PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A RETROSPECTIVE COHORT STUDY OF THE PERFORMANCE OF THE PINNACLE METAL ON METAL (MOM) TOTAL HIP REPLACEMENT: A SINGLE CENTRE INVESTIGATION IN COMBINATION WITH THE FINDINGS OF A NATIONAL RETRIEVAL CENTRE
AUTHORS	Langton, David; Sidaginamale, Raghavendra; Avery, Peter; Waller, Sue; Tank, Ghanshyambhai; Cooke, Nick; Logishetty, Rajesh; Nargol, Antoni

VERSION 1 - REVIEW

REVIEWER	Mr Pedro Foguet UHCW NHS Trust Coventry UK
REVIEW RETURNED	28-Mar-2015

GENERAL COMMENTS	Excellent piece of work. Relevant and clear message. Well written and elegantly questioning the results of some high profile American authors that are too close to the manufacturers to be considered unbiased yet they often get more credit than deserved in the Orthopaedic journals. I have recommend that it is accepted for publication. My only question to the authors, and I think that I know the answer in advance, is whether they did the effort of routinely obtaining consent from the patients whose retrieved implants were sent for analysis? This is relevant given the significant proportion of these prosthesis that are failing prematurely due to design and manufacturing issues that have serious medico-legal implications.

REVIEWER	Mr Michael Whitehouse Consultant Senior Lecturer in Trauma and Orthopaedics University of Bristol, UK I have provided teaching sessions on cemented hip replacement for DePuy International Ltd for which my institution has received payment.
REVIEW RETURNED	09-Sep-2015

GENERAL COMMENTS	The comments below are in the order in which they were encountered in the manuscript: The single surgeon nature of the series limits the generalisability of the work and should therefore be listed as a limitation rather than a strength. No ethical approval for the study is mentioned. Given the nature of the work, the requirements for follow up in this population and the

	jurisdiction of the centre this study would fall under the category of service evaluation. This needs to be clearly stated and the NRES guidance cited in order to justify the lack of ethics committee approval. It should be made clear in the background that the first design factor mentioned in relation to failure of the ASR (a shallow acetabular component (low coverage arc)), does not apply to the Pinnacle. It may be better to remove this factor given the general readership of the journal. As per the instructions to authors "Only papers published or in press should be included in the reference list." The 8th reference is a conference paper that has not been published as a paper as far as I am aware. If the authors wish to cite unpublished data, they should provide a link or source for the data referred to such that it can be reviewed by the reader. "Surgical factors: Cup orientation has been shown to affect wear rates/metal ion release in MoM arthroplasty[9] and devices with lower coverage arcs and sharper articular rims are particularly sensitive to cup position.[15]" the reference to lower coverage arcs and sharper articular rims is not useful as it does not apply to the design considered in this manuscript but to the ASR, another metal on metal implant system with design features not shared with the Pinnacle Metal on Metal system. To avoid confusion for readers that may not be aware of this, I would remove this unless the authors have established a particular link they wish to make, in which case this should be explicitly stated. "devices with no excessive wear with associated ARMD are commonly found in patients with bilateral devices implying a process of sensitisation (unpublished data)." there are a number of published studies looking at bilateral devices, is this statement reflected in any of these studies or this comment unique to the authors' unpublished data? Again, I would prefer that unpublished data referred to is made accessible to the reader, either by providing an online source or providing the data as an appendix. The information regarding the number of surgeons performing the surgeries is inconsistent. the abstract states "All patients implanted with a DePuy Pinnacle MoM hip prostheses by one of the senior authors" but the background states "The aim of this study was to identify variables associated with early failure of the Pinnacle 36mm MoM system by a retrospective analysis of all patients implanted with this device by two experienced hip surgeons at our institution". How many surgeons performed the primary operations and how many were supervised by one or either of the surgeons mentioned? As the journal is targeted at a general readership, the link between the size of the acetabular shell and the thickness of the metal liner used with a 36mm metal on metal bearing should be explained. The authors describe what appears to be two separate phases of follow up - that between 2008 and 2010 where they were using cross sectional imaging and blood metal ion testing "routinely" and that after 2010 when a "full MoM hip recall" was commenced. What follow up was performed, in which individuals at what time points should be clearly stated for each of these phases. Provide a citation of the Harris Hip Score. Was this the only outcome measure used? Why are the results of this not presented? As the paper has been submitted to a journal with a general readership, the authors should clearly state their selection criteria for the use of the SROM stem in the under 70s after 2005. Were all cases of dysplasia and Perthes treated with one or was there selection criteria applied in this cohort and did any other patients receive the SROM. The authors' should also clarify the difference

	between the designs of the stems to explain their selections. When the authors describe their protocol for investigation of these patients, they should state clearly the section criteria they applied to select patients for revision surgery amongst the investigations listed. A flow chart or decision tree would be a useful figure to illustrate this. A patient flow chart should be provided showing the inclusion and exclusion criteria described and the flow of patients through the study to the final numbers. This should start with all patients undergoing primary hip arthroplasty in the unit during the period the procedures were performed. In order to provide context, the authors should also describe the revision rates in their unit for other designs. The authors state that they used 36mm bearings in all cases. They then state "The 160° sub-hemispherical bearing surface does not vary with the thickness of the liner.". According to the manufacturer's design rationale of the Ultamet metal liner and the figures provided in that, the bearing surface is 180 degrees for the 36mm metal liners and is not sub hemispherical. This should be clarified and a source for this information cited. I am concerned by the use of explant analysis to determine if components were outside stated manufacturing tolerances. By definition, these are own bearings and the majority of the cases revised in this series had been undergoing wear in vivo for a number of years. Changes from the original manufacturing specifications and measurements would therefore be expected, particularly at the lower end of the tolerance which is where the authors imply they saw variation. I do not think this is reliable and if the authors wish to determine if components were outside the manufacturing tolerances stated, they would need to look at unworn components. Why was pre and post 2006 implantation selected as a binary variable for the model rather than year of implantation being considered? I assume the authors had a reason for this categorisation. Liner size is given as a variable for the proportional hazards model but elsewhere the authors have stated it is the acetabular shell size, please correct or clarify which measurement was used and explain the implications of this (i.e. given the constant inner diameter, a smaller acetabular shell of a constant thickness results in a thiner liner). Given the authors were concerned about the risk of bias created by the introduction of 36mm metal liners for 50mm outer diameter shells from 2008 onwards, was the model repeated with and without the subsequently excluded 51 cases included to see if the results differed? If not, this should be done. It is not reasonable to state that non-attendees had well functioning prostheses as it is known that those that are lost to follow up have worse overall functional results than those that do. They may also have been revised in other units without the authors knowing and the data capture of retrieval registries is not 100% for patients who have primary hip replacements in their catchment area so revisions may have been performed that were not captured by the NRR. The range (I assume 95% confidence interval) is missing from the Predicted joint survival/survival at 8 years row of the pre 2006 column in table 2. There is a typographical error in the Median (range) blood Cr unilaterals (µg/l) row of table 3, the decimal point is missing from the p value. There is a typographical error in the Median (range) combined wear rates* (mm3 per year) row of table 3, the units have been included in each column containing results as well as the first column, which it has not in other rows.

	The authors state that "Taper wear was associated with a larger female taper angle and an increased femoral head offset". The taper diameters have been stated for the SROM (11/13) and the Corail (12/14) but the taper lengths and angles have not been stated. The authors should either give these values or clearly state which of the taper geometries has the highest angle as this can not be determined by the reader from the diameters alone given the length also varies. The authors give a p value for the difference between survival estimates on page 14 of 42 but have not mentioned what statistical test was used prior to this. They subsequently mention the use of a log rank test on page 15. Statistical tests employed should be described in the methods section before results of the tests are given. "Approximately 93,000 ASRs were sold globally, with around 6,000 reported in the NJR of England and Wales. The 2014 Annual NJR Report lists 11,871 Pinnacle implantations. If England and Wales represent the same proportion of Pinnacle as ASR implantations then it is not unreasonable to suggest that the Pinnacle MoM system has been implanted into over 180,000 patients globally, making it the most commonly used large diameter MoM THR in the world." this is not a reasonable statement to make. The Pinnacle system can be used with a variety of different bearing combinations (metal and ceramic heads; polyethylene, highly cross linked polyethylene, metal and ceramic liners). The implantation of 11,871 Pinnacle shells recorded in the NJR in 2014 can not therefore be reliably used to estimate the the number of Pinnacle shells implanted worldwide that have also had a metal liner in conjunction with a metal head implanted. The proper format for the citation of web sources has not been followed. For the Kaplan Meier figures, the 95% confidence intervals should be displayed and the numbers at risk given under the horizontal axis. If this involves splitting the into multiple panes, this should be done.

REVIEWER	Boomsma M.F. Isala hospital, The Netherlands
REVIEW RETURNED	16-Sep-2015

GENERAL COMMENTS	L.S., Interesting paper, nice write up. I really enjoyed reading this paper. This paper give more insight into the pathological mechanisms related to manufacturing and design. It also seems to give an overview of MoM THA issues because of a lot of relevant issues that are addressed. Nevertheless I would suggest to follow up on my major an minor comments that in my opinion would enhance the quality and readability of this paper for the wider range of readers readers of the BMJ. Major comments: 1:

	Page 7, could a decision rule (flow chart) be given regarding indication for revision be given as functional results of revision are very disappointing. Complaints is a very rough indicator in my opinion alone (off course...) to justify revision for example. So if one decides to revise, on what grounds does the orthopedic surgeon do this. Also high serum ion levels that can lead to cardiac arrhythmia's, brain deposition etc.. Image findings? Combination, 3 out of 3 positive? Please elaborate briefly in the introduction. A list of revised patients in an appendix regarding parameters and their individual outcomes that influenced the choice on revision could be extremely valuable for even the experienced clinician. How were complaints such as pain and discomfort quantified. HHS? How good is this test? Please elaborate. What were the imaging findings on US? Who performed the US, how much experience did they have? 2. In my opinion this paper in the BMJ, which is read by a general audience and not just orthopedic surgeons, would benefit if the authors elaborate in the discussion section on whether it is fair that the whole idea of metal on metal THA should be abandoned or that patient selection and better design and manufacturing, for example a unique patient tailored one piece without tapers design by means of pre-operative 3 D imaging for example could or should be investigated. In summary is there still a future for MoM THA prosthesis in the future? If so, under what conditions regarding patient selection, design and manufacturing? A bridge to future research could be inserted and would not be out of place at this point in time after many reports of THA related problems. 3. No quantification of the pathological capsular reaction is provided, nor taken into account for the analysis. Is this because of the use of ultrasound? Please elaborate on the reason for screening with US instead of CT and MR in the M and M, Also explain why you did not take into the account the expression of the disease by means of pathological capsular reaction. The reason that I am mentioning this is because imaging the pathological capsule by means of CT for example seems to be the strongest predictor for revision and correlates strongly with serum ion levels. 4 As revision rate of MoM cohorts is likely depending on the proportion unilateral/ bilateral patients, please add a survival curve stratified for hips of patients with a unilateral MoM hip replacement and hips of patients with a bilateral MoM hip replacement. As figures 'shell size and median co concentration' and 'Percentage of pinnacle bearing by year' are a comparison of seperate groups, it would be better to depict unconnected dots. 5. Discussion could gain readability from the adagium: what did we find

	find, what's actually new and how does this relate to the literature and then discuss findings that are consistent with other studies which are off course important, but not that new. 6. Please provide METC number. or the waiver... Some minor comments: page 2 line 20: mentions invitation of patients from one senior author, while page 5 mentions implantation of the device by 2 surgeons. page 6 it is stated that there was a full recall and patients were identified...flow chart please regarding inclusion...Please explain or correct. This is confusing. page 4, explain visually low coverage arc and low diametrical clearance page 6, explain visually SROM and Corail system Discussion page 17 line 11 - 24 should be in the introduction not in the discussion. line 42, but without imaging quantification and correlation with of the hip capsule....why? please discuss...

REVIEWER	Rami Madanat MD PhD Helsinki University Hospital Finland
REVIEW RETURNED	19-Sep-2015

GENERAL COMMENTS	This is a retrospective study from a single center assessing the performance of the 36mm Pinnacle metal-on-metal total hip replacement. The study uses both clinical data and implant retrieval data to determine risk factors for revision. The authors found that the device has an unacceptably high revision rate (84.3% at 9 years). They also found that patients with bilateral implants, female patients and patients with components implanted in later years (2006 onwards) have an increased risk of revision. Finally, the authors conclude that a significant number of explanted Pinnacles were manufactured with diameters outside the manufacturer's stated tolerances. General comments The study subject is important and information obtained by combining implant retrieval findings with clinical variables and component information can potentially help elucidate some of the failure mechanisms of metal-on-metal hip implants. My main concern with the study is regarding the use of explanted components to conclude that implants were manufactured outside the manufacturer's tolerances. My second concern is that in its current form the study is difficult to follow as the patient data and retrieval data seem to be very disconnected and the true patient flow

	and selection process is not sufficiently clear. I will address these issues as well as some other concerns below in more detail. The study would greatly benefit from a flow chart showing all stages of the study starting with the entire population including dropouts and deaths and ending with exact numbers of patients revised/unrevised. This should also show how many revised patients had explants studied, ions measured etc. Currently this information is very difficult to follow. The authors need to clarify how the NRR data was used. If possible clarify this in the flow chart (in parallel) or clearly separate this into its own section in the manuscript. Specific comments As mentioned above I do not think it is reliable to make definitive conclusions from explanted components regarding implants being manufactured outside stated tolerances (+/-20 micrometers). The authors themselves mention that components are vulnerable to deflection. This deformation may not only occur during implantation but probably even more likely when components are explanted. Thus, unless the authors have examined a large series of components that have not been implanted I do not think the conclusions they have reached are valid or scientifically sound. If the NRR data was used only to support this argument I would suggest leaving it out and focusing on the retrieval data on the 65 revised implants in the current study. Also, I would suggest not referencing unpublished data if possible. This is done several times in the manuscript. The authors mention that all revisions were for ARMD. Please give more detail on ARMD severity either based on preoperative cross sectional imaging (ultrasound) or intraoperative findings as both should be available. Please use an accepted classification for ALTR/ARMD if possible. Were histological ALVAL scores assessed from tissue samples removed at revision? The authors mention that they collected the Harris Hip Score but do not present any of this data. It would be nice to see the scores for revised and unrevised patients as has been shown for other collected variables. Failed implants had a significantly smaller anteversion (albeit within the recommendations of the surgical manual), can you please discuss if this is clinically relevant. Should table 3 have unilateral vs bilateral in the left column based on the ratio in the second column i.e. more failed bilaterals than unilaterals. Also, can the authors clarify how bilaterals were assessed (failure of either hip, failure of earlier hip) and in how many of these had both hips failed? Please list proportion of simultaneous and sequential cases in the bilaterals. Please provide a table that clarifies for the reader how the thickness of the liner varies according to the shell sizes for the 36mm implant. This would help to understand some of the main arguments especially for readers not so familiar with the implant. In the discussion section the authors compare data and note differences regarding the survival of the Pinnacle based on publications by region (USA vs. Europe). There are two noteworthy

	issues in their literature review table. First, the mean follow up time was somewhat shorter in the studies published from the USA, second, the European studies were retrospective. Thus they are not really comparing apples to apples. I would suggest that the statement regarding differences in financial influence should be clearly substantiated with data. Minor comments: Please clarify where and who performed the measurements of the explants? The abstract mentions that patients were implanted with the Pinnacle device by ONE of the senior authors but the methods section mentions TWO senior authors? I would recommend not using direct quotations from other published papers regarding their conclusions or findings, rather please paraphrase. I believe some of the supplementary material is unnecessary and could be removed.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

My only question to the authors, and I think that I know the answer in advance, is whether they did the effort of routinely obtaining consent from the patients whose retrieved implants were sent for analysis? This is relevant given the significant proportion of these prosthesis that are failing prematurely due to design and manufacturing issues that have serious medico-legal implications.

Our initial successful ethics application focused on hip resurfacings. It was entitled: The Assessment of Failed Hip Resurfacings (reference number 09/H0905/41) and was approved by the LREC County Durham and Tees Valley 2. Subsequently it was extended to include all types of hip replacements and then further still to include all orthopaedic devices. It is hospital policy to consent all patients prior to revision surgery for the storage and analysis of explanted orthopaedic devices.

Reviewer 2

"The single surgeon nature of the series limits the generalisability of the work and should therefore be listed as a limitation rather than a strength."

We have clarified this in the manuscript – this is a two surgeon study. There was no significant difference in the survival of the two surgeons' cohorts.

No ethical approval for the study is mentioned. Given the nature of the work, the requirements for follow up in this population and the jurisdiction of the centre this study would fall under the category of service evaluation. This needs to be clearly stated and the NRES guidance cited in order to justify the lack of ethics committee approval.

Please refer to the above responses.

It should be made clear in the background that the first design factor mentioned in relation to failure of the ASR (a shallow acetabular component (low coverage arc)), does not apply to the Pinnacle. It may be better to remove this factor given the general readership of the journal.

While we appreciate the value of this comment, we believe we clarified this issue in the manuscript. We believe it is important to shift attention away from the impact of cup position – which was largely created from studies on the ASR. As we wrote: "The aim of this study was to identify variables associated with early failure of the Pinnacle 36mm MoM system by a retrospective analysis of all patients implanted with this device by two experienced hip surgeons at our institution. After consideration of the above factors we hypothesized: The Pinnacle system would be relatively

resistant to the effects of cup position in terms of blood metal ion release due to its smoother rim and greater arc of cover conferring protection from edge wear. By extension, ARMD would, in general, be unrelated to cup position."

As per the instructions to authors "Only papers published or in press should be included in the reference list." The 8th reference is a conference paper that has not been published as a paper as far as I am aware. If the authors wish to cite unpublished data, they should provide a link or source for the data referred to such that it can be reviewed by the reader.
This reference is in press. We have also described the findings in detail in the figure and in the Appendix.

"Surgical factors: Cup orientation has been shown to affect wear rates/metal ion release in MoM arthroplasty[9] and devices with lower coverage arcs and sharper articular rims are particularly sensitive to cup position.[15]" the reference to lower coverage arcs and sharper articular rims is not useful as it does not apply to the design considered in this manuscript but to the ASR, another metal on metal implant system with design features not shared with the Pinnacle Metal on Metal system. To avoid confusion for readers that may not be aware of this, I would remove this unless the authors have established a particular link they wish to make, in which case this should be explicitly stated.
While we understand this might be of value, as we described above, the aims of this study were to investigate also the link between metal ions and cup orientation of the Pinnacle as there is far too much emphasis on this variable secondary to the negative publicity over the ASR device: "The aim of this study was to identify variables associated with early failure of the Pinnacle 36mm MoM system by a retrospective analysis of all patients implanted with this device by two experienced hip surgeons at our institution. After consideration of the above factors we hypothesized: The Pinnacle system would be relatively resistant to the effects of cup position in terms of blood metal ion release due to its smoother rim and greater arc of cover conferring protection from edge wear. By extension, ARMD would, in general, be unrelated to cup position."

"devices with no excessive wear with associated ARMD are commonly found in patients with bilateral devices implying a process of sensitisation (unpublished data)." there are a number of published studies looking at bilateral devices, is this statement reflected in any of these studies or this comment unique to the authors' unpublished data? Again, I would prefer that unpublished data referred to is made accessible to the reader, either by providing an online source or providing the data as an appendix.

Reworded to "in our own experience devices with no excessive wear with associated ARMD are commonly found in patients with bilateral devices implying a process of sensitisation and reference added:
Madanat R, Hussey DK, Donahue GS, Potter HG, Wallace R, Bragdon CR, Muratoglu OK, Malchau H. The Symmetry of Adverse Local Tissue Reactions in Patients with Bilateral Simultaneous and Sequential ASR Hip Replacement. J Arthroplasty. 2015 Oct;30(10):1794-8.

The information regarding the number of surgeons performing the surgeries is inconsistent. the abstract states "All patients implanted with a DePuy Pinnacle MoM hip prostheses by one of the senior authors" but the background states "The aim of this study was to identify variables associated with early failure of the Pinnacle 36mm MoM system by a retrospective analysis of all patients implanted with this device by two experienced hip surgeons at our institution". How many surgeons performed the primary operations and how many were supervised by one or either of the surgeons mentioned?
We have clarified this now in the manuscript – two surgeons performed the surgeries - RKL and AVFN.

As the journal is targeted at a general readership, the link between the size of the acetabular shell and the thickness of the metal liner used with a 36mm metal on metal bearing should be explained.
We hope we have clarified this relationship – figure 6 has also been added.

The authors describe what appears to be two separate phases of follow up - that between 2008 and 2010 where they were using cross sectional imaging and blood metal ion testing "routinely" and that after 2010 when a "full MoM hip recall" was commenced. What follow up was performed, in which individuals at what time points should be clearly stated for each of these phases.
Our information and its presentation here was somewhat muddled. We hopefully have addressed this satisfactorily.

As the paper has been submitted to a journal with a general readership, the authors should clearly state their selection criteria for the use of the SROM stem in the under 70s after 2005. Were all cases of dysplasia and Perthes treated with one or was there selection criteria applied in this cohort and did any other patients receive the SROM. The authors' should also clarify the difference between the designs of the stems to explain their selections.
We hope we have addressed the reviewer's comments adequately.

When the authors describe their protocol for investigation of these patients, they should state clearly the section criteria they applied to select patients for revision surgery amongst the investigations listed. A flow chart or decision tree would be a useful figure to illustrate this.
Flow chart figure 5 has now been added.

A patient flow chart should be provided showing the inclusion and exclusion criteria described and the flow of patients through the study to the final numbers. This should start with all patients undergoing primary hip arthroplasty in the unit during the period the procedures were performed.
Flowchart figure 4 has now been added

In order to provide context, the authors should also describe the revision rates in their unit for other designs.
We believe that it should be clear to the reader that the unilateral male Pinnacle results reported in the paper provide the best comparison for the overall results. It is also clearly stated throughout that the results of the Pinnacle at our unit are strikingly similar to those published from other UK centres.

The authors state that they used 36mm bearings in all cases. They then state "The 160° sub-hemispherical bearing surface does not vary with the thickness of the liner.". According to the manufacturer's design rationale of the Ultamet metal liner and the figures provided in that, the bearing surface is 180 degrees for the 36mm metal liners and is not sub hemispherical. This should be clarified and a source for this information cited.
The external shell is hemispherical but the metal bearing liner is a 160 degree sub hemisphere. We know this from extensive experience in the measurements of explanted and new prostheses as well as Depuy's own statements.

I am concerned by the use of explant analysis to determine if components were outside stated manufacturing tolerances. By definition, these are own bearings and the majority of the cases revised in this series had been undergoing wear in vivo for a number of years. Changes from the original manufacturing specifications and measurements would therefore be expected, particularly at the lower end of the tolerance which is where the authors imply they saw variation. I do not think this is reliable and if the authors wish to determine if components were outside the manufacturing tolerances stated, they would need to look at unworn components.
Please refer to the supplementary material we have submitted on this subject in Appendix 3 – as well as inclusion of measurements of unused devices.

Why was pre and post 2006 implantation selected as a binary variable for the model rather than year of implantation being considered? I assume the authors had a reason for this categorisation.
We were prompted to study the effect of year of manufacture/implantation based on our explant findings which appeared to show a marked increase in the number of abnormally low diametrical clearances from 2006 onwards. We have shown evidence of this in the explant analysis section and hence our reason to include all the NRR cases we had measured. The number of non conforming products appears to jump in 2006 and does not increase sequentially thereafter hence the reason to divide into two time periods.

Liner size is given as a variable for the proportional hazards model but elsewhere the authors have stated it is the acetabular shell size, please correct or clarify which measurement was used and explain the implications of this (i.e. given the constant inner diameter, a smaller acetabular shell of a constant thickness results in a thiner liner).
We are sorry for the confusion and have hopefully addressed this.

Given the authors were concerned about the risk of bias created by the introduction of 36mm metal liners for 50mm outer diameter shells from 2008 onwards, was the model repeated with and without the subsequently excluded 51 cases included to see if the results differed? If not, this should be done.
The exclusion of 51 cases was in fact incorrect due to a data processing issue (which we stress made no difference to the actual calculations. We have addressed this issue now and hopefully have provided greater clarity in the flow chart.

Here are the results without any exclusions – for bearing failure:
Variable Value Standard error Wald Chi-Square Pr > Chi² Hazard ratio Hazard ratio Lower bound (95%) Hazard ratio Upper bound (95%)
Shell size -0.137 0.100 1.874 0.171 0.872 0.717 1.061
BILATERAL-Y 0.588 0.375 2.453 0.117 1.801 0.863 3.759
Gender-M -0.713 0.525 1.842 0.175 0.490 0.175 1.372
Stem-SROM 0.109 0.479 0.051 0.821 1.115 0.436 2.853
Early versus late cohort- Early -0.834 0.561 2.208 0.137 0.434 0.145 1.305

For taper failure
Variable Value Standard error Wald Chi-Square Pr > Chi² Hazard ratio Hazard ratio Lower bound (95%) Hazard ratio Upper bound (95%)
Shell size 0.037 0.067 0.303 0.582 1.037 0.910 1.182
BILATERAL-Y 0.690 0.337 4.186 0.041 1.994 1.029 3.861
Gender-M -0.938 0.445 4.436 0.035 0.391 0.164 0.937
Stem-SROM -0.044 0.443 0.010 0.921 0.957 0.402 2.279
Early cohort -0.558 0.460 1.473 0.225 0.572 0.232 1.409

For all failures

Variable Value Standard error Wald Chi-Square Pr > Chi² Hazard ratio Hazard ratio Lower bound (95%) Hazard ratio Upper bound (95%)
Shell size -0.028 0.059 0.230 0.632 0.972 0.866 1.092
BILATERAL-Y 0.678 0.257 6.935 0.008 1.970 1.189 3.263
Gender-M -0.719 0.347 4.295 0.038 0.487 0.247 0.962
Stem-SROM 0.147 0.339 0.188 0.664 1.158 0.596 2.249
Early cohort -1.243 0.418 8.851 0.003 0.289 0.127 0.654

All the same trends are observed – however we do not believe this is the correct statistical approach for reasons clearly explained in the manuscript.

It is not reasonable to state that non-attendees had well functioning prostheses as it is known that those that are lost to follow up have worse overall functional results than those that do. They may also have been revised in other units without the authors knowing and the data capture of retrieval registries is not 100% for patients who have primary hip replacements in their catchment area so revisions may have been performed that were not captured by the NRR
We have clearly stated this limitation in the discussion: “Finally, a significant number of patients were lost to follow up. We have assumed in our survival analyses that these patients are asymptomatic at present. This is a major assumption and joint survival rates reported herein are likely to represent “best outcome scenario”. We also put this as the main limitation at the front of the manuscript. Furthermore, the catchment of the NRR is extremely wide – meaning that if patients had their joints revised at nearby centres it was more likely than not that we would have received the explant. We are not sure from what data the reviewer refers to when he states: “the data capture of retrieval registries is not 100% for patients who have primary hip replacements in their catchment area.” To the best of our knowledge we are one of the few retrieval registries in existence.

The range (I assume 95% confidence interval) is missing from the Predicted joint survival/survival at 8 years row of the pre 2006 column in table 2.
This has been modified now

There is a typographical error in the Median (range) blood Cr unilaterals (µg/l) row of table 3, the decimal point is missing from the p value.
This has been corrected
There is a typographical error in the Median (range) combined wear rates* (mm3 per year) row of table 3, the units have been included in each column containing results as well as the first column, which it has not in other rows.
This has been addressed

The authors state that "Taper wear was associated with a larger female taper angle and an increased femoral head offset". The taper diameters have been stated for the SROM (11/13) and the Corail (12/14) but the taper lengths and angles have not been stated. The authors should either give these values or clearly state which of the taper geometries has the highest angle as this can not be determined by the reader from the diameters alone given the length also varies.
This has been addressed in the results and methods sections

The authors give a p value for the difference between survival estimates on page 14 of 42 but have not mentioned what statistical test was used prior to this. They subsequently mention the use of a log rank test on page 15. Statistical tests employed should be described in the methods section before results of the tests are given.
This has now been changed

"Approximately 93,000 ASRs were sold globally, with around 6,000 reported in the NJR of England and Wales. The 2014 Annual NJR Report lists 11,871 Pinnacle implantations. If England and Wales represent the same proportion of Pinnacle as ASR implantations then it is not unreasonable to suggest that the Pinnacle MoM system has been implanted into over 180,000 patients globally, making it the most commonly used large diameter MoM THR in the world." this is not a reasonable statement to make. The Pinnacle system can be used with a variety of different bearing combinations (metal and ceramic heads; polyethylene, highly cross linked polyethylene, metal and ceramic liners). The implantation of 11,871 Pinnacle shells recorded in the NJR in 2014 can not therefore be reliably

used to estimate the number of Pinnacle shells implanted worldwide that have also had a metal liner in conjunction with a metal head implanted.
The figure of 11,871 refers to metal liners. Please refer to the extracts from the NJR below:

Table 3.10

For the Kaplan Meier figures, the 95% confidence intervals should be displayed and the numbers at risk given under the horizontal axis. If this involves splitting the into multiple panes, this should be done.
We hope we have given enough information now with the inclusion of the numbers at risk and CRRs in Appendix 3.

Reviewer 3

Page 7, could a decision rule (flow chart) be given regarding indication for revision be given as functional results of revision are very disappointing. Complaints is a very rough indicator in my opinion alone (off course...) to justify revision for example. So if one decides to revise, on what grounds does the orthopedic surgeon do this. Also high serum ion levels that can lead to cardiac arrhythmia's, brain deposition etc.. Image findings? Combination, 3 out of 3 positive? Please elaborate briefly in the introduction. A list of revised patients in an appendix regarding parameters and their individual outcomes that influenced the choice on revision could be extremely valuable for even the experienced clinician. How were complaints such as pain and discomfort quantified. HHS? How good is this test? Please elaborate. What were the imaging findings on US? Who performed the US, how much experience did they have?

We have now included a decision flowchart which is hopefully satisfactory. Furthermore we have given greater detail of the revision cases.

In my opinion this paper in the BMJ, which is read by a general audience and not just orthopedic surgeons, would benefit if the authors elaborate in the discussion section on whether it is fair that the whole idea of metal on metal THA should be abandoned or that patient selection and better design and manufacturing, for example a unique patient tailored one piece without tapers design by means of pre-operative 3 D imaging for example could or should be investigated. In summary is there still a future for MoM THA prosthesis in the future? If so, under what conditions regarding patient selection, design and manufacturing? A bridge to future research could be inserted and would not be out of place at this point in time after many reports of THA related problems.
While we agree with this comment and actually considered inserting the commentary we felt that we were already well over the word limit. Perhaps the journal editors could advise here?We have also attempted in the conclusions to mention the fact that it is not necessarily the MoM system itself which is the inherent problem leading to failure.

No quantification of the pathological capsular reaction is provided, nor taken into account for the analysis. Is this because of the use of ultrasound? Please elaborate on the reason for screening with US instead of CT and MR in the M and M, Also explain why you did not take into the account the expression of the disease by means of pathological capsular reaction. The reason that I am mentioning this is because imaging the pathological capsule by means of CT for example seems to be the strongest predictor for revision and correlates strongly with serum ion levels.
These are all excellent points. But…we have a huge population of MoM patients at our hospital – around 2000 patients. It would be simply impractical to routinely use CT or MARS MRI although as the number of patients we have is now diminishing we are conducting MRI scans more frequently. We have also found and have documented as such in our previous published work the link between

operative findings and ultrasound. We believe it to be a reliable tool and in a paper currently under review we explore the link between macroscopic tissue destruction observed at revision surgery, wear, histological response and fluid. Fluid appears to be the strongest link to the presence of advanced tissue destruction and we ae therefore happy with the use of ultrasound scanning. A recent publication is consistent with our clinical approach - Bone Joint J 2015;97-B:1328–37 – which showed clear evdicne that the presence of abnormal fluid is of significant concern.
With regard to quantification of the capsular reaction – we wonder how we could have included this in the statistical analysis? After all, rather than just scan results we had definitive operative and histological confirmation of pathology – while we would have liked to have reported further on this area in this paper we simply did not have enough space and felt that the paper was already approaching over complication.
We have however given greater detail of the operative findings as well as the histo response – factors that we routinely record and study in great depth. While we appreciate that we may have not addressed the reviewer's insightful comments we hope that we have provided enough information within the limits of one manuscript.

As revision rate of MoM cohorts is likely depending on the proportion unilateral/ bilateral patients, please add a survival curve stratified for hips of patients with a unilateral MoM hip replacement and hips of patients with a bilateral MoM hip replacement.
This is included now.

As figures 'shell size and median co concentration' and 'Percentage of pinnacle bearing by year' are a comparison of seperate groups, it would be better to depict unconnected dots.
We have changed this now to the more appropriate box and whisker plots.

Discussion could gain readability from the adagium: what did we find find, what's actually new and how does this relate to the literature and then discuss findings that are consistent with other studies which are off course important, but not that new.
Again we are confined by an already extensive and potentially overlong discussion. The major new finding is that of variation in manufacturing potentially having an impact on clinical outcomes. We believe we have expressed this quite clearly. We are unware of something like this being published before – and if we were to put the message more bluntly we would risk incurring the wrath of reviewers!! We also believe that paticualry table five puts the evidence into context with the existing literature.

Please provide METC number. or the waiver...
Please refer to responses above

page 2 line 20:
mentions invitation of patients from one senior author, while page 5 mentions implantation of the device by 2 surgeons. page 6 it is stated that there was a full recall and patients were identified...flow chart please regarding inclusion...Please explain or correct. This is confusing.
Please refer to explanations above. Flow chart included and surgeons clarified

page 4, explain visually low coverage arc and low diametrical clearance
We have included more visual representations of diametrical clearance though we did not feel there was any more space for the coverage arc. As we believed coverage to be a side issue could we be forgiven for not including this?

page 6, explain visually SROM and Corail system
We have attempted to do this in the methods section and hope this is more satisfactory. There was no significant impact of stem on the survival of the prostheses and did not want to go into too much depth

here.

page 17
line 11 - 24 should be in the introduction not in the discussion.
While we agree with the reviewer's point, could we please keep this section here as it provides a brief overview of the global performance of the device and allows us to then discuss our findings in that context.

line 42, but without imaging quantification and correlation with of the hip capsule....why? please discuss...
We have added the operative and histological findings now in the results section. We would find it extremely difficult to correlate those findings as the paper is already quite complex. We hope the reviewer can sympathise with us here?

Reviewer 4

The study would greatly benefit from a flow chart showing all stages of the study starting with the entire population including dropouts and deaths and ending with exact numbers of patients revised/unrevised. This should also show how many revised patients had explants studied, ions measured etc. Currently this information is very difficult to follow. The authors need to clarify how the NRR data was used. If possible clarify this in the flow chart (in parallel) or clearly separate this into its own section in the manuscript.

We hope we have clarified this in the flowcharts.
The number of metal ions is provided in great depth in Appendix 1.

As mentioned above I do not think it is reliable to make definitive conclusions from explanted components regarding implants being manufactured outside stated tolerances (+/-20 micrometers). The authors themselves mention that components are vulnerable to deflection. This deformation may not only occur during implantation but probably even more likely when components are explanted. Thus, unless the authors have examined a large series of components that have not been implanted I do not think the conclusions they have reached are valid or scientifically sound. If the NRR data was used only to support this argument I would suggest leaving it out and focusing on the retrieval data on the 65 revised implants in the current study. Also, I would suggest not referencing unpublished data if possible. This is done several times in the manuscript.
Yes implants deflect but there is general consensus that this is elastic not plastic deformation. Our techniques are described clearly in the literature and we have shown that it is possible to easily identify changes secondary to deformation. Why also would we see variations in year of implantation when we have used the same methods? The wider explant analysis provided by the NRR explants is essential to the message of the paper – the results even prompted us to begin the survival analysis as we have carried it out. We have addressed the questions of the validity of the measurements in the more extensive attachment. We have also analysed eight unused heads and ten unused liners – these confirmed the trends we had observed with the explants. Please refer to explant section and further information on the analysis of explants to determine original state.

The authors mention that all revisions were for ARMD. Please give more detail on ARMD severity either based on preoperative cross sectional imaging (ultrasound) or intraoperative findings as both should be available. Please use an accepted classification for ALTR/ARMD if possible. Were histological ALVAL scores assessed from tissue samples removed at revision?
This has now been included. We actually developed the term ARMD and we have also produced one of the largest data collections on the histological condition ALVAL. We believe that the descriptions

we have used here and the references are now adequate.

The authors mention that they collected the Harris Hip Score but do not present any of this data. It would be nice to see the scores for revised and unrevised patients as has been shown for other collected variables.
This has now been added

Failed implants had a significantly smaller anteversion (albeit within the recommendations of the surgical manual), can you please discuss if this is clinically relevant.
We have commented on this now to a limited extent in table 3 caption.

Should table 3 have unilateral vs bilateral in the left column based on the ratio in the second column i.e. more failed bilaterals than unilaterals. Also, can the authors clarify how bilaterals were assessed (failure of either hip, failure of earlier hip) and in how many of these had both hips failed? Please list proportion of simultaneous and sequential cases in the bilaterals.
No hips were performed simultaneously – this has been added to the text. Table 3 is correct – more failed unilaterals than bilaterals. From table 3 it can be calculated how many patients had both hips fail.

Please provide a table that clarifies for the reader how the thickness of the liner varies according to the shell sizes for the 36mm implant. This would help to understand some of the main arguments especially for readers not so familiar with the implant.
This has now been addressed in figure 6.

In the discussion section the authors compare data and note differences regarding the survival of the Pinnacle based on publications by region (USA vs. Europe). There are two noteworthy issues in their literature review table. First, the mean follow up time was somewhat shorter in the studies published from the USA, second, the European studies were retrospective. Thus they are not really comparing apples to apples. I would suggest that the statement regarding differences in financial influence should be clearly substantiated with data.
We believe that we have clearly described the differences between the studies in an attempt to explain the differences in outcomes between the US and Europe. We believe this section is appropriate given that we have explicitly stated the differences rather than attempting to compare apples with oranges. Re the financial influence this can clearly be accessed in the conflicts statements in the papers which are referenced and also, with respect to the US studies, the Sunshine Act.

Please clarify where and who performed the measurements of the explants?
This has been clarified.

The abstract mentions that patients were implanted with the Pinnacle device by ONE of the senior authors but the methods section mentions TWO senior authors?
This has been clarified – a bad mistake in the abstract which we apologise about.

I would recommend not using direct quotations from other published papers regarding their conclusions or findings, rather please paraphrase.
These have mostly been removed now.

I believe some of the supplementary material is unnecessary and could be removed.
Could we kindly ask that this be retained? We believe this is essential to the interpretation of the performance of the device and is the unique aspect to this study. It is there for completeness. We firmly believe that the NRR data should be included – it is a novel approach to an orthopaedic study

and, we believe, an entirely appropriate one. In a specialty so dependent on the interaction between manufacturing technology and the human environment it makes sense to present data from a larger database of failed devices. We believe that this should be the future for orthopaedics and indeed there are currently ongoing arrangements to advise that all explanted orthopaedic devices should be analysed on a routine basis due to the information they provide (Parliamentary House Select Investigation 2012).

VERSION 2 – REVIEW

REVIEWER	Mr Pedro Foguet FRCS T&O UHCW NHS Trust, UK
REVIEW RETURNED	06-Dec-2015

GENERAL COMMENTS	Excellent work

REVIEWER	Michael Whitehouse University of Bristol, UK I have provided teaching sessions on cemented hip replacement for DePuy International Ltd for which my institution has received payment.
REVIEW RETURNED	01-Dec-2015

GENERAL COMMENTS	Thank you for the thorough response to my concerns which have been adequately addressed.

REVIEWER	Boomsma MF Isala hospital, Zwolle, The Netherlands
REVIEW RETURNED	20-Dec-2015

GENERAL COMMENTS	L.S., The paper is consistent and results are proportional and well discussed. Although not all my reviewer comments are addressed I understand and accept the author's choices in discussing their results. I admire the author's efforts in putting their focus of investigation in the manufacturing proces and the use of different components.This might improve both future THA designs and create a consistent manufacturing proces of THA in MoM or even in THA in general.