AN NOTATION

A critique of revision rate as an outcome measure

RE-INTERPRETATION OF KNEE JOINT REGISTRY DATA

National registers compare implants by their revision rates, but the validity of the method has never been assessed. The New Zealand Joint Registry publishes clinical outcomes (Oxford knee scores, OKS) alongside revision rates, allowing comparison of the two measurements. In the two types of knee replacement, unicompartimental (UKR) had a better knee score than total replacement (TKR), but the revision rate of the former was nearly three times higher than that of the latter. This was because the sensitivity of the revision rate to clinical failure was different for the two implants. For example, of knees with a very poor outcome (OKS < 20 points), only about 12% of TKRs were revised compared with about 63% of UKRs with similar scores.

Revision therefore is not an objective measurement and should not be used to compare these two types of implant. Furthermore, revision is much less sensitive than the OKS to clinical failure in both types and therefore exaggerates the success of knee replacements, particularly of TKR.

Since their introduction in Sweden, in 1975, the main aim of joint replacement registers has been to compare the performance of competing designs of implant in order to give early warning of those which underperform.1,2 To this end, they have used ‘revision’ as the criterion of failure and the ‘revision rate’ as the measure of the outcome performance of each implant. From a practical point of view, revision has the main advantage that it is the one event which consistently brings the patient back to the hospital and can be registered precisely.2 However, to our knowledge, the usefulness of the method has never been compared with that of other outcome measures in terms of objectivity and sensitivity.

For the past ten years, the New Zealand Joint Registry has reported the performance of knee replacements not only by their revision rates, but also by a validated patient-reported outcome measure, the Oxford knee score (OKS).3-5 The data presented in that register’s comparison of total and unicompartimental replacements allow some relevant questions to be explored.

Comparison of total and unicompartimental knee replacements

In 2009, the New Zealand registry3 reported that the revision rate for primary unicompartimental knee replacement (UKR) (1.51% per year) was nearly three times higher than that for primary total knee replacement (TKR) (0.54% per year). The difference was statistically significant and in accordance with contemporary reports from registers in other countries such as Sweden,6 Australia7 and England and Wales.8

The report also noticed a close correlation, in both types of implant, between the knee scores measured six months after surgery and the risk of revision in the ensuing two years.3,9 This was demonstrated by two graphs, which are shown superimposed in Figure 1 on the two different axes used in the register’s report. As the knee score decreased, the revision rate increased, in a similar pattern for both implants. A UKR scoring less than 20 points had about 40 times the risk of revision of one scoring between 36 and 40 points, and a TKR which scored less than 20 points had a risk of revision of about 30 times that of one scoring between 36 and 40 points.3

The nearly three times higher mean revision rate of UKR might then imply that UKR had a similarly higher proportion of poor results compared with a TKR. This, however, was not the case, and the opposite was true (Fig. 2). The UKR had a better mean knee score (UKR 38.81, TKR 37.01) and a higher proportion of excellent and good results (UKR 79%, TKR 72%). Furthermore, the TKR had a proportion of poor results which
was 1.6 times higher than that of UKR (UKR 7.7%, TKR 12.3%) and more of the worst scores (0 to 1 points) for all the 12 questions in the OKS questionnaire. Figure 3 suggests an explanation. The data are the same as in Figure 1, but with the graphs plotted on a common axis. It can be used to compare the revision rates of UKR and TKR for any given knee score. Its most obvious feature is that, at every level of outcome, a UKR was much more likely to be revised than a TKR with the same score.

Putting these observations together, it appears that the revision rates of TKR and UKR both varied, inversely, with the knee score (Fig. 1), but the difference between their revision rates was substantially independent of the knee score. In other words, in every category of outcome, a UKR was between four and six times more likely to be revised than a TKR with the same outcome (Fig. 3). This difference in the susceptibility to revision of the two implants can account entirely for the difference between their observed mean revision rates and demonstrates how the implant with the better clinical results could nevertheless be the one to be revised more frequently.

**Discussion**

**Lack of objectivity of revision rate.** The reason for the difference in their susceptibility to revision is probably that surgeons are less willing to advise re-operation of a total than a unicompartmental knee replacement because it is a more difficult and a more hazardous procedure. UKR is a much less invasive primary intervention, and can usually be converted to TKR with standard primary components. The replacement of an unsuccessful TKR with another TKR is a major reconstruction of the joint, often requiring special revision implants with wedges and intramedullary stems. Furthermore, conversion of a knee with a painful outcome from a UKR to a TKR, even in the absence of positive evidence of failure of the implant, can still be logically proposed as the next therapeutic step, from a partial to a definitive solution. The rationale for removing a TKR which is painful for no obvious reason, and then implanting another, is not so persuasive. The example of persistent pain is appropriate
because, in New Zealand, it was the most common cause for revision, in both types of implant, and more common in UKR compared with TKR. Attributing the difference between their susceptibilities to revision to the unwillingness of surgeons to re-operate on a TKR, fits well with the data, and we can think of no other plausible explanation. Whatever the reason, it is evident that in the above comparison the revision rate is not an objective measurement of the performance of an implant.

There may be a similar potential for bias whenever revision rates are used to compare other conservative surgical solutions, such as patellofemoral replacement or hip resurfacing, with ‘definitive’ total joint replacement. In addition, the even more invasive prostheses, often used when revising a TKR, may be even less likely to be re-revised no matter how unsatisfactory the clinical outcome.

It is also possible that the decline in the revision rate with increasing patient age, reported by registers \(^3\)\(^\text{6-8}\) for many designs of joint replacement including TKR and UKR, and usually interpreted as a consequence of their less stressful usage in the elderly, might be due instead to the increasing unwillingness of surgeons to advise, and ageing patients to submit to, a further operation, independently of the quality of outcome of the primary implant. This explanation is supported by the lack of any correlation between patient age and the degree of satisfaction reported after TKR.\(^1\)

Studies based on joint registry data have shown that the revision rates of surgeons who perform large numbers of UKRs are much lower than those of surgeons who perform few,\(^3\)\(^\text{14,15}\) a correlation not reported for TKR.\(^3\) If revision rate was an objective measurement of outcome, these differences could be assumed to represent the better clinical results achieved by more experienced practitioners, and a greater dependence of UKR, than TKR, on surgical expertise. However, they could also mean that surgeons with greater experience of UKR are less easily persuaded than beginners to undertake a conversion to TKR, since it may be unnecessary or ineffective. For instance, experience with UKR teaches that pain often improves spontaneously in the first post-operative year, and sometimes as late as the second, particularly if there is no mechanical explanation for it, and that in the absence of such an explanation conversion to a TKR does not reliably relieve the pain.\(^1\)\(^\text{6}\) Yet, the New Zealand Joint Registry reported that more than half of the UKR revisions for pain were performed in the first two post-operative years.\(^3\)

These questions, and others, about the interpretation of joint registry data will continue to remain unanswered for as long as the revision rate is used as the sole measurement of outcome.

**In sensibility of the revision rate.** Even if implant-dependent bias is avoided, as when models in the same class (TKR or UKR) are compared with one another, the revision rate may still be a poor outcome measure, not because of lack of objectivity, but because of the insensitivity of revision to clinical failure.

It has been reported that the mean pre-operative OKS of knees about to undergo TKR is 18 points.\(^5\) Patients with a post-operative knee score of \(\leq 18\) points therefore have had little or no benefit from surgery. The joint registry did not report this exact group, but Figure 3 shows that, of the knees which scored \(< 20\) points on the OKS, 89% of the TKRs and 37% of the UKRs were subsequently not revised. The sensitivity of the revision rate in identifying joint replacements which gave little or no benefit was therefore only about 10% for TKR, and only 60% for UKR. In statistical tables, based on revision rates, a high proportion of the joints recorded as ‘unrevised survivals’ must therefore be more properly regarded as failures.

**Critical comments on these interpretations.** An objection to our conclusions is that they depend on the measurements of the OKS taken six months after surgery, which may be too soon to represent definitive outcome scores. However, OKS questionnaires completed five years after surgery revealed ‘little significant change’ in the scores of either type of implant, and the joint registry report states that, for both TKR and UKR, the score at six months is indicative of the medium-term outcome.\(^3\)

The cut-off date for revisions (two years after the six-month questionnaire or 2.5 years after surgery) may also be thought to be too early. However, over the ten years covered by the registry, the mean times to revision were 2.3 years for TKR and 2.4 years for UKR. The peak time for all the main causes of revision in both types of implant was two years, with the exception of loosening of the tibial component in TKR, which peaked at three years.

The reasons for revision were not the same for the two types of implant. However, the most common reason for revision in both was pain, and it was a more common cause in UKR than in TKR. Loosening of the component was also a common cause of revision in both. The revision profiles of the two implants differed most in the higher proportions of deep infections and patellar complications in TKR.

One of the important determinants of outcome, the pre-operative status of the knee, was not measured by the New Zealand Joint Registry, therefore the better post-operative knee scores of UKR may reflect a similar difference between the series pre-operatively. Furthermore, a survey by the National Joint Registry for England and Wales,\(^1\)\(^\text{7}\) using the OKS questionnaire, found that patients were less likely to be satisfied with their results after UKR than after TKR, although the report warned that further statistical analysis of this observation was required. These provisos, however, seem to have little bearing on our use of the post-operative data in the joint registry to test the two outcome measures, one against the other, in the same series of patients.

**Conclusions.** Using a validated patient-reported outcome score as the yardstick, our conclusions are as follows:

1) The higher revision rate of UKR was not because its results were worse than those of TKR. It seems likely that the former was more often revised because it was easier and
A CRITIQUE OF REVISION RATE AS AN OUTCOME MEASURE

safer to revise than the latter. This lack of objectivity makes the revision rate a misleading outcome measure in this comparison, and perhaps in others.

2) Even in comparisons of implants which are equally readily revised (e.g., TKR with TKR), the low sensitivity of the revision rate to poor results limits its usefulness as an outcome measure.

3) Since most joint replacement registers have used the revision rate as the sole measure of outcome, some of their conclusions need to be reconsidered.

References


