Accepted Manuscript

Systematic review and meta-analysis of the reliability and discriminative validity of cartilage compositional MRI in knee osteoarthritis

Dr James W. MacKay, Dr Samantha B.L. Low, Dr Toby O. Smith, Andoni P. Toms, Professor, Andrew M. McCaskie, Professor, Fiona J. Gilbert, Professor

PII: S1063-4584(18)31069-0
DOI: 10.1016/j.joca.2017.11.018
Reference: YJOCA 4190

To appear in: Osteoarthritis and Cartilage

Received Date: 3 August 2017
Revised Date: 16 October 2017
Accepted Date: 14 November 2017


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
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 MacKay1  
jwm37@cam.ac.uk
Dr Samantha B L Low2  
samanthalow@nnuh.nhs.uk
Dr Toby O Smith3  
toby.smith@uea.ac.uk
Professor Andoni P Toms2  
andonitoms@nnuh.nhs.uk
*Professor Andrew M McCaskie4  
awm41@cam.ac.uk
*Professor Fiona J Gilbert1  
fjg28@medschl.cam.ac.uk

1Department of Radiology, University of Cambridge, Cambridge, UK
2Department of Radiology, Norfolk & Norwich University Hospital, Norwich, UK
3School of Health Sciences, University of East Anglia, Norwich, UK
4Department 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] =
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.

Previous systematic reviews have assessed the reliability and discriminative validity of radiographic and conventional clinical MRI assessment of knee osteoarthritis. 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.
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 Group⁶.
METHODS

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

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.

**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. 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. 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. For the purposes of our review we used an interpretation of ICC values based on Landis and Koch.
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).
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).

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.

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
(II.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. 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. 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. 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.

2. T1rho relaxometry

One study provided intra-observer reliability data, with a CV of 3.8%. 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. Two studies examined test-retest reliability in a multi-center setting, reporting a CVs of 4.9% for major compartments and up to 18.8% for subregional analyses.

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%. Test-retest reliability was excellent (ICCs > 0.85, CV 4.2 to 7.4%) apart from one study comparing different TI mapping techniques for dGEMRIC which reported ICC values as low as 0 and a CV of 11% for a variable flip angle (VFA) technique.

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 2 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\cite{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{93,92}

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.\textsuperscript{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.\textsuperscript{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 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 TIrho 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 (fig28@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


**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
## TABLES

### Table 1 - Characteristics of included reliability studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of OA subjects</th>
<th>Number of other subjects</th>
<th>Mean age</th>
<th>Gender (% female)</th>
<th>MRI field strength (T)</th>
<th>Compositional techniques used</th>
<th>Type of reliability assessed</th>
<th>Reliability statistic**</th>
<th>Reliability value</th>
<th>Assessors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anandacoomarasamy</td>
<td>15</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>3</td>
<td>dGEMRIC</td>
<td>intraO</td>
<td>ICC</td>
<td>0.91</td>
<td>1</td>
</tr>
<tr>
<td>Balamoody</td>
<td>12</td>
<td>49</td>
<td>25</td>
<td>3</td>
<td>T2</td>
<td>TR</td>
<td>RMSCV</td>
<td>2.3 – 6.3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Baum</td>
<td>126</td>
<td>50</td>
<td>50</td>
<td>3</td>
<td>T2</td>
<td>intraO</td>
<td>RMSCV</td>
<td>1.76</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Blumenkrantz</td>
<td>13</td>
<td>17</td>
<td>64</td>
<td>1.5</td>
<td>T2</td>
<td>intraO</td>
<td>CV</td>
<td>1.5 – 2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Bron</td>
<td>11</td>
<td>10</td>
<td>52</td>
<td>3</td>
<td>dGEMRIC</td>
<td>TR</td>
<td>ICC</td>
<td>0.85 – 0.9</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Carballido-Gamio</td>
<td>5</td>
<td>29</td>
<td>20</td>
<td>3</td>
<td>T1rho</td>
<td>TR</td>
<td>CV</td>
<td>3.8 – 12.3†</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Dardzinski</td>
<td>5</td>
<td>52</td>
<td>70</td>
<td>3</td>
<td>T2</td>
<td>TR</td>
<td>RMSCV</td>
<td>3.3 – 6.5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Duryea</td>
<td>10</td>
<td>33</td>
<td>50</td>
<td>3</td>
<td>T2</td>
<td>intraO</td>
<td>TR</td>
<td>CV</td>
<td>1.3</td>
<td>1</td>
</tr>
<tr>
<td>Guha</td>
<td>20</td>
<td>58</td>
<td>54</td>
<td>3</td>
<td>DTI</td>
<td>TR</td>
<td>RMSCV</td>
<td>6.5 – 11.6</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Gupta</td>
<td>20</td>
<td>37</td>
<td>35</td>
<td>3</td>
<td>T1rho</td>
<td>interO</td>
<td>RMSCV</td>
<td>3.9</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Hada</td>
<td>50</td>
<td>57</td>
<td>25</td>
<td>3</td>
<td>T2</td>
<td>intraO</td>
<td>ICC</td>
<td>0.91</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Hannila 2009</td>
<td>20</td>
<td>25</td>
<td>54</td>
<td>3</td>
<td>T2</td>
<td>interO</td>
<td>ICC</td>
<td>1.8 – 14.2†</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Hannila 2015</td>
<td>9</td>
<td>30</td>
<td>44</td>
<td>1.5</td>
<td>T2</td>
<td>TR</td>
<td>RMSCV</td>
<td>2.5 – 22.2†</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Hesper</td>
<td>10</td>
<td>29</td>
<td>70</td>
<td>3</td>
<td>T2*</td>
<td>intraO</td>
<td>ICC</td>
<td>0.97</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**Reliability statistic:** intraO = intraclass correlation coefficient; ICC = intraclass correlation coefficient; RMSCV = root mean squared coefficient of variation; dGEMRIC = direct gadolinium ethoxypropoxyradiolactone intramolecular relaxation correlation coefficient; CV = consistency coefficient; T1rho = T1rho time; TR = temporal resolution; T2* = T2* time.

<table>
<thead>
<tr>
<th>Assessors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
</tr>
<tr>
<td>Experience</td>
</tr>
</tbody>
</table>

Abbreviations:
- mild: mild OA
- sev: severe OA
- NOS: other subjects
- OA: other subjects
- NS: not specified
- T: Tesla
- TR: temporal resolution
- ICC: intraclass correlation coefficient
- RMSCV: root mean squared coefficient of variation
<table>
<thead>
<tr>
<th>Study</th>
<th>Number of OA subjects</th>
<th>Number of other subjects</th>
<th>Mean age</th>
<th>Gender (% female)</th>
<th>MRI field strength (T)</th>
<th>Compositional techniques used</th>
<th>Type of reliability assessed</th>
<th>Reliability statistic</th>
<th>Reliability value</th>
<th>Assessors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mild sev NOS</td>
<td>OA other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holtzman&lt;sup&gt;59&lt;/sup&gt;</td>
<td>- - -</td>
<td>26 - 36</td>
<td>- 38</td>
<td>3</td>
<td>T1rho T2</td>
<td>interO CV</td>
<td></td>
<td>4.3</td>
<td>4.9</td>
<td>2</td>
</tr>
<tr>
<td>Hovis 2011&lt;sup&gt;39&lt;/sup&gt;</td>
<td>- - -</td>
<td>161 - 50</td>
<td>- 59</td>
<td>3</td>
<td>T2</td>
<td>intraO RMSE</td>
<td></td>
<td>0.74 – 1.51</td>
<td>1</td>
<td>NS</td>
</tr>
<tr>
<td>Hovis 2012&lt;sup&gt;39&lt;/sup&gt;</td>
<td>- - -</td>
<td>105 - 65</td>
<td>- 68</td>
<td>3</td>
<td>T2</td>
<td>intraO ICC</td>
<td></td>
<td>0.99</td>
<td>0.99</td>
<td>2</td>
</tr>
<tr>
<td>Jordan&lt;sup&gt;33&lt;/sup&gt;</td>
<td>- - -</td>
<td>8 - 28</td>
<td>- 25</td>
<td>3</td>
<td>T1rho T2</td>
<td>interO CV</td>
<td></td>
<td>3.8</td>
<td>5.7</td>
<td>2</td>
</tr>
<tr>
<td>Joseph 2011&lt;sup&gt;39&lt;/sup&gt;</td>
<td>- - -</td>
<td>145 - 50</td>
<td>- 59</td>
<td>3</td>
<td>T2</td>
<td>intraO RMSCV</td>
<td></td>
<td>0.9 – 2.1</td>
<td>1</td>
<td>NS</td>
</tr>
<tr>
<td>Joseph 2012&lt;sup&gt;34&lt;/sup&gt;</td>
<td>- - -</td>
<td>289 - 51</td>
<td>- 47</td>
<td>3</td>
<td>T2</td>
<td>intraO RMSCV</td>
<td></td>
<td>0.8 – 3.2</td>
<td>1</td>
<td>NS</td>
</tr>
<tr>
<td>Juras 2016&lt;sup&gt;39&lt;/sup&gt;</td>
<td>- - -</td>
<td>23 - 33</td>
<td>- 57</td>
<td>3</td>
<td>T2</td>
<td>interO CV</td>
<td></td>
<td>5.8 – 10.8†</td>
<td>10 years/15 years</td>
<td>NS</td>
</tr>
<tr>
<td>Koll&lt;sup&gt;39&lt;/sup&gt;</td>
<td>80</td>
<td>- - -</td>
<td>58 - 100</td>
<td>1.5</td>
<td>T2</td>
<td>interO RMSCV</td>
<td></td>
<td>2</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>X Li 2005&lt;sup&gt;38&lt;/sup&gt;</td>
<td>- - 9</td>
<td>10 - 52</td>
<td>44 - 40</td>
<td>3</td>
<td>T1rho TR</td>
<td>CV</td>
<td></td>
<td>4.8</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>X Li 2014&lt;sup&gt;37&lt;/sup&gt;</td>
<td>- - -</td>
<td>6 - 22-35</td>
<td>- 50</td>
<td>3</td>
<td>T1rho T2</td>
<td>TR RMSCV</td>
<td></td>
<td>4.2 – 6</td>
<td>5 – 6.3</td>
<td>NS</td>
</tr>
<tr>
<td>X Li 2015&lt;sup&gt;39&lt;/sup&gt;</td>
<td>- - -</td>
<td>18 - NS</td>
<td>- NS</td>
<td>3</td>
<td>T1rho T2</td>
<td>TR&lt;sup&gt;4&lt;/sup&gt; RMSCV</td>
<td></td>
<td>2.3 – 5.1</td>
<td>3.2 – 5.3</td>
<td>NS</td>
</tr>
<tr>
<td>Liebl&lt;sup&gt;50&lt;/sup&gt;</td>
<td>- - -</td>
<td>130 - 59</td>
<td>- 60</td>
<td>3</td>
<td>T2</td>
<td>intraO RMSCV</td>
<td></td>
<td>1.7</td>
<td>1.6</td>
<td>2</td>
</tr>
<tr>
<td>Liess&lt;sup&gt;41&lt;/sup&gt;</td>
<td>- - -</td>
<td>20 - 28</td>
<td>- 20</td>
<td>1.5</td>
<td>T2</td>
<td>TR CV</td>
<td></td>
<td>1.7</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Liu&lt;sup&gt;42&lt;/sup&gt;</td>
<td>- - 14</td>
<td>13 - 53</td>
<td>36 - 23</td>
<td>3</td>
<td>T2</td>
<td>TR CV</td>
<td></td>
<td>3.1 – 10</td>
<td>1</td>
<td>4 years</td>
</tr>
<tr>
<td>Study</td>
<td>Number of OA subjects</td>
<td>Number of other subjects</td>
<td>Mean age</td>
<td>Gender (% female)</td>
<td>MRI field strength (T)</td>
<td>Compositional techniques used</td>
<td>Type of reliability assessed</td>
<td>Reliability statistic**</td>
<td>Reliability value</td>
<td>Assessors</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------</td>
<td>--------------------------</td>
<td>----------</td>
<td>-------------------</td>
<td>-----------------------</td>
<td>-----------------------------</td>
<td>----------------------------</td>
<td>----------------------------</td>
<td>---------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Madelin et al.</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>37</td>
<td>3,7</td>
<td>Sodium</td>
<td>TR</td>
<td>RMSCV</td>
<td>9.1 – 12.3</td>
<td>NS</td>
</tr>
<tr>
<td>Matsubara et al.</td>
<td>10</td>
<td>-</td>
<td>19</td>
<td>57</td>
<td>32</td>
<td>Trho</td>
<td>interO</td>
<td>ICC</td>
<td>0.93</td>
<td>2</td>
</tr>
<tr>
<td>Mosher et al. 1</td>
<td>16</td>
<td>16</td>
<td>18</td>
<td>54</td>
<td>59</td>
<td>Trho</td>
<td>T2</td>
<td>ICC</td>
<td>0.20 – 0.93†</td>
<td>NS</td>
</tr>
<tr>
<td>Mosher et al. 2</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>-</td>
<td>3</td>
<td>T2</td>
<td>intraO</td>
<td>wK</td>
<td>0.77</td>
<td>1</td>
</tr>
<tr>
<td>Multanen et al. 1</td>
<td>78</td>
<td>-</td>
<td>12</td>
<td>59</td>
<td>100</td>
<td>T2</td>
<td>dGEMRIC</td>
<td>ICC</td>
<td>2/3</td>
<td>2</td>
</tr>
<tr>
<td>Multanen et al. 2</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>-</td>
<td>3</td>
<td>dGEMRIC</td>
<td>TR</td>
<td>ICC</td>
<td>0.45 – 0.98†</td>
<td>NS</td>
</tr>
<tr>
<td>Newbould et al. 1</td>
<td>-</td>
<td>-</td>
<td>13</td>
<td>5</td>
<td>77</td>
<td>T2</td>
<td>TR</td>
<td>ICC</td>
<td>0.7 – 0.94</td>
<td>NS</td>
</tr>
<tr>
<td>Newbould et al. 2</td>
<td>-</td>
<td>-</td>
<td>15</td>
<td>5</td>
<td>80</td>
<td>Sodium</td>
<td>TR</td>
<td>ICC</td>
<td>0.67 – 9.9</td>
<td>1</td>
</tr>
<tr>
<td>Nishioka et al. 1</td>
<td>-</td>
<td>-</td>
<td>20</td>
<td>-</td>
<td>90</td>
<td>Trho</td>
<td>interO</td>
<td>CV</td>
<td>&lt; 6</td>
<td>NS</td>
</tr>
<tr>
<td>Nishioka et al. 2</td>
<td>-</td>
<td>-</td>
<td>78</td>
<td>-</td>
<td>77</td>
<td>Trho</td>
<td>T2</td>
<td>interO</td>
<td>11.8 / 12.2</td>
<td>2</td>
</tr>
<tr>
<td>Nishioka et al. 3</td>
<td>-</td>
<td>-</td>
<td>37</td>
<td>-</td>
<td>90</td>
<td>Trho</td>
<td>T2</td>
<td>interO</td>
<td>1.4 / 1.0</td>
<td>2</td>
</tr>
<tr>
<td>Pan et al.</td>
<td>-</td>
<td>-</td>
<td>95</td>
<td>-</td>
<td>61</td>
<td>T2</td>
<td>intraO</td>
<td>TR</td>
<td>1.1 – 1.2</td>
<td>1</td>
</tr>
<tr>
<td>Pedoia et al.</td>
<td>40</td>
<td>-</td>
<td>15</td>
<td>54</td>
<td>48</td>
<td>Trho</td>
<td>T2</td>
<td>CV</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Raya et al. 1</td>
<td>5</td>
<td>-</td>
<td>10</td>
<td>66</td>
<td>31</td>
<td>DTI</td>
<td>TR</td>
<td>RMSCV</td>
<td>2.9 – 5.6</td>
<td>NS</td>
</tr>
<tr>
<td>Raya et al. 2</td>
<td>-</td>
<td>10</td>
<td>16</td>
<td>61</td>
<td>50</td>
<td>DTI</td>
<td>T2</td>
<td>TR</td>
<td>7.3 – 19.3†</td>
<td>1</td>
</tr>
</tbody>
</table>

**Note:** T2: 2 years; T2: 2 years; NS: Not specified.
<table>
<thead>
<tr>
<th>Study</th>
<th>Number of OA subjects</th>
<th>Number of other subjects</th>
<th>Mean age</th>
<th>Gender (% female)</th>
<th>MRI field strength (T)</th>
<th>Compositional techniques used</th>
<th>Type of reliability assessed</th>
<th>Reliability statistic</th>
<th>Reliability value</th>
<th>Assessors</th>
<th>Number</th>
<th>Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schleic^58</td>
<td>- - -</td>
<td>20 - 25</td>
<td>40 3</td>
<td>gagCEST</td>
<td>intraO interO</td>
<td>ICC</td>
<td>0.95 0.95</td>
<td>1 2</td>
<td>8 years 5 years/8 years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serebrakian^59</td>
<td>- - -</td>
<td>127 - 55</td>
<td>56 3</td>
<td>T2</td>
<td>intraO interO</td>
<td>RMSCV</td>
<td>1.1 3.3</td>
<td>2 NS</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singh^60</td>
<td>- - -</td>
<td>8 - 20-35</td>
<td>NS 7</td>
<td>T1rho</td>
<td>TR</td>
<td>ICC CV</td>
<td>0.73 - 0.96 2.3 - 4.3</td>
<td>NS NS</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siversson^61</td>
<td>- - -</td>
<td>9 - 45</td>
<td>56 1.5</td>
<td>dGEMRIC</td>
<td>TR</td>
<td>ICC RMSCV</td>
<td>0 - 0.69 6.0 - 11.6</td>
<td>NS NS</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sritanyaratana^62</td>
<td>II - -</td>
<td>20 - 53 - 32</td>
<td>25 - 3</td>
<td>MT</td>
<td>TR CV</td>
<td>ICC CV</td>
<td>0.5 - 4.6</td>
<td>1 NS</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stehling 2010^63</td>
<td>- - -</td>
<td>120 - 51</td>
<td>50 3</td>
<td>T2</td>
<td>intraO CV</td>
<td>ICC RMSCV</td>
<td>1.2 - 2.8†</td>
<td>2 NS</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stehling 2011^63</td>
<td>- - -</td>
<td>10 - 52</td>
<td>50 3</td>
<td>T2</td>
<td>intraO CV</td>
<td>ICC RMSCV</td>
<td>0.17 - 0.89 0.3 - 0.96</td>
<td>3 5 years/ 6 years/ 13 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surowiec^65</td>
<td>- - -</td>
<td>18 - 18-35</td>
<td>NS 3</td>
<td>T2</td>
<td>T1rho T2</td>
<td>interO ICC</td>
<td>0.81 0.92</td>
<td>2 12 years / 7 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takayama^66</td>
<td>- 16 -</td>
<td>73 - 88</td>
<td>3</td>
<td>T1rho T2</td>
<td>interO ICC</td>
<td>1.5 - 2.6</td>
<td>2 2 medical students, 2 ortho surgeons, 2 radiologists</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tiderius^67</td>
<td>- - -</td>
<td>12 - 24</td>
<td>0 1.5</td>
<td>dGEMRIC</td>
<td>interO CV</td>
<td>ICC</td>
<td>1.8</td>
<td>6</td>
<td>2 medical students, 2 ortho surgeons, 2 radiologists</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Tiel^68</td>
<td>17 - -</td>
<td>50 - 41</td>
<td>3</td>
<td>dGEMRIC</td>
<td>TR ICC</td>
<td>ICC</td>
<td>0.87 - 0.95</td>
<td>1</td>
<td>Medical degree</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Number of OA subjects</td>
<td>Number of other subjects</td>
<td>Mean age</td>
<td>Gender (% female)</td>
<td>MRI field strength (T)</td>
<td>Compositional techniques used</td>
<td>Type of reliability assessed</td>
<td>Reliability statistic**</td>
<td>Reliability value</td>
<td>Assessors</td>
<td>Number</td>
<td>Experience</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------</td>
<td>--------------------------</td>
<td>----------</td>
<td>-------------------</td>
<td>------------------------</td>
<td>-----------------------------</td>
<td>------------------------------</td>
<td>------------------------</td>
<td>--------------------</td>
<td>-----------</td>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>Welsch⁶⁹</td>
<td>- - -</td>
<td>17 - 26 - 24</td>
<td>3, 7</td>
<td></td>
<td></td>
<td>T2, T2*</td>
<td>interO ICC</td>
<td>0.91 - 0.95†</td>
<td>7.2 - 8.7†</td>
<td>3</td>
<td>2 years/10 years/25 years</td>
<td></td>
</tr>
<tr>
<td>Wiener⁷⁰</td>
<td>- - -</td>
<td>25 - 31 - 60</td>
<td>1.5</td>
<td></td>
<td></td>
<td>T1</td>
<td>TR ICC</td>
<td>0.88 - 0.90†</td>
<td>6.8 - 7.8†</td>
<td>1</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>Williams⁷¹</td>
<td>- - -</td>
<td>11 - 28 - 100</td>
<td>3</td>
<td></td>
<td></td>
<td>UTE-T2*</td>
<td>intraO ICC</td>
<td>0.80 - 0.97†</td>
<td>6 - 16†</td>
<td>1</td>
<td>9 years</td>
<td></td>
</tr>
<tr>
<td>Zuo⁷²</td>
<td>- - -</td>
<td>6 - 27-30 - 17</td>
<td>3</td>
<td></td>
<td></td>
<td>T1rho T2</td>
<td>TR CV</td>
<td>1.8 - 5.6 2.9 - 5.7</td>
<td>NS NS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


**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

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of OA subjects</th>
<th>Number of control subjects</th>
<th>Mean age</th>
<th>Gender (% female)</th>
<th>MRI field strength (T)</th>
<th>Compositional techniques used</th>
<th>Definition of OA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunn</td>
<td>20 28 7</td>
<td>7</td>
<td>63/67* 38</td>
<td>65/50* 43</td>
<td>1.5</td>
<td>T2</td>
<td>Radiographs and symptoms</td>
</tr>
<tr>
<td>Eckstein</td>
<td>- - 77 75 58 55</td>
<td>100 100</td>
<td>3</td>
<td>T2, dGEMRIC</td>
<td></td>
<td></td>
<td>Radiographs only</td>
</tr>
<tr>
<td>Hada</td>
<td>50 - 19 57 25 54 37</td>
<td>3</td>
<td>T2</td>
<td>ACR criteria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X Li 2005</td>
<td>- - 9 10 52 30 44 40</td>
<td>3</td>
<td>Tlrho</td>
<td>Radiographs and/or symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X Li 2007</td>
<td>- - 10 16 56 41 30 50</td>
<td>3</td>
<td>Tlrho, T2</td>
<td>Radiographs and symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X Li 2009</td>
<td>- - 10 10 56 41 30 40</td>
<td>3</td>
<td>Tlrho, T2</td>
<td>Radiographs and symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W Li 2010</td>
<td>- - 14 9 62 29 74 64</td>
<td>1.5</td>
<td>dGEMRIC</td>
<td>MRI and/or radiographs and/or symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liu 2015</td>
<td>- - 14 13 53 28 36 23</td>
<td>3</td>
<td>T2</td>
<td>Radiographs and symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Madelin</td>
<td>- - 28 19 64 35 57 42</td>
<td>7</td>
<td>Sodium</td>
<td>ACR criteria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matsubara</td>
<td>10 - 19 57 39 32 0</td>
<td>3</td>
<td>Tlrho</td>
<td>Radiographs and symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mosher</td>
<td>16 16 - 18 51/57* 27</td>
<td>3</td>
<td>Tlrho, T2</td>
<td>Radiographs only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multanen</td>
<td>78 - 12 59 58 100 100</td>
<td>1.5</td>
<td>T2, dGEMRIC</td>
<td>Radiographs and symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newbould 2012</td>
<td>- - 20 17 64 61 65 76</td>
<td>3</td>
<td>T2*</td>
<td>ACR criteria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Newbould 2013</td>
<td>- - 28 23 63 62 NS NS</td>
<td>3</td>
<td>Sodium</td>
<td>ACR criteria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owman</td>
<td>- - 27 7 NS NS NS NS 1.5</td>
<td>dGEMRIC</td>
<td>Radiographs only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Number of OA subjects</td>
<td>Number of control subjects</td>
<td>Mean age</td>
<td>Gender (% female)</td>
<td>MRI field strength (T)</td>
<td>Compositional techniques used</td>
<td>Definition of OA</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------</td>
<td>---------------------------</td>
<td>----------</td>
<td>-------------------</td>
<td>------------------------</td>
<td>-------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td>mild</td>
<td>sev</td>
<td>NOS</td>
<td>OA</td>
<td>control</td>
<td>OA</td>
<td>control</td>
</tr>
<tr>
<td>Raya 201257</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>16</td>
<td>61</td>
<td>31</td>
<td>50</td>
</tr>
<tr>
<td>Raya 201466</td>
<td>5</td>
<td>-</td>
<td>10</td>
<td>66</td>
<td>31</td>
<td>NS</td>
<td>30</td>
</tr>
<tr>
<td>Souza 20136a</td>
<td>-</td>
<td>-</td>
<td>44</td>
<td>19</td>
<td>57</td>
<td>39</td>
<td>NS</td>
</tr>
<tr>
<td>Souza 20146o</td>
<td>-</td>
<td>-</td>
<td>44</td>
<td>93</td>
<td>57</td>
<td>50</td>
<td>61</td>
</tr>
<tr>
<td>Sritanyaratana6a</td>
<td>11</td>
<td>-</td>
<td>20</td>
<td>53</td>
<td>32</td>
<td>36</td>
<td>25</td>
</tr>
<tr>
<td>Stahl 20078a</td>
<td>-</td>
<td>-</td>
<td>8</td>
<td>10</td>
<td>56</td>
<td>58</td>
<td>100</td>
</tr>
<tr>
<td>Stahl 20098a</td>
<td>17</td>
<td>-</td>
<td>20</td>
<td>54</td>
<td>34</td>
<td>53</td>
<td>50</td>
</tr>
<tr>
<td>Wang84</td>
<td>10</td>
<td>-</td>
<td>10</td>
<td>65</td>
<td>36</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>Wirth85</td>
<td>32</td>
<td>-</td>
<td>89</td>
<td>60</td>
<td>55</td>
<td>56</td>
<td>60</td>
</tr>
<tr>
<td>Wyatt86</td>
<td>-</td>
<td>-</td>
<td>12</td>
<td>7</td>
<td>59</td>
<td>50</td>
<td>42</td>
</tr>
<tr>
<td>Yao88</td>
<td>20</td>
<td>20</td>
<td>-</td>
<td>11</td>
<td>48/55*</td>
<td>39</td>
<td>65/60*</td>
</tr>
</tbody>
</table>

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.

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

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