Rubinstein SM, Pool JJM, et al. A systematic review of the diagnostic accuracy of provocative tests of the neck for diagnosing cervical radiculopathy. Eur Spine J 2007;16:307-319.

Design: systematic review of studies of diagnostic test accuracy

Brief summary of results:

- The authors did an electronic search of several databases through June 2005: MEDLINE, EMBASE, CINAHL, Medion, OSTMED, and DARE
- Studies were included if they (1) reported results of any provocative test for diagnosing cervical radiculopathy, (2) compared the diagnostic test to any reference standard, (3) allowed for calculation of sensitivity and specificity of the test, and (4) was a full report
    - Provocative tests were the upper limb tension test, the shoulder abduction test, and Spurling's test
    - Reference standard (AKA gold standard) tests included EMG, plain film, or advanced imaging such as MRI, CT or myelography
    - In the authors' judgment, the ideal reference test would be a combination of electrodiagnostic studies and advanced imaging
- Study quality was based on QUADAS, a commonly used tool which examines numerous aspects of studies of diagnostic tests
    - Was the spectrum of patients in the study representative of the patients who will receive the index test in clinical practice? (spectrum bias)
    - Were the inclusion and exclusion criteria clearly described?
    - Was the reference standard likely to diagnose the target condition?
    - Was the time between application of the reference standard and the index test short enough to be reasonably sure that the status of the target condition did not change in the interval between tests? (disease progression bias)
    - Did all patients receive the reference standard, regardless of the results of the index test?
    - Was the execution of the index test described in enough detail to permit its replication?
    - Was the execution of the reference standard described in enough detail to permit its replication?
    - Were the results of the index test interpreted without knowledge of the reference standard results?

    - Were the results of the reference standard interpreted without knowledge of the index test?
    - When the index test was done, were the same clinical data available (e.g., history and general physical exam) that would be available when the index test is performed in clinical practice?
    - Were uninterpretable/intermediate/unclear index test results reported?
    - Were withdrawals from the study reported or explained?
- Two authors independently reviewed studies for the QUADAS criteria, and were in very substantial agreement in their judgments of the criteria
- Only six studies met the inclusion criteria
- Almost all of the studies suffered from at least one serious kind of bias
    - No study used the optimal reference standard of both electrodiagnostic and advanced imaging studies for cervical radiculopathy
    - Five of the six studies had significant spectrum bias
    - Five of the six studies had significant disease progression bias
    - In two of the studies, the index test was interpreted with knowledge of the results of the reference standard
    - In three of the studies, the reference standard was interpreted with knowledge of the index test results
    - In three of the studies, there was potential bias resulting from withdrawals and dropouts from the study
    - Only one study conducted the examinations in a setting of primary care, where the tests are most likely to be needed; the other studies were done in settings where the patients had been referred for suspected radiculopathy
- The studies did generally score well on the description of inclusion and exclusion criteria, the performance of the index test, and the performance of the reference standard
    - However, there was considerable variability in the performance of the index tests even though they were clearly described; there were four studies of Spurling's test, and no two of them carried out the test in exactly the same manner
- Studies which used imaging as a reference standard, because they classify relatively more patients as diseased, probably overestimated sensitivity and underestimated specificity
- Studies which used EDX as a reference standard, by contrast, classify relatively more patients as non-diseased, probably underestimating sensitivity and overestimating specificity
- There were data to suggest that Spurling's test has a fairly high specificity
    - In the only study which escaped serious problems with bias, which used needle EMG as the reference standard, Spurling's test, done without neck

extension and without cervical rotation, had a sensitivity of 50% and a specificity of 86%; the same test done with ipsilateral rotation had a sensitivity of 50% and a specificity of 74$
    - Other studies, with lower quality scores, also reported fairly high specificity (90% or more) for Spurling's with sensitivities of about 50%
- The same study judged to be of adequate quality also reported high specificity for the traction/neck distraction test (90%) and for the Valsalva test (94%), and reported high sensitivity for the upper limb tension test (97% when done with finger and wrist extension, but only 74% when done with finger and wrist flexion)

Authors' conclusions:

- There were three important obstacles to evaluating the diagnostic value of provocative tests:
    - Only six studies met basic inclusion criteria
    - No study used an optimal reference standard
    - There is no standardization in how the index tests should be performed, making it impossible to compare sensitivities and specificities across studies
- These problems collectively preclude making any strong recommendations for the use of these tests in the primary care setting
- Spurling's test, traction/neck distraction, and Valsalva show high specificity, and the upper limb tension test is reasonably sensitive
- None of the tests had both high sensitivity and specificity
- More high quality studies need to be done before recommendations about provocative tests can be made

Comments:

- In the single study with a low risk of bias, there were three tests with fairly high specificity: Spurling's, the traction/neck distraction test, and Valsalva
    - However, in this study, the reference standard was electrodiagnostic testing, which, as the authors note, may overstate the specificity of these tests
    - By the same token, the upper limb tension test, which had a high sensitivity (97%) under the same reference standard, is likely to be valid, since sensitivities tend to be understated in these conditions
    - The upper limb tension test, for maximum sensitivity, should be done in the classical manner (with finger and wrist extension)
- The lamentable state of the art of diagnostic testing does, as the authors note, preclude any strong recommendations about which provocative tests to use

Assessment: A high quality systematic review, with transparent selection and quality criteria, which provides some evidence that the upper limb tension test has a high sensitivity for detecting

cervical radiculopathy, and that the Spurling, Valsalva, and traction/neck distraction tests may be reasonably specific for cervical radiculopathy