PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cost-effectiveness of total hip arthroplasty versus resurfacing arthroplasty: economic evaluation alongside a clinical trial
AUTHORS	Edlin, Richard ; Tubeuf, Sandy; Achten, Juul; Parsons, Nicholas; Costa, Matthew

VERSION 1 - REVIEW

REVIEWER	Hideki Higashi, PhD Research Fellow School of Population Health University of Queensland Australia I have no competing interests.
REVIEW RETURNED	14-May-2012

GENERAL COMMENTS	This paper aims to evaluate the 12 months incremental cost-effectiveness of hip resurfacing compared to total hip replacement. Although the main interests of readers would be in the 'long-term' cost-effectiveness of resurfacing arthroplasty, this paper would serve as an intermediate step towards that end. Overall the analysis has adequately addressed core issues and the paper is generally well written, albeit with some unclearness. I feel addressing the following issues would strengthen the paper considerably. **Comments** 1. P.15: Table 3 reports the uncertainties around incremental costs and incremental QALYs. First, an explanation should be added to the table what the values in parentheses mean (95% CI or UI?). Second, it is not clear why the uncertainties of ICERs have not been provided if all others have been. Was a non-parametric bootstrap not performed? I believe all necessary information to do this is available from the RCT. If it was done, details should be added to **Methods**. If otherwise, such analysis would be warranted. 2. P.6, 2nd para: Data on personal costs were collected but were not used in the analysis, which is due to the perspective defined in p.4 (NHS and PSS). However, it is not clear why this perspective was chosen. Who is the audience of this study?

	Are patient's costs not relevant for decision-making? It would be useful to see results of additional scenarios that include patient's costs, especially if the main cost difference between THA and RSA lies in the post-surgery period. 3. P.8-9: RSA was more costly "due to further inpatient care after initial discharge (£279) and outpatient care (£83)" (p.8, 3rd para), and "Those in the RSA arm had significantly more outpatient visits than those in the THA arm (5.155 vs. 3.063, P = 0.0054). Here, both the number of physiotherapy sessions and the use of DVT assessments were significantly higher amongst this group (P = 0.002, P = 0.011)" (p.9, 1st para). The higher cost required for RSA post-surgery gives an impression that the health state in this arm was not as good as those in the THA arm. However, QLAY is better in the RSA arm. Some discussions on this disconnect should be added in the **Discussion**. 4. P.11.1st-2nd para: Given the rising debate over the 'potential' side-effects of metal-onmetal implant [1] and the unavailability of other types for resurfacing at this time, this could also have implications on the long-term QALY as compared to THA that has various options. This issue warrants addition to the **Discussion**. 5. P.10 1st para: "…that RSA is the most cost-effective option….". The analysis is comparing only one intervention (RSA) incremental to the comparator (THA), so 'the most' sounds odd (simply 'a'?). 6. Many abbreviations are not spelled-out when they appear first in the text that should be added (e.g. QALY, NICE, RCT, THA, RSA, NHS, HCHS, etc.) **References** 1. Cohen D. How safe are metal-on-metal hip implants? BMJ. 2012;344:e1410.

REVIEWER	Helen Dakin Researcher University of Oxford, UK I have no competing interests, but am involved in trials on knee replacement.
REVIEW RETURNED	15-May-2012

THE STUDY	The manuscript describes the statistical methods for some but not all analyses presented in the paper. Mostly the statistical methods are appropriate, although I am concerned that the analysis underestimates uncertainty by ignoring uncertainty around missing values in adjusted analyses and not capturing uncertainty around implant costs in base case figures. The report addresses most points on the checklist, although many

	are addressed very briefly, or by citation.
RESULTS & CONCLUSIONS	Results are discussed in the light of the clinical results from the same trial, although no previous evidence on the cost-effectiveness of hip resurfacing is discussed. It is unclear whether this is because no previous economic evaluations (model-based or trial-based) have been done.
REPORTING & ETHICS	I assume that the research ethics are addressed in the main clinical paper.
GENERAL COMMENTS	The analysis appears to have generally been well-conducted and the manuscript includes most of the information listed on the checklist (albeit briefly). However, it appears that the analysis may underestimate (or ignore) uncertainty around missing data and implant cost. General comments · The costing methods appear to assume that operation time for the primary procedure and the incidence and cost of complications and re-operations occurring before patients are discharged from hospital are the same for both groups. It would be useful to discuss the extent to which this is true and to acknowledge this as a limitation. · The results are extremely sensitive to the type of THR conducted and the cost of hip prostheses for total hip replacement varies substantially between the three main types of prosthesis. From my experience in knee replacement, there is probably also a large amount of variation within each category and potentially between different types of resurfacing implant. It appears that the uncertainty or variability around implant cost is ignored in the analysis. It is unclear why it wouldn't have been possible to estimate implant costs separately for each patient and do subgroup or adjusted analyses to estimate the impact of implant type on costs and QALYs. · In addition, it is unclear whether metal-on-metal implants are still used in UK clinical practice and (if so) whether it is valid to include such implants in cost-effectiveness estimates. · Adjustment of QALY estimates for baseline utility is always recommended when there is marked imbalance at baseline; it would therefore be extremely useful to see the justification for using unadjusted values as the base case and to present results against metal-on-polyethylene implants and CEACs with and without adjustment for covariates. · The paper does not clearly describe statistical methods for all results presented in the paper. In particular, it is unclear how the confidence intervals in Tables 1-3 and the p-values on pages 8-9 or WebExtra Table 2 are calculated. It is also unclear whether the CEAC shown in Figure 1 is adjusted for baseline characteristics and (if not) how it was calculated, since the methods section only describes the methods used to calculate adjusted CEACs. · Page 6, lines 19-35: The description of missing data suggests that only data on total QALYs were used in multiple imputation, rather than utilities at each timepoint. This implies that patients with utility data at some but not all of these four timepoints were considered to have completely missing data on QALYs. If this is indeed the case, it would be useful to justify why this approach was used, since it could waste a large amount of data and means that one of the covariates explored in subsequent analyses (baseline utility) is omitted from the imputation model, which may introduce bias. It is also unclear from this paragraph whether imputation was conducted on total cost excluding length of stay, or whether different cost components were considered separately. · On page 7 (lines 43-47), it is stated that methods are not established to do seemingly-unrelated regression on multiply

	imputed data. I have occasionally managed to run suest after mim by saving estimates with “mim, storebv”, although the command is temperamental and it may be possible to do this using mi estimate. However, an alternative would be to use bootstrapping, as we have used previously (BMJ Open. 2012 Jan 30;2(1)). Averaging over imputed datasets before running the regression analyses that are used to estimate CEACs and 95% CI means that all uncertainty around the imputed values is excluded from the analsyis. Since the methods used to calculate unadjusted 95% CI are not described, it is unclear whether this problem affects the base case analysis as well as the adjusted figures. Specific comments · Page 3, lines 11-23: Centre-specific implant costs were used for to adjust HRG costs for the primary hospital stay. It is unclear whether the HRGs costs used reflect those for this trust, or the average costs across all English hospitals. · Page 3, lines 35-9: There is typically a large difference between list prices and the price that hospitals pay for prostheses and substantial variation between hospitals in the discounts secured. Were the costs supplied by the hospital trust discounted prices or list prices? · Pages 5-6: The costing section of the methods could be made clearer. It would be useful to say up front what resource data were collected for each patient and, potentially, which information came from patient questionnaires rather than hospital records. · Pages 5-6: The main manuscript does not explicitly state that the analysis includes all resource use incurred by trial participants within 12 months of THR (not just medications, hospitalisations, ambulatory consultations and aids or adaptations related to the study hip). · Page 6, line 3: It is unclear from the text or appendices what unit costs were provided by “relevant RCTs” or which RCTs were used. · Quantities of resources are not reported separately from their unit costs (although unit costs are reported separately from quantities). · There is no indication of how many patients had missing data on costs or QALYs, which makes it difficult to interpret the results of the complete case analysis or assess the impact of the methods used to impute missing data. · Page 8: when presenting the mean difference in costs and QALYs, it would be useful to specify that these figures are not adjusted for baseline characteristics and to give 95% CI around the point estimates, so that the reader can see that these differences are not statistically significant. · Discussion: For a single-centre study such as this, generalisability is a key issue. In particular, do most other centres use the Cormet hip resurfacing implant and is the price of the Cormet implant similar to implants used elsewhere? Are the types (and prices) of THR implants used at this centre comparable with those used elsewhere? · It would be useful to mention the results of the primary endpoint of the clinical trial in the introduction or methods (as per point 9 on the checklist). Minor comments: · Several acronyms are introduced without being first defined, including RSA and THR (page 4, lines 14-16), RCT (pg 3, line 45), as well as NICE, QALY and NHS. · Abstract: “Missing data was assessed” should read “Missing data were imputed” or “dealt with”. · Page 6, line 26: Should this read “[randomised] treatment allocation” rather than “treatment received”?

	· Page 6, line 27: it would be clearer to write out “Oxford Hip Score” and “Harris Hip Score” in full (particularly as “Oxford Knee” would normally imply a particular type of implant, rather than the OKS). · Page 9, lines 45-7: the wording of this statement is currently misleading and implies that the impact of gender and baseline EQ-5D was also to increase QALYs by 0.059. · Page 7, lines 8-11 imply that the complete case analysis represents a “per protocol analysis”, but in Table 3, per protocol results are presented separately from complete case. · Table 3: It would be useful to state that the values in brackets (presumably) indicate 95% CI. It would also be useful to explicitly state which implant was assumed in line 12. · Tables 2-3: it would be useful to state that costs are in £ Sterling. · Web Extra Table 1: it is unclear whether the list of drugs and prices includes all drugs considered in the analysis, or just the most commonly used.

REVIEWER	Callum McBryde MD FRCS (Tr&Orth) Locum consultant orthopaedic surgeon Royal Orthopaedic Hospital and University Hospital's Birmingham, United Kingdom
REVIEW RETURNED	16-May-2012

RESULTS & CONCLUSIONS	I think that the main message from the paper could be made more clear in the discussion. I think a very straightforward rewrite to highlight the main findings in slightly more simple terms would be useful but otherwise a very interesting paper.

REVIEWER	Professor Luigi de Palma Cattedra di Ortopedia e Traumatologia, Universita` Politecnica delle Marche, Azienda Ospedaliero-Universitaria,Ospedali Riuniti di Ancona, Via Conca, Torrette, 60100 Ancona, Italy
REVIEW RETURNED	24-May-2012

GENERAL COMMENTS	The Authors report their experience in cost-effectiveness analysis of total hip arthroplasty versus resurfacing arthroplasty. The manuscript is interesting and well written, but during the lecture of the manuscript some questions may have arisen. In my opinion after 12 months the Authors report to early results for this kind of evaluation, in fact, generally speaking, the clinical results are evaluated after 2 years to be considered expressive. The Authors report that “ this study has a pragmantic nature “, for these reason in cost-effectiveness analysis the Authors have to considerer the cost of complications. Finally, it could be very interesting to report data from patients (indications, age….) to explore the patients cohort.

VERSION 1 – AUTHOR RESPONSE

Reviewer: Hideki Higashi, PhD
Research Fellow
School of Population Health
University of Queensland
Australia

I have no competing interests.

This paper aims to evaluate the 12 months incremental cost-effectiveness of hip resurfacing compared to total hip replacement. Although the main interests of readers would be in the 'long-term' cost-effectiveness of resurfacing arthroplasty, this paper would serve as an intermediate step towards that end. Overall the analysis has adequately addressed core issues and the paper is generally well written, albeit with some unclearness. I feel addressing the following issues would strengthen the paper considerably.

Comments

1. P.15: Table 3 reports the uncertainties around incremental costs and incremental QALYs. First, an explanation should be added to the table what the values in parentheses mean (95% CI or UI?). Second, it is not clear why the uncertainties of ICERs have not been provided if all others have been. Was a non-parametric bootstrap not performed? I believe all necessary information to do this is available from the RCT. If it was done, details should be added to Methods. If otherwise, such analysis would be warranted.

* They're confidence intervals and this has been added – thanks.
* ICERs are a ratio of two different random variables and hence their confidence intervals
* can have strange properties. The probability density for the incremental QALY spans zero:
* above zero, a falling ICER is a good thing; below zero it is not. We are not sure how any
* reader can expect to interpret a confidence interval for an ICER in this case, or quite what
* it adds whenever the uncertainty isn't confined to a single quadrant of the cost-
* effectiveness plane.
*
* As an illustration, take the case where we look at adjustments for baseline EQ-5D. Over
* 10000 bootstrapped observations the ICER is £9K; if we look at the ICERs though as
* coming from a distribution, the estimated mean would be £30K per QALY, with a standard
* deviation of over £2,000,000 per QALY. This is because there are a few observations
* where the incremental QALYs are close to zero and the estimates of ICERs explode. We
* did not report confidence intervals for ICERs as we did not believe it was robust and
* this example probably justifies this view.
*
* As a measure of uncertainty in the ICER was requested, we've added an additional table
* that contains net monetary benefit with a cost-effectiveness threshold of £20k per QALY.
* Probably unsurprisingly, the 95% CI for NMB span zero in all the scenarios considered as
* is almost inevitable within economic analyses.
*
* Bootstrapping here is slightly fraught given the potential differences in baseline quality of
* life. Whilst it is possible to selectively sample to obtain a more balanced sample at
* baseline, our experience having tried to set out the methods for this is was difficult to
* convince reviewers that a robust analysis had been performed and a classical analysis was
* instead requested! As such, we've concentrated on classical analysis in this paper –
* although there is some bootstrapping from a seemingly unrelated regression now. We
* may look to compare and contrast the different methods available in a future paper.

2. P.6, 2nd para: Data on personal costs were collected but were not used in the analysis, which is due to the perspective defined in p.4 (NHS and PSS). However, it is not clear why this perspective was chosen. Who is the audience of this study? Are patient's costs not relevant for decision-making? It would be useful to see results of additional scenarios that include patient's costs, especially if the main cost difference between THA and RSA lies in the post-surgery period.
*

* This has been added as a sensitivity analysis. Briefly, there is little evidence of an effect
* here – whilst a considerable is incurred privately, the resurfacing arm only costs an extra
* £63 over the course of the first year post-treatment (P = 0.967). It is beyond the scope of
* the paper to debate the relative merits and information provided by alternative
* perspectives on costing.

3. P.8-9: RSA was more costly "due to further inpatient care after initial discharge (£279) and outpatient care (£83)" (p.8, 3rd para), and "Those in the RSA arm had significantly more outpatient visits than those in the THA arm (5.155 vs. 3.063, P = 0.0054). Here, both the number of physiotherapy sessions and the use of DVT assessments were significantly higher amongst this group (P = 0.002, P = 0.011)" (p.9, 1st para). The higher cost required for RSA post-surgery gives an impression that the health state in this arm was not as good as those in the THA arm. However, QLAY is better in the RSA arm.
Some discussions on this disconnect should be added in the Discussion.
*
* We have added a couple of sentences to state that it is unclear why this might be the case.
* Whilst the RSA group achieves slightly better health outcomes and requires more
* services, this may be due to heterogeneity in outcomes; if resurfacing works well for
* most but poor for some, then this might produce this type of phenomenon. If so, this
* emphasises the need to follow patients up in the longer term.

4. P.11.1st-2nd para: Given the rising debate over the 'potential' side-effects of metal-on-metal implant [1] and the unavailability of other types for resurfacing at this time, this could also have implications on the long-term QALY as compared to THA that has various options. This issue warrants addition to the Discussion.
*
* The literature on this broader than we can expect to cover within this paper. The reviewer
* should note that the current UK recommendations on MOM THA do not prevent the use
* of resurfacings. Whilst there are outstanding concerns about ion release, this is not
* something that we can do much about and the issues are dealt with in greater depth
* within the BMJ clinical paper. In the discussion we have added the following:
* On the benefit side of the equation, the impact of extending the time period is unclear
* as RSA may improve quality of life in the short term but lead to a quicker deterioration
* once revisions are necessary, or require additional monitoring or revisions by
* virtue of its 'metal-on-metal' nature.
*
* We have added a new sensitivity analysis that considers the impact of removing the MOM
* from the available options in THA.
* The current recommendations against the use of metal on metal THA prostheses is
* addressed by setting all 'metal on metal' implants to missing, estimating which alternative
* THA prosthesis each patient will receive using multiple imputation, and considering the cost
* implications of these estimates.
*
* Overall, since more ceramic on ceramic implants are used than metal on polyethylene, the
* net effect of this analysis is an increase in THA costs and a reduction in the incremental
* costs of RSA. We have considered this further in the discussion as:
* Where the sensitivity analysis assumed the use of the cheapest metal-on-
* polyethylene implants (without incorporating a possible impact on quality of
* life), RSA was no longer cost-effective within-trial. However, this is somewhat
* unrealistic to assume, as the main alternative to metal on metal THA implants
* appears to be the more expensive ceramic on ceramic type. Restrictions in
* the use of MOM THA implants within the UK are likely to lead to more of

* these (likely) less cost-effective implants being used, and so an increase in
* the cost-effectiveness of resurfacing implants.

5. P.10 1st para: "…that RSA is the most cost-effective option….". The analysis is comparing only one intervention (RSA) incremental to the comparator (THA), so 'the most' sounds odd (simply 'a'?).
*
* This has been fixed. Note that given the modified assumption used when costing prostheses and the correction of the error when costing LOS, this sentence has changed considerably.

6. Many abbreviations are not spelled-out when they appear first in the text that should be added (e.g. QALY, NICE, RCT, THA, RSA, NHS, HCHS, etc.)
*
* This has been done.

References
1. Cohen D. How safe are metal-on-metal hip implants? BMJ. 2012;344:e1410.

Reviewer: Helen Dakin
Researcher
University of Oxford, UK

I have no competing interests, but am involved in trials on knee replacement.

The manuscript describes the statistical methods for some but not all analyses presented in the paper.

Mostly the statistical methods are appropriate, although I am concerned that the analysis underestimates uncertainty by ignoring uncertainty around missing values in adjusted analyses and not capturing uncertainty around implant costs in base case figures.
*
* We believe that this has been addressed (details below).

The report addresses most points on the checklist, although many are addressed very briefly, or by citation.

Results are discussed in the light of the clinical results from the same trial, although no previous evidence on the cost-effectiveness of hip resurfacing is discussed. It is unclear whether this is because no previous economic evaluations (model-based or trial-based) have been done.
* We did not find any trial-based evaluations in our systematic search (not reported in the
* article), and the (very few) model-based evaluations that have been done inevitably focus on
* longer term issues that cannot be a natural focus of this paper. Thus existing literature
* cannot easily inform this paper.

I assume that the research ethics are addressed in the main clinical paper.
*
* Yes. This information was provided to the BMJ Open where requested in the
* submission process and we expect this will appear automatically. We added a reference to this
* within the paper.

The analysis appears to have generally been well-conducted and the manuscript includes most of the information listed on the checklist (albeit briefly). However, it appears that the analysis may underestimate (or ignore) uncertainty around missing data and implant cost.
*
* We have attempted to deal with these issues as best we can in the manuscript.
* Please see below.

General comments
• The costing methods appear to assume that operation time for the primary procedure and the incidence and cost of complications and re-operations occurring before patients are discharged from hospital are the same for both groups. It would be useful to discuss the extent to which this is true and to acknowledge this as a limitation.

* The information on minor complications etc within the initial operative period is picked up
* already within the length of stay, we do not believe it requires further comment.
* The total 'operative time' is picked up within the hospital system and id defined by when * patients enters the anaesthetic room to when they leave the operating the
* theatre. There was no real evidence of any difference here (both are about 2.5
* hours). This wasn't explicitly costed but a paragraph has been added to the discussion.
* Beyond the issues surrounding the definition of THA, the trial is inevitably unable to
* consider all possible cost items. The trial did not explicitly consider any differences in
* operative time between the RSA and THA arms; no difference was expected and an
* informal analysis of the data suggests very similar operative times between the arms.

• The results are extremely sensitive to the type of THA conducted and the cost of hip prostheses for total hip replacement varies substantially between the three main types of prosthesis. From my experience in knee replacement, there is probably also a large amount of variation within each category and potentially between different types of resurfacing implant. It appears that the uncertainty or variability around implant cost is ignored in the analysis. It is unclear why it wouldn't have been possible to estimate implant costs separately for each patient and do subgroup or adjusted analyses to estimate the impact of implant type on costs and QALYs.
*
* The trial upon which the economic evaluation was based were pragmatic in the sense that
* surgeons had freedom over which implant type (e.g. metal/metal, metal/polyethylene,
* ceramic/ceramic) to use, and the precise implant to use. In the methodology used for the
* initial submission, NHS unit costs were used for the THA arm of the trial. These figures
* include prosthesis costs and are based on very large number of procedures, there is
* relatively little uncertainty surrounding the mean figure and this was used without
* accounting for any uncertainty due to prosthesis costs. The resurfacing arm was then
* modified to incorporate the average difference in prosthesis costs between the arms, and
* we acknowledge that this may (very slightly) underestimate the uncertainty in the
* resurfacing arm as we cannot know by how much the RSA and THA arms differ.
*
* Following your query we have revised the methodology for costing prostheses. As before,
* the basic cost of admission for a primary hip replacement is based on NHS costs.
* Using the data from those receiving THRs, we calculated an average prosthesis cost using
* the three main types covering the vast majority of cases. This average is deducted from
* the NHS reference cost estimate to provide a prosthetic-exclusive cost. In those cases
* where the implant type is not available or not captured within the three main THA types,
* the prosthetic cost is treated as missing and is estimated within the multiple imputation.
* In this way, variation within the sample is picked up, although we acknowledge this
* won't necessarily pick up within-class variation in the cost of each device.

* This evaluation was also unable to consider the impact of variation in cost within each
* type of prostheses (i.e. within the three THRs, or beyond the single RSA used in the
* trial) as this information is not generally available.

In addition, it is unclear whether metal-on-metal implants are still used in UK clinical practice and (if so) whether it is valid to include such implants in cost-effectiveness estimates.
*
* As this is a report based on an existing clinical trial, we believe that it would be
* inappropriate to retrospectively modify the trial data within the base case. Metal on
* metal RSAs have not been withdrawn from clinical practice, and so the resurfacing arm
* must remain as-is.
*
* In order to address your point, we have incorporated an additional sensitivity analysis
* that treats prosthesis cost data from MOM as missing and so imputes data. Note that
* few of these metal-on-polyethylene implants are used, so that if metal-on-metal THA
* implants were excluded, we find lower incremental costs as the more expensive
* ceramic on ceramic implants are more common. This point is made explicitly in the paper:
* Where the sensitivity analysis assumed the use of the cheapest metal-on-
* polyethylene implants (without incorporating a possible impact on quality of
* life), RSA was no longer cost-effective within-trial. However, this is somewhat
* unrealistic to assume, as the main alternative to metal on metal THA implants
* appears to be the more expensive ceramic on ceramic type. Restrictions in
* the use of MOM THA implants within the UK are likely to lead to more of
* these (likely) less cost-effective implants being used, and so an increase in
* the cost-effectiveness of resurfacing implants.

Adjustment of QALY estimates for baseline utility is always recommended when there is marked imbalance at baseline; it would therefore be extremely useful to see the justification for using unadjusted values as the base case and to present results against metal-on-polyethylene implants and CEACs with and without adjustment for covariates.
*
* The baseline differences were not found to be statistically significant in the main trial
* analysis, so that whilst the likelihood is that they do exist it is a judgement call as to
* whether they can be considered ‘marked’. To treat them as definitively different
* introduces inconsistency between the published trial data and the health economic
* analysis that we believe could be confusing to some readers.

The paper does not clearly describe statistical methods for all results presented in the paper. In particular, it is unclear how the confidence intervals in Tables 1-3 and the p-values on pages 8-9 or WebExtra Table 2 are calculated.
*
* Thank you – in a desire to try to produce text that is accessible to clinical readers, we
* have clearly provided too little methodology here. We have added additional detail:
*
*quality of life
* Where comparisons between the RSA and THA arms are based use non-imputed data,
* a two-sample t-test assuming equal variances is used.
*
* Resource use and valuation
* Where statistical tests analyse resource usage data, t-tests are used to test for
* differences in expected usage (assuming equal variance and non-imputed data).

* Elsewhere, the analysis uses regressions based on imputed data.
* A new 'missing data' section is added containing (mostly moved) details for the STATA
* ice command
*
* The same form of words is (again) used when identifying uncertainty relating to net
* monetary benefit.

It is also unclear whether the CEAC shown in Figure 1 is adjusted for baseline characteristics and (if not) how it was calculated, since the methods section only describes the methods used to calculate adjusted CEACs.
*
* As per your comment below (thanks), we've now used the sureg command and so can
* use imputed data throughout. Given the simplification of the methods in this section,
* we don't believe that there is likely to be any ambiguity here.

• Page 6, lines 19-35: The description of missing data suggests that only data on total QALYs were used in multiple imputation, rather than utilities at each timepoint. This implies that patients with utility data at some but not all of these four timepoints were considered to have completely missing data on QALYs. If this is indeed the case, it would be useful to justify why this approach was used, since it could waste a large amount of data and means that one of the covariates explored in subsequent analyses (baseline utility) is omitted from the imputation model, which may introduce bias. It is also unclear from this paragraph whether imputation was conducted on total cost excluding length of stay, or whether different cost components were considered separately.
*
* We acknowledge that this could occur, although it did not do so here. We had complete
* QALY data for 118 of the 122 patients included in the analysis. At baseline, 3 months, 6
* months and 1 year we had data for 122, 119, 122 and 119 patients. It is highly unlikely
* that much data will have been lost. The 'N=118' and 'N=122' figures have been added
* into Table 1 (and the former in the text too) to make it clear that little data has been lost.

• On page 7 (lines 43-47), it is stated that methods are not established to do seemingly-unrelated regression on multiply imputed data. I have occasionally managed to run suest after mim by saving estimates with "mim, storebv", although the command is temperamental and it may be possible to do this using mi estimate. However, an alternative would be to use bootstrapping, as we have used previously (BMJ Open. 2012 Jan 30;2(1)). Averaging over imputed datasets before running the regression analyses that are used to estimate CEACs and 95% CI means that all uncertainty around the imputed values is excluded from the analsyis. Since the methods used to calculate unadjusted 95% CI are not described, it is unclear whether this problem affects the base case analysis as well as the adjusted figures.
*
* Thank you for this – and we agree that this was a concern. We've taken your suggestion
* but have found that it's a little temperamental.
* On fuller inspection though, we've got it to work using mim, category(fit): sureg …
* As such, we haven't had to 'collapse' the data and we hope that we have addressed your
* (understandable) concerns about underestimating uncertainty.

Specific comments
• Page 3, lines 11-23: Centre-specific implant costs were used for to adjust HRG costs for the primary hospital stay. It is unclear whether the HRGs costs used reflect those for this trust, or the average costs across all English hospitals.
*
* The costs here reflect the average across all English hospitals and this has been made clear

* in the text.

• Page 3, lines 35-9: There is typically a large difference between list prices and the price that hospitals pay for prostheses and substantial variation between hospitals in the discounts secured. Were the costs supplied by the hospital trust discounted prices or list prices?
*
* We acknowledge that this is a potential issue – we have requested more information and
* believe that these are list prices. In the discussion we acknowledge this issue to a greater degree.
* Whilst the list price of the Cormet prosthesis is similar to other prostheses
* available locally, prices are hospital-specific and so some caution is warranted
* when seeking to generalise findings to other locations.

• Pages 5-6: The costing section of the methods could be made clearer. It would be useful to say up front what resource data were collected for each patient and, potentially, which information came from patient questionnaires rather than hospital records.
*
* The only information to come from hospital records was the LOS, treatment received/
* implant type. This has been made clearer.
* Data regarding length of stay and implant received were obtained from hospital
* records, with the remainder of the costing information obtained from patient-reported data.

Pages 5-6: The main manuscript does not explicitly state that the analysis includes all resource use incurred by trial participants within 12 months of THA (not just medications, hospitalisations, ambulatory consultations and aids or adaptations related to the study hip).
*
* The analysis in the original submission did not include patient costs, and so this additional
* cost incurred by the trial participants was not incorporated into the analysis plan. Given
* the additional sensitivity analysis requested by Reviewer 1, this has been changed.

Page 6, line 3: It is unclear from the text or appendices what unit costs were provided by “relevant RCTs” or which RCTs were used.
*
* This has been clarified in the text. In Table 1, the appropriate references are to the unit
* costs for acupuncture and a chiropractor and we believe these are sufficiently clear.

Quantities of resources are not reported separately from their unit costs (although unit costs are reported separately from quantities).
*
* Web Table 1 provides unit costs, with Web Table 2 providing the quantities of each
* resource used. We apologise that Web Table 2 had an incorrect header which appears to
* have cause unnecessary confusion.

• There is no indication of how many patients had missing data on costs or QALYs, which makes it difficult to interpret the results of the complete case analysis or assess the impact of the methods used to impute missing data.
*
* The data – private costs aside – aren’t too bad. A column has been added to Table 2 to
* identify the level of imputed data; we have at least N=109 (89%) on all variables and
* complete data on N=98 (80%) patients. Private costs aside, we have complete cost data
* on N=101 (83%) patients and complete QALY data on N=118 (97%) patients.

• Page 8: when presenting the mean difference in costs and QALYs, it would be useful to specify that

these figures are not adjusted for baseline characteristics and to give 95% CI around the point estimates, so that the reader can see that these differences are not statistically significant.
*
* The 95%CI are given in the tables, which is probably sufficient (we'll check with the
* editor re: house style on this point). In terms of the adjustment – given the section that
* they're in, this should be clear enough. We've revised the labelling in the table to hopefully
* make this clearer.

• Discussion: For a single-centre study such as this, generalisability is a key issue. In particular, do most other centres use the Cormet hip resurfacing implant and is the price of the Cormet implant similar to implants used elsewhere? Are the types (and prices) of THA implants used at this centre comparable with those used elsewhere?
*
* We cannot know for certain as these prices are normally hospital-specific. However, our hospital * is a high-volume arthroplasty centre and is representative of such centre where most 'young * person' arthroplasty is performed … the best we can do is to add a caveat in the discussion.
* The clinical trial upon which this analysis is based used a single type of Cormet
* prosthesis that has been used in the UK for around 15 years. As such, our findings
* are not necessarily generalisable to other types of resurfacing and we cannot
* identify the most cost-effective type of resurfacing as this is beyond the scope of the
* trial. Whilst the list price of Cormet prosthesis is similar to other prostheses
* available locally, prices are hospital-specific and so some caution is
* warranted when seeking to generalise findings to other locations.

• It would be useful to mention the results of the primary endpoint of the clinical trial in the introduction or methods (as per point 9 on the checklist).

* This has been added into the introduction.

Minor comments:
• Several acronyms are introduced without being first defined, including RSA and THA (page 4, lines 14-16), RCT (pg 3, line 45), as well as NICE, QALY and NHS.
*
* Thanks. These have been fixed.

• Abstract: "Missing data was assessed" should read "Missing data were imputed" or "dealt with".
*
* Done.

• Page 6, line 26: Should this read "[randomised] treatment allocation" rather than "treatment received"?
*
* Where we have detail on the treatment received, we've used this in the treatment
* allocation although they are analysed according to the allocated treatment. In the event
* of a protocol violation, what people get is a better predictor than the arm they were
* allocated to.

• Page 6, line 27: it would be clearer to write out "Oxford Hip Score" and "Harris Hip Score" in full (particularly as "Oxford Knee" would normally imply a particular type of implant, rather than the OKS).
*
* Done
•

Page 9, lines 45-7: the wording of this statement is currently misleading and implies that the impact of gender and baseline EQ-5D was also to increase QALYs by 0.059.
*
* Yes, that was pretty ugly. Thanks

• Page 7, lines 8-11 imply that the complete case analysis represents a “per protocol analysis”, but in Table 3, per protocol results are presented separately from complete case.
*
* This has been split into two sentences to make it clear that they’re separate analyses. Thanks.

• Table 3: It would be useful to state that the values in brackets (presumably) indicate 95% CI. It would also be useful to explicitly state which implant was assumed in line 12.
*
* Done. Thanks.

• Tables 2-3: it would be useful to state that costs are in £ Sterling.
*
* This is stated in the paper, and I’m not sure it adds much. We’ll ask the editor for this, too.

• Web Extra Table 1: it is unclear whether the list of drugs and prices includes all drugs considered in the analysis, or just the most commonly used.
*
* The patients provided the data – so this is the full list.

Reviewer: Callum McBryde MD FRCS (Tr&Orth)
Locum consultant orthopaedic surgeon
Royal Orthopaedic Hospital and University Hospital's Birmingham, United Kingdom

I think that the main message from the paper could be made more clear in the discussion. I think a very straightforward rewrite to highlight the main findings in slightly more simple terms would be useful but otherwise a very interesting paper.

*
* Thank you for your comments. We’ve taken a look at the discussion and have tried to
* highlight the main findings more clearly. We don’t want to firm up any findings
* in this study and need to concentrate on caveats as it’s all very preliminary within a
* 1-year analysis. We have a subsequent paper that attempts to deal with things more
* definitively.
*
* We feel the first and final sentences of the discussion outline the status quo pretty clearly:
* In comparison to standard total hip arthroplasty, hip resurfacing arthroplasty appears to
* provide a modest QALY gain for a modest sum within the first 12 months from surgery;
* whilst the additional costs of RSA are statistically significant, the additional benefits are not.
*
* Despite the conclusions of the within-trial analysis, it seems clear that until such work
* is done and further data is available, the cost-effectiveness of resurfacing arthroplasty
* in a UK context remains potentially promising but as yet unproven.
*
* The major changes in the discussion relate to setting out some of the caveats to the
* findings more clearly, and talking about the impact of recent changes to recommendations
* regarding MOM total hip arthroplasties.

Reviewer: Professor Luigi de Palma
Cattedra di Ortopedia e Traumatologia, Universita` Politecnica delle Marche, Azienda Ospedaliero-Universitaria,Ospedali Riuniti di Ancona, Via Conca, Torrette, 60100 Ancona, Italy

The Authors report their experience in cost-effectiveness analysis of total hip arthroplasty versus resurfacing arthroplasty.
The manuscript is interesting and well written, but during the lecture of the manuscript some questions may have arisen.
In my opinion after 12 months the Authors report to early results for this kind of evaluation, in fact, generally speaking, the clinical results are evaluated after 2 years to be considered expressive.
*
* In order to look at the cost-effectiveness, we really need 5, 10 or 15 years to work as a
* starting point, with a time horizon much further off into the future. This paper is intended
* to provide an intermediate step and is in no way meant to be definitive.
*
* We currently have 79% of our data at 2 years and at this point there is no difference in
* clinical function (no difference was detected at 1 year either), so that there's no sign
* of deterioration in relative function as yet. There are no specifically 'economic' data in
* there, and as the specific MoM concerns tend to happen slightly later, it's unclear
* that two years will give us that much more information.

The Authors report that " this study has a pragmantic nature ", for these reason in cost-effectiveness analysis the Authors have to considerer the cost of complications. Finally, it could be very interesting to report data from patients (indications, age….) to explore the patients cohort.
*
* Thanks. For brevity, we have noted in the manuscript that details of the cohort are
* provided in the main BMJ clinical paper. We have also added the following:
* The sample was representative of the broader population undergoing resurfacing in
* the UK; no significant differences were identified between those who took part and
* those who were eligible but chose not to take part. Further details on both the
* ethical approval for the study and the demographics of the patients are provided
* in the clinical paper.

No competing interests.

VERSION 2 – REVIEW

REVIEWER	Hideki Higashi, PhD Research Fellow School of Population Health University of Queensland Australia I have no competing interests.
REVIEW RETURNED	23-Jul-2012

REPORTING & ETHICS	Ethics addressed in the RCT paper (it was made clear in text)
GENERAL COMMENTS	The authors have generaly well responded to the previous comments. I have only one minor comment: "As is normally the case in economic evaluations," (p.12 last para) is not true. This should be deleted.

REVIEWER	Helen Dakin Senior Researcher Health Economics Research Centre, University of Oxford UK
REVIEW RETURNED	12-Jul-2012

GENERAL COMMENTS	I congratulate the authors on making a large number of amends to their manuscript in a short space of time. The revised manuscript is now clearer and more complete, although a few other amends would help ensure that the methods and clinical implications of the paper are fully described. • The safety issues around MOM hips could have profound implications for this paper: particularly as the Cormet resurfacing system evaluated in the trial is MOM and as there appear to be no other resurfacing materials available at present. The sensitivity analysis excluding MOM replacements is very worthwhile as it shows how the cost-effectiveness of MOM RSA would change if the mix of implants used in the comparator arm changed. However, this sensitivity analysis does raise the larger question of whether resurfacing would remain a real option if MOM were discontinued entirely. The safety concerns around MOM hip resurfacing should be mentioned in the abstract along with the caveat about short follow-up so that readers can weigh up the short-term quality of life and efficiency benefits against the medium/long-term risks. The authors should also discuss this issue briefly in the discussion and comment on whether (or to what extent) their findings could generalise to any non-metal-on-metal resurfacing technologies that were used in the past or may be developed in the future. • Page 5, lines 16-19: The authors should explicitly state whether the analysis only included the cost of inpatient care, ambulatory consultations and medications that are related to the study hip, or whether they also included the costs of hospitalisations, consultations and drugs for other conditions: e.g. were drugs for heart disease, GP visits for flu and hospitalisations for cancer included in the costing analysis or excluded? From the list of drugs in Web Table 1, it appears that the analysis included all costs regardless of whether they were related to the hip surgery, but it would be useful to explicitly state this. • Page 5, line 28: It would be useful to insert the word “metal-on-metal” after the word “Cormet” to make it clear earlier in the document (as well as in the discussion) that the RSA technology is also MOM • Page 5, line 33: replace “costs” with “list prices”? • Page 9, “Adjustment for potential baseline differences” section: The authors should explicitly state (e.g. on line 40) that the baseline adjustment is only done in a sensitivity analysis (not in the base case analysis) and state the justification for not adjusting for baseline utility in the base case analysis. • Page 14, line 53-page 15 line 11: The sentences added don’t appear to follow from the ones that precede them and appear to relate to a different sensitivity analysis. • Table 3: The contents of some cells in this table are incompletely shown as they extend below the bottom of the cell.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

I congratulate the authors on making a large number of amends to their manuscript in a short space

of time. The revised manuscript is now clearer and more complete, although a few other amends would help ensure that the methods and clinical implications of the paper are fully described.

* Thank you for your quick review of the paper – we do feel that
* it was improved by your comments.

• The safety issues around MOM hips could have profound implications for this paper: particularly as the Cormet resurfacing system evaluated in the trial is MOM and as there appear to be no other resurfacing materials available at present. The sensitivity analysis excluding MOM replacements is very worthwhile as it shows how the cost-effectiveness of MOM RSA would change if the mix of implants used in the comparator arm changed. However, this sensitivity analysis does raise the larger question of whether resurfacing would remain a real option if MOM were discontinued entirely.

* Clearly, if the MOM resurfacing was restricted generally in the
* way that MOM THAs have been restricted in the UK, then we
* would agree the point of the paper in the UK is lost as only
* the non-MOM THA analysis stands. The only established alternative to
* these implants would then be to refuse surgery and as this is
* rarely opted for long-term in clinical practice (once it is
* established that an operation is necessary), there is little
* natural history data on its consequences.

The safety concerns around MOM hip resurfacing should be mentioned in the abstract along with the caveat about short follow-up so that readers can weigh up the short-term quality of life and efficiency benefits against the medium/long-term risks.

* We have modified the conclusions of the study to include the
* statement that:
* “The short-term follow-up in this trial should be noted, particularly
* in light of the concerns raised regarding adverse reactions to metal
* debris from MOM bearing surfaces in the longer term.”

The authors should also discuss this issue briefly in the discussion and comment on whether (or to what extent) their findings could generalise to any non-metal-on-metal resurfacing technologies that were used in the past or may be developed in the future.

* The discussion already includes the statement that as we extend the
* timeframe of analysis, we may require additional monitoring or revision
* due to the metal-on-metal nature of the resurfacings that we consider.
* This seems to suffice; whilst we acknowledge the concerns,
* any further response is potentially alarmist if the regulatory bodies have
* not seen fit to act on the availability/clinical use of MOM resurfacings.
* In our approach we are trying to reflect the apparent clinical consensus
* that resurfacings appear to remain a treatment option, albeit with
* some specific concerns attached.
*
* In terms of generalisations – this is always slightly difficult.
* There are some early trials of ceramic on ceramic (CoC)
* resurfacings but these provide relatively little data. Historic
* MOM resurfacings do exist, but these tended to perform
* very poorly relative to current versions, and it is unlikely

* that these would to be considered as clinical options. The
* current evidence base is poor, they have not been adopted
* as yet, and there is no real indication of performance relative
* to MOM resurfacings, let alone robust head to head studies.
* As such, it is very difficult to generalise reliably – although
* this problem is common to very many economic evaluations,
* and especially where new treatments are being developed.
*
* Clearly, this is all more in-depth than we could reasonably try
* to incorporate into the discussion. For the purposes of brevity,
* we have revised a paragraph to include the statement that:
*
* "We note also that our findings are not necessarily generalizable
* to other types of resurfacing, including emerging technologies
* such as ceramic on ceramic resurfacings. Whilst the cost-effectiveness
* of these newer treatments may differ from current resurfacings, we
* cannot identify the most cost-effective type of resurfacing as this was
* beyond the scope of the trial and relatively little data exists on which to
* base even a preliminary estimate. To the degree that this may prove
* possible, it is an issue for subsequent decision analytic modelling."

• Page 5, lines 16-19: The authors should explicitly state whether the analysis only included the cost of inpatient care, ambulatory consultations and medications that are related to the study hip, or whether they also included the costs of hospitalisations, consultations and drugs for other conditions: e.g. were drugs for heart disease, GP visits for flu and hospitalisations for cancer included in the costing analysis or excluded? From the list of drugs in Web Table 1, it appears that the analysis included all costs regardless of whether they were related to the hip surgery, but it would be useful to explicitly state this.

* We've added a couple of lines into Page 5 to clarify this point as:
*
* "The analysis considered inpatient and outpatient attendances for
* all reasons, and requested details of other resource usage only
* where it related to pain or hip surgery."

* We excluded a couple of items that were provided by patients
* where Matt Costa (as PI on the RCT) felt that there was no
* plausible connection with pain or hip surgery. The drugs in Table 1
* deal relate to pain relief, antibiotics (which can be given for a
* variety of reasons, and can't be automatically said to be
* irrelevant), sleeping pills (relevant if pain is preventing sleep),
* arthritis (relevant to pain and/or the operation), and
* hydrocortisone (dermatitis at wound sites). We don't feel that any
* paper changes to the paper are necessary to justify these specific
* items.

• Page 5, line 28: It would be useful to insert the word "metal-on-metal" after the word "Cormet" to make it clear earlier in the document (as well as in the discussion) that the RSA technology is also MOM

* We have done as you've requested here.

• Page 5, line 33: replace “costs” with “list prices”?
* Done.

• Page 9, “Adjustment for potential baseline differences” section: The authors should explicitly state (e.g. on line 40) that the baseline adjustment is only done in a sensitivity analysis (not in the base case analysis) and state the justification for not adjusting for baseline utility in the base case analysis.

* We have added a new paragraph at the beginning of this section,
* and slightly modified the start of the succeeding paragraph to
* leave the flow of the paper intact.

* “The base case analysis was conducted to allow for comparability
* between this within-trial analysis and the reporting of the main
* RCT. These quality of life and gender-based analyses are
* conducted as sensitivity analyses to allow comparability with the
* main RCT, which did not find a significant difference in baseline
* quality of life and did not test for an interaction between efficacy
* and gender. Given that these issues may be important within the
* economic evaluation, they are considered as sensitivity analyses.”

• Page 14, line 53-page 15 line 11: The sentences added don’t appear to follow from the ones that precede them and appear to relate to a different sensitivity analysis.

* In order to clarify the earlier example, we have added a sentence
* relating that appears just after the metal-on-polythene case is
* introduced under costs. “However, if we consider both types of
* non-MOM implants (ceramic-on-ceramic and metal-on-polythene),
* this difference disappears entirely as the non-MOM implants were
* slightly more expensive on average than the MOM ones.” This
* provides, we hope, a clearer link to the two non-MOM sensitivity
* analyses. Accordingly, we’ve also cut some of text in the
* discussion and made it slightly clearer that both analyses are
* considered in light of the ban on MOM THAs.

* The section in the discussion now reads:
*
* “Where the sensitivity analysis assumed the use of the cheapest
* metal-on-polyethylene implants (without incorporating a possible
* impact on quality of life), RSA was no longer cost-effective within-
* trial. However, these implants were used relatively rarely in
* practice, and the main alternative to metal on metal THA implants
* was the more expensive ceramic on ceramic type. Restrictions in
* the use of MOM THA implants within the UK are likely to lead to
* more costly THA implants being used, and so a net increase in the
* cost-effectiveness of resurfacing by comparison.”

• Table 3: The contents of some cells in this table are incompletely shown as they extend below the bottom of the cell.

* Thank you, this has been corrected.

Reviewer 2.
Thank you for your quick review of the paper.
Ethics addressed in the RCT paper (it was made clear in text)

* We're not sure if this is a comment that needs addressing, but
* we've slightly modified the statement on Page 4 to make clear
* that the ethics details etc are provided in both the RCT's protocol
* and reporting papers.

The authors have generally well responded to the previous comments. I have only one minor comment: "As is normally the case in economic evaluations," (p.12 last para) is not true. This should be deleted.

* We have amended the statement as suggested as it is not critical.
*
* We do query the reviewer's interpretation slightly as whilst many
* studies have significantly positive/negative incremental costs
* and benefits, the picture provided by net benefit (at least
* at relevant thresholds) is normally much more mixed. The
* reviewer's comment, if true, suggests that most studies have
* power to detect classically significant results – and we do not
* believe that this holds for either most EEACTs or most decision
* analytic models.