Original Article

Dixon T2 imaging of vertebral bone edema: reliability and comparison with short tau inversion recovery

Acta Radiologica
1–11

DOI: 10.1177/02841851221146130
journals.sagepub.com/home/acr


Magnhild H Dagestad[1,2], Monika K Toppe[1], Per M Kristoffersen[1,2], Elisabeth Gjefsen[3,4], Erling Andersen[5], Jörg Assmus[6], Nils Vetti[1,2] and Ansgar Espeland[1,2]

Abstract

Background: It is uncertain whether T2-weighted Dixon water images (DixonT2w) and short tau inversion recovery (STIR) are interchangeable when evaluating vertebral bone edema, or if one method is superior or visualizes the edema differently.

Purpose: To compare image quality and Modic change (MC)-related edema between DixonT2w and STIR and estimate inter-observer reliability for MC edema on DixonT2w.

Material and Methods: Consecutive patients (n = 120) considered for the Antibiotics in Modic changes (AIM) trial underwent lumbar 1.5-T magnetic resonance imaging with two-point DixonT2w and STIR. Two radiologists assessed MC-related high-signal lesions on DixonT2w and compared image quality and lesion extent with STIR. Cohen's kappa and mean of differences ± limits of agreement were calculated.

Results: Fat suppression and artefacts were similar on DixonT2w and STIR in 116 of 120 (97%) patients. Lesion conspicuity was similar in 88, better on STIR in 10, and better on DixonT2w in 9 of 107 patients with MC-related high-signal lesions. Contrast-to-noise ratio for STIR versus DixonT2w was 19.1 versus 17.1 (mean of differences 2.0 ± 8.2). Of 228 lesions L4-S1, 215 (94%) had similar extent on DixonT2w and STIR, 11 were smaller/undetected on STIR, and two were smaller/undetected on DixonT2w. Lesions missed on STIR (n = 9) or DixonT2w (n = 1) had a weak signal increase on the other sequence (≤17%; 0% = vertebral body, 100% = cerebrospinal fluid). Inter-observer reliability (mean kappa L4-S1) was very good for presence (0.87), moderate for height (0.44), and good for volume (0.63) of lesions on DixonT2w.

Conclusion: DixonT2w provided similar visualization of MC-related vertebral edema as STIR.

Keywords

Magnetic resonance imaging, spine, imaging sequences, adults, skeletal-axial, observer performance

Date received: 17 November 2020; accepted: 30 November 2022

Introduction

Water-only images from magnetic resonance imaging (MRI) with the Dixon method are used to evaluate edema changes in various parts of the body, including the spine (1–5). The Dixon method separates fat and water and can provide uniform fat suppression with few artefacts (1) and, unlike other fat-suppression techniques, it provides images with and without fat suppression in one acquisition.

Short tau inversion recovery (STIR) is an established fluid-sensitive MRI method (6). Both STIR and T2-weighted Dixon water-only images (DixonT2w) are used to assess bone edema. However, it is uncertain whether these methods can be used interchangeably, or if

[1]Department of Radiology, Haukeland University Hospital, Bergen, Norway
[2]Department of Clinical Medicine, University of Bergen, Bergen, Norway
[3]Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital HF, Oslo, Norway
[4]Faculty of Medicine, University of Oslo, Oslo, Norway
[5]Department of Clinical Engineering, Haukeland University Hospital, Bergen, Norway
[6]Centre for Clinical Research, Haukeland University Hospital, Bergen, Norway

Corresponding author:
Magnhild H Dagestad, Department of Radiology, Haukeland University Hospital, Jonas Liesvei 65, 5021 Bergen, Norway; Department of Clinical Medicine, University of Bergen, P.O. Box 7804, 5020 Bergen, Norway.
Email: magnhild.hammersland.dagestad@helse–bergen.no

one method is superior or may visualize the edema differently. Few have compared DixonT2w and STIR, and results diverge. At 1.5-T MRI of the spine, DixonT2w showed better fat suppression but poorer lesion conspicuity than STIR in one study (2) and a similar lesion depiction at a reduced scan time in another study (4). DixonT2w yielded better fat suppression than STIR in the hands (7), but poorer inter-observer agreement for occult fractures in hip and pelvis (8).

In this study, we compared the use of DixonT2w and STIR to evaluate bone edema related to Modic changes (MCs) in the lumbar spine. MCs are MRI signal changes extending from the endplate into the vertebral body marrow (9). There is conflicting evidence for whether MCs are related to pain (10,11), but those containing edema may be painful (12–14) and MCs have been targets of treatment in selected patients with back pain (15–21). Data are scarce regarding evaluations of MC edema on DixonT2w, but suggest moderate to very good inter-observer reliability for such evaluations on STIR (22). The aim of the present study was to compare image quality and MC-related edema between DixonT2w and STIR and estimate inter-observer reliability for MC edema on DixonT2w.

Material and Methods

This cross-sectional study was based on MRI examinations of 120 consecutive patients with chronic low back pain. All were prospectively recruited from hospital outpatient clinics in Norway, had MCs on a previous clinical MRI, and were considered for possible participation in the Antibiotics In Modic changes (AIM) trial (15,23). Patients preliminarily eligible for the trial based on the previous MRI and the criteria in Appendix Table A1 were included in the present study regardless of their current MRI findings or final trial eligibility. No patient had prior low back surgery, except disc herniation surgery conducted >1 year earlier. No patient had lumbar metal implants. The Regional Committees for Medical Research Ethics in Norway approved this study. All patients gave written informed consent before inclusion. The Regional Committees for Medical Research Ethics in Norway approved this study and the trial it was based on (REC South East, approval number 2014/158). The trial is registered at ClinicalTrials.gov (identifier NCT02323412).

Images

The 120 patients underwent MRI from June 2015 to September 2016 at five centers using identical protocols and 1.5 T scanners (Magnetom Avanto B19; Siemens Healthineers, Erlangen, Germany). This study applied sagittal two-point DixonT2w and STIR sequences acquired during the same scanning session (Table 1). Dixon in-phase, opposed-phase, or fat only images were not used.

Evaluation

Two radiologists, with four and five years of experience and a special interest in musculoskeletal imaging, independently evaluated levels Th12-S1 (12 endplates) using all

Table 1. MRI parameters for sagittal T2-weighted Dixon and STIR of the lumbar spine.

Parameter	Dixon	STIR
Repetition time (ms)	3970	5530
Echo time (ms)	75	70
Echo train length	15	20
No. of acquisitions	1	1
No. of concatenations	1	1
No. of slices	15	15
Matrix (frequency × phase)	320 × 240	320 × 224
FOV	300 × 300	300 × 300
Slice thickness (mm)	4.0	4.0
Interslice gap (mm)	0.4	0.4
Voxel size (mm)	1.3 × 0.9 × 4.0	1.3 × 0.9 × 4.0
Receiver bandwidth (Hz/px)	401	182
Phase encoding direction	Head to feet	Head to feet
Saturation pulses	Anterior, 30 mm	Anterior, 30 mm
Acquisition time	3 min 48 s	1 min 58 s
Coverage	From above Th12 to below S2	From above Th12 to below S2
Phase oversampling (%)	70	70
Inversion time (ms)	NA	160
Parallel acquisition techniques mode	None	None
Filters	Large FOV filter	Large FOV filter, elliptic filter

MRI was performed on 1.5-T Magnetom Avanto scanners (Siemens) with B19 software, using the integrated phased array spine coil.
FOV, field of view; MRI, magnetic resonance imaging; NA, not applicable; STIR, short tau inversion recovery.

sagittal slices. They were blinded to clinical data, except the presence of chronic low back pain. For each patient, first DixonT2w images were interpreted blinded to STIR, then image quality and signal increases were directly compared between DixonT2w and STIR.

Image quality was defined by visual assessments of fat suppression uniformity, artefacts, and lesion conspicuity and by measurements of noise, signal-to-noise ratio (SNR), and contrast-to-noise ratio (CNR), as detailed below. Increased MC-related signal on DixonT2w was defined as a visibly increased signal compared to normal vertebral bone marrow. The signal increase had to be located in or be in contact with a region with MCs on T1-weighted (T1W) and T2-weighted (T2W) fast spin-echo images – or be shaped and located as an MC. The observers had the T1W and T2W images open when assessing DixonT2w images. MCs were defined based on non-fat suppressed T1W/T2W fast-spin echo images alone as signal changes in the vertebral bone marrow that extend from the endplate, as described by Modic et al. (9).

Blinded to STIR, the observers evaluated presence (yes/no), height, volume, and maximum intensity of MC-related high signal on DixonT2w. The largest height of the high signal lesion was measured in millimeters and was recalculated into a percentage of vertebral body marrow height measured in millimeters along the same line on the same image. This height percentage was analyzed both as a continuous variable and categorized (<10%, <25%, 25%–50%, or >50%). The volume of the region with high signal was visually estimated as <10%, <25%, 25%–50%, or >50% of the total volume of the vertebral body marrow.

Signal intensity was measured on DixonT2w in the brightest area of the MC-related high-signal lesion, in normal vertebral body marrow, and in cerebrospinal fluid (CSF) using circular regions of interest (ROIs) available in our PACS. A ROI of size 25 mm^2 was used for measuring the most intense MC-related high signal and the signal of CSF at the level of the affected vertebral unit. A ROI of 44 mm^2 was used to measure the signal intensity in normal vertebral bone marrow, avoiding the central vertebral vein (Fig. 1) (22). Maximum intensity of the MC-related high signal on DixonT2w ("Dixon") in % points was calculated on a scale from normal vertebral body intensity ("Body," 0%) to CSF intensity ("CSF," 100%) as (["Dixon" – "Body"]/["CSF" – "Body"])*100%. We also used corresponding intensity measurements made on the STIR images (blinded to DixonT2w) in a prior reliability study by three musculoskeletal radiologists, each with >10 years of experience (22).

Comparing DixonT2w and STIR, the current observers evaluated fat suppression uniformity (better on Dixon, similar, better on STIR), artefacts (less on Dixon, similar, less on STIR), lesion conspicuity (better on Dixon, similar, better on STIR), noise (see below), and extent of MC-related high signal (larger on Dixon, similar, larger on STIR). Fat-water swap on Dixon (fat presented as water and water presented as fat) (1) was also noted (yes/no) and whether poor fat suppression or artefacts reduced diagnostic confidence on Dixon (yes/no) or STIR (yes/no). Only definite findings or differences, visible on ≥2 image slices were noted.

Noise was defined as the standard deviation (SD) of the background signal intensity outside the patient, after verifying a Rayleigh distribution of the intensity (i.e. mean/SD = 1.9) (24). Noise was measured in grayscale units in the mid-sagittal image (or the next image left or right) in a circular ROI of 44 mm^2 that was 1–2 cm posterior to the skin at the level of L3 (Fig. 1). We calculated the SNR and CNR at L4-S1; SNR = normal vertebral body intensity/noise; CNR = (maximum MC-related signal intensity in grayscale units – vertebral body intensity)/noise (25).

Conclusive findings

Conclusive findings were based on agreement between both radiologists and – in cases of disagreement – consensus with a musculoskeletal radiologist with >10 years of experience. For the presence of MC-related high signal on STIR, the conclusive rating was the majority rating reported (blinded to DixonT2w) by the three radiologists in the previous reliability study (22). For noise, SNR, and CNR on DixonT2w and STIR, the conclusive value was the mean of two radiologists' values.

Pilot study

The two radiologists conducted a pilot study on 14 patients, not included in the present study, in order to refine and find a similar understanding of the MRI evaluation criteria. They interpreted high signal lesions, compared image quality and lesions between DixonT2w and STIR, and discussed disagreements and criteria. At the end of the pilot study, their MRI assessments and the MRI evaluation criteria were concluded to be adequate for starting the main study.

Statistical analyses

The one-sample signed rank sum test was applied to compare fat suppression uniformity, artefacts, lesion conspicuity, and signal extent between DixonT2w and STIR. To compare noise, SNR, and CNR, we used Bland–Altman plots with the mean of differences ± 1.96 SD (limits of agreement [LoA]).

Cohen's kappa was calculated for inter-observer agreement on MC-related high signal on DixonT2w at each endplate L4-S1, and for agreement on conclusive findings between DixonT2w and STIR across L4-S1 (480 endplates). Due to the low prevalence of findings (<10%) in

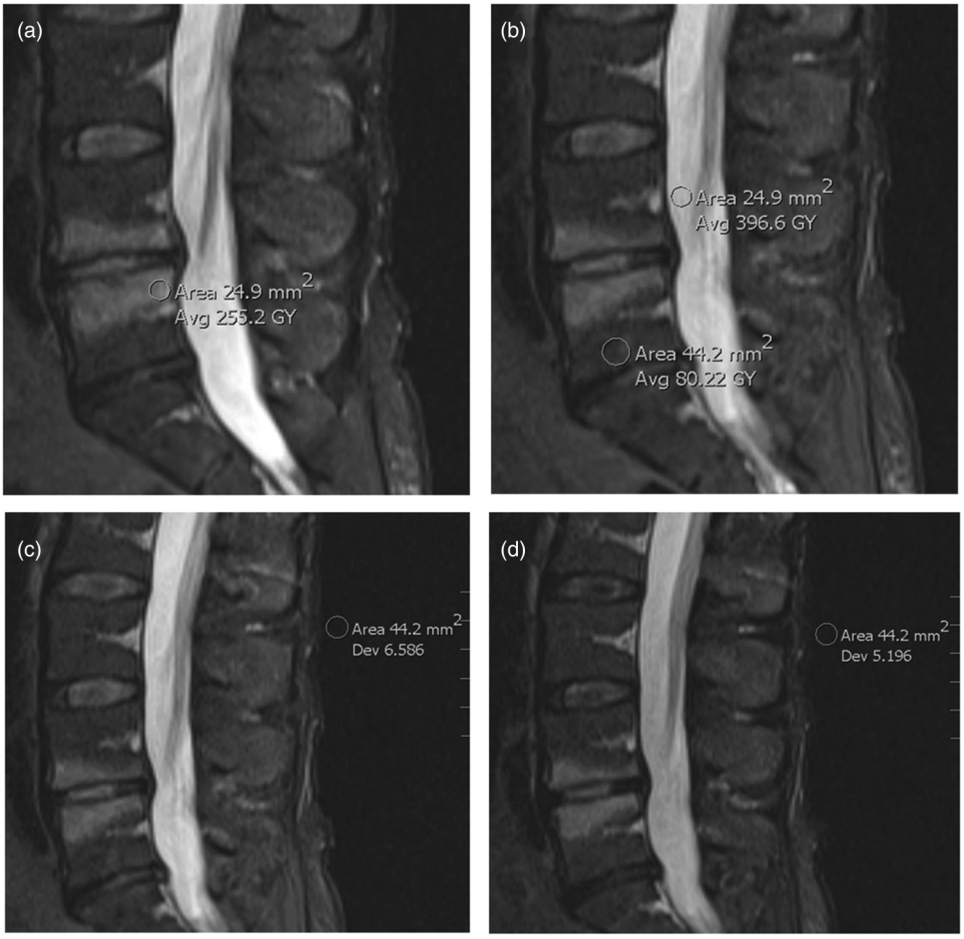

Fig. 1. Measurements of signal intensities and noise. (a–d) A 50-year-old woman with chronic low back pain. Measurements of signal intensity (GY, grayscale units) in (a) the most intense Modic change-related high-signal region and in (b) normal vertebral body and CSF on DixonT2w. Measurement of noise (Dev) on (c) DixonT2w and (d) STIR. Midline slices were used for measuring CSF intensity at the affected level, vertebral body intensity centrally near the endplate in the normal part of the affected or a close vertebra, and noise in air posterior to L3. CSF, cerebrospinal fluid; DixonT2w, T2-weighted Dixon water images; STIR, short tau inversion recovery.

levels Th12-L3, kappa was not calculated for these levels (26). The interpretation of Cohen's kappa was as follows: 0.00–0.20 = poor; 0.21–0.40 = fair; 0.41–0.60 = moderate; 0.61–0.80 = good; and 0.81–1.00 = very good agreement beyond chance (26).

For height and maximum intensity of the high signal on DixonT2w in percentage points, we calculated means of differences between observers with LoA across all endplates L4-S1.

MedCalc 17.6 (MedCalc Software) was used for analyses and Matlab 9.5 (Mathworks) for plots.

Sample size

A sample size of 85 patients is needed to detect an unweighted pairwise kappa value of 0.70 as significantly larger than 0.40 for a finding with 30% prevalence ($\beta = 0.2$, two-sided $\alpha = 0.05$) (26). We included 120 patients to improve the power.

Results

The 120 included patients (72 women) were aged 25–64 years (mean age = 45 years).

Image quality: DixonT2w versus STIR

The fat suppression uniformity was concluded to be similar on DixonT2w and STIR in 116 of 120 (97%) patients and better on Dixon in four patients (Table 2, Fig. 2). Reduced diagnostic confidence due to poor fat suppression was not reported in any patient.

Artefacts were similar on DixonT2w and STIR in 116 patients and less on STIR in four patients by conclusion, and less on STIR in 11 patients according to one radiologist (Table 2). Reduced diagnostic confidence due to artefacts was concluded for both Dixon and STIR in three patients and for Dixon only in four patients. These appeared as pulsation artefacts in all cases, visible posterior to aorta in six of seven cases.

Table 2. Image quality of sagittal T2-weighted Dixon water versus STIR images of the lumbar spine (n = 120) reported independently by two radiologists and in conclusion.

Variable	Radiologist A	Radiologist B	Conclusion*
Fat suppression uniformity			
Better on Dixon	5 (4)	0	4 (3)
Similar	115 (96)	120	116 (97)
Better on STIR	0	0	0
P value	0.06	NA	0.13
Non-uniform fat suppression reduces diagnostic confidence			
On Dixon	0	0	0
On STIR	0	0	0
Artefacts			
Less on Dixon	0	2 (2)	0
Similar	109 (91)	115 (96)	116 (97)
Less on STIR	11 (9)	3 (2)	4 (3)
P value	0.001	0.81	0.13
Artefacts reduce diagnostic confidence			
On Dixon	11 (9)	4 (4)	7 (6)
On STIR	5 (4)	2 (2)	3 (3)
Lesion conspicuity†			
Better on Dixon	16 (14)	12 (11)	9 (8)
Similar	89 (80)	92 (86)	88 (82)
Better on STIR	7 (6)	3 (3)	10 (9)
P value	0.10	0.04	0.86

Values are given as n (%) unless otherwise indicated. P values are from the one-sample signed rank sum test.

*The conclusive rating was based on agreement between both radiologists or consensus with a third radiologist in cases of disagreement.

†Better on Dixon (or STIR) if more lesions have better conspicuity on Dixon vs. STIR (or vice versa); reported only for patients with at least one Modic change-related high-signal lesion at Th12-S1 according to radiologist A (n = 112) or B (n = 107) or in conclusion (n = 107).

NA, not applicable due to few observations; STIR, short tau inversion recovery.

Fat-water swap on Dixon occurred in 49 of 120 (41%) patients. It always affected the whole image stack (the whole part of each image) and did not influence diagnostic confidence.

The conclusive lesion conspicuity was similar in 82% of the patients without favoring DixonT2w or STIR (Table 2). Each radiologist separately tended to report better lesion conspicuity on Dixon.

DixonT2w showed slightly more noise than STIR (mean 5.2 vs. 4.6, LoA ± 1.8), similar SNR (mean 12.1 vs. 12.6, LoA ± 5.9), and slightly lower CNR (mean 17.1 vs. 19.1, LoA ± 8.2) across levels L4-S1 (Fig. 3).

High-intensity lesions: DixonT2w versus STIR

DixonT2w and/or STIR showed 228 MC-related high signal lesions at L4-S1 (480 endplates). DixonT2w missed 1 of 219 high-intensity lesions present on STIR and showed high signal at 9 of 261 endplates without high signal on STIR (kappa = 0.96).

The nine lesions missed on STIR had a low CNR and low maximum intensity on DixonT2w (CNR = 5.7–13.0, mean = 9.2 vs. 17.1 for all lesions; intensity 8%–17%, mean 12% vs. 25% for all lesions; 0% = vertebral body, 100% = CSF). One lesion missed on DixonT2w had a CNR of 5.1 and maximum intensity of 13% on STIR.

Of the 228 lesions L4-S1, 215 (94%) had a similar extent on DixonT2w and STIR, 2 (1%) were smaller or undetected on Dixon, and 11 (5%) were smaller/undetected on STIR (*P* = 0.03) (Fig. 4; Appendix Table A2).

Inter-observer reliability: DixonT2w

The inter-observer agreement (Fig. 5) was very good for the presence of MC-related high signal on DixonT2w (mean kappa = 0.87). The agreement on the categorized height of the high signal ranged from fair to good and was moderate overall (mean kappa = 0.44). For the volume of the region with a high signal, overall inter-observer agreement was good (mean kappa = 0.63).

For the height of the high signal on DixonT2w in the percentage of vertebral body height, the mean of differences between observers ± LoA was 2.1 ± 24.8% points at L4-S1. The corresponding figures were 0.03 ± 7.3% points for maximum intensity of the signal (scale of 0%–100%). Height percentage points were reported in the range of 7–100 (mean = 44) and intensity percentage points in the range of 3–67 (mean = 25) (Appendix Table A3).

Discussion

To our knowledge, this was the first comprehensive study (n = 120) to compare evaluations of MC-related vertebral bone edema between DixonT2w and STIR. The study included assessments of image quality, differences in visualization of bone edema, and inter-observer reliability. We found mostly similar fat-suppression uniformity, artefacts, lesion conspicuity, and diagnostic confidence with both methods. The presence of bone edema was highly similar on DixonT2w and STIR. The inter-observer reliability was very good for presence, moderate for height, and good for volume of MC-related high signal on DixonT2w.

Our findings support the clinical use of DixonT2w to assess vertebral bone edema. They also suggest that these 1.5-T DixonT2w and STIR protocols may be used interchangeably to evaluate the presence and extent of MC-related edema. It seems that no previous study has compared the extent of vertebral high signal between Dixon and STIR. In the present study, 94% of vertebral high signal lesions had similar extent on DixonT2w and STIR.

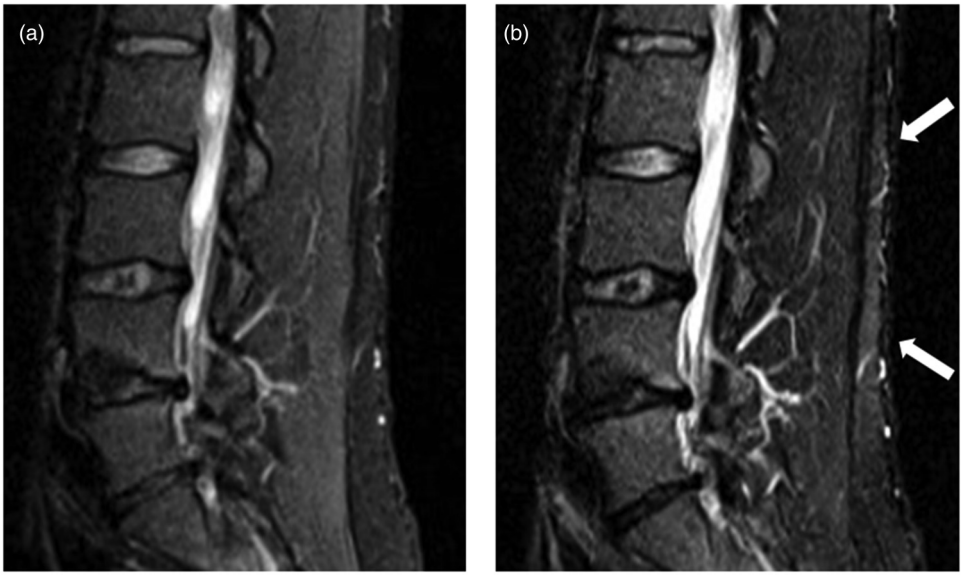

Fig. 2. Difference in fat suppression between DixonT2w and STIR. (a, b) A 31-year-old man with chronic low back pain. More uniform suppression of subcutaneous fat on (a) DixonT2w compared to (b) STIR. DixonT2w, T2-weighted Dixon water images; STIR, short tau inversion recovery.

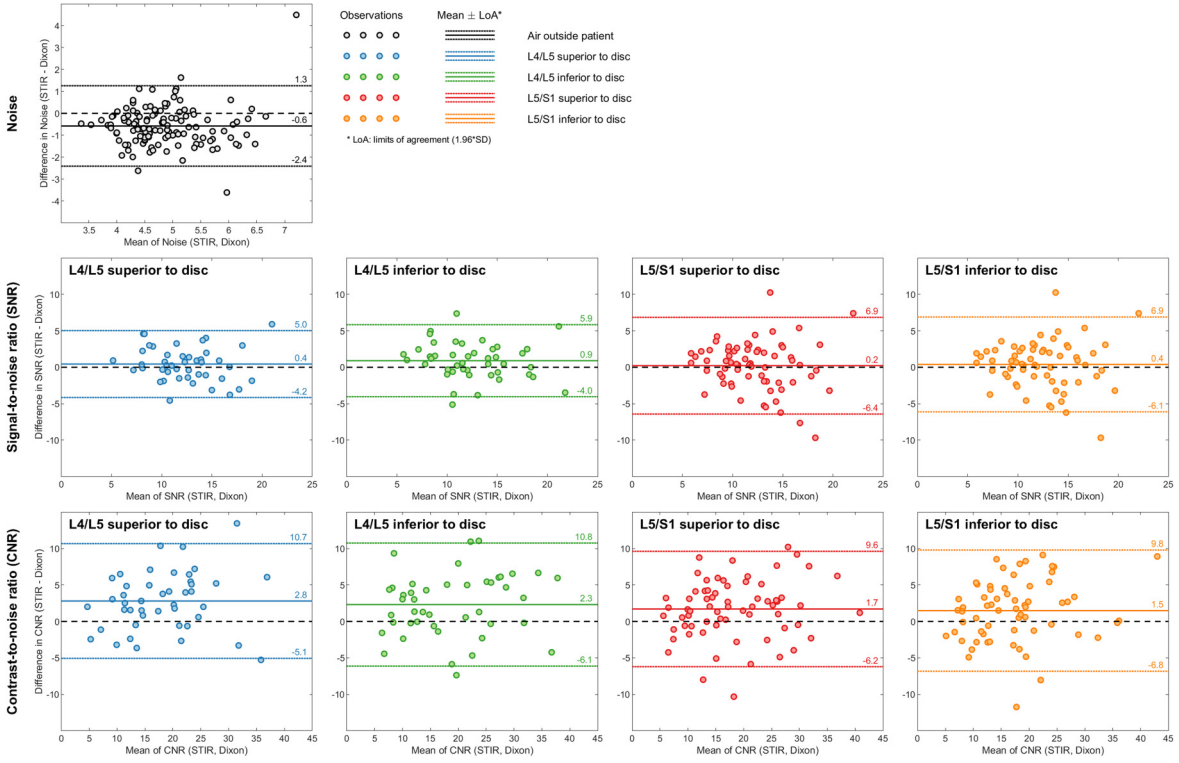

Fig. 3. Differences in noise, SNR, and CNR between STIR and DixonT2w. Noise = the standard deviation of the signal intensity in air outside the patient (in grayscale units). SNR = normal vertebral body intensity in grayscale units/noise. CNR = (maximum MC related signal intensity in grayscale units – vertebral body intensity in grayscale units)/noise. CNR, contrast-to-noise ratio; DixonT2w, T2-weighted Dixon water images; MC, Modic change; SNR, signal-to-noise ratio; STIR, short tau inversion recovery.

Our reliability results also support the use of DixonT2w. The inter-observer reliability for high signal findings on DixonT2w was comparable to the reliability achieved by the more experienced observers for the same findings on STIR in the same sample (22). The kappa on DixonT2w versus STIR was 0.87 versus 0.86 for presence, 0.44

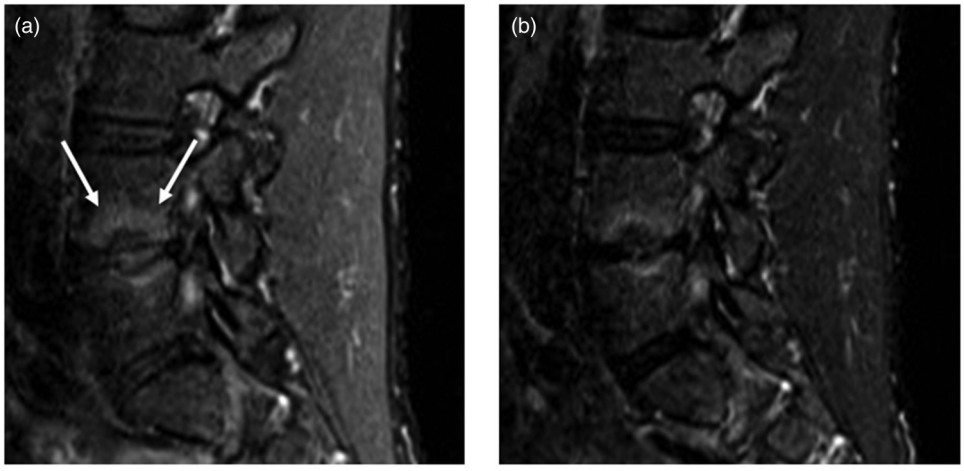

Fig. 4. Difference in lesion conspicuity and extent between DixonT2w and STIR. (a, b) A 38-year-old man with chronic low back pain. (a) DixonT2w shows slightly better conspicuity and somewhat larger extent of high signal (arrows) than (b) STIR. DixonT2w, T2-weighted Dixon water images; STIR, short tau inversion recovery.

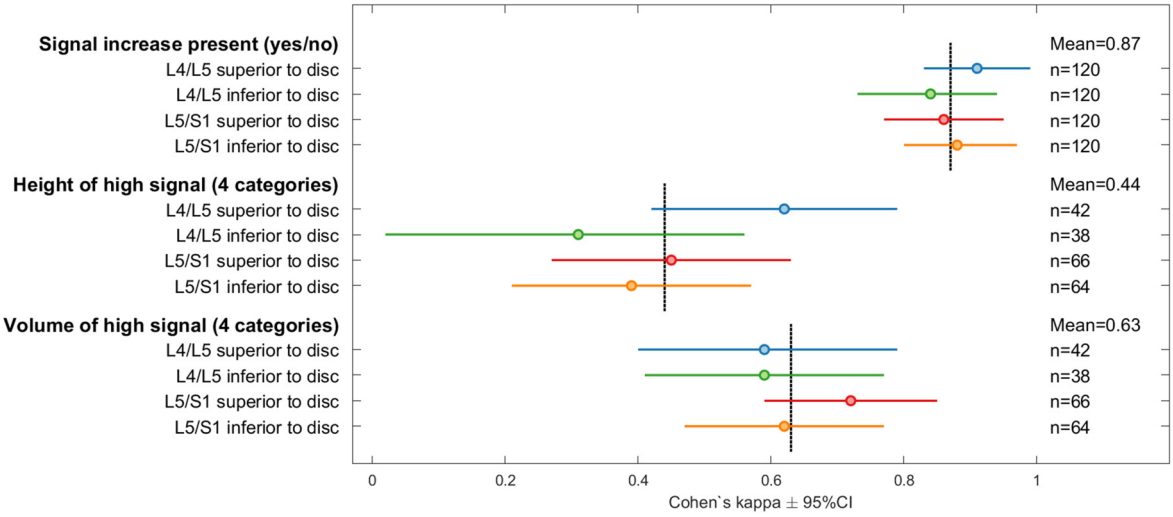

Fig. 5. Inter-observer agreement for Modic change-related signal-increases on DixonT2w. Vertical lines show mean kappa for levels L4–S1. Kappa is linearly weighted for height and volume. CI, confidence interval; DixonT2w, T2-weighted Dixon water images.

versus 0.51 for height, and 0.63 versus 0.56 for volume of high signal; the largest mean of differences ± LoA was 0.03 ± 7.3 vs 0.9 ± 7.6% points for maximum intensity of the signal. The low kappa value for height may be due to gradual tapering of the high signal, making height assessments difficult. These results for reliability when rating high signal lesions on DixonT2w and STIR need confirmation in other studies.

Some minor differences between DixonT2w and STIR need to be discussed. More uniform fat suppression on DixonT2w in a few patients (n = 4/120) may be explained by Dixon being less sensitive to local magnetic field heterogeneity (2). The fat suppression in the lumbar spine was also regarded better on Dixon versus STIR in one small previous study (n = 13) (2). That study did not examine the impact of the difference in fat suppression on diagnostic confidence; we found no impact. Artefacts reduced diagnostic confidence in seven versus three patients on DixonT2w versus STIR. These pulsation artefacts appeared on several sequences. We do not consider them compelling evidence against DixonT2w.

Fat-water swap is a common postprocessing artefact on two-point Dixon images, reported in 17% of patients in a prior study (27). It was more frequent in our study (41%), but the swap always affected the whole image stack and did not affect the diagnostic evaluation. Local fat-water swaps can make the interpretation more difficult (28). Multi-point Dixon methods provide more signal data for improved postprocessing, which makes fat-water swap less likely (1,28).

Slightly more noise and lower CNR on DixonT2w versus STIR did not imply poorer detection of high signal lesions with DixonT2w. Rather, such lesions were somewhat more frequent or slightly larger on DixonT2w compared to STIR, and each observer separately tended to report better lesion conspicuity on DixonT2w (Table 2). However, all lesions missed on STIR had low intensity on DixonT2w (≤17% intensity increase), possibly reducing their clinical relevance.

The acquisition time for Dixon was twice that for STIR (3 min 48 s vs. 1 min 58 s) (Table 1). The MRI protocol included a range of sequences, and each sequence (including STIR) had been shortened to reduce the total scan time while attempting to maintain good image quality. The Dixon sequence provided fat only and in-phase images that may correspond to T1W and non-fat suppressed T2W images (29), compensating for the longer acquisition time.

The strengths of this study include a large sample and standardized MRI scanning. The direct comparison of images was likely to provide reliable evaluation of differences (30).

The present study has some limitations. These include the fact that only MC edema was studied. However, a wide range of edema findings was covered, and the results may be relevant also when assessing edema due to infection, malignancy, or fracture. Our noise measurements were affected by the large field of view filters and multichannel reconstruction applied (we used a phased array coil) (Table 1) (24). We could have used better methods for measuring noise based on repeated acquisitions of the same sequence (24) but this was not feasible. Further, we only looked at two-point Dixon on 1.5-T MRI. The results may not apply to other water-fat separation methods or 3-T MRI. We did not examine the complete Dixon sequence (2, 29, 31), which may have potential to replace T1W images and STIR (25). Findings from 3-T MRI suggest that a Dixon sequence alone may be adequate for detecting vertebral bone metastases (25,32) and assessing degenerative spine conditions (29,33).

In conclusion, we found no important differences in fat suppression uniformity, lesion conspicuity, diagnostic confidence, or high signal detection rate between DixonT2w and STIR. Both methods provided similar visualization of MC-related vertebral bone edema.

Acknowledgements

The authors thank their collaborators in the AIM study group for their contributions.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Norwegian National Program for Clinical Therapy Research, KLINBEFORSK (grant no. 2017201), the Western Norway Regional Health Authority (grant nos. HV 911891, HV 911938), and the South East Norway Regional Health Authority (grant no. 2015-090). They had no role in the study design; collection, analysis, and interpretation of data; writing of the report; or the decision to submit the report for publication.

ORCID iDs

Magnhild H Dagestad https://orcid.org/0000-0003-4856-6321
Per M Kristoffersen https://orcid.org/0000-0001-6471-8861
Erling Andersen https://orcid.org/0000-0001-8268-1206
Ansgar Espeland https://orcid.org/0000-0002-0693-7465

References

1. Guerini H, Omoumi P, Guichoux F, et al. Fat suppression with Dixon techniques in musculoskeletal magnetic resonance imaging: a pictorial review. Semin Musculoskelet Radiol 2015;19:335–347.
2. Brandao S, Seixas D, Ayres-Basto M, et al. Comparing T1-weighted and T2-weighted three-point Dixon technique with conventional T1-weighted fat-saturation and short-tau inversion recovery (STIR) techniques for the study of the lumbar spine in a short-bore MRI machine. Clin Radiol 2013;68:e617–e623.
3. Lee S, Choi DS, Shin HS, et al. FSE T2-weighted two-point Dixon technique for fat suppression in the lumbar spine: comparison with SPAIR technique. Diagn Interv Radiol 2018;24:175–180.
4. Low RN, Austin MJ, Ma J. Fast spin-echo triple echo Dixon: initial clinical experience with a novel pulse sequence for simultaneous fat-suppressed and nonfat-suppressed T2-weighted spine magnetic resonance imaging. J Magn Reson Imaging 2011;33:390–400.
5. van Vucht N, Santiago R, Lottmann B, et al. The Dixon technique for MRI of the bone marrow. Skeletal Radiol 2019;48:1861–1874.
6. Del Grande F, Santini F, Herzka DA, et al. Fat-suppression techniques for 3-T MR imaging of the musculoskeletal system. Radiographics 2014;34:217–233.
7. Kirchgesner T, Perlepe V, Michoux N, et al. Fat suppression at 2D MR imaging of the hands: Dixon method versus CHESS technique and STIR sequence. Eur J Radiol 2017;89:40–46.
8. Heynen B, Tamigneaux C, Pasoglou V, et al. MRI detection of radiographically occult fractures of the hip and pelvis in the elderly: comparison of T2-weighted Dixon sequence with T1-weighted and STIR sequences. Diagn Interv Imaging 2019;100:169–175.
9. Modic MT, Steinberg PM, Ross JS, et al. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 1988;166:193–199.
10. Braten LCH, Schistad EI, Espeland A, et al. Association of Modic change types and their short tau inversion recovery

signals with clinical characteristics- a cross sectional study of chronic low back pain patients in the AIM-study. BMC Musculoskelet Disord 2020;21:368.
11. Herlin C, Kjaer P, Espeland A, et al. Modic changes-their associations with low back pain and activity limitation: a systematic literature review and meta-analysis. PLoS One 2018;13:e0200677.
12. Hanimoglu H, Cevik S, Yilmaz H, et al. Effects of Modic Type 1 changes in the vertebrae on low back pain. World Neurosurg 2019;121:e426–e432.
13. Splendiani A, Bruno F, Marsecano C, et al. Modic I changes size increase from supine to standing MRI correlates with increase in pain intensity in standing position: uncovering the "biomechanical stress" and "active discopathy" theories in low back pain. Eur Spine J 2019;28:983–992.
14. Maatta JH, Karppinen J, Paananen M, et al. Refined phenotyping of Modic changes: imaging biomarkers of prolonged severe low back pain and disability. Medicine (Baltimore) 2016;95:e3495.
15. Braten LCH, Rolfsen MP, Espeland A, et al. Efficacy of antibiotic treatment in patients with chronic low back pain and Modic changes (the AIM study): double blind, randomised, placebo controlled, multicentre trial. Br Med J 2019;367: l5654.
16. Albert HB, Kjaer P, Jensen TS, et al. Modic changes, possible causes and relation to low back pain. Med Hypotheses 2008;70:361–368.
17. Tavares I, Thomas E, Cyteval C, et al. Intradiscal glucocorticoids injection in chronic low back pain with active discopathy: a randomized controlled study. Ann Phys Rehabil Med 2020;64:101396.
18. Beaudreuil J, Dieude P, Poiraudeau S, et al. Disabling chronic low back pain with Modic type 1 MRI signal: acute reduction in pain with intradiscal corticotherapy. Ann Phys Rehabil Med 2012;55:139–147.
19. Cao P, Jiang L, Zhuang C, et al. Intradiscal injection therapy for degenerative chronic discogenic low back pain with end plate Modic changes. Spine J 2011;11:100–106.
20. Cai G, Laslett LL, Aitken D, et al. Effect of zoledronic acid and denosumab in patients with low back pain and Modic change: a proof-of-principle trial. J Bone Miner Res 2018;33:773–782.
21. Koivisto K, Kyllonen E, Haapea M, et al. Efficacy of zoledronic acid for chronic low back pain associated with Modic changes in magnetic resonance imaging. BMC Musculoskelet Disord 2014;15:64.
22. Kristoffersen PM, Vetti N, Storheim K, et al. Short tau inversion recovery MRI of Modic changes: a reliability study. Acta Radiol Open 2020;9:2058460120902402.
23. Storheim K, Espeland A, Grovle L, et al. Antibiotic treatment in patients with chronic low back pain and Modic changes (the AIM study): study protocol for a randomised controlled trial. Trials 2017;18:596.
24. Dietrich O, Raya JG, Reeder SB, et al. Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. J Magn Reson Imaging 2007;26:375–385.
25. Maeder Y, Dunet V, Richard R, et al. Bone marrow metastases: T2-weighted Dixon spin-echo fat images can replace T1-weighted spin-echo images. Radiology 2018;286:948–959.
26. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257–268.
27. Ladefoged CN, Hansen AE, Keller SH, et al. Impact of incorrect tissue classification in Dixon-based MR-AC: fat-water tissue inversion. EJNMMI Phys 2014;1:101.
28. Kirchgesner T, Acid S, Perlepe V, et al. Two-point Dixon fat-water swapping artifact: lesion mimicker at musculoskeletal T2-weighted MRI. Skeletal Radiol 2020;49: 2081–2086.
29. Zanchi F, Richard R, Hussami M, et al. MRI of non-specific low back pain and/or lumbar radiculopathy: do we need T1 when using a sagittal T2-weighted Dixon sequence? Eur Radiol 2020;30:2583–2593.
30. Berg L, Gjertsen O, Hellum C, et al. Reliability of change in lumbar MRI findings over time in patients with and without disc prosthesis--comparing two different image evaluation methods. Skeletal Radiol 2012;41:1547–1557.
31. Ma J. Dixon techniques for water and fat imaging. J Magn Reson Imaging 2008;28:543–558.
32. Hahn S, Lee YH, Suh JS. Detection of vertebral metastases: a comparison between the modified Dixon turbo spin echo T2 weighted MRI and conventional T1 weighted MRI: a preliminary study in a tertiary centre. Br J Radiol 2018;91: 20170782.
33. Sollmann N, Monch S, Riederer I, et al. Imaging of the degenerative spine using a sagittal T2-weighted DIXON turbo spin-echo sequence. Eur J Radiol 2020;131:109204.

Appendix

Dixon T2 imaging of vertebral bone edema: reliability and comparison with short tau inversion recovery

Table A1. Inclusion and exclusion criteria for the trial.

Inclusion criteria	Exclusion criteria
• Age 18–65 years • LBP of >6 months in duration in the area below the 12th rib and above the gluteal folds with a NRS pain intensity score of ≥5 (mean of three 0–10 NRSs: current LBP, the worst LBP within the last 2 weeks, and the usual/ mean LBP within the last 2 weeks) • MRI-confirmed lumbar disc herniation within the preceding 2 years • Type 1 and/or type 2 MC in the vertebral body marrow at the same level as the previously herniated disc. For patients with previous surgery for disc herniation, the MC has to be located at an operated level • Written informed consent	• Allergy to penicillin or cephalosporin • Allergy/hypersensitivity to any of the excipients of the study drug • Current pregnancy or lactation • Kidney (creatinine) or hepatic (ALAT/ASAT) laboratory values above the normal range • Phenylketonuria (Følling's disease) • Mononucleosis or leukemia • Any specific diagnosis that may explain the patient's low back symptoms (e.g. tumor, fracture, spondyloarthritis, infection, spinal stenosis) • Previous low back surgery (L1–S1) for reasons other than disc herniation (e.g. fusion, decompression, disc prosthesis) • Surgery for disc herniation within the last 12 months • Previous surgery for disc herniation, but MC located at level(s) that has/have not been operated on only • Reservation about the intake of gelatin (the capsules used to encapsulate the study medicine contains gelatin, which, among other things, is produced using ingredients derived from pigs) • Regular use of glucocorticoids • Regular use of opioids with the exception of codeine and tramadol • Not understanding Norwegian language • Unlikely to adhere to treatment and/or complete follow-up (e.g. serious ongoing psychiatric disease, drug abuse, plans to move) • Antibiotic treatment within the preceding 1 month before treatment start • Contraindications to MRI (e.g. cardiac pacemaker electrodes, metal implant in the eye or brain, claustrophobia) • Unwilling to participate

LBP, low back pain; MC, Modic change; MRI, magnetic resonance imaging; NRS, Numerical Rating Scale.

Table A2. Presence and extent of Modic change-related signal-increases on DixonT2w reported independently by two radiologists and in conclusion.

Variable	Radiologist A	Radiologist B	Conclusion*
Signal increase present on DixonT2w			
L4/L5 superior to disc	46/120 (38)	43/120 (36)	45/120 (38)
L4/L5 inferior to disc	42/120 (35)	43/120 (36)	44/120 (37)
L5/S1 superior to disc	70/120 (58)	70/120 (58)	71/120 (59)
L5/S1 inferior to disc	70/120 (58)	65/120 (54)	67/120 (56)
Extent of high signal on DixonT2w vs. STIR			
L4/L5 superior to disc			
Larger on Dixon	6/46 (13)	4/43 (9)	3/45 (7)
Similar	39/46 (85)	39/43 (91)	42/45 (93)
Larger on STIR	1/46 (2)	0	0
L4/L5 inferior to disc			
Larger on Dixon	1/42 (2)	3/43 (7)	1/44 (2)

(continued)

Table A2. (continued)

Variable	Radiologist A	Radiologist B	Conclusion*
Similar	40/42 (95)	40/43 (93)	42/44 (96)
Larger on STIR	1/42 (2)	0	1/44 (2)
L5/S1 superior to disc			
Larger on Dixon	10/70 (14)†	4/70 (6)	4/71 (6)
Similar	58/70 (83)	66/70 (94)	67/71 (94)
Larger on STIR	2/70 (3)	0	0
L5/S1 inferior to disc			
Larger on Dixon	6/70 (8)	2/65 (3)	3/68 (4)
Similar	62/70 (89)	63/65 (97)	64/68 (94)
Larger on STIR	2/70 (3)	0	1/68 (2)

Values are given as n (%).
*The conclusive rating was based on agreement between both radiologists or consensus with a third radiologist in cases of disagreement.
†$P = 0.04$ for difference in extent (one-sample signed rank sum test).
DixonT2w, T2-weighted Dixon water images; STIR, short tau inversion recovery.

Table A3. Height, volume, and intensity of Modic change-related signal-increases on DixonT2w reported independently by two radiologists.

Variable	Radiologist A	Radiologist B
Height of high signal region in % of vertebral body height		
L4/L5 superior to disc	49 (17–91)	52 (12–86)
L4/L5 inferior to disc	43 (10–100)	46 (13–86)
L5/S1 superior to disc	49 (15–100)	51 (12–89)
L5/S1 inferior to disc	33 (19–69)	36 (7–66)
Volume of high signal in % of vertebral body volume		
L4/L5 superior to disc		
<10%	23/46 (50)	15/43 (35)
<25%	17/46 (37)	16/43 (37)
25–50%	5/46 (11)	11/43 (26)
>50%	1/46 (2)	1/43 (2)
L4/L5 inferior to disc		
<10%	24/42 (57)	26/43 (61)
<25%	14/42 (33)	7/43 (16)
25–50%	3/42 (7)	10/43 (23)
>50%	1/42 (2)	0/43 (0)
L5/S1 superior to disc		
<10%	35/70 (50)	31/70 (44)
<25%	26/70 (37)	23/70 (33)
25–50%	9/70 (13)	16/70 (23)
>50%	0/70 (0)	0/70 (0)
L5/S1 inferior to disc		
<10%	50/70 (71)	44/65 (68)
<25%	19/70 (27)	14/65 (21)
25–50%	1/70 (1)	7/65 (11)
>50%	0/70 (0)	0/65 (0)
Maximum % signal increase (0%, vertebral body; 100%, CSF)		
L4/L5 superior to disc	26 (6–48)	27 (3–51)
L4/L5 inferior to disc	28 (7–60)	28 (6–58)
L5/S1 superior to disc	24 (7–59)	24 (5–57)
L5/S1 inferior to disc	23 (6–67)	23 (7–67)

Values are given as n (%) or mean (range).
CSF, cerebrospinal fluid; DixonT2w, T2-weighted Dixon water images.