Clinical Orthopaedics and Related Research®
A Publication of The Association of Bone and Joint Surgeons®

CrossMark

CORR Insights

CORR Insights®: Current Pathologic Scoring Systems for Metal-on-metal THA Revisions are not Reproducible

Thomas W. Bauer MD, PhD

Where Are We Now?

Some patients with metal-on-metal (MoM) hip implants, metal-on-polyethylene implants, or other hip arthroplasty constructs develop adverse local tissue reactions (ALTRs). Sometimes those reactions are associated with wear debris particles, macrophages, and osteolysis, but only rare lymphocytes. Other ALTRs contain very few visible particles, but extensive lymphoplasmacytic inflammation and necrosis. Soft-tissue masses and/or effusions may occur, while other times the peri-implant membrane is more linear. Occasionally, these changes are associated with elevated serum metal ion levels. Despite more than a decade of investigations, radiologists, orthopaedic surgeons, pathologists, and biomaterials experts have not yet reached a consensus about the importance of, or even how to describe these reactions. These disagreements are rooted, in part, by a lack of correlation among our disciplines.

Recognizing that tissues around damaged implants show a spectrum of changes, and that any given arthroplasty may show features reflecting more than one mechanism of failure, several groups of researchers have developed grading systems for individual observations [2, 3, 5–7], or combinations of features [1, 4], to semiquantitatively grade the extent to which the morphologic findings in tissue might reflect an adaptive immune response versus infection, mechanical factors, or an innate inflammatory reaction to debris.

In the current study by Smeekes and colleagues, three pathologists tested the reproducibility of two commonly used scoring systems [1, 4], the aseptic lymphocyte vasculitis-associated lesion (ALVAL) score and the modified Oxford ALVAL score. The results provide documentation of what many pathologists have maintained for years: These two scoring systems lack the level of reproducibility that most physicians expect from a routine laboratory test. That does not mean the scoring systems are of no value, but it does illustrate that semiquantitative grading of this type can be difficult and we have room for improvement.

T. W. Bauer MD, PhD (✉)
Department of Pathology and Laboratory Medicine, Hospital for Special Surgery, 535 East 70th St., New York, NY 10021, USA
e-mail: bauerT@hss.edu

CORR Insights

Where Do We Need To Go?

Smeekes and colleagues suggest that a simplified scoring system is needed. While that may be true, they provided no evidence that a more simplified (or for that matter a more complex) system would yield higher interobserver correlation. Future studies should make sure that the involved pathologists concur on how to interpret the various components of any scoring system being evaluated, and they should review a "learning set" of cases before starting the study. Steps like these may increase concordance of pathologists' assessments [1]. From the perspective of a statistician, such advanced preparation should not be needed; that is, a scoring system should stand by itself, but a few paragraphs of descriptive text are unlikely to maximize concordance as effectively as real-time discussion among pathologists over a microscope slide or digital image. Additionally, the Intraclass Correlation Coefficient (ICC) used in the current study is often used for continuous variables, but the components of the Campbell and Oxford score are hardly continuous, and one wonders whether simple measures of agreement might be more effective. And, like misuses of the p value, overreliance on a high ICC could mask observations that may still be clinically meaningful. Further, it is well-

recognized that ALTRs are not uniformly distributed throughout the peri-implant tissue. Similar to grading malignant tumors, most pathologists intentionally select the most extreme, or at least the "most representative" areas of tissue to grade. In the samples of tissue evaluated in this study, the surface and adjacent millimeter or two of the peri-prosthetic membrane would likely be the most useful region of interest, and the Oxford grading system specifically notes that the "score was based on the maximum perivascular lymphoid infiltrate noted in any one specimen" [4], a sampling method not used by the current study authors.

Beyond noting less than ideal correlations among pathologists for selected observations, what we need are correlations among the observations themselves (such as the extent of necrosis or lymphoplasmacytic inflammation), and clinical variables such as imaging findings, a pseudotumor, duration since primary arthroplasty, or the results of revision arthroplasty. Testing those correlations could be of clinical value, even if the correlation coefficients of morphologic grading are suboptimal. Ultimately, one hopes that dissecting the biology of complex adverse tissue reactions will help improve patient selection, implant design, and treatment methodologies resulting in better clinical results and fewer revisions.

How Do We Get There?

First, it is important to recognize that there are different types of ALTRs, and that the morphologic features of those reactions are likely to reflect, to a variable extent, factors related to the host and to the arthroplasty that have led to revision. It is also important to understand that (1) not all clinically unsatisfactory MoM constructs have failed because of an adaptive immune response, (2) not all unsatisfactory metal- or ceramic-on-polyethylene hips have failed due to a macrophage reaction to polyethylene debris, and (3) the extent of ALTRs prevalent around clinically satisfactory implants is unknown. The different patterns of inflammation related to different failure mechanisms can usually be recognized qualitatively, and it may be misleading to infer clinical importance to a semiquantitative scoring system that has been developed for one type of construct if applied to tissue around an arthroplasty of different design and different dominant failure mechanism. Instead, we need prospective studies in which multiple individual morphologic features are correlated with comprehensive information, including clinical findings, serum ion levels, the results of various imaging studies, implant composition and design, intraoperative observations, evaluation of retrieved devices, and the clinical results after

CORR Insights

revision arthroplasty. Finally, a uniform vocabulary needs to be developed, so that surgeons, radiologists, pathologists, and biomechanical engineers use terms like "metallosis", "ALVAL", "adaptive immune response", "osteolysis", "pseudotumor", "polymer reaction", "vasculitis", "lymphoid aggregate", "germinal center", "necrosis", "apoptosis", and "corrosion products", in a uniform way.

References

1. Campbell P, Ebramzadeh E, Nelson S, Takamura K, De Smet K, Amstutz HC. Histological features of pseudotumor-like tissues from metal-on-metal hips. *Clin Orthop Relat Res.* 2010;468:2321–2327.
2. 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;87:18–27.
3. 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;469:1127–1133.
4. Grammatopoulos G, Pandit H, Kamali A, Maggiani F, Glyn-Jones S, Gill HS, Murray DW, Athanasou N. The correlation of wear with histological features after failed hip resurfacing arthroplasty. *J Bone Joint Surg Am.* 2013;95:e81.
5. Natu S, Sidaginamale RP, Gandhi J, Langton DJ, Nargol AV. Adverse reactions to metal debris: Histopathological features of periprosthetic soft tissue reactions seen in association with failed metal on metal hip arthroplasties. *J Clin Pathol.* 2012;65:409–418.
6. Perino G, Ricciardi BF, Jerabek SA, Martignoni G, Wilner G, Maass D, Goldring SR, Purdue PE. Implant based differences in adverse local tissue reaction in failed total hip arthroplasties: A morphological and immunohistochemical study. *BMC Clin Pathol.* 2014;14:39.
7. Willert HG, Buchhorn GH, Fayyazi A, Flury R, Windler M, Koster G, Lohmann CH. Metal-on-metal bearings and hypersensitivity in patients with artificial hip joints. A clinical and histomorphological study. *J Bone Joint Surg Am.* 2005;87:28–36.