J Shoulder Elbow Surg (2017) ■■, ■■–■■

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

A practical, evidence-based, comprehensive (PEC) physical examination for diagnosing pathology of the long head of the biceps

Samuel Rosas, MS[a], Michael K. Krill, BS, ATC[b,c], Kelms Amoo-Achampong, BA[d], KiHyun Kwon, MS[e], Benedict U. Nwachukwu, MD, MBA[f], Frank McCormick, MD[g,h,*]

[a]*Department of Orthopedic Surgery, Baptist Health, Wake Forest University, Winston-Salem, NC, USA*
[b]*Florida Atlantic University Charles E. Schmidt College of Medicine, Boca Raton, FL, USA*
[c]*Jameson Crane Sports Medicine Institute, The Ohio State University Wexner Medical Center, Columbus, OH, USA*
[d]*Icahn School of Medicine at Mount Sinai, New York, NY, USA*
[e]*Florida International University Herbert Wertheim College of Medicine, Miami, FL, USA*
[f]*Department of Orthopedic Surgery, Hospital for Special Surgery, New York, NY, USA*
[g]*Department of Orthopedics, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA*
[h]*Department of Sports Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA*

Background: Clinical examination of the shoulder joint has gained attention as clinicians aim to use an evidence-based examination of the biceps tendon, with the desire for a proper diagnosis while minimizing costly imaging procedures. The purpose of this study is to create a decision tree analysis that enables the development of a clinical algorithm for diagnosing long head of biceps (LHB) pathology.

Methods: A literature review of Level I and II diagnostic studies was conducted to extract characteristics of clinical tests for LHB pathology through a systematic review of PubMed, Medline, Ovid, and Cochrane Review databases. Tests were combined in series and parallel to determine sensitivities and specificities, and positive and negative likelihood ratios were determined for each combination using a subjective pretest probability. The "gold standard" for diagnosis in all included studies was arthroscopy or arthrotomy.

Results: The optimal testing modality was use of the uppercut test combined with the tenderness to palpation of the biceps tendon test. This combination achieved a sensitivity of 88.4% when performed in parallel and a specificity of 93.8% when performed in series. These tests used in combination optimize post-test probability accuracy greater than any single individual test.

Conclusion: Performing the uppercut test and biceps groove tenderness to palpation test together has the highest sensitivity and specificity of known physical examinations maneuvers to aid in the diagnosis of LHB pathology compared with diagnostic arthroscopy (practical, evidence-based, comprehensive examination). A decision tree analysis aides in the practical, evidence-based, comprehensive examination diagnostic accuracy post-testing based on the ordinal scale pretest probability.

Level of evidence: Level II, Systematic Review

Keywords: Biceps tendon; long head; physical examination; pathology; diagnosis; shoulder examination

*Reprint requests: Frank McCormick, MD, 330 Brookline Ave., Boston, MA 02215, USA.
E-mail address: drfrankmccormick@yahoo.com (F. McCormick).

1058-2746/$ - see front matter © 2017 Journal of Shoulder and Elbow Surgery Board of Trustees. All rights reserved.
http://dx.doi.org/10.1016/j.jse.2017.03.002

The physical examination is a requisite and inexpensive component to medical diagnosis. The shoulder examination, in particular, encompasses a myriad of special provocative maneuvers, displaying a wide range of sensitivities and specificities pertaining to diagnostic accuracy. Accurate understanding from the correct sequence of maneuvers or tests increases diagnostic yield.

Clinical diagnosis in the modern era heavily relies on imaging modalities including ultrasound, magnetic resonance imaging (MRI), computed tomography (CT), arthrography, and arthroscopy to diagnose shoulder pathology.[21,33] Current "gold standard" diagnostic testing options have limitations. MRI has poor statistical characteristics for diagnostic accuracy because it is very reader and technician dependent, adds direct and indirect costs, and may be less accurate than the physical examination.[37] Diagnostic arthroscopy is successful in diagnosing intra-articular pathology but is limited in visualization for extra-articular pathology, is costly, and increases patient risk.[37] Increased use of diagnostic imaging contributes to rising health care costs.[14,30,32,38] According to the Centers for Medicare & Medicaid Services, diagnostic imaging costs are significant, accounting for up to 40% of overall health care expenditure increases during the past 10 years.[25] Advanced imaging techniques result in not only higher direct costs but may also increase indirect costs and jeopardize outcomes.[36,39]

As the health care landscape transitions to cost minimization and value-based health care delivery, the development of an efficient, cost-effective, shoulder examination is desired. Shoulder examinations have poor sensitivity or specificity, or both, that makes diagnosing certain pathologies difficult.[4,28,30,33] Thus, evaluating the long head of the biceps brachii tendon (LHB) pathology with high-yield examination maneuvers can aid physicians through increasing the accuracy of shoulder diagnoses and aid in surgical decision making.

Previously published studies focused on the following questions: whether physical examination special tests correlate with surgical findings; whether imaging correlates with surgical findings; and whether physical examination tests are accurate enough to diagnose pathology effectively.[5,9,10,26,28,29,33] Currently, there is a need to develop new algorithms to provide shoulder practitioners with a practical but comprehensive evidence-based approach to diagnose LHB pathology during an office visit and to further reduce the need for diagnostic imaging.[20,22,34]

The purpose of this study was to perform a systematic review and a secondary sensitivity analysis based on preformed likelihood scenarios based on the history of present illness, past medical history, and epidemiology to provide clinicians a practical, evidence-based clinical (PEC) physical examination algorithm to accurately diagnose patients with LHB pathology. Specific objectives were to compile the peak performing physical examination tests extracted from Level I and II studies within the English literature, synthesize the most accurate test combination, develop a clinical algorithm to provide quantify LHB diagnostic accuracy, and create a diagnostic accuracy reference guide.

Materials and methods

A systematic literature review with the terms "proximal," "biceps," "clinical," and "examination" in the PubMed, Ovid, and Cochrane Review databases was completed in May 2015. The searches included the use of Boolean operators such as "and" and "or". The databases were scrutinized independently by 3 authors.

Inclusion criteria included studies that were focused on physical examination tests and compared with the diagnostic "gold standard" from Level I and II studies published in scientific journals. Exclusion criteria were non-English, nonfull text, Level III of evidence or lower, related to superior labrum anterior-to-posterior lesions, investigated rheumatoid arthritis patients, or did not compare tests to a validated "gold standard". The validated "gold standard" used for all included studies and systematic reviews were diagnostic arthroscopy or arthrotomy to confirm anatomic findings.

Relevant studies were independently assessed, and conflicting studies were included only if there were consensus among the authors. References of included studies were evaluated to identify additional articles for inclusion. Applicable data were extracted by reverse calculation where the information desired was not directly stated.

Using Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for systematic reviews (Fig. 1), we retrieved 2086 studies from PubMed, Ovid, and Cochrane Review databases in our original search. A review of references from each article included in the systemic review resulted in 28 additional records. After duplicates were removed, the initial search yielded 2112 studies. Subsequently, 1689 studies were removed for irrelevant titles or abstracts, and an additional 362 were excluded because they were not in English. Lastly, the remaining 61 articles were assessed for eligibility; of these, 14 were excluded for nonfull text, 22 were excluded for not being a Level I or II study, and 18 were excluded for nonrelevant data.

The data extracted were summarized and analyzed according to the statistical methods described by Eusebi et al,[12] focusing on test specificity, sensitivity, positive predictive value, and negative predictive value.

Next, clinical tests were combined to assess improved diagnostic accuracy. The clinical tests were applied in parallel and in series. The first approach, in *parallel* analysis, consists of 2 special tests performed in theory at approximately the same time. The parallel analysis can interpret the findings in an "and" or "or" technique. When a parallel analysis is performed in an "or" technique, the overall sensitivity of the 2 tests is greater than the sensitivity of either special test alone. This parallel analysis allows for 2 opportunities to observe the potential pathology. If both tests are negative, then it is considered a "negative" finding in the algorithm and rules out the pathology, but if just 1 of the 2 special tests is positive, then it is not considered a "negative" result in parallel analysis.[7]

The second approach, in *series* analysis, consists of 2 special tests performed; however, the overall "negative" or "positive" finding depends on the outcomes of both special tests. By using 2 special tests in an "and" technique in series, the specificity for both tests is higher than for either test alone. If both special tests are positive, then it is considered a "positive" result. If either special test is negative, then the in series analysis cannot be considered a "positive" result.[7]