Imaging of glenohumeral osteoarthritis: Reliability and reproducibility of radiological classifications

Umile Giuseppe Longo[a,b], Alessandro de Sire[c,d,*], Sergio De Salvatore[a,b], Giuliano Zollo[a,b], Vincenzo Di Matteo[a,b], Alessandra Corradini[a], Alice Ceccaroli[a], Valerio Ammendolia[c], Giovanni Calabrese[a,b] and Pieter D'Hooghe[e]

[a]*Fondazione Policlinico Universitario Campus Bio-Medico, Rome, Italy*
[b]*Research Unit of Orthopaedic and Trauma Surgery, Department of Medicine and Surgery, Università Campus Bio-Medico di Roma, Rome, Italy*
[c]*Physical Medicine and Rehabilitation Unit, Department of Medical and Surgical Sciences, University of Catanzaro "Magna Graecia", Catanzaro, Italy*
[d]*Research Center on Musculoskeletal Health, MusculoSkeletalHealth@UMG, University of Catanzaro "Magna Graecia", Catanzaro, Italy*
[e]*Aspetar Hospital, Doha, Qatar*

Abstract.
BACKGROUND: Glenohumeral degenerative joint disease may affect up to 20% of the population. There are several classification systems of this disease in the scientific literature.
OBJECTIVE: The aim of this study is to determine the reliability and reproducibility of glenohumeral osteoarthritis classification systems.
METHODS: We assessed glenohumeral plain radiographs performed in a University Hospital. These radiographs were graded into nine radiological classification systems by two observers on two evaluations. Patients who have performed CT/MRI scan were staged according to the Walch classification. The intra-observer and inter-observer reliability of the classification schemes were determined by using Cohen's weighted kappa coefficient.
RESULTS: A total of 752 patients were included in the study (52.4% females and 47.6% males), mean aged 66.2 ± 16.3 years. The intra-observer and inter-observer reliabilities were 0.543 (observer 1), 0.630 (observer 2), and 0.760 (inter-observer) for the Weinstein grading system; 0.661, 0.706, and 0.761 for the Guyette grading system; 0.575, 0.679 and 0.704 for the Kellgren and Lawrence classification; 0.817, 0.816 e 0.871 for the Samilson and Prieto classification; 0.791, 0.811 and 0.847 for the Allain modification; 0.797, 0.842 and 0.860 for the Gerber modification; 0.773, 0.827 and 0.828 for the Buscayret modification; 0.584, 0.648 and 0.755 for the Hawkins and Angelo classification; 0.661, 0.749 and 0.764 for the Rosenberg classification. Intra-observer reliability for MRI was 0.757 (observer 1) and 0.675 (observer 2), while intra-observer reliability for CT was 0.811 (observer 1) and 0.653 (observer 2). Inter-observer reliabilities were 0.790 for MRI and 0.673 for CT.
CONCLUSION: The classification systems according to Weinstein, Guyette, Hawkins and Angelo, Rosenberg and the modifications of the Samilson and Prieto classification according to Allain, Gerber and Buscayret showed a comparable reliability with the commonly used glenohumeral osteoarthritis grading systems, Samilson and Prieto and Kellgren and Lawrence. Thus, they are recommended for clinical and especially scientific purposes.

Keywords: Shoulder, glenohumeral osteoarthritis, radiological classifications, Walch classification, CT/MRI scan

---

*Corresponding author: Alessandro de Sire, Physical Medicine and Rehabilitation Unit, Department of Medical and Surgical Sciences, University of Catanzaro "Magna Graecia", 88100 Catanzaro, Italy. E-mail: alessandro.desire@unicz.it.

1. Introduction

Glenohumeral degenerative joint disease (GJD) affects up to 20% of the elderly population [1]. In the initial stages, treatment consists of a conservative approach, including anti-inflammatory medication, physiotherapy and steroid injections [2]. In advanced disease stages, shoulder arthroplasty is the treatment of choice. This surgical procedure has become increasingly popular in the last decade, with the number of arthroplasties rising exponentially between 2011 and 2017 [3]. Shoulder arthroplasty achieves satisfactory results in the elderly population [4], but at the same time a joint-preserving approach may be beneficial in younger patients [5]. Hence, the classification of the stages of GJD could influence the treatment [6,7].

Numerous classification systems are present in the literature. Determining the most appropriate classification system would allow for the correct treatment of a worldwide spread pathology, such as GJD. According to the current literature, a consensus regarding the reliability and reproducibility of GJD grading systems has yet to be established. The aim of the present study was to determine the repeatability and reproducibility of nine radiological grading systems for GJD and of the Walch Classification System in CT and MRI scans of shoulder.

2. Methods

2.1. Participants

The study was approved by the Institutional Review Board of Campus Bio-Medico University of Rome (Prot. number: 113/20 (OSS) ComEt UCBM) and was performed in accordance with the Declaration of Helsinki and pertinent national and international regulatory requirements.

From 2019 to 2021, 752 consecutive patients who performed radiological shoulder examinations were included in this observational study. The radiographic images examined have followed random selection criteria, not taking into account any previously diagnosed chronic degenerative arthropathies. A computer process of randomizations has been adopted to provide the images to radiologists.

Eligible patients were of any age and gender; with or without rotator cuff injury. In patients who underwent shoulder joint replacement, prior surgery X-rays have been used. Patients were not considered eligible if they presented with fractures, rheumatoid arthritis or with missing or poorly executed medical imaging.

Eligible patients were divided into two groups:

- Group 1, patients included performed shoulder X-Ray in anterior-posterior, outlet, and West Point axillary projection and in intra-rotation and extra-rotation position.
- Group 2, a subgroup of Group 1, where patients were enrolled after performing a CT/MRI shoulder scan, in addition to plain radiographic examinations.

2.2. Intervention

Two orthopaedic surgeons skilled in shoulder surgery independently reviewed the X-ray images twice (after 3 months). The images were classified into nine schemes.

2.2.1. Classification systems

The following classification systems were used: Weinstein (Table 1) [5], Guyette (Table 2) [6], Kellgren and Lawrence (Table 3) [8], Hawkins and Angelo (Table 4) [9], Rosenberg (Table 5) [10], Samilson and Prieto (Table 6) [11], Allain (Table 7) [1], Gerber (Table 8) [12], Buscayret (Table 9) [13] and Walch classification [14] (Table 10). The latter was used by the same two orthopedic surgeons to review the CT/MRI scan twice after three months from the first assessment.

2.3. Statistical analysis

Cohen's weighted kappa was calculated from the two consecutive tests (to assess the intra-observer reliability; i.e., the repeatability) and the two individual tests (to assess the interobserver reliability; i.e., the reproducibility). The coefficient ranges from 0 (no agreement) to 1 (perfect agreement), with values $\leqslant 0.2$ indicating slight agreement, $\leqslant 0.4$ fair agreement, $\leqslant 0.6$ moderate agreement, $\leqslant 0.80$ substantial agreement and > 0.8 almost perfect agreement. All statistical analyses were performed with R software version 4.0.3. (R Core Team, 2020). The significance level was set at 0.05.

3. Results

A total of 752 patients were included in the study, comprising 52.4% females and 47.6% males, with a mean age of 66.2 ± 16.3 years. We assessed the re-

Table 1
Weinstein classification

Stage	Description
I	Normal radiograph. Diagnosis was made at the time of arthroscopy.
II	Minimal joint space narrowing with a concentric head and glenoid.
III	Moderate joint space narrowing with early inferior osteophyte formation.
IV	Severe loss of joint space with osteophyte formation and loss of concentricity between the humeral head and glenoid.

Weinstein classification (Table 1): a classification used specifically for osteoarthritis of the GH joint, based on radiographical features showing damage affecting the joint.

Table 2
Guyette classification

Stage	Description
0	No appreciable signs of arthritis.
I	Mild sclerosis and/or small osteophyte less than 2 mm on only one side of the joint.
II	Large marginal osteophytes or osteophytes on more than one side or surface of the joint, joint space narrowing, and/or the presence of cysts.
III	Joint surface destruction, bone on bone joint space narrowing and/or loose bodies.

Guyette classification (Table 2): this classification, designed specifically for glenohumeral osteoarthritis, takes into account different features that are often observed in joint degeneration.

Table 3
Kellgren and Lawrence classification

Stage	Description
0	No radiological findings of osteoarthritis.
I	Doubtful narrowing of the joint space and possible osteophytic lipping.
II	Definite osteophytes and possible narrowing of the joint space.
III	Moderate multiple osteophytes, definite joint space narrowing, small pseudocystic areas with sclerotic walls and possible bone contour deformity.
IV	Large osteophytes, marked joint space narrowing, severe sclerosis and definite deformity of bone contour.

Kellgren and Lawrence classification (Table 3): this classification, originally designed by Kellgren to evaluate OA of the hand joints, is among the most widely used also for the evaluation of OA of the glenohumeral joint among others. It takes into account features such as osteophytes, cysts and joint space narrowing.

Table 4
Hawkins and Angelo classification

Stage	Description
Mild	Joint space narrowing, sclerosis, osteophytes and/or presence of cysts.
Moderate	Joint space narrowing, sclerosis, osteophytes and/or presence of cysts.
Severe	Joint space narrowing, sclerosis, osteophytes and/or presence of cysts.

Hawkins and Angelo classification (Table 4): this classification was originally designed to evaluate GH osteoarthritis following Putti-Platt repair.

Table 5
Rosenberg classification

Stage	Description
Normal	No reduction of joint space, no osteophytes, no sclerosis.
Mild	Reduction of joint space < 1 mm, mild sclerosis.
Moderate	Reduction of joint space < 2 mm, moderate presence of osteophytes, moderate sclerosis.
Severe	Bone on bone joint space narrowing, many and large osteophytes, severe sclerosis and presence of cysts.

Rosenberg classification (Table 5): this classification was originally designed to grade degenerative changes in patients affected by GH osteoarthritis who underwent Bankart reconstruction for recurrent anterior instability.

peatability and reproducibility for the nine classification schemes.

The results for the intra-observer reliability (repeatability) showed that observer 1 achieved almost perfect reliability for the Samilson and Prieto classification ($k > 0.8$). Substantial reliability was observed for the Guyette, modified Samilson and Prieto according to Allain, Gerber, and Buscayret, and Rosenberg classifi-

Table 6
Samilson and Prieto classification

Grade	Description
1	Inferior humeral or glenoid exostosis, or both, measuring less than 3 mm in height.
2	Inferior humeral or glenoid exostosis, or both, between 3 and 7 mm in height, with slight glenohumeral joint irregularity.
3	Inferior humeral or glenoid exostosis, or both, more than 7 mm in height, with narrowing of the glenohumeral joint and sclerosis.

Samilson and Prieto classification (Table 6): designed for glenohumeral osteoarthritis, this classification evaluates severity based on the presence and size of osteophytes.

Table 7
Modified Samilson and Prieto classification according to Allain

Grade	Description
1	Inferior humeral exostosis between 1 and 3 mm in height.
2	Inferior humeral exostosis between 4 and 7 mm in height.
3	Inferior humeral exostosis more than 7 mm in height.
4	Narrowing of the glenohumeral joint and sclerosis.

Modified Samilson and Prieto classification according to Allain (Table 7): this classification is a modification of the classic Samilson-Prieto, including one more grade and showing similar reliability to its original.

Table 8
Modified Samilson and Prieto classification according to Gerber

Grade	Description
1	Inferior humeral head or glenoid osteophyteof less than 3 mm.
2	Inferior humeral head or glenoid osteophyte between 3 and 5 mm, associated with mild joint line irregularity and subchondral sclerosis.
3	Degenerative changes in the joint greater than above mentioned.

Modified Samilson and Prieto classification according to Gerber (Table 8): the Gerber modification of the Samilson-Prieto classification shows comparable reliability to its original.

Table 9
Modified Samilson and Prieto classification according to Buscayret

Grade	Description
1	Inferior humeral or glenoid exostosis, or both, measuring less than 3 mm in height.
2	Inferior humeral or glenoid exostosis, or both, between 3 and 7 mm in height, with slight glenohumeral joint irregularity.
3	Inferior humeral or glenoid exostosis, or both, more than 7 mm in height, with narrowing of the glenohumeral joint and sclerosis.
4	Bone on bone joint space narrowing with or without osteophytes.

Modified Samilson and Prieto classification according to Buscayret (Table 9): the Buscayret modification of the Samilson-Prieto classification subdivides the severe glenohumeral OA stage into two substages.

Table 10
Walch classification

Category		Description
A	A1	Humeral head centered on the glenoid fossa. No or minor central erosion.
	A2	Humeral head centered on the glenoid fossa. Major central erosion.
B	B1	Posterior subluxation of the humeral head. No posterior bone loss.
	B2	Posterior subluxation of the humeral head. Posterior bone loss resulting in a biconcave glenoid.
C		Glenoid retroversion $> 25°$

Walch classification (Table 10): this classification has been designed to analyze glenoid morphology in glenohumeral OA and includes three different groups with subtypes.

cations ($0.6 < k \leqslant 0.8$). Moderate reliability was noted for the Weinstein, Kellgren and Lawrence, and Hawkins and Angelo classifications.

For observer 2, the intra-observer reliability was almost perfect for the Samilson and Prieto classification, as well as the modified Samilson and Prieto according to Allain, Gerber, and Buscayret ($k > 0.8$). The re-

peatability results were substantial for all other classifications ($0.6 < k \leqslant 0.8$). These findings are summarized in Table 11.

The inter-observer reliability (reproducibility) was almost perfect for the Samilson and Prieto classification and the modified Samilson and Prieto according to Allain, Gerber, and Buscayret ($k > 0.8$). The inter-

Table 11
Results for reliability of radiographic classifications

Radiographic classification	Intra-observer reliability		Inter-observer reliability
	Rater 1	Rater 2	
Weinstein	0.543	0.63	0.76
Guyette	0.661	0.706	0.761
Kellgren and Lawrence	0.575	0.679	0.704
Samilson and Prieto	0.817	0.816	0.871
Allain (SPA)	0.791	0.811	0.847
Gerber (SPG)	0.797	0.842	0.86
Buscayret (SPB)	0.773	0.827	0.828
Hawkins and Angelo	0.584	0.648	0.755
Rosenberg	0.661	0.749	0.764

Intra-observer reliability (repeatability) and inter-observer reliability (reproducibility) for all included classifications except for Walch.

Table 12
Results for reliability of Walch classification

	Intraobserver reliability	Inter-observer reliability
MRI (Walch)	0.757; 0.675	0.790
TC (Walch)	0.811; 0.653	0.673

Results for intra-observer and inter-observer reliability of Walch classification.

observer reliability was substantial for the Weinstein, Guyette, Kellgren and Lawrence, and Hawkins and Angelo classifications ($0.6 < k \leqslant 0.8$). Detailed results can be found in Table 11.

3.1. Repeatability and reproducibility for CT and MRI

The repeatability results for CT scans were almost perfect for observer 1 ($k = 0.811$) and substantial for observer 2 ($k = 0.653$). The reproducibility result for CT scans was substantial ($k = 0.673$), as summarized in Table 12. For MRI scans, the repeatability results were substantial for both observers ($k = 0.757$ and $k = 0.675$), and the reproducibility result was also substantial ($k = 0.790$). Detailed results for MRI evaluations are provided in Table 12.

4. Discussion

The main finding of this study is that intra-observer reliability was almost *perfect* or substantial for all 9 classification schemes. Furthermore, the Samilson and Prieto classification and the modified Samilson and Prieto according to Allain, Gerber and Buscayret achieved almost *perfect* inter-observer reliability. Lastly, the inter-observed reliability was *substantial* for all other classifications.

Osteoarthritis of the glenohumeral joint may affect up to 20% of the elderly population [15]. Pri-

mary forms must be differentiated from secondary osteoarthritis (trauma, [4] inflammatory diseases, [16] infection, [17] instability, [9,18,19] or rotator cuff insufficiency [20]). Treatment strategies include "joint preserving" treatments (NSAIDs or physiotherapy [2], arthroscopic procedures of "debridement" [5] and "capsular release" [21]) and arthroplasty procedures [22]. The choice of a treatment regimen is influenced by the severity of osteoarthritic alterations, highlighted by traditional radiographs.

In plain radiographs, GJD is graded according to the Samilson and Prieto classification scheme, which was originally described for instability arthropathy [11]. In this scheme, GJD stages are predominately determined by the extent of glenohumeral osteophytes located in the lower portion of the glenoid cavity and humeral head. The presence of osteophytes has been described in early stages of GJD [12]. The latest radiologic grading systems described are the Weinstein [5], Guyette grading systems [6], the modified Samilson and Prieto classification according to Allain [23], the modified Samilson and Prieto according to Gerber [24], the modified Samilson and Prieto according to Buscayret [13], Rosenberg and Hawkins and Angelo grading systems [9].

The aim of the study was to determine the reliability and reproducibility of the examined radiographic classifications, which showed a strong intraobserver correlation and a good inter-observer concordance. Finding a unitary and reproducible classification system would be

beneficial for determining the most adequate treatment. Conventionally, the extent of the disease is classified by the classification of Samilson and Prieto [11]. The latter distinguishes three severity stages of the disease, based on the size of the osteophytes located in the lower glenoid cavity and humeral head.

Morphological studies on glenohumeral osteoarthritis have shown that osteophytes are located predominantly in the lower part of the glenohumeral joint [25, 26]. Although osteophytes are a determining radiographic feature of the joint's degenerative disease, the presence and size of these osteophytes is related to the individual osteoblastic response rather than to the extent of joint degeneration [27]. The shoulder joint is not subjected to loads and therefore, the functional limitation of the arthritic glenohumeral joint is more related to the size of osteophytes than to the reduction of the joint space, as observed in this and in other studies [28]. Osteophytes can appear rather late in the course of the glenohumeral arthritic disease, in this way it may be possible to underestimate the extent of the disease [12, 29]. In fact, Weinstein et al. [5] have shown that five out of 25 patients considered had no signs of arthritic glenohumeral disease in preoperative radiographs. Kappe et al. [12] analysed the early radiographic signs of shoulder arthritis in one study and found glenoid sclerosis as the main feature in radiographs of 35 out of 69 patients. Making a correct classification of glenohumeral arthritis, particularly in the early stages of the disease, allows for an appropriate therapeutic choice [2]. Furthermore, it is important to have a good classification system that allows the surgeon to differentiate between early stages of the disease, to which conservative or arthroscopic treatment may be reserved, and late stages to be directed to surgical treatment.

In their experience, Elsharkawi et al. [30] analysed 108 glenohumeral joints affected by osteoarthritis, concluding that the classification systems Samilson and Prieto, Samilson and Prieto modified according to Allain, Samilson and Prieto modified according to Gerber, Weinstein, Guyette and Kellgren and Lawrence are recommended for clinical practice and radiological evaluation. Furthermore, Brox et al. [31] have demonstrated that the Samilson and Prieto classification is generally preferable due to its simpler usage and excellent interobserver agreement. Lastly, as shown by Vo et al. [32] and Nowak et al. [33], the Walch classification system has moderate interobserver reliability and substantial intraobserver reliability, making it an excellent tool in describing glenoid morphology, similar or superior to other commonly used systems. Nonetheless,

Weinstein et al. [5] found no correlation between their grading system and patient outcomes after arthroscopic debridement. Guyette et al. [6] did not find a correlation between the stage of preoperative radiographic evaluations and the outcome after sub acromial decompression. Allain et al. [23] demonstrated a correlation between the degree of postoperative osteoarthritis and functional limitation.

The most important result that emerges from this study is that the Samilson and Prieto classification is the most reliable and reproducible grading and radiographic classification system in radiological diagnostics. The results of the Samilson and Prieto classification system modified according to Gerber [24] observed in this study are slightly better than the Samilson and Prieto classification system. The Samilson and Prieto classification system modified according to Gerber [24], however, is a simplification of the Samilson and Prieto [11] system: both of them, in fact, assume the osteophyte's dimension as a reference element for the staging of the arthritic pathology. The classification of the degenerative pathology is composed of three degrees in both cases, with the only difference that, according to Samilson and Prieto classification, to define grade 3 the osteophyte's thickness must be at least 7 mm, while for the modified Samilson and Prieto classification according to Gerber, a dimension of 5 mm is sufficient. The classification of Kellgren and Lawrence is the most useful to diagnose osteoarthritis in joints under load, such as hip and knee [8]; nevertheless, this classification does not have the same reliability regarding the shoulder joint, especially considering inter-observer variability. The Samilson and Prieto classification and the modified system of Samilson and Prieto according to Gerber discriminate glenohumeral arthritis based on the size of the osteophyte, but, as already mentioned, the latter varies according to the individual osteoblastic response [27]. Some radiographic morphological studies of glenohumeral arthritis show that the osteophyte is formed in the lower portion of the glenoid cavity and humeral head, in a central position [25]. So, when staging a heterogeneous and complex degenerative pathology such as arthritis, relying only on one element (such as the dimension of the osteophyte) would seem to be an excessive simplification. In this way, other equally important elements for diagnosis, such as joint symmetry, joint rupture, geodes, bone sclerosis, loss of joint cartilage, intracapsular pressure, intraosseous pressure and joint contracture would be ignored [34,35]. The Kellgren and Lawrence [8] X-ray classification system fully describes all the different elements of joint

Fig. 1. Female patient, left shoulder X-ray performed at age 85. Wenstein classification, stage 4. Guyette classification, stage 3. Kellgren and Lawrence classification, grade 4. Samilson and Prieto classification, grade 3. SPA classification, grade 4. SPG classification, grade 3. SPB classification, stage 4. Hawkins and Angel classification, severe degree. Rosenberg classification, severe degree.

degenerative pathology and, for this reason, it is universally accepted for diagnosing arthritis. The Rosenberg classification system [10], as well as the Hawkins and Angelo [9], the Weinstein [5] and the Guyette [6] classification systems, also describe other elements, in addition to osteophytes, for radiographical staging of the arthritic pathology. These include the reduction of the articular line (joint space narrowing), bone sclerosis and cystic formations (geodes). From the clinical point of view, however, there is a poor correlation between the severity of symptoms and the radiographic signs identified with these classification systems, particularly in GJD [36]. The knowledge of these classification systems is primary in orthopaedic clinical practice, since the subdivision of patients suffering from glenohumeral degenerative pathology in subgroups, divided by severity, allows for the treatment of patients at different stages, with tailored specific treatments.

This study demonstrates that the Samilson and Prieto radiological classification system for glenohumeral arthritis has intrinsic limitations. The validity of this system, in fact, is dependent on the correct execution of the radiographic examination in antero-posterior position of the shoulder.

The apparent improvement in the reliability and reproducibility of this classification system, based on shoulder position, rotation and on the type of radiograph performed, is explained by the fact that the rotation of

the shoulder could hide the osteophytes of mild and medium size. These would be located at the level of the inferior portion of the glenoid cavity and of the humeral head [34].

On the other hand, osteophytes of severe size are easier to measure during the evaluation of radiographic images; therefore, their grading and staging are not influenced by limb rotation [34] (Fig. 1).

Another aim of this study was to evaluate the reliability of the Walch classification. Given the distribution of CT and MRI, the Pearson Correlation Coefficient (expression of the intra-observer variability) could not be performed by statistical analysis. It was possible to perform, on the same distribution of CT and MRI, the K Cohen coefficient (expression of the inter-observer variability). The statistical results of MRI scans according to the Walch classification indicate a good inter-observer concordance. The statistical results of CT scans according to the Walch classification indicate a discrete inter-observer concordance. The classification proposed by Walch, in fact, is widely used in the pre-operative staging of glenohumeral arthritis [37] (Fig. 2). A further objective of the study was to assess the correlation between radiographic classifications and the Walch classification.

The P value was used to evaluate the agreement rate between radiographic classifications and the Walch classification of CT and MRI scans. For all the other
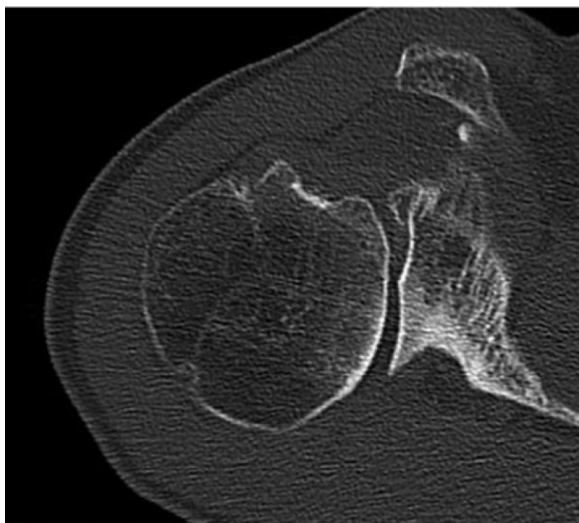
Fig. 2. Female patient, CT scan of right shoulder performed at age 85. The image shows a posterior subluxation of the humeral head, sign of a damaged rotator cuff and consequent eccentric glenohumeral osteoarthritis. The posterior erosion of the glen is responsible for its biconcave shape. Walch classification, category B2.

classifications, a P value lower than the threshold value α (= 0.05) was not displayed; therefore, empirical evidence is not statistically significant. For what concerns the Samilson and Prieto classification and the modified Samilson and Prieto according to Gerber, the study has shown that these systems are not related with the Walch classification of CT/MRI scans.

Kopka et al. [38] have shown that the use of radiography alone, compared with CT and/or MRI scans, has a lower capacity to identify the deterioration pattern of cartilage. This is of extreme relevance, because many surgeons continue to use radiography alone, which, despite having a low cost, is subject to several limitations. These ultimately affect the assessment's reliability. Examples are the angle of the X-ray beam, the experience of the operator and, in case of severe arthritis, the ability of the patient to abduct the arm to the axillary position of West Point (a specific radiographic projection used to assess the anteroinferior glenoid rim). On the other hand, the axillary MRI and CT scans are not subject to these variables and are generally recognized as Gold Standards to define shoulder bone morphology [39,40]. The correlation, in terms of intra-observer and inter-observer variability, between radiographic classifications and the Walch classification, could be improved by optimizing the quality of axillary radiographs and CT/MRI axillary scans [37]. An adequate assessment of shoulder pain and particularly of glenohumeral osteoarthritis could play a key role in planning an adequate surgical or rehabilitative treatment [41,42].

This study's findings carry a substantial implication for clinical practice and future research, as they can be directly integrated into clinical workflows. For instance, the high reliability and reproducibility of the Samilson and Prieto classification system, as well as its modified versions, suggest that these can be effectively used for standardized patient assessments, guiding treatment decisions, and facilitating communication among healthcare professionals. This can lead to more consistent and precise staging of glenohumeral osteoarthritis, which is crucial for tailoring treatment strategies, from conservative management to surgical interventions. Additionally, incorporating these reliable classification systems into electronic health records can enhance the longitudinal tracking of disease progression and treatment outcomes, ultimately improving patient care. Future research could further refine these classification systems and explore their application in diverse clinical settings, ensuring their broad applicability and utility in routine orthopedic practice.

4.1. Strengths and limitations of the study

The strengths of the study are as follows. The high number of patients analyzed for the conduction of a large-scale statistical analysis, in order to increase the generalization of the results, allows for a practical clinical application of this study's results. The database utilized consented to archive X-ray images of the same patient at different points in time, so it was possible to appreciate the evolution of the chronic-degenerative process of arthritis over time. Moreover, the study has provided results that are consistent with those currently available in the literature. These strengths, along with the potential for enhanced diagnostic accuracy, improved treatment planning, and early intervention, underscore the study's significant contributions to clinical practice in managing glenohumeral osteoarthritis. By implementing the study's findings, clinicians can provide more accurate diagnoses, tailor treatments to individual patient needs, and ultimately improve patient outcomes.

This study presents some limitations. The first concerns the fact that patients are selected according to a random criterion and not according to an already established diagnosis of chronic degenerative arthropathy. In this way, the inclusion of younger individuals who may not exhibit pronounced radiographic features of osteoarthritis may influence the study results. This selection bias could lead to an underestimation or overestimation of the reliability and validity of the classifica-

tion systems. However, including younger individuals, who may not exhibit pronounced radiographic features typical of advanced degenerative joint disease, was part of our strategy to understand how these classification systems perform across different age groups. This inclusion enhances the generalizability of the classification systems' applicability and provides meaningful insights into their sensitivity and specificity in a real-world setting.

The second limitation regards a low number of observers who can overestimate the value of K Cohen (inter-observer variability). A larger and more diverse group of observers might capture a broader range of variability, providing a more accurate measure of reliability. In fact, diagnosing arthrosis through radiography is simpler than staging it in universally accepted categories.

The third limitation is the quality of radiographic examination. The reliability of the Samilson and Prieto classification system is highly dependent on the correct execution of the radiographic examination, particularly the antero-posterior positioning of the shoulder. Variability in shoulder positioning, rotation, and the type of radiograph performed can obscure the visibility of osteophytes, particularly those of mild and moderate size. This can lead to misclassification and impact the staging accuracy of glenohumeral osteoarthritis. Clinically, this suggests that standardized radiographic protocols are essential to ensure accurate and consistent OA classification.

The fourth limitation of this study is that assessing only bone spur formation seems to oversimplify the complexity and heterogeneity of osteoarthritis, a chronic degenerative joint pathology. Osteoarthritis involves multiple pathological changes, including cartilage degradation, synovial inflammation, and subchondral bone alterations, which are not fully captured by evaluating bone spurs alone.

Lastly, technological limitations such as computer resolution and image quality can affect the accuracy of classification. Additionally, observer training and the "learning curve" associated with interpreting these images introduce variability and potential inaccuracies in classification.

In the end, the distribution of CT and MRI scans did not allow for the correlation of radiographic classifications to the Pearson Correlation Coefficient. In fact, the correlation was estimated only through the K Cohen coefficient. The main limitations of this study concern the fact that only two observers were included, while the trial was not randomized.

5. Conclusion

All the considered radiographic classifications could be used as an alternative or an aid to the widely diffused Samilson and Prieto classification, in the radiographic diagnosis of glenohumeral osteoarthritis. According to this study, the Samilson and Prieto classification displayed the highest values in terms of reliability and reproducibility, although its real clinical relevance is still under investigation. In fact, the evaluation of osteophytosis alone appears to be too simplifying to frame the complexity and heterogeneity of a chronic-degenerative arthropathy such as osteoarthritis. Eventually, the evidence does not achieve the same significance for the Walch classification, both in terms of intra and inter-observer concordance and in relation to radiological classifications. Therefore, an improved correlation between the Walch classification and radiographic classification systems could be a target for future studies on the imaging of glenohumeral osteoarthritis. The present study is useful for laying the groundwork for establishing the gold-standard of GJD classification systems.

Ethical approval

The Institutional Review Board of Campus Bio-Medico University of Rome ruled that no formal ethics

approval was required in this particular case and the need to obtain informed consent was waived based on the retrospective design and anonymization of patient identifiers (Prot. number: 113/20 (OSS) ComEt UCBM). All methods were performed in accordance with the relevant guidelines and regulations.

Conflict of interest

The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

References

[1]   Clinton J, Franta AK, Lenters TR, Mounce D, Matsen FA, 3rd. Nonprosthetic glenoid arthroplasty with humeral hemi-arthroplasty and total shoulder arthroplasty yield similar self-assessed outcomes in the management of comparable patients with glenohumeral arthritis. J Shoulder Elbow Surg. 2007; 16(5): 534-8.

[2]   Moskowitz RW, Blaine TA. An overview of treatment options for persistent shoulder pain. Am J Orthop (Belle Mead NJ). 2005; 34(12 Suppl): 10-5.

[3]   Wagner ER, Farley KX, Higgins I, Wilson JM, Daly CA, Gottschalk MB. The incidence of shoulder arthroplasty: rise and future projections compared with hip and knee arthroplasty. J Shoulder Elbow Surg. 2020; 29(12): 2601-9.

[4]   Boileau P, Sinnerton RJ, Chuinard C, Walch G. Arthroplasty of the shoulder. J Bone Joint Surg Br. 2006; 88(5): 562-75.

[5]   Weinstein DM, Bucchieri JS, Pollock RG, Flatow EL, Bigliani LU. Arthroscopic debridement of the shoulder for osteoarthritis. Arthroscopy. 2000; 16(5): 471-6.

[6]   Guyette TM, Bae H, Warren RF, Craig E, Wickiewicz TL. Results of arthroscopic subacromial decompression in patients with subacromial impingement and glenohumeral degenerative joint disease. J Shoulder Elbow Surg. 2002; 11(4): 299-304.

[7]   Maffulli N, Longo UG, Berton A, Loppini M, Denaro V. Biological factors in the pathogenesis of rotator cuff tears. Sports Med Arthrosc Rev. 2011; 19(3): 194-201.

[8]   refauthorKellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957; 16(4): 494-502.

[9]   Hawkins RJ, Angelo RL. Glenohumeral osteoarthrosis. A late complication of the Putti-Platt repair. J Bone Joint Surg Am. 1990; 72(8): 1193-7.

[10]   Rosenberg BN, Richmond JC, Levine WN. Long-term followup of Bankart reconstruction. Incidence of late degenerative glenohumeral arthrosis. Am J Sports Med. 1995; 23(5): 538-44.

[11]   Samilson RL, Prieto V. Dislocation arthropathy of the shoulder. J Bone Joint Surg Am. 1983; 65(4): 456-60.

[12]   Kappe T, Elsharkawi M, Floren M, Reichel H, Cakir B. Plain radiographs have limited sensitivity for glenohumeral cartilage lesions. Orthopedics. 2010; 33(5).

[13]   Buscayret F, Edwards TB, Szabo I, Adeleine P, Coudane H, Walch G. Glenohumeral arthrosis in anterior instability before and after surgical treatment: incidence and contributing factors. Am J Sports Med. 2004; 32(5): 1165-72.

[14]   Walch G, Boulahia A, Boileau P, Kempf JF. Primary glenohumeral osteoarthritis: clinical and radiographic classification. The Aequalis Group. Acta Orthop Belg. 1998; 64(Suppl 2): 46-52.

[15]   Parsons IM, Weldon EJ, Titelman RM, Smith KL. Glenohumeral arthritis and its management. Phys Med Rehabil Clin N Am. 2004; 15(2): 447-74.

[16]   Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. Acta Radiol Diagn (Stockh). 1977; 18(4): 481-91.

[17]   Mileti J, Sperling JW, Cofield RH. Shoulder arthroplasty for the treatment of postinfectious glenohumeral arthritis. J Bone Joint Surg Am. 2003; 85(4): 609-14.

[18]   Locher J, Wilken F, Beitzel K, Buchmann S, Longo UG, Denaro V, et al. Hill-sachs off-track lesions as risk factor for recurrence of instability after arthroscopic bankart repair. Arthroscopy. 2016; 32(10): 1993-9.

[19]   Longo UG, Rizzello G, Loppini M, Locher J, Buchmann S, Maffulli N, et al. Multidirectional instability of the shoulder: a systematic review. Arthroscopy. 2015; 31(12): 2431-43.

[20]   Feeley BT, Gallo RA, Craig EV. Cuff tear arthropathy: current trends in diagnosis and surgical management. J Shoulder Elbow Surg. 2009; 18(3): 484-94.

[21]   Richards DP, Burkhart SS. Arthroscopic debridement and capsular release for glenohumeral osteoarthritis. Arthroscopy. 2007; 23(9): 1019-22.

[22]   Ambacher T. [Shoulder replacement in arthritis of the shoulder joint]. Orthopade. 2007; 36(11): 1017-26.

[23]   Allain J, Goutallier D, Glorion C. Long-term results of the Latarjet procedure for the treatment of anterior instability of the shoulder. J Bone Joint Surg Am. 1998; 80(6): 841-52.

[24]   Gerber C, Terrier F, Ganz R. The Trillat procedure for recurrent anterior instability of the shoulder. J Bone Joint Surg Br. 1988; 70(1): 130-4.

[25]   Edelson JG. Patterns of degenerative change in the glenohumeral joint. J Bone Joint Surg Br. 1995; 77(2): 288-92.

[26]   Longo UG, Loppini M, Rizzello G, Romeo G, Huijsmans PE, Denaro V. Glenoid and humeral head bone loss in traumatic anterior glenohumeral instability: a systematic review. Knee Surg Sports Traumatol Arthrosc. 2014; 22(2): 392-414.

[27]   Saito M, Saito S, Ohzono K, Ono K. The osteoblastic response to osteoarthritis of the hip. Its influence on the long-term results of arthroplasty. J Bone Joint Surg Br. 1987; 69(5): 746-51.

[28]   Kircher J, Morhard M, Magosch P, Ebinger N, Lichtenberg S, Habermeyer P. How much are radiological parameters related to clinical symptoms and function in osteoarthritis of the shoulder? Int Orthop. 2010; 34(5): 677-81.

[29]   Maffulli N, Longo UG. Conservative management for tendinopathy: is there enough scientific evidence? Rheumatology (Oxford). 2008; 47(4): 390-1.

[30]   Elsharkawi M, Cakir B, Reichel H, Kappe T. Reliability of radiologic glenohumeral osteoarthritis classifications. J Shoulder Elbow Surg. 2013; 22(8): 1063-7.

[31]   Brox JI, Lereim P, Merckoll E, Finnanger AM. Radiographic classification of glenohumeral arthrosis. Acta Orthop Scand. 2003; 74(2): 186-9.

[32]   Vo KV, Hackett DJ, Gee AO, Hsu JE. Classifications in Brief: Walch Classification of Primary Glenohumeral Osteoarthritis. Clin Orthop Relat Res. 2017; 475(9): 2335-40.

[33]   Nowak DD, Gardner TR, Bigliani LU, Levine WN, Ahmad CS. Interobserver and intraobserver reliability of the Walch classification in primary glenohumeral arthritis. J Shoulder Elbow Surg. 2010; 19(2): 180-3.

[34]   Ilg A, Bankes MJ, Emery RJ. The intra- and inter-observer

reliability of the Samilson and Prieto grading system of glenohumeral arthropathy. Knee Surg Sports Traumatol Arthrosc. 2001; 9(3): 187-90.

[35] Franceschi F, Longo UG, Ruzzini L, Rizzello G, Denaro V. Arthroscopic management of calcific tendinitis of the subscapularis tendon. Knee Surg Sports Traumatol Arthrosc. 2007; 15(12): 1482-5.

[36] Familiari F, Ammendolia A, Rupp MC, Russo R, Pujia A, Montalcini T, Marotta N, Mercurio M, Galasso O, Millett PJ, Gasparini G, de Sire A. Efficacy of intra-articular injections of hyaluronic acid in patients with glenohumeral joint osteoarthritis: A systematic review and meta-analysis. J Orthop Res. 2023; 41: 2345-2358. doi: 10.1002/jor.25648.

[37] Aronowitz JG, Harmsen WS, Schleck CD, Sperling JW, Cofield RH, Sanchez-Sotelo J. Radiographs and computed tomography scans show similar observer agreement when classifying glenoid morphology in glenohumeral arthritis. J Shoulder Elbow Surg. 2017; 26(9): 1533-8.

[38] Kopka M, Fourman M, Soni A, Cordle AC, Lin A. Can glenoid wear be accurately assessed using x-ray imaging? Evaluating agreement of x-ray and magnetic resonance imaging (MRI)

Walch classification. J Shoulder Elbow Surg. 2017; 26(9): 1527-32.

[39] Saliken DJ, Bornes TD, Bouliane MJ, Sheps DM, Beaupre LA. Imaging methods for quantifying glenoid and Hill-Sachs bone loss in traumatic instability of the shoulder: a scoping review. BMC Musculoskelet Disord. 2015; 16: 164.

[40] Gyftopoulos S, Hasan S, Bencardino J, Mayo J, Nayyar S, Babb J, et al. Diagnostic accuracy of MRI in the measurement of glenoid bone loss. AJR Am J Roentgenol. 2012; 199(4): 873-8.

[41] de Sire A, Moggio L, Marotta N, Fortunato F, Spalek R, Inzitari MT, Paolucci T, Ammendolia A. Ultrasound-guided injections and proprioceptive neuromuscular facilitation as shoulder rehabilitation for multiple sclerosis and neuropathic pain. Healthcare. 2022; 10: 1869. doi: 10.3390/healthcare10101869.

[42] de Sire A, Agostini F, Bernetti A, Mangone M, Ruggiero M, Dinatale S, Chiappetta A, Paoloni M, Ammendolia A, Paolucci T. Non-surgical and rehabilitative interventions in patients with frozen shoulder: umbrella review of systematic reviews. J Pain Res. 2022; 15: 2449-2464. doi: 10.2147/JPR.S371513.