DePuy Synthes
JOINT RECONSTRUCTION
COMPANIES OF Johnson & Johnson

Johnson & Johnson response to the Assessment Group's report

MTA: Total hip replacement and resurfacing arthroplasty for the treatment of pain or disability resulting from end stage arthritis of the hip

Key points

- Johnson & Johnson is encouraged by the level of similarity of the findings of the work undertaken by the Assessment Group and the findings of our own modelling presented in our manufacturer submission, namely that THR is a cost effective intervention which dominates hip resurfacing and is associated with significant, sustained quality of life benefits.

- It is felt that the overall findings of the assessment report are strengthened by the level of corroboration achieved when compared to the work undertaken for our submission and model, despite some differences in modelling approaches, data sources and assumptions.

- We are therefore in broad agreement with the Assessment Group's findings; specifically the overarching conclusion that while there is confidence that THR dominates hip resurfacing in all circumstances, a high level of uncertainty around comparisons between classes or categories of THR, such as type of fixation or bearing, make the validity of such comparisons questionable. This is primarily due to the high level of uncertainty around cost-input parameters between categories and the QALY-gains being largely the same in each case.

- It is our belief that NHS reference cost data lacks validity when used in the base-case analysis and the use of manufacturer list prices (per NICE protocol) and costs gleaned from NHS Supply Chain underestimate the significant discounts available to the NHS, consequently overestimating the cost of prostheses overall and should not be used in the base-case.

- We question the validity of using NHS Supply Chain in the base case analysis. According to information supplied to us by NHS Supply Chain directly, it is their belief that only around 1 in 5 of hip prostheses sold in England is procured through them. We know that for some implants this would be well under 1 in 10 procured through that route. As a result, the cost estimates are biased towards implants with a higher share of the market and cannot be relied upon to make any robust conclusions on prostheses costs.

- It is important to recognize the total cost of intervention is much broader than simply device (prosthesis cost). Indeed, the total procedure cost comprises theatre time, staff costs, and ward stay, in addition to the cost of the device and additional consumables.

- Based upon the results of the DePuy Synthes time-in-motion study, a significant degree of variability was observed across the parameters that have a direct impact on procedure times for prostheses implants. This finding highlights an overarching issue inherent in extrapolating micro-level data, in order to reach macro-level conclusions. This was particularly true for the time-in-motion study (see DePuy Synthes submission, Appendix H, pgs. 233-235), which demonstrated that a wide range of factors beyond prostheses selection had a significant impact on procedure times and thus, costs incurred by the Trust.

I. Executive Summary

Johnson & Johnson welcome the opportunity to respond to the Assessment Group's report of 23rd July, 2013. In this response document, it is our intention to highlight areas where we are aligned with their findings, as well as points which we wish to question. Where there were critiques and limitations noted in the context of the DePuy Synthes submission, we have provided additional information for clarification purposes.

Firstly, we agree that, THR represents the dominant treatment strategy as compared to RS – a finding which is robust across a number of assumptions. As stated in the Assessment Group's report, "*Compared to THR, revision rates for RS were higher, mean costs for RS were higher and mean QALYs gained were lower; RS was dominated by THR. Very similar results were obtained for deterministic and probabilistic results and for all age and gender groups and THR was almost 100% cost-effective at any willingness to pay level*" (pp.298). Furthermore, we agree that, given the level of uncertainty in, and inconclusive nature of the evidence base, a granular-level of analysis (e.g. further stratification of THR into Categories A-E) is impractical and does not accurately reflect clinical practice. This view is echoed in the Assessment Group's findings, "*The marginal difference in QALYS is negligible between Categories A to E. On the basis of a negligible difference in QALYs, it is therefore difficult to make a fair comparison between them in terms of outcomes*" (pp. 297). They further state, "*Probabilistic analyses of costs and effectiveness of all categories of THR overlapped markedly confirming that differences are relatively small*" (pp. 298). Accordingly, we assert that any form of prescriptive guidance tied to this level of analysis is problematic and undermines important aspects of the decision-making process; namely, clinical judgment and patient choice. Indeed, hip surgeons have a professional duty to individually assess a patient and discuss with them, the advantages and disadvantages of different implants and fixation choices. This ensures a choice that is fair and consenting, and one which is made with full disclosure to the patient.

In light of NHS England's recent publication of surgeon-level, outcomes data into the public domain, we argue that a level of autonomy in clinical judgment remain intact if these professionals are to be held to account for their decisions. In consulting with an Advisory Board (December 2012) comprised of expert orthopaedic clinicians, there was a strong level of agreement that previous policy decisions resulting in restricted patient and/or clinician choice had a significant and adverse effect on patient outcomes. Specifically, the Advisory Board believed that such decisions impacted negatively on the public's and professionals' confidence in the decision-making process. They further emphasized that guidance must be deployed with caution, ought to be implementable, and needs to be informed by a careful review of registry data. As the evidence base advocating/discouraging the use of cemented technologies is evolving, the Advisory Board cautioned against decisions that force surgeons to adopt a technique for which they are neither familiar nor comfortable using.

We wish to address the distinction between cemented and cementless technologies, especially as it relates to the policy context. The current state of affairs is one in which many Trusts do not use cement, and therefore their trainees have not received appropriate training. It is also true that when surgeons who are unfamiliar with cemented technologies use cemented prostheses, they achieve poor patient outcomes. Thus, policies enforcing the sole use of cemented technologies will require a national training programme on the use of these types of prostheses, both for trainees and their supervisory consultants. Furthermore, additional training will be required for operating department personnel, scrub nurses, anyone in theatre wearing contact lenses (they should not be in the operating theatre while the cement is being mixed), and any of the staff who may possibly be pregnant.

The unintended consequence of forcing hip surgeons to adopt the use of a fixation method that is not their preferred option is a reduction in the number of surgeons available to perform procedures in a clinical area with a high level of demand. This could potentially translate into negative implications for waiting times; however, the converse would be an environment in which surgeons have the freedom to choose their preferred fixation method, and are therefore able to perform more cases on a list, likely with fewer revisions.

1.1. Cost considerations

1.1.1. Additional cost of consumables

Cemented technologies demand the purchase of additional consumables, such as: canal vent, cement restrictor, canal brush and pulse lavage long nozzle. In addition, the reusable kit for cemented implants is typically more costly (as compared to cementless) and rarely supplied by the implant firm. This has a clear implication for the costs incurred by the Trust (e.g. equipment purchase, upkeep, and replacement costs). Some clinicians find the kit for cementless fixation simpler, and they also tend to be integrated with the manufacturer-provided sets. As a result, fewer trays are required for a cementless stem. Arguably, this translates into cost savings related to sterilization, storage, set-up time, trolleys, and manual handling. Anecdotal evidence from Stoke demonstrates a reduction in kit required per cementless procedure, which amounts to a savings of roughly £50/case. For a center that completes 400 cases/year, a sizeable efficiency savings is conferred upon the Trust.

1.1.2. NHS Supply Chain cost data

As it relates to the use of NHS Supply Chain costs in the Assessment Group's health economic model, it is our view that such an approach represents a crude estimation of costs. Estimates obtained directly from NHS Supply Chain suggest only around 1 in 5 hip replacement prostheses are procured through that organization. In addition, an analysis of DePuy Synthes sales channels associates some of the highest volume products with sales channels *outside* of the NHS Supply Chain procurement route. Fewer than one in ten of DePuy Synthes' products are sold through NHS Supply Chain, and in the instances where NHS Supply Chain is involved, prices are substantially higher than our average selling price, due to the lower volumes procured. Taken together, these factors would indicate an overestimation of costs, and while the Assessment Group's analysis demonstrated substantial intra-category cost differences, we would request that the Committee interpret these findings in the context of the evidence presented here.

1.1.3. Variability in total procedure cost (beyond the device)

[redacted]

[redacted]

II. Further explanation of methodology

In the current section, we provide additional information related to the methodology used to construct our health economic models, as well as further justification around the parameterisation of the models.

3.1. Systematic literature review

Abacus International (on behalf of Johnson & Johnson) do not as standard (for time and cost reasons) use two independent reviewers at first and second pass; however, data extraction is undertaken by a second independent reviewer as part of the quality assessment process.

Risk assessment of selection bias was undertaken as part of the quality assessment during the systematic review process, however, it was agreed that due to the increasing length of the submission and the importance of the National Joint Registry data, the full quality assessment would not be written into the final submission document.

We understand that the criticism around exclusion criteria refers to a lack of detail as to the excluded studies and the rationale for doing so. The exclusion criteria were clearly stated in the DePuy Synthes submission document (p.48), however, a full list of excluded studies was not included, for reasons outlined above.

Quality assessment of the included literature was undertaken in full at the data extraction phase and has been provided to Johnson & Johnson in a number of data extraction tables. Again, due to the increasing length of the submission document, full quality assessments were not included in the final submission document.

3.2. Internationally-derived registry data

The use of generalisable national registry data on hip arthroplasty from Australia and Sweden is justified on the basis of time horizon and procedure coverage in order to supplement evidence from the UK NJR. The UK NJR has only been collecting data since April 2003 and is currently celebrating its 10th anniversary. The Swedish Hip Arthroplasty Register started collecting data on primary hip replacement in 1979[b]. Data from Sweden was therefore included in our submission as this registry has the longest time horizon of any hip registry worldwide and is the only published source of generalisable registry revision rates for THR patients in their second decade post-surgery.

Established clinical practice in Sweden supports primary hip arthroplasty that employ cemented femoral and acetabular components, decreasing as a percentage of total procedures from about

[a] Baker, P., Deehan, D., Lees, D., et al. 2012. The Effect of Surgical Factors on Early Patient Reported Outcome Measures (PROM S) Following Total Knee Replacement. *Journal of Bone and Joint Surgery,* 94:1058-66.

[b] Swedish Hip Arthroplasty Register. Available from: http://www.shpr.se/en/default.aspx Last accessed 13 August 2013

95% in 2002, to approximately 85% in 2011. Therefore, this registry's data is more limited vis-à-vis revision rates for either cementless or hybrid/reverse hybrid hip arthroplasties.

The Australian Orthopaedic Association National Joint Replacement Registry was established in September 1999, and thus extends to a longer time horizon than the UK NJR, with the ability to report 11-year outcomes[c]. In addition, clinical practice in Australia has been to carry out primary hip arthroplasty using a much greater proportion of cementless femoral and acetabular components, increasing from about 50% of total procedures in 2003, to about 60% in 2011. Data from the Australian Orthopaedic Association National Joint Replacement Registry was therefore included in our submission regarding outcomes with this technology.

3.3. Cost inputs

Expert clinical advice sought at an Advisory Board in December 2012 suggested the cost of revision was considerably greater compared to a primary procedure - at least double. In the base case, the mean cost of revision (£13,339) was assumed to be double the average cost of a primary THR whereas the Assessment Group used an estimate from Vanhegan et al 2012 and inflated to 2011/12 prices (£16,517).

Sensitivity analysis was conducted to investigate the impact of this assumption, with the mean costs of revision being varied between £6,700 and £20,099 in one-way sensitivity analyses. The net benefit of THR vs. RS in patients suitable for both procedures remained positive, varying between £4,025 and £5,204. Our approach to the cost of revision can be considered conservative compared to that of the Assessment Group, as any increase in the cost of revision would make THR more cost-effective versus RS, and DePuy's THRs more cost-effective vs. other classes of THR.

NHS Reference Costs do not disaggregate costs for cemented and cementless procedures; therefore a micro-costing approach was used to overcome this limitation and to reflect the difference in costs between different procedures. The Assessment Group states "The model by DePuy has several strengths" (pp.316-317) including that resource use was based upon a micro-costing analysis, resource use was based on a detailed bottom-up costing method (i.e. a time in motion study), prostheses costs were based on the manufacturer's list prices rather than the average selling prices available to the NHS, and that the impact of using NHS Reference Costs were investigated in the sensitivity analyses.

Scenario analyses using NHS Reference Costs (DePuy Synthes submission pp.130-136) showed that the incremental costs for each procedure were reduced, however, key conclusions remained unchanged: in patients suitable for both procedures RS was dominated by THR; and in patients not suitable for RS the incremental costs and QALYs were negligible (total costs ranged from £7,297 to £7,381 and total QALYs from 8.799 QALYs to 8.814 QALYs over a lifetime).

3.4. HRQoL parameters

Expert clinical advice suggests patients would have a lower post-operative QoL following revision compared to patients receiving a primary procedure. The mean post-operative utility was assumed to be 0.635 in the 3 month time period following revision based upon Briggs et al 2003. The Assessment Group commented that a utility of 0.635 would not reflect the lower QoL reported by Dawson et al 2001 of 0.62; however, the impact of this assumption was tested in one-way sensitivity analyses (SA's) and the PSA and was found to be minimal. Based upon the standard error reported by Briggs et al 2003 for the coefficient associated with revision from a regression model for estimating QoL scores, post-operative utility was varied between 0.598 and 0.672 in one-way SA's. As reported in Table 42 and Figure 35 (pp.138-139) of the DePuy Synthes submission, the impact of

[c] Australian Orthopaedic Association National Joint Replacement Registry. Annual Report. Adelaide: AOA; 2012

varying the disutility in patients suitable for THR and RS between 0.108 and 0.182 was minimal: THR continued to dominate RS and the net benefit of THR vs. RS remained positive, varying between £4,131 and £5,098.

In summary, using Dawson et al 2001 may have been the more methodologically robust approach, however this assumption has minimal impact upon the results and use of the Briggs et al 2003 disutility value gives a more conservative estimate.

3.5. Choice of curves fitted

The methodology implemented to inform the analysis of revision rates in the DePuy Synthes submission conformed to the current advice issued by the NICE Decision Support Unit (DSU) regarding survival analysis, though it is recognised that the context of this guidance is modelling survival within a single clinical trial and does not address the use of observational data.

A concerted effort was made during the analysis of revision rate data to try and create a 'bathtub' effect from first principles using early/late causes of revision. The reasons why we do not believe this to necessarily be the most appropriate functional form for modelling of the NJR dataset is discussed in Appendix D, Section 4 of our submission. We have not been able to review the Assessment Group Report's Appendix 18 to comment on the methodology employed by the Assessment Group, as it has not been provided to us, but given the effect on the results between bathtub and lognormal distributions, the uncertainty with regards to this area of the modelling must be emphasised emphatically. Past performance of prosthesis does not necessarily provide a suitable representation of future performance (see Australian registry reports for examples) and there therefore exists temporal uncertainty present in the extrapolation, over and above the parameter uncertainty in the coefficients of the revision models. The current probabilistic analysis performed by the Assessment Group does not reflect this temporal aspect of uncertainty, and therefore in our opinion fails to characterise fully the true uncertainty in the outcomes.

3.6. Different indications for revisions (and the associated costs)

KOL opinion sought at an Advisory Board in December 2012 to inform our submission suggested that there are multiple reasons for revision and, in addition, manifold options for types of hips used for revision that it was almost impossible to make any clear assumptions. On this advice it was agreed that the most reasonable assumption would be to create a weighted average cost of revision, with the impact of this assumption being tested in the sensitivity analyses

3.7. New revision rate data

The Assessment Group categorised the various types of THR differently to the approach used in our submission. The Assessment Group identified 5 groups of THR based on expert clinical advice and the most frequently used cup and head component combinations reported in the NJR dataset. Based on expert clinical advice, the DePuy submission categorised THR based on method of fixation which was consistent with the National Joint Registry's 9th Annual Report 2012. In addition, two further classes of DePuy THRs were considered. Comparison of revision rates based on the Assessment Group and the DePuy submission's categorisation of THR is therefore difficult due to the different approaches employed. Table 1 presents a comparison of revision rates at 9 years based on the alternative methods of THR classification, however the limitations of such a comparison should be born in mind.

Table 1: Comparison of THR revision rates at 9 years between the Assessment Group and the DePuy Synthes submission

[redacted]			[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]						

[redacted]

3.8. Subgroup analyses based upon age and gender

The statistical analysis of the NJR dataset in our submission accounted for age and gender through the inclusion of age and gender as covariates in the model selection process (DePuy Synthes submission, pp.106). In addition, alternative functional forms for age were considered by stratifying age at primary procedure (all ages, <70 years and <55 years). Analysis of the NJR dataset found that the mean age and gender varied between prosthesis classes (DePuy Synthes submission, pp.77) and was consistent with expert clinical advice. In order to reflect the differences in the population of patients for whom both procedures are suitable and for whom RS is not suitable (as specified in the Final NICE Scope, Dec 2012), the cost-effectiveness analysis considered two populations Table 2 (DePuy Synthes submission, Table 20, pp.101).

Table 2: Patient populations considered in the model

Population	Mean age (years)	Male, %	Assumption
Hip resurfacing	55.3	70.9	Suitable for THR and resurfacing arthroplasty
THR	70.4	37.5	Not suitable for resurfacing arthroplasty
All patients	69.2	40.5	

The cost-effectiveness of different types of prosthesis in specific age and gender groups was not investigated due to equity concerns and because it was considered outside the Final NICE Scope (Dec 2012). Scenario analyses 4 and 5 investigated the impact of alternative Weibull models of revision stratified by age at primary procedure <70 years and <55 years, respectively (DePuy Synthes submission, pp.129-136), however, conclusions remained the same: (1) THR dominated RS and (2) the marginal difference in QALYs between classes of THR was overshadowed by the uncertainty.

The Assessment Group evaluated the cost-effectiveness of RS vs. THR separately for males and females by age group (40, 50 and 60 years old). For all sub-groups (including both deterministic and probabilistic results) THR dominated RS as it was both less costly and more effective, which was consistent with the base case analysis presented by the Assessment Group and within our submission. In addition the Assessment Group evaluated the cost-effectiveness of different THR categories split by age and gender. The Assessment Group reported that "The marginal difference in QALYS is negligible between Categories A to E" (pp.297) as incremental QALYs were apparent at 2 to 3 decimal places. Given the considerable overlap in the 95% CI's for each of the categories of THR, the negligible QALY differences is overshadowed by the uncertainty around the estimates and therefore any decision to restrict certain categories of THR to specific age and gender groups would be subject to substantial uncertainty.

III. Conclusion

The clinical and cost-effectiveness evidence base for THR compared to RS is well established; THR is more effective and less costly than RS. However, the evidence is less certain at the level of stratification adopted in the Assessment Report for different categories of THR, which shows that there is a negligible QALY difference between different THRs and considerable uncertainty around cost input parameters. While there is confidence that THR dominates hip resurfacing in all circumstances, a high level of uncertainty around comparisons between classes or categories of THR, such as type of fixation or bearing, make the validity of such comparisons questionable. Indeed, the Assessment Group concluded that, "*The marginal difference in QALYS is negligible between Categories A to E. On the basis of a negligible difference in QALYs, it is therefore difficult to make a fair comparison between them in terms of outcomes*" (pp. 297).

We therefore wish to reiterate the importance of clinical judgment and patient choice as the key drivers in decisions related to implant selection. Moreover, it is our view that, preserving a degree of clinical autonomy has important implications for potential costs averted due to the retraining of surgeons, as well as their ability to achieve quality outcomes. As patient choice factors prominently in the new policy landscape of the NHS, we believe it ought to be accorded a level of significant importance in guidance documents as well.

In closing, we acknowledge the Assessment Group's critique of our submission. Where appropriate, we have sought to provide additional detail for the sake of clarity, and above all else, to demonstrate

that we have adopted a methodology that is consistent with the guidelines set forward by NICE; is informed by a robust analysis of the evidence from the literature, together with expert opinion; and in some instances, represents a conservative approach. This point is particularly salient with respect to the cost inputs, where a notable level of variability pervades the evidence base.