OMERACT Hip Inflammation Magnetic Resonance Imaging Scoring System (HIMRISS) Assessment in Longitudinal Study

Jacob L. Jaremko, Robert G.W. Lambert, Susanne J. Pedersen, Ulrich Weber, Duncan Lindsay, Zeid Al-Ani, Kieran Steer, Marcus Pianta, Stephanie Wichuk, and Walter P. Maksymowych

ABSTRACT. Objective. To assess reliability, feasibility, and responsiveness of Hip Inflammation Magnetic resonance imaging Scoring System (HIMRISS) for bone marrow lesions (BML) in hip osteoarthritis (OA).

Methods. HIMRISS was scored by 8 readers in 360 hips of 90 patients imaged pre/post-hip steroid injection. Pre-scoring, new readers trained online to achieve intraclass correlation coefficient (ICC) > 0.80 versus experts.

Results. HIMRISS reliability was excellent for BML status (ICC 0.83–0.92). Despite small changes post-injection, reliability of BML change scores was high in femur (0.76–0.81) and moderate in acetabulum (0.42–0.56).

Conclusion. HIMRISS should be a priority for further assessment of hip BML in OA, and evaluated for use in other arthropathies. (First Release February 15 2019; J Rheumatol 2019;46:1239–42; doi:10.3899/jrheum.181043)

Key Indexing Terms:
- HIP JOINT SCORING METHODS
- OSTEOARTHRITIS OMERACT IMAGING
- BONE MARROW

Semi-quantitative scoring of magnetic resonance imaging (MRI) features of arthritis offers an objective target for therapy. Per the Outcome Measures in Rheumatology (OMERACT) Filter 2.0,123, scoring systems should be carefully evaluated for reliability, feasibility, and discrimination. The Hip Inflammation MRI Scoring System (HIMRISS) evaluates markers of active hip inflammation including bone marrow lesion (BML).45, recognizing that BML is increased T2/short-tau inversion recovery (STIR) signal intensity from inflammatory or non-inflammatory processes.6 In OMERACT 2016, a Web-based training module incorporating real-time iterative feedback calibration (RETIC) improved feasibility of HIMRISS scoring without sacrificing reliability5. Subsequent innovations in HIMRISS include Web-based interface and touch-sensitive electronic overlays to facilitate scoring. For OMERACT 2018, feasibility, reliability, and responsiveness of HIMRISS BML scoring were tested in a multireader scoring exercise on new prospectively obtained longitudinal data in patients receiving steroid injections for hip osteoarthritis (OA). The exercise was performed within the OMERACT MRI in Arthritis Working Group, from January to April 2018, and presented at OMERACT 14 (Terrigal, Australia, May 2018).

MATERIALS AND METHODS
HIMRISS. HIMRISS BML scoring has been described previously.46 Within a Web-based interface (www.carearthritis.com; accounts free to registered users), a reader opens a coronal fluid-sensitive MRI sequence. The reader moves/resizes a semitransparent overlay (with adjustable opacity) to fit the femoral head on a reference slice, then scrolls through slices, identifying each region containing BML on each slice by touching or mouse-clicking on the overlay. This sets the score for that region to 1, while regions not clicked/touched by the user have a default score of 0 (no BML). A spread-to
RESULTS

Patients had a wide range of radiographic OA severity from KL grades 1–4. Pain, functional disability, and stiffness were substantial at baseline (mean WOMAC 44/100), as expected in subjects presenting for steroid injection, but there was only slight improvement at 8 weeks (mean WOMAC 37.7/100; Table 1).

At baseline, HIMRISS BML scoring reliability was excellent between the 2 expert readers (ICC 0.91 femoral head, 0.88 acetabulum), and although somewhat lower among the 6 other readers (0.73, 0.62), reliability was still excellent among all 8 readers (0.83, 0.83). Reliability declined slightly when only the central 5 slices were considered (all 8 readers: 0.82 femoral head, 0.76 acetabulum; Table 1).

Consistent with the small changes observed in WOMAC scores, the magnitude of BML change between baseline and 8 weeks postintraarticular steroid injection was small, averaging 2/65 at femoral head and < 1/35 in the acetabulum. Based on observed reliability, only 45% and 33% of femoral heads and acetabulae (respectively) showed change greater than the SDC (for femoral head SDC = 3.6/65, acetabulum SDC = 1.9/35; Table 2). Despite this, interobserver reliability for change remained high at the femoral head for experts (ICC 0.81) and all readers (ICC 0.76), and fair to moderate at the acetabulum (0.56, 0.42; Table 1). Reliability declined only slightly when only the central 5 slices were considered. These 5 slices contained most of the observed femoral head BML (mean BML score = 11.8 for central 5 slices vs 16.1 for all 15 slices) and half of the acetabular BML (5.3 vs 10.2).

The RETIC training (8 cases) required 2–6 hours for new users to complete. Then, scoring time in the exercise was 5–15 minutes per hip. Despite the lengthy reading task (360 hips), all readers completed the exercise and reader comments were highly positive regarding participation in future OMERACT scoring exercises.

DISCUSSION

In our study, we assessed feasibility, reliability, and discrimination of HIMRISS BML scoring in a multireader exercise on a large prospective longitudinal dataset. We observed excellent reliability, even among new readers, for BML status at baseline, and high reliability for detection of change in BML despite the small changes seen in this dataset, indicating high responsiveness. New reader ICC were only ~0.1–0.2 below that of experts on their first scoring exercise. These high levels of interobserver reliability are likely because of the insistence that new readers first achieve competence on training data using our interactive RETIC system.

In posthoc analysis, we found that simplifying scoring to only 5 central slices would have identified most of the total burden of BML (about 3/4 of femoral head BML, and about half of acetabular BML; Table 2), with only slightly decreased reliability for BML status and change, and a decrease in scoring time (removing 10/15 slices scored per hip), which may improve feasibility. However, it is not known whether restricting the number of slices scored would alter discrimination or effect-size of the tool because of the
exercise of this size (360 hips), it was not feasible to score only active lesions, not structural damage; and in a scoring corresponds to the timing of maximal therapeutic effect, thus according to the underlying condition and therapy.

...whether the timing of maximal temporal change on MRI post-injection, declining rapidly after 3 months. It is unclear better than placebo for hip OA pain relief7, and the magnitude of this effect is usually maximal between 2–8 weeks... discriminative capacity still cannot be assessed because of variation in red marrow distribution.

This study had limitations. Although HIMRISS was sufficiently reliable to detect small changes in BML (Table 2), discriminative capacity still cannot be assessed because of limited change in BML and self-reported clinical outcomes 8 weeks post-injection in this dataset. Intraarticular steroid injections have previously been shown to be significantly better than placebo for hip OA pain relief7, and the magnitude of this effect is usually maximal between 2–8 weeks post-injection, declining rapidly after 3 months. It is unclear whether the timing of maximal temporal change on MRI corresponds to the timing of maximal therapeutic effect, thus the optimal timing of MRI is unknown and may vary according to the underlying condition and therapy.

Other limitations were that HIMRISS scoring considers only active lesions, not structural damage; and in a scoring exercise of this size (360 hips), it was not feasible to score the dataset twice, hence we could not test intraobserver reliability.

Overall, HIMRISS BML scoring is feasible and highly reliable when performed by readers trained using our RETIC interactive online calibration method. HIMRISS should be a priority for further assessment of hip BML in OA and evaluated for use in inflammatory hip arthropathies.

**ACKNOWLEDGMENT**

Thanks to Joanne McGoe for her assistance with study subjects, Joel Paschke for his computer programming skills, and Geoff Bostick and Linda Woodhouse for crucial support establishing the STIHO cohort.

**ONLINE SUPPLEMENT**

Supplementary material accompanies the online version of this article.

**REFERENCES**


