Abstract

**Background** Since 1998 BUPA has used the Short-Form 36 (SF-36) questionnaire to monitor changes in health status after adult elective surgery. Over 70 independent hospitals across the United Kingdom have collected data on over 100,000 patient episodes. SF-36 is one of a number of tools that support clinical governance in the sector. Results are reported confidentially, putting the emphasis on supporting a learning culture.

**Formulation of problems apparent at 3 years** Feedback was sub-optimal: discussions with hospital staff and consultants revealed that the league tables were hard to interpret, and there was uncertainty about the definition of outlier results. The number of patients recruited to the survey was variable across the hospitals. No grouping of low-volume procedures met with agreement.

**Action plan for year 4** Use ‘Shewhart’ control charts to distinguish common and special cause variation in results; substitute a condition-specific tool in one instance; adoption of an ‘alert’ process to prompt local audit of unusual results; use of a reminder letter to improve return rate of follow-up questionnaires; and focus programme on a list of 20 common procedures. Discuss these changes with the managerial and clinical leaders of each of BUPA’s hospitals.

**Current position at year 5** The use of Shewhart charts has been welcomed by managers and clinicians at BUPA’s hospitals. The renewed confidence in the programme has encouraged constructive debate into allowing wider access to previously confidential results. Some changes in clinical practice have occurred.

**Keywords:** quality control, clinical governance, treatment outcome, questionnaires

**Background**

Donabedian suggested in the 1960s that healthcare quality should be seen in terms of structure, process and outcome. Since then much has been done in the United Kingdom independent hospital sector, as in the NHS, on structure and process, and on the systems that these produce. Progress on outcomes has been more problematic. One conventional measure of outcome following elective surgery has been perioperative death, but because of its relative infrequency, this has been of little value in helping consultants and hospitals learn from each other’s successes and failures. Much more progress has been made in the independent sector by developing standard pathways of care and then monitoring and comparing variances from these pathways. The main limitation with such hospital-based measures has been that, whilst rich data are available for the hours or days that the patients spend in hospital, much less is known about how the patient progresses post-discharge. Only the patient understands this with any richness, so the use of a tool that assesses outcome from the patient perspective is essential.

In 1998 the British United Provident Association (BUPA) selected the Short-Form 36 (SF-36) as the patient-centred tool that it would use pre-operatively and at 3 months post-operatively. The SF-36 questionnaire was developed by the Medical Outcomes Study in the United States to appraise physical and mental health status. A British English version of this 36-item instrument is available. It is not specific to one specialty but is a generic tool designed for use in clinical practice and research. The validity and reliability of the SF-36 as a measure of health status, as well as its acceptability to patients, has been confirmed. Its use as a measure of health change has also been validated. Published work illustrates its use with hysterectomies, coronary artery bypass grafts and orthopaedic surgery amongst others. The SF-36 questionnaire has a scoring system that provides eight scores for dimensions of health (‘vitality’, ‘bodily pain’, etc.) and two summary scores for mental and physical health.

The aim of introducing the SF-36 programme was to support learning; this has paradoxically been both reinforced and somewhat amended by the change in political climate whereby UK healthcare organizations now have a statutory duty to seek quality improvement through clinical governance. In addition,
the independent sector is required, under the Care Standards Act 2000, to monitor outcomes and use comparative data to improve standards of care.

The process

The first questionnaire is given to the patient for completion prior to admission and is collected by the hospital. A second is mailed out 3 months later with a freepost envelope. Each of BUPA’s 35 hospitals and, more recently, a number of the other ‘network’ hospitals recognized by the insurance division have joined the programme. To date over 70 hospitals have monitored more than 100,000 cases of elective surgery. The standard measure of outcome is change in SF-36 scores. This information is fed back to each hospital every 3 months; consultants also receive regular feedback. In each case their own results are identifiable and are compared with their peers’ anonymized results.

Formulation of problems apparent at 3 years

When the programme was set up the primary aim was to facilitate continuous quality improvement, so it was considered important that results should not name individual patients and that hospital managers should not be able to identify which consultant’s care had led to which feedback. It was hoped that this confidential process would overcome the concerns of consultants and help hospitals and consultants to understand their work in the context of their peers’ results. The outlier results would also be able to identify where audit of other data (e.g. readmissions, complaints, variance from pathway) would be most likely to yield useful information. This vision proved to be a little optimistic.

(1) The use of league tables and histograms to present feedback was not optimal, as they were perceived as difficult to interpret (Figs 1 and 2), and did not naturally lead to the clear identification of priorities for audit. Matrons, consultants, general managers and audit co-ordinators all reported (during regular meetings at which clinical quality was a frequent agenda item) that the programme’s potential contribution to greater health gain was hampered by this. Histograms failed to highlight changes of statistical significance, and league tables were misleading in the identification of outliers. They also both gave a snapshot view and did not permit the identification of trends in performance over time. The number of patients reporting changes >2 standard deviations (SDs) from the relevant procedure mean were also

![Fig. 1 Example of a league table chart initially used to provide SF-36 feedback to hospitals. (Physical change after tonsillectomy.)](image-url)
OUTCOME MONITORING TO FACILITATE CLINICAL GOVERNANCE

Fig. 2 Example of a histogram initially used to provide SF-36 feedback to hospitals and consultants (An orthopaedic surgeon’s report for total hip replacements.)
contrasted with the number expected (5 per cent of volume). It was, however, difficult for the recipients of such data to understand at what point having ‘more outliers than expected’ became a call to action.

Outcomes are influenced by random variation as well as case mix, which can significantly alter a hospital’s position in a league table. Nonetheless, there is an implicit assumption that hospitals located towards the bottom of a league table provide a worse service. This may be inaccurate and if so, is demoralizing for staff. Variation was illustrated by including 95 per cent confidence intervals for the performance of each hospital. This made feedback confusing, however, and did not appropriately differentiate between important or remarkable variation in practice. Using this system, about 1 in 20 hospitals could rightly or wrongly be identified as outliers. To encourage additional audit when performance is, in fact, within the normal range wastes staff time and antagonizes them towards the collection of outcome data. Similarly, rewarding a presence at the top of a league table is unfair if that performance is actually due to chance.

(2) It was evident from many discussions with consultants that they were dubious about judging, or being judged, on the basis of anything other than conventional quantitative, technical measures (e.g. range of movement of a joint in degrees, visual acuity using a Snellen chart).

(3) Hospital staff were frustrated because they wished to be able to identify any consultants associated with poor performance, and so protect future patients and the reputation of their hospital.

(4) The questionnaire return rates were not sufficiently high to allow much reliance to be placed on the output. There was a view that this poor compliance might be a consequence of hospital staff failing to imbue patients with enthusiasm for the programme because of their frustration, or of their finding the output difficult to interpret – both as mentioned above. Records illustrated that 61.2 per cent (SD 9.4) of eligible patients both received and completed the baseline SF-36 questionnaire. There was, however, wide variation (SD 29.2) between hospitals, with more than a quarter of them yielding recruitment rates in excess of 80 per cent.

(5) Data had been collected on over 1000 different procedures, but many of these occurred so infrequently that if they were to be reported for a consultant, or even a hospital, they would show results for just one or two cases, and this was not considered to be helpful. Attempts had been made to group the procedures according to their allocation to Health Resource Groups, so that procedures expected to have similar effects could be reported together. No method of grouping met with complete confidence, however; consultants were concerned that their results were being skewed by mixing together procedures that could have different rates of recovery (e.g. combining primary hip replacement and removal of internal fixation from bone in HRG H02).

**Action plan for year 4**

**Making output more user-friendly**

Key facets of safety improvements are the accurate identification of performance outliers and trends in performance over time. In order to facilitate this, Shewhart charts, also known as control charts, have been introduced as the method for feedback to hospitals. It has been asserted that these charts, which are commonly used in industry to improve the quality of work, are well suited for use in healthcare. Using retrospective data from the excess deaths after paediatric cardiac surgery seen in Bristol and the murders of general practitioner Harold Shipman, Shewhart charts have been shown to provide a simple yet powerful graphical method to identify outstanding practice at an early stage.12

These charts classify variation as having a common cause or special cause and can therefore guide the user to the most appropriate action to effect improvement. They offer an objective, clear method of presenting outcomes data that is easy to interpret and reduces stigmatization of ‘poor performers’, compared with league tables. These charts present data in relation to bands drawn at 3 SDs above and below the mean. Variation between these control lines is described as being due to common causes, whilst variation outside is ascribed to special causes. Experience in engineering suggests that looking into the background of special cause variation is usually quick and easy, whereas investigating common cause variation is often time consuming and fruitless. These lines are equivalent to a confidence interval of 99.8 per cent, which gives only a 1 in 1000 chance of wrongly identifying an outlier compared with a 1 in 40 chance if 2 SDs are used. Separate reports are created for each of the procedures now selected (generally the two or three highest volume therapeutic procedures in each surgical specialty).

The charts we have developed for SF-36 questionnaire feedback to hospitals (Fig. 3) illustrate outcome data from patients treated in the most recent eight quarters, with each quarter separately depicted on the x-axis. The shaded circles represent the mean change in health status of patients treated at the relevant hospital during each of the last eight quarters, for a given procedure. Hospitals whose mean SF-36 score lies outside these ‘audit lines’ are easy to identify. If present, individual patient scores lying outside 3 SDs are also illustrated. The rectangles marked on each quarter represent the inter-quartile range of all hospitals’ scores – i.e. from the 25th to the 75th centile – against which a hospital’s own mean score can be compared. This allows clear identification of trends in performance rather than just providing a snapshot view. Highlighting variation due to special cause enables one to examine clinical practice, to identify the factors behind apparently exceptional or potentially poor performance and hence facilitate clinical governance.

As part of the management of this change, information packs on the new Shewhart charts were sent to all hospitals in July 2002. They have also received a visit from a member of the team involved in this project. This educational programme aims
to improve accurate interpretation of the charts and to emphasize the importance of staff involvement to the success of this programme

Consultants’ concerns
The decision was taken (Spring 2002) to continue with a cautious ‘step by step’ approach towards winning acceptance from consultants, whilst seeking to address their concerns by making detailed changes to the programme. A condition-specific questionnaire (the ‘Visual Function 14’\(^\text{13}\)) was adopted in place of the SF-36 for phakoemulsification, after concerns were expressed by ophthalmologists that the latter tool was unable to reflect fairly the benefits of cataract surgery. Reports were not issued to consultants until at least 10 of their patient episodes had been followed up, and several changes were made to the reporting to try to show where results were significantly different to the normative data.

Staff frustration
A process has been developed that will, whilst observing consultant confidentiality, alert hospital managers if there are indications of serious performance issues, which the hospital will then be asked to investigate and confirm through the many other sources of audit data available locally.

Return rate
The programme was enhanced so that a postal reminder was sent to patients who failed to return the postal questionnaire at 3 months. This improved the net return rate to 75 per cent of those who had completed the initial questionnaire.

Low volume procedures
The focus of the programme was narrowed to 20 common procedures, selected to cover the main surgical specialties, so that reporting of results could then be on the basis of individual procedures, and to ensure that results would be based on robust numbers of cases.

Current position: year 5
Encouraging and acting upon staff feedback has been essential to the success of this programme. Staff feedback since the introduction of Shewhart charts has indicated that they now look on data collection more positively and that perceived value of the programme has risen. This is attributed to the nationwide educational exercise, which ensured that relevant staff were confident of their ability to interpret Shewhart charts. Acknowledging the importance of staff at all levels was paramount. Administrators who send out the first questionnaire, nurses and consultants who can encourage patients to complete and return the forms and matrons and general managers who receive and act on the results all play an important role in ensuring the success of this project.

Two examples illustrate the value of the data to hospitals and consultants. A BUPA hospital was concerned to see low change scores for its patients in one specialty and further investigation

---

Fig. 3 Example of a Shewhart chart providing feedback to a hospital from their completed SF-36 questionnaire data.
(A hospital’s physical summary scores for inguinal hernia repair.)
revealed a run of complaints about post-operative pain relief for one procedure (hysteroscopy). The relevant care pathway was modified and outcomes are now being monitored for evidence of improvement. The second example concerns a consultant who was prompted by his relatively low scores to discuss the matter with the hospital’s head of nursing. It emerged that the consultant gave his patients an unusually optimistic expectation of recovery, and he is now monitoring his outcome results to observe the effect of more deliberate expectation setting.

**Conclusion**

The programme has shown that clinical outcomes can be measured objectively and collected systematically using a comparatively simple process. It has also demonstrated that feedback must be presented in an understandable and user-friendly manner if it is to influence clinicians.

Outcome measurement is not just about the identification of potentially poor performance but should be primarily about learning from and disseminating good practice. Individual hospitals use their outcome results to target audit programmes and focus on areas of practice to investigate. The value is to be found in identifying factors behind any exceptional practice and disseminating the lessons learned from this to improve quality across the whole sector.

**Competing interests**

All authors are employed by BUPA

**Statement of contribution**

A.V.O. introduced and B.M. manages the SF-36 outcomes project. S.C., V.W. and B.M. wrote the paper. V.W. saw the applicability of Shewhart charts in this context. All authors commented on further drafts.

**References**


Accepted 16 January 2004