Continuous monitoring of the performance of hip prostheses

S. L. Hardoon, J. D. Lewsey, P. J. Gregg, B. C. Reeves, J. H. P. van der Meulen

From The Clinical Effectiveness Unit of The Royal College of Surgeons of England, London, England and London School of Hygiene and Tropical Medicine, London, England

New brands of joint prosthesis are released for general implantation with limited evidence of their long-term performance in patients. The CUSUM continuous monitoring method is a statistical testing procedure which could be used to provide prospective evaluation of brands as soon as implantation in patients begins and give early warning of poor performance. We describe the CUSUM and illustrate the potential value of this monitoring tool by applying it retrospectively to the 3M Capital Hip experience.

The results show that if the clinical data and methodology had been available, the CUSUM would have given an alert to the underperformance of this prosthesis almost four years before the issue of a Hazard Notice by the Medical Devices Agency. This indicates that the CUSUM can be a valuable tool in monitoring joint prostheses, subject to timely and complete collection of data. Regional or national joint registries provide an opportunity for future centralised, continuous monitoring of all hip and knee prostheses using these techniques.

Continuous monitoring methods are defined as statistical testing procedures which are used for prospective monitoring of an intervention, once the intervention is in use. They help to identify unacceptable and poor performance of the intervention as early as possible. The methods have been used widely for many years in manufacturing and industry, but less extensively in medicine. Rogers et al.[1] have given an overview of continuous monitoring methods applied to cardiac surgical performance. Other applications include monitoring annual death rates in a general practitioners’ practice,[2] 30-day mortality after heart transplantation[3] and after myocardial infarction[4] and patient flow into hospital.[5]

There are many different types of prosthesis available for hip replacement. According to the National Joint Registry for England and Wales[6] in 2004, 88 different acetabular components and 101 femoral components were used. All hip replacement prostheses used in the United Kingdom are required by law to be ‘CE marked’. This indicates that a prosthesis complies with European standards as detailed in the Medical Devices Directive.[7] Manufacturers are required to provide evidence that essential aspects of safety and performance are met (see Annexe I of the Medical Devices Directive).[7] However, a manufacturer can place the CE mark on a joint replacement implant before the medium- and long-term clinical performance of that implant is known.[8] This highlights a need for continuous evaluation of new prostheses from the time of introduction into clinical practice and a mechanism to identify poorly performing components as soon as possible.

Our aim was to demonstrate how a continuous monitoring method could be used to evaluate hip prostheses prospectively in terms of revision rate and give early warning of poor performance. We illustrate the potential value of continuous monitoring by applying a particular method to data on a prosthesis known to have performed poorly.

The CUSUM continuous monitoring method

There are a variety of different types of statistical testing procedure which fall under the umbrella term of continuous monitoring methods. We focus on the CUmulative SUM (CUSUM) continuous monitoring method as described by Grigg et al[12] and Steiner, Cook and Farewell.[9] It is important to note that there is some overlap in terminology and the term ‘CUSUM’ encompasses several different continuous monitoring methods.

The specific outcome of interest for a prosthesis under surveillance is its rate of revision. A prosthesis will be said to be ‘performing
poorly’ if the revision rate is unacceptably higher than a chosen target rate.

Monitoring of a prosthesis begins as soon as the first implantation has taken place. A cumulative statistic is updated at regular intervals as more and more outcome data are collected. If the cumulative statistic falls below zero, it is reset to zero. Results can be displayed in a chart representing the cumulative statistic plotted against time. As soon as the cumulative statistic exceeds a pre-specified boundary level an alert is issued that the performance level is unacceptable.

The resetting to zero is a feature particular to this CUSUM monitoring technique. It corresponds to the use of the CUSUM to test only whether performance is unacceptable and not to ascertain whether target performance levels have been met. An implication of the resetting process is that it is possible for poorly performing prostheses to be identified more quickly by the CUSUM than by alternative monitoring methods which test for both poor performance and target performance.

The cumulative statistic is computed from the observed number of revisions in the interval, the expected number of revisions according to the target revision rate, and the rate above which the revision rate is considered to be unacceptable. The Poisson distribution is used to calculate the expected number of revisions in each interval, given the target revision rate and the total follow-up time of all patients in that interval, with the prosthesis under surveillance. Patients who die and have not had a revision are censored at date of death. In other words, their follow-up time is terminated prematurely.

Even if revisions occur at a rate consistent with the target rate, it is possible for alerts to be issued (cumulative statistic crosses the boundary level) due to chance alone. The boundary level is set to keep the probability of this occurring within a certain time to a minimum, while at the same time keeping the probability of signalling an alert high, if revisions occur at an unacceptable rate. Computer simulations are used to estimate ‘false’ and ‘true’ detection probabilities for different choices of boundary level.

An example: Application to the 3M Capital Hip System

Background. The 3M Capital Hip System hip prosthesis (3M Health Care Ltd, Loughborough, United Kingdom) was first implanted in patients in 1990 and marketed in the United Kingdom from 1991 onwards. In 1995, at the British Orthopaedic Association Annual Conference, Massoud et al. spoke of their concerns about the performance of the femoral component of the 3M Capital Hip prosthesis, prompting the Medical Devices Agency (MDA; now the Medicines and Healthcare Products Regulatory Agency) to commence an investigation in the summer of 1996. In 1998 the MDA issued a Hazard Notice advising that all patients with a 3M Capital Hip prosthesis should be recalled for review, in light of reports of the poor short-term performance of the femoral component of this prosthesis. By this time, 4688 3M Capital Hips had been supplied in the United Kingdom. A detailed account of the official investigation carried out in 2001 into the performance of the 3M Capital Hip system can be found elsewhere.11

Description of the dataset. We made use of the dataset from the official 3M Capital Hip System investigation which aimed to include all patients in the United Kingdom who had received this prosthesis from the time that the system was introduced into clinical practice. Of the 4688 3M Capital Hip Systems which were supplied in the United Kingdom, 3688 (78.7%) were included in the database. The remainder were untraceable (706, 15.1%), unused (35, 0.7%), or the patient refused consent or did not respond (259, 5.5%). This number of prostheses corresponded to 3494 patients since some patients had bilateral implants and others had both primary and revision 3M Capital Hip prostheses.

The hip replacements were implanted between 19 January 1990 and 6 November 1997. Of the 3651 hips in the database with a known implant date, 3364 (92%) were implanted in the first six years. It is likely that the Hazard Notice issued on 19 February 1998 would have influenced decisions on whether revisions were necessary. For this reason, we only used follow-up information until this date and not beyond.

Of the 3688 prostheses, 3440 (93%) in 3268 patients were for primary replacement procedures. These are the focus of the analyses. Of these 3268 patients, 2124 (65%) were women and the mean age at implantation was 71 years (20 to 94, SD 9.9). Of the 3440 primary implants, primary osteoarthritis was the main indication for hip replacement for 2672 (78%).

The 3M Capital Hip System had four different types of femoral component: modular flanged, modular round-back, monobloc flanged and monobloc round-back. The modular flanged stem was used in 1211 (35%) implants, the modular round-back in 551 (16%), the monobloc flanged in 585 (17%), and the monobloc round-back in 258 (8%). The type of femoral component used for the remaining 835 (24%) primary implants was unknown.

Survival from revision. Figure 1 shows the Kaplan-Meier survival curves for the different femoral components. The percentage of prostheses not revised (95% confidence interval) at three years was 93% (91 to 94) for the modular flanged stem, 95% (93 to 96) for the modular round-back stem, 97% (95 to 98) for the monobloc flanged stem, and 98% (94 to 99) for the monobloc round-back stem.

Application of the CUSUM. We applied the CUSUM retrospectively to the data. We calculated the cumulative statistic at intervals of three months, after the introduction of the 3M Capital Hip System into the market. The target revision rate was chosen to be consistent with the current National Institute of Clinical Excellence benchmark (first proposed in April 2000) of a rate of 10% or less at ten years post-operatively.12 The revision rate considered to be unacceptable was chosen as 1.5 times the target rate or above. An
alert boundary level of 3 was used. It was found from computer simulations that setting a boundary level of 3 corresponded to a probability of 0.17 of a ‘false alert’ (CUSUM alerts but revision rate is at the target level) occurring within eight years of follow-up and a probability of 0.99 of a ‘true alert’ (CUSUM alerts when the revision rate is 1.5 times higher than the target level) within eight years of follow-up.

Separate CUSUM charts for each type of femoral component, and then for all components combined are shown in Figure 2. Using data on all femoral components combined, an alert to possible unacceptable performance would have been given on 19 July 1994. This is four years and six months after the introduction of the prosthesis in January 1990 and three years seven months before the issue of the Hazard Notice.

It can be seen in Figure 2 that, if each femoral component had been monitored separately, an alert would have triggered first for the modular flanged stem (19 October 1993). This is nine months before the alert from the CUSUM for all components combined. Alerts would have followed, assuming the continuation of monitoring, for the modular round-back stem (19 July 1996) and the monobloc flanged stem (19 October 1997). An alert would not have been given for the monobloc round-back stem before the time of the Hazard Notice.

**Discussion**

The CUSUM continuous monitoring method can be tailored to monitor the performance of a hip prosthesis in terms of the rate of revision after specification of certain parameters including target and unacceptable revision rates. Use of the CUSUM can provide much needed prospective evaluation of prostheses from the time of introduction onto the market and could give an early warning if a prosthesis is performing at an unacceptable level. Indeed, retrospective application to the 3M Capital Hip System showed that had the method been available, the CUSUM method would have given strong evidence of the underperformance of this implant at an early stage. The CUSUM alert for all femoral components combined in mid-July 1994 would have supported the suspicions of surgeons at that time, regarding the poor performance of this brand of prosthesis. Over 40% (1569) of the 3651 3M Capital Hip prostheses in our dataset with known implant dates were implanted after this time and would therefore have been avoided. By January 2000, 218 (14%) of these subsequent implants had been revised.

**Methodological limitations.** The times to alert are based on the strong assumption of complete and immediately available data on revisions. This is unlikely to occur in practice and therefore alert times would have been slightly later than those found in our study. Furthermore, in estimating the number of implants prevented it was assumed that there was no time delay between issuing an alert and withdrawing the hip system from use. This would not have been the case. Any alert issued should not be taken as absolute evidence of unacceptable performance because of the possibility of a ‘false alert’. Therefore an investigation would have first been instigated to eliminate other possible reasons for

---

**Fig. 1**

Kaplan-Meier survival curves for the femoral components of the 3M Capital Hip System.
CUSUM charts for a) modular flanged stem, b) modular round-back stem, c) monobloc flanged stem, d) monobloc round-back stem and e) all femoral components combined. Used prospectively (as intended), the CUSUM monitoring would stop as soon as an alert is issued (i.e. as soon as the boundary line is crossed). The graph is continued beyond this point (dotted section) only because the CUSUM method has been applied here in a retrospective context (see text for further explanation).
the alert, such as a statistical ‘false alert’, data errors or case-mix severity. Nevertheless, even allowing a year for such checks to be completed, the alert would still have led to early withdrawal of the 3M prosthesis (i.e. in July 1995, almost three years before the 1998 Hazard Notice).

Comparisons with other studies. This is the first study to consider the application of continuous monitoring to evaluate revision rates of hip prostheses. In a recent study by Nizard et al., an alternative version of the CUSUM monitoring method was used to evaluate a CT-based navigation system for total knee replacement. In that report the aspect of interest was surgical technique rather than the prosthesis, and the (immediate) outcomes were alignment of the lower limb, positioning of the prosthesis and operating time. In agreement with our findings, the authors concluded that the CUSUM was a useful tool for assessing new surgical developments. One difference between our approach and the CUSUM used by Nizard et al. was that we updated the CUSUM at regular time intervals, while their CUSUM was updated on a person-by-person basis. When the outcomes are far in the future, as is the case for revision but not for the outcomes examined in their study, a CUSUM updating at time intervals, assuming a Poisson model, is more appropriate.

A report on the performance of the 3M Capital Hip system was published in 2001. It is important to note that our study is not a re-investigation but provides an alternative analysis to that found in the report. Here we evaluated a statistical monitoring technique designed to provide ongoing real-time analysis while the prosthesis was in use. Future research. There are further issues which need to be addressed. First, the choice of statistical model to calculate the expected outcome (numbers of revisions) is important. A key drawback of the Poisson approach used here is that it assumes that the revision rate is constant over time from implantation. Survival curves of prostheses suggest that this is not the case. A specific issue is therefore whether more complex alternative statistical models need to be used to allow for a varying revision rate.

A second issue is whether approaches which allow some adjustment for ‘case-mix’ should be adopted or whether unadjusted approaches with post-alert case-mix severity checks are sufficient. Risk-adjusted methods allow adjustment of the expected number of revisions in any given interval depending on the specific characteristics of the patients. The particular CUSUM method which we have employed may be adapted to adjust for case-mix, provided that an adequate prognostic risk model can be constructed.

Finally, the question remains whether separate monitoring should take place for different elements of the system (e.g. prosthesis type/comination, surgical technique, hospital). Our study showed that individual CUSUMs for the different femoral components would have had a different interpretation compared with the overall combined CUSUM, because the femoral components had different revision rates.

Implications. Large-scale regional or national joint registries, of which the National Joint Registry for England and Wales and the Swedish National Hip Arthroplasty Register are examples, offer a platform for centralised, prospective continuous monitoring in the future of all hip and knee systems used in clinical practice. This puts high demands on the quality of the data in these registries and especially on the ability to link primary replacements with subsequent revisions.

This project was supported financially by the National Joint Registry for England and Wales. Jan H. P. van der Meulen received an NHS Public Health Career Scientist Award.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

References