

# Measuring hospital performance: are we asking the right questions?

Martin McKee MD FRCP<sup>1</sup> Anne-Marie Rafferty BSc DPhil<sup>1</sup> Linda Aiken PhD<sup>2</sup>

*J R Soc Med* 1997;90:187-191

No one, I think, who brings ordinary powers of observation to bear on the sick and maimed, can fail to observe a remarkable difference in the aspect of cases in their duration and their termination in different hospitals.—Florence Nightingale<sup>1</sup>

The observation that survival differs between hospitals was first made by Florence Nightingale over a century ago<sup>2</sup>, and in the intervening period this issue has resurfaced intermittently, largely through the efforts of enthusiastic individuals. Only in the late 1980s and 1990s has it emerged as a major policy issue. Several factors have contributed to this. The most important has been the need in certain industrialized countries to make difficult decisions about resource allocation because of accelerating hospital costs. A recent report from the American Office of Technology Assessment notes that 'The period since 1980 has seen constant change in the role of hospitals all over the world, reflecting both the dynamism of medicine and the tightening financial climate'<sup>3</sup>. This report records how governments are not only pursuing aggregate cost containment policies but are also focusing on more efficient production of hospital services so that they can obtain more benefit for the same investment. Studies for the European Commission and the World Health Organization came to similar conclusions<sup>4,5</sup>, which have important implications for various groups of staff: hospitals are highly labour intensive and staff costs typically account for over half of all expenditure. Unfortunately there is little to guide decisions about where to invest in hospitals since there is scant evidence about which factors make a difference to outcomes. In particular, it is unclear whether it is better to invest in equipment or staff and which is the most effective mix of staff. Consequently, better information on outcomes is seen as an essential tool to enable managers and clinicians make good decisions.

The second factor driving this issue onto the agenda is promotion of consumer choice and thus the provision of information to the public. This is a central tenet of current government policy in the UK and is based on a view that

market forces are the most effective means of improving quality and reducing costs throughout the public sector. Asymmetry of information, whereby the provider is in a much better position than the 'customer' to judge the quality of a product—be it education, health care or some other service—is recognized as a barrier to the creation of such markets<sup>6</sup> and great efforts have been put into overcoming it, despite evidence that this may be very difficult or even impossible<sup>7</sup>. The principal tools have been the production of charters, setting out defined services that should be provided, and performance tables, which are intended to measure the extent to which the service levels are achieved.

In the National Health Service (NHS), these developments are manifest as the Patient's Charter and the consequent tables of hospital performance. In England, until now, the measures of performance have excluded standards of clinical care. This omission has been criticized by several commentators<sup>8</sup>, especially since the Scottish Health Department has already published measures of clinical outcome by hospital<sup>9</sup>, and the English Department of Health now proposes to augment the existing information with clinical indicators in 1998<sup>10</sup>.

## MONITORING QUALITY OF CARE

The dilemmas faced by those who would publish tabulated measures of clinical outcome have now been examined by several researchers. Several important questions arise: are the available data appropriate or of sufficient quality to support such comparisons? Is it possible adequately to adjust for severity? Are the numbers sufficient to draw meaningful conclusions? A study designed to address these questions directly examined deaths following eight common conditions in hospitals in one region<sup>11</sup>. The answer to each question was no. Furthermore, in view of the way those in the NHS had responded to previous attempts to use such information to change behaviour, the paper concluded that publication of death rates was likely to create perverse incentives, concerning both treatment decisions and methods of recording data, that could actually have adverse consequences for patients. A major difficulty, of the government's own making, is the use of the finished consultant episode (defined as the period a patient spends

<sup>1</sup>Health Services Research Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, England; <sup>2</sup>Center for Health Services and Policy Research, School of Nursing, University of Pennsylvania, 420 Guardian Drive, Philadelphia 19104-6096, USA

Correspondence to: Professor Martin McKee

under the care of a single consultant, with the possibility of several accruing during a single admission) as a measure of activity. This makes it difficult, if not impossible, to obtain an accurate denominator<sup>12</sup>. For example, simply having a set of case notes re-coded can increase the number of recorded episodes by 22%<sup>13</sup> as additional episodes are "found". There are now many examples from the NHS of how, when an indicator has been adopted as a performance measure, the way in which it is collected undergoes a change—usually to display those involved in a more favourable light<sup>14,16</sup>. Indeed, the phenomenon by which an indicator ceases to be of any value once it is used in this way is now enshrined in economic theory on the basis of experience with a succession of rapidly obsolete monetarist targets during the early 1980s.

This research was not, however, completely negative since it did identify two examples of performance that seemed to merit further investigation when results were discussed with staff from the hospitals concerned. In one, a hospital appeared to have a much better outcome than others for a certain procedure, and staff who had worked there and elsewhere suggested that the surgeons involved were renowned among junior staff for their attention to both operative and postoperative detail. In another, which had inexplicably poor results, there was evidence of poor relationships between the medical staff that were held by some to hamper team working. In neither instance, however, could it be determined whether these findings were due to chance or to differences in severity.

Several other teams have subsequently elaborated on these issues. A study of patients treated for gastro-intestinal haemorrhage showed that the ranking of hospital performance changed when crude death rates were adjusted for severity, as derived from information available to an admitting doctor, but changed further when information from endoscopy was included<sup>17</sup>. The possibility that even more information might further change rankings could not be excluded. Other studies that have progressively increased the amount of information used to adjust for severity have also indicated how this leads to substantial changes in rankings—for example, a British study of stroke units<sup>18</sup> and an American study of overall hospital mortality<sup>19</sup>. This problem is exactly analogous to non-randomized comparisons, in which it seems impossible to be sure that one has eliminated confounding. One can never be certain that any remaining variation, after adjustment for severity, is attributable to the hospital rather than undetected differences in patient severity.

The statistical issues related to league tables have also been examined in detail, in a paper combining research in the health and education fields<sup>20</sup>. The authors concluded that 'the current official support for output league tables, even adjusted, is misplaced'. The paper was presented at

the Royal Statistical Society and it is noteworthy that, of over 30 individuals invited to contribute to the published commentary that accompanied it, only one, from the Department of Health, argued in favour of publication of such tables.

The scope for manipulation of data and creation of perverse incentives has been examined in a study of the data used to derive the New York 'report cards' that describe death rates for cardiac surgeons in the State<sup>21</sup>. Even though the information is derived from highly specified, complex data sets, well beyond that available from routine data in the UK, there was a substantial increase over time in reported co-morbidity with, for example, rates of chronic obstructive pulmonary disease reported in 6.9% of patients in 1989 and 17.4% in 1991. Rates of congestive heart failure increased from 1.7% to 7.6% in the same period. There was no significant change in less ambiguous variables, such as age, that might explain the apparent increase in severity and consequent improvement in severity-adjusted performance. In some hospitals the increases were even more spectacular, such as the one in which the rate of chronic obstructive pulmonary disease increased from 1.8% to 52.9%. The authors concluded that it was not possible to determine whether the apparently improved performance following publication of the death rates<sup>22</sup> reflected a true improvement or simply a combination of greater selection of patients and the consequences of inflation of recorded severity.

The methodological limitations and the potential disadvantages also emerged from an American survey of cardiologists and cardiac surgeons in Pennsylvania, where death rates for individual surgeons are made public. 87% of cardiologists reported that the tables had no or negligible influence on their referral decisions and less than 10% reported discussing the information with more than one-tenth of their patients who were candidates for surgery. More than half expressed concerns about the scope for manipulation of data. More worrying was that a similar proportion also reported greater difficulty than before in finding surgeons willing to operate on high-risk patients. This was supported by the finding that two-thirds of surgeons reported that they were less willing to do so<sup>23</sup>.

### THE REAL QUESTIONS?

Although it is widely agreed that league tables are flawed, both in terms of their meaningfulness and their vulnerability to manipulation<sup>24</sup>, the fact remains that hospitals differ in their performance in ways that cannot be explained. There is cause for concern; but we argue that those seeking to identify erring hospitals in the UK are asking the wrong questions. The idea of publishing hospital death rates emanates from the USA, where clinical practice is very

different from that in the UK. The commercial ethos there has ensured that many surgeons continue to undertake a much wider range of procedures than would a British surgeon. If there have been any benefits from league tables they seem largely to have arisen from forcing out of business surgeons undertaking very small numbers of procedures, a practice encouraged by a fee-for-service system but discouraged by a wide range of policies in the UK, which has been a world leader in the development of team-working. The days when all the surgeons in a district general hospital treated the same mix of patients, each performing small numbers of widely differing procedures, have largely gone with the growth of sub-specialization in, typically, vascular, breast and abdominal surgery. Instead, we propose, the questions that now need to be answered relate to the organization of care. Two specific issues arise. The first is the optimum size for a particular service and, related to that, how services should be organized to ensure that this is achieved. The second is whether there are aspects of organizational culture that have an impact on care.

### **The size of a service**

The first question has received considerable attention elsewhere. A detailed review was published in 1990<sup>25</sup>, since supplemented by many other papers. In brief, for a wide range of interventions, there is a clear relationship between volume and outcome, with better results obtained in those centres treating larger numbers of patients, although the nature of the relationship observed varies. There are several unresolved questions—not least whether this association is causal. One possibility is that those hospitals treating more patients provide better treatment, whether through practice or the availability of standardized routines, better equipment, or some other factor. The second possibility is that they obtain better results because they treat less seriously ill patients. The latter argument receives some support from a reanalysis of studies of volume and outcome for coronary artery bypass grafting in which the relative advantage of high volume decreased with improved adjustment for case-mix<sup>26</sup>—although this defined 200 procedures per year as the threshold for designation as 'high volume', leaving open the possibility of a causal relationship at lower volumes. Given the limitations of severity measures noted above, this issue is very difficult to resolve but it is of great importance since there are several factors emerging that will create pressure for change in the configuration of British hospital services. These include the shift of minor procedures to primary care, the government's support for a new form of cottage hospital, and the consequences of the Calman recommendations for junior medical staffing<sup>27</sup> (which, taken with policies designed to

reduce hours, threaten the viability of many smaller units). The ultimate shape that these services will take is still far from clear<sup>28</sup>, in view of the often contradictory consequences of the different policies, but decisions must be informed by the best possible evidence on how service configuration is likely to affect quality of care.

### **Organizational culture**

The second question—whether organizational culture can influence the quality of clinical care—is the subject of many anecdotes but remarkably little research. Stories abound of hospitals that have a reputation for either good or bad care. In the recent example of the Treliske Hospital, which has suffered a series of highly publicized mishaps<sup>29</sup>, the absence of good comparative data makes it impossible to know whether Treliske really is performing worse than other hospitals or, as the local medical staff argue, it is merely experiencing the effects of a situation reminiscent of the early discussions about apparent leukaemia clusters near nuclear facilities.

Research in this area has been sparse and much has arisen from nursing rather than medicine. An early study arose from work commissioned by the British Government on nurse training and recruitment, undertaken in anticipation of the creation of the NHS<sup>30,31</sup>. The coauthor of the minority report of this initiative was John Cohen, a psychologist from the Cabinet Office. In attempting to measure the effectiveness of nursing care, Cohen took the novel criterion of patients' length of stay as an outcome variable. He contended that length of stay was associated with the quality of trained staff and represented a valid measurement of the effectiveness of nursing care; thus Cohen was one of the first to analyse the relationship between nurse staffing skill mix and patient outcome<sup>32</sup>.

After many years of dormancy, this question has attracted interest. An investigation in the USA<sup>33</sup> was based on earlier work that had identified certain hospitals ('magnet' hospitals) widely regarded by nurses as offering a good environment in which to practise nursing; outcomes had not been studied, so could not have been a criterion for selection. The hospitals were characterized by greater nursing autonomy and better relationships between doctors and nurses. These 39 hospitals were matched with 195 controls having the same characteristics, by use of a complex multivariate sampling procedure. After adjustment for severity, the magnet hospitals achieved a statistically significant 4.6% lower inpatient mortality rate. This investigation suggests that those factors that lead to a hospital being deemed effective in organizational terms may also contribute to better quality of care. Importantly, this study indicated that, when factors such as board certification of physicians and availability of technology were taken into

account, there was significant residual variation in outcome that could be attributed to nursing organization and staffing.

Other work has reached similar conclusions. The authors of a study that examined the relationship between a measure of organizational culture and a range of variables including implementation of quality assurance activities, charges, length of stay, and perceived patient outcome concluded that there were tangible patient benefits from a culture that was supportive and encouraged flexibility<sup>34</sup>. Another study showed that organizational and professional job satisfaction among nurses is a strong predictor of process measures of quality of care<sup>35</sup>.

Finally, several investigations of intensive care units, taking advantage of the much greater scope for severity adjustment in the patients, indicate that units with apparently good and bad results cannot be distinguished on the basis of global judgments based on site visits or organization or structural factors. They do, however, differ in terms of certain practices such as the presence of a patient-centred culture, strong medical and nursing leadership, effective communication and collaboration, and an open approach to conflict resolution and problem solving<sup>36</sup>. Subsequent work with a larger number of units confirmed the importance of organizational culture as well as low nurse turnover. This work also suggested that diagnostic diversity was associated with a worse outcome, thus linking with the work on volume and outcome discussed above<sup>37</sup>.

Another link with the debate on volume and outcome emerges from a study of AIDS units. This revealed high-quality care both in non-specialized units in the magnet hospitals mentioned above and in specialized units situated in non-magnet hospitals; thus, it may not be necessary to establish specialized units to achieve excellent outcomes if the organizational culture is right<sup>38</sup>.

In the UK, the National Confidential Enquiry into Peri-operative Deaths regularly provides evidence that organizational factors can affect clinical outcome, although this is largely by identifying lapses from accepted practice<sup>39</sup>. It has not, of course, been able to examine the question of what might make a hospital perform better than expected.

## THE FUTURE

Hospital services everywhere are changing, and the revolution provides both opportunities and challenges. There is a growing body of evidence that the quality of clinical care can be influenced by organizational structures and cultures but many questions are unanswered. Although low treatment rates seem less likely to achieve the desired results, the points at which further improvements occur for particular interventions are not clear. We need a better understanding of the relationship between volume and outcome. We also need to know much more about the

contribution made by different groups to outcomes of care—especially in view of American findings on the importance of optimum nurse staffing and organization. The lessons of Cohen's 1947 work were never absorbed, so history may repeat itself if nursing is not factored adequately into analyses of patient outcome<sup>40</sup>. The answers are essential if we are to introduce policies that make the quality of care better rather than worse, especially as we face a climate of constraints on health care expenditure in which decision-makers are asking themselves about optimal skill levels. We need a major co-ordinated research programme, drawing on experience in countries facing similar challenges—in particular the USA, which has experienced a substantial change in the structure of the nursing workforce<sup>41</sup>. Until then, our limited resources should not be diverted towards pointless attempts to meet meaningless targets.

*Acknowledgment* This paper is based in part on discussion at a workshop on restructuring hospital workforces held at Bellagio, Italy, with funding from the Rockefeller Foundation and the Baxter Foundation.

## REFERENCES

- 1 Nightingale F. *Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army: Formed Chiefly on the Experience of the Late War*. London: Harrisons, 1858:1
- 2 Cohen IB. Florence Nightingale. *Sci Am* 1984;250:128
- 3 Office of Technology Assessment, US Congress. *Hospital Financing in Seven Countries*. Washington DC: US Government Printing Office, 1995
- 4 Abel-Smith B, Figueras J, Holland W, McKee M, Mossialis E. *Choices in Health Policy: An Agenda for the European Union*. Aldershot/Luxembourg: Dartmouth Press/Office for Official Publications of the European Communities, 1995.
- 5 World Health Organization. *European Health Care Reforms: Analysis of Current Strategies (Summary)*. Copenhagen: WHO, 1996
- 6 Roberts JA. Managing markets. *J Publ Health Med* 1993;15:305–10
- 7 Anonymous. Leap of faith over the data tap. *Lancet* 1995;345:1449–51
- 8 Dickson N. League tables: use for patients. *Qual Health Care* 1995;4:1
- 9 Clinical Outcomes Working Group. *Clinical Outcome Indicators*. Edinburgh: Scottish Office, 1995
- 10 Department of Health. *New Citizen's Charter Commitments for the 21st Century* [press release]. London: Department of Health Press, 19 September 1996
- 11 McKee M, Hunter D. Mortality league tables: do they inform or mislead? *Qual Health Care* 1995;4:5–12
- 12 Clarke A, McKee M. The consultant episode: an unhelpful measure. *BMJ* 1992;305:1307–8
- 13 Cleary R, Beard R, Coles J, Devlin B, Hopkins A, Schumacher D, et al. Comparative hospital databases: value for management and quality. *Qual Health Care* 1994;3:3–10
- 14 Hall C. NHS waiting list figures "might have been fiddled". *Daily Telegraph*, 24 November 1996
- 15 Smith P. Outcome related performance indicators and organisational control in the public sector. *Br J Management* 1993;4:135–51

- 16 Edhouse JA, Wardrope J. Do the national performance tables really indicate the performance of accident and emergency departments? *J Accid Emerg Med* 1996;13:123-6
- 17 Rockall TA, Logan RF, Devlin HB, Northfield TC. Variation in outcome after acute upper gastrointestinal haemorrhage. The National Audit of Acute Upper Gastrointestinal Haemorrhage. *Lancet* 1995;346:346-50
- 18 Davenport RJ, Dennis MS, Warlow CT. Effect of correcting outcome data for case mix: an example from stroke medicine. *BMJ* 1996;312:1503-5
- 19 Green J, Passman LJ, Wintfeld N. Analyzing hospital mortality: the consequences of diversity in patient mix. *JAMA* 1991;265:1849-53
- 20 Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *J R Statist Soc A* 1996;159:385-443
- 21 Green J, Wintfeld N. Report cards on cardiac surgeons: assessing New York State's approach. *N Engl J Med* 1995;332:1229-32
- 22 Hannan EL, Kilburn H, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA* 1994;271:761-6
- 23 Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care: a survey of cardiovascular specialists. *N Engl J Med* 1996;335:251-6
- 24 Anonymous. Dicing with death rates. *Lancet* 1993;341:1183-4
- 25 Black N, Johnston A. Volume and outcome in hospital care: evidence, explanations and implications. *Health Serv Management Res* 1990;3:108-14
- 26 Sowden AJ, Deeks JJ, Sheldon TA. Volume and outcome in coronary artery bypass graft surgery: true association or artefact. *BMJ* 1995;311:151-5
- 27 Working Group on Specialist Training for Hospital Doctors. *Hospital Doctors: Training For the Future (The Calman Report)*. London: Department of Health, 1993
- 28 Coulter A. Why should health services be primary-care led? *J Health Serv Res Policy* 1996;1:122-4
- 29 Millar S. There will always be human error in hospital. *Guardian*, 21 December 1996:3
- 30 Ministry of Health. Department of Health for Scotland, Ministry of Labour and National Service. *Working Party on the Recruitment and Training of Nurses*. London: HMSO, 1947
- 31 Ministry of Health, Department of Health for Scotland, Ministry of Labour and National Service. *Working Party on the Recruitment and Training of Nurses, Minority Report*. London: HMSO, 1948
- 32 Anonymous. Nursing problems and Dr Cohen: Comments on the Ministry W.P. Report. *Nursing Mirror* 1948;87:1-2
- 33 Aiken LH, Smith HL, Lake ET. Lower Medicare mortality among a set of hospitals known for good nursing care. *Med Care* 1994;32:771-87
- 34 Shortell SM, O'Brien JL, Carman JM, Foster RW, Hughes EF, Boerstler H, O'Connor EJ. Assessing the impact of continuous quality improvement/total quality management: concept versus implementation. *Health Serv Res* 1995;30:377-401
- 35 Leveck ML, Jones CB. The nursing practice environment, staff retention, and quality of care. *Res Nurs Health* 1996;19:331-43
- 36 Zimmerman JE, Shortell SM, Rousseau DM, et al. Improving intensive care: observations based on organisational case studies in nine intensive care units: a prospective, multicenter study. *Crit Care Med* 1993;21:1443-51
- 37 Shortell SM, Zimmerman JE, Rousseau DM, Gillies RR, Wagner DP, Draper EA, Knaus WA, Duffy J. The performance of intensive care units: does good management make a difference? *Med Care* 1994;32:508-25
- 38 Aiken LH, Sloane DM. *Effects of Specialization and Client Differentiation on the Status of Nurses: The Case of AIDS*. Philadelphia: Center for Health Services and Policy Research, 1996
- 39 NCEPOD. *The Fifth Report of a Confidential Enquiry into Perioperative Deaths*. London: RCS, 1996
- 40 Rafferty AM. *The Politics of Nursing Knowledge*. London: Routledge, 1996:180
- 41 Aiken LH, Sochalski J, Anderson GF. Downsizing the hospital nursing workforce. *Health Affairs* 1996;14:88-92