



## ■ HIP

# Blood metal ion levels are not a useful test for adverse reactions to metal debris: a systematic review and meta-analysis

**M. Pahuta,  
J. M. Smolders,  
J. L. van Susante,  
J. Peck,  
P. R. Kim,  
P. E. Beaulé**

*University of Ottawa/  
The Ottawa Hospital,  
Ottawa, Canada*

## Objectives

Alarm over the reported high failure rates for metal-on-metal (MoM) hip implants as well as their potential for locally aggressive Adverse Reactions to Metal Debris (ARMDs) has prompted government agencies, internationally, to recommend the monitoring of patients with MoM hip implants. Some have advised that a blood ion level  $>7 \mu\text{g/L}$  indicates potential for ARMDs. We report a systematic review and meta-analysis of the performance of metal ion testing for ARMDs.

## Methods

We searched MEDLINE and EMBASE to identify articles from which it was possible to reconstruct a  $2 \times 2$  table. Two readers independently reviewed all articles and extracted data using explicit criteria. We computed a summary receiver operating curve using a Bayesian random-effects hierarchical model.

## Results

Our literature search returned 575 unique articles; only six met inclusion criteria defined a priori. The discriminative capacity of ion tests was homogeneous across studies but that there was substantial cut-point heterogeneity. Our best estimate of the “true” area under curve (AUC) for metal ion testing is 0.615, with a 95% credible interval of 0.480 to 0.735, thus we can state that the probability that metal ion testing is actually clinically useful with an  $\text{AUC} \geq 0.75$  is 1.7%.

## Conclusion

Metal ion levels are not useful as a screening test for identifying high risk patients because ion testing will either lead to a large burden of false positive patients, or otherwise marginally modify the pre-test probability. With the availability of more accurate non-invasive tests, we did not find any evidence for using blood ion levels to diagnose symptomatic patients.

**Cite this article:** *Bone Joint Res* 2016;5:379–386

**Keywords:** Metal-on-metal; Metal ion; Hip implants

■ M. Pahuta, MD, Orthopaedic Resident,  
■ J. Peck, MD, Orthopaedic Resident,  
■ P. R. Kim, MD, Associate Professor of Surgery,  
■ P. E. Beaulé, MD, Professor of Surgery, Division Head, University of Ottawa, The Ottawa Hospital, 501 Smyth Road, Ottawa, Ontario, Canada.  
■ J. M. Smolders, MD, PhD, Orthopaedic Fellow,  
■ J. L. van Susante, MD, Professor of Surgery, Rijnstate Hospital, Arnhem, The Netherlands.

Correspondence should be sent to P. E. Beaulé; email: pbeaule@toh.ca

doi: 10.1302/2046-3758.59.BJR-2016-0027.R1

*Bone Joint Res* 2016;5:379–386.

Received: 19 January 2016

Accepted: 19 May 2016

## Article focus

■ We report a systematic review and meta-analysis of the performance of blood metal ion testing for adverse reactions to metal debris (ARMDs) in patients with metal-on-metal (MoM) hip implants.

## Key messages

■ Blood metal ion levels are not useful as a screening test for identifying high-risk patients because ion testing will either lead to a large burden of false positive results or otherwise marginally modify the pre-test probability.

■ With the availability of more accurate non-invasive tests such as MRI, there is no role for using blood ion levels to diagnose symptomatic patients.

## Strengths and limitations

■ Out of 575 references identified by the literature search, only six met selection criteria.  
■ These six studies were homogeneous in terms of diagnostic performance estimates.

## Introduction

Despite the fact that total hip arthroplasty (THA) has been touted as the operation of the 21st century, orthopaedic researchers continue to propose new designs in an effort to improve implant longevity and patient function.<sup>1</sup> One such purported improvement was the use of a large femoral head with a metal-on-metal (MoM) bearing; this design failed and has become a significant public health concern leading to withdrawal and even recall of certain implant designs.<sup>2,3</sup> This is yet another example of the flawed cycle of innovation in arthroplasty where new implants actually underperform relative to the existing standard.<sup>4</sup> This alarm over MoM hip implants stems from reported high failure rates, and the potential for locally aggressive ion-induced local tissue reactions such as pseudotumours, a type of adverse reaction to metal debris (ARMD).<sup>5</sup> It has been estimated that over a million patients have received these implants worldwide, thus posing a significant concern in regard to monitoring and advising patients on both short and long term performance.<sup>5</sup>

In the 1980s, MoM bearings were introduced as an improvement on standard metal-on-polyethylene (MoP) bearings. MoM bearings generate less volumetric wear which may translate into longer implant survival.<sup>6</sup> In addition, MoM bearings facilitate greater hip stability and range of movement from larger head sizes. Furthermore, MoM bearings allow for bone-conserving hip resurfacing (HR) which is important for younger patients who will eventually require revision surgery.<sup>7</sup> Both MoP and MoM bearings have been associated with elevated systemic levels of metal ions and ARMDs, however, ion levels have been shown to be consistently higher in patients with MoM bearings, and ARMDs appear to be more common in these patients.<sup>8-11</sup>

ARMDs span a spectrum of aseptic necrotic effusions that include pseudotumours as well as aseptic lymphocyte-dominated vasculitis-associated lesions (ALVL) (pericapsular hypersensitivity reactions associated with osteolysis).<sup>12-14</sup> When patients present with mechanical symptoms or pain following a MoM hip implant, unlike with a MoP hip implant, expectant management may not be advisable due to the potential for ARMD lesions which can lead to progressive tissue destruction that compromises reconstructive options.<sup>15-17</sup> As these lesions are diagnosed with ultrasound (US) or metal artefact reduction sequence magnetic resonance imaging (MARS MRI), there has been interest in developing an inexpensive and rapid laboratory test.<sup>18</sup>

Government agencies worldwide have published recommendations on the surveillance and work-up of patients with MoM hip implants. The United Kingdom's Medicines and Healthcare Regulatory Agency (Medicines and Healthcare Regulatory Agency 2012),<sup>19</sup> the United States' Food and Drug Administration (Food and Drug Administration 2013),<sup>20</sup> the European Commission's Scientific

Committee on Emerging Newly Identified Health Risks (European Commission Scientific Committee on Emerging Newly Identified Health Risks 2014),<sup>21</sup> Health Canada (Health Canada 2012)<sup>22</sup> and the Therapeutic Goods Administration of Australia (Therapeutic Goods Administration 2012)<sup>23</sup> have recommended close follow-up of patients, even those with well-functioning implants. Surveillance with metal ions has been recommended for patients with the ASR implant (Depuy Synthes, Warsaw, Indiana) in the United Kingdom, with large-diameter THA and small-diameter hip resurfacing (HR) in Australia, and with large-diameter THA and any HR in Europe. The recommended work-up of symptomatic patients by each of these organisations includes blood metal ion assessment. To date, the most detailed guidelines were put forth by the United Kingdom's Medicines and Healthcare Regulatory Agency which advised that a blood metal ion level > 7 µg/L indicates potential for soft-tissue reaction. As noted by others, ion level cut-offs are arbitrary and not supported by scientific data.<sup>5,24-26</sup>

Given the uncertain utility of laboratory testing in surveillance and investigation of patients with MoM hip implants, there is a need to synthesise the evidence for the measurement of blood cobalt and chromium ion concentrations. In this paper we report a systematic review and meta-analysis of the screening and diagnostic value of metal ion testing for ARMDs.

## Materials and Methods

**Literature search and data extraction.** We conducted an electronic search to identify relevant articles that reported original research findings including blood ion concentrations in patients with total hip arthroplasty (THA) or hip resurfacing (HR). MEDLINE (between January 1946 and 15th February 2015) and EMBASE (between January 1974 and February 2015) were searched for relevant publications with the assistance of a clinical librarian. The electronic search was individually tailored to each database to maximise sensitivity (see Appendix 1). We supplemented the electronic search by obtaining referenced articles and articles citing articles were each of the articles ultimately included in the meta-analysis through Scopus. Two readers (a fellowship-trained joint reconstruction surgeon and orthopaedic surgery resident) independently reviewed all articles using explicit criteria, and recorded assessments using a standard computerised form. A third reader (a fellowship-trained joint reconstruction surgeon) resolved any disagreements. Readers screened titles and abstracts to exclude animal and basic science studies, review articles, guidelines, and editorials. The readers identified articles evaluating a hip prosthesis (either HR or THA) and reporting cobalt and/or chromium blood ion concentrations from the full text.

We included articles in the meta-analysis if it was possible to reconstruct a 2 × 2 table for the use of blood ion

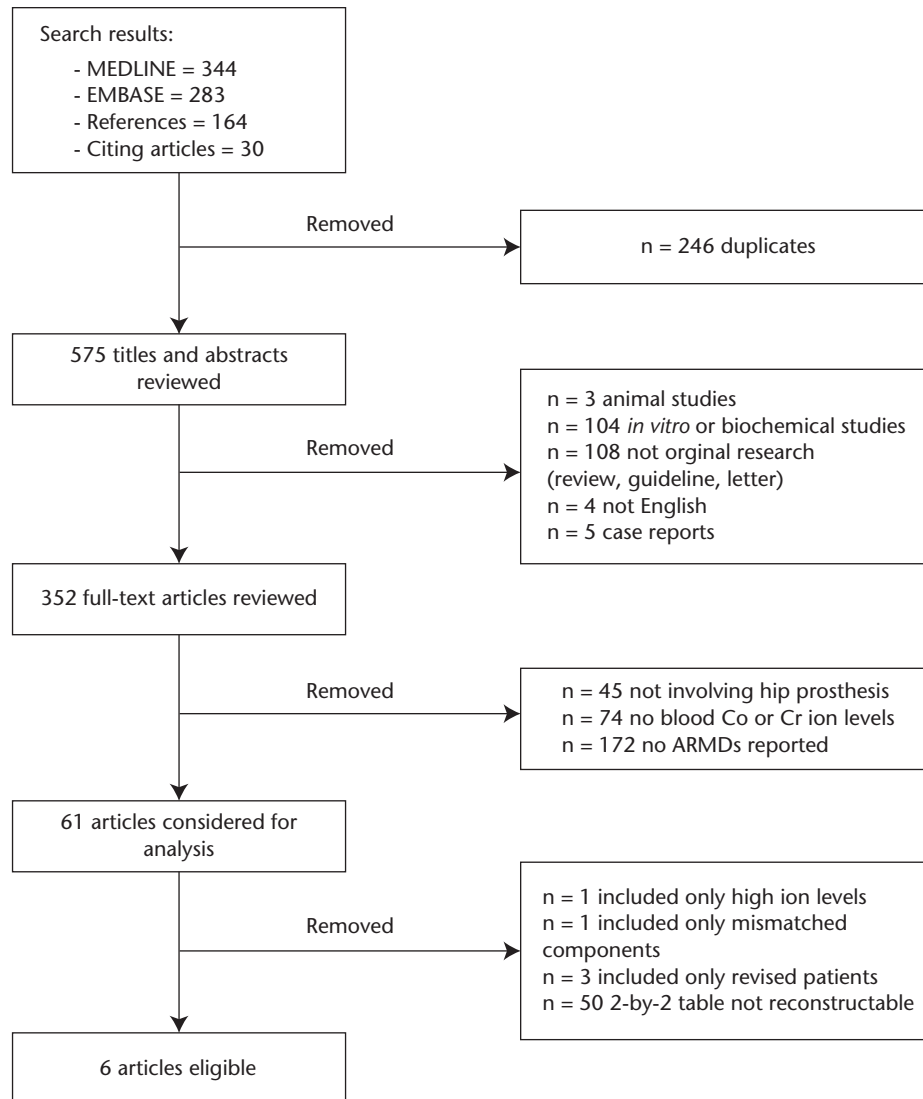


Fig. 1

Flow of studies through selection process.

measurements as a test for ARMDs. Only studies that evaluated metal ion levels for screening or for diagnosis of symptomatic patients were eligible. Studies that only recruited patients who underwent revision surgery were excluded (three studies, 81 hips). In these papers, the decision making for revision was not clearly described. Consequently we felt that there was a high risk of bias from the spectrum effect as ion levels were likely used in the decision-making process.<sup>27</sup> Eligible measures of diagnostic performance included: sensitivity; specificity; predictive values; likelihood ratios; diagnostic odds ratios; and receiver operating characteristic (ROC) curves. If an odds ratio for an ion cut-point used as a covariate in logistic regression was reported, we deemed the study eligible. Reviewers collected the following covariates: country of study; inclusion criteria; benchmark test; index test; number of patients; ARMD prevalence; number of

revisions; and prevalence of symptomatic patients. Reviewers also assessed the quality of studies using the QUADAS-2 tool.<sup>28</sup>

**Statistical analysis.** We selected the diagnostic performance meta-analytic technique according to the algorithm proposed by Chappell, Raab and Wardlaw.<sup>29</sup> Ultimately, we computed a summary receiver operating characteristic (SROC) curve using a random-effects hierarchical SROC (HSROC) model controlling for the different cut-points reported (see Appendix 2).<sup>30,31</sup> We quantified heterogeneity by comparing the widths of 95% confidence intervals (CIs) and 95% prediction intervals (PIs).<sup>32</sup> **Implementation.** We performed Bayesian computation for both the diagnostic performance and normative ion level meta-analyses using R (R Foundation, Vienna, Austria) and the Bayesian modeling language, Stan.<sup>33</sup> We ran four chains for 5000 iterations, discarded the

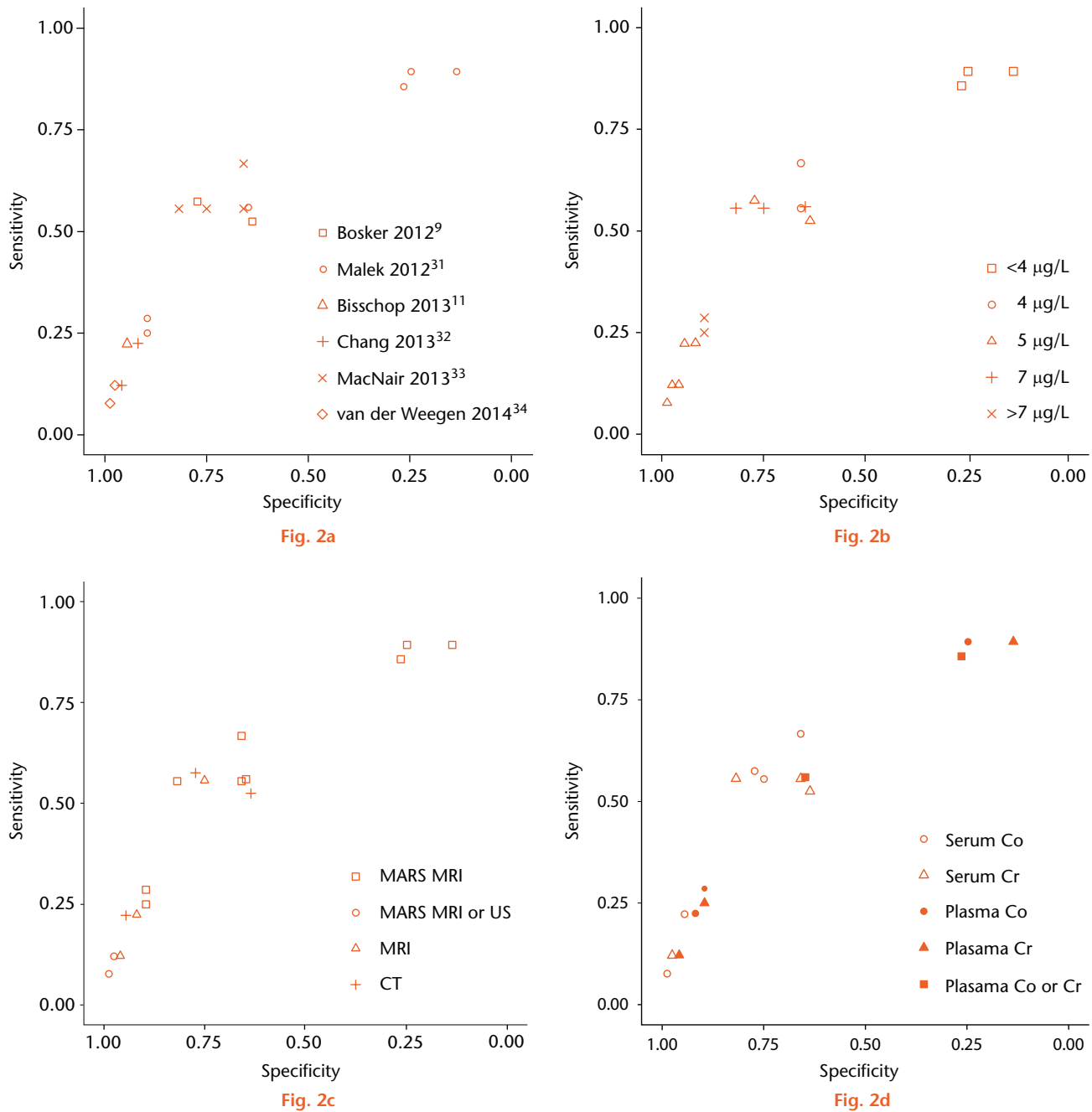


Fig. 2

first 2500 and used a thinning interval of five iterations. We assessed appropriate sampling of chains graphically by ensuring mixing on trace plots, and convergence by ensuring the Gelman-Rubin statistic was  $< 1.2$ .<sup>34,35</sup> We used non-informative prior distributions.

## Results

Our literature search identified 575 unique references (Fig. 1). Six met the selection criteria defined *a priori* (see Appendix 1).<sup>9,11,36-39</sup> We contacted the authors of two

studies<sup>9,11</sup> reporting logistic regression with a  $> 5 \mu\text{g/L}$  blood ion cut-point for additional data.

These six studies included a total of 898 hips, of which 376 had an ARMD. The prevalence of ARMDs ranged from 29% to 69%, and the prevalence of symptoms ranged from 23.6% to 100% (Figs 2f and 2g, Appendices 3, 4, 5, 6). Studies differed in the blood fraction tested, ion measured and cut-point used. Only 50% of studies<sup>36,38,39</sup> used MARS MRI as the benchmark for diagnosis. Only one study<sup>36</sup> used blood ion levels in a diagnostic

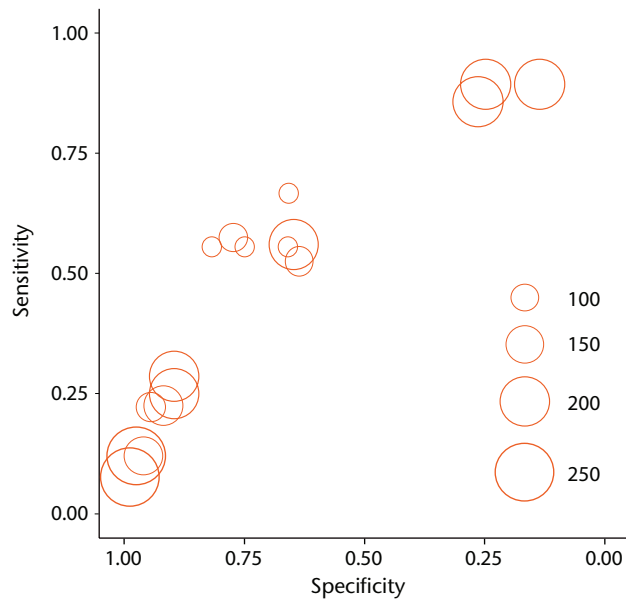


Fig. 2e

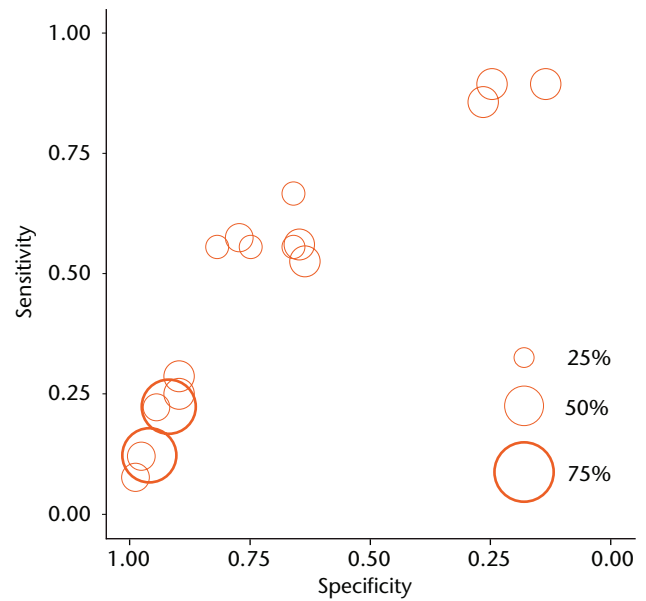


Fig. 2f

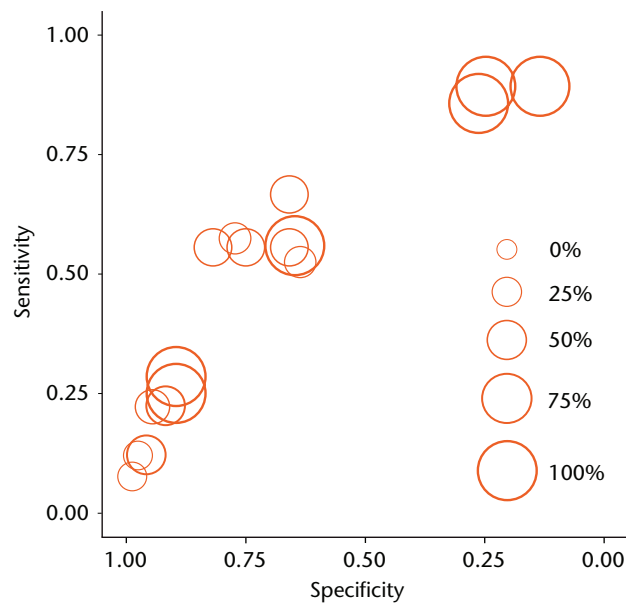


Fig. 2g

Fig. 2 (continued)

Individual reports of sensitivity and specificity plotted in receiver operating characteristic space with (a) study, (b) ion threshold, (c) benchmark test, (d) index test characteristics, (e) number of hips, (f) adverse reactions to metal debris prevalence, and (g) prevalence of symptomatic patients highlighted. Three studies reported more than one estimate (a). Circle size is proportional to sample size in (e) and prevalence in (f) and (g).

context (symptomatic patients) whereas the remaining four studies<sup>9,11,37,39</sup> used blood ion levels in a screening (undifferentiated patients) context. The three studies<sup>9,11,37</sup> not using MARS MRI were deemed at risk of bias (Table I, Fig. 2c, Appendix 3). Two studies<sup>36,37</sup> used plasma, rather than serum, for ion testing and were therefore deemed to have concerns regarding applicability (Table I, see also Appendix 3). No study described the time interval between ion testing and imaging, and therefore all were deemed to have concerns regarding applicability (Table I,

see also Appendix 3). All studies were of either Level I or Level II quality.<sup>40</sup> Prior to proceeding with meta-analysis, we investigated whether clinical, methodological and quality variability manifested in heterogeneity in the estimates of diagnostic accuracy obtained from each study.

Study-specific estimates of specificity and sensitivity all appear to lie close to a common smooth ROC curve (Fig. 2a), however, there appeared to be variability in the performance of specific cut-points across different studies (Fig. 2b). This suggests that the discriminative capacity of the

**Table I.** QUADAS-2 risk of bias table for included studies

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Bosker et al 2012 <sup>9</sup>	☺	☺	☹	???	☺	☺	☺
Malek et al 2012 <sup>36</sup>	☺	☺	☺	???	☺	???	☺
Bisschop et al 2013 <sup>11</sup>	☺	☺	☹	???	☺	☺	☺
Chang et al 2013 <sup>37</sup>	☺	☺	☹	???	☺	???	☺
MacNair et al 2013 <sup>38</sup>	☺	☺	☹	???	☺	☺	☺
Van der Weegen et al 2014 <sup>39</sup>	☺	☺	☺	???	☺	☺	☺

☺ Low risk; ☹ High risk; ??? Unclear risk

ion test is homogeneous across studies but that there is substantial cut-point heterogeneity. The cut-point heterogeneity may be due to heterogeneity in the benchmark modality used: the studies reporting unexpectedly high specificity and low sensitivity at 5 µg/L and 7 µg/L cut-points did not exclusively use MARS MRI as the benchmark (Fig. 2c). Sample size, ARMD prevalence, prevalence of symptomatic patients, and ion test characteristics were not associated with cut-point heterogeneity (Figs 2d, 2e, 2f and 2g). Given the homogeneity in discrimination capacity but implicit cut-point heterogeneity we pursued SROC meta-analysis without meta-analysing cut-points (see Appendix 2).

Our best estimate of the “true” ROC curve for metal ion test is the mean SROC curve plotted in Figure 3. However, due to random variability the “truth” may not be the same in all studies. Accounting for this random variability, we have 95% confidence that the study-specific “truth” will lie within the 95% prediction region. The prediction and credible regions have similar widths, which further supported minimal heterogeneity. The area under the curve (AUC) for the SROC curve was 0.615 (95% CI 0.480 to 0.735), thus we can state that the probability that metal ion testing is actually clinically useful with  $AUC \geq 0.75$  is 1.7% (see Appendix 2).<sup>41</sup>

Due to implicit cut-point heterogeneity, we did not perform meta-analysis of cut-point performance (see Appendix 2). Therefore, the SROC curve in Figure 3 does not relate cut-points to a particular specificity and sensitivity. However, diagnostic performance at any given cut-point will lie somewhere on the SROC curve in Figure 3 – we just do not know where. Hence, our meta-analysis can be used to evaluate the overall performance of ion tests, without reference to a particular cut point.

## Discussion

This systematic review and meta-analysis is the first synthesis of evidence for the use of blood ion measurements as a test for ARMDs in patients with MoM hip implants. We identified minimal heterogeneity in the inherent discrimination capacity of ion tests used in each study. Our meta-analysis indicates that blood ion levels are a poor

test for classifying patients as having or not having an ARMD.

All but one study included in our review evaluated blood ion levels in a screening context. Estimates of diagnostic accuracy obtained from high prevalence/symptomatic samples can be biased upwards due to the spectrum effect.<sup>22</sup> The prevalence of ARMDs and symptomatic patients in the included studies spanned a wide range (29% to 69%), therefore, we could graphically evaluate for a spectrum effect. Since Figures 2f and 2g demonstrated that symptom prevalence was not associated with the operating point, we concluded that there was an absence of spectrum effect in our meta-analysis.

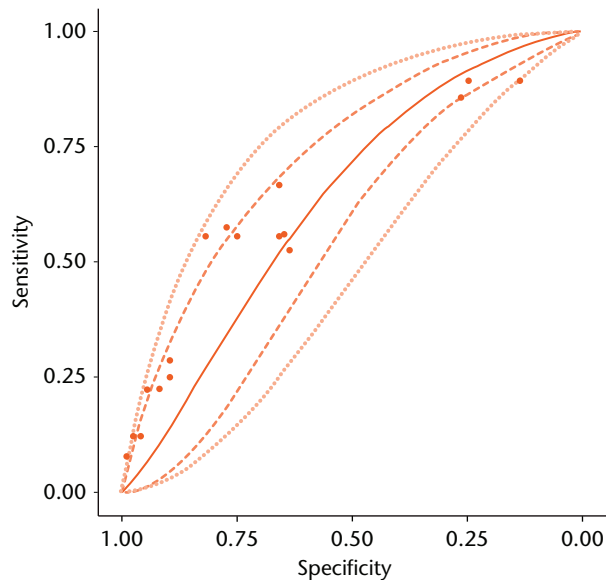
Based on a mean AUC of 0.615, blood ion levels are a poor, and not clinically useful, test for classifying patients as having or not having an ARMD.<sup>41-44</sup> It has been suggested that a clinically useful test has an  $AUC \geq 0.75$ .<sup>45</sup> Considering the reconstructive consequences of delayed diagnosis, a false negative result could harm patients.<sup>15-17</sup> With the availability of non-invasive tests which definitively determine the presence of an ARMD, we see no role for using blood ion levels to diagnose symptomatic patients.<sup>18</sup>

Screening is the process of identifying high-risk patients in the general population. Since screen-positive patients will undergo further testing, screening tests need not be as accurate as diagnostic tests. Screening can use two different approaches: exclude patients with very low probability of disease from further testing by maximising the negative predictive value (NPV), or identify high-risk patients for further testing by maximising the positive predictive value (PPV).<sup>46</sup> The performance of ion testing using these two approaches is shown in Table II. Calculations were made using the SROC curve plotted in Figure 3 and using the mean prevalence of ARMDs in the studies included in this review (41%). Indeed, maximising NPV is burdensome because 99% of patients will test positive. Furthermore, test-positive patients have the same probability of disease as they did prior to undergoing the test. On the other hand, maximising PPV does not reassure test negative patients because they still have a 21% probability of having an ARMD. Test-positive patients are hardly “high risk” because the risk of an

**Table II.** Performance of ion levels in screening

	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)	Prevalence of positive test result (%)
<b>Maximise NPV</b>	99	1	99	41	99
<b>Maximise PPV</b>	81	48	79	52	64

PPV, positive predictive value; NPV, negative predictive value

**Fig. 3**

Summary receiver operating characteristic curve for meta-analysis. Mean curve (—), 95% credible region (---), and 95% prediction region (...) are shown. Sensitivity and specificity reported by individual studies (\*).

ARMD is marginally different from the pre-test probability (52% versus 41%). Aside from statistical concerns, screening for ARMDs is problematic on theoretical grounds. The World Health Organization recommends that screening only be performed if patients will be offered treatment.<sup>47</sup> We are unaware of any evidence supporting revision on asymptomatic patients with ARMDs and thus screening would serve no clinical purpose.

We have synthesised the totality of evidence for the diagnostic value of metal ion levels for ARMDs in patients with MoM hip implants. We conclude that blood ion levels have no role in the diagnostic algorithm for ARMDs. The probability that we have incorrectly calculated the AUC to be less than 0.75 is 1.7%. Given the strength and consistency of the findings of our meta-analysis, and the improbability that the results of our meta-analysis are incorrect, further study of metal ion testing for the diagnosis of ARMDs would be an inefficient use of research resources.

A perceived limitation of our study may be that conclusions are based on a small number of studies, half of which did not use MARS MRI as the benchmark modality. We therefore carefully assessed for, and controlled for, heterogeneity. We used a powerful meta-analytic technique that allowed us to partition results into a “cut-point effect” and “accuracy effect” (see Appendix 2).

The methodological heterogeneity only manifested in heterogeneity in the cut-point effect, and not in the accuracy effect. Due to heterogeneity, our meta-analysis cannot be used to determine a useful cut-point. However, this is a moot point because the accuracy of the test is so poor. It was remarkable that these methodologically heterogeneous studies formed a smooth ROC curve (Fig. 2). Therefore, there was substantial homogeneity among these studies in the accuracy effect. This homogeneity is further reflected in the fact that the prediction intervals and confidence intervals were nearly equivalent in width (Fig. 3). In other words, our results are tantamount to those from a single study with 898 hips and 376 ARMDs.

We emphasise that our systematic review evaluated the use of blood metal ion levels for the diagnosis of ARMDs. Our findings do not apply to the investigation of the systemic consequences of metal ion exposure which are believed to occur at levels > 60 µg/L.<sup>48</sup> Further research should be directed to determining how blood ion measurements should be used to investigate cobaltism.<sup>49</sup>

We conclude that the available evidence does not support existing guidelines, which recommend the use of blood ion measurements for both screening and diagnosis of ARMD.

### Supplementary material

Appendices showing the Medline search strategy, ROC curves, study and patient characteristics, pseudotumour detection and concentration results can be found alongside the paper online at <http://www.bjr.boneandjoint.org.uk/>

### References

1. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. *Lancet* 2007;370:1508-1519.
2. Cohen D. Out of joint: the story of the ASR. *BMJ* 2011 Jan;342:d2905.
3. Cohen D. How safe are metal-on-metal hip implants? *BMJ* 2012;344:e1410
4. Huiskes R. Failed innovation in total hip replacement. Diagnosis and proposals for a cure. *Acta Orthop Scand* 1993;64:699-716.
5. Kwon Y, Lombardi A V, Jacobs JJ, et al. Risk stratification algorithm for management of patients with metal-on-metal hip arthroplasty: consensus statement of the American Association of Hip and Knee Surgeons, the American Academy of Orthopaedic Surgeons, and the Hip Society. *J Bone Joint Surg [Am]* 2014;96-A:e4.
6. Sieber HP, Rieker CB, Köttig P. Analysis of 118 second-generation metal-on-metal retrieved hip implants. *J Bone Joint Surg [Br]* 1999;81-B:46-50.
7. Beaulé PE, Mussett SA, Medley JB. Metal-on-Metal Bearings in Total Hip Arthroplasty. *Instr Course Lect* 2010;59:17-25.
8. Qu X, Huang X, Dai K. Metal-on-metal or metal-on-polyethylene for total hip arthroplasty: a meta-analysis of prospective randomized studies. *Arch Orthop Trauma Surg* 2011;131:1573-1583.

9. **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.
10. **Walsh AJ, Nikolaou VS, Antoniou J.** Inflammatory pseudotumor complicating metal-on-highly cross-linked polyethylene total hip arthroplasty. *J Arthroplasty* 2012;27:324.e5-e8.
11. **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.
12. **Willert H-G, Buchhorn GH, Fayyazi A, et al.** Metal-on-metal bearings and hypersensitivity in patients with artificial hip joints. A clinical and histomorphological study. *J Bone Joint Surg [Am]* 2005;87-A:28-36.
13. **Pandit H, Glyn-Jones S, McLardy-Smith P, et al.** Pseudotumours associated with metal-on-metal hip resurfacings. *J Bone Joint Surg [Br]* 2008;90-B:847-851.
14. **Langton DJ, Jameson SS, Joyce TJ, et al.** 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-B:38-46.
15. **Grammatopoulos G, Pandit H, Kwon Y-M, et al.** Hip resurfacings revised for inflammatory pseudotumour have a poor outcome. *J Bone Joint Surg [Br]* 2009;91-B:1019-1024.
16. **de Steiger RN, Miller LN, Prosser GH, et al.** Poor outcome of revised resurfacing hip arthroplasty. *Acta Orthop* 2010;81:72-76.
17. **Whiteside LA.** Surgical technique: transfer of the anterior portion of the gluteus maximus muscle for abductor deficiency of the hip. *Clin Orthop Relat Res* 2012;470:503-510.
18. **Garbuz DS, Hargreaves BA, Duncan CP, et al.** The John Charnley Award: diagnostic accuracy of MRI versus ultrasound for detecting pseudotumors in asymptomatic metal-on-metal THA. *Clin Orthop Relat Res* 2014;472:417-423.
19. **No authors listed.** Medical safety alert: Metal-on-metal (MoM) hip replacements - updated advice with patient follow ups; 2012. <https://www.gov.uk/drug-device-alerts/medical-device-alert-metal-on-metal-mom-hip-replacements-updated-advice-with-patient-follow-ups> (date last accessed 04 August 2016).[[bibmisc]]
20. **No authors listed.** Concerns about Metal-on-Metal Hip Implants; 2013. <http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/ImplantsandProsthetics/MetalonMetalHipImplants/ucm241604.htm> (date last accessed 04 August 2016).[[bibmisc]]
21. **No authors listed.** Final opinion on the safety of Metal-on-Metal joint replacements with a particular focus on hip implants; 2014. <https://www.tga.gov.au/metal-metal-hip-replacement-implants#recommend> (date last accessed 04 August 2016).[[bibmisc]]
22. **No authors listed.** Metal-on-Metal Hip Implants – Information for Orthopaedic Surgeons Regarding Patient Management Following Surgery – For the Public; 2012. <http://www.healthycanadians.gc.ca/recall-alert-rappel-avis/hc-sc/2012/14120a-eng.php> (date last accessed 04 August 2016).[[bibmisc]]
23. **No authors listed.** Metal-on-metal hip replacement implants; 2012. <https://www.tga.gov.au/metal-metal-hip-replacement-implants> (date last accessed 04 August 2016).[[bibmisc]]
24. **Lombardi AV, Jr, Barrack RL, Berend KR, et al.** The Hip Society: algorithmic approach to diagnosis and management of metal-on-metal arthroplasty. *J Bone Joint Surg [Br]* 2012;94-B:14-18.
25. **No authors listed.** Food and Drug Agency. FDA Safety Communication: Metal-on-Metal Hip Implants. <http://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/ucm335775.htm> (date last accessed 20 May 2016).
26. **Van Der Straeten C, Grammatopoulos G, Gill HS, et al.** The 2012 Otto Aufranc Award: the interpretation of metal ion levels in unilateral and bilateral hip resurfacing. *Clin Orthop Relat Res* 2013;471:377-385.
27. **Zhou XH, Obuchowski NA, McClish DK.** *Statistical Methods in Diagnostic Medicine*. Second ed. New York: Wiley & Sons, 2011. [[bibmisc]]
28. **Whiting PF, Rutjes AWS, Westwood ME, et al; QUADAS-2 Group.** QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-536.
29. **Chappell FM, Raab GM, Wardlaw JM.** When are summary ROC curves appropriate for diagnostic meta-analyses? *Stat Med* 2009;28:2653-2668.
30. **Rutter CM, Gatsonis CA.** A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;20:2865-2884.
31. **Dukic V, Gatsonis C.** Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds. *Biometrics* 2003;59:936-946.
32. **Higgins JPT, Thompson SG.** Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-1558.
33. **Stan Development Team.** Stan Modeling Language Users Guide and Reference Manual, Version 2.10.0. <http://mc-stan.org> (date last accessed 12 July 2016).
34. **Cowles MK, Carlin BP.** Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *J Am Stat Assoc* 1996;91:883-904.
35. **Brooks SP, Gelman A.** General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 1998;7:434-455.
36. **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.
37. **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.
38. **Macnair RD, Wynn-Jones H, Wilmhurst 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.
39. **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.
40. **Wright JG.** A practical guide to assigning levels of evidence. *J Bone Joint Surg [Am]* 2007;89-A:1128-1130.
41. **Fan J, Upadhye S, Worster A.** Understanding receiver operating characteristic (ROC) curves. *CJEM* 2006;8:19-20.
42. **Hanley JA.** Receiver operating characteristic (ROC) methodology: the state of the art. *Crit Rev Diagn Imaging* 1989;29:307-335.
43. **Hanley JA, McNeil BJ.** The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
44. **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.
45. **Fan J, Upadhye S, Worster A.** Understanding receiver operating characteristic (ROC) curves. *CJEM* 2006;8:19-20.
46. **Raffle AE, Gray JAM.** *Screening: Evidence and practice*. Oxford: Oxford University Press, 2007.
47. **Wilson JM, Jungner YG.** [Principles and practice of mass screening for disease]. *Bol Oficina Sanit Panam* 1968;65:281-393. [In Spanish]
48. **Tower SS.** Arthroprosthetic cobaltism associated with metal on metal hip implants. *BMJ* 2012;344:e430.
49. **Devlin JJ, Pomerleau AC, Brent J, et al.** Clinical features, testing, and management of patients with suspected prosthetic hip-associated cobalt toxicity: a systematic review of cases. *J Med Toxicol* 2013;9:405-415.

#### Funding Statement

- P. Beaulé declares that the authors have received funding, grants and royalties for work and consultancy unrelated to this paper from MicroPort, MEDACTA, Corin, Zimmer Biomet, Depuy, J & J, and the Canadian Institute of Health Research.
- P.R. Kim has received funding from Stryker which is unrelated to this paper.

#### Author Contribution

- M. Pahuta: Conception and design, Acquisition of data, Analysis and Interpretation of data, Drafting of manuscript, Critical revision
- J. M. Smolders: Reviewed manuscript, Acquisition of data
- J. L. van Susante: Manuscript preparation, Critical revision, Study design, Acquisition of data
- J. Peck: Data collection/analysis, Manuscript preparation, Critical revision
- P. R. Kim: Reviewed manuscript, Acquisition of data
- P. E. Beaulé: Content, Manuscript preparation, Data analysis

#### ICMJE conflict of interest

- None declared

© 2016 Pahuta et al. This is an open-access article distributed under the terms of the Creative Commons Attribution licence (CC-BY-NC), which permits unrestricted use, distribution, and reproduction in any medium, but not for commercial gain, provided the original author and source are credited.