NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

Total hip replacement and hip resurfacing arthroplasty for the treatment of pain and disability resulting from arthritis of the hip for which non-surgical management has failed

DePuy International Ltd

Multiple Technology Appraisal (MTA)

Submission date: 1st February 2013

Executive Summary

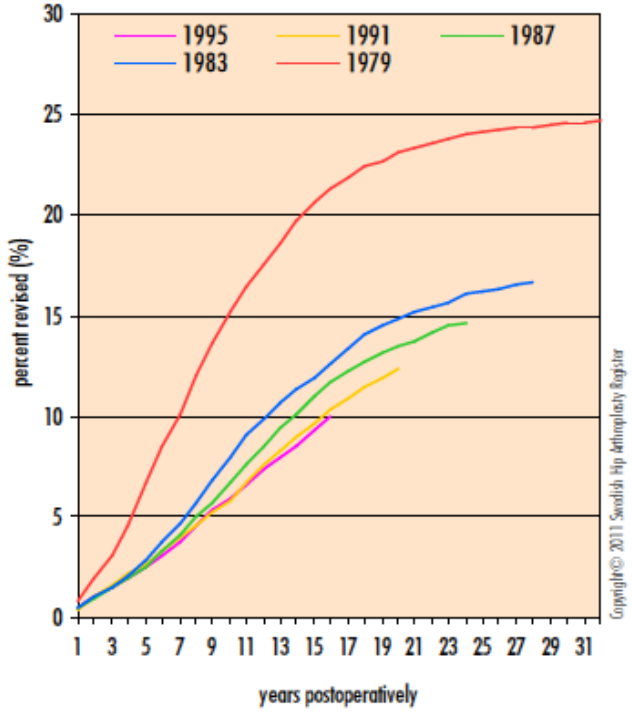
Hip replacement surgery is a well-established and cost-effective procedure for patients with arthritis of the hip for whom non-surgical management has failed

Pain and disability resulting from arthritis of the hip most commonly occurs due to osteoarthritis (OA) and rheumatoid arthritis (RA). It has been estimated that more than 0.75 million people in the UK have painful arthritis of the hip due to OA or RA (1-3). Pain and the resulting loss of function can confer a substantial burden to patients, leading to a decline in quality of life (QoL) that is greater than observed with other chronic conditions such as diabetes, angina and heart disease (4). There are a variety of pharmacological and non-pharmacological treatments for arthritis of the hip, collectively termed non-surgical management (NSM), which aim to control pain and allow patients to retain function and QoL. If patients reach a point where NSM has failed, the most effective approach to reduce pain and restore function is surgery to replace the damaged joint.

Hip replacement surgery is a well-established procedure that has been successfully performed for decades (5). The mortality risk associated with surgery is low (approximately 0.5% in the 30 days following surgery (6)) and recovery is rapid, with patients able to resume normal activities within 6-12 weeks (7). Substantial improvements in QoL following hip replacement surgery have been reported through the collection of patient reported outcome measures (PROMs) relating to routine NHS procedures; patients in England that underwent hip replacement procedures in 2010-2011 reported an average post-operative EQ-5D score of 0.762, compared with an average preoperative score of 0.357 (8). Hip replacement procedures were associated with the highest health gains following surgery of all the procedures covered by the National PROMs Programme (also including knee replacements, hernia and varicose vein procedures) (see Appendix G). Hip replacement is considered a highly cost-effective intervention and total hip replacement has been reported to be more cost-effective than life-extending interventions such as coronary artery bypass surgery and renal dialysis (9).

The benefits of hip replacement are maintained over the long-term. The NICE scope for this appraisal (issued in December 2012) states that “currently, artificial hip joints last on average for 10 to 15 years”, however this may be a substantial underestimation and is not aligned with the existing NICE guidance (10). In the Swedish Hip Arthroplasty Register, which extends follow-up to the third decade post-surgery, cumulative percent revision rates of less than 25% at 32 years are reported even for much older generations of implants and surgical practice. Revision rates for the most recent cohort of patients (those that received hip replacements in 1995) are on track to be well below 20% at 32 years post-procedure (see Figure 1) (11).

Figure 1: Cumulative frequency of revision - all diagnoses and reasons (11)



The purpose of the NICE MTA is to appraise the clinical and cost effectiveness of THR and surface replacement within their CE marked indications for the treatment of pain or disability resulting from end stage arthritis of the hip and therefore to review the guidance provided in TA2 and TA44. The scope was subsequently updated by NICE on 19th December 2012 to review the population of people with pain or disability resulting from arthritis of the hip for which non-surgical management has failed, and to investigate the following comparators:

1. Different types of primary total hip replacement and hip resurfacing arthroplasty compared with each other for people in whom both procedures are suitable.
2. Different types of primary total hip replacement compared with each other for people in whom hip resurfacing arthroplasty is not suitable.

Total hip replacement (THR) allows a range of fixation methods and articulating surface combinations to satisfy individual patient characteristics and surgeon choice

The two broad classes of hip replacement surgery currently available are hip resurfacing arthroplasty and total hip replacement (THR). Hip resurfacing arthroplasty involves replacement of the surface of the head of the femur with a metal cap and a monobloc metal cup which fits into the acetabulum. The number of hip resurfacing arthroplasty procedures performed annually in England and Wales has fallen dramatically in recent years. In 2011, hip resurfacing arthroplasty accounted for just 2% of all primary hip procedures (12), down from 10% of all procedures performed in 2004 (13). Increasingly, hip resurfacing arthroplasty is only considered suitable for use in a subgroup of patients; typically active males under 55 years.

THR involves the removal of the head and neck of the femur and the insertion of an artificial femoral stem and head along with a cup component which fits into the acetabulum of the pelvis. There are currently four classes of THR used in England and Wales:

- Cemented prostheses are held in place with bone cement and typically consist of three components; a femoral stem with a modular femoral head which articulates with an acetabular cup. The femoral head can be either metal or ceramic and the acetabular cup is typically polyethylene. Cemented prostheses are often used in older patients who are less active and may be less likely to receive revision surgery. In 2011, 36% of primary THR procedures that took place in England and Wales and recorded by the NJR were cemented (Table 2.5 from the NJR 9th Annual Report) and the average age of patients receiving a cemented prosthesis was 72.8 years (12).
- Cementless prostheses are implanted without the need for bone cement and rely on initial press-fit fixation followed by natural bone growth over time to secure the prosthesis in place. Cementless prostheses typically consist of four components; a femoral stem, femoral head, acetabular cup shell and acetabular liner. The femoral head can be either metal or ceramic and the acetabular liner is typically ceramic or polyethylene, therefore allowing a greater combination of bearings than exist with cemented polyethylene prostheses. Hard bearing surfaces, such as ceramics, are considered particularly suitable for young and active patients who require substantial wear resistance from their prosthesis. An additional advantage of cementless prostheses is the thinness of the acetabular liners, which allows the use of larger femoral heads (which have been demonstrated to have lower post-operative dislocation rates (14-19)). In recent years, cementless THR has overtaken cemented THR as the most commonly used primary procedure in England and Wales; in 2011, 44% of primary hip replacement procedures were cementless compared with 36% cemented procedures (Table 2.5 from the NJR 9th Annual Report) (38). Other countries have experienced a similar trend towards a greater use of cementless THR; for example, the Australian Orthopaedic Association reported an increase in the use of cementless fixation from 51.3% in 2003 to 63.5% in 2011 (20) and the Swedish Hip Arthroplasty Register has also reported an increase in the use of cementless THR between 2000 and 2010 (11). Patients that receive cementless THR are typically younger than patients receiving cemented THR, reflecting a greater demand on the wear rate and longevity of the prosthesis. The average age of patients that received cementless prostheses in 2011 was 65.4 years (38).
- Hybrid THR consists of a cemented femoral stem and a cementless acetabular cup.
- Reverse hybrid uses a cementless femoral stem and a cemented acetabular cup. In 2011, 18% of all primary hip procedures were hybrid or reverse hybrid (Table 2.5 from the NJR 9th Annual Report). Of these, 85% were hybrid and 15% were reverse hybrid.

The different classes of prostheses and bearing surface combinations available (see Table 1) reflect individual patient requirements and characteristics and the continual need for improved outcomes following hip replacement surgery. Surgeons make choices based on their surgical training, the quality of clinical evidence, a class of prosthesis and a technology that they have experience with as well as the characteristics of the patient such as age, activity levels and body mass index (BMI). A large range of options available to the surgeon allows the surgeon to match the proven technology to the patient's individual requirements, therefore increasing the likelihood of the best possible outcome for the patient.

Table 1: Summary of bearing surface combinations

Articulation Type	Characteristics	Benefits

CoC	• Hardest bearing surface material. • The most expensive bearing combination.	• Can be used with large femoral heads. • Associated with the lowest rate of wear and thus the lowest revision rate in more active patients.
CoP	• Combines a hard ceramic femoral head with a polyethylene cup.	• Lower wear rate compared with MoP bearings. • Generally cheaper than CoC.
MoP	• Most commonly used bearing surface combination. • Associated with osteolysis due to polyethylene wear, but the use of cross-linked polyethylene has improved wear rates.	• The cheapest bearing combination. • Good for patients with lower activity levels who will not need a substantial amount of wear from their prosthesis.
MoM	• Hard bearing surface. • Associated with high revision rates.	• Can be used with large femoral heads, however due to high revision rates they are no longer recommended for use.

Abbreviations: CoC, ceramic-on-ceramic; CoP, ceramic-on-polyethylene; MoM, metal-on-metal; MoP, metal-on-polyethylene.

The Orthopaedic Data Evaluation Panel (ODEP) provides independent ratings of the available clinical evidence for hip prostheses. ODEP could be strengthened through formal recognition by NICE, extension to 15 years and through the differentiation of implants that outperform their respective class

The Orthopaedic Data Evaluation Panel (ODEP) provides an independent assessment of the quality and number of years of available evidence for prosthesis brands evaluated against the original NICE guidance (TA2) (10). NICE should consolidate the role that ODEP ratings currently play by formally endorsing the current rating system and recommending its extension to 15 years in line with patient expectations and the maturity of some devices. Further development to these ratings to recognise products performing better than their class would help signpost the NHS to purchase the most clinically effective devices.

The Corail® Stem/Pinnacle® Cup cementless prosthesis from DePuy is the most widely used implant combination in England and Wales according to the National Joint Registry, and the C-Stem®/Marathon™ Cup is a well-established cemented prosthesis combination from DePuy

DePuy's Corail® Stem/Pinnacle® Cup cementless prosthesis is currently the most widely used prosthesis combination according to the National Joint Registry (38). A range of femoral heads and acetabular liners allows metal-on-polyethylene, ceramic-on-polyethylene and ceramic-on-ceramic prosthesis combinations within the Corail®/Pinnacle® cementless system, including larger head diameter options. This range of articulating surface combinations allows the surgeon the flexibility to tailor the prostheses to the individual patient and the Corail® stem and Pinnacle® cup can also be used in hybrid and reverse hybrid procedures with other DePuy products, within their CE marks. The Corail® cementless femoral stem has an ODEP rating of 10A (at least 10 years of the highest quality evidence) and the Pinnacle® cup has an ODEP rating of 7A (7 years of the highest quality evidence). The Corail® stem has demonstrated survivorship of 97.8% at 10 years in patients under 50 years old, demonstrating its suitability for younger patients (21).

The C-Stem® and C-Stem®AMT/Marathon™ Cup system is a suitable option for patients with poorer quality bone stock, for whom surgeons feel that additional fixation with cement is

required, or for patients with low levels of activity or elderly patients who would not be expected to require revision surgery. The moderately cross-linked Marathon™ material has a lower rate of wear compared with non-cross-linked high molecular weight polyethylene (22-24), therefore reducing the risk of acetabular loosening due to osteolysis. In addition, the C-Stem®/Marathon™ components are compatible with hybrid procedures. The C-Stem® and C-Stem®AMT/Marathon™ Cup-also provides a suitable option for surgeons whose training and expertise has been gained in performing cemented procedures. The C-Stem®, C-Stem®AMT and Marathon™ Cup implants have ODEP ratings of 10A, 5A and 3A respectively.

Surgical training courses are provided for registrars and consultants on the safe and effective use of DePuy prostheses. A variety of courses are provided, dependent on the needs and experience of the surgeon or theatre staff. These courses have the highest independent ratings of all medical device courses offered in the UK and are well attended (see Appendix I).

The main source of clinical effectiveness data for THR and hip resurfacing arthroplasty in England and Wales is the National Joint Registry

For the purpose of the NICE appraisal, a systematic review was conducted to identify randomised controlled trials (RCTs) reporting on the clinical effectiveness of THR vs hip resurfacing arthroplasty and cemented THR vs cementless THR. On 19th December 2012, NICE issued an updated scope, however, the systematic reviews were conducted according to the brief of the original scope issued in October 2012 along with the invitation to make a submission, and therefore the population considered was ‘adult patients with end stage arthritis of the hip’. The studies that were identified were heterogeneous in terms of the outcomes that they reported and the length of follow-up that was conducted. In addition, there were no consistent differences between THR and hip resurfacing arthroplasty or cemented THR and cementless THR in the clinical trial results that were reported, therefore the RCT data were not considered suitable to inform the economic model.

The clinical effectiveness of THR and hip resurfacing arthroplasty has been well established through the collection of observational data. A number of national arthroplasty registries throughout Europe and elsewhere collect data on the clinical effectiveness of hip replacement prostheses, including the National Joint Registry (NJR) in the UK which was established in 2002, is funded by the Department of Health (25) and contains data on hip replacement procedures performed over the last 8 years.

The main clinical outcome reported by the NJR is the revision rate; NICE has previously set a benchmark revision rate of 10% at 10 years and prostheses are monitored to ensure that they meet this criterion through the NJR. There are multiple reasons why a patient may need revision surgery, with the most common reasons reported in 2011 as aseptic loosening (42% of revisions), pain (24% of revisions), osteolysis (13% of revisions) and dislocation/subluxation (13% of revisions) (12). Possible causes of revision include prosthesis failure, surgical error, surgical complications and quality of rehabilitation care.

Analysis of patient level data from the NJR and via NJR Supplier Feedback data demonstrates lower rates of revision for all types of THR compared with hip resurfacing arthroplasty and little clinically relevant difference in revision rates between different types of THR

For this appraisal, the NJR provided DePuy with patient level data on primary hip procedures performed since 2003. This very large dataset, comprising in excess of 470,000 records,

contained information on prosthesis class and bearings used but did not contain manufacturer or brand information. In addition, an NJR Supplier Feedback dataset for DePuy manufactured prostheses was also made available to inform the economic model following permission granted by the NJR. This dataset contained over 140,000 records.

In line with data reported in the NJR 9th Annual Report (12), the patient level dataset confirmed that hip resurfacing arthroplasty is associated with substantially higher revision rates than any class of THR. Qualitative observation of revision rates over time show a lower revision rate at all years with Corail®/Pinnacle® compared with cementless prostheses as a class (Figure 2). It should be noted that due to the high percentage of Corail®/Pinnacle® used in the cementless class, the overall cementless revision rates are artificially low as they include the best in class Corail®/Pinnacle® revision rates. Qualitative observation shows that revision rates with DePuy cemented prostheses were consistent with the overall cemented class (Figure 2).

Figure 2: [redacted]

As the NJR only contains data on revision rates for up to 8 years post procedure, it was necessary to perform extrapolation of the patient level and Supplier Feedback data to predict revision rates for the lifetime of patients. This was achieved by fitting parametric survival curves to the data in order to provide the probability of revision at any required time point. In order to account for the two populations specified in the NICE scope (i.e. patients who are suitable for THR or hip resurfacing arthroplasty and patients not suitable for hip resurfacing arthroplasty), extrapolations were performed based on patient age in the respective groups (i.e. 55 years and 70 years). Extrapolations for up to 30 years post-procedure showed that Corail®/Pinnacle® is associated with lower revision rates than cemented or cementless prostheses classes, but slightly higher than hybrid prostheses in both patient populations (Figure 3).

Figure 3: [redacted]

The NJR provides an excellent resource of data collected over several years, however, there are a number of limitations associated with the dataset that need to be taken into consideration when making comparisons between different prosthesis classes.

- The NJR offers the benefits of being able to reflect clinical decision making in relation to individual patient need, but is subject to bias due to the non-randomised, non-controlled nature of prosthesis class and bearing allocation; this is highlighted in the different average ages of patients that receive cemented THR (average patient age 73 years) and cementless THR (mean patient age 65 years). The trend for the use of cementless prostheses in younger patients and cemented prostheses used in older patients may result in bias against cementless prostheses in terms of revision rates. Younger patients are typically more active than older patients and are therefore likely to gain more use from their prosthesis, leading to a greater risk of wear and subsequent need for revision. In addition, younger patients are expected to live for longer, and would therefore be

exposed to the risk of revision for a longer period of time. Older patients who have cemented THR may be considered unfit for revision surgery due to frailty or comorbid conditions, potentially resulting in an under reporting of hip failures in this population. Additionally, the NJR does not consider surgical factors such as surgeon experience, the training of the clinical team, learning curves or patient factors, such as activity levels and general health when comparing revision rates between different classes of THR.

- When the NJR was first established, the registration of procedures was not mandatory, which may also provide an additional source of bias for the earliest NJR data on which the longest-term follow-up is based.
- The NJR reports revision rates based on revisions recorded due to any cause. In reality, the impact of a revision on the patient and the NHS may vary depending on the reason for the revision and the component of the prosthesis that needs to be replaced. For example, an acetabular liner in a cementless prosthesis that needs to be replaced due to wear may be relatively easy to replace with minimal impact on the subsequent functioning of the patient as the liner can simply be removed from the fixed shell without additional bone loss. Conversely, because the cup and liner comprise a single unit, the removal of a cemented acetabular cup may be a more complicated procedure with a greater amount of bone loss. Revision of ceramic implants may be associated with better patient outcomes, as ceramics are associated with low rates of osteolysis and therefore may have less bone loss compared with conventional polyethylene components.
- The NJR patient level/Supplier Feedback data does not capture hip dislocations which are a serious adverse event associated with substantial costs and impact on QoL and patient reported outcomes (26, 27). Consequently, the NJR data do not allow differences in dislocation rates due to the femoral head size used to be accounted for. Evidence has demonstrated that the use of larger head sizes available for cementless prostheses (i.e. 32 mm and above) are associated with a reduction in the rate of post-operative dislocation) (14-19).
- In 2011 there were 119 different brands of acetabular cups and 142 brands of femoral stems used in primary and revision procedures. The NJR does not make a full breakdown of revision rates by manufacturer and brand publically available, but there is likely to be inter-class variability in revision rates based on individual product performance based on results from other registries (28).

The extrapolation of long-term revision rates from the existing 8 year NJR dataset is associated with considerable uncertainty. In spite of the need for extrapolation, the NJR dataset is considered to be the most relevant to this appraisal, containing real world data from patients who have received hip replacement procedures in England and Wales. Many other nations around the world have well established joint registries that contain revision rates for a greater length of time than the NJR (for example the Australian Orthopaedic Association National Joint Replacement Registry and the Swedish Hip Arthroplasty Register). However, there are differences in outcomes observed from country to country that occur as a result of the historically variable use of cemented and cementless THR, the associated training and products used in each class and variation in the prosthesis brands used.

A micro-costing model of cemented and cementless THR showed great variation in cost

across and between different classes of THR, therefore resulting in a high level of uncertainty around input costs to the economic model

Due to the absence of UK-specific costing data, a micro-costing analysis was conducted to determine individual factors contributing to the overall cost of a THR procedure. Surgical procedures conducted in different centres were assessed for staff utilisation, theatre time and resource use. Prosthesis acquisition costs were also considered. Whilst the acquisition cost of a cementless prosthesis was higher, due to the use of list prices (which is considered the worst possible scenario as discounts are available to the NHS), cemented [redacted]. There was considerable variability in the overall cost of conducting a THR procedure regardless of the class - this is likely due to natural variability between surgeons and centres, as well as due to surgeons with different levels of experience performing the operations. A time in motion study conducted by DePuy has shown that registrars take longer to perform an operation than consultants and inexperienced theatre staff take more time over procedures than experienced staff. In addition to these factors, the individual characteristics of the patient may considerably affect operating time; for example, a patient with a high BMI or comorbidities may require a more complicated surgical procedure, thus suggesting that the simplest procedure is likely to be most efficient.

The conclusion of the micro-costing analysis was that the procedural cost differential is driven by the length of the procedure. There are various stand-alone observations which were shown to contribute to the overall procedure time, some of which (such as the mixing and setting time of cement) are easier to quantify than others (such as variation due to individual patient characteristics). There is not only uncertainty around each stand-alone parameter, but also around the combination of these factors occurring at the same time in any one case.

The relative cost-effectiveness of the different classes of THR is highly uncertain due to substantial and overlapping variability in costs and comparable quality of life gains

As per the NICE scope, an economic model was developed to determine the cost-effectiveness of different types of THR and hip resurfacing arthroplasty in patients suitable for both procedures, and to determine the cost-effectiveness of different types of THR in patients not suitable for hip resurfacing arthroplasty. The two populations were distinct in terms of age and gender mix; patients suitable for THR and hip resurfacing arthroplasty were mostly male (70.9%) with a mean age of 55.3 years. Conversely, patients not suitable for hip resurfacing arthroplasty were mostly female (62.5%) with an average age of 70.4 years.

Different classes of THR that were considered in the model included cemented THR, cementless THR, hybrid THR and reverse hybrid THR. In addition, DePuy cementless (Corail®/Pinnacle®) and DePuy cemented prostheses were also considered. Efficacy data was taken from the NJR and cost data from a micro-costing analysis. In line with the NICE reference case, list prices for prosthesis acquisition were used in the base case analysis. This is considered very conservative as the list price is not representative of the prices that DePuy products are available to the NHS. Depending on the combination of components and the volume acquired, the acquisition cost of a DePuy cementless hip can be very similar to a cemented hip. As a result the cost-effectiveness calculations for medical devices always reflect the worst case scenario, given that discounts are provided to the NHS and prices

decrease over time unlike pharmaceutical products which are protected by patent and more regulated pricing until expiry of the patent.

An incremental analysis was performed to compare the cost-effectiveness of each class of THR and hip resurfacing arthroplasty in patients suitable for both procedures, and the cost-effectiveness of each class of THR in patients not suitable for hip resurfacing arthroplasty.

[redacted]

Hip resurfacing arthroplasty was more costly and less effective than each class of THR. In both patient populations, there was a marginal difference in the QALY gain between the different classes of THR, suggesting that the clinical performance of different constructs in terms of revision rates are very similar, considering the limitations and heterogeneity of the data available.

The relative cost-effectiveness is therefore largely driven by differences in costs, which are highly uncertain as demonstrated in the micro-costing analysis and it is therefore difficult to compare relative cost-effectiveness. The high level of uncertainty undermines the ability to analyse the cost-effectiveness of the different prosthesis classes.

In the base case analysis, a pairwise comparison demonstrated that All THR dominated hip resurfacing arthroplasty in the population of patients that were suitable for both procedures. Univariate sensitivity analysis showed that THR remained the dominant treatment strategy in all analyses performed.

[redacted]

Hip resurfacing arthroplasty was more costly and less effective than each class of THR.

Uncertainty around model inputs was assessed in probabilistic sensitivity analysis (PSA). Hip resurfacing was associated with substantially higher costs and fewer QALYs compared with all classes of THR and can be seen as an outlier on the cost-effectiveness frontier in patients suitable for THR and hip resurfacing (Figure 4). In patients suitable for THR and hip resurfacing arthroplasty, here is a negligible difference in QALYs between all of the THR classes of approximately 0.21 QALYs; equivalent to 7-8 quality-adjusted life days over a period of more than 11 quality-adjusted life years. It can be seen that there is substantial overlap in the 95% credible intervals for all of the THR classes indicating that that uncertainty and mean costs and QALYs may be considered equivalent. Any overlap in the credible intervals between hip resurfacing and the group of THR technologies is negligible; therefore in general the costs and benefits of THR can be considered different to those of hip resurfacing.

For patients not suitable for hip resurfacing, results from the PSA (Figure 4) suggest that any long term incremental benefits associated with a particular class of THR may not be observed due to the older age of patients who are not suitable for hip resurfacing. DePuy cemented was the least costly technology, at a cost of £7,700 for approximately 8.1 QALYs. There is substantial overlap between each of the technologies in terms of costs and QALYs, and that the incremental differences are negligible; thus all classes of THR may be considered equivalent.

Figure 4: [redacted]

Conclusion

Patient-centred clinician decision making is the means of ensuring an efficient service provision and the best possible outcomes for patients and therefore all types of THR should be recommended for use in the NHS.

The uncertainty around the costs associated with different classes of THR highlight the variability that exists in the cost of performing a procedure between surgeons and centres. On the basis of the results presented in this analysis, any recommendations on prosthesis selection should not only consider the results of a cost-utility analysis (such as the one presented here), but should also take into account the operational issues associated with providing a THR service, the impact of training, the variability of costs and results between centres and the preference of different centres for the use of particular constructs on the basis of effectiveness, efficiency and costs at a local level. This approach was favoured by surgical experts (see Section 4.3.3.2 for details).

The current range of choices within THR including the different prosthesis classes, the bearing combinations available, the different stem designs and femoral head sizes reflect the need for patient-centred decision making, driven by the choice, training and expertise of the surgeon according to patient need. Of critical importance is the safety and performance of products within each class, therefore we propose that NICE strengthens the role ODEP ratings play by formally endorsing the current rating system and considers its extension to 15 years in line with patient expectations. Further development to these ratings to recognise products that outperform their class would help signpost the NHS to purchase the most clinically effective devices. There is additional value in a prosthesis that has excellent quality data over a substantial length of time.

The orthopaedic industry is currently working with ODEP and the orthopaedic profession to set up an additional data collection service (Beyond Compliance) to link NJR, PROMs and Hospital Episode Statistics (HES) data with clinical imaging and surgical information on newly introduced implants with independent review, an approach which we believe should be endorsed by NICE to strengthen the evidence base for newly introduced technology. This additional data is expected to feed into ODEP ratings.