Establishing Inter-hospital Comparisons of Outcomes

ROBERT CLEARY
CASPE Research, London, UK

This paper describes the ongoing development of a region-wide system for monitoring the short-term outcomes of total knee replacement surgery. The system aims to collect data from a dozen hospitals, and relies on a close collaboration with locally based surgeons and medical audit staff. The intention is to provide routine comparative information based on a broad conception of outcome, that includes both clinical/technical measures and patient-derived assessments of general health status. These data may be reported in the context of relatively detailed case mix information. To date, both data capture and clinical support for the project have been good. Example reports are presented together with a discussion of current limitations and possible future developments.

Key words: Outcomes, joint replacement, comparative data, health status measures.

INTRODUCTION

Total knee replacement (TKR) is widely used to reduce the pain and improve the physical functioning of patients suffering from severe arthritis of the knee. This paper describes the development of a system for monitoring of the outcomes of TKR. The system is being established across 12 units performing TKR within the UK's Northern Regional Health Authority (NRHA). The NRHA serves a population of around 1.2 million over 45 years old and in doing so undertakes around 1000 TKRs per annum. In addition to being a common procedure, knee replacement is relatively costly — the Region estimates an average cost of around £4000 and a typical 19 days length of stay [1].

Our purpose in working with 12 hospitals simultaneously is not simply one of collecting data on large numbers of cases. Rather, our primary interest is to provide a system that gives surgeons at each unit comparative data on their performance. Our interest in comparative data arises from the results of our earlier work in the development of routine outcome monitoring, undertaken at the Freeman Hospital in Newcastle upon Tyne [2,3]. Over the course of a 3-year project at the Freeman, CASPE worked with clinicians across a wide variety of specialities to develop measures and methods that would allow the monitoring of condition-specific outcomes within a single hospital. Across the different specialities and conditions, a consensus developed with respect to the kind of measures required to obtain a reasonably complete picture of the outcome obtained. In general terms, each of the systems attempted to monitor adverse events (e.g. deaths or re-admissions); the amelioration of symptoms; clinical/technical measures (e.g. blood glucose control in diabetes or joint stability in TKR) and some patient-derived measure of general health status. The combined use of clinical and patient-derived measures was felt to be particularly important — giving voice to the patient's view of the outcome, but in a context that may be more readily related to clinical practice. In the case of the surgical interventions monitored in this way at the Freeman (of which TKR was...
one), baseline information was collected pre-operatively from surgeon and patient, with equivalent data being captured at 3 and 12 months post-discharge, allowing longitudinal comparisons to be made for each patient.

The scheme described above proved practical and was successful on a number of occasions in yielding surprising results, some of which lead on to measurable changes in practice. The clinicians' response to the monitoring systems was generally enthusiastic. However, a criticism voiced on more than one occasion was that some of the results were difficult to interpret, and therefore to act upon, in the absence of valid comparative information. While this was felt to be a particular problem in the case of relatively novel measures, such as patient-completed health status questionnaires, there are potential problems with even established clinical measures. While the latter may be featured in the relevant literature, it may be that the cases reported are not truly comparable with the population passing through the hospitals of northern England. The lack of trusted comparators can have a significant effect on the impact of data describing disappointing outcomes — there is often a feeling that "our patients are older/sicker/different than the norm", but there is no way to examine the validity of this as an explanation. While, compared to a monitoring system in a single unit, a system offering inter-unit comparisons could be more effective in informing practice, it is clear that it must be put in the context of solid information on the case mix underlying the results. The system described below attempts to do this.

**METHODS**

**Project management and staffing**

The project, running over the course of 2 years in the first instance, is directed by a steering group chaired by an orthopaedic surgeon who has been active in the development of medical audit within the region. The group includes at least one surgeon from each participating site. Their role has been to develop and refine the dataset which the system collects, agree on the methods of data collection and approve the format of reports from the system. We believe that this broad-based clinical involvement in the direction of the project has been a crucial factor in its early success. The central staffing of the project (amounting to around 1.5 whole time equivalents) covers: the development and maintenance of software for the central database; statistical analysis and reporting; the day to day management of the project, including the induction of new units; and data entry.

**Dataset**

The data set consists of three main components:
- casemix variables including age, sex, the type and extent of arthritis, the identity of any comorbidities, and details of previous orthopaedic surgery. This information is obtained from a pro-forma completed by the surgeon pre-operatively;
- the Knee Society Clinical Rating [4] (the "knee score"). This is a clinical assessment in two halves, one covering pain and technical issues such as the stability, range of motion and alignment of the joint, the other being a functional assessment in terms of walking distance, ability to climb stairs etc. This is again a pro-forma completed by the surgeon, but it is not only completed pre-operatively but also at 3- and 12-month follow-up clinics;
- the Nottingham Health Profile [5] (NHP). The NHP is one of the more widely used, and better validated measures of general health status [6]. It is a patient-completed questionnaire consisting of around 40 statements intended to reflect "problems people may have in their daily life", each requiring a yes/no response [e.g. "I feel I am a burden on others (Y/N)"]. Rather than yielding a single number the NHP provides a six-dimensional measure of health that may be displayed as a profile of scores relating to energy, pain, emotional reaction, sleep, social isolation and mobility. Patients are asked to complete a pre-operative NHP just after their admission, and post-operative NHPs are administered by post, to coincide with the 3- and 12-month clinical follow-ups.

In addition to these three main components, the surgeon was asked to record the type of prosthe-
sis used, and details of any complications on a post-operative form.

Data collection

At each participating unit, the process of data collection had to fit in with the established procedures of a given orthopaedic department. As such procedures may vary in many different ways between units we felt it would be important to draw on some local knowledge in the setup and day-to-day administration of the project on the ground. Clinical input to the process, in terms of form-filling, was already at a relatively high level and it seemed unlikely that surgeons would have time to manage the project locally themselves. Therefore, a unit's participation in the study was made conditional on the availability of support from local medical audit staff. This resource covers the identification of relevant admissions and follow-up appointments, the distribution and retrieval of pro-formae and the forwarding of data to the study's central office for analysis and reporting. The maximum requirement for local audit staff time was estimated at the outset as being 0.2 whole time equivalent — a level felt to be appropriate for only the most active departments. While not yet formally surveyed, it appears that in practice rather less audit staff time is required.

Feedback

Copies of completed data forms are forwarded to the study's central office for entry on computer and subsequent analysis and reporting. Feedback to participating surgeons and audit staff has taken a number of different forms: a comprehensive report on the whole database some 6 months into the project; general and topical presentations to the project steering group; a quarterly newsletter for contributors and others, and presentations to local orthopaedic audit meetings. This last forum has given the opportunity to present a detailed analysis of a unit's casemix and outcomes in the context of regional norms. Wherever comparisons between individual units are presented, these are made on an anonymous basis, with units identified by a code letter known only to their staff and the central team.

RESULTS

Data collection

At the time of writing (15 months into the study), 10 of the 12 units identified at the outset have joined the project. This has been an incremental process with the results obtained from early contributors being used to demonstrate the potential of the system to other units. None of the units so far approached have declined to take part. In general, all the surgeons at a given unit are contributing to the database, but at a couple of sites the take up has not been complete. So far, our success in capturing all relevant admissions has been checked at three units, by comparing the database with the contents of the units' own theatre records. In each case the system has pre-operative data on at least 98% of the TKRs undertaken. The capture of 3-month follow-up has been less reliable — with complete data on around 70% of cases being returned to the central office. However, it seems that the follow-up rate at a given unit tends to improve after the first few months of participation, and a lot of any initial shortfall reflects teething troubles in the local data collection process. The response rate to the postal follow-up of patients with the NHP has remained consistently above 85%. The first few 12-month follow-ups are now starting to return to the central office, but it is too early to judge how complete these data are.

Outcomes

Although the development program is still very much "in progress", it is possible to give an example of the kind of reports we have been presenting to the contributors. Of the 563 replacements currently registered, 3-month follow-up data is available and has been analysed for around 250. In general, data from individual units have been presented in terms of comparisons with either the regional mean or the overall distribution of scores. Figure 1 shows an example of the latter approach, where a scatterplot has been used to compare the 3 month technical outcomes obtained at one unit (unit "X"), with all the equivalent scores obtained from other units. Each point on the plot represents one knee replacement. Its position on the horizontal axis indicates the technical
component of the knee score measured pre-operatively. A low score on this scale indicates a knee that is in bad shape mechanically, and/or painful. As might be expected, the scores cluster in the lower half of the pre-operative scale. The vertical axis plots a replacement's corresponding score as measured at 3-month follow-up. At this point in time, the bulk of scores are towards the upper end of the scale. Thus, in rough terms, the most effective replacements are found in the upper-left quadrant of the figure. Those replacements that have yielded no improvement in the technical knee score lie on the diagonal, and those replacements that have seen, to 3 months, a decline in technical status and/or an increase in knee pain, are shown below the line. It is interesting to note that both pre-operative and follow-up scores are spread quite widely and, for example, that a few of those undergoing TKR have pre-operative scores of 70 or more. While a plot of this kind has a number of limitations, we have found contributing surgeons keen to see their own results in this kind of context — there is a great desire to know how one "measures up".

Figure 2 offers a similar comparison between unit X and the region — this time focusing on the patients' view of the outcome. The histogram plots the mean NHP scores (higher scores indicating poorer health) for the region, as obtained on admission and again at 3 months. It is immediately apparent that the NHP is sensitive to a change in health status associated with the TKR. Substantial improvements are seen across the dimensions, with pain benefitting particularly. To illustrate how the health status of patients at unit X compares with the norm, the histogram has been over-printed with markers indicating the mean NHP obtained for the first 20 cases at that unit. As can be seen, the admission scores are in each case marginally higher than the regional mean. However, on four of the six dimensions the 3 month score falls to the regional level. A wide variety of variables might explain such an effect. However, while controlling for the underlying diagnosis, age, comorbidity, primary replacement vs revision, the pattern of results has persisted. On a number of occasions, and in relation to both clinical and patient-derived outcomes, contributing surgeons have been surprised as to the negligible impact on outcome associated with some quite large differences in casemix.

Figure 3 illustrates one way in which we have attempted to bring together the clinical outcome, represented by the technical knee score, and the patients' report of their health status.
Outcome Monitoring

Against a regional background of the kind used in Figure 2, the graph plots the NHP scores for “patient A”. Patient A is the 80-year-old osteo-arthritis sufferer from unit X, whose knee-score outcome can be seen on the diagonal of Figure 1 (admission score 58, 3-month score 58). This disappointing technical outcome was something of an outlier for unit X, and the NHP scores are not inconsistent with it: little improvement was seen in pain or mobility, and there was a real worsening on the other dimensions. Of course, this being a general measure it may be contaminated with a change in health unassociated with the TKR [7]. For this reason, the consideration of aggregate NHPs — within groups of patients — should not be overlooked. However, the discussion of NHPs relating to such individual knee-score outliers (often providing positive news absent from the technical outcome) has been useful in developing a feeling amongst the contributors that such patient-derived measures are a valuable component of the system.

DISCUSSION

In terms of case numbers, these are early days for a system that aims to enable contributors to make meaningful, casemix-controlled comparisons between the various units providing TKR. However, we believe we have developed a system for data collection that, in mixing local knowledge with central coordination, has good chances of being successful over the long haul. Considerable work remains in terms of evaluating and refining the process of data collection. For example, we know little about the reliability of the clinical measures and so we are now beginning to address the question of whether different surgeons complete the knee score in comparable ways. Additionally, we would like to try and reduce the amount of data collection currently required of the surgeons. There are of course routine information systems within the NHS that hold much of the casemix information for which our system currently demands clinical time. However, such routine information systems have in the past found to be lacking [8], and it certainly is the case that several of our contributors have little faith in the information stored in their hospitals' databases. We now have the opportunity to validate these contents retrospectively, using our database. Where the results are favourable we will have the option of using the hospital system as a source.

One limitation of the current system is its short-term nature. Orthopaedic surgeons have been monitoring their outcomes for many years, but have tended to restrict the definition of “outcome” to the long-term (10-year rather than 12-month) survival of the prosthesis. There is of course a role for a database such as ours in providing a detailed cohort for the later analysis of joint survival. That said, we find a recognition amongst our contributors that rich information about the short-term outcome is in itself of value: some TKRs go wrong fairly quickly or never achieve the intended level of performance; different groups of patients fare differently over the initial months, and the surgeons are keen to identify those patients who will “do well”.

Above all, the success of the system must be judged in terms of the use to which the data are put. Although the development of a “data collection system” can be quite challenging, it is ultimately pointless if it is not part of a system of feedback that is designed to inform decision-making. (Such a system does not exist in a vacuum, and therefore does not have to provide all the answers. While our own database contains little process information it is able to provide outcome information that may raise questions best addressed by the local consideration, or indeed audit, of practice at a given unit.) Amongst many possible developments of the system — extension to hip replacements; inclusion of patient satisfaction and/or cost data; comparisons with US databases [9] — the most beneficial might be a widening of the audience of the data we collect. Both purchasers and patients could be said to have a stake in the information we collect, and it is not implausible that as clinical confidence in this kind of outcome monitoring grows, it will have a role beyond medical audit.

Acknowledgements: The financial support of Northern Regional Health Authority, and the active cooperation of orthopaedic surgeons and medical audit staff around Northern Region are acknowledged.
REFERENCES


