Supplementary Material

Appendix 1

Medline Search Strategy
1. Arthroplasty, Replacement, Hip/
2. Hip Prosthesis/
3. Metal-on-Metal Joint Prostheses/
4. 1 or 2 or 3
5. ((joint or hip) adj2 (arthroplast$ or replacement$ or resurface$ or implant$ or prosthes$)).tw.
6. 4 or 5
7. Cobalt/
8. Chromium/
9. (chromium or cobalt).tw.
10. 7 or 8 or 9
11. Clinical Laboratory Techniques/
12. Biological Markers/an, bl, ch, du [Analysis, Blood, Chemistry, Diagnostic Use]
13. (blood or serum or plasma).tw.
14. 11 or 12 or 13
15. 4 and 6 and 10 and 14
16. animals/ not humans/
17. 15 not 16
18. (case reports or comment or editorial or english abstract).pt.
19. 17 not 18
20. limit 19 to english language

Embase Search Strategy
1. hip arthroplasty/
2. metal on metal joint prosthesis/
3. hip prosthesis/
4. ((joint or hip) adj2 (arthroplast$ or replacement$ or resurface$ or implant$ or prosthes$)).tw.
5. 1 or 2 or 3 or 4
6. cobalt/
7. chromium/
8. (chromium or cobalt).tw.
9. 6 or 7 or 8
10. biological marker/
11. blood analysis/
12. chemical analysis/
13. (blood or serum or plasma).tw.
14. cobalt blood level/
15. chromium blood level/
16. 10 or 11 or 12 or 13 or 14 or 15
17. 5 and 9 and 16
18. limit 17 to (amphibia or ape or bird or cat or cattle or chicken or dog or "ducks and geese" or fish or "frogs and toads" or goat or guinea pig or "hamsters and gerbils" or horse or monkey or mouse or "pigeons and doves" or "rabbits and hares" or rat or reptile or sheep or swine)
19. 17 not 18
20. case report/
21. 19 not 20
22. limit 21 to (conference abstract or conference paper or "conference review" or editorial or note or conference proceeding)
23. 21 not 22
24. limit 23 to english language

Appendix 2

ROC Curves

A standard receiver operative characteristic (ROC) curve plots specificity and sensitivity measured at different cut-points from a single study (Fig. a).[1] The points are generated by an underlying three step mechanism: (1) increasing cut-point, (2) increases specificity, and (3) decreases sensitivity (the opposite is also true). We can quantify the discriminating power of a test by computing the area under the ROC curve (AUC); which measures 0.79 for Figure a.
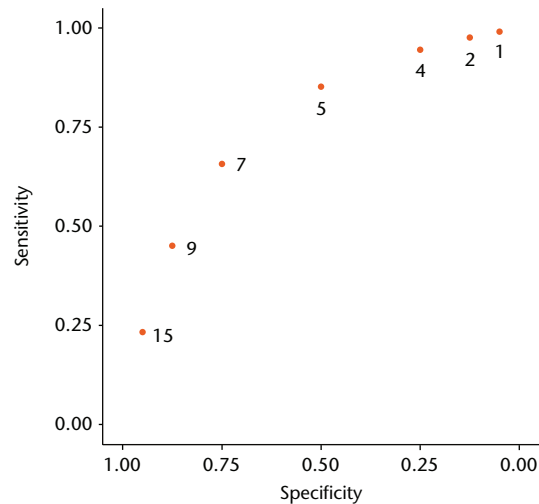
Fig. a

Standard receiver operative characteristic curve for a hypothetical study measuring specificity and sensitivity of metal ion concentrations for the diagnosis of adverse reactions to metal debris. The plot is annotated with the cut-point (µg/L) corresponding to the estimate of diagnostic accuracy.

The AUC is the probability that the test will correctly classify a patient as having or not having disease. Figure b plots three different ROC curves. Curve #1 is for a very good test; Curve #2 is the ROC curve from Figure b, it is not perfect but clinically useful test; Curve #3 is a diagonal line and has an AUC of 0.5. The used to generate Curve #3 is not useful because flipping a coin to diagnose patients would have equivalent discrimination performance. A common scale for interpreting the discriminative value of the AUC is listed in Table i.[2-4] A clinically useful test has an AUC ≥ 0.75.[5]
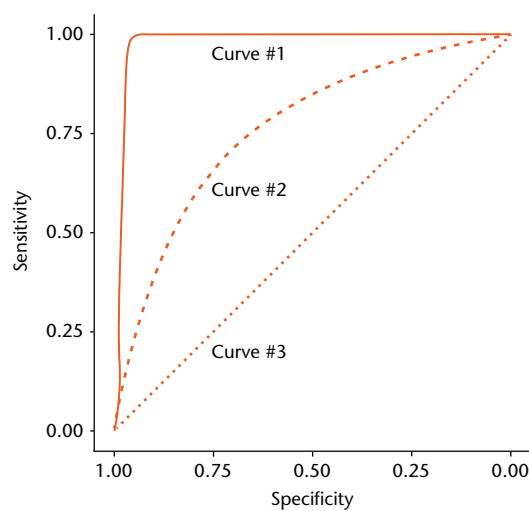
Fig. b

Three hypothetical receiver operative characteristic curves. Curve #1 indicates a high discrimination performance with an AUC of 0.98. Curve #2 indicates moderate discrimination performance with an area under curve (AUC) of 0.79. Curve #3 indicates poor discrimination performance with an AUC of 0.5.

Table i. Scale for Interpreting the area under curve. "[" and "]" indicates upper and lower limits, respectively, included in interval. "(" and ")" indicates upper and lower limits, respectively, excluded from interval

Discrimination Capacity	AUC
Excellent	[0.9, 1.0]
Good	[0.8, 0.9)
Fair	[0.7, 0.8)
Poor	[0.5, 0.7)
No	[0, 0.5)

Summary receiver operative characteristic (SROC) CURVES

The first step in diagnostic test meta-analysis is to plot specificity and sensitivity, measured at any cut-point, reported by different studies. In Figure c we see an example of such a plot. Studies report seemingly random sensitivities and specificities. We say these studies are "heterogeneous" because they do not seem to be reporting a common sensitivity and specificity.[6]
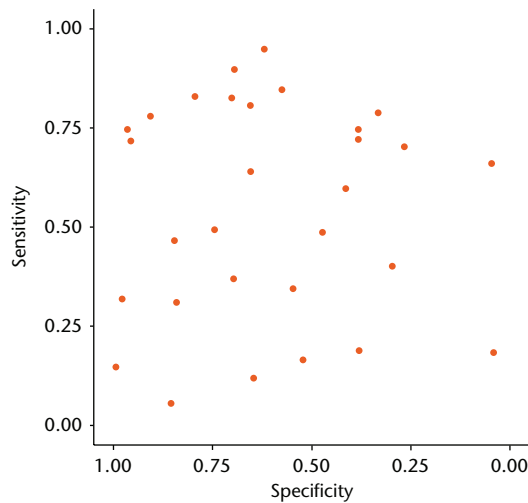

Fig. c

Hypothetical meta-analysis of studies reporting the diagnostic accuracy of ion levels for adverse reactions to metal debris. The random scatter indicates a high level of heterogeneity.

In Figure d the points appear to cluster around a single specificity and sensitivity. We say these studies are "homogeneous" because they seem to be reporting a common sensitivity and specificity. In this scenario, we can meta-analyse the individual estimates and compute a summary specificity and sensitivity.[7]
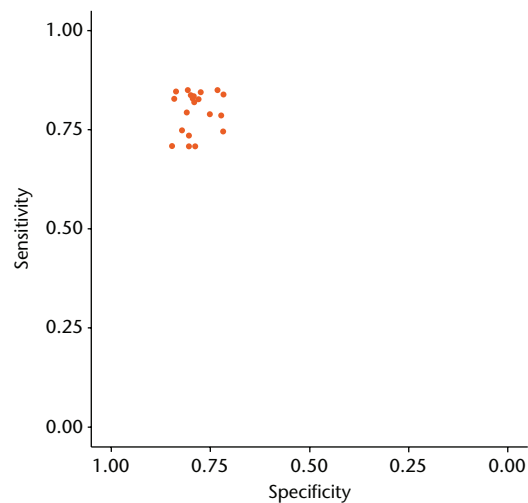

Fig. d

Hypothetical meta-analysis of studies reporting the diagnostic accuracy of ion levels for adverse reactions to metal debris. The clustering indicates a high level of homogeneity.

If each study reports diagnostic accuracy at a different cut-point, we might obtain a plot similar to Figure e. In this scenario, it does not make sense to compute summary specificity and sensitivity; we instead compute a summary ROC (SROC) curve. The points in Figure e are identical to the points in Figure a. If the cut-points used by individual studies are equivalent to the cut-points in Figure a, we can interpret the SROC curve in the same way as a standard ROC curve.
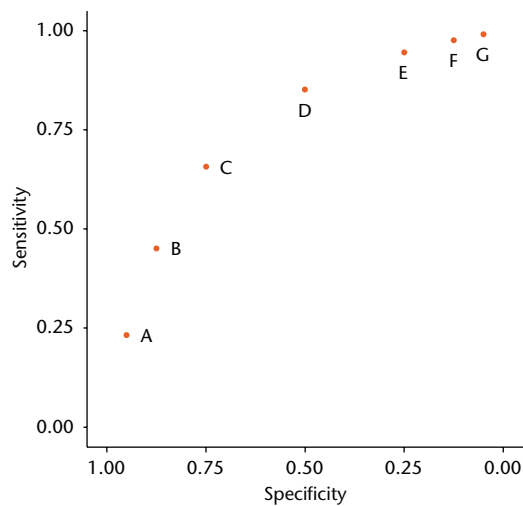

Fig. e

Hypothetical meta-analysis of seven studies reporting the diagnostic accuracy of ion levels for adverse reactions to metal debris.

In Figure e, the points lie on a smooth curve indicating homogeneity in the intrinsic discrimination performance of the test used in each study. There is heterogeneity in the cut-point, either implicit or explicit.

It is also possible to generate Figure e from a set of studies using the same cut-point. This could occur if the variation between studies is due to, for example, calibration of mass spectrophotometers used to take ion measurements. All studies report different sensitivities and specificities at the 7 μg/L threshold. However when plotted in ROC space they form a perfect curve. This occurs because, for example, a 7 μg/L cut-point in study A corresponds to the "true" a 15 μg/L cut-point, while a 7 μg/L cut-point in study B corresponds to the "true" a 9 μg/L cut-point, and so forth. In this situation where individual studies use a common explicit cut-point, but form an ROC curve, we conclude that an implicit cut-point is the source of heterogeneity. Any source of methodological, clinical and statistical heterogeneity can lead to an implicit cut-point effect.[6]

In the situation of implicit cut-point heterogeneity, we can still perform a meta-analysis using an SROC curve. An ROC curve plots the inherent discrimination capacity of a test. If studies report sensitivities and specificities which lie on a smooth ROC curve, each study is using a test with identical inherent discrimination capacity. Inherent discrimination capacity is independent of cut-point; this is evidenced by the fact that cut-point is not used to compute AUC. Therefore, we can still use SROC technique to meta- analyse the shape of ROC curve.[6]

Appendix 3
Characteristics of Studies

Bosker 2012

Study Characteristics

Patient sampling	116 patients (117 hips) patients enrolled in a previous RCT on surgical approach receiving MoM THA (Biomet M2A-Magnum) were considered for study. Excluded patients: death from cancer, death from cardiovascular disease, emigration, refusal, cemented acetabular component, revision prior to ion measurement. 105 patients (106 hips) were recruited into the study.
Patient characteristics and setting	Laboratory tests and imaging was obtained prospectively as part of this study. Patients were a mean of 43.2 months from the index surgery. 30.0% of patients were experiencing "non-specific groin pain" and 24.1% were ultimately revised.
Index tests	Serum cobalt and chromium measured from venous blood. Technique not described.
Target condition and reference standard(s)	Two radiologists read CT scans. Pseudotumour was defined as "a (semi)-solid or cystic peri-prosthetic soft-tissue mass with a diameter ≥ 2 cm that could not be attributed to an infection, malignancy, bursa or scar tissue. [...] A thickened capsule was recorded but not considered to constitute a pseudotumour."
Flow and timing	Unknown
Comparative	
Notes	

Methodological Quality

Item	Authors' judgment	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control study design avoided	Yes		
Did the study avoid inappropriate exclusions	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Unclear	

Malek 2012

Study Characteristics

Patient sampling	All patients with unexplained pain or limp identified through a regional surveillance program (269 hips) were considered for study. Excluded patients: bilateral implants (N=60). 209 patients were recruited into the study.
Patient characteristics and setting	As part of a regional surveillance program, all symptomatic patients underwent blood metal ion testing and MARS MRI. This study was a retrospective review. Patients were a mean 22.8 months from the index surgery.
Index tests	Plasma cobalt and chromium measured from venous blood. Measured by inductively coupled-plasma mass spectroscopy.
Target condition and reference standard(s)	Pseudotumour was defined as "presence of an abnormal, thick walled large fluid collection or a soft-tissue mass directly communicating with hip joint capsule" on MARS-MRI.
Flow and timing	Unknown
Comparative	
Notes	

Methodological Quality

Item	Authors' judgment	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control study design avoided	Yes		
Did the study avoid inappropriate exclusions	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
		Low	Unclear
DOMAIN 3: Reference Standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Low	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Unclear	

Bisschop 2013

Study Characteristics

Patient sampling	All patients who received a Smith & Nephew BHR hip resurfacing implant at one hospital (129 patients, 149 hips) were considered for this study. Excluded patients: bilateral implants (N=16), could not be contacted (N=4). 109 patients were recruited into the study.
Patient characteristics and setting	Laboratory tests and imaging was obtained prospectively as part of this study. Patients were a mean 41 months from the index surgery. 37.6% of the 125 patients with unilateral or bilateral implants had "complaints," groin pain, groin discomfort, mass, or neurologic symptoms. 2.1% of the 125 patients with unilateral or bilateral implants were revised.
Index tests	Serum cobalt measured from venous blood. Measured by inductively coupled-plasma mass spectroscopy.
Target condition and reference standard(s)	CT scan without metal suppression protocol. Pseudotumour was defined "solid, semisolid, or cystic eccentric extension of the capsule, resulting in an increase in the volume of the capsule that could not be attributed to an infection, malignancy, bursa, or scar tissue."
Flow and timing	Unknown
Comparative	
Notes	

Methodological Quality

Item	Authors' judgment	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control study design avoided	Yes		
Did the study avoid inappropriate exclusions	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		High	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Unclear	

Chang 2013

Study Characteristics	
Patient sampling	All patients who presented to a single orthopaedic surgeon after the recall of the DePuy ASR implant (227 patients) were considered for this study. Excluded patients: bilateral implants (N=18); lost to follow-up (N=26); failed to complete both investigations, decline to participate, contraindication to MRI, could not tolerate MRI (N=27). 156 patients were recruited into the study.
Patient characteristics and setting	As part of surveillance following implant recall, all symptomatic patients underwent blood metal ion testing and MRI. This study was a retrospective review. Patients were a mean 43 months from the index surgery. 48.7% of experienced pain ormechanical symptoms. It was not reported if any patients underwent revision.
Index tests	Plasma cobalt and chromium measured from venous blood. Measured by inductively coupled-plasma mass spectroscopy.
Target condition and reference standard(s)	MRI scans for each patient were retrospectively reviewed with consensus interpretation by two fellowship-trained musculoskeletal radiologists. Pseudotumor was defined as "a periprosthetic collection of any size, either of fluid or solid signal intensity, excluding iliopsoas bursal distention."
Flow and timing	Unkown.
Comparative	
Notes	

Methodological Quality			
Item	**Authors' judgment**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control study design avoided	Yes		
Did the study avoid inappropriate exclusions	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
		Low	Unclear
DOMAIN 3: Reference Standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Unclear	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Unclear	

Macnair 2013

Study Characteristics

Patient sampling	76 patients (77 hips) who received THA or HR (DePuy ASR) at one hospital were considered for this study. Excluded patients: underwent revision (N=13); failed to complete investigations (N=2); declined to participate (N=4). 57 patients (62 hips) were recruited into the study.
Patient characteristics and setting	As part of surveillance of MoM implants, all patients underwent MARS MRI. Subsequently, patients underwent blood metal ion testing as part of this study. Patients were a mean 45 months from the index surgery. 44.6% of patients were deemed symptomatic (OHS <= 44); assuming a normal distribution of OHS with mean 43 and standard deviation 8.7.
Index tests	Serum cobalt and chromium measured from venous blood. Measured by inductively coupled-plasma mass spectroscopy
Target condition and reference standard(s)	MARS MRI reviewed by 2 musculoskeletal radiologists. Pseudotumour was defined as a periprosthetic soft tissue mass with no hyperintense T2W fluid signal or fluid-filled periprosthetic cavity.
Flow and timing	Unknown
Comparative	
Notes	

Methodological Quality

Item	Authors' judgment	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control study design avoided	Yes		
Did the study avoid inappropriate exclusions	Unclear		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Low	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Unclear	

Van der Weegen 2014

Study Characteristics

Patient sampling	Prospective cohort of 258 patients (296 hips) who received HR (Biomet ReCap) at one hospital. Excluded: 6 patients who died, 17 patients who were revised, 3 patients lost to follow-up, 4 patients unable to participate in study procedures, 7 patients who refused. 221 patients (256 hips) patients were recruited into the study.
Patient characteristics and setting	All patients underwent blood metal ion testing and MARS MRI or ultrasound as part of surveillance for MoM impants. Patients were a mean 55.2 months from the index surgery. 23.6% of 256 hips were painful.
Index tests	Serum cobalt and chromium measured from venous blood. Measured by atomic absorption spectrometry.
Target condition and reference standard(s)	MARS MRI reviewed by 2 musculoskeletal radiologists. Pseudotumour was defined as a periprosthetic soft tissue mass with no hyperintense T2W fluid signal or fluid-filled periprosthetic cavity. In patients' with contra-indication to MRI, ultrasound was used.
Flow and timing	Unkown
Comparative	
Notes	

Methodological Quality

Item	Authors' judgment	Risk of bias	Applicability concerns
DOMAIN 1: Patient Selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control study design avoided	Yes		
Did the study avoid inappropriate exclusions	Yes		
		Low	Low
DOMAIN 2: Index Test All tests			
Were the index test results interpreted without knowledge of the results of the reference standard	Unclear		
If a threshold was used, was it pre-specified?	Yes		
		Low	Low
DOMAIN 3: Reference Standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
		Low	Low
DOMAIN 4: Flow and Timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive a reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
		Unclear	

Appendix 4

Demographic and patient characteristics

Study	N/n	Mean Age	Gender % Male	Country	THA:HR (%:%)	Bilateral (%)	Implant	Time from implantation (measure, range)	Definition of symptomatic	% Patients Symptomatic	Sampling
Bosker 2012	105/106	60	48	Netherlands	106:0 (100:0)	0 (0)	M2A Magnum (Biomet)	3.6 years (median, 2.1-4.5)	Groin/buttock/leg/thigh pain, clicking, swelling	30	All eligible patients from referenced RCT
Malek 2012	209/209	58	40	UK	190:19 (91:9)	0 (0)	Various	1.9 years (median, 0.5-5.5)	Unexplained pain or limp	100	Retrospective consecutive cohort of symptomatic MoM hip patients at one institution
Bisschop 2013	109/109	54	46	Netherlands	0:109 (0:100)	0 (0)	BHR (Smith and Nephew)	4.1 months (mean, 10- 82)	Unclear (groin pain, groin discomfort, mass, neurologic symptoms	38	Prospective cohort of patient with BHR implants at one institution
Chang 2013	156/156	65	44	USA	0:156 (0:100)	0 (0)	ASR (Depuy)	43 months (median, range 12- 59)	Pain or mechanica l symptoms	49	Retrospective review of all patients with ASR (Depuy) implants implanted by one surgeon
Macnair 2013	57/62	54	83	UK	46:16 (74:26)	18 (29)	ASR ASR/Co rail (Depuy)	45 months (mean, 31- 64)	Oxford hip score ≤44	45	Retrospective review of all patients who received ASR implants at one institution
Van der Weegen 2014	221/256	54	Not reported	Netherlands	0:256 (0:100)	35 (16)	ReCap resurfaci ng hip (Biomet)	4.6 years (mean, 1-8.2)	Hip pain	24	Prospective cohort of all patients who received HR at one institution

N=number of patients included in analysis; n=number of hips included in analysis

Appendix 5

Reference test for pseudotumor detection

Study	Diagnostic imaging modality	Number of image interpreters	Definition of pseudotumor	Blinding to clinical or ion information	ARMD prevalence (% of hips)
Bosker 2012	CT, confirmed by MRI/US	2	Semi-solid or cystic periprosthetic soft-tissue mass with a diameter ≥ 2cm that could not be attributed to an infection, malignancy, bursa or scar tissue	Unclear	40 (38)
Malek 2012	MARS-MRI	Unclear	Presence of an abnormal, thick walled large fluid collection or a soft-tissue mass directly communicating with hip joint capsule	Unclear	84 (40)
Bisschop 2013	CT	Unclear	Solid, semisolid, or cystic eccentric extension of the capsule, resulting in an increase in the volume of the capsule that could not attributed to an infection, malignancy, bursa or scar tissue	Unclear	36 (33)
Chang 2013	MRI	2	Periprosthetic collection of any size, either fluid or solid signal intensity, excluding iliopsoas bursal distension	Unclear	107 (69)
Macnair 2013	MARS-MRI	2	periprosthetic soft tissue mass with no hyperintense T2W fluid signal or fluid-filled periprosthetic cavity	Unclear	18 (29)
Van der Weegen 2014	MARS-MRI, US (if MRI contraindicate d)	2	periprosthetic soft tissue mass with no hyperintense T2W fluid signal or fluid-filled periprosthetic cavity	Unclear	91 (36)

Appendix 6

Index test and metal ion concentration results

Study	Metal ions included in analysis	Blood fraction measured	Method of ion measurement	Metal ion threshold (pre-specified, yes/no)	Metal ion results at testing threshold(s)
Bosker 2012	Co, Cr	Serum	Inductively coupled plasma mass spectrometry	>5μg/L (Yes)	[Co] > 5μg/L DOR 4.05 (95% CI: 1.63 to 10.06), sensitivity 0.575, specificity 0.773 [Cr] > 5μg/L DOR 1.96 (95% CI: 0.87 to 4.41), sensitivity 0.525, specificity 0.636
Malek 2012	Co, Cr	Plasma	Inductively coupled plasma mass spectrometry	Various >2μg/L, >3.5μg/L, >7μg/L, >11μg/L, >16μg/L (Yes)	[Co or Cr] > 7μg/L sensitivity 0.57, specificity 0.65 [Co or Cr] > 3.5μg/L sensitivity 0.86, specificity 0.27 [Co] > 2μg/L sensitivity 0.9, specificity 0.25 [Cr] > 2μg/L sensitivity 0.9, specificity 0.14 [Co] > 16μg/L sensitivity 0.29, specificity 0.9 [Cr] > 11μg/L sensitivity 0.25, specificity 0.9
Bisschop 2013	Co	Serum	Inductively coupled plasma mass spectrometry	>85nmol/L (Yes)	[Co] > 85nmol/L DOR 4.9, sensitivity 0.222, specificity 0.945
Chang 2013	Co, Cr	Plasma	Inductively coupled plasma mass spectrometry	>5ppb, equivalent to >5μg/L (Yes)	[Co] > 5μg/L sensitivity 0.22, specificity 0.92 [Cr] > 5μg/L sensitivity 0.12, specificity 0.96
Macnair 2013	Co, Cr	Serum	Inductively coupled plasma mass spectrometry	>4μg/L, >7μg/L (Yes)	[Co] > 7μg/L sensitivity 0.56, specificity 0.76 [Co] > 4μg/ L sensi ti vi t y 0.72, specificity 0.66 [Cr] > 7μg/L sensitivity 0.56, specificity 0.83 [Cr] > 4μg/L sensitivity 0.61, specificity 0.66
Van der Weegen 2014	Co, Cr	Serum	Atomic absorption spectrometry	>7μg/L (Yes)	[Co] > 7μg/L sensitivity 0.08, specificity 0.99 [Cr] > 7μg/L sensitivity 0.12, specificity 0.98

References

1. **Zhou XH, Obuchowski NA, McClish DK.** *Statistical Methods in Diagnostic Medicine.* Second ed. New York: Wiley & Sons, 2011.[[bibmisc]]
2. **Hanley JA, McNeil BJ.** The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
3. **Hanley JA.** Receiver operating characteristic (ROC) methodology: the state of the art. *Crit Rev Diagn Imaging* 1989;29:307-335.
4. **Muller MP, Tomlinson G, Marrie TJ, et al.** Can routine laboratory tests discriminate between severe acute respiratory syndrome and other causes of community-acquired pneumonia? *Clin Infect Dis* 2005;40:1079-1086.
5. **Fan J, Upadhye S, Worster A.** Understanding receiver operating characteristic (ROC) curves. *CJEM* 2006;8:19-20.
6. **Littenberg B, Moses LE.** Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Med Decis Making.* 1993;13:313–321.
7. **Chappell FM, Raab GM, Wardlaw JM.** When are summary ROC curves appropriate for diagnostic meta-analyses? *Stat Med* 2009;28:2653-2668.
8. **Bosker BH, Ettema HB, Boomsma MF, et al.** High incidence of pseudotumour formation after large-diameter metal-on-metal total hip replacement: A prospective cohort study. *J Bone Joint Surg [Br]* 2012;94-B:755-761.
9. **Malek IA, King A, Sharma H, et al.** The sensitivity, specificity and predictive values of raised plasma metal ion levels in the diagnosis of adverse reaction to metal debris in symptomatic patients with a metal-on-metal arthroplasty of the hip. *J. Bone Joint Surg [Br]* 2012;94-B:1045–1050.
10. **Bisschop R, Boomsma MF, Van Raay JJ, et al.** High prevalence of pseudotumors in patients with a Birmingham Hip Resurfacing prosthesis: a prospective cohort study of one hundred and twenty-nine patients. *J Bone Joint Surg [Am]* 2013;95-A:1554-1560.
11. **Chang EY, McAnally JL, Van Horne JR, et al.** Relationship of plasma metal ions and clinical and imaging findings in patients with ASR XL metal-on-metal total hip replacements. *J Bone Joint Surg [Am]* 2013;95-A:2015-2020.
12. **Macnair RD, Wynn-Jones H, Wimhurst JA, Toms A, Cahir J.** Metal ion levels not sufficient as a screening measure for adverse reactions in metal-on-metal hip arthroplasties. *J Arthroplasty* 2013;28:78-83.
13. **Van der Weegen W, Sijbesma T, Hoekstra HJ, et al.** Treatment of pseudotumors after metal-on-metal hip resurfacing based on magnetic resonance imaging, metal ion levels and symptoms. *J Arthroplasty* 2014;29:416-421.