Supplementary material

Development of NTH aseptic lymphocyte-dominated vasculitis-associated lesion (ALVAL) scoring system

Introduction

The first comprehensive descriptions of ALVAL were published in 2005. It was recognized at the time that the major new histological findings were those of a development of a heavy perivascular lymphocytic infiltrate and varying degrees of synovial surface membrane damage. These responses were different to those classically associated with metal-on-polyethylene (MoP) devices, which were typically limited to macrophagic infiltration.

The NTH grading system was developed between the years 2007 and 2010 as data were collected on an ever-increasing number of metal-on-metal (MoM) failures. Various parameters were investigated, their relationships to each other were examined, and the results were described in multiple publications. These parameters included histiocytic sheet thickness,[1] perivascular lymphocytic cuff thickness,[1,2] synovial surface membrane disruption,[1,2] particulate load,[3,4] joint fluid cobalt (Co) and chromium (Cr) concentrations,[3,4] the extent of macroscopic soft tissue and volumetric wear of the explanted prostheses.[3,5]

We found that the amount of metal exposure, whether measured in terms of volumetric wear rates or joint fluid metal ion concentrations, was not directly related to the extent of surface membrane necrosis or lymphocyte cuff thickness. However, we found a highly significant correlation between the mean diameter of perivascular lymphocytic cuffs and the extent of surface necrosis determined both micro- and macroscopically. We also identified patients with bilateral MoM prostheses where extensive soft-tissue necrosis was identified in association with relatively well functioning devices. These observations indicated that soft-tissue injury pathology was directly associated with the ALVAL process. We believed, therefore, that the extent of surface membrane necrosis should form an integral part of an ALVAL assessment.

Comparison with existing grading systems and relevant studies (Table i)[6,7]

Campbell ALVAL score. Campbell et al[6] produced a comprehensive multi-part scoring system which incorporates both the macrophage and lymphocyte response, as well as the integrity of the tissue architecture, including the extent of surface membrane disruption. We chose not to adopt this grading scale at our unit for two reasons. First, the complexity of the score was felt to be too unwieldy to use in the routine clinical environment. This concern has been borne out to some extent by a recent study.[8] Second, for research purposes, we did not feel it was appropriate to use a system which accrued points in the absence of perivascular cuffs, one of the key features of ALVAL. As described above, we have not been able to identify a relationship between macrophage infiltration and either lymphocyte infiltration or surface membrane damage.

Oxford ALVAL score. In 2013, the Oxford group[7] produced a paper which aimed to determine whether the amount of wear generated from MoM hip resurfacings correlated with the intensity of the ALVAL response. For reasons similar to ours, Oxford used a grading system which differed from the Campbell score due to concerns that the Campbell score assessed a number of histological features that are commonly seen in periprosthetic tissues with other implant biomaterials. The Oxford score essentially only differs from the NTH score in the fact that it does not include an assessment of the extent of synovial surface membrane damage. Therefore, while all tissues categorized as 'grade 3/severe ALVAL' using the Oxford score would be categorized as a minimum of grade 2/moderate using the NTH score, not all would achieve an NTH severe score because the Oxford system does not rely on the presence of extensive surface membrane necrosis.

Supplementary references are listed to show the studies in which the authors published an in-depth description of their ALVAL grading systems and the examination of the interobserver reliability scores. It also includes a description of the histological features which form part of the grading assessment (Table i).

In summary, while there are minor differences in the assessment of moderate ALVAL responses, the fundamentals of the scoring systems are similar. It is impossible for tissues which had a 'low' or 'no ALVAL' score using our NTH system to achieve a corresponding high grade or severe ALVAL score using any other grading system.

Table i. Comparison of aseptic lymphocyte-dominated vasculitis-associated lesion grading systems

	Campbell et al[6]	Oxford group[7]	NTH
Subjects	32	56	65
Interobserver kappa coefficient	0.71	0.74	0.71
Co-observer	Biomedical scientist	Consultant pathologist	Orthopaedic trainee
Macrophage infiltration	Yes	No	No
Perivascular cuffs	Yes	Yes	Yes
Surface membrane destruction	Yes	No	Yes

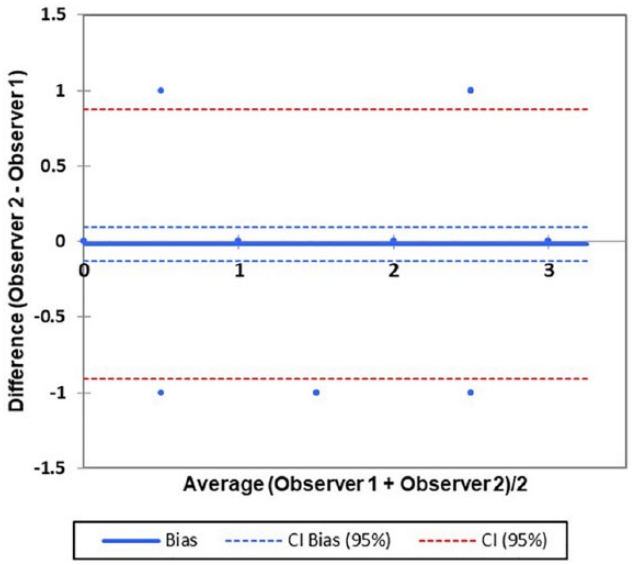

Fig. a

Bland-Altman plot of aseptic lymphocyte-dominated vasculitis-associated lesion scoring comparison (CI, confidence interval).

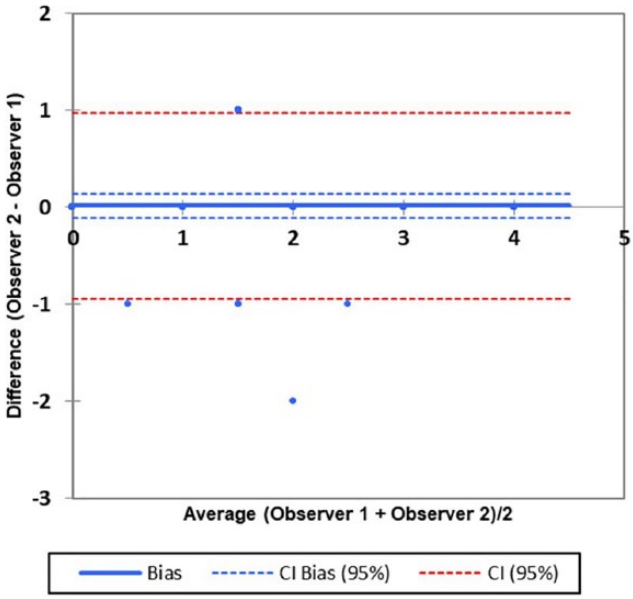

Fig. b

Bland-Altman plot of particle scoring comparison (CI, confidence interval).

Repeatability of NTH ALVAL score

To test the interobserver reliability of the NTH ALVAL grading system, the first 65 consecutive patients were reassessed independently by another observer. The second observer, previously inexperienced in histopathology, was given limited training by the pathologist who developed the grading system.

The second observer was blinded to the results and the agreement was assessed by a third investigator. Repeatability testing using Cohen's kappa[9] demonstrated a highly significant interobserver ($k = 0.71$, $p < 0.001$) correlation coefficient (k).

Cohen's kappa does not account for the extent of disagreement, however. For this reason, we plotted the results using an approach normally used for continuous variables, Bland-Altman analysis (Fig. a).

This confirmed no bias in the differences between the two observers and that there was no single occasion where the scores differed by more than one grade.

Repeatability of particulate grading

The methodology described above was used to assess the interobserver repeatability of the particulate grading scale. Cohen's kappa coefficient was found to be 0.74, again indicating substantial interobserver agreements. The Bland-Altman plot is shown in Figure b.

Processing of joint fluid samples prior to inductively coupled plasma mass spectrometry (ICPMS)

In the current study, joint fluid samples were not acid digested prior to ICPMS analysis. The advantage of this approach, we believe, is that these concentrations more accurately gauge the periprosthetic environment's exposure to free metal ions, and to smaller particulate debris which may be protein-bound (the process of ICPMS liberates metal from proteins). The disadvantage of this approach is that larger, inert particles are not measured. Acid digestion transforms all particulate debris into ionic form, allowing detection of all Co and Cr in the fluid by ICPMS. There is debate as to which method is preferable.

In order to examine the effect of acid digestion, we randomly selected 25 samples which had been measured using our normal approach, then compared these measurements with those obtained following acid digestion. As Cr tends to be found in larger particulate masses, and Co compounds are more soluble, we expected that there would be a greater change in Cr concentrations than in Co, and that Co concentrations would be relatively unchanged between measurements.

We then conducted linear regression analysis to compare the results of the two methods.

The equation for Co was log post-digestion Co concentration $= 0.315 + (0.961 \times$ Log pre-digested Co), with an R^2 value of 93% (Fig. c).

There was greater variability in the effect of digestion with respect to Cr concentrations. Simple linear regression returned a lower R^2 value (0.78), and there was an increase in measured Cr post digestion (Log Cr post digestion $= 1.215 + 0.673 \times$ Log Cr pre- digestion) (Fig. d).

Albumin concentrations in patients with adverse reaction to metal debris (ARMD)

To determine whether the fluid effusions seen in association with ALVAL contain greater concentrations of

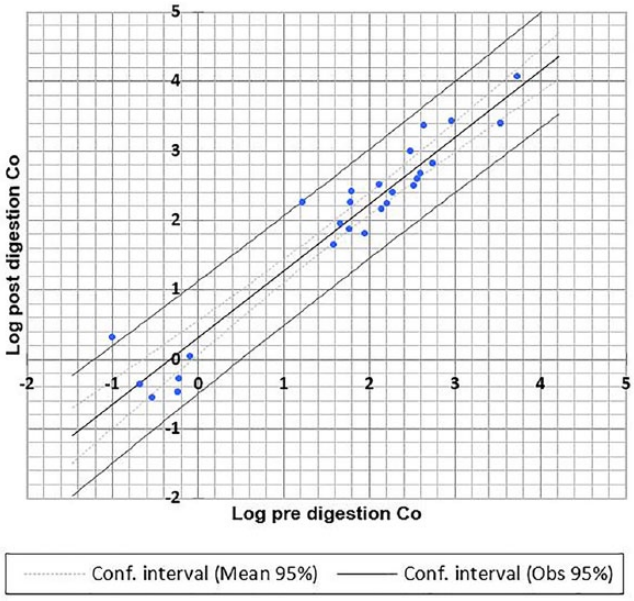

Fig. c

Cobalt (Co) concentrations in synovial fluid samples pre- and post-acid digestion. Values have been log normalized (Conf., confidence; Obs, observation).

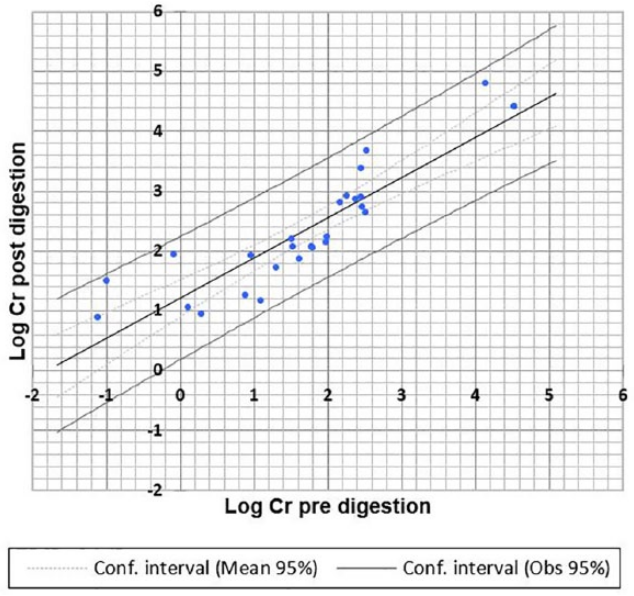

Fig. d

Chromium (Cr) concentrations pre- and post-acid digestion. Values have been log normalized (Conf., confidence; Obs, observation).

albumin than are reported in the literature, we analyzed 50 stored consecutive hip aspirate samples from patients with ARMD[10] and compared them with albumin concentrations measured using the same methodology in ten consecutive patients revised for aseptic loosening of non-MoM devices. The range of albumin concentrations in ARMD patients was 10.8 g/l to 104.2 g/l, with a median value of 27.2 g/l. The median albumin concentration in the non-MoM patients was 18.4 g/l which compared well

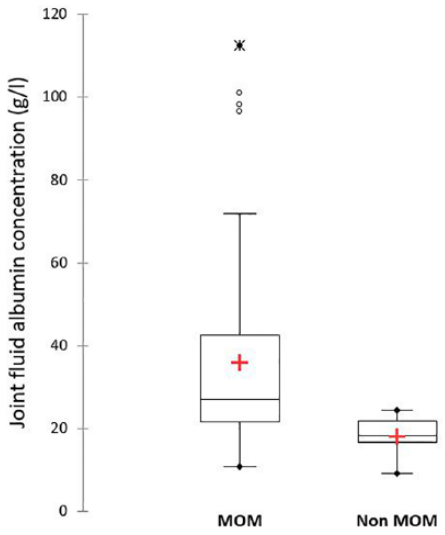

Fig. e

Albumin concentrations in metal-on-metal (MoM) and non-MoM synovial fluid samples.

with the reported mean (SD) concentration in 19 patients with osteoarthritis of the knee, which was 18 g/l (SD 0.08) (Fig. e).[10]

Fluid metal ion concentrations in context with weight of material loss

Basic calculations were performed to put into context the metal ion concentrations reported in the manuscript. With respect to the normal composition of a MoM hip (American Society for testing Materials (ASTM) classifications and our own testing), contemporary cobalt-chromium-molybdenum alloys are composed of approximately 65% Co and 30% Cr by weight.

We performed a series of theoretical calculations. Theoretical volume losses from implants were converted into weights of material loss assuming a density of 8.3 g/cm^3. These values were then converted to daily rates of material loss, i.e. 100 g of total material loss in an implant in service for 100 days would equate to a daily weight loss of 1 g per day.

For the purposes of these calculations, we assumed no metal ions were present in the joint fluid before metal exposure and that no metal ions were cleared from the fluid in a day. The resulting concentrations of Co and Cr metal ions that had accumulated in the joint at a single moment in time were then calculated for three different scenarios:

- theoretical wear rates with 5 ml of joint fluid, i.e. the ‘expected’ amount of metal released from the components;
- theoretical wear rates with 20 ml of fluid, i.e. the ‘expected’ amount of metal released from the components;

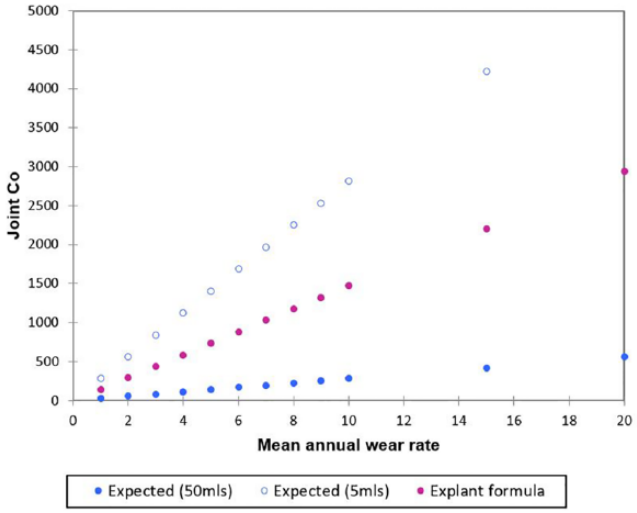

Fig. f

Comparison of synovial fluid cobalt (Co) concentrations as derived from explant data regression analysis with theoretical concentrations calculated as described.

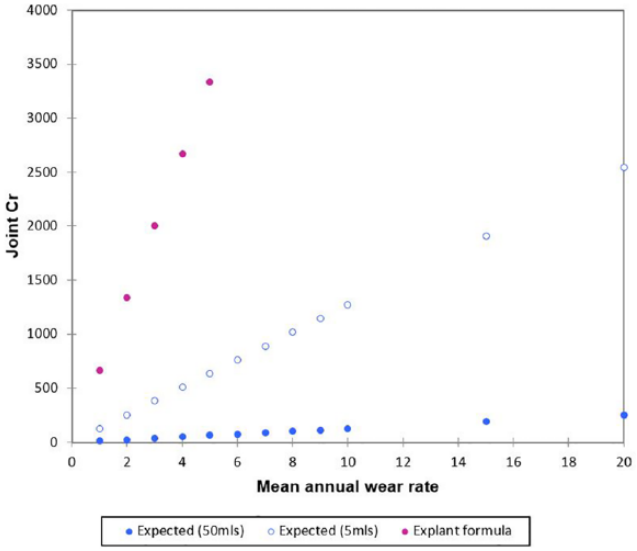

Fig. g

Comparison of synovial fluid chromium (Cr) concentrations as derived from explant data regression analysis with theoretical concentrations calculated as described above.

- actual Co and Cr values calculated from the regression analyses described in the paper (no aseptic lymphocyte-dominated vasculitis-associated lesion).

The results are plotted Figure f. One can see that the rate of clearance of Co is comparable with the theoretical calculations, i.e. Co appears to be cleared rapidly from the joint.

For Cr, however, the values obtained from patients are far greater than would be predicted from daily clearance of metal load (Fig. g).

References

1. **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.
2. **Langton DJ, Joyce TJ, Mangat N, et al.** Reducing metal ion release following hip resurfacing arthroplasty. *Orthop Clin North Am* 2011;42:169-180.
3. **Langton DJ, Joyce TJ, Jameson SS, et al.** Adverse reaction to metal debris following hip resurfacing: the influence of component type, orientation and volumetric wear. *J Bone Joint Surg [Br]* 2011;93-B:164-171.
4. **Sidaginamale RP, Joyce TJ, Bowsher JG, et al.** The clinical implications of metal debris release from the taper junctions and bearing surfaces of metal-on-metal hip arthroplasty: joint fluid and blood metal ion concentrations. *Bone Joint J* 2016;98:925-933.
5. **Langton DJ, Sidaginamale RP, Avery P, et al.** Retrospective cohort study of the performance of the Pinnacle metal on metal (MoM) total hip replacement: a single-centre investigation in combination with the findings of a national retrieval centre. *BMJ Open* 2016;6:e007847.
6. **Campbell P, Ebramzadeh E, Nelson S, et al.** Histological features of pseudotumor-like tissues from metal-on-metal hips. *Clin Orthop Relat Res* 2010;468:2321-2327.
7. **Grammatopoulos G, Pandit H, Kamali A, et al.** The correlation of wear with histological features after failed hip resurfacing arthroplasty. *J Bone Joint Surg [Am]* 2013;95:e81.
8. **Smeekes C, Cleven AHG, van der Wal BCH, et al.** Current Pathologic Scoring Systems for Metal-on-metal THA Revisions are not Reproducible. *Clin Orthop Relat Res* 2017;475:3005-3011.
9. **Watson PF, Petrie A.** Method agreement analysis: a review of correct methodology. *Theriogenology* 2010;73:1167-1179.
10. **Miyaguchi M, Kobayashi A, Kadoya Y, et al.** Biochemical change in joint fluid after isometric quadriceps exercise for patients with osteoarthritis of the knee. *Osteoarthritis Cartilage* 2003;11:252-259.