HIR
Healthcare Informatics Research

Design and Validation of a Computer Application for Diagnosis of Shoulder Locomotor System Pathology

Albert Bigorda-Sague[1,2,3], Javier Trujillano Cabello[2,3,4], Gemma Ariza Carrio[2,3,4], Carmen Campoy Guerrero[2,5]

[1]Department of Rehabilitation, Santa Maria University Hospital, Lleida, Spain
[2]University of Lleida, Lleida, Spain
[3]The Lleida Biomedical Research Institute's Dr. Pifarré Foundation (IRBLleida), Lleida, Spain
[4]Intensive Care Unit and Department of Rehabilitation, Arnau de Vilanova University Hospital, Lleida, Spain
[5]Department of Rehabilitation, Onze de Setembre Primary Care Center, Lleida, Spain

Objectives: To design and validate a computer application for the diagnosis of shoulder locomotor system pathology. **Methods:** The first phase involved the construction of the application using the Delphi method. In the second phase, the application was validated with a sample of 250 patients with shoulder pathology. Validity was measured for each diagnostic group using sensitivity, specificity, and positive and negative likelihood ratio (LR(+) and LR(−)). The correct classification ratio (CCR) for each patient and the factors related to worse classification were calculated using multivariate binary logistic regression (odds ratio, 95% confidence interval). **Results:** The mean time to complete the application was 15 ± 7 minutes. The validity values were the following: LR(+) 7.8 and LR(−) 0.1 for cervical radiculopathy, LR(+) 4.1 and LR(−) 0.4 for glenohumeral arthrosis, LR(+) 15.5 and LR(−) 0.2 for glenohumeral instability, LR(+) 17.2 and LR(−) 0.2 for massive rotator cuff tear, LR(+) 6.2 and LR(−) 0.2 for capsular syndrome, LR(+) 4.0 and LR(−) 0.3 for subacromial impingement/rotator cuff tendinopathy, and LR(+) 2.5 and LR(−) 0.6 for acromioclavicular arthropathy. A total of 70% of the patients had a CCR greater than 85%. Factors that negatively affected accuracy were massive rotator cuff tear, acromioclavicular arthropathy, age over 55 years, and high pain intensity ($p < 0.05$). **Conclusions:** The developed application achieved an acceptable validity for most pathologies. Because the tool had a limited capacity to identify the full clinical picture in the same patient, improvements and new studies applied to other groups of patients are required.

Keywords: Software, Medical Informatics Applications, Self-Examination, Shoulder, Sensitivity and Specificity

Corresponding Author
Albert Bigorda-Sague
Department of Rehabilitation, Santa Maria University Hospital, Alcalde Rovira Roure, Street 44, Lleida 25198, Spain. Tel: +34-65-63-60-033, E-mail: albertbs@gmail.com (https://orcid.org/0000-0001-7282-0938)

I. Introduction

Shoulder problems rank third as the reason for primary care visits due to musculoskeletal causes [1]. In many cases, physical therapy is beneficial; however, the number of patients who do not receive treatment or receive it late is high. A British study [2] found that only 14% of patients who request care for shoulder pain are referred for physical therapy over a period of three years from the first consultation. Furthermore, the wait time for physical therapy is usually delayed, being more than three months in half of patients [3]. There is evidence linking delayed diagnosis and care with a worse outcome [4].

Hence, it is necessary to develop systems or methods that allow patients to avoid excessive delays in the search for specialized care [5,6]. Subacute or chronic shoulder pain according to its clinical characteristics is likely to benefit from a computer application that detects the presence of musculoskeletal pathology considering that conditions of the shoulder often follow the same symptomatic patterns. Expert systems are an excellent tool, with low cost and easy accessibility, to alert patients about a pathological condition and refer them to a professional who can assess and treat them. The use of this type of application in the health field has increased considerably with the explosion in the market of mobile devices and the generalization of internet access. Currently, there are more than 260,000 indexed health applications [7], and more than 70% of the world population is estimated to be interested in accessing at least one of them [8]. Similarly, the search for online advice has shown remarkable growth. A study conducted in the United States found that 35% of the population has consulted the internet specifically to find out what disease they or someone they knew had, and 53% of users who searched for an online diagnosis consulted a doctor to discuss what they had found on the internet [9].

A computer application offering pathological guidance on the shoulder and refers the patient to a rehabilitation center for assessment and optional physical therapy could save time, cost, and suffering among users and improve patient management for treatment [10]. However, expert systems have focused on general medicine and are not useful or specific for locomotor system pathologies. Thus far, there is no known self-administered application for the diagnosis of musculoskeletal shoulder pathology that has undergone validation. Additionally, the content, appearance, and optimal algorithms for such an application are unknown. To this end, the objective of our study was to design a computer application for suspicion of shoulder locomotor system pathology and to validate it in a sample of patients.

II. Methods

The study was performed in two phases. The first phase comprised the design of the application, and the second phase involved the validation of the application in a group of patients with diagnosed shoulder pathology.

1. Phase 1 (Design of the Application)

We conducted a literature review through an electronic search of the online databases of PubMed, Scopus, Web of Science, and the Cochrane Library for relevant studies. We used the following keywords: shoulder, physical examination, sensitivity, and specificity. Articles (41 out of 246) that fulfilled the condition of including the sensitivity and specificity values of the tests were selected. The Delphi method (with 3 rounds) was used to design the application. A panel of experts was selected, consisting of a physician specializing in rehabilitation, one specializing in traumatology, two physical therapists, a clinical psychologist and a methodology specialist. All of these professionals had more than 10 years of experience in their respective fields. The first Delphi round was used to determine the clinical exploratory tests to be included in the application. Tests were selected by consensus (unanimity or majority agreement) based on the survey answers of the experts. In the second round, the first prototype of the application was enabled to be self-administered with seven diagnostic groups [11]: subacromial impingement or rotator cuff tendinopathy (SI), acromioclavicular arthropathy (AA), capsular syndrome (CS), massive rotator cuff tear (MRCT), glenohumeral arthritis (GA), glenohumeral instability (GI), and cervical radiculopathy (R). Also, two subgroups of rotator cuff tear were analyzed: external rotator cuff tear (ERCT) and internal rotator cuff tear (IRCT). This first prototype was assessed with a pilot sample of 36 patients with a diagnosis of shoulder pathology. The test-retest reliability (with an interval of 3 to 4 days) was assessed using the kappa coefficient. The mean time to complete the application was recorded. The changes suggested by the patients (aimed at making the questions and the type of exploratory test to be performed more comprehensible) were incorporated into the application. In the third Delphi round, the algorithms that were considered redundant were removed, and the final format of the application was completed.

The application was designed for a tablet-type device (10.1-inch touchscreen) with the PowerPoint program with internal logic programmed in Visual Basic and a graphic interface that included images and videos demonstrating the requested exploratory tests.

2. Phase 2 (Validation of the Application)

To validate the application, a cohort (study period between July 2016 and March 2017) of patients diagnosed with shoulder pathology by the Traumatology and Rehabilitation Services of the Arnau de Vilanova and Santa Maria de Lleida University Hospitals was used. The inclusion criterion was the presence of shoulder pain for more than 6 weeks. The exclusion criteria were being under 18 years of age, a deficit

or physical disability that made it impossible to perform the assessments, not knowing how to read or a language barrier that would prevent understanding the text, history of shoul-

der surgery, shoulder fracture occurring less than 6 months prior, central nervous system diseases, hemiplegia or/and paralysis of the upper limb, and insufficient cognitive capac-

Table 1. Summary of exploratory tests to evaluate the diagnostic groups

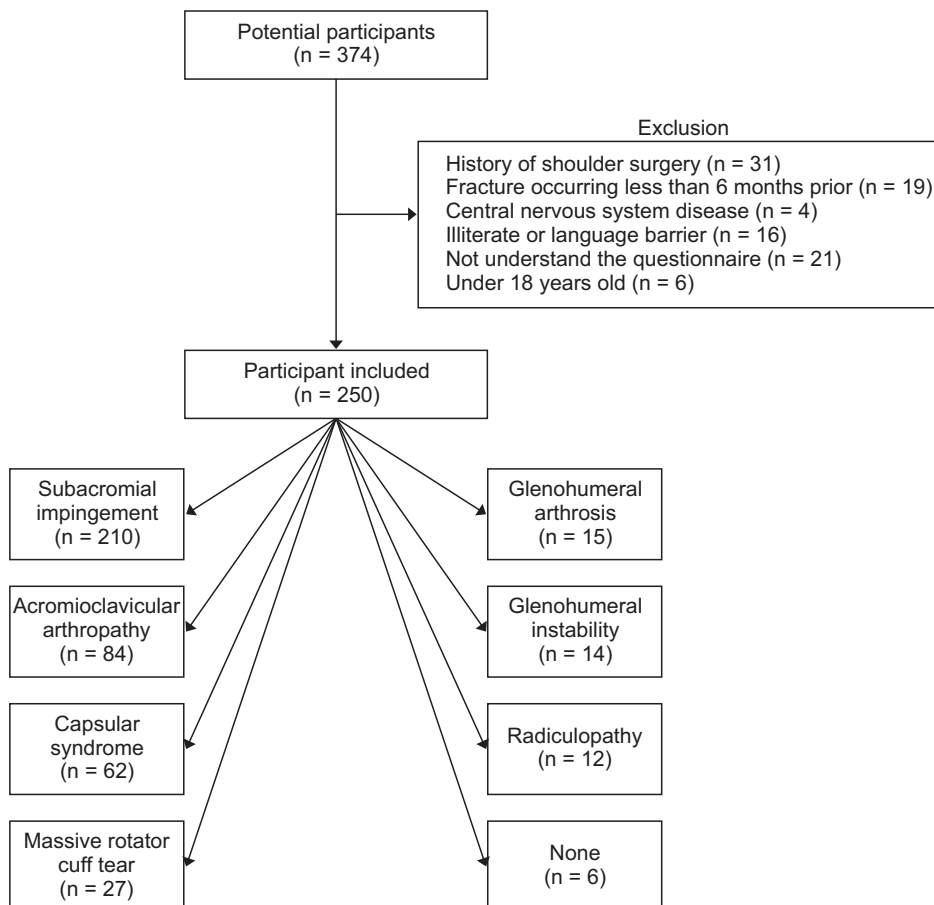| Group	Abbreviation	Description
Radiculopathy	R	Symptoms in the shoulder by the active mobilization of the cervical spine
Glenohumeral arthrosis	GA	Differential clinical assessments of glenohumeral arthrosis
Glenohumeral instability	GI	Clinical assessment of instability, history of joint dislocation, or positivity in the adapted Apprehension test
Massive rotator cuff tear	MRCT	Clinical assessment of muscular deficit, adaptation of the Drop Arm test, and adaptation of the Drop Sign test (with passive mobility preserved)
Capsular syndrome	CS	Clinical assessment of restriction or joint blockade together with a passive external rotation deficit
Acromioclavicular arthropathy	AA	Clinical assessment of pain primarily on mobilization of the shoulder, specifically in the area of the joint and during the performance of an adaptation of the Cross-Body Adduction test
Subacromial impingement	SI	Clinical assessment of pain primarily on mobilization of the shoulder, specifically in the anterior/lateral aspect of the arm during the performance of an adaptation of the Cross-Body Adduction test or by means of an adaptation of the Painful Arc test

Figure 1. Flow chart of the validation group (n = 250).

ity to complete the questionnaire. A convenience sample was used to ensure a minimum number of members for each diagnostic group of more than 10 patients.

Demographic variables (age, sex), educational level (primary, basic, secondary, and higher education), and shoulder pain characteristics (time since onset of pain, intensity according to the visual analogue scale [VAS], and laterality) were collected.

The study was approved by the ethics committees of both hospitals. Patients agreed to participate by signing an informed consent form.

Discrete variables were described by absolute and relative frequency, and continuous variables were described by mean and standard deviation. The sensitivity, specificity, and positive and negative likelihood ratio (LR) were measured for the seven diagnostic groups. If a diagnostic group could be established with different algorithms in the application, the algorithms were evaluated separately.

The diagnostic accuracy of the application, referring to the patients, was determined by the correct classification rate (CCR) in the presence or absence of the seven diagnostic groups in the same patient. A patient was classified as a diagnostic failure (false diagnosis-FALSED) when his or her CCR did not exceed 85%. A multiple binary logistic regression model with calculation of the odds ratio and its 95% confidence interval were used to determine the factors related to FALSED. The demographic variables, pain intensity, and diagnostic groups were included in the model. A stepwise variable selection procedure was used. The SPSS version 24.0 program was used for the statistical calculation.

III. Results

1. Application Design Process
Supplementary Table S1 of the article includes screenshots of the final version of the application. This also includes the bibliographic support for each screen and specifies the most important changes that were made in the different rounds of the Delphi method and the pilot test. Table 1 lists the exploratory tests used to determine each diagnostic group, and Supplementary Figure S1 includes the set of algorithms that constitute the final application. In the pilot group (n = 36), the test-retest reliability obtained a kappa index of 0.66, and the mean time to complete the application was 15 ± 7 minutes.

2. Application Validation
The selection and exclusion criteria of the validation group

(n = 250) are shown in Figure 1. The distribution of the different diagnostic groups was not uniform; there were 210 cases of subacromial impingement or rotator cuff tendinopathy (84%), 84 cases of arthropathy of the acromioclavicular joint (34%), 62 cases of capsular syndrome (25%), 27 cases of massive rotator cuff tear (11%), 15 cases of glenohumeral arthrosis (6%), 14 cases of glenohumeral instability (6%), 12 cases of cervical radiculopathy (5%), and 6 cases of unknown cause or other pathologies (2%). It should be noted that 57% of the patients had two or more diagnoses in the same shoulder.

Table 2 shows the demographic and shoulder pain characteristics. It is important to note that patients reported high pain intensity with 12% of the patients reporting a score >8 on the VAS.

Table 2. Demographic and pain characteristics (n = 250)

Variable	Value
Sex, female	143 (57.2)
Age (yr)	
18–25	7 (2.8)
26–35	6 (0.4)
36–45	30 (12.0)
46–55	61 (24.4)
56–65	63 (25.2)
66–75	51 (20.4)
>75	32 (12.8)
Education level	
Primary	87 (34.8)
Basic	61 (24.4)
Secondary	61 (24.4)
Higher	41 (16.4)
Time since onset of pain	
6–8 weeks	13 (5.2)
2–3 months	13 (5.2)
3–6 months	35 (14)
6 months–1 year	70 (28)
1–3 years	73 (29.2)
3–6 years	19 (7.6)
>6 years	27 (10.8)
Bilateral	41 (16.4)
Diagnostic groups	1.7 ± 1
Pain intensity (VAS)	6.4 ± 2

Values are presented as number (%) or mean ± standard deviation.
VAS: visual analogue scale.

Table 3 shows the diagnostic accuracy for each of the seven diagnostic groups. Some of the groups required more than one algorithm to reach their diagnosis.

The accuracy measured for each patient with the correct classification ratio served to differentiate patients with a good classification (32.4% of users with full accuracy and 37.6% with accuracy greater than 85%) compared with those with a bad classification (30% of users with accuracy worse than 85%) that were assigned to the FALSED group.

Table 4 shows the factors that were independently associated with a worse classification (FALSED). There was one patient-dependent factor (age), one factor related to shoulder pain intensity (VAS > 8), and two factors that belonged to the most difficult to classify diagnostic groups (MRCT and AA).

IV. Discussion

When comparing our application with other published ones, we did not find any similar that followed a construction and validation process. We also did not find applications that fulfilled the objective of being self-administered or that would work in a graphical environment with a tactile system for selecting responses. We only found some related applications.

Gartner et al. [12] designed a questionnaire with 55 clinical history-related questions to diagnose the 24 most frequent shoulder pathologies through a computer program. The program correctly classified the diagnoses in 55.4% of the total sample, and the correct diagnosis was among the first three suggestions of the program in 80% of the patients. It is

important to note that all patients in that study had a single diagnosis (compared to only 43% in our sample). Unlike our study, the questionnaire was printed on a sheet of paper without illustrations or animations, its transcription to a computer was required for interpretation, and it only asked about elements of the patient's history (did not request tests or auto-examinations). Farmer et al. designed a tool that diagnosed the six most common musculoskeletal pathologies in the shoulder. In a first trial [13], it obtained a validity of 88%; and in a second trial, with an updated version [14], it obtained a validity of 91%. However, this tool was managed by a medical professional and not self-administered by the patient as in our study, which may explain the much higher values obtained. It is important to note that the diagnostic groups chosen by the researchers to define musculoskeletal shoulder pathology were the same as those selected in our study. Honekamp and Hanseroth [15] evaluated an online

Table 4. Multivariate logistic regression analysis of a worse diagnostic (FALSED)

Variable	OR (95% CI)
MRCT group	4.5 (1.7–11.8)
AA group	2.0 (1.1–3.7)
Older than 55 years	2.9 (1.4–5.8)
Pain intensity (VAS > 8)	4.5 (1.8–11.1)

MRCT: massive rotator cuff tear, AA: acromioclavicular arthropathy, VAS: visual analogue scale, OR: odds ratio, CI: confidence interval.

Table 3. Accuracy measures for the diagnostic groups

	Sensitivity	Specificity	LR(+)	LR(−)
R	91.7 (59.8–99.6)	88.2 (82.8–92.2)	7.8 (5.2–11.8)	0.1 (0.0–0.6)
GA	66.7 (38.7–87.0)	83.8 (78.4–88.2)	4.1 (2.6–6.5)	0.4 (0.2–0.8)
GI (1)	100 (73.2–99.3)	98.3 (95.4–99.5)	59.0 (22.3–156)	-
GI (2)	14.3 (2.5–43.8)	100 (98.0–100)	-	0.9 (0.7–1.1)
GI (3)	78.6 (48.8–94.3)	94.9 (91.1–97.2)	15.5 (8.4–28.6)	0.2 (0.1–0.6)
MRCT (1)	83.3 (50.9–97.1)	95.1 (91.0–97.5)	17.2 (8.9–33.1)	0.2 (0.0–0.6)
MRCT (2)	61.5 (32.3–84.9)	98.1 (94.9–99.4)	32.5 (11.2–93.8)	0.4 (0.2–0.8)
CS	79.0 (66.5–87.9)	87.2 (81.4–91.5)	6.2 (4.2–9.2)	0.2 (0.1–0.4)
SI (1)	32.9 (26.6–39.7)	92.5 (78.5–98.0)	4.4 (1.5–13.2)	0.7 (0.6–0.8)
SI (2)	76.2 (69.7–81.7)	80.0 (63.9–90.4)	3.8 (2.0–7.1)	0.3 (0.2–0.4)
SI (3)	80.0 (73.8–85.1)	80.0 (63.9–90.4)	4.0 (2.1–7.5)	0.3 (0.2–0.3)
AA	50.0 (39.0–61.0)	80.1 (73.1–85.7)	2.5 (1.7–3.7)	0.6 (0.5–0.8)

The values in parentheses represent the 95% confidence interval. Sensitivity and specificity are presented in percentage.
LR(+): positive likelihood ratio, LR(−): negative likelihood ratio. Other abbreviations are listed in Table 1.

information system focused on shoulder pathologies. The aim of their study was to show that a guided search offered better and more satisfactory results than generic search engines or health portals. To do this, a search engine prototype was used at the Zittau/Görlitz University of Applied Sciences in which the participants had to perform a search based on a hypothetical clinical history. The results demonstrated that the searches in the experimental group were less difficult, twice as fast, and more effective (74% success compared with 23% in the control group). Again, this computer tool differs from our application because the sequence of screens does not follow a predetermined logic and because the web portal does not ask about the patient's condition to offer a diagnosis; rather, the patient must search for a diagnosis through an information portal. However, this study confirmed that guided searches in shoulder pathology, such as ours, are faster, friendlier, and more effective than those carried out by the majority of the population in generic search portals. There are other applications aimed at the diagnosis of pathologies typical of general medicine [8,16], such as iTriage, WebMD, Isabel, DxPlain, Diagnosis Pro, and PEPID, but they are designed for use by physicians and are not specific to musculoskeletal shoulder pathology.

During the study, the possibility of incurring a selection bias was considered when using a convenience sample from rehabilitation and traumatology services. This recruitment was used in these centers to select a sample similar to the one proposed as the final user of the application (patients with shoulder pain for more than 6 weeks requesting care, especially physical therapy, and who could exhibit the seven pathologies evaluated by the application). It should be noted that the distribution of the pathologies with said sampling was similar to that found in a primary care setting [11].

One interesting result that should be considered for future applications is that patients usually belong to several diagnostic groups at once and may exhibit bilateral involvement. This finding requires that diagnostic algorithms, as in our application, have the capacity for multiple entries for the same patient.

We also found factors that decrease the diagnostic accuracy. The first of these is older age, which will require us to improve the applications so that older people feel more comfortable with their use and understand the statements more easily. Second, the group of patients with more intense pain in the shoulder have more problems in correctly discriminating specific clinical examinations. Third, we must improve the algorithm's ability to suspect massive rotator cuff tear and acromioclavicular arthropathy.

Our application achieved good reliability and acceptable validity for some pathologies, but it had insufficient capacity to diagnose the entire clinical picture in the same patient. We intend to continue improving the application's design and to validate the application with an impact study on the end user. It should not be forgotten that the ultimate goal is to ensure that the application becomes an accessible and self-administered tool capable of advising patients to seek specific care without increasing overdiagnosis [17].

This study should serve as the pilot test for future research in this field and to improve possible applications that may be developed. It is highlighted that this is the first study of these characteristics.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

Supplementary Materials

Supplementary Materials can be found via http://doi.org/10.4258/hir.2019.25.2.82. Table S1. Supporting information: content, bibliographic justification and validity values of the questionnaire screens. Figure S1. The set of algorithms that constitute the final application.

ORCID

Albert Bigorda-Sague (http://orcid.org/0000-0001-7282-0938)
Javier Trujillano Cabello (http://orcid.org/0000-0001-8071-7711)
Gemma Ariza Carrio (http://orcid.org/0000-0002-2020-8947)
Carmen Campoy Guerrero (http://orcid.org/0000-0002-2052-5424)

References

1. Urwin M, Symmons D, Allison T, Brammah T, Busby H, Roxby M, et al. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. Ann Rheum Dis 1998;57(11):649-55.
2. Linsell L, Dawson J, Zondervan K, Rose P, Randall

T, Fitzpatrick R, Carr A. Prevalence and incidence of adults consulting for shoulder conditions in UK primary care; patterns of diagnosis and referral. Rheumatology (Oxford) 2006;45(2):215-21.

3. Kennedy CA, Manno M, Hogg-Johnson S, Haines T, Hurley L, McKenzie D, et al. Prognosis in soft tissue disorders of the shoulder: predicting both change in disability and level of disability after treatment. Phys Ther 2006;86(7):1013-37.

4. Kooijman MK, Barten DJ, Swinkels IC, Kuijpers T, de Bakker D, Koes BW, et al. Pain intensity, neck pain and longer duration of complaints predict poorer outcome in patients with shoulder pain: a systematic review. BMC Musculoskelet Disord 2015;16:288.

5. Kooijman M, Swinkels I, van Dijk C, de Bakker D, Veenhof C. Patients with shoulder syndromes in general and physiotherapy practice: an observational study. BMC Musculoskelet Disord 2013;14:128.

6. Lee JH. Future of the smartphone for patients and healthcare providers. Healthc Inform Res 2016;22(1):1-2.

7. Research2Guidance. mHealth app developer economics 2016 [Internet]. Berlin, Germany: Research2Guidance; c2019 [cited at 2018 Apr 15]. Available from: https://research2guidance.com/product/mhealth-app-developer-economics-2016/.

8. Silva BM, Rodrigues JJ, de la Torre Diez I, Lopez-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015;56:265-72.

9. Fox S, Duggan M. Health online 2013 [Internet]. Washington (DC): Pew Research Center; 2013 [cited at 2019 Apr 15]. Available from: http://www.pewinternet. org/2013/01/15/health-online-2013/.

10. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematic review of randomized controlled trials. Healthc Inform Res 2018;24(3):207-26.

11. Ostor AJ, Richards CA, Prevost AT, Speed CA, Hazleman BL. Diagnosis and relation to general health of shoulder disorders presenting to primary care. Rheumatology (Oxford) 2005;44(6):800-5.

12. Gartner J, Blauth W, Hahne HJ. The value of the anamnesis for the probability diagnosis of shoulder pain. Z Orthop Ihre Grenzgeb 1991;129(4):322-5.

13. Farmer N, Schilstra MJ. A knowledge-based diagnostic clinical decision support system for musculoskeletal disorders of the shoulder for use in a primary care setting. Shoulder Elbow 2012;4(2):141-51.

14. Farmer N. An update and further testing of a knowledge-based diagnostic clinical decision support system for musculoskeletal disorders of the shoulder for use in a primary care setting. J Eval Clin Pract 2014;20(5):589-95.

15. Honekamp W, Hanseroth S. Evaluation of an online information system for shoulder pain patients. Int J Comput Inf Technol 2013;2(3):442-6.

16. Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential diagnosis generators: an evaluation of currently available computer programs. J Gen Intern Med 2012;27(2):213-9.

17. Lee YS. Value-based health technology assessment and health informatics. Healthc Inform Res 2017;23(3):139-40.