instrument in identifying where there is a significant divergence between what is done and what should be done. The results can then be used to see whether introducing guidelines is worthwhile and to set the objectives for audit. Purchasers can also use utilisation review to monitor this aspect of quality over time. A few health authorities in Britain have expressed interest in this type of work and will be working towards a pilot. However, this will require considerable research and development and investment of will and resources, for it means more involvement in the clinical arena, which so far purchasers have been rather reluctant to pursue. It is also important that national organisations with a remit to monitor and audit how public money is spent (for example, the Audit Commission and the National Audit Office) start to take a closer interest in this part of health service activity.

Obtaining data on the appropriateness of health care could also be fed back for use in developing better large scale performance indicators, such as the efficiency index. If activity were to be weighted by some measure of its appropriateness it would be more sensitive to the effectiveness of health care activity rather than just the level of activity, an attractive option, but one requiring careful thought, piloting, and evaluation before being introduced for routine monitoring, to avoid, for example, the risk of stimulating improved performance in the targeted areas at the expense of the non-targeted areas.

\textbf{Effectiveness at provider level}

To compare the quality of medical care not only efficacy (processes which are associated with improved outcomes in trials) but also effectiveness as it affects actual measures of health status need to be examined. This is because doing the “right thing” is not the only factor likely to influence outcome. There may be considerable differences in skills, experience, facilities, and support available in hospitals and other health care settings which have an impact on quality.

There has been considerable debate about comparing patient outcomes (for example, mortality) between providers. The fundamental problem is that like is usually not being compared with like: patients are not randomly assigned to hospitals, clinicians, or treatments, and thus differences in outcomes over time or between providers are difficult to interpret since they are affected by differences in case mix. Providers dealing with the more severe cases may seem to be performing least well.\textsuperscript{17} The key challenge is to disentangle the effects of case mix from those of providers in order to attribute cause, and this requires adjusting the outcomes for differences in severity or patient mix.\textsuperscript{18} Statistical adjustment of observational data to remove the confounding effect of case mix is difficult because we may not know the important factors affecting prognosis and the data needed to make such an adjustment are unlikely to be routinely available or are often poorly coded.\textsuperscript{19} Given these caveats, however, it does not make sense to compare waiting times, which are also sensitive to differences in variables like severity, throughput, and demand and are a much weaker measure of quality, while ignoring what may be significant variation in mortality rates or other health status measures.

If adjustment for case mix and the accuracy of data are adequate, comparisons may be used to identify statistically significant outliers, and these can be investigated for evidence of significant deficits in quality of care. For example, Hannan et al identified significant outlier hospitals providing open heart surgery after adjusting for several key variables and found during site visits that poor performers had problems with quality.\textsuperscript{20} This demonstrates the importance of supplementing the calculation and comparison of risk-adjusted mortality rates with reviews of actual care before drawing any conclusions about the effectiveness of care.\textsuperscript{21}

There are a few such examples in Britain in which the results have been used to inform contracting. For example, data on treatment and outcome for men with non-seminomatous germ cell testicular (or scrotal) tumours were used in Scotland after adjustment using validated adjustment factors to examine variations in survival. The death rate in the one specialist unit in western Scotland was found to be significantly lower than the rates in the other four units grouped together.\textsuperscript{22} The non-specialist units had nearly three times the death rate of the specialist one (p < 0.001), and as a result the local health board modified its contracting decisions accordingly. Similar work is also being conducted on the treatment of colorectal cancer.

These methods of analysis which examine provider outcomes have several attractions compared to other approaches. Firstly, they allow first approximate evaluations of the quality of packages of care rather than just the individual treatments. Secondly, they may facilitate the identification of collateral factors such as volume and skills, which affect quality over and above the actual treatment given. This is important because clinical trials have not been used to examine these factors. This
44

We import some evidence that patients treated by clinicians and units carrying out procedures (for example, coronary artery bypass grafts) frequently have better outcomes than those treated by low volume providers. Nevertheless, we should be cautious in assuming that all the association is real and causal and generalisable to the United Kingdom and across all procedures. Thus use of measures of mortality and, possibly, readmission and complication rates or indices adjusted for case mix may have a significant role in allowing us to begin comparing the quality of care. This is a major research issue which deserves more and thorough investigation rather than the polarised and often poorly informed polemic around league tables. Unfortunately, health services research in Britain has been unduly focused on evaluation of treatments to the neglect of other health service factors, some of which have been considered here.

... health services research ... has been unduly focused on evaluation of treatments to the neglect of other health service factors ...

Effectiveness at population level

There is interest in promoting the use of population based health status measures, usually disease specific mortality rates linked to Health of the Nation targets for each district. “However, though the rates are reliable, and valid, they have repeatedly been shown to be insensitive to medical care variation and have limited applicability to assessing the clinical effectiveness of medical care.” Therefore it is important to concentrate on examining individual levels of appropriateness and provider related outcomes. However, these population measures remind us that when considering quality it is important not to focus so closely on the delivery of health care as to lose a more global or population perspective. We know that health care is only one factor, sometimes only a small factor, influencing the health of populations. Within a population perspective, a large scale view is taken of the health care system and its ability to effect changes in the health of the population.

As stated, the drive to reduce waiting times, often for mild conditions, the push to improve the standard of amenity, in short, the concentration on what may be of marginal importance to health may be wasteful. Viewed against the backdrop of poor and, in places extreme, social conditions which directly and adversely affect health status (for example high unemployment, poor housing, and poor nutrition), the expenditure of large sums to achieve these relatively unimportant changes seems highly inefficient and even unethical. A high incidence of smoking and an increasing incidence in young women, as a direct result of targeted marketing by the tobacco industry; poor conditions of health and safety in the workplace; and the increasing inequalities between rich and poor are among the factors which ensure poor health status for large groups in the population. Many of these problems are only amenable to sensible action at a large scale social policy level, and it is unfortunate that public health medicine in the United Kingdom has not operated more convincingly at this level.

Effectiveness is most efficiently enhanced by operating at all levels, political and social, provider and individual treatment level. To concentrate exclusively on the micro level without due regard to the large scale context is likely to compromise effectiveness and reduce efficiency. In summary, though the notion of “quality of care” has become fashionable, most of the focus has been on initiatives such as the patient’s charter, waiting times, quality of the physical environment, patient centredness in outcomes measurement, etc. Nevertheless, at the heart of quality must be the effectiveness and cost effectiveness of interventions. Without ensuring that health technologies are effective and are delivered appropriately then many of the other dimensions of quality may simply be window dressing. Substantial variations in the rates of procedures, the way in which similar patients are treated, and the degree to which professionals often ignore the best scientific evidence have all been well documented. The NHS needs methods for ensuring that the effectiveness dimension of quality is brought to the fore and becomes a routine part of quality assessment and activity. Clinical autonomy can no longer be an excuse for inappropriate care. The challenge for the future is twofold: to increase the amount of health technology assessment carried out and to develop methods of ensuring that health care converges with this best practice – that is, the promotion of evidence based practice. By introducing evidence based clinical guidelines and associated utilisation review and persuading purchasers to “purchase protocols” rather than just procedures the effectiveness dimension may become more routine, but it will require a radical rethink of the type of data collected and the way in which the purchaser provider split is managed.