



MODERN METAL-ON-METAL HIP IMPLANTS A TECHNOLOGY OVERVIEW

Adopted by the American Academy of Orthopaedic Surgeons

Board of Directors

December 2, 2011

This *Technology Overview* was prepared using systematic review methodology, and summarizes the findings of studies published as of July 15, 2011 on modern metal-on-metal hip implants. The publication of studies and registry reports subsequent to this date will enhance understanding of this topic. As a summary, this document does not make recommendations for or against the use of metal-on-metal hip implants, and it should not be construed as an official position of the American Academy of Orthopaedic Surgeons. Readers are encouraged to consider the information presented in this document and reach their own conclusions about metal-on-metal hip implants.

The American Academy of Orthopaedic Surgeons has developed and is providing this *Technology Overview* as an educational tool. Patient care and treatment should always be based on a clinician's independent medical judgment given the individual patient's clinical circumstances.

Table of Contents

Table of Contents	ii
Summary of Published Results	iv
Introduction.....	1
Methods Overview	2
Included Articles	2
Quality of the Literature.....	2
Outcomes Considered	2
Question #1: What are the clinical outcomes in patients with metal-on-metal hip replacements in comparison with other bearing surface combinations?	4
Overall Revision Rates	4
Validated Patient-Oriented Outcome measures	9
Question #2: What are the patient, implant, and surgical factors that best predict successful/unsuccessful outcomes of metal-on-metal hip replacement?	11
Metal-on-metal Total Hip Arthroplasty (THA)	11
Implant Head Size Factors	11
Patient Age Group Factors.....	13
Gender Factors	14
Age and Femoral Component Size Interaction.....	15
Gender and Femoral Component Size Interaction.....	16
U.K./Wales Results.....	17
Metal-on-Metal Hip Resurfacing.....	18
English Outcomes	18
Australian Outcomes.....	18
Question #3: What is the prevalence of adverse clinical problems from metal-on-metal hip replacement compared to other bearing surface combinations?	24
Local Metal Debris Release	24
Pseudotumors	25
Imaging	26
Other Complications	26
Appendices.....	27
Appendix I	27
Inclusion Criteria	27
Appendix II.....	28
Search Strategies and Databases	28
Appendix III.....	30
Study Attrition Flowchart	30
Included Studies.....	31
Excluded Studies.....	32
Appendix IV.....	37
Appraising Evidence Quality and Applicability	37
Studies of Interventions	37
Studies of Screening and Diagnostic Tests.....	40
Studies of Prognostics.....	43
Studies of Incidence and Prevalence.....	45
Studies of Joint Registry Evaluation.....	48

Appendix V.....	49
Study Quality and Applicability	49
Appendix VI.....	58
Supplemental Tables and Figures	58
Tables and Figures for Question #1	58
Tables and Figures for Question #2	59
Tables and Figures for Question #3	62
Appendix VII	69
AAOS Bodies that Approved this Technology Overview	69
References	72
Excluded Articles Bibliography.....	74

Summary of Published Results

Summaries of the data that pertain to the three key questions addressed in this Technology Overview are:

Question #1: What are the clinical outcomes in patients with metal-on-metal hip replacements in comparison with other bearing surface combinations?

- Analyses conducted on objective patient-oriented outcomes by two joint registries indicate that, overall, patients who receive metal-on-metal total hip arthroplasty and hip resurfacing are at greater risk for revision than patients who receive total hip arthroplasty (THA) using a different bearing surface combination. The Australian¹ and U.K./Wales² registries both indicated that metal-on-metal THA using larger femoral head components places individuals at a greater revision risk than individuals with smaller femoral head components.
- Studies evaluating validated, patient-oriented outcomes included a high strength study¹⁵ using Harris Hip Score, WOMAC, and SF-12; a high strength study²³ evaluating Harris Hip Score and Oxford Hip Score; a moderate strength study²² evaluating WOMAC.
 - Two^{15, 23} studies compared metal-on-metal total hip arthroplasty (MoM THA) with metal-on-polyethylene (MoP THA) and one²² compared MoM THA with hip resurfacing.
 - In each case of MoM THA vs. MoP THA, both treatment arms improved significantly, but there were no significant differences between groups.
 - The comparison between hip resurfacing and MoM THA yielded statistically significantly better one and two year WOMAC scores for hip resurfacing, but there was no clinical relevance in the difference, post-operative protocols were different, and patient demographics differed between treatment arms.
- Rigorous multivariate statistical analysis that would allow determination of whether one particular type of patients fare better than others with metal-on-metal THA has not yet been conducted by the registries or in other peer reviewed literature. Such analysis must simultaneously account for the effects of all patient and device characteristics of interest and also take into account any interactions between relevant variables.

Question #2: What are the patient, implant, and surgical factors that best predict successful/unsuccessful outcomes of metal-on-metal hip replacement?

- Data from the Australian¹ and U.K./Wales² registries addressed patient age and gender, and implant head size as variables that are predictive of success/failure in metal-on-metal total hip arthroplasty and hip resurfacing.

- Data from both registries reported that larger femoral head components used in metal-on-metal total hip arthroplasty have higher revision rates and revision risk after adjusting for age and gender.
- The U.K./Wales registry reports that increased age is associated with increased revision risks of Large-Head Metal-on-Metal Total Hip Arthroplasty (LHMoM THA). The Australian registry also reported that patients older than 65 years with larger femoral head components had a greater revision risk than patients of the same age, but with smaller femoral head components.
- The U.K./Wales registry reported hip resurfacing patients in all age groups, except males <55 years of age, were at an increased revision risk compared to cemented total hip arthroplasty with an unspecified bearing surface. The Australian registry reported hip resurfacing patient ≥ 65 years of age to have the highest revision risk.
- Head size and risk of revision for hip resurfacing procedures were inversely related to each other. Those receiving the smallest femoral head components had the greatest risk of revision. There were no gender differences reported in hip resurfacing outcomes, but the implant size was associated with poorer outcomes when gender/implant size interaction was analyzed.
- The registry analyses were limited to one variable in Cox regression models, therefore conclusions should be made cautiously. Further research employing rigorous multivariate analysis is needed to determine which variables best predict failure/success. Not including all variables that could potentially confound interpretations increases the chances of reaching spurious conclusions.

Question #3: What is the prevalence of adverse clinical problems from metal-on-metal hip replacement compared to other bearing surface combinations?

- Limited data exists comparing the prevalence of adverse clinical problems with metal-on-metal hip systems and other bearing surfaces. Several studies noted a correlation between suboptimal hip implant positioning and higher wear rates, local metal debris release, and consequent local tissue reactions to metal debris (e.g., “pseudotumors”). Several studies reported elevated serum metal ion (cobalt and chromium) concentrations in patients with metal-on-metal hip articulations, especially in patients with malpositioned implants. The clinical significance of elevated serum metal ion concentrations remains unknown.
- Registry reports present data on adverse clinical problems and reasons for revision surgery, but do not identify the specific bearing surface combination in all cases. The U.K./Wales registry did begin gathering data on soft tissue reactions in July of 2009, but had too little data when the most recent report was published.

INTRODUCTION

Metal-on-metal hip implants consist of a ball, stem, and shell, all made of metal materials, while hip resurfacing consists of a trimmed femoral head capped with a metal covering. This Technology Overview addresses the former while a previous AAOS Technology Overview addressed the latter ([link](#)). Artificial hip systems have inherent risks associated with the procedure such as reactions to anesthesia, wound infections, excessive bleeding, dislocation, and blood clots. Metal-on-metal hip systems, in addition to the inherent risks, have additional unique risks. The metal ball and cup slide against each other while walking or running, and have the potential to release metal particles in the periprosthetic space and/or systemically. While precautions can be taken to optimize the positioning of the hip system, there is no way to fully prevent the production of metal debris.

There have been occurrences of patients experiencing less than optimal outcomes from metal-on-metal hip systems, sometimes termed “adverse local tissue reactions (ALTR),” or “adverse reaction to metal debris (ARMD).” Patients react to metal particles in different ways, and it is not possible to predict which patients will experience an adverse event with metal-on-metal hip systems. These reactions have led to earlier failure rates for the affected patients.

In February 2011, the U.S. Food and Drug Administration (FDA) issued a public health communication about hip replacement components that have both a metal ball and metal socket (metal-on-metal hip systems), and on May 6, 2011, issued an order for manufacturers to conduct post-market surveillance on total metal-on-metal hip replacement devices to monitor adverse events.

The goals of this Technology Overview are to use the tools of evidence-based medicine to summarize information on the indications, effectiveness, prevalence of adverse events, and failure rates of modern metal-on-metal hip replacement technology. The findings will compare and contrast modern metal-on-metal hip replacement technology to the current generation of total hip technology using other bearing surfaces to highlight areas of overlap and inform the practicing orthopaedist concerning important differences.

METHODS OVERVIEW

This report was developed using the methods of a systematic review. We began by having a panel of physicians frame three Key Questions, and next developed rules (inclusion criteria) for determining what information we would include (The full list of criteria appears in Appendix I). Finally, we conducted comprehensive literature searches (Appendix II) to ensure that the data we considered are not biased in favor of any particular point of view. Thereafter, we evaluated the quality of the relevant studies, including their methods of analysis, considered their results, and summarized this information in graphs and tables.

INCLUDED ARTICLES

Our searches identified 3038 hip replacement citations and data from 8 joint registries that were potentially relevant to this Overview and that could potentially meet our inclusion criteria. Of these, 19 articles and the data from 2 joint registry reports met these criteria. This information forms the dataset that we used to address the questions below.

We emphasize that this is a Technology Overview of metal-on-metal hip replacements, and not of all hip replacement procedures. The THA outcomes that we consider are only from the most recent registry reports^{1,2}. We have chosen this strategy because the most recent reports provide direct statistical comparisons of metal-on-metal, hip resurfacing, and other bearing surface hip replacements.

QUALITY OF THE LITERATURE

Assessing the quality of evidence is an important step in a systematic review. Readers can have more confidence in the results of high quality studies than low quality studies. To assess quality, we used a domain-based approach which allows for the evaluation of studies of all designs. Design alone, however, does not adequately reflect the quality of a study. Therefore, we assessed study quality using a system that considers not only design, but also several aspects of how well a study was conducted. Details about this system are provided in Appendix IV.

We also considered the quality of the statistical analyses provided in each registry report, and we discuss these considerations in the text of the main body of this Overview.

The current available studies do not allow us to determine whether one particular type of patient fares better than others. Subgroup analysis requires rigorous statistical analysis; including simultaneously accounting for the effects of all patient and device characteristics of interest and performing statistical tests for interactions between the relevant variables. To date, such analyses have not been conducted.

OUTCOMES CONSIDERED

We included patient-oriented outcomes to assess clinical effectiveness of metal-on-metal hip replacements. Validated “paper-and-pencil” outcome measures (e.g., Harris Hip Score, etc.) were used as subjective measures and revision rates were used as objective measures of effectiveness. These outcomes matter to patients (i.e., patient-oriented) and indicate whether an intervention is effective. Serum metal ions may serve as potential surrogate markers for implant failure and thus were included as an outcome in this Overview. Outcomes based on imaging of hip implants for positioning and imaging of periprosthetic tissues were also included in this

Overview. These surrogate markers were agreed upon by the physician task force *a priori* to literature retrieval and quality evaluation to minimize bias.

Question #1: What are the clinical outcomes in patients with metal-on-metal hip replacements in comparison with other bearing surface combinations?

Different aspects of the question are addressed by two^{1,2} high quality joint registries, the Australian Orthopaedic Association National Joint Replacement Registry, and the National Joint Registry for England and Wales; and two^{15,23} high strength and one²² moderate strength primary literature studies (see Appendix V for quality evaluation of these registries and articles). Therefore one could have high confidence in the included evidence from the registries and studies rated as high strength, and have moderate confidence in studies rated as moderate strength.

Patient oriented outcomes were reported in the registries and studies. Registry data provided objective patient oriented outcomes such as revision rates, and the primary literature studies provided validated, subjective patient oriented outcomes including data from instruments such as the Harris Hip Score, and others.

None of the registry data presented on revision rates were from registry conducted analyses that employed statistical adjustments that could have accounted for all possibly important variables. Therefore, one cannot conclusively determine how including all potentially important variables would affect the outcomes presented. The best available evidence is comprised of each registry's analysis with the greatest number of statistical adjustments.

OVERALL REVISION RATES

The data from the two registries found revision rates are higher among metal-on-metal hip systems than other bearing surface systems. Each registry performed different statistical adjustments and reached similar conclusions. The quantitative results showing the overall revision rates, and risk for revision, expressed as a hazard ratio, are shown in Tables 1-3.

Data from the Australian registry show that metal-on-metal THA has the highest revision rate of 1.14 hips per 100 observed years and highest cumulative percent revision of 7.7% over nine years (Table 1). These results indicate statistically significant hazard ratios when comparing each bearing surface to metal-on-metal. The U.K. and Wales registry concluded that the revision rates at three and five years were statistically different comparing Large-Head Metal-on-Metal Total Hip Arthroplasty (LHMoM THA) with other hip systems including those fixed with and without cement (there were no data given as to the kind of bearing surface used in each of those systems). The U.K./Wales registry indicated metal-on-metal hip resurfacing had greater revision risks as well.

Table 1. Revision rates after Metal-on-Metal hip replacement and other bearing surfaces from the Australian joint registry (Primary diagnosis OA)

Bearing Surface	N revised	N total	Obs. Years	Revisions/100 Obs. Yrs	95% CI
C/C	793	29945	111047	0.71	0.67, 0.77
C/P	903	34560	143224	0.63	0.59, 0.67
M/M	667	17808	58503	1.14	1.06, 1.23
M/P	1818	62550	250414	0.73	0.69, 0.76
CM/P	93	5807	17248	0.54	0.44, 0.66
Other	10	393	1018	0.98	0.47, 1.81
Cumulative % Revision	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
C/C	1.4 (1.3, 1.6)	2.5 (2.3, 2.7)	3.2 (3.0, 3.5)	4.0 (3.7, 4.4)	4.3 (3.9, 4.7)
C/P	1.3 (1.1, 1.4)	2.2 (2.0, 2.3)	2.8 (2.6, 3.1)	3.7 (3.4, 4.0)	5.1 (4.4, 5.9)
M/M	1.6 (1.4, 1.8)	3.9 (3.6, 4.2)	5.2 (4.8, 5.7)	6.3 (5.7, 6.9)	7.7 (6.0, 9.7)
M/P	1.5 (1.4, 1.6)	2.5 (2.3, 2.6)	3.3 (3.2, 3.5)	4.2 (4.0, 4.5)	5.2 (4.8, 5.5)
CM/P	1.1 (0.9, 1.4)	1.7 (1.4, 2.1)	2.1 (1.6, 2.6)	NR	NR

*C/C is ceramic-on-ceramic; M/M is metal-on-metal; C/P is ceramic-on-polyethylene; M/P is metal-on-polyethylene; CM/P is ceramicised metal-on-polyethylene; NR means not reported because this was the first time ceramicised metal was reported and there were no data for these periods. These data should be interpreted with caution as ceramicised metal has been largely combined with modified polyethylene.

Table 2. Risk of revision after Metal-on-Metal and other bearing surfaces from the Australian joint registry (Primary diagnosis OA)

Surfaces compared	Time period	HR**	95% CI	p value
C/C* vs. M/M*	6 Mth+	0.57	0.5, 0.64	<0.001
C/P* vs. M/M	Entire period	0.60	0.54, 0.67	<0.001
M/P* vs. M/M	9 Mth+	0.59	0.53, 0.66	<0.001
CM/P* vs. M/M	3 Mth+	0.32	0.24, 0.43	<0.001

*C/C is ceramic-on-ceramic; M/M is metal-on-metal; C/P is ceramic-on-polyethylene; M/P is metal-on-polyethylene; CM/P is ceramicised metal-on-polyethylene

**Hazard Ratio Analyses were conducted by the Australian joint registry using Cox regression adjusting for age and gender.

Table 3. Revision Rate after Metal-on-Metal Hip Replacement and Hip Resurfacing from the U.K. and Wales Registry

Prosthesis Type	N	One Yr Revision Rt	Three Yr. Revision Rt	Five Yr. Revision Rt.
LHMoM THR*	8882	1.3 (1.1, 1.6)	4.7 (4.2, 5.4)†	7.8 (6.6, 9.3)†
Resurfacing	13853	2.1 (1.9, 2.3)	4.3 (4.0, 4.8)†	6.3 (5.7, 7.0)†
Cemented**	99359	0.6 (0.6, 0.7)	1.4 (1.3, 1.5)	2.0 (1.8, 2.1)
Cementless**	62937	1.3 (1.2, 1.4)	2.5 (2.4, 2.7)	3.4 (3.2, 3.7)
Hybrid**	31662	0.9 (0.8, 1.0)	1.8 (1.6, 1.9)	2.7 (2.4, 3.0)

*Large head metal-on-metal total hip resurfacing; Large head is defined as ≥ 36 mm

**No indication to the type of bearing surface used was given

†U.K./Wales' reported values that were statistically different from the other implants used

Hip systems are not all the same as different hip components are used in different patients. Further analyses considering additional variables are necessary, as these variables as well as others, are important in clinical decision-making and, therefore, warrant additional discussion. Many known and unknown variables potentially affect the outcome of hip implantation, and the most meaningful results would be obtained from a randomized controlled trial accounting for the variables of interest, thereby reducing the chance of spurious conclusions.

The Australian registry dichotomized hip systems into those with a femoral head component ≤ 28 mm and > 28 mm which led to similar revision rates among non-metal-on-metal bearing surfaces. For metal-on-metal bearing surfaces, there was a clear difference between the two sizes with those > 28 mm and ≤ 28 mm having 1.33 and 0.65 hips revised per 100 observed years, respectively. This finding supports the U.K./Wales registry showing LHMoM THA having the highest revision rate when compared to other hip systems at five years. The Australian registry had data for > 28 mm metal-on-metal hip through seven years, and at that point, metal-on-metal had the highest cumulative percent revision of 7.7%. These results are shown in Table 4 below.

After the Australian registry conducted a Cox regression analysis dichotomizing femoral head components by those ≤ 28 mm and > 28 mm, the findings shown in Tables 5-6 were statistically significant for femoral heads > 28 mm. In each instance, having a metal-on-metal femoral head > 28 mm, compared to other bearing surfaces of the same size, had greater risk of revision. The results were not statistically significant for femoral head components ≤ 28 mm.

Table 4. Revision Rates after Metal-on-Metal and other bearing surfaces stratified by component head size from the Australian Registry (Primary diagnosis OA)

Bearing Surface	Head Size	N revised	N total	Obs. Years	Revisions/100 Obs. Yrs	95% CI
C/C	≤28mm	215	5259	28088	0.77	0.67, 0.87
C/C	>28mm	578	24686	82959	0.7	0.64, 0.76
C/P	≤28mm	744	24325	119017	0.63	0.58, 0.67
C/P	>28mm	159	10235	24207	0.66	0.56, 0.77
M/M	≤28mm	108	2746	16618	0.65	0.53, 0.78
M/M	>28mm	559	15062	41885	1.33	1.23, 1.45
M/P	≤28mm	1405	40236	202616	0.69	0.66, 0.73
M/P	>28mm	413	22314	47797	0.86	0.78, 0.95
CM/P	≤28mm	43	1819	6817	0.63	0.46, 0.85
CM/P	>28mm	50	3988	10431	0.48	0.36, 0.63
Cumulative % Revision	Head Size	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
C/C	≤28mm	1.9 (1.6, 2.3)	3.2 (2.7, 3.7)	4.1 (3.5, 4.7)	4.8 (4.2, 5.5)	4.9 (4.3, 5.6)
C/C	>28mm	1.3 (1.2, 1.5)	2.3 (2.1, 2.5)	3.0 (2.7, 3.3)	3.9 (3.5, 4.3)	4.3 (3.7, 4.8)
C/P	≤28mm	1.4 (1.2, 1.5)	2.3 (2.1, 2.5)	3.0 (2.8, 3.2)	3.8 (3.5, 4.1)	5.2 (4.5, 6.0)
C/P	>28mm	1.0 (0.8, 1.2)	1.9 (1.6, 2.2)	2.3 (1.9, 2.8)	4.1 (2.9, 5.9)	NR
M/M	≤28mm	1.4 (1.0, 1.9)	2.8 (2.3, 3.5)	3.6 (3.0, 4.4)	4.4 (3.6, 5.3)	5.1 (3.8, 6.9)
M/M	>28mm	1.6 (1.4, 1.9)	4.2 (3.8, 4.6)	5.9 (5.3, 6.5)	7.7 (6.6, 9.0)	NR
M/P	≤28mm	1.5 (1.4, 1.7)	2.5 (2.4, 2.7)	3.4 (3.2, 3.6)	4.4 (4.1, 4.6)	5.3 (5.0, 5.7)
M/P	>28mm	1.5 (1.3, 1.6)	2.3 (2.1, 2.5)	2.9 (2.5, 3.2)	3.2 (2.7, 3.8)	NR
CM/P	≤28mm	1.3 (0.9, 1.9)	2.2 (1.6, 3.0)	2.8 (2.0, 3.9)	NR	NR
CM/P	>28mm	1.0 (0.8, 1.4)	1.4 (1.0, 1.9)	1.5 (1.1, 2.0)	NR	NR

*C/C is ceramic-on-ceramic; M/M is metal-on-metal; C/P is ceramic-on-polyethylene; M/P is metal-on-polyethylene; CM/P is ceramicised metal-on-polyethylene; NR means not reported because this was the first time ceramicised metal was reported and there were no data for these periods. These data should be interpreted with caution as ceramicised metal has been largely combined with modified polyethylene.

Table 5. Risk of Revision after Metal-on-Metal and other bearing surfaces stratified by component head size from the Australian Registry (Primary diagnosis OA)

≤ 28mm Head size	Time Period	HR*	95% CI	p value
C/C vs. M/M	Entire period	1.18	0.93, 1.49	p=0.164
	C/P vs. M/M	Entire period	0.94	0.77, 1.16
M/P vs. M/M	Entire period	1.06	0.87, 1.29	p=0.563
	CM/P vs. M/M	Entire period	0.84	0.59, 1.20

*Hazard Ratio

Table 6. Risk of Revision after Metal-on-Metal and other bearing surfaces stratified by component head size from the Australian Registry (Primary diagnosis OA)

> 28mm Head size	Time Period	HR*	95% CI	p value
C/C vs. M/M	1.5yrs+	0.41	0.34, 0.49	<0.001
	C/P vs. M/M	Entire period	0.46	0.39, 0.55
M/P vs. M/M	9Mth+	0.32	0.26, 0.39	<0.001
	CM/P vs. M/M	3Mth+	0.19	0.12, 0.30

*Hazard Ratio

Based on these data, metal-on-metal total hip arthroplasty and hip resurfacing have higher revision rates than other commonly used bearing surfaces. Some caution should be exercised when interpreting the results as registries did not employ rigorous multivariate analyses to address all potential confounding variables. One can still have a high level of confidence in the results because these registries were evaluated as high quality.

VALIDATED PATIENT-ORIENTED OUTCOME MEASURES

More research, particularly randomized controlled trials, is needed to compare metal-on-metal THA with other commonly used bearing surfaces before conclusions can be reached as to the non-inferiority of metal-on-metal THA.

A table summarizing the results of three studies is shown in Table 7.

- Two high strength studies^{15,23} used the Harris Hip Score comparing metal-on-metal total hip arthroplasty with metal-on-polyethylene total hip arthroplasty. Each of these studies reported that baseline outcome measurements were not statistically different from each other. After two, five, and ten years, all treatment arms had statistically significant improvement, but were not statistically different from each other.
- One high strength study²³ used the Oxford Hip Score to compare metal-on-metal total hip arthroplasty with metal-on-polyethylene total hip arthroplasty. Baseline outcome measures were not statistically different from each other. At five and ten years, each group had statistically significant improvement, but was not statistically different from each other.
- One high strength study¹⁵ and one moderate strength study²² used the WOMAC instrument. One compared metal-on-metal hip resurfacing with metal-on-metal total hip arthroplasty, and the other compared metal-on-metal total hip arthroplasty with metal-on-polyethylene total hip arthroplasty. Patients receiving hip resurfacing had statistically significantly better one and two year WOMAC scores than metal-on-metal THA, but hip resurfacing patients had different post-operative protocols, a lower BMI, were slightly younger, and there was no clinical relevance. No significant differences were seen at two years between metal-on-metal THA and metal-on-polyethylene THA patients.
- Finally, one high strength study¹⁵ used the Short-Form 12 to compare metal-on-metal THA with metal-on-polyethylene THA. At baseline there were no statistically significant differences between groups, and each group had statistically significant improvement at two years, but was not statistically different from each other.

Table 7. Summary of Subjective Outcome Measures from Primary Literature

Author	N	Outcome Measure Used	Comparison	Group 1 Baseline Avg. (SD)	Group 2 Baseline Avg. (SD)	P-value	Follow-up Years	Group 1 Avg. (SD) at Follow-up	Group 2 Avg. (SD) at Follow-up	P-value
MacDonald 2003¹⁵	41	Harris Hip Score	MoM THA vs MoP THA	46.5 (13.4)	46.6 (12.4)	p > .05	2	91.6 (11.5)	92(12.5)	p > .05
Zijlstra 2009²³	195	Harris Hip Score	MoP THA vs MoM THA	46 (13)	48 (15)	0.746	5	87 (13)	90 (7)	0.791
Zijlstra 2009²³	195	Harris Hip Score	MoP THA vs MoM THA	46 (13)	48 (15)	0.746	10	87 (10)	86 (10)	0.441
Zijlstra 2009²³	195	Oxford Hip Score	MoP THA vs MoM THA	40 (8)	40 (8)	0.661	5	18 (8)	19 (8)	0.515
Zijlstra 2009²³	195	Oxford Hip Score	MoP THA vs MoM THA	40 (8)	40 (8)	0.661	10	24 (9)	27 (13)	0.494
Vendittoli 2010²²	137	WOMAC	MoM HR vs THA	52.7 (15.4)	54.4 (18.3)	0.548	2	5.7 (8.6)	9 (11.9)	0.007
MacDonald 2003¹⁵	41	WOMAC	MoM THA vs MoP THA	58.7 (15.3)	59.1 (14.5)	p > .05	2	17.3 (15.1)	19.9 (18.9)	p > .05
MacDonald 2003¹⁵	41	SF-12	MoM THA vs MoP THA	Phy: 32.8 (10.2), M: 53.5 (9.6)	Phy: 31.1 (4.7), M: 53.0 (11.6)	p > .05	2	Phy: 47.2 (9.6), M: 55.9 (6.2)	Phy: 55 (11.7), M: 53.3 (11.7)	p > .05

Question #2: What are the patient, implant, and surgical factors that best predict successful/unsuccessful outcomes of metal-on-metal hip replacement?

To address this question, we included data from the Australian and U.K./Wales registries addressing patient and implant characteristics that determined success/failure of metal-on-metal hip systems.

The registries stratified risk according to patient gender and age, and implant head size using Cox regression analysis. Registry analyses did not account for all possible variables that could predict success/failure of metal-on-metal hip systems. Additional variables could include BMI, smoking status, ASA score, socioeconomic variables, etc. Rigorous multivariate analysis is needed to determine prognostic variables that best determine success or failure. To fully interpret a multiple regression analysis authors should: clearly identify the variables of interest, report results on all these variables, ensure a correct sample size for the number of variables included and the outcome examined, explain if variables are coded so there is no ambiguity when interpreting results, address potential issues of collinearity, state the type of model used (e.g. forward inclusion, hierarchical), report a goodness-of-fit statistic so readers can interpret the model and validate the model (e.g., using split-half analysis).^{3,4}

METAL-ON-METAL TOTAL HIP ARTHROPLASTY (THA) IMPLANT HEAD SIZE FACTORS

In metal-on-metal total hip arthroplasty, using Australian registry data, differences were reported when considering head size as a variable and revision rates as the outcome (see Tables 8-11). Metal-on-metal THA using head sizes ≤ 28 mm yielded the lowest number of revisions and head sizes > 40 mm had the highest number of revisions per 100 observed years. After Cox regression analysis by the registry, head components > 32 mm were 2.42 times as likely to be revised as head components ≤ 32 mm. Components 36-40mm were 1.75 times as likely to be revised as components ≤ 28 mm, and components greater than 40mm were 2.37 times as likely to be revised as components ≤ 28 mm. These comparisons are only for metal-on-metal total hip arthroplasty from the Australian registry adjusting for age and gender. Randomized controlled trials or a more rigorous multivariate analysis is needed to help control for variables that potentially confound interpretation.

Table 8. Revision rates by implant head size group (Primary diagnosis OA)

Metal-on-Metal THA	N Revised	N total	Obs Years	Revisions/100 Obs Years	95% CI
≤ 28 mm	108	2746	16618	0.65	0.53, 0.78
30-32mm	57	1843	5873	0.97	0.74, 1.26
36-40mm	152	3262	11320	1.34	1.14, 1.57
> 40 mm	350	9957	24692	1.42	1.27, 1.57
Cumulative % Revision	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
≤ 28 mm	1.4 (1.0, 1.9)	2.8 (2.3, 3.5)	3.6 (3.0, 4.4)	4.4 (3.6, 5.3)	5.1 (3.8, 6.9)
30-32mm	1.7 (1.2, 2.4)	3.3 (2.5, 4.3)	4.2 (3.2, 5.6)	NR	NR
36-40mm	2.1 (1.6, 2.6)	4.2 (3.5, 5.0)	6.0 (5.1, 7.1)	8.2 (6.6, 0.1)	NR
> 40 mm	1.5 (1.3, 1.8)	4.4 (3.9, 4.9)	6.4 (5.5, 7.4)	NR	NR

Table 9. Revision Rates Comparing Implant Heads ≤32mm and >32mm (Primary diagnosis OA)

Metal-on-Metal THA	N Revised	N Total	Obs Yrs	Revisions/100 Obs Yrs	95% CI
≤ 32mm	165	4589	22491	0.73	0.63, 0.85
> 32mm	502	13219	36012	1.39	1.27, 1.52
Cumulative % Revision	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
≤ 32mm	1.5 (1.2, 1.9)	3.0 (2.5, 3.6)	3.8 (3.3, 4.5)	4.5 (3.9, 5.3)	5.3 (4.0, 6.9)
> 32mm	1.6 (1.4, 1.9)	4.3 (3.9, 4.7)	6.2 (5.6, 6.9)	8.4 (7.1, 9.9)	NR

Table 10. Hazard ratios for metal-on-metal THA comparing implant head sizes (Primary diagnosis OA)

Metal-on-Metal THA	Time Period	HR*	95% CI	p value
30-32mm vs. ≤ 28mm	Entire period	1.24	0.90, 1.73	0.191
36-40mm vs. ≤ 28mm	Entire period	1.75	1.36, 2.25	<0.001
> 40mm vs. ≤ 28mm	1.5Yrs+	2.37	1.78, 3.14	<0.001

*HR is the hazard ratio determined using Cox proportional hazard modeling adjusting for age and gender

Table 11. Hazard ratios for metal-on-metal THA comparing implant head sizes ≤ 32mm and > 32mm (Primary diagnosis OA)

Metal-on-Metal THA	Time Period	HR*	95% CI	p value
> 32mm vs. ≤ 32mm	2 Yrs+	2.42	1.79, 3.29	<0.001

*HR is the hazard ratio determined using Cox proportional hazard modeling adjusting for age and gender

PATIENT AGE GROUP FACTORS

Revision rates across age groups did not differ significantly (see Table 12). Patients 65-74 had the lowest revision rate of 1.08 per 100 observed years and patients ≥ 75 had the highest with 1.22 per 100 observed years. Using patients aged 65-74 as the referent group seen in Table 13, no statistically significant differences in risk of revision were seen. As stated above, this analysis did not address all potential confounding variables, and only adjusted for gender differences. A more rigorous multivariate analysis would need to be employed to mitigate the chances of drawing spurious conclusions.

Table 12. Revision Rates by Age Group (Primary diagnosis OA)

Metal-on-Metal	N Revised	N total	Obs Years	Revisions/100 Obs Years	95% CI
<55	139	3425	11708	1.19	1.00, 1.40
55-64	220	5880	19417	1.13	0.99, 1.29
65-74	201	5564	18634	1.08	0.93, 1.24
≥ 75	107	2939	8744	1.22	1.00, 1.48
Cumulative % Revision	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
<55	1.6 (1.2, 2.1)	4.4 (3.7, 5.3)	5.6 (4.7, 6.7)	6.4 (5.3, 7.8)	6.4 (5.3, 7.8)
55-64	1.4 (1.1, 1.7)	3.8 (3.3, 4.4)	5.3 (4.6, 6.2)	6.9 (5.8, 8.1)	8.9 (5.8, 13.6)
65-74	1.7 (1.4, 2.1)	3.7 (3.2, 4.3)	4.9 (4.2, 5.7)	5.7 (4.9, 6.8)	NR
≥ 75	1.9 (1.5, 2.5)	3.9 (3.1, 4.7)	5.3 (4.2, 6.5)	5.8 (4.6, 7.3)	NR

Table 13. Hazard ratios for metal-on-metal THA comparing age groups (Primary diagnosis OA)

	Time Period	HR*	95% CI	p value
<55 vs. 65-74	Entire period	1.14	0.92, 1.41	0.241
55-64 vs. 65-74	Entire period	1.05	0.87, 1.27	0.617
≥ 75 vs. 65-74	Entire period	1.06	0.84, 1.34	0.641

*Hazard ratios adjusted for gender.

GENDER FACTORS

After Cox regression analysis, it was demonstrated that females were 1.32 times as likely to need revision as males. This regression analysis only accounted for age differences, and not all potentially important variables that may confound the interpretation.

Table 14. Revision Rates by Gender (Primary diagnosis OA)

Metal-on-Metal	N Revised	N Total	Obs Yrs	Revisions/100 Obs Yrs	95% CI
Male	322	9937	32271	1	0.89, 1.11
Female	345	7871	26232	1.32	1.18, 1.46
Cumulative % Revision	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
Male	1.4 (1.2, 1.6)	3.3 (2.9, 3.7)	4.5 (4.0, 5.1)	4.5 (3.9, 5.3)	8.1 (5.6, 11.6)
Female	1.9 (1.6, 2.2)	4.6 (4.1, 5.2)	6.1 (5.5, 6.9)	8.4 (7.1, 9.9)	6.9 (6.1, 7.9)

Table 15. Hazard ratios for metal-on-metal THA comparing gender (Primary diagnosis OA)

	Time Period	HR*	95% CI	p value
Female vs. Male	Entire Period	1.32	1.14, 1.54	<0.001

*HR is adjusted for age differences

AGE AND FEMORAL COMPONENT SIZE INTERACTION

The Australian registry addressed the interaction between age and femoral head size. The data indicated that head size was an important variable concerning implant survival shown in Tables 16-17. Patients under 65 years with implants ≤ 32 mm had the lowest revision rate of 0.66 per 100 observed years. The highest revision rate was for patients younger than 65 with implants greater the 32mm. When these two groups were compared, the hazard ratio showed that being < 65 with ≤ 32 mm head size reduced that chance of revision by 48% over the study period. Comparing individuals ≥ 65 and components ≤ 32 mm with individuals of the same age, but with > 32 mm head sizes yielded significant results in favor of those with the smaller head size; they had a 23% reduction in revision risk over the study period. Having a larger femoral head component put individuals at higher revision risk than individuals of the same age, but with smaller femoral head components. This analysis adjusted for gender differences, but not all potential confounding variables.

Table 16. Revision Rates by Implant Head Size and Age Group (Primary diagnosis OA)

Metal-on-Metal	N Revised	N total	Obs Years	Revisions/100 Obs Years	95% CI
≤ 32 mm and < 65	80	2392	12160	0.66	0.52, 0.82
≤ 32 mm and ≥ 65	85	2197	10332	0.82	0.66, 1.02
> 32 mm and < 65	279	6913	18966	1.47	1.30, 1.65
> 32 mm and ≥ 65	223	6306	17046	1.31	1.14, 1.49
Cumulative % Revision	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
≤ 32 mm and < 65	1.2 (0.9, 1.8)	2.8 (2.1, 3.5)	3.5 (2.8, 4.5)	4.3 (3.4, 5.3)	4.4 (3.5, 5.5)
≤ 32 mm and ≥ 65	1.8 (1.3, 2.4)	3.3 (2.6, 4.2)	4.1 (3.3, 5.2)	4.8 (3.9, 6.0)	NR
> 32 mm and < 65	1.5 (1.2, 1.8)	4.6 (4.0, 5.2)	6.7 (5.8, 7.7)	9.6 (7.7, 11.9)	NR
> 32 mm and ≥ 65	1.8 (1.5, 2.1)	4.0 (3.4, 4.6)	5.7 (4.9, 6.7)	6.9 (5.4, 8.8)	NR

Table 17. Hazard ratios for metal-on-metal THA comparing the interaction between age and implant head size (Primary diagnosis OA)

	Time Period	HR*	95% CI	p value
< 65 ys ≤ 32 mm vs. < 65 ys > 32 mm	Entire period	0.52	0.40, 0.67	<0.001
< 65 ys > 32 mm vs. ≥ 65 ys > 32 mm	Entire period	1.16	0.97, 1.38	0.107
< 65 ys ≤ 32 mm vs. ≥ 65 ys ≤ 32 mm	Entire period	0.83	0.61, 1.13	0.243
≥ 65 ys ≤ 32 mm vs. ≥ 65 ys > 32 mm	Entire period	0.77	0.60, 1.00	0.049

*HR adjusted for gender differences

GENDER AND FEMORAL COMPONENT SIZE INTERACTION

Seen in Tables 18-19, the Australian registry indicated there may be a gender/implant size effect. Males and females with components > 32mm had the highest revision rates. Females and males with components ≤ 32mm had no significant differences in revision risk. Comparing males with components ≤ 32mm to males with components >32mm yielded borderline results suggesting that ≤ 32mm is protective against revision. The same comparison done in females yielded statistically significant results showing females with components ≤ 32mm reduces the risk of revision 50%. These results support data previously presented. Individuals receiving metal-on-metal total hip arthroplasty with larger femoral head components, regardless of age and gender, have a higher revision risk. A more rigorous multivariate analysis is needed to avoid misinterpretation due to confounding variables.

Table 18. Revision Rates by Gender and Implant Head Size (Primary diagnosis OA)

Metal-on-Metal	N Revised	N total	Obs Years	Revisions/100 Obs Years	95% CI
Male ≤32mm	88	2363	11913	0.74	0.59, 0.91
Male >32mm	234	7574	20358	1.15	1.01, 1.31
Female ≤32mm	77	2226	10578	0.73	0.57, 0.91
Female >32mm	268	5645	15654	1.71	1.51, 1.93
Cumulative % Revision	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
Male ≤32mm	1.6 (1.2, 2.2)	3.0 (2.4, 3.9)	3.8 (3.1, 4.8)	4.7 (3.8, 5.9)	5.7 (3.9, 8.4)
Male >32mm	1.3 (1.1, 1.6)	3.4 (3.0, 4.0)	5.0 (4.3, 5.9)	8.2 (6.3, 10.6)	NR
Female ≤32mm	1.4 (1.0, 2.0)	3.0 (2.3, 3.8)	3.8 (3.0, 4.8)	4.3 (3.4, 5.4)	NR
Female >32mm	2.1 (1.7, 2.5)	5.4 (4.8, 6.2)	7.7 (6.7, 8.9)	8.8 (7.2, 10.7)	NR

Table 19. Hazard ratios for metal-on-metal THA comparing the interaction between gender and implant head size (Primary diagnosis OA)

	Time Period	HR*	95% CI	p value
Male ≤32mm vs. Male >32mm	Entire period	0.78	0.61, 1.00	0.05
Male >32mm vs. Female >32mm	Entire period	0.66	0.56, 0.79	<0.001
Male ≤32mm vs. Female ≤32mm	Entire period	1.04	0.76, 1.41	0.815
Female ≤32mm vs. Female >32mm	Entire period	0.5	0.39, 0.64	<0.001

*HR is adjusted for age differences

U.K./WALES RESULTS

The U.K./Wales registry reported the revision rates and hazard ratios of Large-Head ($\geq 36\text{mm}$) Metal-on-Metal Total Hip Arthroplasty (LHMoM THA) compared to a cemented total hip arthroplasty with an unspecified bearing surface shown in Tables 20-21. For each age group addressed in the registry, LHMoM THA had statistically significant hazard ratios showing increases in revision risk between 52-273%. (Revision rates and hazard ratios for the referent group (cemented THA) can be found in Appendix VI) These data are consistent with the data from the Australian registry suggesting individuals with large femoral head components, regardless of age and gender, have significant increases of revision risk associated with metal-on-metal total hip arthroplasty.

Table 20. Revision rate and hazard ratio for the interaction between being male, age group and implant size from the U.K./Wales Registry**

Male LHMoM* THA	N	3 Yr. Revision Rt.	5 Yr. Revision Rt.	HR†
<55	1564	4.5 (3.3, 6.1)	6.4 (4.5, 8.9)	1.52 (1.01, 2.28)
55-64	1857	4.0 (2.9, 5.4)	5.5 (3.9, 7.8)	2.19 (1.56, 3.07)
65+	1444	3.4 (2.2, 5.1)	6.4 (3.7, 11.0)	2.21 (1.55, 3.16)

*Large head metal-on-metal defined as $\geq 36\text{mm}$

**U.K./Wales only provides data pertaining to metal-on-metal THA for large head implants

†Hazard ratio calculated using cemented total hip arthroplasty with unspecified bearing surface as referent group

Table 21. Revision rate and hazard ratio for the interaction between being female, age group, and implant head size from the U.K./Wales Registry**

Female LHMoM* THA	N	3 Yr. Revision Rt.	5 Yr. Revision Rt.	HR†
<55	1037	6.3 (4.6, 8.6)	9.2 (5.9, 14.1)	2.22 (1.48, 3.32)
55-64	1526	6.7 (5.2, 8.7)	9.5 (7.4, 12.3)	3.73 (2.80, 4.96)
65+	1454	3.6 (2.5, 5.2)	10.5 (6.1, 17.8)	3.65 (2.67, 5.00)

*Large head metal-on-metal defined as $\geq 36\text{mm}$

** U.K./Wales only provides data pertaining to metal-on-metal THA for large head implants

† Hazard ratio calculated using cemented total hip arthroplasty with unspecified bearing surface as referent group

METAL-ON-METAL HIP RESURFACING ENGLISH OUTCOMES

Results from the U.K./Wales registry for hip resurfacing showed statistically significant increases in revision risk compared to cemented total hip arthroplasty with an unspecified bearing surface. Females older than 65 years were 6.59 times as likely to have their hip revised as females of the same age with cemented total hip arthroplasty. Males older than 65 years were 3.85 times as likely to have their hip revised as males of the same age with cemented total hip arthroplasty. Borderline results were found for males <55 years suggesting they, too, were at an increased revision risk.

Table 22. Results from U.K./Wales registry on hip resurfacing

Male Hip Resurfacing	N	3 Yr. Revision Rt.	5 Yr. Revision Rt.	HR†
<55	4215	3.6 (3.0, 4.3)	5.6 (4.5, 6.9)	1.37 (0.98, 1.92)
55-64	3852	3.3 (2.7, 3.9)	4.5 (3.6, 5.6)	2.14 (1.63, 2.80)
65+	980	5.4 (4.1, 7.3)	6.0 (4.4, 8.2)	3.85 (2.83, 5.23)
Female Hip Resurfacing	N	3 Yr. Revision Rt.	5 Yr. Revision Rt.	HR
<55	2647	5.4 (4.5, 6.5)	8.3 (6.8, 10.0)	2.32 (1.67, 3.21)
55-64	1854	5.4 (4.4, 6.6)	7.9 (6.3, 9.8)	3.61 (2.78, 4.68)
65+	305	7.1 (4.5, 11.1)	8.8 (5.2, 14.7)	6.59 (4.17, 10.41)

† Hazard ratio calculated using cemented total hip arthroplasty with unspecified bearing surface as referent group

AUSTRALIAN OUTCOMES

The Australian registry addressed a greater number of variables regarding hip resurfacing compared to the U.K./Wales registry. Like metal-on-metal total hip arthroplasty, hip resurfacing data addressed patient age and gender, and implant head size. Unlike total hip arthroplasty, hip resurfacing analyses addressed revision rates and hazard ratios comparing patients by primary diagnosis for hip replacement.

The cumulative percent revision for metal-on-metal hip resurfacing was 7.2% over nine years, second only to metal-on-metal total hip arthroplasty (7.7%).

Table 23. Revision Rate for Metal-on-metal Hip Resurfacing (Primary diagnosis OA)

Cumulative % Revision of Hip Resurfacing	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
Total Resurfacing	1.8 (1.6, 2.1)	3.1 (2.8, 3.4)	4.2 (3.9, 4.7)	5.8 (5.3, 6.5)	7.2 (6.2, 8.4)

REVISION RATES BY PRIMARY DIAGNOSIS

Revision rates and Cox regression models calculated the revision risk stratified by primary diagnosis seen in Table 24. The most common reason for hip resurfacing was osteoarthritis. The only primary diagnosis that led to significantly increased revision risk was developmental dysplasia of the hip (DDH). DDH patients were 1.94 times as likely to have revision surgery as patients with osteoarthritis.

Table 24. Revision Rates for Metal-on-Metal Hip Resurfacing by Primary Diagnosis

Revision Rates for THR by Primary Diagnosis	N Revised	N Total	Obs. Years	Revision/100 Obs. Yrs	95% CI
Osteoarthritis	490	12587	52020	0.94	0.86, 1.03
Developmental Dysplasia	35	359	1682	2.08	1.45, 2.89
AVN	13	246	1192	1.09	0.58, 1.87
Other	10	115	527	1.9	0.91, 3.49
Cumulative % Revision of THR by Primary Diagnosis	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
Osteoarthritis	1.8 (1.6, 2.1)	3.1 (2.8, 3.4)	4.2 (3.9, 4.7)	5.8 (5.3, 6.5)	7.2 (6.2, 8.4)
Developmental Dysplasia	2.5 (1.3, 4.8)	5.7 (3.7, 8.8)	11.1 (7.9, 15.6)	14.0 (10.0, 19.5)	NR
AVN	2.5 (1.1, 5.4)	4.7 (2.6, 8.4)	6.0 (3.5, 10.2)	6.0 (3.5, 10.2)	NR
Other	2.6 (0.8, 7.9)	5.7 (2.6, 12.2)	9.5 (5.0, 17.6)	NR	NR

*THR means total hip resurfacing

Table 25. Risk of Revision Comparing Primary Diagnoses

Age and Gender Adjusted Resurfacing HRs by Primary Diagnosis	Time Period	HR*	95% CI	p value
DDH vs. Osteoarthritis	Entire period	1.94	1.36, 2.77	<0.001
AVN vs. Osteoarthritis	Entire period	1.41	0.81, 2.47	0.226
DDH vs. AVN	Entire period	1.37	0.72, 2.61	0.331

*HR is adjusted for age and gender

REVISION RATES BY PATIENT AGE

Patients ≥ 65 years old had the highest revision rate and were 3.51 times as likely to have their hip revised as patients <55 years old. Another significant difference was found comparing patients ≥ 65 years old with patients 55-64, as those older than 65 were 1.49 times as likely to have their hip revised.

The findings from the U.K./Wales and Australian registries agree indicating that the oldest patient populations receiving metal-on-metal hip resurfacing have the greatest revision risk. The analysis done by the Australian registry was only on hip resurfacing patients. Hazard ratios could be higher if the comparison was made using a different bearing surface as the comparator as was done in the U.K./Wales registry. These analyses did not employ a rigorous multivariate approach needed to control for all potential confounding variables.

Table 26. Revision Rates of Metal-on-Metal Hip Resurfacing by Age Groups (Primary diagnosis OA)

Hip Resurfacing Revision Rates by Age	N Revised	N total	Obs Years	Revisions/100 Obs Years	95% CI
<55	227	6377	26170	0.87	0.76, 0.99
55-64	199	5004	20677	0.96	0.83, 1.11
≥ 65	64	1206	5173	1.24	0.95, 1.58
Cumulative % Revision of Hip Resurfacing by Age	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
<55	1.5 (1.2, 1.9)	2.7 (2.3, 3.2)	4.1 (3.5, 4.7)	5.6 (4.8, 6.5)	7.1 (5.6, 9.1)
55-64	1.8 (1.5, 2.2)	3.1 (2.6, 3.6)	4.2 (3.6, 4.9)	5.8 (4.9, 6.8)	NR
≥ 65	3.5 (2.6, 4.7)	4.6 (3.5, 6.0)	5.4 (4.2, 7.0)	7.3 (5.5, 9.7)	NR

Table 27. Risk of Revision Comparing Age Groups (Primary diagnosis OA)

Gender Adjusted Resurfacing HRs by Age	Time Period	HR*	95% CI	p value
55-64 vs. <55	Entire period	1.15	0.95, 1.39	0.154
≥ 65 vs. <55	0-3Mth	3.51	2.22, 5.54	<0.001
≥ 65 vs. <55	3Mth+	1.29	0.91, 1.82	0.154
≥ 65 vs. 55-64	Entire period	1.49	1.12, 1.98	0.006

*HR adjusted for age

REVISION RATES BY IMPLANT HEAD SIZE

Implant head size has shown to be an important factor across both metal-on-metal THA and hip resurfacing. However, individuals receiving hip resurfacing with components ≤ 44 mm were 5.87 times as likely to have their hip revised as individuals receiving an implant ≥ 55 mm. Similar results were found when individuals with 45-49mm components were compared to individuals with ≥ 55 mm. No significant differences were found when 50-54mm femoral heads were compared to those greater ≥ 55 mm.

Table 28. Revision Rates for Metal-on-Metal Hip Resurfacing by Implant Head Size (Primary diagnosis OA)

Revision Rates for THR by Head Size	N Revised	N Total	Obs. Years	Revision/100 Obs. Yrs	95% CI
≤ 44 mm	116	1139	5077	2.28	1.89, 2.74
45-49mm	167	3083	12291	1.36	1.16, 1.58
50-54mm	193	7543	31736	0.61	0.53, 0.70
≥ 55 mm	14	822	2917	0.48	0.26, 0.81
Cumulative % Revision of THR by Head Size	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
≤ 44 mm	3.4 (2.5, 4.6)	7.2 (5.7, 8.9)	10.4 (8.6, 12.6)	13.8 (11.4, 16.7)	NR
45-49mm	2.5 (2.0, 3.2)	4.3 (3.6, 5.1)	5.9 (5.0, 6.9)	8.8 (7.3, 10.5)	NR
50-54mm	1.4 (1.1, 1.7)	2.1 (1.8, 2.4)	2.8 (2.4, 3.2)	3.7 (3.1, 4.3)	4.3 (3.5, 5.3)
≥ 55 mm	1.1 (0.6, 2.2)	1.6 (0.9, 2.9)	2.2 (1.3, 3.7)	2.2 (1.3, 3.7)	NR

Table 29. Risk of Revision for Metal-on-Metal Hip Resurfacing Comparing Implant Head Size (Primary diagnosis OA)

Age and Gender Adjusted Hip Resurfacing HRs by Head Size	Time Period	HR*	95% CI	p value
≤ 44 mm vs. ≥ 55 mm	Entire period	5.87	3.21, 10.71	<0.001
45-49mm vs. ≥ 55 mm	Entire period	3.21	1.78, 5.47	<0.001
50-54mm vs. ≥ 55 mm	Entire period	1.31	0.76, 2.25	0.336

HR is adjusted for age and gender

REVISION RATES BY GENDER AND IMPLANT HEAD SIZE

When incorporating femoral component size, no significant gender differences were found, though intra-gender differences did become evident. Males receiving an implant <50mm were 2.54 times as likely to have their hips revised as males receiving an implant ≥50mm. The same comparison done with females yielded similar results with the smaller component having 3.41 times the revision risk.

Differences between total hip arthroplasty and hip resurfacing were indicated by the data. Larger femoral heads in THA were shown to increase revision risk as opposed to hip resurfacing where larger femoral heads may be protective against revision.

Table 30. Revision Rates for Metal-on-Metal Hip Resurfacing by Gender and Implant Head Size (Primary diagnosis OA)

Revision Rates for THR by Gender and Head Size	N Revised	N Total	Obs. Years	Revision/100 Obs. Yrs	95% CI
	Male <50mm	90	1555	5693	1.58
Male ≥50mm	197	7940	32563	0.6	0.52, 0.70
Female <50mm	193	2667	11675	1.65	1.43, 1.90
Female ≥50mm	10	425	2090	0.48	0.23, 0.88
Cumulative % Revision of THR by Gender and Head Size	1 Yr	3 Yrs	5 Yrs	7 Yrs	9 Yrs
Male <50mm	3.3 (2.5, 4.3)	4.9 (3.9, 6.2)	6.5 (5.2, 8.2)	10.2 (7.9, 13.3)	NR
Male ≥50mm	1.4 (1.2, 1.7)	2.1 (1.8, 2.4)	2.8 (2.4, 3.2)	3.5 (3.0, 4.2)	3.9 (3.2, 4.7)
Female <50mm	2.5 (1.9, 3.1)	5.1 (4.3, 6.1)	7.5 (6.4, 8.7)	10.3 (8.8, 12.1)	NR
Female ≥50mm	0.5 (0.1, 1.9)	1.0 (0.4, 2.7)	1.8 (0.8, 4.1)	3.3 (1.5, 6.9)	NR

Table 31. Risk of Revision for Metal-on-Metal Hip Resurfacing Comparing the Interaction between Gender and Implant Head Size (Primary diagnosis OA)

Age Adjusted Hip Resurfacing HRs by Gender and Head Size	Time Period	HR*	95% CI	p value
Male <50mm vs. Male ≥50mm	Entire period	2.54	1.98, 3.26	<0.001
Male ≥50mm vs. Female ≥50mm	Entire period	1.18	0.63, 2.24	0.601
Male <50mm vs. Female <50mm	Entire period	0.88	0.69, 1.14	0.338
Female <50mm vs. Female ≥50mm	Entire period	3.41	1.80, 6.43	<0.001

*HR is adjusted for age

Question #3: What is the prevalence of adverse clinical problems from metal-on-metal hip replacement compared to other bearing surface combinations?

We sought to determine the prevalence of adverse clinical problems comparing metal-on-metal total hip arthroplasty/resurfacing to other bearing surfaces. All metal-on-metal hip implants wear and cause a release of metal ions²⁴. This wear and elevated ions is believed to cause soft tissue reactions in the periprosthetic space called “Adverse Reactions to Metal Debris (ARMD),” and/or “Adverse Local Tissue Reactions (ALTR).”

The incidence/prevalence of complications, metal ions, and tissue reactions to metal is addressed using five high strength and 13 moderate strength studies.

LOCAL METAL DEBRIS RELEASE

Metal debris release is a concern regarding metal-on-metal hip replacements. Five moderate strength studies^{5, 8, 15, 19, 23} measured cobalt, chromium, and titanium concentrations in patients receiving hip replacements using several bearing surfaces including metal-on-metal. A table summarizing the results is presented in Appendix VI.

- Studies comparing metal-on-metal THA to metal-on-polyethylene showed statistically significant increases in cobalt in the metal-on-metal groups. Cobalt levels ranged from 2.2-6.47 times greater in metal-on-metal groups versus metal-on-polyethylene groups.
- Studies comparing metal-on-metal THA/hip resurfacing to a hip system without a metal component (i.e., ceramic-on-polyethylene, ceramic-on-ceramic) showed statistically significant differences in cobalt concentrations in each instance. Metal-on-metal groups had cobalt levels 1.66-5 times greater than non-metal hip systems.
- Metal-on-metal THA compared to metal-on-polyethylene showed statistically significant increases in chromium. Levels in metal-on-metal patients were 1.92-2.92 times greater than metal-on-polyethylene.
- Metal-on-metal hip resurfacing compared to ceramic-on-ceramic THA showed statistically significant increases in chromium. Levels in hip resurfacing patients were 1.88-3.13 times greater than ceramic-on-ceramic.
- Metal-on-metal THA compared to metal-on-polyethylene showed no significant differences in titanium concentrations.

One moderate strength study¹¹ addressed ion levels in metal-on-metal THA patients with and without pseudotumors, and metal-on-polyethylene patients as controls.

- In each case, cobalt and chromium levels were significantly higher in metal-on-metal patients compared to metal-on-polyethylene.

- In the metal-on-metal group, patients with pseudotumors had cobalt levels 4.84 times higher and chromium levels 5.71 times higher than metal-on-metal patients without pseudotumors.
- Hip joint aspirate was collected from patients with metal-on-metal THA. Significant differences were seen between the two groups with cobalt and chromium levels 13.7 and 7.69 times higher, respectively, in the group with pseudotumors.

One moderate strength study¹⁷ compared metal-on-metal THA and hip resurfacing. No significant differences were seen between treatment arms that would suggest one hip system released more ions than the other. However, ion concentrations were consistent with concentrations found in previous studies.

One moderate strength study¹⁰ addressed prognostic variables of ion release in patients receiving metal-on-metal hip resurfacing. The study found the relationship between metal ions and cup inclination to be exponential, with a sharp increase in ion release when the inclination angle was greater than 45 degrees. The relationship between metal ions and version was negatively exponential, with a sharp decrease in ions when the version angle was less than 20 degrees. The final model predicted that for a unit increase in inclination, the average levels of Co and Cr increased by 4.5%, and 3.3%, respectively. The model also predicted that for a unit increase in version, the average levels of Co and Cr decreased by 2.5 % and 2.0%, respectively. A table summarizing the results with the final model is presented in Appendix VI.

PSEUDOTUMORS

This section will summarize the incidence of pseudotumors and other associated metal ion reactions using two high strength^{7, 12} and two moderate strength studies^{11, 18}.

The incidence of pseudotumors and associated metal ion reactions are summarized in Tables 32-33. The evidence comparing metal-on-metal THA/ hip resurfacing to other bearing surfaces is sparse and drawing conclusions is difficult. More research is needed comparing the incidence of pseudotumors in patients with different bearing combinations. The U.K./Wales registry just began reporting adverse soft tissue reactions in July 2009. Too little data are currently available for reliable statistical analyses.

Table 32. Incidence of Pseudotumors

Column1	N	N of pseudotumors	% pseudotumors
Kwon 2011	158	7	4.43
Glyn-Jones 2009	1224	26	2.12
Langton 2010	660	17	2.58
Ng 2011	206	58	28.15

Table 33. Number of Pseudotumors by Bearing Surface

Column1	Type of implant	N	N of pseudotumors	% pseudotumors
Kwon 2011	MoM HR	194	7	3.61
Glyn-Jones 2009	MoM HR	1267	26	2.05
Ng 2011	MoM HR	6	4	66.67
Ng 2011	MoM THA	26	15	57.69
Ng 2011	CoP THA	8	1	12.5
Ng 2011	MoP THA	166	31	18.67
Langton 2010	MoM HR	573	13	3.2
Langton 2010	MoM THA	87	5	6

IMAGING

Computed Tomography findings from a moderate strength study⁹ reported a median inclination angle of 55° and median version angle of 31° in patients with MRI confirmed periprosthetic lesions from metal-on-metal implants. These findings are supportive of the moderate strength prognostic study¹⁰ on ion release mentioned above suggesting increased ion release in patients with an inclination angle greater than 45° and a version angle greater than 20°. Additional quality research with larger sample sizes is needed to validate the predictive value of component angles and whether or not they correlate with increased incidence of periprosthetic tissue reactions. Imaging study results can be found in Appendix VI.

OTHER COMPLICATIONS

This section summarizes other complications associated with metal-on-metal total hip arthroplasty. Many of the studies included different complications, therefore, four high strength^{6, 13, 16, 22} and three moderate^{14, 18, 21} strength studies were included. Several joint registries report common complications associated with hip arthroplasty in general, but do not separate complications by bearing surfaces.

Complications included groin pain, femoral loosening, acetabular loosening, femoral neck fracture, periprosthetic fracture, acetabular fracture, impingement, sepsis, metal allergy, osteolysis, heterotopic ossification, squeaking, dislocation, aseptic loosening, DVT, neurapraxia (sciatic), and symptomatic leg length discrepancy. Differences in study design, treatment arms, patient demographics, clinical circumstances and small sample size make comparisons and conclusions difficult to draw. A table summarizing this information is presented in Appendix VI.

APPENDICES

APPENDIX I INCLUSION CRITERIA

We used the following criteria to determine whether studies should be included in this systematic review:

1. Study must be of humans.
2. Study must have enrolled 10 or more patients in any arm.
3. Study must have been published in 1995 or later.
4. Study must have a two year follow-up.
5. Study must be published in English.
6. Study must quantitatively express its results.
7. Study must be a full report, published in a peer-reviewed journal. Meeting abstracts, traditional reviews, and text chapters will not be included. Manufacturer marketing information is also excluded.
8. If there are duplicate publications of the same study, the most recent publication will be included unless the earlier publication contains information not in the later one. In this latter case, both publications will be included.
9. Study must use metal to metal, contemporary hybrid fixation resurfacing implants.
10. FDA database for adverse reporting will not be used as rates cannot be determined by the data reported.
11. National registries can be used if enough data is reported.
12. International devices not currently approved for use in the US will be included.
13. The task force will specify what devices have been abandoned and will not consider products that are “off” the market.
14. Only studies of the highest level of available evidence will be included, assuming that there are two or more studies of that higher level. For example, if there are two High Quality studies that address a question, Moderate and Low Quality studies will not be included. If there is only one High Quality study, Moderate Quality studies will be included.
15. We will not consider data on intermediate outcomes. Data on patient-oriented outcomes will take precedence over data on surrogate outcomes. Similarly, we will not consider biomechanical or *in vitro* studies.

APPENDIX II SEARCH STRATEGIES AND DATABASES

To identify studies for this Overview we searched PubMed, EMBASE, and the Cochrane Library through June 24, 2011.

Our PubMed search strategy was:

("Hip"[Mesh] OR "Hip Joint"[Mesh] OR "Hip Prosthesis"[Mesh] OR "Arthroplasty, Replacement, Hip"[Mesh] OR "Arthroplasty, Replacement, Hip/adverse effects"[Mesh] OR "Hip Dislocation"[Mesh] OR "Hip Fractures"[Mesh] OR "Hip Injuries"[Mesh] OR hip[title/abstract] OR hips[Title/Abstract] OR hip[Text Word] OR hips[Text Word])

AND

(metal[Title/Abstract] OR metal[Text Word] OR cobalt[Title/Abstract] OR cobalt[Text Word] OR chromium[Title/Abstract] OR chromium[Text Word])

AND (English[lang])

NOT

(comment[pt] OR editorial[pt] OR letter[pt] OR addresses[pt] OR news[pt] OR "newspaper article"[pt] OR "historical article"[pt] OR "case reports"[pt])

Sorted by study type

#5

#1 AND (Meta-Analysis[ptyp] OR "systematic review"[title/abstract])

#6

#1 AND (Clinical Trial[ptyp] OR Randomized Controlled Trial[ptyp])

#7

#1 NOT (#5 OR #6)

Our EMBASE search strategy was:

('hip prosthesis'/exp OR 'hip arthroplasty'/exp OR 'hip injury'/exp OR 'hip'/exp)

AND

(metal:ab OR metal:ti OR cobalt:ab OR cobalt:ti OR cobalt:de OR chromium:de OR chromium:ab OR chromium:ti)

AND

('hypersensitivity concepts'/exp OR 'reoperation'/exp OR 'toxicity'/exp OR 'pseudotumor'/de OR 'metal implantation'/exp OR 'osteolysis'/exp OR 'joint instability'/exp OR 'postoperative complication'/exp OR 'adverse outcome'/exp OR 'metal ion'/exp)

AND

[english]/lim

AND

([article]/lim OR [conference paper]/lim OR [review]/lim) AND [english]/lim AND [humans]/lim

Sorted by study type

#1 [meta analysis]/lim OR [systematic review]/lim

(10, 1 de-duplicated, Ref IDs 2990-2996)

#2 [controlled clinical trial]/lim OR [randomized controlled trial]/lim

(63, 12 de-duplicated, Ref IDs 2997-3055)

NOT #1 OR #2

(1255, 326 de-duplicated, Ref IDs 3056-4308)

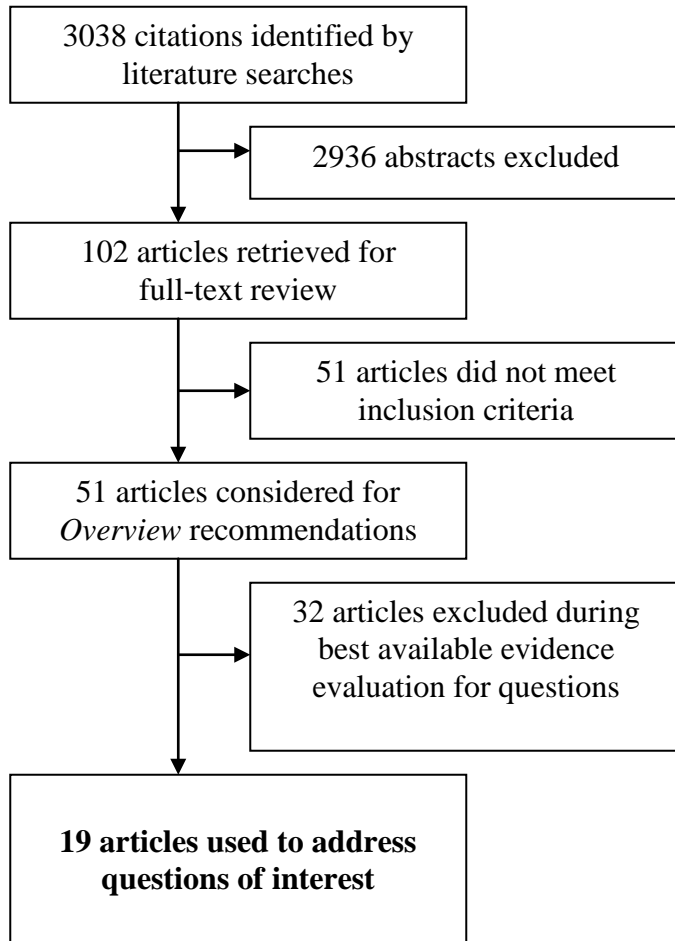
Our Cochrane Library search strategy was:

(hip OR hips OR (hip arthroplasty) OR (hip dislocation) OR (hip fracture) OR femoral OR femur OR acetabul*)

AND

(metal OR cobalt OR chromium)

APPENDIX III STUDY ATTRITION FLOWCHART



INCLUDED STUDIES

Table 34. Author/Year and Title of included studies

Author/Year	Title
Brodner 2003	Serum cobalt levels after metal-on-metal total hip arthroplasty
Ebramzadeh 2011	Failure Modes of 433 Metal-on-Metal Hip Implants: How, Why, and Wear
Glyn-Jones 2009	Risk factors for inflammatory pseudotumour formation following hip resurfacing
Hailer 2011	Elevation of circulating HLA DR(+) CD8(+) T-cells and correlation with chromium and cobalt concentrations 6 years after metal-on-metal hip arthroplasty
Hart 2009	The painful metal-on-metal hip resurfacing
Hart 2011	Insufficient Acetabular Version Increases Blood Metal Ion Levels after Metal-on-metal Hip Resurfacing
Kwon 2011	'Asymptomatic' Pseudotumors After Metal-on-Metal Hip Resurfacing Arthroplasty Prevalence and Metal Ion Study
Langton 2010	Early failure of metal-on-metal bearings in hip resurfacing and large-diameter total hip replacement: A consequence of excess wear
Lavigne 2011	Residual groin pain at a minimum of two years after metal-on-metal THA with a twenty-eight-millimeter femoral head, THA with a large-diameter femoral head, and hip resurfacing
Lazennec 2009	Outcome and serum ion determination up to 11 years after implantation of a cemented metal-on-metal hip prosthesis
MacDonald 2003	Metal-on-metal versus polyethylene in hip arthroplasty: a randomized clinical trial
Marker 2007	Femoral Neck Fractures After Metal-on-Metal Total Hip Resurfacing. A Prospective Cohort Study
Moroni 2008	Does ion release differ between hip resurfacing and metal-on-metal THA?
Ng 2011	Perivascular lymphocytic infiltration is not limited to metal-on-metal bearings
Pattyn 2011	Whole blood metal ion concentrations in correlation with activity level in three different metal-on-metal bearings
Sabah 2011	Magnetic resonance imaging findings in painful metal-on-metal hips: a prospective study
Steffen 2009	Femoral Neck Fractures After Hip Resurfacing
Vendittoli 2010	A comparison of clinical results of hip resurfacing arthroplasty and 28 mm metal on metal total hip arthroplasty: A randomised trial with 3-6 years follow-up
Zijlstra 2009	No superiority of cemented metal-on-metal vs metal-on-polyethylene THA at 5-year follow-up

EXCLUDED STUDIES

Table 35. Author/year, Title, and Reason for Exclusion of Full-Text Reviewed Articles

Author	Title	Exclusion Reason
Amstutz 2006	Metal-on-metal hybrid surface arthroplasty. Surgical Technique	Surgical technique paper
Antoniou 2008	Metal ion levels in the blood of patients after hip resurfacing: a comparison between twenty-eight and thirty-six-millimeter-head metal-on-metal prostheses	Only 1 year follow-up
Antoniou 2008	Metal ion levels in the blood of patients after hip resurfacing: a comparison between twenty-eight and thirty-six-millimeter-head metal-on-metal prostheses	Duplicate Study
Baker 2011	A medium-term comparison of hybrid hip replacement and Birmingham hip resurfacing in active young patients	Low quality
Beaule 2011	A Prospective Metal Ion Study of Large-Head Metal-on-Metal Bearing: A Matched-Pair Analysis of Hip Resurfacing Versus Total Hip Replacement	No baseline values
Benoit 2009	Hueter anterior approach for hip resurfacing: assessment of the learning curve	No patient oriented outcomes
Bozic 2010	Risk of complication and revision total hip arthroplasty among Medicare patients with different bearing surfaces	Low quality
Brodner 1997	Elevated serum cobalt with metal-on-metal articulating surfaces	Less than 2 year follow-up
Brodner 2004	Cup inclination and serum concentration of cobalt and chromium after metal-on-metal total hip arthroplasty	Less than 2 year follow-up
Clarke 2003	Levels of metal ions after small- and large- diameter metal-on-metal hip arthroplasty	Less than 2 year follow-up
Daniel 2004	Metal-on-metal versus polyethylene in hip arthroplasty: a randomized clinical trial	Letter to the Editor
Daniel 2009	Six-year results of a prospective study of metal ion levels in young patients with metal-on-metal hip resurfacings	No baseline values
Daniel 2007	The validity of serum levels as a surrogate measure of systemic exposure to metal ions in hip replacement	Low quality
Daniel 2006	The effect of the diameter of metal-on-metal bearings on systemic exposure to cobalt and chromium	Less than 2 year follow-up
Dastane 2008	Metal-on-metal hip arthroplasty does equally well in osteonecrosis and osteoarthritis	Irrelevant comparison
Davies 2005	An unusual lymphocytic perivascular infiltration in tissues around contemporary metal-on-metal joint replacements	Fewer than 10 patients per group
Dunstan 2005	Metal ion levels after metal-on-metal proximal femoral replacements: a 30-year follow-up	Fewer than 10 patients per group
Fujishiro 2011	Perivascular and diffuse lymphocytic inflammation are not specific for failed metal-on-metal hip implants	Not relevant; no MoM

Author	Title	Exclusion Reason
Grammatopoulos 2010	Optimal acetabular orientation for hip resurfacing	No patient oriented outcomes
Grammatopoulos 2009	Hip resurfacings revised for inflammatory pseudotumour have a poor outcome	Outside scope of TO
Grubl 2007	Long-term follow-up of metal-on-metal total hip replacement	No baseline values
Hailer 2009	Significant increases in serum concentrations of chromium and cobalt but not of nickel and manganese two years after metal-on-metal alloarthroplasty of the hip: a prospective, randomised study comparing metal-on-metal with metal-on-polyethylene bearings	Abstract only
Hall 2009	Patient-reported outcome following metal-on-metal resurfacing of the hip and total hip replacement	Six month follow-up
Hall 2009	Patient-reported outcome following metal-on-metal resurfacing of the hip and total hip replacement	Duplicate Study
Hallows 2011	Serum Metal Ion Concentration: Comparison Between Small and Large Head Metal-on-Metal Total Hip Arthroplasty	Low quality
Hart 2006	The association between metal ions from hip resurfacing and reduced T-cell counts	Less than 2 year follow-up
Hart 2009	Circulating levels of cobalt and chromium from metal-on-metal hip replacement are associated with CD8+ T-cell lymphopenia	Low quality
Hart 2009	Circulating levels of cobalt and chromium from metal-on-metal hip replacement are associated with CD8(+) T-cell lymphopenia	Duplicate Study
Jacobs 1996	Cobalt and chromium concentrations in patients with metal on metal total hip replacements	Fewer than 10 patients per group
Jacobs 1998	Metal release in patients who have had a primary total hip arthroplasty. A prospective, controlled, longitudinal study	All MoP comparisons
Jacobs 2004	Three- to six-year results with the Ultima metal-on-metal hip articulation for primary total hip arthroplasty	Better available evidence
Kim 2011	Cobalt and chromium levels in blood and urine following hip resurfacing arthroplasty with the conserve plus implant	No control group and incomplete statistics
Kim 2008	Causes of Early Failure in a Multicenter Clinical Trial of Hip Resurfacing	1 year follow-up
Korovessis 2003	Zweymueller with metal-on-metal articulation: Clinical, radiological and histological analysis of short-term results	Low quality
Kwon 2010	Lymphocyte proliferation responses in patients with pseudotumors following metal-on-metal hip resurfacing arthroplasty	No patient oriented outcomes
Kwon 2010	Lymphocyte proliferation responses in patients with pseudotumors following metal-on-metal hip resurfacing arthroplasty	Duplicate Study

Author	Title	Exclusion Reason
Ladon 2004	Changes in metal levels and chromosome aberrations in the peripheral blood of patients after metal-on-metal hip arthroplasty	Low quality
Latteier 2011	Gender is a Significant Factor for Failure of Metal-on-Metal Total Hip Arthroplasty	Retrospective review
Lavigne 2011	Comparison of whole-blood metal ion levels in four types of metal-on-metal large-diameter femoral head total hip arthroplasty: the potential influence of the adapter sleeve	Low quality
Lavigne 2008	[Return to sport after hip resurfacing or total hip arthroplasty: a randomized study]	Irrelevant comparison
Le Duff 2007	Metal-on-metal hip resurfacing for obese patients	Not an outcome of interest
Le Duff 2009	Range of motion after stemmed total hip arthroplasty and hip resurfacing: A clinical study	Surrogate outcome (RoM)
Lhotka 2003	Four-year study of cobalt and chromium blood levels in patients managed with two different metal-on-metal total hip replacements	Irrelevant comparison
Li 2009	Hip resurfacing arthroplasty for ankylosing spondylitis	Duplicate Study
Li 2009	Hip Resurfacing Arthroplasty for Ankylosing Spondylitis	Irrelevant study group
Lombardi 2001	Short-term results of the M ² a-taper metal-on-metal articulation	Duplicate Study
Lombardi 2004	Mid-term results of a polyethylene-free metal-on-metal articulation	Low quality
Lombardi 2001	Short-term results of the M(2)a-taper metal-on-metal articulation	Low quality
Lombardi 2004	Mid-term results of a polyethylene-free metal-on-metal articulation	Duplicate Study
MacDonald 2005	Metal on metal versus metal on polyethylene in total hip arthroplasty - a prospective randomised clinical trial	Abstract only
Maezawa 2002	Cobalt and chromium concentrations in patients with metal-on-metal and other cementless total hip arthroplasty	Low quality
Maezawa 2006	Early failure of modern metal-on-metal total hip arthroplasty using a Wagner standard cup	Low quality
Maezawa 2010	Seven years of chronological changes of serum chromium levels after Metasul metal-on-metal total hip arthroplasty	Low quality
Marker 2008	Metal-on-metal hip implants: do they impair renal function in the long-term? A 10-year follow-up study	Irrelevant comparison
McBryde 2008	The influence of surgical approach on outcome in Birmingham hip resurfacing	Very low quality
McBryde 2008	The influence of surgical approach on outcome in Birmingham hip resurfacing	Very low quality
McMinn 2005	Mini-incision resurfacing arthroplasty of hip through the posterior approach	Retrospective review

Author	Title	Exclusion Reason
Migaud 2004	Cementless metal-on-metal hip arthroplasty in patients less than 50 years of age: comparison with a matched control group using ceramic-on-polyethylene after a minimum 5-year follow-up	Low quality
Migaud 2011	Cementless metal-on-metal versus ceramic-on-polyethylene hip arthroplasty in patients less than fifty years of age: a comparative study with twelve to fourteen-year follow-up	Low quality
Mont 2009	Resurfacing is comparable to total hip arthroplasty at short-term followup	Low quality
Naudie 2004	Metal-on-metal versus metal-on-polyethylene bearings in total hip arthroplasty: a matched case-control study	Low quality
Pabinger 2003	Migration of metal-on-metal versus ceramic-on-polyethylene hip prostheses	No patient oriented outcomes
Pattyn 2008	Primary ceramic-on-ceramic total hip replacement versus metal-on-metal hip resurfacing in young active patients	No quantitative results
Pollard 2006	Treatment of the young active patient with osteoarthritis of the hip	Retrospective review
Rasquinha 2006	Serum metal levels and bearing surfaces in total hip arthroplasty	Low quality
Sandiford 2010	Metal on metal hip resurfacing versus uncemented custom total hip replacement--early results	Less than 2 year follow-up
Savarino 2002	Ion release in patients with metal-on-metal hip bearings in total joint replacement: a comparison with metal-on-polyethylene bearings	Low quality
Savarino 2006	Differences in ion release after ceramic-on-ceramic and metal-on-metal total hip replacement. Medium-term follow-up	Low quality
Savarino 2008	Serum ion levels after ceramic-on-ceramic and metal-on-metal total hip arthroplasty: 8-year minimum follow-up	Low quality
Schaffer 1999	Increased blood cobalt and chromium after total hip replacement	Incomplete Statistics
Schmalzried 1996	Factors correlating with long term survival of McKee-Farrar total hip prostheses	Low quality
Sikes 2008	Instability after total hip arthroplasty: treatment with large femoral heads vs constrained liners	Lit review/Retro case study
Stulberg 2010	Early return to function after hip resurfacing: is it better than contemporary total hip arthroplasty?	Low quality
Stulberg 2010	Early Return to Function After Hip Resurfacing. Is It Better Than Contemporary Total Hip Arthroplasty?	Low quality
Stulberg 2008	Results and lessons learned from a United States hip resurfacing investigational device exemption trial	Low quality
Vail 2006	Metal-on-metal hip resurfacing compares favorably with THA at 2 years followup	Low quality

Author	Title	Exclusion Reason
Visuri 2010	A retrospective comparative study of mortality and causes of death among patients with metal-on-metal and metal-on-polyethylene total hip prostheses in primary osteoarthritis after a long-term follow-up	Very low quality
Witzleb 2006	Exposure to chromium, cobalt and molybdenum from metal-on-metal total hip replacement and hip resurfacing arthroplasty	Fewer than 10 patients per group
Witzleb 2006	Exposure to chromium, cobalt and molybdenum from metal-on-metal total hip replacement and hip resurfacing arthroplasty	Fewer than 10 patients per group
Yalcin 2011	Crowe Type I and II DDH managed by large diameter metal-on-metal total hip arthroplasty	Low quality
Zerahh 2011	A prospective randomised study of periprosthetic femoral bone remodeling using four different bearings in hybrid total hip arthroplasty	No outcome of interest
Zhang 2010	Large-diameter metal-on-metal cementless total hip arthroplasty in the elderly	Low quality
Zijlstra 2010	No Superiority of Cemented Metal-on-Metal Over Metal-on-Polyethylene THA in a Randomized Controlled Trial at 10-Year Follow-up	Duplicate Study

APPENDIX IV APPRAISING EVIDENCE QUALITY AND APPLICABILITY STUDIES OF INTERVENTIONS

QUALITY

We judged quality using questions specified before this topic was selected and a computer program determined the final quality rating. We separately evaluated the quality of evidence for each outcome reported by each study. This follows the suggestion of the GRADE working group. We evaluated quality using a domain-based approach using a scheme to allow for evaluation of intervention studies of all designs. The domains we used are whether:

- The study was prospective (with prospective studies, it is possible to have an *a priori* hypothesis to test; this is not possible with retrospective studies.)
- The study was of low statistical power
- The assignment of patients to groups was unbiased
- There was blinding to mitigate against a placebo effect
- The patient groups were comparable at the beginning of the study
- The intervention was delivered in such a way that any observed effects could reasonably be attributed to that intervention
- Whether the instruments used to measure outcomes were valid
- Whether there was evidence of investigator bias

Each quality domain is addressed by one or more questions that are answered “Yes,” “No,” or “Unclear.” These questions and the domains that each addresses are shown below.

To arrive at the quality of the evidence for a given outcome, all domains except the “Statistical Power” domain are termed as “flawed” if one or more questions addressing any given domain are answered “No” for a given outcome, or if there are two or more “Unclear” answers to the questions addressing that domain. The “Statistical Power” domain is considered flawed if a given study did not enroll enough patients to detect a standardized difference between means of 0.2.

Domain flaws lead to corresponding reductions in the quality of the evidence. The manner in which we conducted these reductions is shown in the table below. For example, the evidence reported in a randomized controlled trial (RCT) for any given outcome is rated as “High” quality if zero or one domain is flawed. If two or three domains are flawed for the evidence addressing this outcome, the quality of evidence is reduced to “Moderate,” and if four or five domains are flawed, the quality of evidence is reduced to “Low.” The quality of evidence is reduced to “Very Low” if six or more domains are flawed.

Some flaws are so serious that we automatically term the evidence as being of “Very Low” quality, regardless of a study’s domain scores. These serious design flaws are:

- Non-consecutive enrollment of patients in a case series
- Case series that gave patients the treatment of interest AND another treatment
- Measuring the outcome of interest one way in some patients and measuring it in another way in other patients
- Low statistical power

Quality Questions and Domains for Four Designs of Studies of Interventions

Domain	Question:	Parallel, Contemporary Controls	Crossover Trials	Historical Controls	Case Series
Group Assignment	Stochastic	Yes	Yes	No	No
Group Assignment	Quasi-random Assignment	No	No	No	na*
Group Assignment	Matched Groups	No	No	Yes	No
Group Assignment	Consecutive Enrollment	na	na	na	Yes
Prospective	Prospective	Yes	Yes	Yes	Yes
Blinding	Blinded Patients	Yes	Yes	No	No
Blinding	Blinded Assessors	Yes	Yes	No	No
Blinding	Blinding Verified	Yes	Yes	No	No
Group Comparability	Allocation Concealment	Yes	Yes	No	No
Group Comparability	>80% Follow-up	Yes	Yes	No	Yes
Group Comparability	<20% Completion Difference	Yes	Yes	No	No
Group Comparability	Similar Baseline Outcome Values	Yes	na	Yes	No
Group Comparability	Comparable Pt. Characteristics	Yes	na	Yes	No
Group Comparability	Same Control Group Results	na	Yes	na	na
Group Comparability	Same Experimental Group Results	na	Yes	na	na
Treatment Integrity	Same Centers	Yes	Yes	Yes	No
Treatment Integrity	Same Treatment Duration in and across All Groups	Yes	Yes	Yes	No
Treatment Integrity	Same Concomitant Treatment to All Groups (controlled studies only)	Yes	Yes	Yes	na
Treatment Integrity	No Confounding Treatment (case series only)	na	na	na	Yes
Measurement	Same Instruments	Yes	Yes	Yes	Yes
Measurement	Valid Instrument	Yes	Yes	Yes	Yes
Bias	Article & Abstract Agree	Yes	Yes	Yes	Yes
Bias	All Outcomes Reported	Yes	Yes	Yes	Yes
Bias	A Priori Analysis	Yes	Yes	Yes	Yes
Statistical Power	Statistically Significant	High	High	High	High
Statistical Power	Number of patients in analysis	See below for further information			

*"na" means "not asked"

Relationship between Quality and Domain Scores for Studies of Interventions

Number of Flawed Domains	Quality
0-1	High
2-3	Moderate
4-5	Low
>5	Very Low

APPLICABILITY

We rated the applicability (also called “generalizability” or “external validity”) of the evidence for each outcome reported by each study. As with quality, a computer program that used predetermined questions about specific applicability domains determined applicability ratings. We rated applicability as either “High”, “Moderate”, or “Low” depending on how many domains are flawed. As with quality, a domain is “flawed” if one or more questions addressing that domain is answered “No: or if two or more are answered “Unclear.” We characterized a domain as “flawed” if one or more questions addressing any given domain are answered “No” for a

given outcome, or if there are two or more “Unclear” answers to the questions addressing that domain

Our questions and domains about applicability are those of the PRECIS instrument. The instrument was originally designed to evaluate the applicability of randomized controlled trials, but it can also be used for studies of other design. The questions in this instrument fall into four domains. These domains and their corresponding questions are shown below. The applicability of a study is rated as “High” if it has no flawed domains, as “Low” if all domains are flawed, and as “Moderate” in all other cases as shown in the table below.

Applicability Questions and Domains for Studies of Interventions

Question	Domain
All Types of Patients Enrolled	Participants
Flexible Instructions to Practitioners	Interventions and Expertise
Full Range of Expt'l Practitioners	Interventions and Expertise
Usual Practice Control	Interventions and Expertise
Full Range of Control Practitioners	Interventions and Expertise
No Formal Follow-up	Interventions and Expertise
Usual and Meaningful Outcome	Interventions and Expertise
Compliance Not Measured	Compliance and Adherence
No Measure of Practitioner Adherence	Compliance and Adherence
All Patients in Analysis	Analysis

Relationship between Applicability and Domain Scores for Interventions

Number of Flawed Domains	Applicability
0	High
1, 2, 3	Moderate
4	Low

STUDIES OF SCREENING AND DIAGNOSTIC TESTS

QUALITY

As with our appraisal of the quality of studies of intervention, our appraisal of studies of screening and diagnostic tests is a domain-based approach conducted using *a priori* questions in the table below and scored by a computer program. The questions we used are those of the QUADAS instrument and the six domains we employed are :

- Participants (whether the spectrum of disease among the participants enrolled in the study is the same as the spectrum of disease seen in actual clinical practice)
- Reference Test (whether the reference test , often a “gold standard,” and the way it was employed in the study ensures correct and unbiased categorization of patients as having or not having disease)
- Index Test (whether interpretation of the results of the test under study, often called the “index test”, was unbiased)
- Study Design (whether the design of the study allowed for unbiased interpretation of test results)
- Information (whether the same clinical data were available when test results were interpreted as would be available when the test is used in practice)
- Reporting (whether the patients, tests, and study protocol were described well enough to permit its replication)

Quality Questions and Domains for Studies of Screening and Diagnostic Tests

QUADAS Question:	Domain
Full Patient Spectrum	Participants
Patient Selection Criteria Described	Reporting
Ref. Std. Classifies Condition	Reference Test
Disease Progression Absent	Study Design
Partial Verification Avoided	Study Design
Differential Verification Avoided	Study Design
Independent Ref. Std. and Index Test	Reference Test
Index Test Execution Described	Reporting
Reference Std. Execution Described	Reporting
Index Test Interpreted without Ref. Std. Results	Index Test
Ref. Test Interpreted without Index Test Results	Reference Test
Usual Clinical Data Available	Information
Uninterpretable /Indeterminate Results Reported	Reporting
Withdrawals Explained	Reporting

We considered some design flaws as so serious that their presence automatically guarantees that a study is characterized as being of “Very Low” quality regardless of its domain scores. These flaws are:

- The presence of spectrum bias (occurs when a study does not enroll the full spectrum of patients who are seen in clinical practice. For example, a diagnostic case control study

enrolls only those known to be sick and those known to be well, a patient population quite different from that seen in practice. Because diagnostic case control studies enroll only the easy to diagnose patients, these kinds of studies typically overestimate the abilities of a diagnostic test.)

- Failure to give all patients the reference standard regardless of the index test results
- Non-independence of the reference test and the index text

Because the QUADAS instrument contains reporting questions, quality questions, and questions about whether a study flaw was serious, we arrive at quality ratings in a stepwise answer. First, we determine if one or more serious flaw is present. First, we determine whether any serious flaws are present. If so the quality of evidence is automatically set to “Very Low”. “Serious flaws” are present only if the relevant QUADAS question is answered “No”. We do not use “Unclear” answers to indicate the presence of a serious flaw.

If no serious flaws are present, we then determine a quality rating using all domains *except* the reporting domain. A domain is considered flawed if there are one or more “No” answer or two or more “Unclear” to the questions that address that domain. The relationship between the five quality domains and the reporting domain are shown in the table below:

Relationship between Quality and Domain Scores for Studies of Screening/Diagnostic Tests

Number of Flawed Domains	Quality
0	High
1	Moderate
2	Low
≥3	Very Low

Finally, we use the reporting domain to modify the quality determined in the second step. If one or two of the five QUADAS reporting questions are answered “No”, the quality rating is not changed. If three questions are answered “No” the quality is reduced by one category (e.g., from “High” to “Moderate”), if four reporting questions are “No”, the quality is reduced by two categories (e.g., from “High” to “Low”), and if all five reporting questions are answered “No” the quality is reduced by three categories. (e.g., from “High” to “Very Low”). Two “Unclear” answers are counted as equivalent to one “No” answer in the reporting domain. We also set “floor” so that no study can ever have less than “Very Low” quality. For example, evidence classed as “Low” quality at the second step of our quality appraisal cannot be reduced below “Very Low” even if all of the reporting questions are answered “No.”

APPLICABILITY

We judged the applicability of evidence pertinent to screening and diagnostic tests using a modified version of the PRECIS instrument, implying that the questions are determined *a priori*. As before, scoring was accomplished by a computer. The applicability domains we employed for screening and diagnostic tests were:

- Patients (i.e., whether the patients in the study are like those seen in actual clinical practice)
- Index Test (i.e., whether the test under study could be used in actual clinical practice and whether it was administered in a way that reflects its use in actual practice)

- Directness (i.e., whether the study demonstrated that patient health is affected by use of the diagnostic test under study)
- Analysis (i.e., whether the data analysis reported in the study was based on a large enough percentage of enrolled patients to ensure that the analysis was not conducted on “unique” or “unusual” patients)

The specific questions we used, and the domains to which they pertain are shown in the table below:

Applicability Questions and Domains for Studies of Screening and Diagnostic Tests

Question	Domain
All Types of Patients Enrolled	Participants
Flexible Instructions about Index Test Methods to Practitioners	Index Test
Full Range of Practitioners & Settings	Index Test
Full Range of Index Text Readers	Index Test
Index Test Usable in Routine Practice	Index Test
Patient’s Outcomes Measured	Directness
All Patients in Analysis	Analysis

We characterized a domain as “flawed” if one or more questions addressing any given domain are answered “No” for a given screening/diagnostic/test, or if there are two or more “Unclear” answers to the questions addressing that domain. We characterized the applicability of a screening/diagnostic test as “High” if none of its domains are flawed, “Low” if all of its domains are flawed, and “Moderate” in all other cases.

Relationship between Domain Scores and Applicability for Studies of Screening and Diagnostic Tests

Number of Flawed Domains	Applicability
0	High
1,2, 3	Moderate
4	Low

STUDIES OF PROGNOSTICS

QUALITY

Our appraisal of studies of prognostics is a domain-based approach conducted using *a priori* questions, and scored by a computer program. The five domains we employed are:

- Prospective (A variable is specified as a potential prognostic variable *a priori*. This is not possible with retrospective studies.)
- Power (Whether the study had sufficient statistical power to detect a prognostic variable as statistically significant)
- Analysis (Whether the statistical analyses used to determine that a variable was rigorous to provide sound results)
- Model (Whether the final statistical model used to evaluate a prognostic variable accounted for enough variance to be statistically significant)
- Whether there was evidence of investigator bias

Quality Questions and Domains for Studies of Prognostics

Question	Domain
Prospective	Prospective
At Least 10 Patients per Important Variable	Power
At Least 10 Events*	Power
All Important Variables Screened for Entry Into Model	Analysis
Interactions Tested	Analysis
Collinearity Absent	Analysis
Primary Analysis (not subgroup or post hoc)	Analysis
Statistically Significant Fit	Model
Article and Abstract Agree	Investigator Bias
Results Reported for All Variables Studies	Investigator Bias
Blinded Data Analysts**	Investigator Bias

*Asked only if the variable predicted by the prognostic is dichotomous.

**Asked only if the prognostic variable is derived from a study that attempts to predict which patients respond best to a treatment.

We separately determined a quality score for each prognostic reported by a study. We characterized the evidence relevant to that prognostic variable as being of “High” quality if there are no flaws in any of the relevant domains, as being of “Moderate” quality if one of the relevant domains is flawed, as “Low” quality if there are two flawed domains, and as “Very Low” quality if three or more relevant domains are flawed. We characterized a domain as “flawed” if one or more questions addressing any given domain are answered “No” for a given prognostic variable, or if there are two or more “Unclear” answers to the questions addressing that domain.

Relationship between Quality and Domain Scores for Studies of Prognostics

Number of Flawed Domains	Quality
0	High
1	Moderate
2	Low
≥3	Very Low

APPLICABILITY

We separately evaluated the applicability of each prognostic variable reported in a study, and did so using a domain-based approach that involves predetermined questions and computer scoring. The domains we used for the applicability of prognostics are:

- Patients (i.e. whether the patients in the study and in the analysis were like those seen in actual clinical practice)
- Analysis (i.e., whether the analysis was not conducted in a way that was likely to describe variation among patients that might be unique to the dataset the authors used)
- Outcome (i.e., whether the prognostic was a predictor of a clinically meaningful outcome)

Applicability Questions and Domains for Studies of Prognostics

Question	Domain
Full Spectrum of Patients	Patients
All Patients in Analysis	Patients
No Stepwise Analysis	Analysis
Unambiguous Coding Scheme	Analysis
Model Validated	Analysis
Clinically Meaningful Outcome	Outcome

We characterized the evidence relevant to that prognostic as being of “High” applicability if there are no flaws in any of the relevant domains, as being of “Low” applicability if all three domains are flawed, and as of “Moderate” applicability in all other cases. We characterized a domain as “flawed” if one or more questions addressing any given domain are answered “No” for a given prognostic variable, or if there are two or more “Unclear” answers to the questions addressing that domain.

Relationship between Domain Scores and Applicability for Studies of Prognostics

Number of Flawed Domains	Applicability
0	High
1,2	Moderate
3	Low

STUDIES OF INCIDENCE AND PREVALENCE

QUALITY

Our appraisal of studies of incidence and prevalence is a domain-based approach conducted using *a priori* questions and scored by a computer program. The four domains we employed are:

- Outcome (whether the study is measuring the incidence/prevalence of a clinically meaningful event)
- Measurement (whether the study measured the disease/disorder/condition in a way that would lead to accurate estimates of incidence or prevalence)
- Participants (whether those who were studied were representative of the population of interest)
- Investigator Bias (whether author biases could have prejudiced the results)

Quality Questions and Domains for Studies of Incidence and Prevalence

Question	Domain	Incidence	Prevalence
Outcome Could Occur >1 Time in a Participant	None*	Yes	Yes
Study of Proportions or Number of Episodes	None	Yes	Yes
Only First Episode Counted	Measurement	Yes	Yes
Standard Methods for Collecting Outcomes Data	Outcome	Yes	Yes
Consistent Outcome Definitions	Outcome	Yes	Yes
Data Obtained from People or Records	None	Yes	Yes
Free from Response Bias	Measurement	Yes	Yes
Free from Information Bias	Measurement	Yes	Yes
Valid Instrument	Measurement	Yes	Yes
Valid Database Entries	Measurement	Yes	Yes
Study of In-Hospital Events	None	Yes	Yes
Use of Medical Records/Administrative Databases	Measurement	Yes	Yes
Appropriately Timed Outcome	Measurement	Yes	No
Chronic or Acute Disease	None	No	Yes
Study of Point Prevalence	None	No	Yes
Can Estimate Be Affected by Disease Severity	None	Yes	Yes
Correction for Disease Severity	Measurement	Yes	Yes
Population or Sample Data	None	Yes	Yes
Random Selection of Participants	Participants	Yes	Yes
>80% of Patients in Analysis	Participants	Yes	Yes
Free of Financial Conflicts of Interest	Bias	Yes	Yes
A Priori Analysis	Bias	Yes	Yes
Consistent Abstract, Results, Discussion	Bias	Yes	Yes

*An entry of “None” means that the question is not used in determining quality but, rather, is used for other purposes. A “Yes” entry in the above table means that a question is asked, a “No” entry means that it is not asked.

We characterized a study that has no flaws in any of its domains as being of “High” quality, a study that has one flawed domain as being of “Moderate” quality, a study with two flawed domains as being of “Low” quality, and a study with three or more flawed domains as being of “Very Low” quality. We characterized a domain as “flawed” if one or more questions addressing any given domain are answered “No” for a given screening/diagnostic/test, or if there are two or more “Unclear” answers to the questions addressing that domain.

We considered some design flaws as so serious that their presence automatically guarantees that a study is characterized as being of “Very Low” quality regardless of its domain scores. These flaws are:

- The outcome of interest could have occurred more than once in a person during the course of the study, and more than the first episode of the outcome was counted in the incidence/prevalence estimate
- The study was a study of the proportion (or number) of people who have a disease, and the study was not a study of point prevalence.

Relationship between Quality and Domain Scores for Studies of Incidence and Prevalence

Number of Flawed Domains	Quality
0	High
1	Moderate
2	Low
≥3	Very Low

APPLICABILITY

We separately evaluated the applicability of prevalence and incidence studies, and did so using a domain-based approach that involves predetermined questions and computer scoring. The domains we used for the applicability of prognostics are:

- Participants (i.e. whether the participants in the study were like those seen in the population of interest)
- Analysis (i.e., whether participants were appropriately included and excluded from the analysis)
- Outcome (i.e., whether the incidence/prevalence estimates being made were of a clinically meaningful outcome)

Applicability Questions and Domains for Studies of Incidence and Prevalence

Question	Domain
Full Spectrum of Patients	Patients
All Patients in Analysis	Patients
No Stepwise Analysis	Analysis
Unambiguous Coding Scheme	Analysis
Model Validated	Analysis
Clinically Meaningful Outcome	Outcome

We characterized a domain as “flawed” if one or more questions addressing any given domain are answered “No” for a given screening/diagnostic/test, or if there are two or more “Unclear” answers to the questions addressing that domain. We characterized the applicability of a screening/diagnostic test as “High” if none of its domains are flawed, “Low” if all of its domains are flawed, and “Moderate” in all other cases.

Relationship between Applicability and Domain Scores for Studies of Incidence and Prevalence

Number of Flawed Domains	Applicability
0	High
1,2	Moderate
3	Low

STUDIES OF JOINT REGISTRY EVALUATION

We judged the quality of joint registry reports using questions specified before this topic was selected and a computer program determined the final quality rating. We asked the following questions:

- Is submission of data to the registry mandatory?
- Is the registry population-based?
- Have the registry data been validated against hospital patient records or state/territory health department records?
- Does the registry ensure >90% of the patients are captured in the database?
- Does the registry perform statistical quality control measures to ensure the quality of the data?
- Does the registry perform checks to ensure that the diagnosis of patients who are in the registry is correct OR is the diagnosis unambiguous?

We characterized the evidence from a registry report as being of “High” quality if only one (or none) of the questions is answered “No”. If two or three questions are answered “No”, the evidence from the registry report is characterized as “Moderate” quality. If four or five questions are answered “No”, the evidence from the registry report is characterized as “Low” quality. If all six questions are answered “No” the evidence from the registry is characterized as “Very Low” quality.

Relationship between Joint Registry Evaluation Questions and Quality for Joint Registry Reports

Number of Questions answered “No”	Quality
0, 1	High
2,3	Moderate
4,5	Low
6	Very Low

**APPENDIX V
STUDY QUALITY AND APPLICABILITY**

Table 34. Symbols used in study quality summary tables

Symbol	Description
●	Yes
○	No

JOINT REGISTRIES

Table 36. Study design quality - Joint registries

Registry	Is submission of data to the registry mandatory?	Is the registry population-based?	Have the registry data been validated against hospital patient records or state/territory health department records?	Does the registry ensure >90% of the patients are captured in the database?	Does the registry perform statistical quality control measures to ensure the quality of the data?	Does the registry perform checks to ensure that the diagnosis of patients who are in the registry is correct OR is the diagnosis unambiguous?	Quality
Australia (2010)	●	●	●	●	●	●	High
England & Wales (2010)	●	●	●	○	●	●	High
Italian (Regional) (2010)	○	●	●	○	●	●	Moderate

PRIMARY LITERATURE STUDIES

Table 37. Quality, Applicability and Strength of Treatment Studies for Validated Patient-Oriented Outcome Measures

●: Domain free of flaws

○: Domain flaws present

Study	Outcome	Prospective	Power	Group Assignment	Blinding	Group Comparability	Treatment Integrity	Measurement	Investigator Bias	Quality	Participants	Intervention and Expertise	Compliance and Adherence	Analysis	Applicability	Study Strength
MacDonald 2003	Harris Hip Score	●	●	●	●	●	●	●	○	High	●	○	●	●	Moderate	High
Zijlstra 2009	Harris Hip Score (5yr)	●	●	●	●	○	●	●	●	High	●	○	●	●	Moderate	High
Zijlstra 2009	Harris Hip Score (10yr)	●	●	●	●	○	●	●	●	High	●	○	●	●	Moderate	High
Zijlstra 2009	Oxford Hip Score (5yr)	●	●	●	●	○	●	●	●	High	●	○	●	●	Moderate	High
Zijlstra 2009	Oxford Hip Score (10yr)	●	●	●	●	○	●	●	●	High	●	○	●	●	Moderate	High
Vendittoli 2010	WOMAC	●	●	●	○	○	●	●	○	Moderate	○	○	●	●	Moderate	Moderate
MacDonald 2003	WOMAC	●	●	●	●	●	●	●	○	High	●	○	●	●	Moderate	High
MacDonald 2003	SF-12	●	●	●	●	●	●	●	○	High	●	○	●	●	Moderate	High

Table 38. Quality, Applicability and Strength of Diagnostic Studies for Imaging Periprosthetic Tissues and Implant Positioning

●: Domain free of flaws

○: Domain flaws present

Study	Test	Reporting (Penalty)	Index Test	Reference Test	Participants	Information	Study Design	Quality	Participants	Index Test	Directness of Results	Analysis	Applicability	Study Strength
Sabah 2011	MRI of Periprosthetic Tissues	○	●	○	●	●	●	Moderate	●	○	●	●	Moderate	Moderate
Hart 2009	MRI of Periprosthetic Tissues	○	●	○	●	●	●	Moderate	●	○	●	●	Moderate	Moderate
Hart 2009	CT of Implant Positioning	○	●	○	●	●	●	Moderate	●	○	●	●	Moderate	Moderate

Table 39. Quality, Applicability and Strength of Prognostic Studies for Metal Ion Release

- : Domain free of flaws
- : Domain flaws present

Study	Prognostic	Prospective	Power	Analysis	Model	Investigator Bias	Quality	Patients	Analysis	Outcomes	Applicability	Study Strength
Hart 2011	Inclination Angle on Cobalt Release	●	●	●	●	○	Moderate	●	○	●	Moderate	Moderate
Hart 2011	Inclination Angle on Chromium Release	●	●	●	●	○	Moderate	●	○	●	Moderate	Moderate
Hart 2011	Version Angle on Cobalt Release	●	●	●	●	○	Moderate	●	○	●	Moderate	Moderate
Hart 2011	Version Angle on Chromium Release	●	●	●	●	○	Moderate	●	○	●	Moderate	Moderate
Hart 2011	Gender on Cobalt Release	●	●	●	●	○	Moderate	●	○	●	Moderate	Moderate
Hart 2011	Gender on Chromium Release	●	●	●	●	○	Moderate	●	○	●	Moderate	Moderate

Table 40. Quality, Applicability and Strength of Treatment Studies for Cobalt Ion Concentrations

- : Domain free of flaws
- : Domain flaws present

Study	Outcome	Prospective	Power	Group Assignment	Blinding	Group Comparability	Treatment Integrity	Measurement	Investigator Bias	Quality	Participants	Intervention and Expertise	Compliance and Adherence	Analysis	Applicability	Study Strength
MacDonald 2003	Cobalt Concentration	●	●	●	●	●	●	○	○	Moderate	●	○	●	●	Moderate	Moderate
Brodner 2003	Cobalt Concentration (2yr)	●	●	●	○	●	●	○	○	Moderate	●	○	●	●	Moderate	Moderate
Brodner 2003	Cobalt Concentration (5yr)	●	●	●	○	●	●	○	○	Moderate	●	○	●	●	Moderate	Moderate
Pattyn 2011	Cobalt Concentration	●	●	●	○	●	○	●	●	Moderate	●	○	●	●	Moderate	Moderate
Zijlstra 2010	Cobalt Concentration (5yr)	●	●	●	●	○	●	○	●	Moderate	●	○	●	●	Moderate	Moderate
Zijlstra 2010	Cobalt Concentration (10yr)	●	●	●	●	○	●	○	●	Moderate	●	○	●	●	Moderate	Moderate
Hailer 2011	Cobalt Concentration	●	●	○	●	○	●	●	●	Moderate	●	○	●	●	Moderate	Moderate
Moroni 2008	Cobalt Concentration	●	●	○	○	●	●	○	●	Moderate	●	○	●	●	Moderate	Moderate
Kwon 2011	Cobalt Concentration (Blood)	●	●	○	●	○	●	●	●	Moderate	●	○	○	●	Moderate	Moderate
Kwon 2011	Cobalt Concentration (Hip Aspirate)	●	●	○	●	○	●	●	●	Moderate	●	○	○	●	Moderate	Moderate

Table 41. Quality, Applicability and Strength of Treatment Studies for Chromium Ion Concentrations

- : Domain free of flaws
- : Domain flaws present

Study	Outcome	Prospective	Power	Group Assignment	Blinding	Group Comparability	Treatment Integrity	Measurement	Investigator Bias	Quality	Participants	Intervention and Expertise	Compliance and Adherence	Analysis	Applicability	Study Strength
MacDonald 2003	Chromium Concentration	●	●	●	●	●	●	○	○	Moderate	●	○	●	●	Moderate	Moderate
Pattyn 2011	Chromium Concentration	●	●	●	○	●	○	●	●	Moderate	●	○	●	●	Moderate	Moderate
Hailer 2011	Chromium Concentration	●	●	○	●	○	●	●	●	Moderate	●	○	●	●	Moderate	Moderate
Moroni 2008	Chromium Concentration	●	●	○	○	●	●	○	●	Moderate	●	○	●	●	Moderate	Moderate
Kwon 2011	Chromium Concentration (Blood)	●	●	○	●	○	●	●	●	Moderate	●	○	○	●	Moderate	Moderate
Kwon 2011	Chromium Concentration (Hip Aspirate)	●	●	○	●	○	●	●	●	Moderate	●	○	○	●	Moderate	Moderate

Table 42. Quality, Applicability and Strength of Treatment Studies for Titanium Ion Concentrations

- : Domain free of flaws
- : Domain flaws present

Study	Outcome	Prospective	Power	Group Assignment	Blinding	Group Comparability	Treatment Integrity	Measurement	Investigator Bias	Quality	Participants	Intervention and Expertise	Compliance and Adherence	Analysis	Applicability	Study Strength
MacDonald 2003	Titanium Concentration	●	●	●	●	●	●	○	○	Moderate	●	○	●	●	Moderate	Moderate

Table 43. Quality, Applicability, and Strength of Treatment Studies for Molybdenum Ion Concentrations

- : Domain free of flaws
- : Domain flaws present

Study	Outcome	Prospective	Power	Group Assignment	Blinding	Group Comparability	Treatment Integrity	Measurement	Investigator Bias	Quality	Participants	Intervention and Expertise	Compliance and Adherence	Analysis	Applicability	Study Strength
Moroni 2008	Molybdenum Concentration	●	●	○	○	●	●	○	●	Moderate	●	○	●	●	Moderate	Moderate

Table 44. Quality, Applicability and Strength of Incidence/Prevalence Studies of Complications Due to MoM Implants

- : Domain free of flaws
- : Domain flaws present

Study	Outcome	Investigator Bias				Quality	Analysis			Applicability	Study Strength
		Outcome	Measurement	Participants	Investigator Bias		Analysis	Outcome	Participants		
Marker 2007	Incidence of Complications	●	●	●	●	High	●	●	●	High	High
Steffen 2009	Incidence of Femoral Neck Fracture	●	○	●	○	Low	●	●	●	High	Moderate
Lavigne 2011	Incidence of Groin Pain	●	●	●	●	High	●	●	●	High	High
Vendittoli 2010	Incidence of Complications	●	●	●	●	High	●	●	●	High	High
Lazennec 2009	Incidence of Acetabular Loosening	●	●	○	●	Moderate	●	●	○	Moderate	Moderate
Ng 2011	Incidence of Complications	●	○	●	○	Low	●	●	●	High	Moderate
Ebramzadeh 2011	Incidence of complications	●	●	●	○	Moderate	●	●	●	High	High

Table 45. Quality, Applicability, and Strength of Incidence/Prevalence Studies of Pseudotumors

●: Domain free of flaws

○: Domain flaws present

Study	Outcome	Outcome	Measurement	Participants	Investigator Bias	Quality	Analysis	Outcome	Participants	Applicability	Study Strength
Kwon 2011	Incidence of Pseudotumors	●	●	○	○	Moderate	●	○	●	Moderate	Moderate
Glyn-Jones 2009	Incidence of Pseudotumors	●	●	●	○	Moderate	●	●	●	High	High
Langton 2010	Incidence of Pseudotumors	●	●	●	○	Moderate	●	●	●	High	High
Ng 2011	Incidence of Pseudotumors	●	○	●	○	Low	●	●	●	High	Moderate

**APPENDIX VI
SUPPLEMENTAL TABLES AND FIGURES**

TABLES AND FIGURES FOR QUESTION #1

Table 46. Number of Patients at risk used to calculate hazard ratios for “Risk of revision after Metal-on-Metal and other bearing surfaces from the Australian joint registry”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
C/C	29945	24473	20086	16402	13129	9857	6765	3996	1602	286
C/P	34560	31103	26556	21902	17256	12950	8879	5112	2057	530
M/M	17808	15347	11922	8579	5727	3617	2317	1446	670	125
M/P	62550	50943	42915	36221	29931	23592	17416	11539	5913	1651
CM/P	5807	4764	3762	2780	1793	1051	309	0	0	0

Table 47. Number of Patients at risk used to calculate “Risk of Revision after Metal-on-Metal and other bearing surfaces stratified by component head size from the Australian Registry”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
C/C	5259	4948	4624	4206	3754	3152	2400	1584	685	145
C/P	24325	23230	21207	18695	15621	12108	8514	4985	2048	530
M/M	2746	2636	2523	2382	2185	1941	1663	1242	572	97
M/P	40236	36115	33169	29959	26157	21730	16705	11325	5878	1645
CM/P	1819	1659	1468	1195	867	554	177	0	0	0

Table 48. Number of Patients at risk used to calculate “Risk of Revision after Metal-on-Metal and other bearing surfaces stratified by component head size from the Australian Registry”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
C/C	24686	19525	15462	12196	9375	6705	4365	2412	917	141
C/P	10235	7873	5349	3207	1635	842	365	127	9	0
M/M	15062	12711	9399	6197	3542	1676	654	204	98	28
M/P	22314	14828	9746	6262	3774	1862	711	214	35	6
CM/P	3988	3105	2294	1585	926	497	132	0	0	0

TABLES AND FIGURES FOR QUESTION #2

Table 49. Number of patients used to calculate “Hazard ratios for metal-on-metal THA* comparing implant head sizes”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
≤28mm	2746	2636	2523	2382	2185	1941	1663	1242	572	97
30-32mm	1843	1576	1265	950	696	388	119	3	0	0
36-40mm	3262	2776	2222	1749	1341	896	422	172	97	28
>40mm	9957	8359	5912	3498	1505	392	113	29	1	0

Table 50. Number of patients used to calculate “Hazard ratios for metal-on-metal THA comparing ≤32mm and >32mm”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
≤32mm	4589	4212	3788	3332	2881	2329	1782	1245	572	97
>32mm	13219	11135	8134	5247	2846	1288	535	201	98	28

Table 51. Number of patients used to calculate “Hazard ratios for metal-on-metal THA comparing age groups ”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
<55	3425	2968	2325	1697	1141	758	530	369	196	50
55-64	5880	5048	3901	2797	1918	1228	819	509	240	49
65-74	5564	4838	3781	2784	1895	1209	733	443	196	24
≥75	2939	2493	1915	1301	773	422	235	125	38	2

Table 52. Number of patients used to calculate “Hazard ratios for metal-on-metal THA comparing gender”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
Male	9937	8501	6548	4672	3107	1958	1292	848	414	75
Female	7871	6846	5374	3907	2620	1659	1025	598	256	50

Table 53. Number of patients used to calculate “Hazard ratios for metal-on-metal THA comparing the interaction between age and implant head size”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
≤32mm and <65	2392	2206	1977	1727	1527	1287	1032	759	376	78
≤32mm and ≥65yrs	2197	2006	1811	1605	1354	1042	750	486	196	19
>32mm and <65	6913	5810	4249	2767	1532	699	317	119	60	21
>32mm and ≥65yrs	6306	5325	3885	2480	1314	589	218	82	38	7

Table 54. Number of Patients used to calculate “Hazard ratios for metal-on-metal THA comparing the interaction between gender and implant head size

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
Male ≤32mm	2363	2163	1954	1726	1513	1256	997	736	356	59
Male >32mm	7574	6338	4594	2946	1594	702	295	112	58	16
Female ≤32mm	2226	2049	1834	1606	1368	1073	785	509	216	38
Female >32mm	5645	4797	3540	2301	1252	586	240	89	40	12

Table 55. Number of patients used to calculate “Revision Rate for Metal-on-metal Hip Resurfacing”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
Total Resurfacing	12587	11186	9707	8115	6449	4735	3205	1851	661	81

Table 56. Number of Patients used to calculate “Risk of Revision Comparing Primary Diagnoses While Adjusting for Age and Gender”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
Osteoarthritis	12587	11186	9707	8115	6449	4735	3205	1851	661	81
Developmental Dysplasia of the Hip	359	335	311	269	211	169	116	67	18	4
Avascular Necrosis	246	231	202	177	162	130	94	55	17	2

Table 57. Number used to calculate “Risk of Revision Comparing Age Groups While Adjusting for Gender”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
<55	6377	5633	4846	4054	3207	2390	1649	956	353	52
55-64	5004	4483	3907	3238	2583	1855	1223	705	254	26
≥65	1206	1070	954	823	659	490	333	190	54	3

Table 58. Number used to calculate “Risk of Revision for Metal-on-Metal Hip Resurfacing Comparing Implant Head Size”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
≤44mm	1139	1035	917	761	639	505	367	209	92	12
45-49mm	3083	2733	2334	1914	1482	1069	699	409	165	13
50-54mm	7543	6736	5893	4998	3998	2942	1990	1130	372	50
≥55mm	822	682	563	442	330	219	149	103	32	6

Table 59. Number used to calculate “Risk of Revision for Metal-on-Metal Hip Resurfacing Comparing the Interaction between Gender and Implant Head Size”

Number at Risk	0 Yr	1 Yr	2 Yrs	3 Yrs	4 Yrs	5 Yrs	6 Yrs	7 Yrs	8 Yrs	9 Yrs
Male <50mm	1555	1333	1093	853	647	478	303	167	78	3
Male ≥50mm	7940	7015	6084	5111	4039	2944	1987	1141	376	54
Female <50mm	2667	2435	2158	1822	1474	1096	763	451	179	22
Female ≥50mm	425	403	372	329	289	217	152	92	28	2

Table 60. Data on the referent group used to calculate hazard ratios in the U.K./Wales Registry

Male Total replacement using cement	N	3 Yr. Revision Rt.	5 Yr. Revision Rt.	HR
<55	1669	2.7 (1.9, 3.8)	4.4 (3.2, 6.0)	1
55-64	5560	1.7 (1.3, 2.1)	2.6 (2.1, 3.2)	1
65+	27694	1.5 (1.3, 1.7)	2.0 (1.8, 2.3)	1
Female Total replacement using cement	N	3 Yr. Revision Rt.	5 Yr. Revision Rt.	HR
<55	2,174	2.5 (1.8 to 3.4)	3.6 (2.7 to 4.8)	1
55-64	8,239	1.6 (1.3 to 2.0)	2.5 (2.1 to 3.0)	1
65+	54,023	1.1 (1.0 to 1.2)	1.6 (1.5 to 1.8)	1

TABLES AND FIGURES FOR QUESTION #3
TABLES SUMMARIZING METAL ION DATA

Table 61. Primary Literature Data Summarizing Cobalt Ion Levels

Study (Cobalt)	Strength of Evidence	N	Bearing Surface	Preoperative Concentration (µg/L)	Analysis Time Span	Increase Factor	Final Concentration (µg/L)	Intergroup Significant
MacDonald2003	Moderate	22	Metal/Metal	0.14 (0.09-0.17)	2 Years	6.47	1.10 (0.66-2.43)	
	Moderate	18	Metal/Poly	0.11 (0.09-0.15)	2 Years		0.17 (0.12-0.23)	<0.05
Brodner2003	Moderate	50	Metal/Metal	0.15 (0.15-0.15)	2 years	5.00	0.75 (0.40-2.85)	
	Moderate	50	Ceramic/Poly	0.15 (0.15-0.15)	2 Years		0.15 (0.15-0.15)	<0.05
	Moderate	50	Metal/Metal	0.15 (0.15-0.15)	5 Years	4.66	0.70 (0.23-1.75)	
	Moderate	50	Ceramic/Poly	0.15 (0.15-0.15)	5 Years		0.15 (0.15-0.15)	<0.05
Pattyn2011	Moderate	20	M/M (BHR)	0.48 (0.48-0.48)	2 Years	3.33	1.60 (0.70-2.05)	<0.05
	Moderate	22	M/M (Durom)	0.48 (0.48-0.48)	2 Years	1.66	0.80 (0.60-0.90)	<0.05
	Moderate	10	M/M (Metasul)	0.48 (0.48-0.48)	2 Years	2.81	1.35 (0.90-2.00)	<0.05
	Moderate	18	C/C	0.48 (0.48-0.48)	2 Years		0.48 (0.48-0.48)	
Zijlstra2010	Moderate	17	Metal/Metal	0.18 (0.18-1.77)	5 years	2.93	0.88 (0.29-7.02)	
	Moderate	13	Metal/Poly	0.24 (0.18-0.65)	5 Years		0.30 (0.29-1.65)	<0.05
	Moderate	17	Metal/Metal	0.18 (0.18-1.77)	10 Years	2.2	1.10 (0.50-11.0)	
	Moderate	13	Metal/Poly	0.24 (0.18-0.65)	10 Years		0.50 (0.40-1.30)	<0.05
Hailer2011	Moderate	19	Metal/Metal	0.09 (0.06-0.12)	6 Years	3.58	0.86 (0.49-1.22)	
	Moderate	17	Metal/Poly	0.16 (0.1-0.23)	6 Years		0.24 (0.12-0.36)	<0.05

*In all metal-on-metal study groups, increases in metal ion level from baseline to follow-up were statistically significant

**MacDonald and Zijlstra found statistically significant increases in metal ions in the metal-on-polyethylene groups from baseline to follow-up

Table 62. Primary Literature Data Summarizing Chromium and Titanium Ion Levels

Study	Strength of Evidence	N	Bearing Surface	Preoperative Concentration (µg/L)	Analysis Time Span	Final Concentration	Increase Factor	Intergroup P-value
MacDonald2003 (Chromium)	Moderate	22	Metal/Metal	1.09 (0.67-1.96)	2 Years	2.50 (1.18-3.13)	1.92	
	Moderate	18	Metal/Poly	0.58 (0.38-1.02)	2 Years	1.30 (1.00-1.83)		<0.05
Pattyn2011 (Chromium)	Moderate	20	M/M (BHR)	0.48 (0.48-0.48)	2 Years	1.50 (0.68-2.20)	3.13	<0.05
	Moderate	22	M/M (Durom)	0.48 (0.48-0.48)	2 Years	0.90 (0.50-1.40)	1.88	<0.05
	Moderate	10	M/M (Metasul)	0.48 (0.48-0.48)	2 Years	1.20 (0.83-1.70)	2.5	<0.05
	Moderate	18	C/C	0.48 (0.48-0.48)	2 Years	0.48 (0.48-0.48)		
Hailer2011 (Chromium)	Moderate	19	Metal/Metal	0.32 (0.22-0.42)	6 Years	1.05 (0.53-1.6)	2.92	
	Moderate	17	Metal/Poly	0.31 (0.20-0.42)	6 Years	0.36 (0.15-0.58)		<0.05
MacDonald2003 (Titanium)	High	22	Metal/Metal	1.04 (0.72-1.24)	2 Years	1.80 (1.15-2.75)	1.2	
	High	18	Metal/Poly	1.23 (0.94-1.78)	2 Years	1.50 (1.15-2.08)		>0.05

*In all metal-on-metal study groups, increases in metal ion level from baseline to follow-up were statistically significant

Table 63. Kwon et. al. Data Summarizing Patients w/ and w/o Pseudotumors

Blood Concentrations	Strength of Evidence	N	Bearing Surface	Analysis Time Span	Final Concentration	Intergroup P-value
Kwon (Cobalt)	Moderate	7	M/M w/ Pseudotumor	5 Years	9.2 (4.8-22.5)	<0.05
	Moderate	151	M/M w/o Pseudotumor	5 Years	1.9 (0.5-10.8)	<0.05
Kwon (Chromium)	Moderate	20	M/P	5 Years	0.6 (0.5-3.1)	
	Moderate	7	M/M w/ Pseudotumor	5 Years	12.0 (3.8- 22.8)	<0.05
Hip Aspirate	Moderate	151	M/M w/o Pseudotumor	5 Years	2.1 (0.5-14.3)	<0.05
	Moderate	20	M/P	5 Years	0.5 (0.5-1.1)	
Kwon(Cobalt)	Moderate	7	M/M w/ Pseudotumor	5 Years	1182(206-1802)	
	Moderate	151	M/M w/o Pseudotumor	5 Years	86.2 (1.0-158)	0.003
Kwon(Chromium)	Moderate	7	M/M w/ Pseudotumor	5 Years	883 (221-1322)	
	Moderate	151	M/M w/o Pseudotumor	5 Years	114.8 (3-230)	0.001

Table 64. Summary of Evidence Comparing Metal-on-Metal THA to Hip Resurfacing

Study	Strength of Evidence	N	Bearing Surface	Analysis Time Span	Final Concentration	P-value
Moroni2008 (Cobalt)	Moderate	20	MoM Resurfacing	24 Months	1.40 (0.08-8.96)	<0.05*
	Moderate	26	MoM THA	26.3 Months	1.33 (0.34-5.32)	>0.05**
Moroni2008 (Chromium)	Moderate	48	Control		0.29 (0.08-0.86)	
	Moderate	20	MoM Resurfacing	24 Months	2.30 (0.69-7.24)	<0.05*
	Moderate	26	MoM THA	26.3 Months	1.73 (0.22-6.65)	>0.05**
Moroni2008 (Molybdenum)	Moderate	48	Control		0.25 (0.06-0.67)	
	Moderate	20	MoM Resurfacing	24 Months	0.90 (0.83-1.30)	<0.05*
	Moderate	26	MoM THA	26.3 Months	0.96 (0.83-1.73)	>0.05**
	Moderate	48	Control		BDL†	

*P-values are result of comparing hip resurfacing to healthy control w/o implants

**P-values are result of comparing hip resurfacing to metal-on-metal THA

†BDL stands for below detection limit

Table 65. Prognostic Study of Metal Ion Release

Explanatory Variable	Strength of Evidence	N	Ion of Interest	Median Angle (Range)	Estimated Effect	P-value	95% CI
Inclination(Hart2011)	Moderate	100	Cobalt	49 (13-78)	1.045	<0.05	1.021-1.069
	Moderate		Chromium		1.033	<0.05	1.021-1.069
Version(Hart2011)	Moderate	100	Cobalt	27 (-47-69)	0.975	<0.05	0.962-0.989
	Moderate		Chromium		0.98	<0.05	0.967-0.994
Gender(Hart2011)	Moderate	100	Cobalt		1.68	<0.05	1.23-2.78
	Moderate		Chromium		1.873	<0.05	1.149-3.056

*Final model for Co= EXP (-1.067 + 0.517*Gender + 0.044*Inclination – 0.025*Version)

*Final model for Cr= EXP (-0.643 + 0.628*Gender + 0.032*Inclination – 0.020*Version)

*Male = 0; Female = 1

IMAGING RESULTS

Table 66. Magnetic Resonance Imaging Results

MRI Findings	Strength of Evidence	N	Types of Abnormalities
Sabah2011	Moderate	31	23 (74%) periprosthetic tissue lesions
	Moderate	31	21 (68%) Fluid-like lesions
	Moderate	31	2 (6%) Solid lesions
	Moderate	29	26 (90%) had muscle atrophy present
Hart2009	Moderate	26	16 (58%) periprosthetic tissue lesions
	Moderate	26	14 (54%) Fluid-like lesions
	Moderate	26	2 (8%) Solid masses

Table 67. Computed Tomography Results

CT Findings	Column1	N	Median Inclination angle	Median version angle
Hart2009	Moderate	16	55 (39-78)	31 (-47-48)

TABLES FOR COMPLICATIONS

Table 68. Summary of complications from Marker et. al.

	Marker 2007
Study Strength	High
Bearing Surface	MoM HR*
femoral loosening	3/550 (0.55%)
acetabular loosening	9/550 (1.64%)
femoral neck fracture	13/550 (2.36%)
osteolysis	1/550 (0.18%)
infection in joint	4/550 (0.73%)
implant breakage	1/550 (0.18%)
acetabular protrusion	2/550 (0.36%)
abductor rupture	1/550 (0.18%)

*HR stands for hip resurfacing

Table 69. Summary of Complications from Ebramzadeh et. al.

Column1	Ebramzadeh 2011	Ebramzadeh 2011
Study Strength	High	High
Bearing Surface	MoM HR*	MoM THA*
femoral loosening	66/279 (23.66%)	4/99 (4.04%)
acetabular loosening	55/279 (19.71%)	47/99 (47.47%)
femoral neck fracture	62/279 (22.22%)	1/99 (1.01%)
impingement	8/279 (2.87%)	8/99 (8.08%)
sepsis	18/279 (6.45%)	5/99 (5.05%)
metal allergy	13/279 (4.66%)	7/99 (7.07%)
osteolysis	2/279 (0.72%)	3/99 (3.03%)

*HR stands for hip resurfacing; THA stands for total hip arthroplasty

Table 70. Summary of Complications from Steffen et. al.

Column1	Steffen 2009
Study Strength	Moderate
Bearing Surface	MoM HR
femoral neck fracture	15/828 (1.81%)

Table 71. Summary of Complications from Ng et. al.

Column1	Ng 2011	Ng 2011	Ng 2011	Ng 2011
Study strength	Moderate	Moderate	Moderate	Moderate
Bearing Surface	MoM THA	MoM HR	MoP THA*	CoP THA*
periprosthetic fracture	0/26 (0.00%)	1/6 (16.67%)	9/166 (5.42%)	1/8 (12.50%)
dislocation	2/26 (7.69%)	0/6 (0.00%)	17/166 (10.24%)	1/8 (12.50%)
aseptic loosening	11/26 (42.31%)	2/6 (33.33%)	115/166 (69.28%)	5/8 (62.50%)
infection in joint	4/26 (15.38%)	2/6 (33.33%)	21/166(12.65%)	1/8 (12.50%)
implant breakage	0/26 (0.00%)	0/6 (0.00%)	1/166 (0.60%)	0/8 (0.00%)

*MoP stands for metal-on-polyethylene; CoP stands for ceramic-on-polyethylene

Table 72. Summary of Complications from Lavigne et. al.

Column1	Lavigne 2011	Lavigne 2011	Lavigne 2011
Study Strength	High	High	High
Comparison	MoM HR	MoM Large Head	MoM 28mm
groin pain	15/101 (14.85%)	15/89 (16.9%)	11/85 (12.9%)

Table 73. Summary of Complication from Lazennec et. al.

Column1	Lazennec 2009
Study Strength	Moderate
Bearing Surface	MoM THA
acetabular loosening	8/109 (7.34%)
2/109 (1.83%)	2/109 (1.83%)

Table 74. Summary of Complications from Vendittoli et. al.

Column1	Vendittoli 2010	Vendittoli 2010
Study Strength	High	High
Bearing Surface	MoM THA	MoM HR
groin pain	0/100 (0.00%)	4/109 (3.67%)
femoral loosening	0/100 (0.00%)	6/109 (5.50%)
femoral neck fracture	4/100 (4.00%)	0/109 (0.00%)
acetabulum fracture	0/100 (0.00%)	2/109 (1.83%)
impingement	0/100 (0.00%)	2/109 (1.83%)
ossification	0/100 (0.00%)	2/109 (1.83%)
squeaking	2/100 (2.00%)	0/109 (0.00%)
dislocation	4/100 (4.00%)	0/109 (0.00%)
infection in joint	5/100 (5.00%)	0/109 (0.00%)
DVT	3/100 (3.00%)	1/109 (0.92%)
neurapraxia (sciatic)	2/100 (2.00%)	1/109 (0.92%)
symptomatic leg length discrepancy	1/100 (1.00%)	0/109 (0.00%)

APPENDIX VII
AAOS BODIES THAT APPROVED THIS TECHNOLOGY OVERVIEW

TASK FORCE

Kevin J. Bozic, MD, MBA

University of California, San Francisco
School of Medicine
Department of Orthopaedic Surgery
Philip R. Lee Institute for Health Policy Studies
500 Parnassus Ave, MU-320W, Box 0728
San Francisco, CA 94143

James Browne, MD

University of Virginia
Department of Orthopaedic Surgery
PO Box 800159
Charlottesville, VA 22908

Chris J. Dangles, MD

University of Illinois Urbana-Champaign
1107 W. University Ave
Champaign Illinois 61821

Paul A. Manner, MD, FRCSC

Department of Orthopaedics and Sports Medicine
University of Washington School of Medicine
1959 Pacific Street NE
Box 356500
Seattle, WA 98195-6500

Adolph J. Yates, Jr., MD

Department of Orthopaedic Surgery
University of Pittsburgh Medical Center
Suite #415
5200 Centre Avenue
Pittsburgh, PA 15232

Technology Overview Oversight Chair

Kristy L. Weber, MD

Chair, AAOS Council on Research and Quality

Johns Hopkins Outpatient Center
Department of Orthopaedic Surgery
601 N Caroline St Rm 5251
Baltimore, MD 21287

CONFLICT OF INTEREST

All members of the AAOS task force listed above disclosed any conflicts of interest prior to the development of the key questions for this Technology Overview. Conflicts of interest are disclosed with the American Academy of Orthopaedic Surgeons via a private on-line reporting database and also verbally at the recommendation approval meeting.

Kevin John Bozic, MD, MBA: 9 (AAOS; Agency for Healthcare Research and Quality (AHRQ); American Association of Hip and Knee Surgeons; American Joint Replacement Registry; American Orthopaedic Association; California Joint Replacement Registry Project; California Orthopaedic Association; Orthopaedic Research and Education Foundation); Submitted on: 10/20/2011.*

James Andrew Browne, MD: (n) Submitted on: 06/29/2011.*

Chris John Dangles, MD: (n) Submitted on: 07/01/2011.*

Paul A Manner, MD: 8 (Orthopedics); Submitted on: 06/29/2011.*

Adolph J Yates Jr, MD: 9 (American Association of Hip and Knee Surgeons); Submitted on: 07/05/2011.*

Kristy L Weber, MD: 8 (AAOS Now); 9 (Musculoskeletal Tumor Society; Ruth Jackson Orthopaedic Society); Submitted on: 10/20/2011.*

***Disclosure Items:** (n) = Respondent answered 'No' to all items indicating no conflicts.
1= Royalties from a company or supplier; 2= Speakers bureau/paid presentations for a company or supplier; 3A= Paid employee for a company or supplier; 3B= Paid consultant for a company or supplier; 3C= Unpaid consultant for a company or supplier; 4= Stock or stock options in a company or supplier; 5= Research support from a company or supplier as a PI; 6= Other financial or material support from a company or supplier; 7= Royalties, financial or material support from publishers; 8= Medical/Orthopaedic publications editorial/governing board; 9= Board member/committee appointments for a society.

GUIDELINES OVERSIGHT COMMITTEE

The AAOS Guidelines Oversight Committee (GOC) consists of ten AAOS members. The overall purpose of this Committee is to oversee the development of the clinical practice guidelines, performance measures, health technology assessments and utilization guidelines.

EVIDENCE BASED PRACTICE COMMITTEE

The AAOS Evidence Based Practice Committee (EBPC) consists of nine AAOS members. This Committee provides review, planning, and oversight for all activities related to quality improvement in orthopaedic practice, including, but not limited to evidence-based guidelines, performance measures, and outcomes..

COUNCIL ON RESEARCH AND QUALITY

To enhance the mission of the AAOS, the Council on Research and Quality promotes the most ethically and scientifically sound basic, clinical, and translational research possible to ensure the future care for patients with musculoskeletal disorders. The Council also serves as the primary resource to educate its members, the public, and public policy makers regarding evidenced-based medical practice, orthopaedic devices and biologics regulatory pathways and standards development, patient safety, occupational health, technology assessment, and other related areas of importance.

BOARD OF DIRECTORS

The 17 member AAOS Board of Directors manages the affairs of the AAOS, sets policy, and determines and continually reassesses the Strategic Plan.

DOCUMENTATION OF APPROVAL

AAOS Task Force Draft Completed:	October 26, 2011
Manufacturer Review Completed:	November 15, 2011
AAOS Guidelines Oversight Committee:	November 21, 2011
AAOS Evidence Based Practice Committee:	November 21, 2011
AAOS Council on Research and Quality:	November 21, 2011
AAOS Board of Directors:	December 2, 2011

AAOS STAFF

Kevin M. Boyer – Manager, Clinical Practice Guidelines and Appropriate Use Criteria

Paul Zemaitis, MPH – Research Analyst, Clinical Practice Guidelines

Anne Woznica – AAOS Medical Librarian

Charles M. Turkelson, PhD – former Director of Research and Scientific Affairs

Janet L. Wies, MPH – former Manager, Clinical Practice Guidelines

REFERENCES

1. Australian Orthopaedic Association National Joint Replacement Registry. Annual Report 2010. Adelaide: Australian Orthopaedic Association; 2010.
2. National Joint Registry for England and Wales. 7th Annual Report. Hemel Hempstead, UK: National Joint Registry Center; 2010.
3. Altman DG. Statistics and ethics in medical research. VI--Presentation of results. *Br Med J* 1980;281(6254):1542-1544.
4. Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol* 2001;54(10):979-985.
5. Brodner W, Bitzan P, Meisinger V, Kaider A, Gottsauner WF, Kotz R. Serum cobalt levels after metal-on-metal total hip arthroplasty. *The Journal of bone and joint surgery American volume* 2003;85-A:2168-2173.
6. Ebramzadeh E, Campbell PA, Takamura KM et al. Failure modes of 433 metal-on-metal hip implants: how, why, and wear. *Orthop Clin North Am* 2011;42(2):241-50, ix.
7. Glyn-Jones S, Pandit H, Kwon YM, Doll H, Gill HS, Murray DW. Risk factors for inflammatory pseudotumour formation following hip resurfacing. *J Bone Joint Surg Br* 2009;91(12):1566-1574.
8. Hailer NP, Blaheta RA, Dahlstrand H, Stark A. Elevation of circulating HLA DR(+) CD8(+) T-cells and correlation with chromium and cobalt concentrations 6 years after metal-on-metal hip arthroplasty. *Acta Orthop* 2011;82(1):6-12.
9. Hart AJ, Sabah S, Henckel J et al. The painful metal-on-metal hip resurfacing. *J Bone Joint Surg Br* 2009;91(6):738-744.\
10. Hart AJ, Skinner JA, Henckel J, Sampson B, Gordon F. Insufficient Acetabular Version Increases Blood Metal Ion Levels after Metal-on-metal Hip Resurfacing. *Clin Orthop Relat Res* 2011.
11. Kwon YM, Ostlere SJ, McLardy-Smith P, Athanasou NA, Gill HS, Murray DW. "Asymptomatic" Pseudotumors After Metal-on-Metal Hip Resurfacing Arthroplasty Prevalence and Metal Ion Study. *J Arthroplasty* 2011;26(4):511-518.
12. Langton DJ, Jameson SS, Joyce TJ, Hallab NJ, Natsu S, Nargol AV. Early failure of metal-on-metal bearings in hip resurfacing and large-diameter total hip replacement: A consequence of excess wear. *J Bone Joint Surg Br* 2010;92(1):38-46.
13. Lavigne M, Laffosse JM, Ganapathi M, Girard J, Vendittoli P. Residual groin pain at a minimum of two years after metal-on-metal THA with a twenty-eight-millimeter femoral

- head, THA with a large-diameter femoral head, and hip resurfacing. *J Bone Joint Surg Am* 2011;93 Suppl 2:93-98.
14. Lazennec JY, Boyer P, Poupon J et al. Outcome and serum ion determination up to 11 years after implantation of a cemented metal-on-metal hip prosthesis. *Acta Orthop* 2009;80(2):168-173.
 15. MacDonald SJ, McCalden RW, Chess DG et al. Metal-on-metal versus polyethylene in hip arthroplasty: a randomized clinical trial. *Clin Orthop* 2003;282-296.
 16. Marker DR, Seyler TM, Jinnah RH, Delanois RE, Ulrich SD, Mont MA. Femoral neck fractures after metal-on-metal total hip resurfacing: a prospective cohort study. *J Arthroplasty* 2007;22(7 Suppl 3):66-71.
 17. Moroni A, Savarino L, Cadossi M, Baldini N, Giannini S. Does ion release differ between hip resurfacing and metal-on-metal THA? *Clin Orthop Relat Res* 2008;466(3):700-707.
 18. Ng VY, Lombardi AV, Jr., Berend KR, Skeels MD, Adams JB. Perivascular lymphocytic infiltration is not limited to metal-on-metal bearings. *Clin Orthop Relat Res* 2011;469(2):523-529.
 19. Pattyn CA, Lauwagie SN, Verdonk RC. Whole blood metal ion concentrations in correlation with activity level in three different metal-on-metal bearings. *J Arthroplasty* 2011;26(1):58-64.
 20. Sabah SA, Mitchell AW, Henckel J, Sandison A, Skinner JA, Hart AJ. Magnetic resonance imaging findings in painful metal-on-metal hips: a prospective study. *J Arthroplasty* 2011;26(1):71-6, 76.
 21. Steffen R-T, Foguet PR, Krikler SJ, Gundle R, Beard DJ, Murray DW. Femoral Neck Fractures After Hip Resurfacing. *J Arthroplasty* 2009;24(4):614-619.
 22. Vendittoli PA, Ganapathi M, Roy AG, Lusignan D, Lavigne M. A comparison of clinical results of hip resurfacing arthroplasty and 28 mm metal on metal total hip arthroplasty: a randomised trial with 3-6 years follow-up. *Hip Int* 2010;20(1):1-13.
 23. Zijlstra WP, van Raay JJ, Bulstra SK, Deutman R. No Superiority of Cemented Metal-on-Metal Over Metal-on-Polyethylene THA in a Randomized Controlled Trial at 10-Year Follow-up. *Orthopedics* 2010;154-161.
 24. Witzleb WC, Ziegler J, Krummenauer F, Neumeister V, Guenther KP. Exposure to chromium, cobalt and molybdenum from metal-on-metal total hip replacement and hip resurfacing arthroplasty. *Acta Orthop* 2006;77(5):697-705.

EXCLUDED ARTICLES BIBLIOGRAPHY

Amstutz HC, Beaulé PE, Dorey FJ, Le Duff MJ, Campbell PA, Gruen TA. Metal-on-metal hybrid surface arthroplasty. Surgical Technique. *J Bone Joint Surg Am* 2006;88 Suppl 1 Pt 2:234-249.

Antoniou J, Zukor DJ, Mwale F, Minarik W, Petit A, Huk OL. Metal ion levels in the blood of patients after hip resurfacing: a comparison between twenty-eight and thirty-six-millimeter-head metal-on-metal prostheses. *J Bone Joint Surg Am* 2008;90 Suppl 3:142-148.

Baker RP, Pollard TC, Eastaugh-Waring SJ, Bannister GC. A medium-term comparison of hybrid hip replacement and Birmingham hip resurfacing in active young patients. *J Bone Joint Surg Br* 2011;93(2):158-163.

Beaulé PE, Kim PR, Hamdi A, Fazekas A. A prospective metal ion study of large-head metal-on-metal bearing: a matched-pair analysis of hip resurfacing versus total hip replacement. *Orthop Clin North Am* 2011;42(2):251-7, ix.

Benoit B, Gofton W, Beaulé PE. Hueter anterior approach for hip resurfacing: assessment of the learning curve. *Orthop Clin North Am* 2009;40(3):357-363.

Bozic KJ, Ong K, Lau E et al. Risk of complication and revision total hip arthroplasty among Medicare patients with different bearing surfaces. *Clin Orthop Relat Res* 2010;468(9):2357-2362.

Brodner W, Bitzan P, Meisinger V, Kaider A, Gottsauner WF, Kotz R. Elevated serum cobalt with metal-on-metal articulating surfaces. *The Journal of bone and joint surgery British volume* 1997;79:316-321.

Brodner W, Grubl A, Jankovsky R, Meisinger V, Lehr S, Gottsauner-Wolf F. Cup inclination and serum concentration of cobalt and chromium after metal-on-metal total hip arthroplasty. *J Arthroplasty* 2004;19(8 Suppl 3):66-70.

Clarke MT, Lee PT, Arora A, Villar RN. Levels of metal ions after small- and large-diameter metal-on-metal hip arthroplasty. *J Bone Joint Surg Br* 2003;85(6):913-917.

Daniel J, Pynsent PB, McMinn DJ. Metal-on-metal versus polyethylene in hip arthroplasty: a randomized clinical trial. *Clin Orthop* 2004;271-272.

Daniel J, Ziaee H, Salama A, Pradhan C, McMinn DJ. The effect of the diameter of metal-on-metal bearings on systemic exposure to cobalt and chromium. *J Bone Joint Surg Br* 2006;88(4):443-448.

- Daniel J, Ziaee H, Pynsent PB, McMinn DJ. The validity of serum levels as a surrogate measure of systemic exposure to metal ions in hip replacement. *J Bone Joint Surg Br* 2007;89(6):736-741.
- Daniel J, Ziaee H, Pradhan C, McMinn DJ. Six-year results of a prospective study of metal ion levels in young patients with metal-on-metal hip resurfacings. *J Bone Joint Surg Br* 2009;91(2):176-179.
- Dastane MR, Long WT, Wan Z, Chao L, Dorr LD. Metal-on-metal hip arthroplasty does equally well in osteonecrosis and osteoarthritis. *Clin Orthop Relat Res* 2008;466(5):1148-1153.
- Davies AP, Willert HG, Campbell PA, Learmonth ID, Case CP. An unusual lymphocytic perivascular infiltration in tissues around contemporary metal-on-metal joint replacements. *J Bone Joint Surg Am* 2005;87(1):18-27.
- Dunstan E, Sanghrajka AP, Tilley S et al. Metal ion levels after metal-on-metal proximal femoral replacements: a 30-year follow-up. *J Bone Joint Surg Br* 2005;87(5):628-631.
- Fujishiro T, Moojen DJ, Kobayashi N, Dhert WJ, Bauer TW. Perivascular and diffuse lymphocytic inflammation are not specific for failed metal-on-metal hip implants. *Clin Orthop Relat Res* 2011;469(4):1127-1133.
- Grammatopoulous G, Pandit H, Kwon YM et al. Hip resurfacings revised for inflammatory pseudotumour have a poor outcome. *J Bone Joint Surg Br* 2009;91(8):1019-1024.
- Grammatopoulos G, Pandit H, Kwon Y-M et al. Hip resurfacings revised for inflammatory pseudotumour have a poor outcome. *Journal of Bone and Joint Surgery - Series B* 2009;91(8):1019-1024.
- Grammatopoulos G, Pandit H, Glyn-Jones S et al. Optimal acetabular orientation for hip resurfacing. *J Bone Joint Surg Br* 2010;92(8):1072-1078.
- Grubl A, Marker M, Brodner W et al. Long-term follow-up of metal-on-metal total hip replacement. *J Orthop Res* 2007;25(7):841-848.
- Hailer N, Dahlstrand H, Stark A. Significant increases in serum concentrations of chromium and cobalt but not of nickel and manganese two years after metal-on-metal alloarthroplasty of the hip: a prospective, randomised study comparing metal-on-metal with metal-on-polyethylene bearings. *The Journal of Bone and Joint Surgery* 2009;91-B:132-13a.
- Hall DP, Srikantharajah D, Anakwe RE, Gaston P, Howie CR. Patient-reported outcome following metal-on-metal resurfacing of the hip and total hip replacement. *Hip Int* 2009;19(3):245-250.

Hallows RK, Pelt CE, Erickson JA, Peters CL. Serum Metal Ion Concentration: Comparison Between Small and Large Head Metal-on-Metal Total Hip Arthroplasty. *J Arthroplasty* 2011.

Hart AJ, Hester T, Sinclair K et al. The association between metal ions from hip resurfacing and reduced T-cell counts. *J Bone Joint Surg Br* 2006;88(4):449-454.

Hart AJ, Skinner JA, Winship P et al. Circulating levels of cobalt and chromium from metal-on-metal hip replacement are associated with CD8+ T-cell lymphopenia. *J Bone Joint Surg Br* 2009;91(6):835-842.

Jacobs JJ, Skipor AK, Doorn PF et al. Cobalt and chromium concentrations in patients with metal on metal total hip replacements. *Clin Orthop Relat Res* 1996;(329 Suppl):S256-S263.

Jacobs JJ, Skipor AK, Patterson LM et al. Metal release in patients who have had a primary total hip arthroplasty. A prospective, controlled, longitudinal study. *J Bone Joint Surg Am* 1998;80(10):1447-1458.

Jacobs M, Gorab R, Mattingly D, Trick L, Southworth C. Three- to six-year results with the Ultima metal-on-metal hip articulation for primary total hip arthroplasty. *J Arthroplasty* 2004;19(7 Suppl 2):48-53.

Kim PR, Beaulé PE, Laflamme GY, Dunbar M. Causes of early failure in a multicenter clinical trial of hip resurfacing. *J Arthroplasty* 2008;23(6 Suppl 1):44-49.

Kim PR, Beaulé PE, Dunbar M et al. Cobalt and chromium levels in blood and urine following hip resurfacing arthroplasty with the conserve plus implant. *J Bone Joint Surg Am* 2011;93 Suppl 2:107-117.

Korovessis P, Petsinis G, Repanti M. Zweymueller with metal-on-metal articulation: Clinical, radiological and histological analysis of short-term results. *Arch Orthop Trauma Surg* 2003;123(1):5-11.

Kwon YM, Thomas P, Summer B et al. Lymphocyte proliferation responses in patients with pseudotumors following metal-on-metal hip resurfacing arthroplasty. *J Orthop Res* 2010;28(4):444-450.

Ladon D, Doherty A, Newson R, Turner J, Bhamra M, Case CP. Changes in metal levels and chromosome aberrations in the peripheral blood of patients after metal-on-metal hip arthroplasty. *J Arthroplasty* 2004;19(8 Suppl 3):78-83.

Latteier MJ, Berend KR, Lombardi AV, Jr., Ajluni AF, Seng BE, Adams JB. Gender is a Significant Factor for Failure of Metal-on-Metal Total Hip Arthroplasty. *J Arthroplasty* 2011.

- Lavigne M, Masse V, Girard J, Roy AG, Vendittoli PA. [Return to sport after hip resurfacing or total hip arthroplasty: a randomized study]. *Revue de chirurgie orthopédique et réparatrice de l'appareil moteur* 2008;94:361-367.
- Lavigne M, Belzile EL, Roy A, Morin F, Amzica T, Vendittoli PA. Comparison of whole-blood metal ion levels in four types of metal-on-metal large-diameter femoral head total hip arthroplasty: the potential influence of the adapter sleeve. *J Bone Joint Surg Am* 2011;93 Suppl 2:128-136.
- Le Duff MJ, Amstutz HC, Dorey FJ. Metal-on-metal hip resurfacing for obese patients. *J Bone Joint Surg Am* 2007;89(12):2705-2711.
- Le Duff MJ, Wisk LE, Amstutz HC. Range of motion after stemmed total hip arthroplasty and hip resurfacing: A clinical study. *Bulletin of the NYU Hospital for Joint Diseases* 2009;67(2):177-181.
- Lhotka C, Szekeres T, Steffan I, Zhuber K, Zweymuller K. Four-year study of cobalt and chromium blood levels in patients managed with two different metal-on-metal total hip replacements. *J Orthop Res* 2003;21(2):189-195.
- Li J, Xu W, Xu L, Liang Z. Hip resurfacing arthroplasty for ankylosing spondylitis. *J Arthroplasty* 2009;24(8):1285-1291.
- Lombardi AV, Mallory TH, Alexiades MM et al. Short-term results of the M2a-taper metal-on-metal articulation. *The Journal of arthroplasty* 2001;16:122-128.
- Lombardi AV, Mallory TH, Cuckler JM, Williams J, Berend KR, Smith TM. Mid-term results of a polyethylene-free metal-on-metal articulation. *The Journal of arthroplasty* 2004;19:42-47.
- Lombardi AV, Jr., Mallory TH, Alexiades MM et al. Short-term results of the M²-taper metal-on-metal articulation. *J Arthroplasty* 2001;16:122-128.
- Lombardi J, Mallory TH, Cuckler JM, Williams J, Berend KR, Smith TM. Mid-term results of a polyethylene-free metal-on-metal articulation. *J Arthroplasty* 2004;19(7 SUPPL.):42-47.
- MacDonald SJ, Bourne RB, Rorabeck CH, Cleland D, Chess DG, McCalden RW. Metal on metal versus metal on polyethylene in total hip arthroplasty - a prospective randomised clinical trial. *The Journal of Bone and Joint Surgery* 2005;87-B:321-32b.
- Maezawa K, Nozawa M, Hirose T et al. Cobalt and chromium concentrations in patients with metal-on-metal and other cementless total hip arthroplasty. *Arch Orthop Trauma Surg* 2002;122(5):283-287.

Maezawa K, Nozawa M, Matsuda K, Yuasa T, Shitoto K, Kurosawa H. Early failure of modern metal-on-metal total hip arthroplasty using a Wagner standard cup. *J Arthroplasty* 2006;21(4):522-526.

Maezawa K, Nozawa M, Yuasa T, Aritomi K, Matsuda K, Shitoto K. Seven years of chronological changes of serum chromium levels after Metasul metal-on-metal total hip arthroplasty. *J Arthroplasty* 2010;25(8):1196-1200.

Marker M, Grubl A, Riedl O, Heinze G, Pohanka E, Kotz R. Metal-on-metal hip implants: do they impair renal function in the long-term? A 10-year follow-up study. *Arch Orthop Trauma Surg* 2008;128(9):915-919.

McBryde CW, Revell MP, Thomas AM, Treacy RB, Pynsent PB. The influence of surgical approach on outcome in Birmingham hip resurfacing. *Clin Orthop Relat Res* 2008;466(4):920-926.

Migaud H, Jobin A, Chantelot C, Giraud F, Laffargue P, Duquennoy A. Cementless metal-on-metal hip arthroplasty in patients less than 50 years of age: comparison with a matched control group using ceramic-on-polyethylene after a minimum 5-year follow-up. *J Arthroplasty* 2004;19(8 Suppl 3):23-28.

Migaud H, Putman S, Krantz N, Vasseur L, Girard J. Cementless metal-on-metal versus ceramic-on-polyethylene hip arthroplasty in patients less than fifty years of age: a comparative study with twelve to fourteen-year follow-up. *J Bone Joint Surg Am* 2011;93 Suppl 2:137-142.

Mont MA, Marker DR, Smith JM, Ulrich SD, McGrath MS. Resurfacing is comparable to total hip arthroplasty at short-term follow-up. *Clin Orthop Relat Res* 2009;467(1):66-71.

Naudie D, Roeder CP, Parvizi J, Berry DJ, Egli S, Busato A. Metal-on-metal versus metal-on-polyethylene bearings in total hip arthroplasty: a matched case-control study. *J Arthroplasty* 2004;19(7 Suppl 2):35-41.

Pabinger C, Biedermann R, Stöckl B, Fischer M, Krismer M. Migration of metal-on-metal versus ceramic-on-polyethylene hip prostheses. *Clin Orthop* 2003;103-110.

Pattyn C, De Smet KA. Primary ceramic-on-ceramic total hip replacement versus metal-on-metal hip resurfacing in young active patients. *Orthopedics* 2008;31(11):1078.

Pollard TC, Baker RP, Eastaugh-Waring SJ, Bannister GC. Treatment of the young active patient with osteoarthritis of the hip. A five- to seven-year comparison of hybrid total hip arthroplasty and metal-on-metal resurfacing. *J Bone Joint Surg Br* 2006;88(5):592-600.

Rasquinha VJ, Ranawat CS, Weiskopf J, Rodriguez JA, Skipor AK, Jacobs JJ. Serum metal levels and bearing surfaces in total hip arthroplasty. *J Arthroplasty* 2006;21(6 Suppl 2):47-52.

Sandiford NA, Muirhead-Allwood SK, Skinner JA, Hua J. Metal on metal hip resurfacing versus uncemented custom total hip replacement--early results. *J Orthop Surg Res* 2010;5:8.

Savarino L, Granchi D, Ciapetti G et al. Ion release in patients with metal-on-metal hip bearings in total joint replacement: a comparison with metal-on-polyethylene bearings. *J Biomed Mater Res* 2002;63(5):467-474.

Savarino L, Greco M, Cenni E et al. Differences in ion release after ceramic-on-ceramic and metal-on-metal total hip replacement. Medium-term follow-up. *J Bone Joint Surg Br* 2006;88(4):472-476.

Savarino L, Padovani G, Ferretti M et al. Serum ion levels after ceramic-on-ceramic and metal-on-metal total hip arthroplasty: 8-year minimum follow-up. *J Orthop Res* 2008;26(12):1569-1576.

Schaffer AW, Pilger A, Engelhardt C, Zweymueller K, Ruediger HW. Increased blood cobalt and chromium after total hip replacement. *J Toxicol Clin Toxicol* 1999;37(7):839-844.

Schmalzried TP, Szuszczewicz ES, Akizuki KH, Petersen TD, Amstutz HC. Factors correlating with long term survival of McKee-Farrar total hip prostheses. *Clin Orthop Relat Res* 1996;(329 Suppl):S48-S59.

Sikes CV, Lai LP, Schreiber M, Mont MA, Jinnah RH, Seyler TM. Instability after total hip arthroplasty: treatment with large femoral heads vs constrained liners. *J Arthroplasty* 2008;23(7 Suppl):59-63.

Stulberg BN, Trier KK, Naughton M, Zadzilka JD. Results and lessons learned from a United States hip resurfacing investigational device exemption trial. *J Bone Joint Surg Am* 2008;90 Suppl 3:21-26.

Stulberg BN, Fitts SM, Bowen AR, Zadzilka JD. Early return to function after hip resurfacing: is it better than contemporary total hip arthroplasty? *J Arthroplasty* 2010;25(5):748-753.

Stulberg BN, Fitts SM, Bowen AR, Zadzilka JD. Early Return to Function After Hip Resurfacing. Is It Better Than Contemporary Total Hip Arthroplasty? *J Arthroplasty* 2010;25(5):798-753.

Vail TP, Mina CA, Yergler JD, Pietrobon R. Metal-on-metal hip resurfacing compares favorably with THA at 2 years followup. *Clin Orthop Relat Res* 2006;453:123-131.

Visuri T, Borg H, Pulkkinen P, Paavolainen P, Pukkala E. A retrospective comparative study of mortality and causes of death among patients with metal-on-metal and metal-on-polyethylene total hip prostheses in primary osteoarthritis after a long-term follow-up. *BMC Musculoskelet Disord* 2010;11:78.

Witzleb WC, Ziegler J, Krummenauer F, Neumeister V, Guenther KP. Exposure to chromium, cobalt and molybdenum from metal-on-metal total hip replacement and hip resurfacing arthroplasty. *Acta Orthop* 2006;77(5):697-705.

Yalcin N, Kilicarslan K, Cicek H, Kayaalp C, Yildirim H. Crowe Type I and II DDH managed by large diameter metal-on-metal total hip arthroplasty. *Hip Int* 2011;21(2):168-175.

Zerahn B, Borgwardt L, Ribel-Madsen S, Borgwardt A. A prospective randomised study of periprosthetic femoral bone remodeling using four different bearings in hybrid total hip arthroplasty. *Hip Int* 2011;21(2):176-186.

Zhang X, Xu W, Li J, Fang Z, Chen K. Large-diameter metal-on-metal cementless total hip arthroplasty in the elderly. *Orthopedics* 2010;33(12):872.

Zijlstra WP, Cheung J, Sietsma MS, Van Raay JJAM, Deutman R. No superiority of cemented metal-on-metal vs metal-on-polyethylene THA at 5-year follow-up. *Orthopedics* 2009;32(7):479.