Measuring and reporting of vertebral endplate bone marrow lesions as seen on MRI (Modic changes): Recommendations from the ISSLS Degenerative Spinal Phenotypes Group

Aaron J Fields  
*University of California, San Francisco*

Michele C Battié  
*University of Western Ontario*

Richard J Herzog  
*Hospital for Special Surgery*

Jeffrey G Jarvik  
*University of Washington*

Roland Krug  
*University of California, San Francisco*

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

Please let us know how this document benefits you.

**Recommended Citation**

Fields, Aaron J; Battié, Michele C; Herzog, Richard J; Jarvik, Jeffrey G; Krug, Roland; Link, Thomas M; Lotz, Jeffrey C; O’Neill, Conor W; Sharma, Aseem; and ISSLS Degenerative Spinal Phenotypes Group, "Measuring and reporting of vertebral endplate bone marrow lesions as seen on MRI (Modic changes): Recommendations from the ISSLS Degenerative Spinal Phenotypes Group," *European spine journal*. 28, 10. 2266 - 2274. (2019).  
[https://digitalcommons.wustl.edu/open_access_pubs/11887](https://digitalcommons.wustl.edu/open_access_pubs/11887)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Measuring and reporting of vertebral endplate bone marrow lesions as seen on MRI (Modic changes): recommendations from the ISSLS Degenerative Spinal Phenotypes Group

Aaron J. Fields1 · Michele C. Battie2 · Richard J. Herzog3 · Jeffrey G. Jarvik4 · Roland Krug5 · Thomas M. Link5 · Jeffrey C. Lotz1 · Conor W. O’Neill1 · Aseem Sharma6 · for the ISSLS Degenerative Spinal Phenotypes Group

Abstract

Purpose The positive association between low back pain and MRI evidence of vertebral endplate bone marrow lesions, often called Modic changes (MC), offers the exciting prospect of diagnosing a specific phenotype of chronic low back pain (LBP). However, imprecision in the reporting of MC has introduced substantial challenges, as variations in both imaging equipment and scanning parameters can impact conspicuity of MC. This review discusses key methodological factors that impact MC classification and recommends guidelines for more consistent MC reporting that will allow for better integration of research into this LBP phenotype.

Methods Non-systematic literature review.

Results The high diagnostic specificity of MC classification for a painful level contributes to the significant association observed between MC and LBP, whereas low and variable sensitivity underlies the between- and within-study variability in observed associations. Poor sensitivity may be owing to the presence of other pain generators, to the limited MRI resolution, and the imperfect reliability of MC classification.

Conclusions Comparison of MC data between studies can be problematic. Various methodological factors impact detection and classification of MC, and the lack of reporting guidelines hinders interpretation and comparison of findings. Thus, it is critical to adopt imaging and reporting standards that codify acceptable methodological criteria.

Graphic abstract

These slides can be retrieved under Electronic Supplementary Material.

Keywords Low back pain · Modic changes · Bone marrow lesion · Endplate damage · Magnetic resonance imaging

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00586-019-06119-6) contains supplementary material, which is available to authorized users.

Extended author information available on the last page of the article
**What are Modic changes and why are they important?**

Modic et al. [1] and de Roos et al. [2] separately observed that the bone marrow near damaged endplates in LBP patients had a distinct appearance when viewed with MRI. Compared to the adjacent normal bone marrow, the endplate lesions with active inflammation and fibrovascular replacement of the hematopoietic marrow appeared hypointense on T1-weighted images and hyperintense on T2-weighted images (type 1 changes; MC1); the endplate lesions with fatty replacement of the marrow appeared hyperintense on T1-weighted images and isointense to slightly hyperintense on conventional T2-weighted images (type 2 changes; MC2). A third type of endplate lesion, characterized by sclerotic subchondral bone, appeared hypointense on both T1- and T2-weighted images (type 3 changes; MC3) [3]. The etiology, risk factors, and management of Modic changes are the focus of a large body of work that has been reviewed elsewhere [4–8]. Briefly, it appears Modic changes may be either transient or permanent stages of a chronic pathologic process [8–11]; they associate with age and disk degeneration [5, 12]; and they predominate in the lower lumbar spine (L4–S1) [12]. Perhaps most importantly, systematic reviews suggest a positive association between LBP and Modic changes [5, 6].

Yet, the clinical relevance of Modic changes varies widely between studies [4, 5, 13], and controversies remain. Such inconsistencies may be due in part to misclassification and to imprecision in reporting of Modic changes. While some investigators follow standardized imaging and grading procedures, many utilize equipment, imaging sequences, or reporting methods that can inadvertently misrepresent Modic changes. Also, the frequent omission of basic methodological details makes study comparisons difficult. As a result, discrepancies between study findings may reflect differences in imaging procedures, particularly between older and newer imaging technologies. This issue is becoming more apparent and more important as clinical interest in Modic changes grows. Thus, we recommend that investigators reporting Modic change data follow simple reporting guidelines (Box 1). These guidelines are not meant to constrain investigators to one particular methodology; instead, they are meant to ensure that the methodological details that impact the appearance and detection of Modic changes are clearly reported.

**Identifying Modic changes: the high cost of misclassification**

Misclassification prevents a clear understanding of the clinical significance of Modic changes and is a potentially important source of discrepancy between studies. The conventional MRI assessment of Modic changes amounts to a binary classification test. Owing to the difficulty of measuring test performance against histopathologic findings as ground truth, prior studies compared Modic classification to clinical pain diagnoses from provocative discography. Although the usefulness and safety of provocative discography are controversial [14, 15], the accuracy of provocative discography can be quite high (specificity of 0.94 and false-positive rate of 6%) if performed using a low-pressure technique [16]. A review of the performance of Modic classification shows that high specificity may underlie the significant associations between Modic changes and LBP. Specifically, among six studies that reported the diagnostic performance of Modic classification for identifying a painful disk concordant with a positive discogram, the specificity was over 95% [17–21] in all but one [22] (Table 1). In other words, among patients with chronic LBP receiving discography, observation of a Modic change has a low false-positive rate at a given spinal level. The specificity may be higher for MC1 than for MC2 or MC3, although it is difficult to compare lesion types since some lesions show mixed elements of both MC1 and MC2 or MC 2 and MC 3 [23].

### Box 1  Suggested guidelines for presentation of Modic change data

| MR unit: | Identify the MRI instrument and magnetic field strength, plus any surface coils or specialized tables used to collect and amplify signal. Include model number and manufacturer. For longitudinal or multi-center studies, use scanners with the same field strength. |
| Sequences: | Specify the T1- and T2-weighted sequence parameters, including the type of spin echo (i.e., fast vs. conventional), repetition time/echo time, field of view, matrix size, slice thickness/spacing, number of echoes, and the type of fat suppression that was applied (T2 only). If fat suppression was used, the Modic classification for type 2 changes should be defined as having hyperintense signal on T1-weighted images and hypointense on fat-saturated T2-weighted images. If additional sequences, e.g., STIR/Dixon, will be used for classifying Modic changes, these sequences should be reported in addition to the T1- and T2-weighted sequences. |
| Image evaluation: | Describe which image slices were rated and which levels were evaluated. |
| Rater agreement: | Report the inter-rater and intra-rater kappa statistics for categorical Modic classification. Also, report the inter-rater and intra-rater intra-class correlation coefficients (ICC) if using quantitative measurements, e.g., lesion size, cerebrospinal fluid-normalized intensity, etc. |
Conversely, the low and variable sensitivity of Modic classification for detecting discography–concordant pain may contribute to the weak associations with symptoms and to inter-study variability. For example, among the same six studies, the sensitivity of Modic classification for identifying a painful disk was less than 50%; moreover, it was highly variable between studies ranging from 14 to 48% (Table 1). Thus, the absence of a Modic change is not sufficient for ruling out pain at a given spinal level.

The low and variable sensitivity can potentially reflect the existence of other pain generators. It may also be that poor sensitivity of Modic changes is related to limited spatial resolution, signal-to-noise, or lack of fat saturation in the MRI examination. For example, one study reported that only 11% of fibrovascular marrow lesions and 62% of fatty marrow lesions identified on histologic sections were visible with a conventional imaging protocol [24]. As we discuss in the next section, this finding underscores the importance of reporting technical imaging parameters when interpreting the reported associations between Modic changes and LBP (Box 1).

Finally, the imperfect reliability of Modic classification directly impacts diagnostic sensitivity and thereby contributes to inter-study variability. Jones et al. [25] reported good inter-reader reliability ($\kappa = 0.85$) for five raters, with intra-reader reliability for the individual raters ranging from $\kappa = 0.71–1.00$. Other studies reported modest-to-good agreement between raters ($\kappa = 0.64–0.85$ [9, 12, 26–28]; Table 2). The impact of rater agreement and its variation between studies highlights the need to measure and report inter-rater values (Box 1).

To emphasize the impact of imperfect reliability, consider the hypothetical effects of false-negative classifications on the relationship between Modic changes and LBP (Fig. 1). In three scenarios with different sample sizes, rater disagreements resulting in false-negative classifications were used to calculate inter-rater reliabilities and odds ratios (ORs). For a given sample size, ORs were highly sensitive to inter-rater reliability, especially for inter-rater agreements below $\kappa = 0.80$. Moreover, for the average inter-rater agreement of the six studies reported in Table 2 ($\kappa = 0.788$), associations between Modic changes and LBP ranged from a mean OR = 3.83 (CI 0.43–34.85, not significant) to a mean OR = 9.99 (CI 2.14–52.16, p < 0.001). Together these data suggest that imperfect reliability of Modic classification could be source of discrepancies between studies. Reliability is expected to improve by implementation of standardized imaging protocols and adoption of quantitative classification schemes that are based on a continuous rather than categorical measurement scale, thereby increasing sensitivity and bringing much-needed clarity to the complex relationship between endplate bone marrow lesions and LBP.

### The importance of MRI field strength, pulse sequence, and fat suppression

The ability of MRI to interrogate the endplate bone marrow is limited by instrument precision. MRI precision is influenced by field strength, acquisition matrix, and pulse

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Levels</th>
<th>Age (years)</th>
<th>Field strength</th>
<th>Specificity (%)</th>
<th>Sensitivity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Braithwaite [17]</td>
<td>1998</td>
<td>290</td>
<td>42</td>
<td>0.5T, 1.5T</td>
<td>97</td>
<td>23</td>
</tr>
<tr>
<td>Ito [18]</td>
<td>1998</td>
<td>101</td>
<td>37</td>
<td>1.5T</td>
<td>95</td>
<td>22</td>
</tr>
<tr>
<td>Kokkonen [22]</td>
<td>2002</td>
<td>103</td>
<td>40</td>
<td>NR</td>
<td>64</td>
<td>41</td>
</tr>
<tr>
<td>Thompson [20]</td>
<td>2009</td>
<td>2457</td>
<td>43</td>
<td>NR</td>
<td>95</td>
<td>26</td>
</tr>
<tr>
<td>Weishaupt [21]</td>
<td>2001</td>
<td>116</td>
<td>42</td>
<td>1.0T</td>
<td>96</td>
<td>48</td>
</tr>
</tbody>
</table>

Table 1 Summary of diagnostic performance of Modic classification for discography–concordant pain

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Subjects</th>
<th>Field strength</th>
<th>$\kappa$, inter-rater</th>
<th>$\kappa$, intra-rater</th>
<th>Raters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones [25]</td>
<td>2005</td>
<td>50</td>
<td>NR</td>
<td>0.85</td>
<td>0.71–1.00</td>
<td>5</td>
</tr>
<tr>
<td>Chung [26]</td>
<td>2004</td>
<td>59</td>
<td>0.5T–1.5T</td>
<td>0.83</td>
<td>NR</td>
<td>2</td>
</tr>
<tr>
<td>Karchevsky [12]</td>
<td>2005</td>
<td>100</td>
<td>1.5T</td>
<td>0.81</td>
<td>NR</td>
<td>3</td>
</tr>
<tr>
<td>Kuismia [9]</td>
<td>2006</td>
<td>60</td>
<td>1.5T</td>
<td>0.64</td>
<td>0.90</td>
<td>2</td>
</tr>
<tr>
<td>Peterson [27]</td>
<td>2007</td>
<td>51</td>
<td>0.6T</td>
<td>0.81</td>
<td>0.70–0.86</td>
<td>2</td>
</tr>
<tr>
<td>Wang [28]</td>
<td>2011</td>
<td>83</td>
<td>1.5T</td>
<td>0.79</td>
<td>0.88</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2 Summary of the inter- and intra-rater reliability of Modic classification

$NR$ Not reported
sequence parameters because these factors directly affect image resolution, signal-to-noise, and contrast-to-noise ratios. In addition, field strength and fat saturation (for T2) also alter the T1 and T2 relaxation times and tissue visualization. All of these effects impact the appearance of water and fat signals, which are used to distinguish MC1 from MC2. For example, when magnets with field strengths less than 1.0T are used to image the vertebral bone marrow, they tend to greatly mitigate chemical shift, susceptibility artifacts, and flow artifacts compared to 1.5T magnets. This improves the clarity of MC1, but makes it more difficult to distinguish MC2. Conversely, the 1.5T magnets make it more difficult to identify MC1 because marrow inhomogeneities are more pronounced compared to magnets with field strengths less than 1.0T [29]. Magnets with field strengths of 3.0T or higher may have better conventional fat saturation than 1.5T magnets [30]. This reduces the tendency to overlook the marrow edema that occurs with MC1, although it is unclear whether this results in a clinically important difference. Variable field strengths also cause discrepancies in the observed frequency of different types of Modic changes. For example, whereas MC1 were 3–4 times more prevalent on 0.3T scanners, MC2 were identified twice as often with 1.5T scanners [29]. Generally speaking, the higher spatial resolution and greater dynamic range and signal- and contrast-to-noise ratios of 1.5T and 3.0T magnets, as compared to weaker magnets, make it easier to identify Modic changes. Thus, magnetic field strength should be reported in studies of Modic changes (Box 1).

Pulse sequence parameters also play a key role because they can dramatically alter the marrow signal. On T1-weighted images, fat is bright and fluid is dark. This makes fat conspicuous and helps identify boundaries between bone marrow lesions and normal marrow. Also, the short echo time of T1-weighted images provides a high intrinsic signal-to-noise ratio, which enhances anatomic detail. On T2-weighted images, fluid appears bright and fat is variable, depending on whether the images are acquired as conventional spin echo (moderate-to-high fat signal) vs. fast spin echo (high fat signal). It is important to recognize that the original studies of Modic et al. [1] and de Roos et al. [2] did not use fat-saturated T2-weighted sequences; hence, MC1 and MC2 both were defined as appearing hyperintense on T2-weighted MRI. However, many investigators now use fat-saturated T2-weighted sequences, which cause MC1 to appear hyperintense and MC2 to appear hypointense (Fig. 2). Spectral fat suppression has the added benefit of increasing the dynamic range, and if combined with modest echo times of conventional T2-weighted sequences (60–80 ms), it can do so while preserving anatomic detail. The efficiency of fat suppression for Modic classification depends on the field homogeneity, which is better in the center of the magnet bore and improves with higher-order shimming. Spectral fat suppression can only be effectively used at field strengths greater than 1.0T. A number of fat suppression sequences are often used to supplement T1- and T2-weighted sequences, including chemical shift imaging sequences (such as Dixon) and short T1 inversion recovery (STIR) sequences. STIR sequences can be used at low field strengths, and they provide more uniform fat saturation than fast spin-echo sequences. Depending on the type of spectral fat suppression, STIR sequences can be less susceptible to magnetic field inhomogeneities. Authors should report the type of fat suppression used (Box 1).

**Advances in the classification and detection of endplate bone marrow lesions**

Recent advances in image analysis techniques may improve the reliability of Modic classification and provide quantitative methodologies that are needed to evaluate treatments.
These advanced techniques are exploratory, with small studies performed mostly at single imaging centers. One strategy involves contouring the shape of the Modic change, which allows classification to be based on continuous measurements of size rather than on “present or absent” categorizations (Fig. 3). For example, Wang et al. [28] reported outstanding intra- and inter-rater reliability values for three quantitative indices measured from manual contours of mid-sagittal slices: affected/unaffected vertebral area ratio (intra-rater: 0.96; inter-rater: 0.81), cerebrospinal fluid-adjusted mean signal intensity of the Modic change (intra-rater: 0.99; inter-rater: 0.92), and total signal intensity of the Modic change (intra-rater: 0.96; inter-rater: 0.92). Semi-automated contouring approaches [31] improve measurement efficiency and may further enhance the reliability of indices based on Modic change size and intensity.

A second strategy for quantitative and objective classification of endplate bone marrow lesions is based on the assessment of bone marrow composition rather than on lesion size or structure. Measuring bone marrow composition may be especially advantageous for evaluating lesion progression and for monitoring the response to treatments, where biochemical changes in the marrow compartment are likely to precede any visible structural changes. MR imaging based on chemical shift encoding-based water–fat imaging enables the spatially resolved assessment of bone marrow fat at trabecular sites with heterogeneous red marrow distribution [32]. For example, multi-echo gradient echo acquisitions with echo time steps that result in a water–fat phase difference different from 0 and $2\pi$ give robust water–fat separation, and accurate fat quantification is possible with techniques that incorporate a precalibrated multi-peak fat spectrum in the signal model [33]. These methods may be useful for quantifying the extent of marrow edema and fatty replacement that coincide with MC1 and MC2 (Fig. 3). Compared to single-voxel spectroscopy, water–fat MRI sequences also facilitate measuring the spatial heterogeneity in marrow content, which could be useful for classifying lesions that exhibit both edematous and fatty changes. Currently, classification of these “mixed-type” Modic changes [34, 35] with T1- and T2-weighted sequences alone is highly subjective.

A third strategy involves diffusion-weighted imaging, which may help differentiate patients that have endplate bone marrow lesions with degenerative versus infectious etiologies. Using diffusion-weighted MRI, Patel et al. [36] found that patients with well-marginated, linear regions of high signal intensity situated within adjacent vertebrae at the interface between normal and abnormal bone marrow were predominantly infection free; conversely, the absence of this “claw sign” was associated with discitis/osteomyelitis. Those authors hypothesized that a gradual, progressive degenerative process results in a well-defined border between the normal and affected bone marrow, although this remains unconfirmed. Apparent diffusion coefficient (ADC) maps remove the T2 shine-through of diffusion-weighted imaging and thereby provide quantifiable signal that is directly proportional to the diffusivity of water inside the tissue. ADC values may accurately distinguish between infectious spondylitis and MC1 [37, 38]. One limitation with this approach is the sensitivity of image quality and ADC values to chosen strength and timing of the gradients ($b$-values) [38].

A complementary strategy for identifying spinal levels with endplate bone marrow lesions involves using pulse sequences that enhance visualization of endplate damage. Damage to the cartilage endplate and subchondral bone is believed to be an important factor in the etiology of

**Fig. 2**  
**a** Type 2 Modic changes seen on sagittal T1- and T2-weighted images of a fresh cadaveric lumbar spine. Use of fat suppression reverses the hyperintense signal on the T2-weighted images. Images were acquired at 3.0T with a fast spin-echo sagittal T2 sequence (repetition time msec/echo time msec 4282/85; 27 cm field of view; 3 mm slice thickness) and a sagittal T1 sequence (556/14, 27 cm field of view, 3 mm slice thickness). **b** Matching sagittal histologic section of L5-S1 level indicating endplate bone marrow lesion with fatty replacement of the hematopoietic elements. Heidenhain tri-chrome stain.
endplate bone marrow lesions because damage promotes cross talk between inflammatory factors expressed in the disk and the quiescent bone marrow [8, 39]. However, conventional T2-weighted sequences used in the spine (echo time = 60–80 ms) are unable to show the cartilage endplate because the cartilage has short T2 values, and thus, its signal is not captured in sequences with long echo times. Newer sequences such as ultra-short time-to-echo [40–42] (UTE) can overcome this limitation. For example, Law et al. [42] first demonstrated the feasibility of assessing cartilage endplate integrity using UTE MRI, which simultaneously improves visualization of endplate morphology (Fig. 3). In addition to assessing the morphology of the cartilage endplate, UTE also has the capability of assessing its biochemical composition [41], which could potentially be used to assess early degeneration. In the future, quantitative measurements of cartilage endplate degeneration and damage may provide a more comprehensive evaluation of endplate bone marrow lesions [43].

Summary and recommendations

Comparison of Modic change data between studies can be problematic. Depending on various technical factors such as imaging sequence and magnetic field strength, MC1 can be detected at greatly varying frequencies and MC2 can appear hypo- rather than hyperintense. These variations may result in inconsistencies in the apparent relationship between Modic changes and LBP. Problems with comparability can be even greater in longitudinal studies where the comparison is between images acquired at baseline with older MR units that have lower signal-to-noise ratios and images acquired at follow-up with newer MR units with improved signal-to-noise ratios and fat-saturated sequences. Even when identical sequences are used with the same imager at all time points, the subjective and categorical nature of Modic classification limits inter-rater and intra-rater reliability. All of these factors affect the reported associations between Modic changes and LBP and may underlie the wide variability between studies.

Overall, as research and clinical interest in Modic changes increases, and as stronger magnets and newer sequences gradually replace older ones, it will be critical to appreciate how technological advancements can influence reported clinical associations and comparability of study results. It will also be necessary to ensure the methodological details that accompany Modic change data are sufficiently documented to understand which technologies and techniques were used for image acquisition and analysis. This technical information is essential for consistent interpretation of Modic change data. Therefore, it is critical to adopt imaging and reporting standards that codify acceptable methodological information that is necessary to accompany Modic change data (Box 1).
Finally, while qualitative Modic classification with T1 and T2 sequences is currently the norm, novel quantitative methods show potential for assessing the severity of changes in marrow composition and for characterizing endplate structure and function in a more objective and operator- and scanner-independent manner. These developments may form the basis for more accurate future classification systems.

Acknowledgements The Degenerative Spinal Phenotypes Group of the International Society for the Study of the Lumbar Spine was co-chaired by Dino Samartzis, Joseph Assheur, Bradley Weiner, and Michele Battì, with the goal of investigating and promoting common concepts, nomenclature, definitions, and core measures for degenerative spinal phenotypes to advance knowledge in the field. The imaging data accompanying this commentary were acquired with support from the National Institutes of Health, Grants R01AR070198 (AJF), R01AR063705 (JCL), and P30AR066262 (JCL). The authors acknowledge Dave Glidden, PhD, for providing biostatistics consultation.

Compliance with Ethical Standards
Conflict of interest Dr. Fields reports grants from NIH/NIAMS during the conduct of the study; consultation for Relevent MedSystems outside of the current work. Dr. Jarvik reports the following relationships: Section Editor and consultant for UpToDate; co-editor of book for Springer Publishing; and Faculty Board of Review for GE-Association of University Radiologists Radiology Research Academic Felowship (GERRAF). Dr. Krug reports grants from General Electric outside of the current work; consultation for Relievant MedSystems outside of the current work. Dr. Lotz reports grants from NIH/NIAMS during the conduct of the study; consultation for Relevent MedSystems outside of the current work. The other authors report no potential conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the creative commons license, and indicate if changes were made.

References

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Affiliations

Aaron J. Fields1, Michele C. Battie2, Richard J. Herzog3, Jeffrey G. Jarvik4, Roland Krug5, Thomas M. Link5, Jeffrey C. Lotz1, Conor W. O’Neill1, Aseem Sharma6, for the ISSLS Degenerative Spinal Phenotypes Group

Aaron J. Fields
aaron.fields@ucsf.edu

1 Department of Orthopaedic Surgery, University of California, 513 Parnassus Avenue, S-1161, Box 0514, San Francisco, CA 94143-0514, USA

2 Faculty of Health Sciences and Western’s Bone and Joint Institute, University of Western Ontario, London, ON, Canada

3 Department of Radiology, Hospital for Special Surgery, New York, NY, USA

4 Departments of Radiology, Neurosurgery and Health Services, and the Comparative Effectiveness, Cost and Outcomes Research Center, University of Washington, Seattle, WA, USA

5 Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, USA

6 Mallinckrodt Institute of Radiology, Washington University School of Medicine in St. Louis, St. Louis, MO, USA