Comparison of the accuracy of telehealth examination versus clinical examination in the detection of shoulder pathology

Kendall E. Bradley, MD[a],*, Chad Cook, PT, PhD, MBA, FAPTA[b], Emily K. Reinke, PhD[c], Emily N. Vinson, MD[d], Richard C. Mather III, MD, MBA[c], Jonathan Riboh, MD[c], Tally Lassiter, MD, MHA[c], Jocelyn R. Wittstein, MD[c]

[a]Department of Orthopaedic Surgery, Duke University Medical Center, Durham, NC, USA
[b]Division of Physical Therapy, Department of Orthopaedic Surgery, Duke University Medical Center, Durham, NC, USA
[c]Division of Sports Medicine, Department of Orthopaedic Surgery, Duke University Medical Center, Durham, NC, USA
[d]Musculoskeletal Radiology, Duke University Medical Center, Durham, NC, USA

Hypothesis/Background: In 2017, the American Orthopaedic Association advocated for the increased use of telehealth as an assessment and treatment platform, and demand has significantly increased during the coronavirus disease 2019 pandemic. Diagnostic effectiveness (also called overall diagnostic accuracy) and reliability of a telehealth clinical examination vs. a traditional shoulder clinical examination (SCE) has not been established. Our objective is to compare the diagnostic effectiveness of a telehealth shoulder examination against an SCE for rotator cuff tear (RCT), using magnetic resonance imaging (MRI) as a reference standard; secondary objectives included assessing agreement between test platforms and validity of individualized tests. We hypothesize that tests provided in a telehealth platform would not have inferior diagnostic effectiveness to an SCE.
Methods: The study is a case-based, case-control design. Two clinicians selected movement, strength, and special tests for the SCE that are associated with the diagnosis of RCT and identified similar tests to replicate for a simulated telehealth-based examination (STE). Consecutive patients with no prior shoulder surgery or advanced imaging underwent both the SCE and STE in the same visit using 2 separate assessors. We randomized the order of the SCE or STE. A blinded reader assessed an MRI, to be used as a reference standard. We calculated diagnostic effectiveness, which provides values from 0% to 100% as well as agreement statistics (Kappa) between tests by an assessment platform, and sensitivity, specificity, and likelihood ratios for individual tests in both SCE and STE. We compared the diagnostic effectiveness (overall) of the SCE and STE with a Mann-Whitney U test.
Results: We included 62 consecutive patients with shoulder pain, aged 40 or older; 50 (81%) received an MRI as a reference standard. The diagnostic effectiveness of stand-alone tests was poor regardless of the group, with the exception of a few tests with high specificity. None had greater than 70% accuracy. There was no significant difference between the overall diagnostic effectiveness of the STE and SCE ($P = .98$). Overall agreement between the STE tests and the SCE tests ranged from poor to moderate (Kappa, 0.07-0.87).
Conclusion: This study identified initial feasibility and noninferiority of the physician-guided, patient-performed STE when compared with an SCE in the detection of RCTs. Although these results are promising, larger studies are needed for further validation of an STE assessment platform.

*Reprint requests: Kendall E. Bradley, MD, Department of Orthopaedic Surgery, Duke University Medical Center, DUMC Box 104002, Durham, NC 27710, USA.
E-mail address: Bradley.kendall@gmail.com (K.E. Bradley).

Shoulder pain is a common cause of disability in the adult population, with rotator cuff tendon tears increasing in frequency each decade after the age of 40 years and affecting approximately 25% of adults over the age of 50.[40,44] Differentiating shoulder pain requires a careful assessment of patient report, physical examination, and imaging. Meta-analyses have demonstrated that the stand-alone diagnostic utility of movement, physical testing, and special tests of the shoulder are poor.[15,16] Imaging fares better and improves the ability to identify full-thickness rotator cuff tears (RCTs) primarily in addition to other shoulder pathologies.[27] These challenges suggest that a shoulder specialist is required to distinguish between the prevalent shoulder pathologies, variabilities in test findings, and imaging findings.[6,17]

Unfortunately, there is a lack of access to subspecialized orthopedic care, including shoulder specialists, in rural settings as compared with urban settings.[21,23,26,28,33] Access issues will worsen with a projected shortfall of 20,000-30,000 surgical specialists by 2030.[18] To address this shortfall, some shoulder surgeons have integrated telehealth as an alternative to conventional clinical care.[22] Telehealth evaluations are cost-effective and provide access to specialized care in a variety of orthopedic conditions in the United States and abroad.[11,19,24,29,38,42,43] In Finland, tele-consultations allowed general practitioners to examine and diagnose 25% of patients who otherwise would have required referral to an outside provider.[11] The United States Army began using telehealth medicine for orthopedic surgery in July 2007 for soldiers deployed overseas.[2] The use of telehealth as an assessment and treatment platform has been advocated by the American Orthopaedic Association.[41]

Public demand for orthopedic telehealth services has surged with the arrival of the coronavirus disease 2019 (COVID-19) pandemic in early 2020. To date, most telehealth-based studies are survey oriented or involve physician and patient perceptions of care. There are no studies that compare the diagnostic effectiveness (also known as overall diagnostic accuracy) of a telehealth examination platform with a standard clinical examination (SCE) of the shoulder. One study examined the accuracy of self-administered hip examination for FAI and actually showed that the accuracy of telehealth-based assessment was slightly higher.[32] Our objective is to compare the diagnostic effectiveness of a simulated telehealth shoulder examination (STE) against an SCE for RCT, using magnetic resonance imaging (MRI) as a reference standard. Secondary objectives included assessing agreement between test platforms and diagnostic validity of individualized tests (eg, sensitivity, specificity, likelihood ratios). We hypothesize that the STE would be noninferior to an SCE in accurately diagnosing RCTs.

Materials and methods

Study design

The study is a case-based, case-control design. We used the Standards for Reporting Diagnostic Accuracy Studies (STARD) reporting standards to guide this study.[3] All care in outpatient, orthopedic specialists' practices was provided by 3 orthopedic surgeons (J.R.W., T.L., J.R.). Patients provided electronic informed consent and then underwent both the SCE and STE during the same visit but by 2 different providers to avoid confirmation bias. We randomized the order in which this examination was performed.

Eligibility and exclusion criteria

We recruited consecutive patients, 40 years of age or older, presenting with shoulder pain and seen in the Duke Sports Medicine Clinic when research staff were available for the study. We excluded patients from the study if the shoulder being evaluated had prior shoulder arthroplasty, instability, prior imaging revealing a rotator cuff injury, history of fracture/dislocation prior advanced imaging, or contraindications to advanced imaging. We did not compensate subjects. We provided MRIs for all patients, paid for through research funds, but not change other patient management for the standard of care.

Index testing

This study involved 2 sets of index tests for 2 assessment platforms (SCE and STE). We selected the SCE tests if they were commonly used tests and measures from the literature and practice, and if they achieved reasonable diagnostic accuracy in the summated study.[16] We designed tests to identify all types of RCTs (ie, supraspinatus, infraspinatus, subscapularis) and selected tests for either having high sensitivity or specificity, or when available, both.

The senior author, an orthopedic surgeon with 10 years of experience, and a physical therapist who is PhD and has specialized in diagnostic accuracy research for 15 years created the STE tests. The goal of both clinicians was to identify tests that reflected a clinical examination; each SCE test had an analogous "sister" STE test they created to reflect its purpose in clinical practice. We included tests that detected other shoulder pathologies in order to provide the standard of care to patients for a shoulder examination. The goal was also to create tests that were

transferable to any telehealth setting with a video feed. A description of these testing procedures is given in Table I.

Clinical testing

We randomized patients to undergo the STE or SCE first. Randomization was assigned using randomly assorted blocks of sizes 4 and 6 and stratified by the attending surgeon.[7]

All patients underwent both examinations. For the SCE, 1 of 3 fellowship-trained sports medicine orthopedic surgeons (J.R.W., T.L., J.R.) performed a set of shoulder examination procedures in the predetermined order. They performed the shoulder examination maneuvers in the same order as the analogous STE maneuvers. To avoid bias, the senior surgeon obtained a detailed history and reviewed radiographs only after all examination procedures were complete.

Simulated telehealth testing

For the STE, time permitting, we invited the patient to view a tutorial video of the STE. A research coordinator used a portable electronic device (Apple iPad; Apple, Cupertino, CA, USA) equipped with a video camera to image the patient, and a senior orthopedic resident (PGY4-6) served as the telehealth examiner performing the physician-guided, patient-performed telehealth examination from a separate room at a desktop interface (Cisco Webex DX80 or Cisco Webex DX70; Cisco, San Jose, CA, USA). We used a senior in order to eliminate the bias of the attending surgeon might have if they performed both examinations. A total of 9 senior residents and fellows participated in the telehealth version of the examination (45% by 4 fellows, 32% by 2 PGY5 residents, and 23% by 3 PGY4 residents). These were the residents and fellows on service with the attending. Both devices had video capability such that the patient and examiner could see, hear, and observe each other. For the STE, the senior author developed a script to ask standardized questions regarding the quality of the shoulder pain in order to minimize differences between trainees. The resident then directed the patient through a series of self-examination maneuvers, using the script provided. The examinations were meant to mimic traditional clinic testing.

Data management

We collected and managed all study data, with the exception of the MRI images, using REDCap electronic data capture tools hosted at Duke University. REDCap (Research Electronic Data Capture; Vanderbilt University, Nashville, TN, USA) is a secure web-based software platform designed to support data capture for research studies, providing: (1) an interface for validated data capture, (2) audit trails for tracking data manipulation and export procedures, (3) export procedures for data downloads to statistical packages, and (4) procedures for data integration and interoperability with external sources.[12,13]

Reference standard

We compared the STE of shoulder pain using patient self-examination with the SCE using MRI as the gold standard[16] to determine the accuracy of detecting RCTs. Fifty (81%) patients had shoulder imaging obtained using MRI with a dedicated shoulder coil. All 50 patients underwent a nonarthrographic shoulder MRI examination performed on a 3.0-Tesla MR scanner (Trio TIM; Siemens Healthcare, Erlangen, Germany) using a phased array 8-channel shoulder coil (Invivo). The imaging protocol consisted of the following sequences: axial, oblique sagittal, and oblique coronal fat-suppressed fast spin-echo T2-weighted sequences (slice thickness, 3.0 mm; FOV, 16 cm; TR/TE, 3000/65); axial fat-suppressed fast spin-echo intermediate-weighted sequence (slice thickness, 3.0 mm; FOV, 16 cm; TR/TE, 3000/23); and oblique sagittal T1-weighted sequence (slice thickness, 3.0 mm; FOV, 16 cm; TR/TE, 688/11). A single musculoskeletal radiologist who was blinded to the clinical findings reviewed the research MRIs prospectively and recorded the presence or absence of partial and complete RCTs as well as tear locations. MRI was considered positive if the tear was complete.

Sample size estimate

When determining the sample size of a noninferiority trial, one aim is to show that a new testing platform is not unacceptably worse than an older one (SCE).[34] To do so, we need to select a noninferiority margin, calculate the confidence window around the difference between the treatments, and then determine the acceptability of difference one is willing to accept. We estimated the sample size with the intention to treat analyses, and these forms of analyses typically lead to noninferiority between groups.[34] We assumed a 20% difference in overall diagnostic effectiveness between groups as unacceptable, and at 95% power, our projected use of a Mann-Whitney U test for comparison of differences, and an error probability of .05, this led to a projected sample size of 60.

Because of COVID-19 and the subsequent closure of the MRI facility, 50 of these patients completed MRIs before the publication of this paper, whereas the remaining 12 subjects were indefinitely unable to access the research scanner. Complete data for all 62 were available in 100% of physical examination test items.

Data analysis and statistical considerations

We performed the analysis using SPSS (V26.0; IBM, Armonk, NY, USA) and a publicly available online software calculator from the University of Illinois, Chicago (http://araw.mede.uic.edu/cgi-bin/testcalc.pl). We first tabulated summary values for age, gender, and diagnoses based on a radiology read.

We calculated agreement between tests using Cohen's Kappa. Cohen's Kappa calculates the chance-corrected agreement between 2 or more raters. Kappa values may range from 0 (perfect lack of agreement) to +1.0 (perfect agreement). It is possible for the statistic to be negative, which suggests that the agreement is worse than random. Although arbitrary, Landis and Koch[25] provided cutoff values for interpretation as follows: <0, no agreement; 0-0.20, slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; and 0.81-1, almost perfect agreement.[25]

We calculated diagnostic accuracy measures of sensitivity (SN), specificity (SP), positive likelihood ratio (LR+), and negative likelihood ratio (LR−) for each component of the STE and SCE. SN is the ability of the test to identify a positive finding when the targeted diagnosis is present. SP is the ability of the test

Table I Shoulder examination

Examination maneuver	Description	Telehealth modification	Positive test
Rotator cuff—supraspinatus/infraspinatus			
Drop arm	Patient actively lowers the arm from the abducted position to the side	No modification	Inability to control the arm while lowering it from flexion
Shoulder shrug[24]	Inability to abduct the arm to 90° without elevation of the scapula	No modification	Scapula elevates
ER lag sign[18]	Elbow passive flexed to 90° with shoulder near maximum external rotation in abduction. Wrist is released	Arm supported by a table; patient passively externally rotates the affected shoulder to the maximal ER, and then releases the support	Shoulder internally rotates/forearm falls toward the table
Active elevation deficit	Active and passive forward flexion recorded, with notation of deficit	Passive forward flexion performed by gripping the wand and pushing the affected shoulder into maximal passive flexion using the contralateral arm	Deficit of active flexion relative to passive flexion
ER weakness	Subjective rating of MMT	Isometric ER testing with resistance by the contralateral hand at the dorsal wrist	MMT <4+ Patient report of subjective weakness
Abduction weakness[26]	–	Isometric abduction testing against the wall	–
ER pain with strength testing	Patient asked if test painful	No modification	+ Pain
Abduction pain with strength testing	–	No modification	–
Rotator cuff—subscapularis			
Lift-off sign[10]	Patient attempts to lift the hand away from his or her back	No modification	Patient unable to lift the hand off back
Belly press[9]	Patient presses the hands to the abdomen while the elbows push forward	No modification	Elbows drop backward with wrist flexion
IR weakness[26]	Subjective rating of MMT	Isometric IR testing with resistance by the contralateral fist at the palmar aspect of the hand	MMT <4+ Patient report of subjective weakness
IR pain with strength testing	Patient asked if test painful	No modification	+ Pain
Impingement			
Hawkins Kennedy[14]	Arm flexed 90° while passively internally rotating the shoulder	Patient flexes the shoulder, supports the elbow, and passively internally rotates the shoulder with the contralateral arm	Pain
Neer's sign[38]	Arm flexed causing impingement of the greater tuberosity against the acromion	Patient forward flexes the shoulder with the hand pronated	Pain at the anterior edge of the acromion

Painful arc[28]	Shoulder actively abducted from 60° to 180°	No modification	Pain with the elevation of the shoulder
Adhesive capsulitis PROM flexion	Examiner passively maximally flexes the arm while the patient standing and controls movement of the scapula	Patient holds the stick vertically and the opposite arm assists forward flexion	0-max degrees (continuous) Positive test was a 25% limitation of the affected side to the nonaffected side
ER affected to contralateral limitation	With elbows at side and elbow flexed to 90°, the examiner maximally externally rotates and repeats on the contralateral side	Patient holds the stick horizontally and the opposite arm assists to externally rotate	0-max degrees (continuous) Positive test was a 10° reduction of the affected side to contralateral
IR limitation	Patient turns around and attempts to reach toward the scapula	No modification	By max level (to lateral thigh, buttock, LS region, waist, T12, interscapular region). Positive test was inability to lift to T12

ER, external rotation; *MMT*, manual motor testing; *IR*, internal rotation; *PROM*, passive range of motion.

to identify a negative finding when the targeted is negative. LR+ indicates a shift in probability supporting the existence of a disorder if the test is found to be positive. Values greater than 1 indicate an increased ability to diagnose a condition with a positive finding. LR− indicates a shift in probability supporting the absence of a disorder if the test is found to be negative. Values less than 1 and closer to 0 are used to rule out a diagnosis. Jaeschke et al[20] provided arbitrary cutoffs for likelihood variables, suggesting that LR+ of >10.0 and LR− of <0.10 have large increases or decreases in the likelihood of the disease. LR+ of >5 to 10 and LR− of <0.20 to 0.10 have moderate increases and decreases in the likelihood of the disease.[20] For this study, we did not attempt to differentiate location of tear during determination of accuracy of each test (ie, subscapularis vs. supraspinatus), although location can be noted.

For our primary objective, we analyzed overall diagnostic effectiveness $((TP + TN)/(TP + TN + FP + FN) \times 100)$.[1,37] We calculated this for both SCE and STE tests and summated to a grand mean. For each individual test, values can range from 0% (completely inaccurate) to 100% (accurate in ruling in and out). We compared summated mean diagnostic effectiveness using a Mann-Whitney U nonparametric test. We defined a statistically significant finding as $P < .05$.

Results

From August 2019 to March 2020, 96 consecutive patients who met the inclusion criteria were considered for enrollment (Fig. 1). Of the 96 patients, 34 declined to participate in the study. Of the remaining 62 patients, 2 were unable to tolerate an enclosed MRI or had a contraindication to MRI, as discovered by the MRI technician. As previously stated, we included 62 patients for the analysis of agreement between tests (secondary objective) and 50 (81%) for diagnostic effectiveness analysis (primary objective).

The mean age was 57.9 years (±11.2) and 31 (51.7%) of these patients were women. The 50 patients who received an MRI exhibited a similar demographic distribution: 52% women, with the mean age of 58.2 years. The final diagnosis, per official radiology read, indicated that 22% of patients had a full-thickness supraspinatus tear, whereas 62% of patients had partial tearing of one of their rotator cuff tendons. All of the full-thickness supraspinatus tears were accompanied by other complete or partial tears, most commonly of the infraspinatus tendon. There were no isolated full-thickness tears of subscapularis nor infraspinatus (Fig. 2).

Table II shows the agreement between the clinical and telehealth testing and includes only those tests that have been represented as rotator cuff–related tests or movements in previous studies.[15] Night pain reported by patients had almost perfect agreement between telehealth examination and clinical testing. The painful arc test, shoulder shrug with active abduction, and active internal rotation limitation all had a moderate amount of agreement. External rotation (ER) lag sign, drop arm test, Neer's sign, abduction
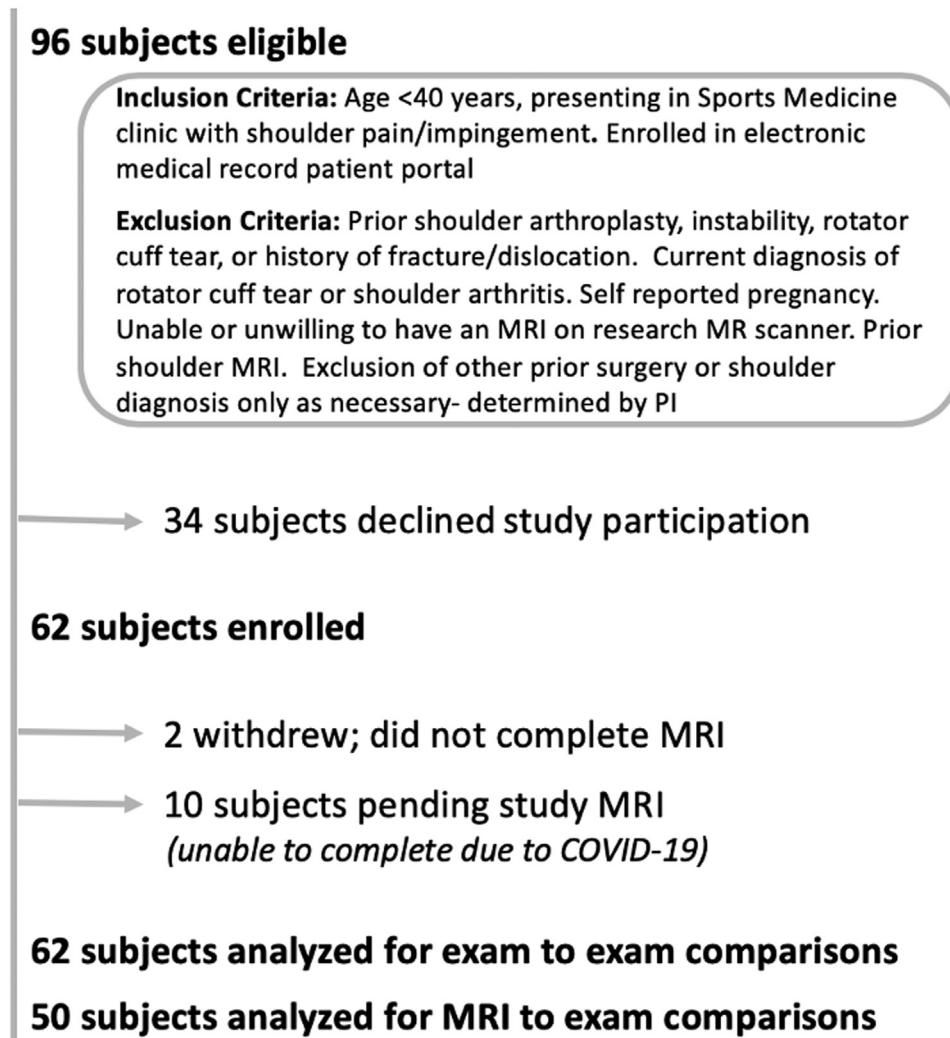
96 subjects eligible

Inclusion Criteria: Age <40 years, presenting in Sports Medicine clinic with shoulder pain/impingement. Enrolled in electronic medical record patient portal

Exclusion Criteria: Prior shoulder arthroplasty, instability, rotator cuff tear, or history of fracture/dislocation. Current diagnosis of rotator cuff tear or shoulder arthritis. Self reported pregnancy. Unable or unwilling to have an MRI on research MR scanner. Prior shoulder MRI. Exclusion of other prior surgery or shoulder diagnosis only as necessary- determined by PI

→ **34 subjects declined study participation**

62 subjects enrolled

→ **2 withdrew; did not complete MRI**

→ **10 subjects pending study MRI**
(unable to complete due to COVID-19)

62 subjects analyzed for exam to exam comparisons
50 subjects analyzed for MRI to exam comparisons

Figure 1   Patient enrollment flowchart.

weakness with strength testing, active to passive flexion limitation, and passive ER affected to contralateral limitation all had fair agreement.

The STE and SCE both had large LR+ for the drop arm test and active to passive flexion limitation (Table III). For the SCE, the ER lag sign also had a high LR+, whereas strength testing with ER had a moderate likelihood ratio, LR+ = 4.92. The remaining tests did not provide moderate to large changes in post-test probability but did exhibit differences depending on the platform. The belly press test and IR weakness test had LR+ around 2.0 for both SCE (2.11 and 2.46, respectively) and STE (2.28 and 2.46, respectively). Whereas the SCE had an increased LR+ for abduction weakness (1.93), the STE did not (1.17) (Table III).

When comparing the diagnostic effectiveness (overall accuracy) for examination maneuvers, the overall accuracy was similar; neither examination was very accurate (Table IV). The mean diagnostic effectiveness for SCE was

45.53% compared with STE, which was 45.72 % (*P* = .98). The SCE and STE both had accuracies of around 60% for the painful arc test and night pain. The STE performed poorly for strength testing, with all values being less than 50%. The SCE performed slightly better for abduction weakness, with a value of 62%. The STE performed better for the lift-off sign (54% vs. 26%) and Neer's sign (66% vs 56%), although these tests still had low accuracy.

Discussion

This study endeavored to compare the diagnostic effectiveness (overall accuracy) of an STE with that of an SCE. In this case-based study, we identified commonly used clinical tests and created tests that would complement the clinical tests but could be used in a simulated telehealth setting. Further, 50 of the 62 subjects received an MRI as a reference standard, in which the rater was blinded to the
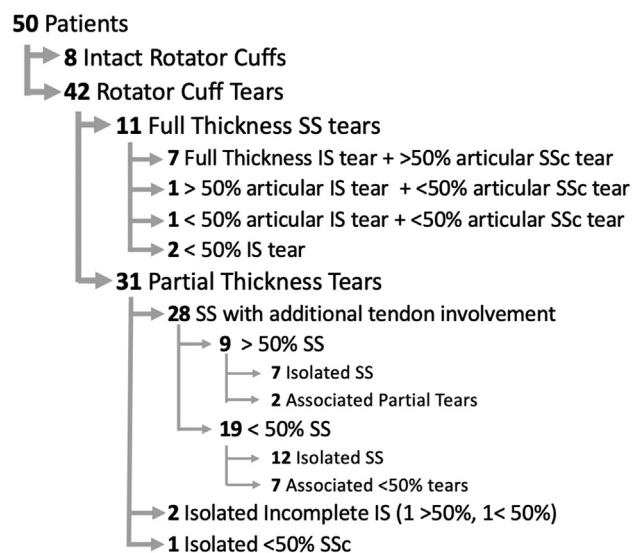
50 Patients
- **8** Intact Rotator Cuffs
- **42** Rotator Cuff Tears
  - **11** Full Thickness SS tears
    - **7** Full Thickness IS tear + >50% articular SSc tear
    - **1** > 50% articular IS tear + <50% articular SSc tear
    - **1** < 50% articular IS tear + <50% articular SSc tear
    - **2** < 50% IS tear
  - **31** Partial Thickness Tears
    - **28** SS with additional tendon involvement
      - **9** > 50% SS
        - **7** Isolated SS
        - **2** Associated Partial Tears
      - **19** < 50% SS
        - **12** Isolated SS
        - **7** Associated <50% tears
    - **2** Isolated Incomplete IS (1 >50%, 1< 50%)
    - **1** Isolated <50% SSc

Figure 2 Location of rotator cuff tear on MRI as read by the radiologist. *SS*, supraspinatus; *IS*, infraspinatus; *SSc*, subscapularis.

clinical findings of the patient. Our goal was to determine if the overall diagnostic effectiveness, which reflects whether true positives and true negatives were better in one group or another. We think our findings are promising and timely, especially during a global pandemic, when virtual appointments become increasingly important. Several areas we feel are worth discussing.

Agreement

There was fair to good agreement between the SCE and STE. Night pain, as expected, had almost perfect agreement. This test was a question to the patient, and it is interesting that the answer changed in the minutes between the SCE and STE for a few patients. Some patients gave a conditional response, such as "I have pain at night, but it doesn't keep me awake," or their answer changed while they were talking. Intuitively, tests that required minimal intervention by an examiner also had higher agreement. For example, the painful arc test, shoulder shrug with active abduction, and active internal rotation limitation all had a moderate amount of agreement. Tests that required explanation or subjective grading, such as the ER lag sign, drop arm test, abduction weakness with strength testing, active to passive flexion limitation, and ER affected to contralateral limitation still exhibited fair agreement.

Diagnostic effectiveness (overall accuracy)

We found that neither the SCE nor STE was accurate to identify an RCT. The tests were usually either sensitive or specific, and in some cases, the tests were neither. Although the accuracy between the SCE and STE was low, these findings are consistent with past meta-analytic literature.[15,16] Particularly, this is well represented in studies with smaller sample sizes that did not differentiate tendon

Table II Agreement between clinical testing and telehealth testing

Tests	Kappa statistic	Strength of agreement	P value
ER lag sign	0.32	Fair	.007
Painful arc test	0.42	Moderate	<.001
Shoulder shrug	0.59	Moderate	<.001
Drop arm test	0.36	Fair	.005
Belly press test	0.17	Slight	.167
Lift-off sign	0.03	Slight	.738
Hawkins Kennedy test	0.07	Slight	.601
Neer's sign	0.22	Fair	.085
Night pain	0.87	Almost perfect	<.001
ER pain with strength testing	0.15	Slight	.243
IR pain with strength testing	0.19	Slight	.107
Abduction pain with strength testing	0.17	Slight	.148
ER weakness with strength testing	0.04	Slight	.758
IR weakness with strength testing	0.14	Slight	.268
Abduction weakness with strength testing	0.32	Fair	.007
IR limitation	0.51	Moderate	<.001
Active to passive flexion limitation	0.35	Fair	.005
ER affected to contralateral limitation	0.27	Fair	.027

ER, external rotation; *IR*, internal rotation.

tear type (we did not) or which included a case-based designed as we did in our study. It is worth noting that emerging data suggest that differentiating tendon type may also lack accuracy.[35,36] Combining tests or examination findings may improve accuracy for the identification of shoulder pathology, although this was not the purpose of our study, nor did we have the power to combine tests to look at conditions.[14,31]

Perception

Most studies evaluating telehealth have focused on the surgeon's subjective perception of the quality of the examination. We did not ask the providers their perception of the SCE and STE in this study. In Norway, surgeons were asked to evaluate how they felt the quality of the appointment was compared with an in-office consultation. A total of 98% of these surgeons felt that the consultations were good or very good, whereas the visits took about the same amount of time.[5] Both visits took about equal amount of time. There have been some studies that have tried to look at a more objective comparison, such as examination with smartphone photography compared with a manual goniometer to measure elbow range of motion.[30] Future studies with this model should also evaluate the difference in time of visit as well as both patient and clinician satisfaction.

Table III Diagnostic accuracy of each clinical and telehealth test and measure

Test	Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio
Clinical testing (SCE)				
ER lag sign	2.70	100	Infinite	0.97
Painful arc test	75.7	23.1	.98	1.05
Shoulder shrug	35.1	69.2	1.14	0.94
Drop arm test	16.2	100	Infinite	0.84
Belly press test	16.2	92.3	2.11	0.91
Lift-off sign	5.4	84.6	0.35	1.12
Hawkins Kennedy test	62.2	46.2	1.15	0.82
Neer's sign	67.6	23.1	0.88	1.41
Night pain	73.0	15.4	0.86	1.76
ER pain with strength testing	48.6	53.8	1.05	0.95
IR pain with strength testing	32.4	30.8	0.47	2.20
Abduction pain with strength testing	64.9	15.4	0.77	2.28
ER weakness with strength testing	37.8	92.3	4.92	0.67
IR weakness with strength testing	18.9	92.3	2.46	0.88
Abduction weakness with strength testing	59.5	69.2	1.93	0.59
IR limitation	54.1	69.2	1.76	0.66
Active to passive flexion limitation	13.5	100	Infinite	0.86
ER affected to contralateral limitation	37.8	46.2	0.7	1.35
Telehealth testing (STE)				
ER lag sign	8.1	92.3	1.05	1.00
Painful arc test	75.7	23.1	0.98	1.05
Shoulder shrug	37.8	76.9	1.64	0.81
Drop arm test	8.3	100	Infinite	0.92
Belly press test	35.1	84.6	2.28	0.77
Lift-off sign	45.9	76.9	1.99	0.70
Hawkins Kennedy test	72.2	15.4	0.85	1.81
Neer's sign	75.7	38.5	1.23	0.63
Night pain	75.7	23.1	0.98	1.05
ER pain with strength testing	45.9	61.5	1.19	0.88
IR pain with strength testing	27.0	84.6	1.76	0.86
Abduction pain with strength testing	48.6	38.5	0.79	1.34
ER weakness with strength testing	21.6	53.8	0.47	1.46
IR weakness with strength testing	18.9	92.3	2.46	0.88
Abduction weakness with strength testing	27.0	76.9	1.17	0.95
IR limitation	18.9	76.9	0.82	1.05
Active to passive flexion limitation	8.1	100	Infinite	0.92
ER affected to contralateral limitation	51.4	61.5	1.34	0.79

SCE, shoulder clinical examination; *ER*, external rotation; *IR*, internal rotation; *STE*, simulated telehealth-based examination.

Table IV Comparative analysis of diagnostic effectiveness (overall accuracy) with 50 patients for rotator cuff tears

The tests	Clinical values of diagnostic accuracy (% correct)	Telehealth values of diagnostic accuracy (% correct)
ER lag sign	28	30
Painful arc test	62	62
Shoulder shrug	44	48
Drop arm test	38	33
Belly press test	34	46
Lift-off sign	26	54
Hawkins Kennedy test	58	56
Neer's sign	56	66
Night pain	58	62
ER pain with strength testing	50	50
IR pain with strength testing	32	42
Abduction pain with strength testing	48	46
ER weakness with strength testing	52	30
IR weakness with strength testing	38	38
Abduction weakness with strength testing	62	40
IR limitation	58	34
Active to passive flexion limitation	36	32
ER affected to contralateral limitation	40	54
Mean diagnostic accuracy	45.53%	45.72%

ER, external rotation; *IR*, internal rotation.
P value = .961 (no significant difference).

Limitations

This study had a relatively small sample size, and although the results are promising, larger studies will be needed for validation. Our study was also limited by only being able to complete 50 of 62 (81%) MRIs before restrictions on research being placed due to COVID-19. We calculated our power analysis using a 20% difference in overall diagnostic effectiveness. Although this may seem high, diagnostic tests are typically either sensitive or specific, with the majority of tests ranging between 50% and 70% accuracy; thus a 20% difference in plausible. That said, we are comfortable in reporting that there are no differences among groups. If we used the current data, the same statistical analyses, and differences between groups, our post hoc power analysis suggests that we would need over 60,000 subjects in each group to show a statistically significant difference.

Although we describe the locations of the specific tendon tears, we did not analyze overall accuracy for full-thickness tears by individual tendon or other shoulder pathologies. Emerging data question the utility of this with clinical tests.[35,36] Out of concern for asymptomatic cuff tears, we have intentionally included only full-thickness tears and not partial-thickness tears in our analysis of accuracy for detecting cuff tears. We felt that new shoulder pain in patients over the age of 40 with full-thickness RCTs would likely be symptomatic, possibly with other pathologies as well. Future research into this data set will address alternative diagnoses.[4,9,16,35,39] We performed pooled accuracy, rather than dividing specific tests by tear type, as it is well known that clinical tests do not effectively distinguish tears of the rotator cuff across all tendon types. Indeed, cross-over positive findings are very common with some diagnostic clinical tests being used for multiple conditions.[10] For example, a drop arm test will be positive for a supraspinatus tendon injury, but it is also positive for infraspinatus problems as well as impingement.

Finally, different trainees performed telehealth examinations, which may have led to some bias on how patients were "coached." Subsequent larger studies may benefit from using a single telehealth examiner for consistency of administration and for this examination to be performed by a second senior surgeon instead of a trainee. The age of the trainee may have led to increased ease of use of telehealth technology, as compared with an attending surgeon. We attempted to minimize the effect of the telehealth examiner variation by providing an instructional video to the patients before participation and by using a script for the administration of the examination. The senior author also reviewed the examination with the trainee before their administration of the examination. We felt that it was important for the senior author to perform all in-person examinations as part of routine clinical care; however, randomization of the senior author to perform either a telehealth examination or an in-person examination may decrease bias in the future.

Future directions

We felt that, although we were unable to complete all 62 MRI studies, the utility of this study revealing non-inferiority of a telehealth examination was important in the wake of the current global pandemic. Future studies should expand the sample size, consider analyzing tests specifically for each RCT lesion, consider accuracy for other shoulder pathologies, and determine cost-effectiveness and management consequences examination using the STE. In addition, assessing patient satisfaction is important as telehealth continues to increase in use. It will also be important to note the impacts of misdiagnosis and malpractice as the telehealth use increases.[8] We also feel that the data gleaned from this study will be useful in future studies of clinical decision-making based on telehealth examination findings and encounters.

Conclusion

This study demonstrated the noninferiority of an STE for rotator cuff pathology. This may increase the geographic footprint of health care networks and give providers an opportunity to evaluate patients in the midst of a pandemic. Future studies are underway to test the accuracy of the STE for different shoulder pathologies, assess clinical decision-making based on the STE, evaluate patient satisfaction, and calculate the cost-effectiveness of this tool.

Disclaimer

Acknowledgments

References

1. Baratloo A, Hosseini M, Negida A, El Ashal G. Part 1: simple definition and calculation of accuracy, sensitivity and specificity. Emerg (Tehran) 2015;3:48-9.

2. Blank E, Lappan C, Belmont PJ Jr, Shaun Machen M, Ficke J, Pope R, et al. Early analysis of the United States Army's telemedicine orthopaedic consultation program. J Surg Orthop Adv 2011;20:50.

3. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Clin Chem 2015;61:1446-52. https://doi.org/10.1373/clinchem.2015.246280

4. Brockmeyer M, Schmitt C, Haupert A, Kohn D, Lorbach O. Limited diagnostic accuracy of magnetic resonance imaging and clinical tests for detecting partial-thickness tears of the rotator cuff. Arch Orthop Trauma Surg 2017;137:1719-24. https://doi.org/10.1007/s00402-017-2799-3

5. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Quality of care for remote orthopaedic consultations using telemedicine: a randomised controlled trial. BMC Health Serv Res 2016;16:483. https://doi.org/10.1186/s12913-016-1717-7

6. Dinnes J, Loveman E, McIntyre L, Waugh N. The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review. Health Technol Assess 2003;7:1-166.iii. https://doi.org/10.3310/hta7290

7. Sealed Envelope. Create a blocked randomisation list. Available at: https://www.sealedenvelope.com/simple-randomiser/v1/lists; 2019. Accessed February 11, 2015.

8. Fogel AL, Kvedar JC. Reported cases of medical malpractice in direct-to-consumer telemedicine. JAMA 2019;321:1309-10. https://doi.org/10.1001/jama.2019.0395

9. Gismervik SØ, Drogset JO, Granviken F, Rø M, Leivseth G. Physical examination tests of the shoulder: a systematic review and meta-analysis of diagnostic test performance. BMC Musculoskelet Disord 2017;18:41. https://doi.org/10.1186/s12891-017-1400-0

10. Hanchard NC, Lenza M, Handoll HH, Takwoingi Y. Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement. Cochrane Database Syst Rev 2013;2013:CD007427. https://doi.org/10.1002/14651858.CD007427.pub2

11. Harno K, Arajärvi E, Paavola T, Carlson C, Viikinkoski P. Clinical effectiveness and cost analysis of patient referral by videoconferencing in orthopaedics. J Telemed Telecare 2001;7:219-25.

12. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208. https://doi.org/10.1016/j.jbi.2019.103208

13. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377-81. https://doi.org/10.1016/j.jbi.2008.08.010

14. Hegedus EJ, Cook C, Lewis J, Wright A, Park J-Y. Combining orthopedic special tests to improve diagnosis of shoulder pathology.

Phys Ther Sport 2015;16:87-92. https://doi.org/10.1016/j.ptsp.2014.08.001

15. Hegedus EJ, Goode A, Campbell S, Morin A, Tamaddoni M, Moorman CT, et al. Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. Br J Sports Med 2008;42:80-92. https://doi.org/10.1136/bjsm.2007.038406

16. Hegedus EJ, Goode AP, Cook CE, Michener L, Myer CA, Myer DM, et al. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. Br J Sports Med 2012;46:964-78. https://doi.org/10.1136/bjsports-2012-091066

17. Iannotti JP, Ciccone J, Buss DD, Visotsky JL, Mascha E, Cotman K, et al. Accuracy of office-based ultrasonography of the shoulder for the diagnosis of rotator cuff tears. J Bone Joint Surg Am 2005;87:1305-11. https://doi.org/10.2106/JBJS.D.02100

18. IHS Markit. The complexities of physician supply and demand 2017 update: projections from 2015 to 2030. Washington, DC: Association of American Medical Colleges; 2017.

19. Jacklin P, Roberts J, Wallace P, Haines A, Harrison R, Barber J, et al. Virtual Outreach Project Group Virtual outreach: economic evaluation of joint teleconsultations for patients referred by their general practitioner for a specialist opinion. BMJ 2003;327:84. https://doi.org/10.1136/bmj.327.7406.84

20. Jaeschke R, Guyatt GH, Sackett DL, Guyatt G, Bass E, Brill-Edwards P, et al. Users' guides to the medical literature: III. How to use an article about a diagnostic Test B. What are the results and will they help me in caring for my patients? JAMA 1994;271:703-7.

21. Jetty A, Moore MA, Coffman M, Petterson S, Bazemore A. Rural family physicians are twice as likely to use telehealth as urban family physicians. Telemed J E Health 2018;24:268-76. https://doi.org/10.1089/tmj.2017.0161

22. Kane LT, Thakar O, Jamgochian G, Lazarus MD, Abboud JA, Namdari S, et al. The role of telehealth as a platform for postoperative visits following rotator cuff repair: a prospective, randomized controlled trial. J Shoulder Elbow Surg 2020;29:775-83. https://doi.org/10.1016/j.jse.2019.12.004

23. Kirch DG, Petelle K. Addressing the physician shortage: the peril of ignoring demography. JAMA 2017;317:1947-8. https://doi.org/10.1001/jama.2017.2714

24. Koutras C, Bitsaki M, Koutras G, Nikolaou C, Heep H. Socioeconomic impact of e-health services in major joint replacement: a scoping review. Technol Health Care 2015;23:809-17. https://doi.org/10.3233/THC-151036

25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

26. Lee ACW, Harada N. Telehealth as a means of health care delivery for physical therapist practice. Phys Ther 2012;92:463-8. https://doi.org/10.2522/ptj.20110100

27. Lenza M, Buchbinder R, Takwoingi Y, Johnston RV, Hanchard NC, Faloppa F. 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 Syst Rev 2013;2013:CD009020. https://doi.org/10.1002/14651858.CD009020.pub2

28. Lishner DM, PhD MR, Levine P, Patrick D. Access to primary health care among persons with disabilities in rural areas: a summary of the literature. J Rural Health 1996;12:45-53.

29. McGill A, North J. An analysis of an ongoing trial of rural video-conference fracture clinics. J Telemed Telecare 2012;18:470-2. https://doi.org/10.1258/jtt.2012.gth110

30. Meislin MA, Wagner ER, Shin AY. A comparison of elbow range of motion measurements: smartphone-based digital photography versus goniometric measurements. J Hand Surg Am 2016;41:510-5.e511. https://doi.org/10.1016/j.jhsa.2016.01.006

31. Moulton SG, Greenspoon JA, Millett PJ, Petri M. Suppl 1: M3: risk factors, pathobiomechanics and physical examination of rotator cuff tears. Open Orthop J 2016;10:277. https://doi.org/10.2174/1874325001610010277

32. Owusu-Akyaw KA, Hutyra CA, Evanson RJ, Cook CE, Reiman M, Mather RC. Concurrent validity of a patient self-administered examination and a clinical examination for femoroacetabular impingement syndrome. BMJ Open Sport Exerc Med 2019;5:e000574. https://doi.org/10.1136/bmjsem-2019-000574

33. Regan S, Wong ST. Patient perspectives on primary health care in rural communities: effects of geography on access, continuity and efficiency. Rural Remote Health 2009;9:1142.

34. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. Trials 2011;12:106. https://doi.org/10.1186/1745-6215-12-106

35. Sgroi M, Loitsch T, Reichel H, Kappe T. Diagnostic value of clinical tests for supraspinatus tendon tears. Arthroscopy 2018;34:2326-33. https://doi.org/10.1016/j.arthro.2018.03.030

36. Sgroi M, Loitsch T, Reichel H, Kappe T. Diagnostic value of clinical tests for infraspinatus tendon tears. Arthroscopy 2019;35:1339-47. https://doi.org/10.1016/j.arthro.2018.12.003

37. Šimundić A-M. Measures of diagnostic accuracy: basic definitions. EJIFCC 2009;19:203.

38. Taylor M, Caffery LJ, Scuffham PA, Smith AC. Economic modelling of telehealth substitution of face-to-face specialist outpatient consultations for Queensland correctional facilities. Aust Health Rev 2018;42:522-8. https://doi.org/10.1071/AH17135

39. van Kampen DA, van den Berg T, van der Woude HJ, Castelein RM, Scholtes VA, Terwee CB, et al. The diagnostic value of the combination of patient characteristics, history, and clinical shoulder tests for the diagnosis of rotator cuff tear. J Orthop Surg Res 2014;9:70. https://doi.org/10.1186/s13018-014-0070-y

40. Vos T, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017;390:1211-59. https://doi.org/10.1016/S0140-6736(17)32154-2

41. Wongworawat MD, Capistrant G, Stephenson JM. The opportunity awaits to lead orthopaedic telehealth innovation: AOA critical issues. J Bone Joint Surg Am 2017;99:e93. https://doi.org/10.2106/JBJS.16.01095

42. Wood G, Naudie D, MacDonald S, McCalden R, Bourne R. An electronic clinic for arthroplasty follow-up: a pilot study. Can J Surg 2011;54:381. https://doi.org/10.1503/cjs.028510

43. Wootton R. Telemedicine and isolated communities: a UK perspective. J Telemed Telecare 1999;5:27-34.

44. Yamamoto A, Takagishi K, Osawa T, Yanagawa T, Nakajima D, Shitara H, et al. Prevalence and risk factors of a rotator cuff tear in the general population. J Shoulder Elbow Surg 2010;19:116-20. https://doi.org/10.1016/j.jse.2009.04.006