

Lenza M, Buchbinder R, et al. Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered. Cochrane Database of Systematic Reviews 2013, Issue 9. Art # CD009020

Design: Meta-analysis of studies of diagnostic test accuracy

Main study question: What is the best imaging modality for rotator cuff tears when the clinical examination identifies the patient as a likely candidate for surgery?

PICOS:

- Patient population: people with shoulder pain suspected of having a rotator cuff tear for whom surgery was being considered
 - o Studies which included healthy controls, patients with previously diagnosed specific shoulder pain (osteoarthritis, frozen shoulder, rheumatoid arthritis, tumors) were excluded
 - o Studies which supplied insufficient information to allow construction of two by two tables (test results vs. surgery results) were also excluded
- Intervention: ultrasound (US), MR arthrography (MRA), and MR imaging (MRI)
- Comparison intervention: reference (gold) standard for rotator cuff tear was either arthroscopic or open (including mini-open) surgical findings
- Outcomes: diagnostic test accuracy, including sensitivity, specificity, positive likelihood ratios, and negative likelihood ratios
- Study types: only full reports from prospective studies
 - o Retrospective studies and studies which had an excessively long (one year or more) delay between imaging and surgery were excluded

Study selection:

- Electronic databases included MEDLINE, EMBASE, LILACS, and the Cochrane Register of Diagnostic Test Accuracy Studies
- Search dates went through February of 2011
- Study selection and quality assessment were done by two authors working independently
- Multiple categories of rotator cuff tears are often reported (full thickness tear, partial thickness tear, no tear); in order to construct two by two tables, the authors grouped cuff tears as follows:
 - o To create tables for partial thickness tears, data for full thickness tears were combined with data for no tears

- To create tables for “any” tears, full thickness and partial thickness tears were combined into a single category to compare with no tears
- Study quality was assessed by numerous criteria, including the following
 - Spectrum of patients appropriate: recruitment of patients was consecutive and all patients had shoulder pain suspected of being a surgically treatable rotator cuff tear
 - Acceptable reference standard: arthroscopy or open surgery
 - Open surgery was considered less accurate than arthroscopic surgery because open surgery cannot assess and identify joint surface and inferior surface tears
 - Acceptable delay between tests: average interval between imaging and surgery was one month or less
 - Partial verification avoided: did all patients who had an imaging test go on to receive verification at surgery?
 - Differential verification bias: did all patients have the same surgical operation regardless of the results of the imaging test?
 - Index test blinded: imaging tests interpreted without knowledge of surgical results (imaging not retrospectively re-interpreted after results of surgery became known)
 - Reference test blinded: surgeon was not aware of imaging test results when operation was done
 - Relevant clinical information: if clinical data (Neer’s test, etc) which would normally be available to the interpreter of the imaging test were available to the interpreters in the study in question
 - Uninterpretable results reported: if the number of results which could not be interpreted were reported
 - Withdrawals explained
 - Learning curve for imaging and reference tests reported: did the test interpreters have experience in diagnosis of musculoskeletal disease and experience in operating on shoulder conditions
 - Imaging and reference tests have clear criteria for what constituted a positive result
- Pooling of data from studies was conducted following the *Cochrane Handbook for Diagnostic Test Accuracy*
 - For each test and target condition, a summary receiver operating characteristic (ROC) curve was fitted, and an estimate was made of overall sensitivity, specificity, and positive/negative likelihood ratios (LR) were calculated following Cochrane statistical methods

Main results:

- 2198 records were screened from the electronic literature search; 270 full-text articles were assessed for eligibility, and 20 studies were included for the meta-analysis; 25 articles published in languages other than English were set aside for translation and later classification
- Most comparisons of different imaging tests were indirect comparisons (MRI, US, and MRA being done in different study populations); however, six studies made direct comparisons (two imaging techniques done in the same population and compared with the results of surgery)
- The methodological quality of most studies was judged to be low or unclear for most categories of quality assessment
 - o Half of the studies were considered to have an unclear reference standard, mostly because of surgery being done with an open rather than an arthroscopic operation
 - o Spectrum (prospective, consecutive recruitment of surgical candidates) was done in only 6 of the 20 studies
 - o More than half of the studies poorly described such criteria as the time period between imaging tests and operation, differential verification bias, relevant clinical information, and learning curves for imaging and surgical experience
- For detection of “any” rotator cuff tears, MRA, MRI, and US had some studies included in the analysis
 - o 6 studies of 346 patients assessed MRI; the average sensitivity was 98%; average specificity was 79%; positive LR was 5, and negative LR was 0.03
 - o 13 studies of 848 patients assessed US; the average sensitivity was 91%; average specificity was 85%; positive LR was 6, and negative LR was 0.11
 - o 3 studies of 183 patients assessed MRA; meta-analysis was not done (one study had MRA performing worse than chance prediction would perform; the other two studies reported very high sensitivity and specificity)
 - o Comparison of MRI and US did not show a clear superiority of one study over another for sensitivity or specificity; although there were 3 studies which made a head-to-head comparison of MRI and US, they had conflicting results, and there was no statistically significant difference between the two tests
 - o For head-to-head comparison of MRA and US, 2 studies of 127 patients reported higher sensitivity of MRA over US, but specificity results were conflicting (one study reporting that US had 100% specificity and the other reporting 45%)
- For detection of full thickness tears, MRI, MRA, and US all had relevant studies
 - o In 7 studies with 367 patients assessing MRI, average sensitivity was 94%; specificity was 93%, positive LR was 13, and negative LR was 0.06
 - o In 10 studies with 723 patients assessing US, average sensitivity was 92%, specificity was 93%; positive LR was 12, and negative LR was 0.09

- In 3 studies of 183 patients assessing MRA, average sensitivity was 94%, specificity was 92%; positive LR was 12, and negative LR was 0.06
- Comparing MRI and US, no differences in diagnostic accuracy were found, including comparisons based on head-to-head comparison in 4 studies
- As with “any” cuff tears, head-to-head comparisons of MRA and US reported higher sensitivity with MRA but conflicting comparisons of specificity
- For detection of partial thickness tears, all 3 imaging studies were studied
 - In 6 studies of 346 patients assessing MRI, average sensitivity was 74%, specificity was 93%, positive LR was 10, and negative LR was 0.28
 - In 8 studies of 654 patients assessing US, average sensitivity was 52%, specificity was 93%; positive LR was 8, and negative LR was 0.52
 - In 2 studies, sensitivity was only 13%, but in one study, sensitivity was 100%
 - In 4 studies of 233 patients assessing MRA, meta-analysis was not performed and average sensitivity/specificity were not calculated
 - Comparisons of MRI and US did not show one to be any better than the other, but both modalities had lower sensitivity for partial than for “any” or for full thickness tears
- There was greater variability of diagnostic accuracy across all studies for US than for MRI, but the authors were not able to investigate sources of heterogeneity due to a lack of data
- The relevance of the various quality criteria was investigated with sensitivity analyses, in which diagnostic accuracy was compared in studies which met and failed to meet different criteria
 - There were too few studies of MRI and MRA to allow for these analyses
 - However, for US, two quality criteria were investigated: acceptable reference standard and blinding of index test results
 - A small difference was found for US in detecting partial thickness tears, depending on whether the reference standard was acceptable: sensitivity was only 52% when all studies (open and arthroscopic surgery) were combined, but when only studies with acceptable reference standards (arthroscopic repair only) were combined, the sensitivity rose to 62%

Authors’ conclusions:

- For full thickness tears, MRI, MRA, and US perform well as diagnostic tests, and any one of them could be used in patients in whom surgery is being considered
- For partial thickness tears, MRI and US had lower sensitivity than for “any” tears and full thickness tears; US was only marginally better than chance in excluding a partial thickness tear

- The results of these studies only apply to patients for whom surgery is being considered; they do not apply to settings in primary care where asymptomatic changes in the rotator cuff are common and over-reliance on early imaging is also common
- The findings were based on small studies with poor reporting standards
- Sources of heterogeneity could not be systematically explored, but the experience of the clinician is important in diagnostic test accuracy, especially for US, where the results were highly heterogeneous
- There are 25 studies in other languages which await translation; the results of these studies, if appropriate for inclusion, will be considered in a future update
- There have been other systematic reviews of the same topic, such as de Jesus 2009 and Smith 2012, with results which are generally consistent with the current analysis
- The studies could only assess diagnostic test accuracy; the important issue of whether different diagnostic testing strategies lead to different benefits in terms of pain relief and shoulder function could not be addressed

Comments:

- Special statistical methods were used to create summary sensitivities and specificities; these were not separately averaged, since separate averaging of sensitivity and specificity obscures the trade-offs that occur between the two
- Some quality considerations are not clearly appropriate in the settings of real world practice
 - o That is, blinding of the reference test means that surgeons are expected to operate without reviewing the imaging studies
 - o Similarly, to avoid verification bias, all patients with no cuff tear on imaging ought to have arthroscopic surgery anyway in order to have complete data; however desirable this is in a research setting, it may not be appropriate to expect in all settings
- The diagnostic performance of the imaging tests for partial thickness tears was lower than for full thickness tears, but there was a peculiar feature of the analysis which was not discussed by the authors
 - o Specifically, for partial thickness tears, full thickness tears were combined with the absence of tears into a single category
 - o This analysis, done for the purpose of creating convenient two by two tables, creates a situation in which it is not clear whether imaging does poorly in telling a partial from a full thickness tear, telling a partial tear from no tear, or a bit of both
- The authors also combined full and partial thickness tears into a single category of “any” tear in order to create a convenient two by two table for telling these tears apart from having no rotator cuff tear

- The statistical issues in meta-analysis of diagnostic tests are much more complex than for meta-analysis of randomized clinical trials of treatment, and often require input from professional statisticians (Mackaskill 2010); these methods appear to have been followed by the authors
- There is considerable heterogeneity in diagnostic performance of tests which measure continuous variables with variable cutoff points (such as the level of fasting blood glucose which define the presence or absence of diabetes), but at least in these studies, instrument calibration is likely to have reasonable consistency between samples
- However, in studies of imaging tests, there is no equivalent measurement which lends itself to creation of cutoff values between presence and absence of disease; this is likely to create even more heterogeneity between studies due to differences among radiologists who report the results of imaging tests
- Positive and negative likelihood ratios are generally considered to be useful when the positive LR is greater than 10 (a positive result means that the odds of a disease becomes 10 times as likely than it was before the test was done) or when the negative LR is less than 0.1 (a negative test result means that the odds of a disease is only one tenth as likely after the test than it was before the test)
 - o However, these are only general rules, and the usefulness may depend on the consequences of getting a wrong answer
 - o For example, if a false positive result leads to an invasive and inappropriate intervention, then it makes good sense to decide on a minimum acceptable specificity, and to estimate the sensitivity at that level of specificity
- It is not clear from the authors' discussion whether patients are worse off if arthroscopic surgery is being done when the cuff is not torn, or if surgery is not done when the cuff is torn
 - o If it is true that delays in repair may convert cuff tears from reparable to irreparable, then high sensitivity is more valuable than high specificity
- The results of the diagnostic performance of MRA for "any" tears and for partial thickness are not easy to interpret due to conflicting results, and the lack of data to compute average sensitivity and specificity
 - o For example, MRA for any tear is graphed in Figure 5 on page 14; two studies place MRA near the upper left corner of the ROC curve, but one study had MRA below the diagonal line (suggesting that MRA is worse than a coin toss)
- There are additional sources of variability in the fact that being "considered for surgery" does not define a particular shoulder condition, and may differ in different health care settings
- There is one point in the Discussion section which does not follow from the Results section

- The authors suggest in the Discussion that MRI may miss fewer cases of a rotator cuff tear than US, but their Results section, they report no significant difference between MRI and US
- In spite of all of the difficulties in data presentation and analysis, there does appear to be support for the conclusion that all three modalities are well suited to diagnose full thickness tears in patients whose clinical picture suggests that these tears may lead to a need for surgery

Assessment: A high-quality meta-analysis which supports good evidence that MRI, MRA, and US are all accurate at identifying full thickness rotator cuff tears in patients whose history and physical examination makes them candidates for possible surgery, and that there is no evidence to suggest that any of the three is superior for this purpose. Inadequate for evidence about the comparative accuracy in partial thickness tears (due to the way that clinically very different categories were combined in the analysis, leading to equivocal interpretation of the findings)

References:

De Jesus JO, Parker L, et al. Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. *American Journal of Roentgenology* 2009;192(6):1701–7.

Macaskill P, Gatsonis C, et al. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration, 2010.

Smith TO, Daniell H, et al. The diagnostic accuracy of MRI for the detection of partial and full-thickness rotator cuff tears in adults. *Magnetic Resonance Imaging* 2012;30(3):336–46.