Elsevier Editorial System(tm) for Osteoarthritis and Cartilage

Manuscript Draft

Manuscript Number: OAC8307R1

Title: Systematic Review and Meta-Analysis of the Reliability and Discriminative Validity of Cartilage Compositional MRI in Knee Osteoarthritis

Article Type: Review

Keywords: Knee osteoarthritis; Magnetic resonance imaging; Cartilage composition; Quantitative cartilage imaging; Cartilage mapping

Corresponding Author: Dr. James MacKay,

Corresponding Author's Institution: University of Cambridge

First Author: James MacKay

Order of Authors: James MacKay; Samantha Low; Toby Smith; Andoni Toms; Andrew McCaskie; Fiona Gilbert

Abstract: Objective

To assess reliability and discriminative validity of cartilage compositional magnetic resonance imaging (MRI) in knee osteoarthritis (OA).

Design

The study was carried out per PRISMA recommendations. We searched MEDLINE and EMBASE (1974 - present) for eligible studies. We performed qualitative synthesis of reliability data. Where data from at least 2 discrimination studies were available, we estimated pooled standardized mean difference (SMD) between subjects with and without OA. Discrimination analyses compared controls and subjects with mild OA (Kellgren-Lawrence (KL) grade 1-2), severe OA (KL grade 3-4) and OA not otherwise specified (NOS) where not possible to stratify. We assessed quality of the evidence using QAREL and QUADAS-2 tools.

Results

Fifty-eight studies were included in the reliability analysis and 25 studies were included in the discrimination analysis, with data from a total of 1,989 knees. Intra-observer, inter-observer and test-retest reliability of compositional techniques were excellent with most intraclass correlation coefficients > 0.8 and coefficients of variation < 10%. T1rho and T2 relaxometry were significant discriminators between subjects with mild OA and controls, and between subjects with OA (NOS) and controls ($p < 0.001$). T1rho showed best discrimination for mild OA (SMD [95% CI] = 0.73 [0.40 to 1.06], $p < 0.001$) and OA (NOS) (0.60 [0.41 to 0.80], $p < 0.001$). Quality of evidence was moderate for both parts of the review.

Conclusions

Cartilage compositional MRI techniques are reliable and, in the case of T1rho and T2 relaxometry, can discriminate between subjects with OA and controls.

TITLE:

SYSTEMATIC REVIEW AND META-ANALYSIS OF THE RELIABILITY AND DISCRIMINATIVE VALIDITY OF CARTILAGE COMPOSITIONAL MRI IN KNEE OSTEOARTHRITIS

MANUSCRIPT TYPE:

Systematic review and meta-analysis

AUTHORS:

Dr James W MacKay[1]	jwm37@cam.ac.uk
Dr Samantha B L Low[2]	samantha.low@nnuh.nhs.uk
Dr Toby O Smith[3]	toby.smith@uea.ac.uk
Professor Andoni P Toms[2]	andoni.toms@nnuh.nhs.uk
*Professor Andrew M McCaskie[4]	awm41@cam.ac.uk
*Professor Fiona J Gilbert[1]	fjg28@medschl.cam.ac.uk

[1]Department of Radiology, University of Cambridge, Cambridge, UK
[2]Department of Radiology, Norfolk & Norwich University Hospital, Norwich, UK
[3]School of Health Sciences, University of East Anglia, Norwich, UK
[4]Department of Trauma & Orthopaedics, University of Cambridge, Cambridge UK
*joint senior authors

CORRESPONDING AUTHOR:

Dr James W MacKay
Department of Radiology
University of Cambridge
Cambridge Biomedical Campus, Box 218
Hills Road
Cambridge CB2 0QQ
United Kingdom
Tel: +44 788 245 9094
Fax: +44 1223 330 915
Email: jwm37@cam.ac.uk

RUNNING TITLE

CARTILAGE MRI: SYSTEMATIC REVIEW

ABSTRACT

Objective

To assess reliability and discriminative validity of cartilage compositional magnetic resonance imaging (MRI) in knee osteoarthritis (OA).

Design

The study was carried out per PRISMA recommendations. We searched MEDLINE and EMBASE (1974 – present) for eligible studies. We performed qualitative synthesis of reliability data. Where data from at least 2 discrimination studies were available, we estimated pooled standardized mean difference (SMD) between subjects with and without OA. Discrimination analyses compared controls and subjects with mild OA (Kellgren-Lawrence (KL) grade 1-2), severe OA (KL grade 3-4) and OA not otherwise specified (NOS) where not possible to stratify. We assessed quality of the evidence using QAREL and QUADAS-2 tools.

Results

Fifty-eight studies were included in the reliability analysis and 26 studies were included in the discrimination analysis, with data from a total of 2,007 knees. Intra-observer, inter-observer and test-retest reliability of compositional techniques were excellent with most intraclass correlation coefficients > 0.8 and coefficients of variation < 10%. T1rho and T2 relaxometry were significant discriminators between subjects with mild OA and controls, and between subjects with OA (NOS) and controls ($p < 0.001$). T1rho showed best discrimination for mild OA (SMD [95% CI] =

0.73 [0.40 to 1.06], p < 0.001) and OA (NOS) (0.60 [0.41 to 0.80], p < 0.001). Quality of evidence was moderate for both parts of the review.

Conclusions

Cartilage compositional MRI techniques are reliable and, in the case of T1rho and T2 relaxometry, can discriminate between subjects with OA and controls.

KEY WORDS

Knee osteoarthritis; Magnetic resonance imaging; Cartilage composition; Quantitative cartilage imaging; Cartilage mapping

INTRODUCTION

Breakdown of articular cartilage is an important feature of knee osteoarthritis (OA). The earliest changes in articular cartilage are alterations in the biochemical composition of the extracellular matrix (ECM), a network of collagen fibrils and glycoproteins. These compositional changes may predispose to the development of focal defects, which in turn may lead to more diffuse cartilage loss associated with established OA.

Cartilage compositional MRI techniques such as T1rho relaxometry, T2 relaxometry and delayed gadolinium enhanced MRI of cartilage (dGEMRIC) are sensitive to changes in cartilage ECM composition, and provide a way to detect degeneration before gross morphological changes become apparent. This contrasts with conventional clinical MRI which can detect focal defects and diffuse cartilage loss but is limited in its ability to detect earlier changes in cartilage composition. Compositional MRI techniques may therefore allow identification of individuals suitable for intervention at an earlier stage, before irreversible changes occur. They also have the potential to assess response to treatments designed to repair or regenerate cartilage or slow degradation[1].

Previous systematic reviews have assessed the reliability and discriminative validity of radiographic and conventional clinical MRI assessment of knee osteoarthritis[2–4]. However, there has been no systematic review which has evaluated the reliability or discriminative validity of cartilage compositional techniques. This was identified as a gap in the literature in recent Osteoarthritis Research Society International (OARSI) guidelines for the use of imaging in the setting of OA clinical trials[5].

As cartilage compositional MRI techniques grow in popularity it is important to understand how reliable the techniques are and how well they can distinguish cartilage in individuals with OA compared to cartilage in healthy controls. Accordingly, the aim of this study was to assess the reliability and discriminative validity of cartilage compositional MRI in knee osteoarthritis.

For the purposes of this review, we use the term "reliability" to encompass both repeatability (measurement precision with conditions remaining unchanged between repeat measurements e.g. same observer, same MR platform) and reproducibility (measurement precision with conditions changing between repeat measurements, e.g. change in observer, change in MR platform) as defined by the Radiological Society of North America Quantitative Imaging Biomarkers Alliance (RSNA-QIBA) Metrology Working Group[6].

METHODS

This systematic review and meta-analysis undertaken in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations[7].

Protocol & Registration

The study review protocol was prospectively registered on PROSPERO, the international prospective register of systematic reviews (available at www.crd.york.ac.uk/prospero/display_record.asp?ID=CRD42016045250).

Eligibility criteria

We considered *in-vivo* studies in human subjects involving at least one cartilage compositional MRI technique at the knee. The list of compositional techniques considered included, but was not limited to, T1rho relaxometry, T2 relaxometry, T2* relaxometry, delayed gadolinium enhanced MRI of cartilage (dGEMRIC), sodium imaging, glycosaminoglycan chemical exchange saturation transfer imaging (gagCEST), diffusion weighted imaging (DWI) and diffusion tensor imaging (DTI). For a study to be included, it needed to provide reliability data on the technique used (either in subjects with OA or healthy controls or both) or provide measurements comparing subjects with OA to a control group (i.e. discrimination data), or provide both reliability and discrimination data. We considered only full-text papers reporting original data. Conference abstracts, review papers, letters to the editor and opinion pieces were excluded. We limited included studies to those published in English. Studies using animal models or human tissues *ex-vivo* were excluded.

Information sources

We searched MEDLINE (1946 – February 2017) and EMBASE (1974 – February 2017) via OVID. We also searched the databases OpenGrey, Clinicaltrials.gov and the World Health Organization (WHO) International Clinical Trials Registry for additional studies. The search strategy for the OVID search is presented in the Supplemental Material. We further scrutinized the reference lists of full-text manuscripts obtained, personal databases and the contents tables of key journals for any omitted studies.

Study selection

Two researchers performed initial screening to identify potentially eligible studies per inclusion and exclusion criteria. The full-texts of all potentially eligible papers were then evaluated to enable a final decision on inclusion. Discrepancies were resolved by discussion between the reviewers.

Data extraction and list of items

Data extraction was performed by a single researcher using a piloted electronic data collection form and subsequently verified by a second researcher, with disagreements resolved by discussion. Where a study was considered potentially eligible but data were not presented in an extractable format (e.g. presented in a figure without raw values), the corresponding study author was contacted by email to attempt to obtain the relevant data.

Data extracted for all studies included the following: year of publication, number of participants, age and sex of participants, study design, definition of OA used by the study (if applicable), details of the MR acquisition protocol, MR field strength, experience and training of image analysts, blinding of image analysts to additional

clinical information (e.g. OA/control status) and type of regional or subregional analysis performed.

We divided study participants with OA into those with mild OA (Kellgren-Lawrence (KL) grades 1-2), severe OA (KL grades 3-4) or OA not otherwise specified (NOS) when the study did not provide the information required to stratify[8].

Risk of bias in primary studies

The risk of bias for studies of reliability was performed by a single reviewer using a modification of the Quality Appraisal of Diagnostic Reliability (QAREL) tool relevant to our analysis[9]. Assessment of risk of bias in studies of discriminative validity was performed using a modification of the revised Quality Assessment of Diagnostic Accuracy (QUADAS-2) tool[10]. Full details of the modifications made to QAREL and QUADAS-2 tools are presented in the Supplemental Materials

Data analysis

The primary endpoint for the reliability assessment was a narrative summary of the reliability statistics for intra and inter-observer and test-retest reliability.

A meta-analysis was not appropriate due to the heterogeneity in methods for calculating reliability metrics from the included studies. For example, the coefficient of variation (CV) may be presented as a single value, or the root-mean-square average of several values (RMSCV). There are numerous approaches to computing the intraclass correlation coefficient (ICC) which prevent pooling, and directly comparing ICC across different populations could be misleading[11,12]. For the purposes of our review we used an interpretation of ICC values based on Landis and Koch[13].

The primary endpoints for the discriminative validity part of this review were estimates of standardized mean difference (average difference between groups divided by the pooled standard deviation of the two groups, analogous to effect size) between subjects with OA and normal controls for each compositional technique studied. Secondary endpoints were estimates of standardized mean difference between subjects with OA and normal controls for each compositional technique studied, limited to studies where control group participant age had been matched to the OA group. We used the standardized mean difference to allow comparison across different compositional techniques with values which vary considerably in magnitude. Where there were less than two studies available for any given comparison, a narrative analysis was conducted.

We assessed the appropriateness for meta-analysis by assessing the data extraction table for study heterogeneity in cohort characteristics, imaging technique, analysis technique and study processes. Where study heterogeneity was evident for one or more of these factors, a narrative analysis was undertaken. When this did not occur, a meta-analysis was undertaken. In each analysis, statistical heterogeneity was calculated through the I^2 statistic. Fixed effects models were used to pool outcome measures with low heterogeneity ($I^2 \leq 10\%$), whereas random effects models were used to pool outcome measures with high heterogeneity ($I^2 > 10\%$). We used a strict I^2 threshold as we wished to minimize the risk of any 'unknown' heterogeneity from influencing the interpretation of our analyses, particularly as we placed emphasis on excluding 'known' heterogeneity in our assessment of study characteristics.

All meta-analyses were performed using Review Manager version 5.3 (The Cochrane Collaboration)[14].

[FIGURE 1]

RESULTS

Study selection

Database searching identified 665 citations, with an additional 79 citations identified through other sources (personal databases, reference lists of included studies, contents tables of key journals). The full-text version of 192 articles was retrieved for detailed review. Forty-eight articles were included in the reliability assessment, 16 articles were included in the discrimination validity assessment and 10 articles were included in both evaluations (Figure 1)[15–88].

Reliability study characteristics

Characteristics of included studies are reported in Table 1. Data from 1,473 subjects were included in the reliability analysis. The most commonly used compositional technique was T2 relaxometry, featuring in 36 of 58 (62%) studies. The number of participants in each study ranged from five to 289 (median 20). The mean (standard deviation, SD) age of participants was 46.2 (14.5) years. Fifty-three percent of included subjects were female.

[TABLE 1]

Discriminative validity study characteristics

Characteristics of included studies are reported in table 2. Data from 766 subjects were included in the discriminative validity analysis. The most commonly used compositional technique was T2 relaxometry, featuring in 17 of 26 (65%) studies. The number of participants in each study ranged from 15 to 152 (median 33). The mean (SD) age of OA subjects was 58.3 (4.9) years compared to 40.9

(11.7) years in control subjects with females representing 56% of OA subjects compared to 52% of controls. Eight studies included subjects with mild OA, three studies included both subjects with mild OA and subjects with severe OA and 15 studies did not stratify OA severity and were considered as OA (NOS) for our analyses.

[TABLE 2]

Risk of bias in primary studies

Full results of quality assessments of reliability and discriminative validity studies are presented in the Supplemental Materials. Overall, the quality of the evidence was moderate for the reliability assessment and moderate for the discriminant validity assessment. Recurrent weaknesses for the reliability data included the assessment of reliability in only healthy volunteer subjects, lack of information regarding image analyst experience or training, and lack of information regarding image analyst blinding to previous results for studies of intra and inter-observer reliability. Recurrent weaknesses for the discrimination validity data included the use of unmatched control subjects and the potential lack of blinding of image analysts to subject group.

Reliability outcomes

The results of the reliability analysis are presented in Table 1 and Figure 2.

1. **T2 relaxometry**

Intra-observer ICCs ranged from 0.30 to 0.99 and CVs ranged from 0.8 to 4.7%. Studies featuring multiple subregional analyses tended to report lower ICC values [65]. Inter-observer ICCs ranged from 0.17 to 0.99 and CVs ranged from 1.0 to 12.2%. Again, studies which performed analysis on multiple small subregions and analysis of multiple

cartilage layers had poorer agreement[26,65]. CVs for test-retest reliability where analyses were performed on major compartments (e.g. medial femur, medial tibia etc) were 2.3 to 6.5%, with higher values (up to 22%) again seen where smaller subregional or laminar analyses were performed[26]. Three studies examined test-retest reliability in a multi-center setting, reporting CVs between 2.3 and 5.3% for major compartments and up to 14% for subregional analyses[16,39,45].

2. T1rho relaxometry

One study provided intra-observer reliability data, with a CV of 3.8%[32]. All studies of inter-observer reliability reported ICC values in the 'excellent' range (> 0.8). Inter-observer CV values ranged from 1.4 to 11.8%. Test-retest reliability was good-to-excellent (ICCs 0.73 to 0.96, CV 2.3 to 6.1%) when major compartments were analysed but poorer (ICCs as low as 0.2, CV as high as 19%) in two studies where laminar analysis was performed[20,45]. Two studies examined test-retest reliability in a milti-center setting, reporting a CVs of 4.9% for major compartments and up to 18.8% for subregional analyses[39,45].

3. dGEMRIC

Intra- and inter-observer reliability data were reported by two studies each, with excellent agreement (ICCs > 0.9) and CVs of less than 3%[15,48,67]. Test-retest reliability was excellent (ICCs > 0.85, CV 4.2 to 7.4%) apart from one study comparing different T1 mapping techniques for dGEMRIC which reported ICC values as low as 0 and a CV of 11% for a variable flip angle (VFA) technique[61].

4. Others

Intra- and inter-observer reliability data were also reported for sodium imaging, gagCEST, T2* relaxometry, T1 relaxometry (without contrast), magnetization transfer (MT) imaging and ultrashort TE T2* (UTE-T2*) relaxometry, with excellent agreement (ICCs > 0.8) for gagCEST, T2*, MT and UTE-T2*, CVs of 5.1 to 5.9% for T1 mapping and CVs of 8.1 to 11.4% for sodium imaging.

Test-retest reliability data were reported for the above techniques as well as DTI. Excellent test-retest reliability ICCs were demonstrated for T2* (0.93) and sodium imaging (0.91)[49,50]. Test-retest CVs were generally less than 10% except for sodium imaging which had test-retest CVs between 9.1 and 12.3%[32,43].

[FIGURE 2]

Synthesis of results – discriminative validity

Results of the meta-analysis of discriminative validity are presented in Table 3 and Figure 3. Individual forest plots are presented in the Supplemental Materials.

1. **Mild OA**

Both T2 and T1rho relaxometry demonstrated significant discrimination between subjects with mild OA and controls ($p < 0.001$), with a greater standardized mean difference (SMD) for T1rho (0.73, 95% CI 0.40 to 1.06) than for T2 (0.49, 0.30 to 0.67). dGEMRIC did not show significant discrimination for mild OA with a SMD of 0.13 (95% CI -0.23 to 0.49, $p = 0.47$). Single studies evaluating MT imaging in patellar cartilage and DTI imaging did discriminate significantly between subjects with mild OA and controls in some compartments[56,62].

2. **Severe OA**

Pooling of discrimination data was possible for T2 relaxometry only, which demonstrated significant discrimination between subjects with severe OA and controls ($p < 0.001$) with, as expected, a greater SMD of 1.24 (0.63 to 1.85) when compared to the mild OA data. Single studies evaluated discrimination validity for T1rho, dGEMRIC and MT, demonstrating significant differences between groups for T1rho and dGEMRIC but not MT.

3. OA (Not Otherwise Specified)

T2 relaxometry and T1rho relaxometry demonstrated significant differences between subjects with OA (NOS) and controls ($p < 0.001$). As for mild OA subjects, T1rho relaxometry had a higher SMD (0.60, 0.41 to 0.80) than T2 relaxometry (0.48, 0.34 to 0.62). dGEMRIC ($p = 0.18$, SMD = -0.31, -0.78 to 0.15) and sodium imaging ($p = 0.17$, SMD = -0.20, -0.50 to 0.09) did not significantly discriminate OA subjects from controls. Single studies of T2* relaxometry and DTI demonstrated significant discrimination in some compartments[49,57].

4. Additional analyses

When we restricted our analysis to studies that had control groups matched to OA subjects for age, we could pool data for T2 and T1rho relaxometry comparing OA (NOS) subjects with controls. Both techniques retained significant discrimination between subjects with OA and controls ($p < 0.001$), but with lower SMD values of 0.32 (0.20 to 0.44) for T2 and 0.34 (0.17 to 0.51) for T1rho.

[TABLE 3]

[FIGURE 3]

DISCUSSION

This systematic review has shown that cartilage compositional MRI techniques perform well across the domains of intra-observer, inter-observer and test-retest reliability. T2 and T1rho relaxometry demonstrated discrimination validity in mild OA and non-specific OA populations.

Reliability values were generally high across all techniques studied with ICC values in the 'excellent' range (> 0.8) and low CVs (< 5% in most cases). Reliability was slightly poorer for sodium imaging. This is probably explained by the fact that the images being analyzed are of lower signal-to-noise ratio (SNR) than proton images. The reliability values here are commensurate with those for established quantitative measures of joint structure such as cartilage volume and quantitative imaging biomarkers in other body systems[2,89,90]. This suggests that cartilage compositional MRI is suitably reliable for use in the assessment of knee OA, particularly for the techniques of T2 and T1rho relaxometry where there are most data available. Analyses performed on small cartilage subregions or involving laminar analysis (where cartilage is split into 2 or 3 layers from deep to superficial) tended to be less reliable than those assessing larger cartilage regions. This is probably due to the effects of noise and partial volume with adjacent synovial fluid and subchondral bone which are likely to be exacerbated in small regions of interest (ROIs), together with increased scope for subjective positioning differences between observers. This should be borne in mind when such analyses are interpreted. Few studies examined test-retest reliability in a multi-center setting. To facilitate use of compositional techniques in large-scale clinical trials, more studies in this area are needed.

T1rho and T2 relaxometry demonstrated the ability to discriminate between subjects with mild OA and controls, and subjects with OA (NOS) and controls. T1rho demonstrated larger SMD values in both populations suggesting that it has superior discrimination validity, in keeping with the results of previous *in vivo* and *ex vivo* studies[66,84,91]. dGEMRIC and sodium imaging demonstrated smaller and non-significant SMDs between OA subjects and controls. This is contrasts with previous work showing better correlation between dGEMRIC and glycosaminoglycan (GAG, the main polysaccharide side chains of the proteoglycan molecules) content of articular cartilage than between T1rho values and GAG content[92]. Possible reasons for the poorer performance of dGEMRIC in this meta-analysis include variation in imaging protocols between studies which may have affected results[9392].

For all techniques studied, there was significant statistical heterogeneity between different cartilage regions. This concurs with previous work demonstrating significant spatial variation in articular cartilage compositional values at the knee, and potentially suggests that changes in cartilage composition due to OA also show substantial spatial variation[70,76,94].

The principal role for cartilage compositional MRI is the detection of adverse changes in cartilage composition prior to morphological damage. Therefore, they are of greatest potential utility in subjects with mild OA and are of questionable value once the disease is more advanced[95]. SMD values for T1rho and T2 relaxometry, which showed significant discrimination validity for mild OA population, correspond to relatively small absolute differences between mild OA subjects and controls of 3.5 ms (95% CI 1.9 – 5.2 ms) for T1rho and 1.9 ms (1.2 – 2.7 ms) for T2. Although statistically significant differences have been demonstrated between groups in this meta-analysis, it is

questionable whether these differences are of of clinical significance, that is of sufficient magnitude to provide useful clinical discrimination when interpreting a single measurement. Moreover, these small differences may be exaggerated because few control groups were matched to OA subjects for important characteristics such as age or sex.

This study has some unavoidable limitations. First, we have considered reliability and discrimination validity of cartilage compositional MRI, but not responsiveness to change. This will be an important factor to consider when using cartilage compositional MRI as an imaging biomarker of response to treatment, and from a clinical utility point of view it may be that the magnitude of intra-subject change is more important than absolute mean differences between OA and control groups. However, at present, there have not been sufficient studies in this area to permit pooled analysis. Moreover, the majority of studies which have used cartilage compositional MRI in a longitudinal setting do not report sufficient data to allow calculation of standardized response means for pooling. Second, the quality of included studies was variable for both reliability and discriminative validity studies. The reliability values reported across different studies were consistent suggesting that substantial bias affecting the results of this part of the review had not been introduced by these factors. However, discriminative validity results did vary between studies, as indicated by moderate heterogeneity in the meta-analyses.

In conclusion, this systematic review and meta-analysis has demonstrated that cartilage compositional MRI techniques are reliable and, in the case of T2 and T1rho relaxometry, able to discriminate between subjects with OA and controls.

ACKNOWLEDGEMENTS

The research team acknowledges the support of the National Institute for Health Research Cambridge Biomedical Research Centre. We thank all corresponding authors who responded to requests for data. No writing assistance was used.

AUTHOR CONTRIBUTIONS

Conception and design:	JM, TS, AT, AM, FG
Analysis and interpretation of the data:	JM, SL
Drafting of the article:	JM, SL, TS, AT, AM, FG
Critical revision of the article:	JM, SL, TS, AT, AM, FG
Final approval of the article:	JM, SL, TS, AT, AM, FG
Provision of study materials or patients:	N/A
Statistical expertise:	JM, TS
Obtaining of funding:	JM, AM, FG
Administrative, technical, or logistic support:	SL, TS, AT, AM, FG
Collection and assembly of data:	JM, SL

Guarantors of entire study: Dr James MacKay (jwm37@cam.ac.uk), Professor Fiona Gilbert (fjg28@medschl.cam.ac.uk)

DECLARATION OF FUNDING AND ROLE OF FUNDING SOURCE

This study was funded by the Addenbrooke's Charitable Trust via a Clinical Research Training Fellowship grant. The funder had no role in the study design, data collection and analysis, preparation of the manuscript or the decision to submit the manuscript for publication.

COMPETING INTERESTS STATEMENT

The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

REFERENCES

1. Guermazi A, Alizai H, Crema MD, Trattnig S, Regatte RR, Roemer FW. Compositional MRI techniques for evaluation of cartilage degeneration in osteoarthritis. *Osteoarthritis Cartilage*. 2015;23(10):1639-1653. doi:10.1016/j.joca.2015.05.026.

2. Hunter DJ, Zhang W, Conaghan PG, et al. Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence. *Osteoarthritis Cartilage*. 2011;19(5):589-605. doi:10.1016/j.joca.2010.10.030.

3. Hunter DJ, Zhang W, Conaghan PG, et al. Systematic review of the concurrent and predictive validity of MRI biomarkers in OA. *Osteoarthritis Cartilage*. 2011;19(5):557-588. doi:10.1016/j.joca.2010.10.029.

4. Reichmann WM, Maillefert JF, Hunter DJ, Katz JN, Conaghan PG, Losina E. Responsiveness to change and reliability of measurement of radiographic joint space width in osteoarthritis of the knee: a systematic review. *Osteoarthritis Cartilage*. 2011;19(5):550-556. doi:10.1016/j.joca.2011.01.023.

5. Hunter DJ, Altman RD, Cicuttini F, et al. OARSI Clinical Trials Recommendations: Knee imaging in clinical trials in osteoarthritis. *Osteoarthritis Cartilage*. 2015;23(5):698-715. doi:10.1016/j.joca.2015.03.012.

6. Sullivan DC, Obuchowski NA, Kessler LG, et al. Metrology Standards for Quantitative Imaging Biomarkers. *Radiology*. 2015;277(3):813-825. doi:10.1148/radiol.2015142202.

7. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P, others. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.

8. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis*. 1957;16(4):494-502.

9. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol*. 2010;63(8):854-861. doi:10.1016/j.jclinepi.2009.10.002.

10. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536.

11. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428.

12. Obuchowski NA, Barnhart HX, Buckler AJ, et al. Statistical issues in the comparison of quantitative imaging biomarker algorithms using pulmonary nodule volume as an example. *Stat Methods Med Res*. 2015;24(1):107-140. doi:10.1177/0962280214537392.

13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.

14. The Nordic Cochrane Centre. *Review Manager (RevMan)*. Copenhagen: The Cochrane Collaboration; 2014.

15. Anandacoomarasamy A, Leibman S, Smith G, et al. Weight loss in obese people has structure-modifying effects on medial but not on lateral knee articular cartilage. *Ann Rheum Dis*. 2012;71(1):26-32. doi:10.1136/ard.2010.144725.

16. Balamoody S, Williams TG, Wolstenholme C, et al. Magnetic resonance transverse relaxation time T2 of knee cartilage in osteoarthritis at 3-T: a cross-sectional multicentre, multivendor reproducibility study. *Skeletal Radiol*. 2013;42(4):511-520. doi:10.1007/s00256-012-1511-5.

17. Baum T, Joseph GB, Arulanandan A, et al. Association of magnetic resonance imaging–based knee cartilage T2 measurements and focal knee lesions with knee pain: Data from the Osteoarthritis Initiative. *Arthritis Care Res*. 2012;64(2):248-255. doi:10.1002/acr.20672.

18. Blumenkrantz G, Lindsey CT, Dunn TC, et al. A pilot, two-year longitudinal study of the interrelationship between trabecular bone and articular cartilage in the osteoarthritic knee. *Osteoarthritis Cartilage*. 2004;12(12):997-1005. doi:10.1016/j.joca.2004.09.001.

19. Bron EE, van Tiel J, Smit H, et al. Image registration improves human knee cartilage T1 mapping with delayed gadolinium-enhanced MRI of cartilage (dGEMRIC). *Eur Radiol*. 2013;23(1):246-252. doi:10.1007/s00330-012-2590-3.

20. Carballido-Gamio J., Link T.M., Majumdar S. New techniques for cartilage magnetic resonance imaging relaxation time analysis: Texture analysis of flattened cartilage and localized intra- and inter-subject comparisons. *Magn Reson Med*. 2008;59(6):1472-1477.

21. Dardzinski BJ, Schneider E. Radiofrequency (RF) coil impacts the value and reproducibility of cartilage spin-spin (T2) relaxation time measurements. *Osteoarthritis Cartilage*. 2013;21(5):710-720. doi:10.1016/j.joca.2013.01.006.

22. Duryea J, Cheng C, Schaefer LF, Smith S, Madore B. Integration of accelerated MRI and post-processing software: a promising method for studies of knee osteoarthritis. *Osteoarthritis Cartilage*. 2016;24(11):1905-1909. doi:10.1016/j.joca.2016.06.001.

23. Guha A, Wyatt C, Karampinos DC, Nardo L, Link TM, Majumdar S. Spatial variations in magnetic resonance-based diffusion of articular cartilage in knee osteoarthritis. *Magn Reson Imaging*. 2015;33(9):1051-1058. doi:10.1016/j.mri.2015.06.004.

24. Gupta R, Virayavanich W, Kuo D, et al. MR T(1)ρ quantification of cartilage focal lesions in acutely injured knees: correlation with arthroscopic evaluation. *Magn Reson Imaging*. 2014;32(10):1290-1296. doi:10.1016/j.mri.2014.07.015.

25. Hada S, Kaneko H, Sadatsuki R, et al. The degeneration and destruction of femoral articular cartilage shows a greater degree of deterioration than that of the tibial and patellar articular cartilage in early stage knee osteoarthritis: a cross-sectional study. *Osteoarthritis Cartilage*. 2014;22(10):1583-1589. doi:10.1016/j.joca.2014.07.021.

26. Hannila I, Räinä SS, Tervonen O, Ojala R, Nieminen MT. Topographical variation of T2 relaxation time in the young adult knee cartilage at 1.5 T. *Osteoarthritis Cartilage*. 2009;17(12):1570-1575. doi:10.1016/j.joca.2009.05.011.

27. Hannila I, Lammentausta E, Tervonen O, Nieminen MT. The repeatability of T2 relaxation time measurement of human knee articular cartilage. *Magn Reson Mater Phys Biol Med*. 2015;28(6):547-553. doi:10.1007/s10334-015-0494-3.

28. Hesper T, Miese FR, Hosalkar HS, et al. Quantitative T2(*) assessment of knee joint cartilage after running a marathon. *Eur J Radiol*. 2015;84(2):284-289. doi:10.1016/j.ejrad.2014.11.021.

29. Holtzman DJ, Theologis AA, Carballido-Gamio J, Majumdar S, Li X, Ma CB. T1ρ and T2 Quantitative Magnetic Resonance Imaging Analysis of Cartilage Regeneration Following Microfracture and Mosaicplasty Cartilage Resurfacing Procedures. *J Magn Reson Imaging*. 2010;32(4):914-923. doi:10.1002/jmri.22300.

30. Hovis KK, Stehling C, Souza RB, et al. Physical activity is associated with magnetic resonance imaging-based knee cartilage T2 measurements in asymptomatic subjects with and those without osteoarthritis risk factors. *Arthritis Rheum*. 2011;63(8):2248-2256. doi:10.1002/art.30419.

31. Hovis KK, Alizai H, Tham S-C, et al. Non-traumatic anterior cruciate ligament abnormalities and their relationship to osteoarthritis using morphological grading and cartilage T2 relaxation times: data from the Osteoarthritis Initiative (OAI). *Skeletal Radiol*. 2012;41(11):1435-1443. doi:10.1007/s00256-012-1379-4.

32. Jordan CD, McWalter EJ, Monu UD, et al. Variability of CubeQuant T1ρ, quantitative DESS T2, and cones sodium MRI in knee cartilage. *Osteoarthritis Cartilage*. 2014;22(10):1559-1567. doi:10.1016/j.joca.2014.06.001.

33. Joseph GB, Baum T, Carballido-Gamio J, et al. Texture analysis of cartilage T 2 maps: individuals with risk factors for OA have higher and more heterogeneous knee cartilage MR T 2 compared to normal controls-data from the osteoarthritis initiative. *Arthritis Res Ther*. 2011;13(5):1.

34. Joseph GB, Baum T, Alizai H, et al. Baseline mean and heterogeneity of MR cartilage T2 are associated with morphologic degeneration of cartilage, meniscus, and bone marrow over 3 years – data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage*. 2012;20(7):727-735. doi:10.1016/j.joca.2012.04.003.

35. Juras V, Bohndorf K, Heule R, et al. A comparison of multi-echo spin-echo and triple-echo steady-state T2 mapping for in vivo evaluation of articular cartilage. *Eur Radiol*. 2016;26:1905-1912. doi:10.1007/s00330-015-3979-6.

36. Koli J, Multanen J, Kujala UM, et al. Effects of Exercise on Patellar Cartilage in Women with Mild Knee Osteoarthritis. *Med Sci Sports Exerc*. 2015;47(9):1767-1774. doi:10.1249/MSS.0000000000000629.

37. Li X, Wyatt C, Rivoire J, et al. Simultaneous acquisition of T1ρ and T2 quantification in knee cartilage: Repeatability and diurnal variation. *J Magn Reson Imaging*. 2014;39(5):1287-1293. doi:10.1002/jmri.24253.

38. Li X, Han ET, Ma CB, Link TM, Newitt DC, Majumdar S. In vivo 3T spiral imaging based multi-slice T(1rho) mapping of knee cartilage in osteoarthritis. *Magn Reson Med*. 2005;54(4):929-936.

39. Li X, Pedoia V, Kumar D, et al. Cartilage T1ρ and T2 relaxation times: longitudinal reproducibility and variations using different coils, MR systems and sites. *Osteoarthritis Cartilage*. 2015;23(12):2214-2223. doi:10.1016/j.joca.2015.07.006.

40. Liebl H, Joseph G, Nevitt MC, et al. Early T2 changes predict onset of radiographic knee osteoarthritis: data from the osteoarthritis initiative. *Ann Rheum Dis*. 2015;74(7):1353-1359.

41. Liess C, Lüsse S, Karger N, Heller M, Glüer C-C. Detection of changes in cartilage water content using MRI T2-mapping in vivo. *Osteoarthritis Cartilage*. 2002;10(12):907-913. doi:10.1053/joca.2002.0847.

42. Liu F, Choi KW, Samsonov A, et al. Articular Cartilage of the Human Knee Joint: In Vivo Multicomponent T2 Analysis at 3.0 T. *Radiology*. 2015;277(2):477-488. doi:10.1148/radiol.2015142201.

43. Madelin G, Babb JS, Xia D, Chang G, Jerschow A, Regatte RR. Reproducibility and repeatability of quantitative sodium magnetic resonance imaging in vivo in articular cartilage at 3 T and 7 T. *Magn Reson Med*. 2012;68(3):841-849. doi:10.1002/mrm.23307.

44. Matsubara H, Okazaki K, Takayama Y, et al. Detection of early cartilage deterioration associated with meniscal tear using T1ρ mapping magnetic resonance imaging. *BMC Musculoskelet Disord*. 2015;16:22. doi:10.1186/s12891-015-0487-4.

45. Mosher TJ, Zhang Z, Reddy R, et al. Knee Articular Cartilage Damage in Osteoarthritis: Analysis of MR Image Biomarker Reproducibility in ACRIN-PA 4001 Multicenter Trial. *Radiology*. 2011;258(3):832-842. doi:10.1148/radiol.10101174.

46. Mosher TJ, Liu Y, Yang QX, et al. Age dependency of cartilage magnetic resonance imaging T2 relaxation times in asymptomatic women. *Arthritis Rheum*. 2004;50(9):2820-2828. doi:10.1002/art.20473.

47. Multanen J, Rauvala E, Lammentausta E, et al. Reproducibility of imaging human knee cartilage by delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) at 1.5 Tesla. *Osteoarthritis Cartilage*. 2009;17(5):559-564. doi:10.1016/j.joca.2008.12.001.

48. Multanen J, Heinonen A, Häkkinen A, et al. Bone and cartilage characteristics in postmenopausal women with mild knee radiographic osteoarthritis and those without radiographic osteoarthritis. *J Musculoskelet Neuronal Interact*. 2015;15(1):69-77.

49. Newbould RD, Miller SR, Toms LD, et al. T2* measurement of the knee articular cartilage in osteoarthritis at 3T. *J Magn Reson Imaging*. 2012;35(6):1422-1429.

50. Newbould RD, Miller SR, Tielbeek JAW, et al. Reproducibility of sodium MRI measures of articular cartilage of the knee in osteoarthritis. *Osteoarthritis Cartilage*. 2012;20(1):29-35. doi:10.1016/j.joca.2011.10.007.

51. Nishioka H, Hirose J, Nakamura E, et al. T1ρ and T2 mapping reveal the in vivo extracellular matrix of articular cartilage. *J Magn Reson Imaging*. 2012;35(1):147-155. doi:10.1002/jmri.22811.

52. Nishioka H, Hirose J, Nakamura E, et al. Detecting ICRS grade 1 cartilage lesions in anterior cruciate ligament injury using T1ρ and T2 mapping. *Eur J Radiol*. 2013;82(9):1499-1505. doi:10.1016/j.ejrad.2013.04.038.

53. Nishioka H, Hirose J, Okamoto N, et al. Evaluation of the relationship between T1ρ and T2 values and patella cartilage degeneration in patients of the same age group. *Eur J Radiol*. 2015;84(3):463-468. doi:10.1016/j.ejrad.2014.11.015.

54. Pan J, Pialat J-B, Joseph T, et al. Knee cartilage T2 characteristics and evolution in relation to morphologic abnormalities detected at 3-T MR imaging: a longitudinal study of the normal control cohort from the Osteoarthritis Initiative. *Radiology*. 2011;261(2):507-515. doi:10.1148/radiol.11102234.

55. Pedoia V, Li X, Su F, Calixto N, Majumdar S. Fully automatic analysis of the knee articular cartilage T1ρ relaxation time using voxel-based relaxometry. *J Magn Reson Imaging*. 2016;43(4):970-980. doi:10.1002/jmri.25065.

56. Raya JG, Dettmann E, Notohamiprodjo M, Krasnokutsky S, Abramson S, Glaser C. Feasibility of in vivo diffusion tensor imaging of articular cartilage with coverage of all cartilage regions. *Eur Radiol*. 2014;24(7):1700-1706. doi:10.1007/s00330-014-3155-4.

57. Raya JG, Horng A, Dietrich O, et al. Articular Cartilage: In Vivo Diffusion-Tensor Imaging. *Radiology*. 2012;262(2):550-559. doi:10.1148/radiol.11110821.

58. Schleich C, Bittersohl B, Miese F, et al. Glycosaminoglycan chemical exchange saturation transfer at 3T MRI in asymptomatic knee joints. *Acta Radiol*. 2016;57(5):627-632. doi:10.1177/0284185115598811.

59. Serebrakian AT, Poulos T, Liebl H, et al. Weight loss over 48 months is associated with reduced progression of cartilage T2 relaxation time values: Data from the osteoarthritis initiative. *J Magn Reson Imaging*. 2015;41(5):1272-1280. doi:10.1002/jmri.24630.

60. Singh A, Haris M, Cai K, Kogan F, Hariharan H, Reddy R. High resolution T1ρ mapping of in vivo human knee cartilage at 7T. *PloS One*. 2014;9(5):e97486. doi:10.1371/journal.pone.0097486.

61. Siversson C, Tiderius C-J, Neuman P, Dahlberg L, Svensson J. Repeatability of T1-quantification in dGEMRIC for three different acquisition techniques: two-dimensional inversion recovery, three-dimensional look locker, and three-dimensional variable flip angle. *J Magn Reson Imaging*. 2010;31(5):1203-1209. doi:10.1002/jmri.22159.

62. Sritanyaratana N, Samsonov A, Mossahebi P, Wilson JJ, Block WF, Kijowski R. Cross-relaxation imaging of human patellar cartilage in vivo at 3.0T. *Osteoarthritis Cartilage*. 2014;22(10):1568-1576. doi:10.1016/j.joca.2014.06.004.

63. Stehling C, Baum T, Mueller-Hoecker C, et al. A novel fast knee cartilage segmentation technique for T2 measurements at MR imaging--data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage*. 2011;19(8):984-989. doi:10.1016/j.joca.2011.04.002.

64. Stehling C, Liebl H, Krug R, et al. Patellar Cartilage: T2 Values and Morphologic Abnormalities at 3.0-T MR Imaging in Relation to Physical Activity in Asymptomatic Subjects from the Osteoarthritis Initiative. *Radiology*. 2010;254(2):509-520. doi:10.1148/radiol.09090596.

65. Surowiec RK, Lucas EP, Fitzcharles EK, et al. T2 values of articular cartilage in clinically relevant subregions of the asymptomatic knee. *Knee Surg Sports Traumatol Arthrosc*. 2014;22(6):1404-1414. doi:10.1007/s00167-013-2779-2.

66. Takayama Y, Hatakenaka M, Tsushima H, et al. T1p is superior to T2 mapping for the evaluation of articular cartilage denaturalization with osteoarthritis: radiological-pathological correlation after total knee arthroplasty. *Eur J Radiol*. 2013;82(4):e192-8.

67. Tiderius CJ, Tjörnstrand J, Åkeson P, Södersten K, Dahlberg L, Leander P. Delayed gadolinium-enhanced MRI of cartilage (dGEMRIC): intra- and interobserver variability in standardized drawing of regions of interest. *Acta Radiol*. 2004;45(6):628-634. doi:10.1080/02841850410008379.

68. van Tiel J, Bron EE, Tiderius CJ, et al. Reproducibility of 3D delayed gadolinium enhanced MRI of cartilage (dGEMRIC) of the knee at 3.0 T in patients with early stage osteoarthritis. *Eur Radiol*. 2013;23(2):496-504. doi:10.1007/s00330-012-2616-x.

69. Welsch GH, Apprich S, Zbyn S, et al. Biochemical (T2, T2* and magnetisation transfer ratio) MRI of knee cartilage: feasibility at ultra-high field (7T) compared with high field (3T) strength. *Eur Radiol*. 2010;21(6):1136-1143. doi:10.1007/s00330-010-2029-7.

70. Wiener E, Pfirrmann CWA, Hodler J. Spatial variation in T1 of healthy human articular cartilage of the knee joint. *Br J Radiol*. 2010;83(990):476-485. doi:10.1259/bjr/62779246.

71. Williams A, Qian Y, Chu CR. UTE-T2* mapping of human articular cartilage in vivo: a repeatability assessment. *Osteoarthritis Cartilage*. 2011;19(1):84-88. doi:10.1016/j.joca.2010.10.018.

72. Zuo J, Li X, Banerjee S, Han E, Majumdar S. Parallel imaging of knee cartilage at 3 Tesla. *J Magn Reson Imaging*. 2007;26(4):1001-1009.

73. Dunn TC, Lu Y, Jin H, Ries MD, Majumdar S. T2 Relaxation Time of Cartilage at MR Imaging: Comparison with Severity of Knee Osteoarthritis1. *Radiology*. 2004;232(2):592-598. doi:10.1148/radiol.2322030976.

74. Eckstein F, Le Graverand MPH, Charles HC, et al. Clinical, radiographic, molecular and MRI-based predictors of cartilage loss in knee osteoarthritis. *Ann Rheum Dis*. 2011;70(7):1223-1230. doi:10.1136/ard.2010.141382.

75. Li W, Scheidegger R, Wu Y, et al. Delayed contrast-enhanced MRI of cartilage: Comparison of nonionic and ionic contrast agents. *Magn Reson Med*. 2010;64(5):1267-1273. doi:10.1002/mrm.22555.

76. Li X, Pai A, Blumenkrantz G, et al. Spatial distribution and relationship of T1rho and T2 relaxation times in knee cartilage with osteoarthritis. *Magn Reson Med*. 2009;61(6):1310-1318.

77. Li X, Ma CB, Link TM, et al. In vivo T1rho and T2 mapping of articular cartilage in osteoarthritis of the knee using 3 Tesla MRI. *Osteoarthritis Cartilage*. 2007;15(7):789-797. doi:10.1016/j.joca.2007.01.011.

78. Madelin G, Babb J, Xia D, et al. Articular cartilage: evaluation with fluid-suppressed 7.0-T sodium MR imaging in subjects with and subjects without osteoarthritis. *Radiology*. 2013;268(2):481-491. doi:10.1148/radiol.13121511.

79. Owman H, Ericsson YB, Englund M, et al. Association between delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) and joint space narrowing and osteophytes: a cohort study in patients with partial meniscectomy with 11 years of follow-up. *Osteoarthritis Cartilage*. 2014;22(10):1537-1541. doi:10.1016/j.joca.2014.02.929.

80. Souza RB, Kumar D, Calixto N, et al. Response of knee cartilage T1rho and T2 relaxation times to in vivo mechanical loading in individuals with and without knee osteoarthritis. *Osteoarthritis Cartilage*. 2014;22(10):1367-1376. doi:10.1016/j.joca.2014.04.017.

81. Souza RB, Feeley BT, Zarins ZA, Link TM, Li X, Majumdar S. T1rho MRI relaxation in knee OA subjects with varying sizes of cartilage lesions. *The Knee*. 2013;20(2):113-119. doi:10.1016/j.knee.2012.10.018.

82. Stahl R, Blumenkrantz G, Carballido-Gamio J, et al. MRI-derived T2 relaxation times and cartilage morphometry of the tibio-femoral joint in subjects with and without osteoarthritis during a 1-year follow-up. *Osteoarthritis Cartilage*. 2007;15(11):1225-1234. doi:10.1016/j.joca.2007.04.018.

83. Stahl R, Luke A, Li X, et al. T1rho, T2 and focal knee cartilage abnormalities in physically active and sedentary healthy subjects versus early OA patients--a 3.0-Tesla MRI study. *Eur Radiol*. 2009;19(1):132-143.

84. Wang L, Regatte RR. Quantitative mapping of human cartilage at 3.0T: parallel changes in T2, T1p, and dGEMRIC. *Acad Radiol*. 2014;21(4):463-471.

85. Wirth W, Maschek S, Roemer FW, Eckstein F. Layer-specific femorotibial cartilage T2 relaxation time in knees with and without early knee osteoarthritis: Data from the Osteoarthritis Initiative (OAI). *Sci Rep*. 2016;6:34202. doi:10.1038/srep34202.

86. Wyatt C, Guha A, Venkatachari A, et al. Improved differentiation between knees with cartilage lesions and controls using 7T relaxation time mapping. *J Orthop Transl*. 2015;3(4):197-204. doi:10.1016/j.jot.2015.05.003.

87. Newbould RD, Miller SR, Upadhyay N, et al. T1-weighted sodium MRI of the articulator cartilage in osteoarthritis: a cross sectional and longitudinal study. *PloS One*. 2013;8(8):e73067. doi:10.1371/journal.pone.0073067.

88. Yao W, Qu N, Lu Z, Yang S. The application of T1 and T2 relaxation time and magnetization transfer ratios to the early diagnosis of patellar cartilage osteoarthritis. *Skeletal Radiol*. 2009;38(11):1055-1062. doi:10.1007/s00256-009-0769-8.

89. Athelogou M, Kim HJ, Dima A, et al. Algorithm Variability in the Estimation of Lung Nodule Volume From Phantom CT Scans: Results of the QIBA 3A Public Challenge. *Acad Radiol*. 2016;23(8):940-952. doi:10.1016/j.acra.2016.02.018.

90. Ng CS, Raunig DL, Jackson EF, et al. Reproducibility of perfusion parameters in dynamic contrast-enhanced MRI of lung and liver tumors: effect on estimates of patient sample size in clinical trials and on individual patient responses. *AJR Am J Roentgenol*. 2010;194(2):W134-140. doi:10.2214/AJR.09.3116.

91. Regatte RR, Akella SVS, Lonner J h., Kneeland J b., Reddy R. T1ρ relaxation mapping in human osteoarthritis (OA) cartilage: Comparison of T1ρ with T2. *J Magn Reson Imaging*. 2006;23(4):547–553. doi:10.1002/jmri.20536.

92. van Tiel J, Kotek G, Reijman M, et al. Is T1ρ Mapping an Alternative to Delayed Gadolinium-enhanced MR Imaging of Cartilage in the Assessment of Sulphated Glycosaminoglycan Content in Human Osteoarthritic Knees? An in Vivo Validation Study. *Radiology*. 2016;279(2):523-531. doi:10.1148/radiol.2015150693.

93. Burstein D, Velyvis J, Scott KT, et al. Protocol issues for delayed Gd(DTPA)2--enhanced MRI (dGEMRIC) for clinical evaluation of articular cartilage. *Magn Reson Med*. 2001;45(1):36-41. doi:10.1002/1522-2594(200101)45:1<36::AID-MRM1006>3.0.CO;2-W.

94. Smith HE, Mosher TJ, Dardzinski BJ, et al. Spatial variation in cartilage T2 of the knee. *J Magn Reson Imaging*. 2001;14(1):50-55. doi:10.1002/jmri.1150.

95. Jungmann PM, Kraus MS, Nardo L, et al. T(2) relaxation time measurements are limited in monitoring progression, once advanced cartilage defects at the knee occur: longitudinal data from the osteoarthritis initiative. *J Magn Reson Imaging*. 2013;38(6):1415-1424.

FIGURE CAPTIONS

Figure 1. PRISMA flow chart demonstrating process for selection of included studies.

Figure 2. Summary of reliability data (top panel – coefficient of variation (%), bottom panel – intraclass correlation coefficient) for the three most commonly studied compositional technique. Central point represents median value, lines represent ranges. Where a study provided more than one estimate, a mean value is used.

Figure 3. Summary forest plot comparing standardized mean differences between subjects with and without OA for each technique where pooling of data was possible.

Supplemental figures

Supplemental Figure 1. Results of QAREL assessment of reliability studies

Supplemental Figure 2. Results of QUADAS-2 assessment of discrimination studies

Supplemental figure 3a. Forest plot for studies comparing T2 values in subjects with mild OA vs controls

Supplemental figure 3b. Forest plot for studies comparing T2 values in subjects with severe OA vs controls

Supplemental figure 3c. Forest plot for studies comparing T2 values in subjects with OA not otherwise specified (NOS) vs controls

Supplemental figure 4a. Forest plot for studies comparing T1rho values in subjects with mild OA vs controls

Supplemental figure 4b. Forest plot for studies comparing T1rho values in subjects with OA (NOS) vs controls

Supplemental figure 5a. Forest plot for studies comparing dGEMRIC values in subjects with mild OA vs controls

Supplemental figure 5b. Forest plot for studies comparing dGEMRIC values in subjects with OA (NOS) vs controls

Supplemental figure 6. Forest plot for studies comparing Sodium values in subjects with OA (NOS) vs controls

Figure 1

Click here to download high resolution image

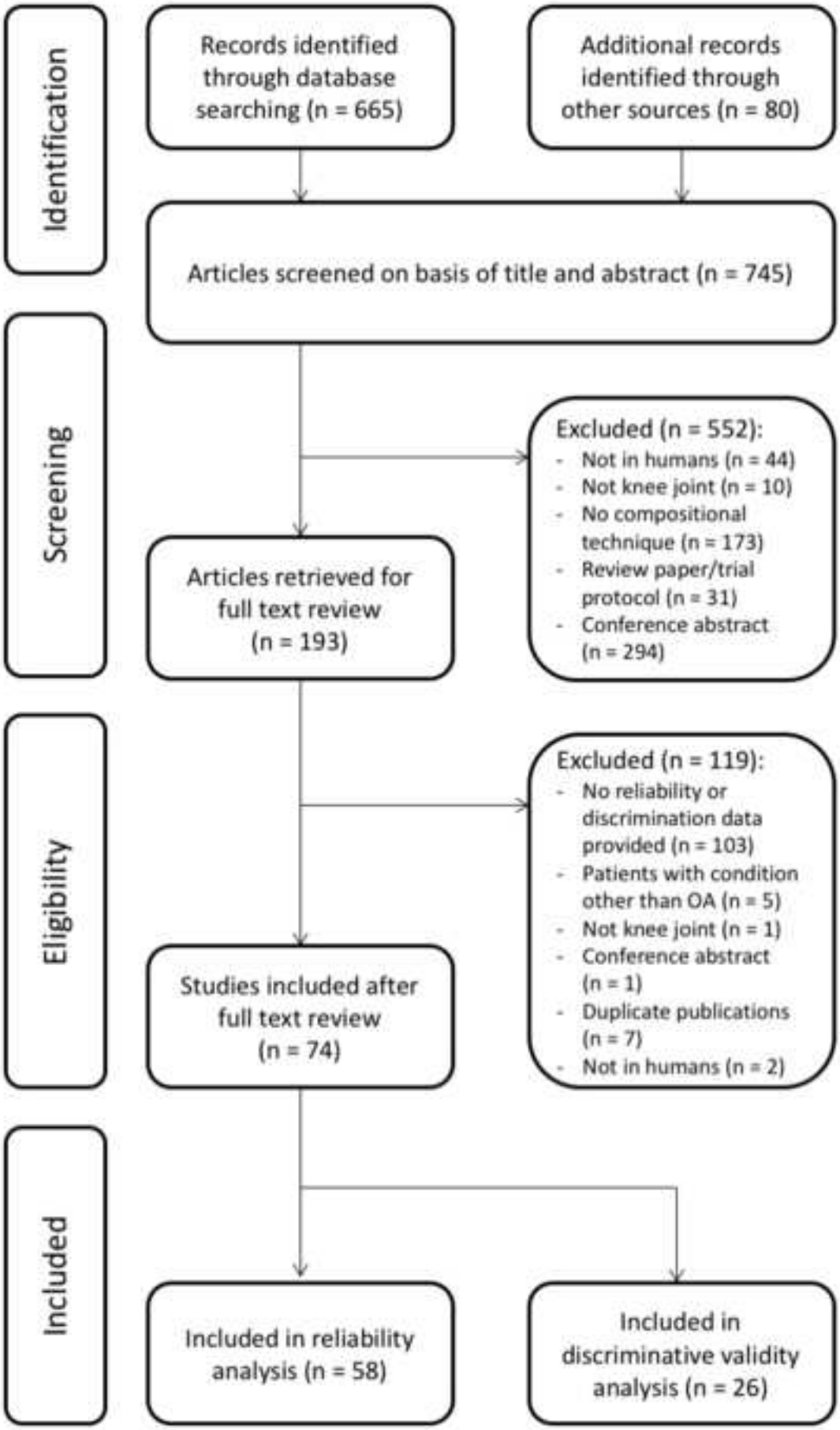


Figure 2

Click here to download high resolution image

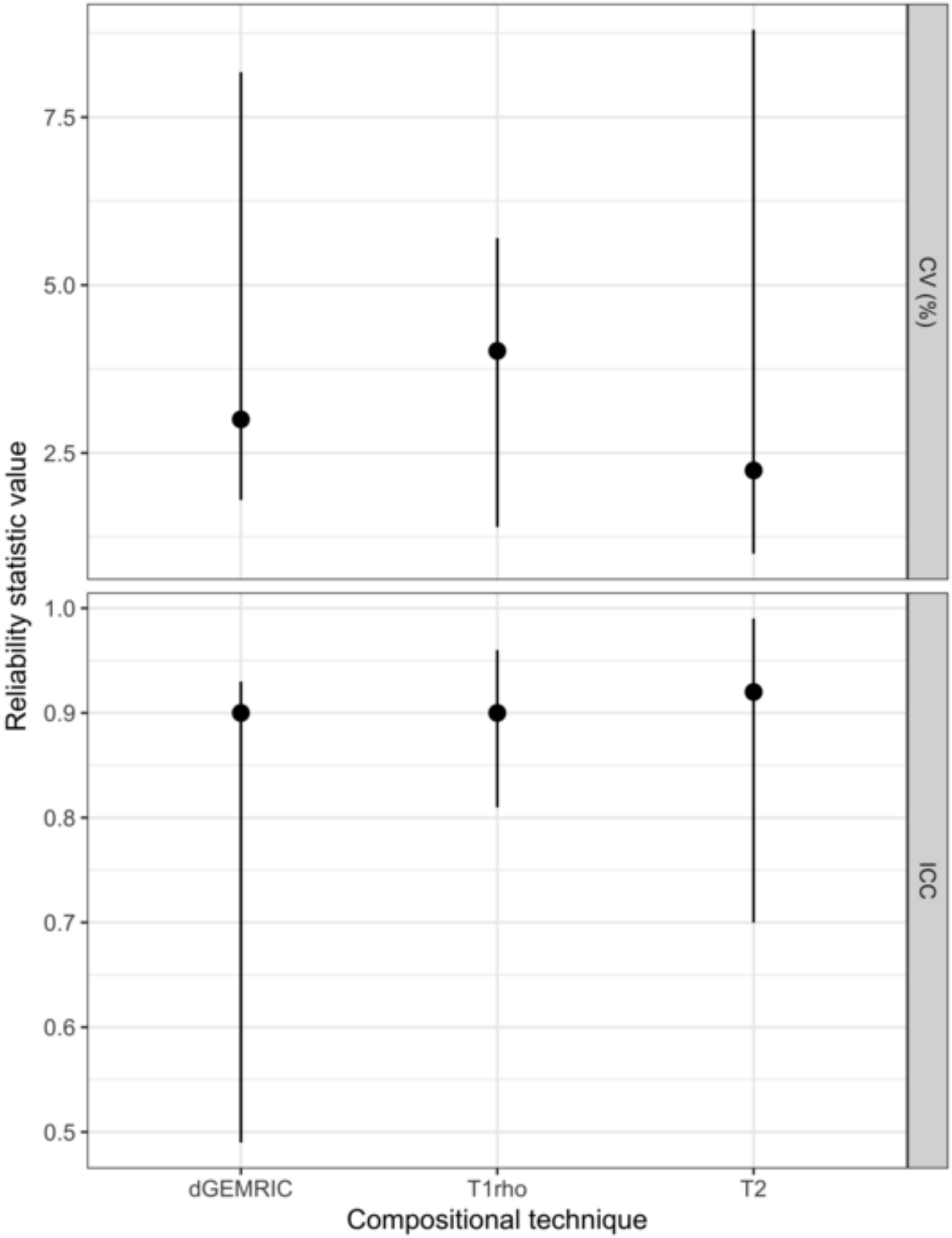
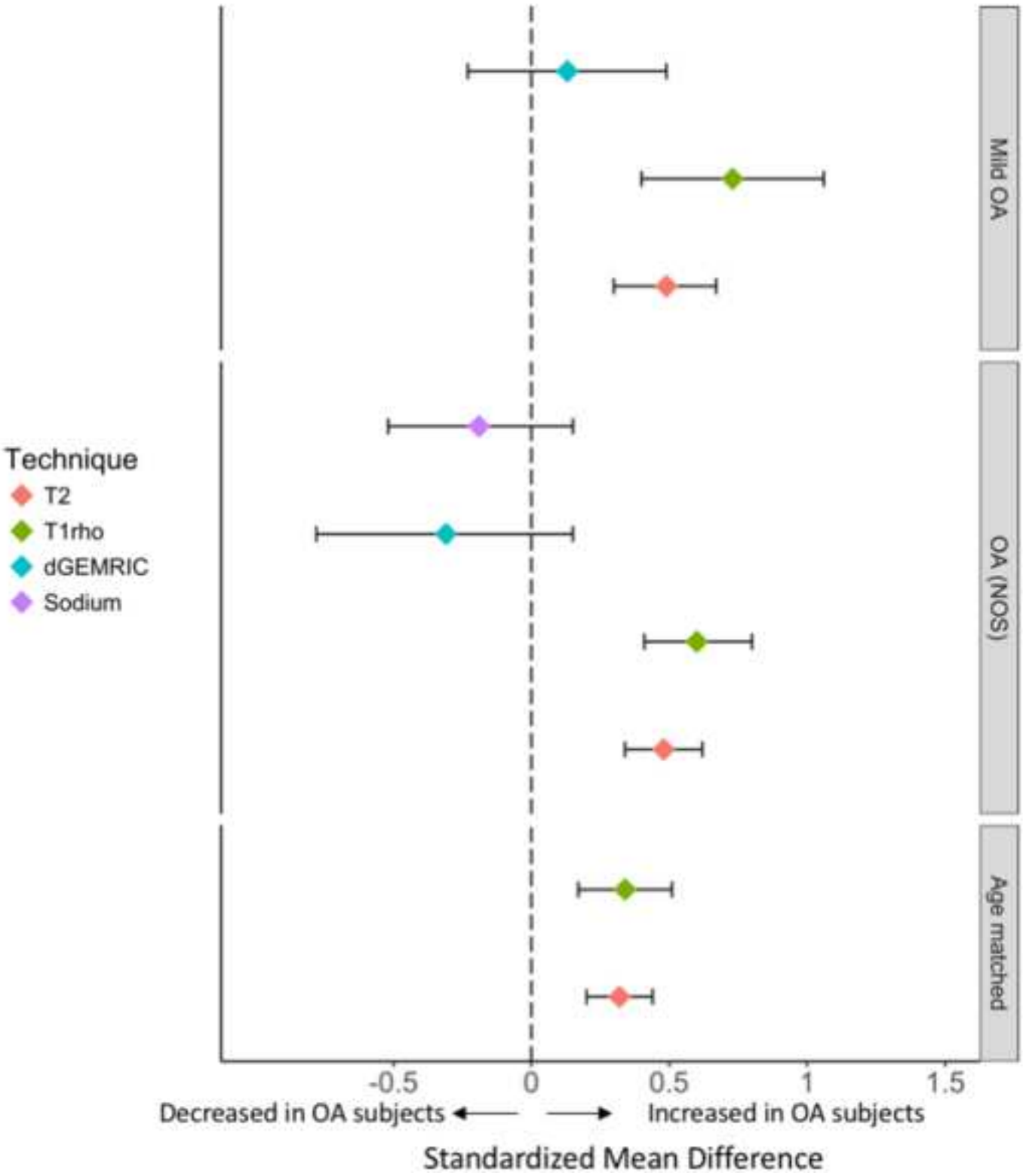


Figure 3

Click here to download high resolution image



TABLES

Table 1 - Characteristics of included reliability studies

Study	Number of OA subjects			Number of other subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Type of reliability assessed	Reliability statistic**	Reliability value	Assessors	
	mild	sev	NOS		OA	other	OA	other						Number	Experience
Anandacoomarasamy[15]	-	-	-	20	-	NS	-	NS	3	dGEMRIC	intraO	ICC	0.91	1	NS
Balamoody[16]	-	-	12	-	49	-	25	-	3	T2	TR[#]	RMSCV	2.3 – 6.3	1	NS
Baum[17]	-	-	-	126	-	50	-	50	3	T2	intraO	RMSCV	1.76	1	NS
Blumenkrantz[18]	13	17	-	-	64	-	39	-	1.5	T2	intraO	CV	1.5 – 2	1	NS
Bron[19]	-	-	11	10	52	27	64	40	3	dGEMRIC	TR	ICC	0.85 – 0.9	1	NS
Carballido-Gamio[20]	-	-	-	5	-	29	-	20	3	T1rho	TR	CV	3.8 – 12.3†	NS	NS
Dardzinski[21]	-	-	5	5	52		70		3	T2	TR	RMSCV	3.3 – 6.5	1	NS
Duryea[22]	-	-	-	10	-	33	-	50	3	T2	intraO TR	CV ICC CV ICC	1.3 0.99 5.8 0.92	1	NS
Guha[23]	-	-	20	20	58	54	50	60	3	DTI	TR	RMSCV	6.5 – 11.6	NS	NS
Gupta[24]	-	-	-	20	-	37	-	35	3	T1rho	interO	RMSCV ICC	3.9 0.96	2	NS
Hada[25]	50	-	-	19	57	25	54	37	3	T2	intraO interO	ICC ICC	0.91 0.87	2	NS
Hannila 2009[26]	-	-	-	20	-	23	-	50	1.5	T2	interO	RMSCV ICC	1.8 – 14.2† 0 – 0.98†	2	NS
Hannila 2015[27]	-	-	-	9	-	30	-	44	1.5	T2	TR	RMSCV ICC	2.5 – 22.2† 0 – 0.98†	1	NS
Hesper[28]	-	-	-	10	-	29	-	70	3	T2*	intraO	ICC	0.97	1	8 years

Study	Number of OA subjects			Number of other subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Type of reliability assessed	Reliability statistic**	Reliability value	Assessors	
	mild	sev	NOS		OA	other	OA	other						Number	Experience
Holtzman[29]	-	-	-	26	-	36	-	38	3	T1rho T2	interO	CV	4.3 4.9	2	At least 25 datasets
Hovis 2011[30]	-	-	-	161	-	50	-	59	3	T2	intraO	RMSE	0.74 – 1.51	1	NS
Hovis 2012[31]	-	-	105	-	65	-	68	-	3	T2	intraO interO	ICC	0.99 0.99	2	NS
Jordan[32]	-	-	-	8	-	28	-	25	3	T1rho T2 Sodium	intraO interO TR intraO interO TR intraO interO TR	RMSCV	3.8 5.7 4.6 – 6.1 4.7 6.7 6.3 – 10.7 8.1 11.4 11.3 – 12.9	2	NS
Joseph 2011[33]	-	-	-	145	-	50	-	59	3	T2	intraO	RMSCV	0.9 – 2.1	1	NS
Joseph 2012[34]	-	-	-	289	-	51	-	47	3	T2	intraO	RMSCV	0.8 – 3.2	1	NS
Juras 2016[35]	-	-	-	23	-	33	-	57	3	T2	interO	CV	5.8 – 10.8†	2	10 years/ 15 years
Koli[36]	80	-	-	-	58	-	100	-	1.5	T2	interO	RMSCV	2	NS	NS
X Li 2005[38]	-	-	9	10	52	30	44	40	3	T1rho	TR	CV	4.8	NS	NS
X Li 2014[37]	-	-	-	6	-	22-35	-	50	3	T1rho T2	TR	RMSCV	4.2 – 6 5 – 6.3	NS	NS
X Li 2015[39]	-	-	-	18	-	NS	-	NS	3	T1rho T2	TR[#]	RMSCV	2.3 – 5.1 3.2 – 5.3	NS	NS
Liebl[40]	-	-	-	130	-	59	-	60	3	T2	intraO interO	RMSCV	1.7 1.6	2	NS
Liess[41]	-	-	-	20	-	28	-	20	1.5	T2	TR	CV	1.7	NS	NS
Liu[42]	-	-	14	13	53	28	36	23	3	T2	TR	CV	3.1 - 10	1	4 years

Study	Number of OA subjects			Number of other subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Type of reliability assessed	Reliability statistic**	Reliability value	Assessors	
	mild	sev	NOS		OA	other	OA	other						Number	Experience
Madelin[43]	-	-	-	6		37	-	50	3, 7	Sodium	TR	RMSCV	9.1 – 12.3	NS	NS
Matsubara[44]	10	-	-	19	57	39	32	0	3	T1rho	interO	ICC	0.93	2	8 years/ 9 years
Mosher 2011[45]	16	16	-	18	54	27	59	67	3	T1rho T1rho T2 T2	TR[#]	ICC RMSCV ICC RMSCV	0.20 – 0.93† 7.23 – 18.83† 0.61 – 0.98† 4.36 – 14†	NS NS	NS Computer tutorial
Mosher 2004[46]	-	-	-	30	-	22-86	-	100	3	T2	intraO	wK	0.77	1	8 years
Multanen 2015[48]	78	-	-	12	59	58	100	100	1.5	T2 dGEMRIC	interO	RMSCV	2 3	2	6 years/ 12 years
Multanen 2009[47]	-	-	-	10	-	32	-	50	1.5	dGEMRIC	TR	ICC RMSCV	0.45 – 0.98† 4.7 – 12. 9†	NS	NS
Newbould 2012[49]	-	-	13	5	64	62	77	80	3	T2*	TR	ICC CV	0.7 – 0.94 3.2 – 7.7	NS	NS
Newbould 2012b[50]	-	-	15	5	64	62	80	80	3	Sodium	TR	ICC CV	0.67 – 0.94 3.6 – 9.9	1	NS
Nishioka 2012[51]	-	20	-	-	77	-	90	-	3	T1rho T2	interO	CV	< 6 < 6	NS	NS
Nishioka 2015[53]	-	-	78	-	69	-	77	-	3	T1rho T2	interO	RMSCV	11.8 12.2	2	NS
Nishioka 2013[52]	-	-	-	37	-	23	-	59	3	T1rho T2	interO	RMSCV	1.4 1.0	2	NS
Pan[54]	-	-	-	95	-	55	-	61	3	T2	intraO	RMSCV	1.1 – 1.2	1	NS
Pedoia[55]	40	-	-	15	54	48	NS	NS	3	T1rho	TR	CV	2	2	2 years/ 4 years
Raya 2014[56]	5	-	-	10	66	31	NS	30	7	DTI	TR	RMSCV	2.9 – 5.6	NS	NS
Raya 2012[57]	-	-	10	16	61	31	50	44	7	DTI T2	TR	RMSCV	7.3 – 19.3† 5.5 – 6.9†	1	5 years

Study	Number of OA subjects			Number of other subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Type of reliability assessed	Reliability statistic**	Reliability value	Assessors	
	mild	sev	NOS		OA	other	OA	other						Number	Experience
Schleich[58]	-	-	-	20	-	25	-	40	3	gagCEST	intraO interO	ICC	0.95 0.95	1 2	8 years 5 years/ 8 years
Serebrakian[59]	-	-	-	127	-	55	-	56	3	T2	intraO interO	RMSCV	1.1 3.3	2	NS
Singh[60]	-	-	-	8	-	20-35	-	NS	7	T1rho	TR	ICC CV	0.73 – 0.96 2.3 – 4.3	NS	NS
Siversson[61]	-	-	-	9	-	45	-	56	1.5	dGEMRIC	TR	ICC RMSCV	0 – 0.69 6.0 – 11.6	NS	NS
Sritanyaratana[62]	11	-	-	20	53	32	36	25	3	MT	TR	CV	0.5 – 4.6	1	NS
Stehling 2010[64]	-	-	-	120	-	51	-	50	3	T2	intraO	CV	1.2	1	NS
Stehling 2011[63]	-	-	-	10	-	52	-	50	3	T2	interO	RMSCV	1.2 – 2.8†	2	NS
Surowiec[65]	-	-	-	18	-	18-35	-	NS	3	T2	intraO interO	ICC	0.17 – 0.89 0.3 – 0.96	3	5 years/ 6 years/ 13 years
Takayama[66]	-	16	-	-	73	-	88	-	3	T1rho T2	interO	ICC	0.81 0.92	2	12 years / 7 years
Tiderius[67]	-	-	-	12	-	24	-	0	1.5	dGEMRIC	intraO interO	CV	1.5 – 2.6 1.8	6	2 medical students, 2 ortho surgeons, 2 radiologists
Van Tiel[68]	17	-	-	-	50	-	41	-	3	dGEMRIC	TR	ICC	0.87 – 0.95	1	Medical degree

Study	Number of OA subjects			Number of other subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Type of reliability assessed	Reliability statistic**	Reliability value	Assessors	
	mild	sev	NOS		OA	other	OA	other						Number	Experience
Welsch[69]	-	-	-	17	-	26	-	24	3, 7	T2	interO	ICC	0.91 – 0.95†	3	2 years/ 10 years/ 25 years
											TR	CV	7.2 – 8.7†		
										T2*	interO	ICC	0.88 – 0.90†		
											TR	CV	6.8 – 7.8†		
										MT	interO	ICC	0.84 – 0.91†		
											TR	CV	9.2 – 10.8†		
Wiener[70]	-	-	-	25	-	31	-	60	1.5	T1	TR	CV	5.1 – 5.9	1	NS
Williams[71]	-	-	-	11	-	28	-	100	3	UTE-T2*	intraO	ICC	0.80 – 0.97†	1	9 years
											TR	RMSCV	6 - 16†		
Zuo[72]	-	-	-	6	-	27-30	-	17	3	T1rho	TR	CV	1.8 – 5.6	NS	NS
										T2			2.9 – 5.7		

Abbreviations: sev - severe, NS – not specified, intraO – intra-observer, interO – inter-observer, TR – test-retest.
**Reliability values are presented as ranges when values were provided separately for different cartilage ROIs. RMSCV and CV values are provided as percentages.
†Laminar (e.g superficial/deep cartilage layers) analysis performed.
[#] Multicenter study

Table 2 - Characteristics of included discrimination studies

Study	Number of OA subjects			Number of control subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Definition of OA
	mild	sev	NOS		OA	control	OA	control			
Dunn[73]	20	28	-	7	63/67*	38	65/50*	43	1.5	T2	Radiographs and symptoms
Eckstein[74]	-	-	77	75	58	55	100	100	3	T2, dGEMRIC	Radiographs only
Hada[25]	50	-	-	19	57	25	54	37	3	T2	ACR criteria
X Li 2005[38]	-	-	9	10	52	30	44	40	3	T1rho	Radiographs and/or symptoms
X Li 2007[77]	-	-	10	16	56	41	30	50	3	T1rho, T2	Radiographs and symptoms
X Li 2009[76]	-	-	10	10	56	41	30	40	3	T1rho, T2	Radiographs and symptoms
W Li 2010[75]	-	-	14	9	62	29	74	64	1.5	dGEMRIC	MRI and/or radiographs and/or symptoms
Liu 2015[42]	-	-	14	13	53	28	36	23	3	T2	Radiographs and symptoms
Madelin[78]	-	-	28	19	64	35	57	42	7	Sodium	ACR criteria
Matsubara[44]	10	-	-	19	57	39	32	0	3	T1rho	Radiographs and symptoms
Mosher[45]	16	16	-	18	51/57*	27	44/75*	67	3	T1rho, T2	Radiographs only
Multanen[48]	78	-	-	12	59	58	100	100	1.5	T2, dGEMRIC	Radiographs and symptoms
Newbould 2012[49]	-	-	20	17	64	61	65	76	3	T2*	ACR criteria
Newbould 2013[86]	-	-	28	23	63	62	NS	NS	3	Sodium	ACR criteria
Owman[79]	-	-	27	7	NS	NS	NS	NS	1.5	dGEMRIC	Radiographs only

Study	Number of OA subjects			Number of control subjects	Mean age		Gender (% female)		MRI field strength (T)	Compositional techniques used	Definition of OA
	mild	sev	NOS		OA	control	OA	control			
Raya 2012[57]	-	-	10	16	61	31	50	44	7	DTI, T2	ACR criteria
Raya 2014[56]	5	-	-	10	66	31	NS	30	7	DTI	ACR criteria
Souza 2013[81]	-	-	44	19	57	39	NS	NS	3	T1rho, T2	Radiographs and symptoms
Souza 2014[80]	-	-	44	93	57	50	61	58	3	T1rho, T2	Radiographs and symptoms
Sritanyaratana[62]	11	-	-	20	53	32	36	25	3	MT	Radiographs and symptoms
Stahl 2007[82]	-	-	8	10	56	58	100	100	3	T2	Radiographs and symptoms
Stahl 2009[83]	17	-	-	20	54	34	53	50	3	T1rho, T2	ACR criteria
Wang[84]	10	-	-	10	65	36	20	40	3	T1rho, T2	Radiographs only
Wirth[85]	32	-	-	89	60	55	56	60	3	T2	Radiographs only
Wyatt[86]	-	-	12	7	59	50	42	71	3, 7	T1rho, T2	Radiographs and symptoms
Yao[88]	20	20	-	11	48/55*	39	65/60*	27	3	T2, dGEMRIC, MT	Radiographs and symptoms

Abbreviations: sev - severe, ACR – American College of Rheumatology, NS – not specified.
*Mean ages and female percentage for mild/severe OA groups provided separately

Table 3 - Summary of discrimination between subjects with OA and controls for included compositional techniques. Only techniques with data available for pooling (i.e. at least 2 studies per comparison) are included in this table.

Compositional technique	Standardised Mean Difference (95% CI)					
	Mild OA vs controls	*n*	Severe OA vs controls	*n*	OA (NOS) vs controls	*n*
T2	**0.49 (0.30, 0.67)**	8	**1.24 (0.63, 1.85)**	3	**0.48 (0.34, 0.62)**	10
T1rho	**0.73 (0.40, 1.06)**	4	-	-	**0.60 (0.41, 0.80)**	7
dGEMRIC	0.13 (-0.23, 0.49)	2	-	-	-0.31 (-0.78, 0.15)	3
Sodium	-	-	-	-	-0.19 (-0.52, 0.15)	3

n – number of studies pooled for standardised mean difference estimate
Pooled comparisons in **bold** were statistically significant ($p < 0.05$)

SUPPLEMENTAL MATERIAL

1. SEARCH STRATEGY FOR MEDLINE AND EMBASE

1	exp osteoarthritis/
2	osteoarth*.tw.
3	or/1-2
4	knee.tw.
5	exp knee/
6	or/4-5
7	exp magnetic resonance imaging/
8	MRI.ti,ab.
9	magnetic resonance.ti,ab.
10	or/7-9
11	map*.tw.
12	relax*.tw.
13	T2.tw.
14	T2*.tw.
15	T1rho.tw.
16	T1*.tw.
17	sodium.tw.
18	dGEMRIC.tw.
19	gadolinium.tw.
20	gadopentetate.tw.
21	diffusion.tw.
22	DWI.tw.
23	DTI.tw.
24	gagCEST.tw.
25	collagen.tw.
26	proteoglycan.tw.
27	glycosaminoglycan.tw.
28	composition*.tw.
29	or/11-28
30	exp sensitivity/
31	exp specificity/
32	reliab*.tw.
33	valid*.tw.
34	sensitiv*.tw.
35	specific*.tw.
36	discim*.tw.
37	reproduc*.tw.
38	variab*.tw.
39	or/30-38
40	and/3,6,10,29,39

2. QUALITY ASSESSMENT OF RELIABILITY AND DISCRIMINATIVE VALIDITY STUDIES

[SUPPLEMENTAL FIGURE 1]

Modified QAREL items as follows:

1. Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied?
2. Was the test interpreted by analysts who were representative of those to whom the authors intended the results to be applied?
3. Were analysts blinded to the findings of other analysts during the study (only inter-observer reliability studies)?
4. Were analysts blinded to their own prior findings of the test under evaluation (only intra-observer reliability studies)?
5. Were analysts blinded to clinical information e.g. age, sex, OA status (only for studies featuring both OA subjects and controls)?
6. Was the time interval between repeated measurements compatible with the stability of the measurement (only test-retest reliability studies)?
7. Was the MR protocol described reproducibly?
8. Were appropriate statistical measures of agreement used?

[SUPPLEMENTAL FIGURE 2]

Modified QUADAS-2 items as follows:

1. Patient selection

Risk of bias

Could the selection of patients have introduced bias?

Signaling question 1: Was an attempt made to match control and OA populations for important baseline characteristics (e.g. age, sex)?

Signaling question 2: Did the study avoid inappropriate exclusions?

Applicability

Are there concerns that the included patients do not match the review question?

2. Index test

Risk of bias

Could the conduct or interpretation of the index test have introduced bias?

Signaling question 1: Were image analysts blinded to additional clinical cues (e.g. age, sex, OA status)?

Signaling question 2: Is the MRI acquisition protocol described in sufficient detail for it to be reproduced?

Signaling question 3: Is the image analysis procedure described in sufficient detail for it to be reproduced?

Applicability

Is there concern that the index test, its conduct or interpretation differ from the review question?

3. Definition of OA

Risk of bias

Signaling question 1: Was the study definition of OA based on clinical symptoms, imaging findings or both? Were standard (ACR) criteria applied?

Applicability

Is there concern that the definition of OA used differs from the review question?

4. Flow and timing

Risk of bias

Signaling question 1: Were all participants included in the analysis?

3. FOREST PLOTS

[SUPPLEMENTAL FIGURE 3A]

[SUPPLEMENTAL FIGURE 3B]

[SUPPLEMENTAL FIGURE 3C]

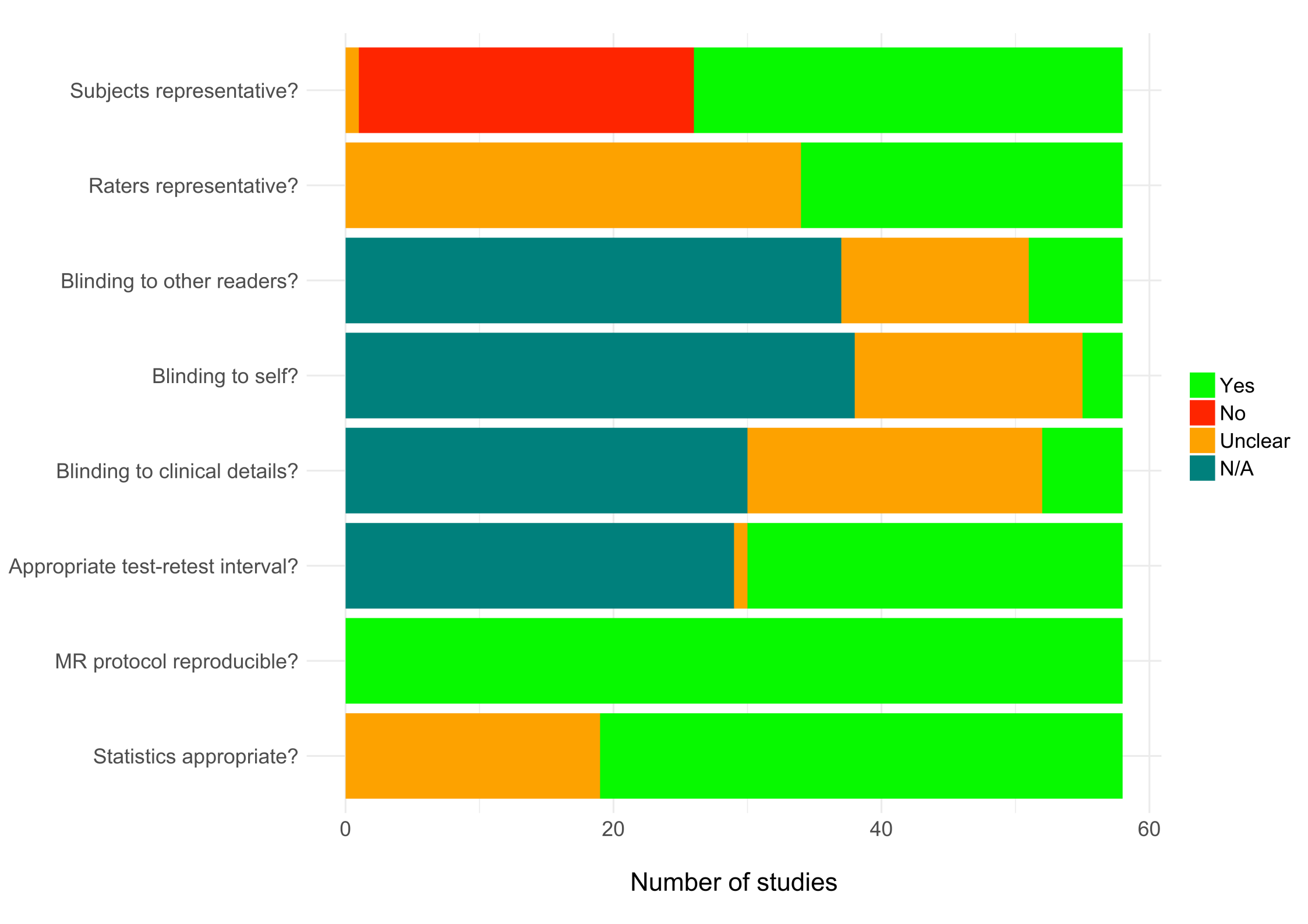
[SUPPLEMENTAL FIGURE 4A]

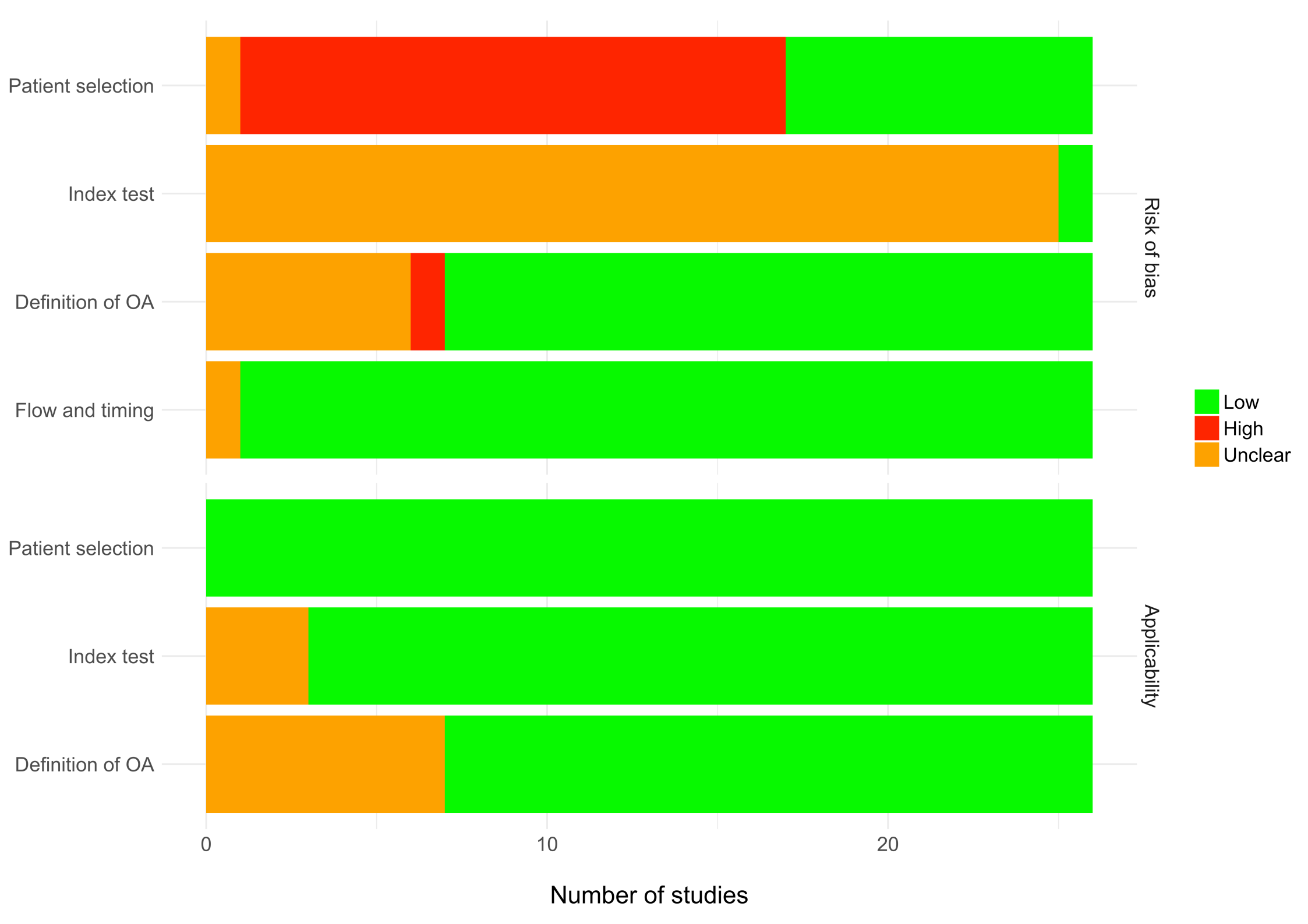
[SUPPLEMENTAL FIGURE 4B]

[SUPPLEMENTAL FIGURE 5A]

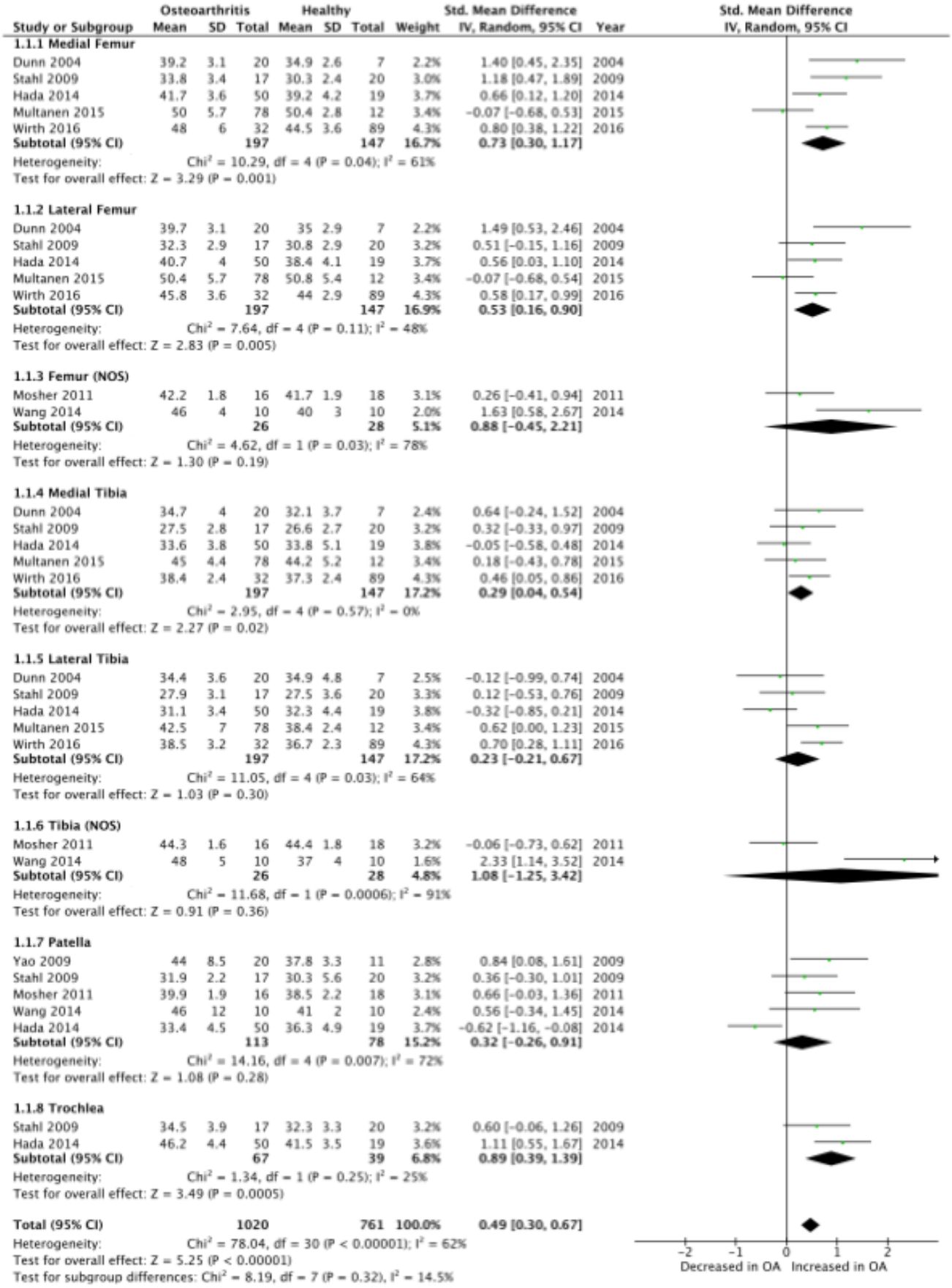
[SUPPLEMENTAL FIGURE 5B]

[SUPPLEMENTAL FIGURE 6]

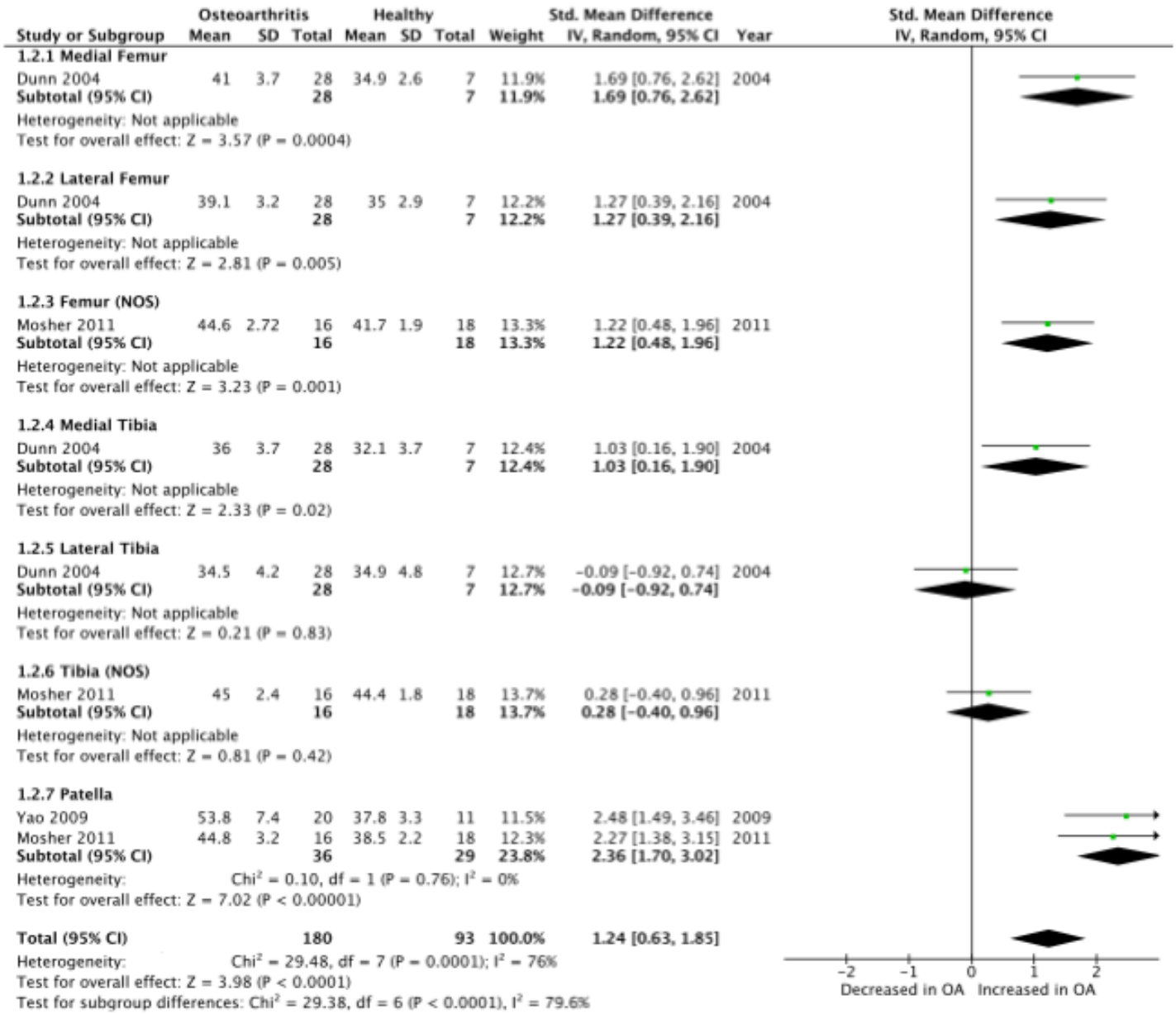

Subjects representative?
Raters representative?
Blinding to other readers?
Blinding to self?
Blinding to clinical details?
Appropriate test-retest interval?
MR protocol reproducible?
Statistics appropriate?
0
20
40
60
Number of studies
Yes
No
Unclear
N/A


Patient selection
Index test
Definition of OA
Flow and timing
Patient selection
Index test
Definition of OA
Risk of bias
Applicability
Low
High
Unclear
0
10
20
Number of studies

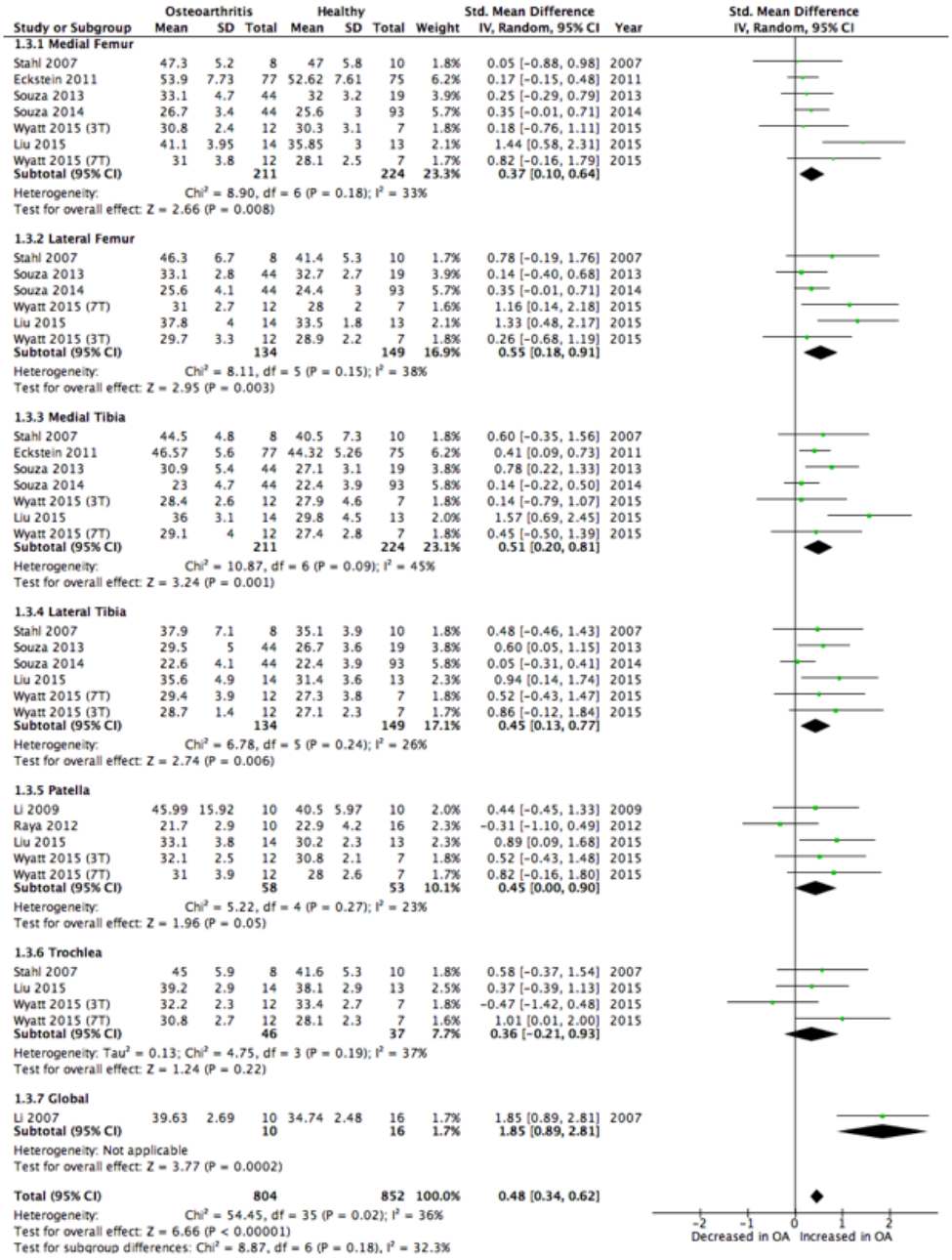
T2: mild OA vs controls



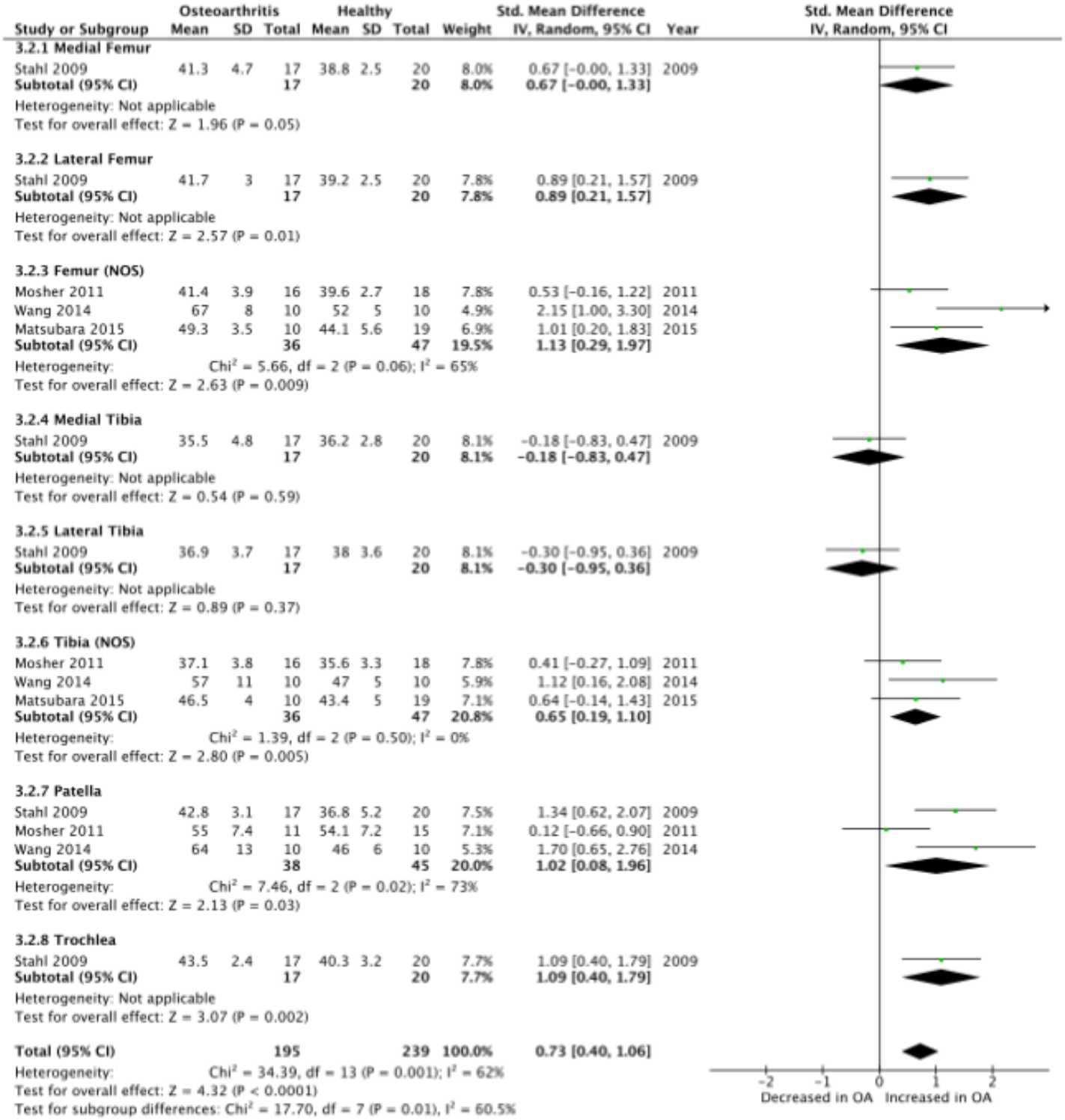
T2: severe OA vs controls



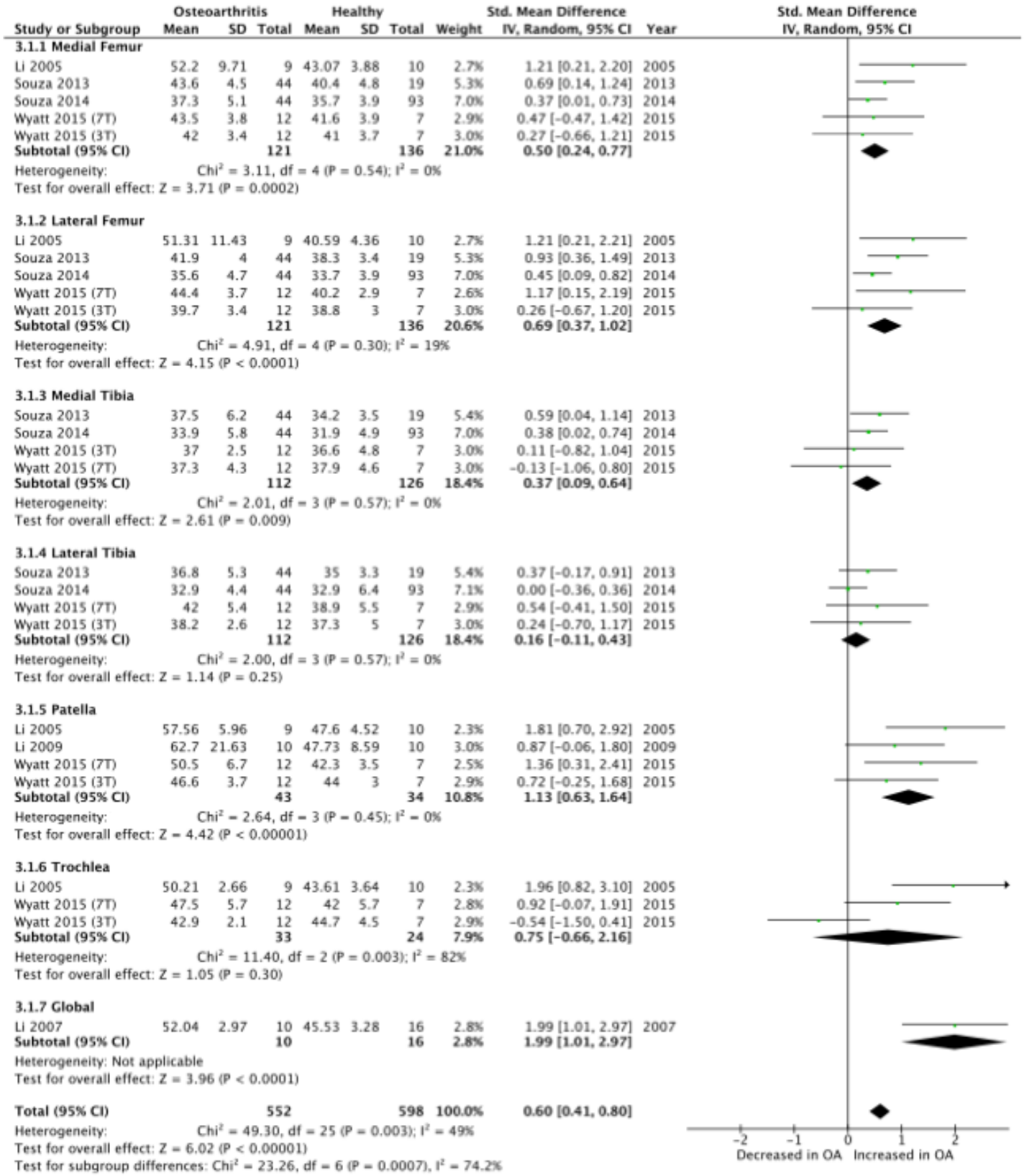
T2: OA (NOS) vs controls



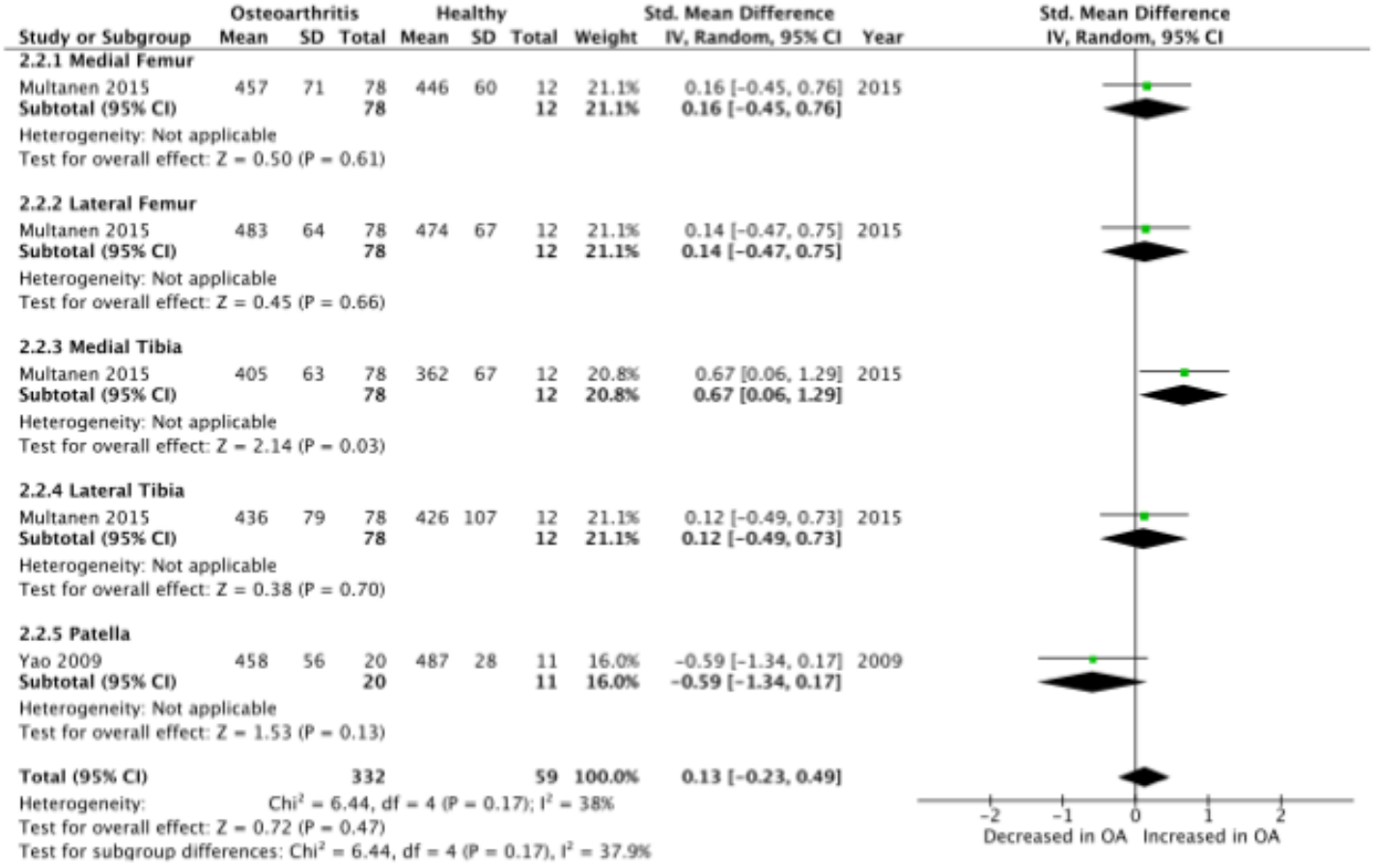
T1rho: mild OA vs controls



T1rho: OA (NOS) vs controls



dGEMRIC: mild OA vs controls



dGEMRIC: OA (NOS) vs controls