Reliability and Validity of Acetabular and Femoral Bone Loss Classification Systems in Total Hip Arthroplasty: A Systematic Review

Alex Gu, BS · Marco Adriani, MD · Michael-Alexander Malahias, MD, PhD · Safa C. Fassihi, MD · Allina A. Nocon, MPH · Mathias P. Bostrom, MD · Peter K. Sculco, MD

Received: 9 December 2019/Accepted: 25 May 2020 /Published online: 18 June 2020
© Hospital for Special Surgery 2020

Abstract *Background:* A variety of classification systems have been developed to help surgeons treat patients with acetabular or femoral bone loss in total hip arthroplasty, yet no “gold standard” for classification has been agreed upon. Furthermore, the reliability and validity of the available classification systems remain unknown. *Questions/Purpose:* The aims of our study were to determine the reliability and validity of the three most common acetabular and femoral bone loss classification systems (Paprosky, American Academy of Orthopaedic Surgeons [AAOS], and Saleh and Gross). *Methods:* A systematic review of the literature was performed to identify studies that reported on the reliability or validity (or both) of the acetabular and femoral components of the three bone loss classification systems. *Results:* In all, seven articles met our inclusion criteria. Six studies reported on the reliability (all six studies) or validity (three studies) of acetabular bone loss rating systems (286 acetabula), and five analyzed reliability (all five studies) or validity (three studies) of femoral bone loss classification systems (364 femurs). In studies in which either the Paprosky or AAOS acetabular bone loss classifications were used, the classification systems were considered unreliable in 75%

and 100% of them, respectively. On the femoral side, the Paprosky classification demonstrated moderate interobserver and good intraobserver reliability. The AAOS femoral bone classification was found to have good intraobserver reliability but poor interobserver reliability. The Saleh and Gross acetabular and femoral bone loss classification systems yielded mixed results, but each was considered reliable in one of the studies looking at these aspects of the systems. *Conclusion:* Although surgical techniques, treatment options, and advanced imaging available to the surgeon have evolved over the past few decades, the acetabular and femoral bone loss classification systems, first developed in the 1990s, have remained largely unchanged. Our results indicate that improvements to these systems are necessary in order for them to be as useful as possible in planning the surgical course.

Keywords classification · reliability · validity · systematic review · revision hip arthroplasty · femoral bone loss · acetabular bone loss

Introduction

Total hip arthroplasty (THA) is one of the most commonly performed surgical procedures, and projections show a drastic increase in the number of THAs performed in the next few decades [15]. Although the development of highly cross-linked polyethylene has decreased the incidence of osteolysis and aseptic loosening failures, the rates of periprosthetic infection and peri-prosthetic fracture have remained unchanged over the past decade [28]. As the number of THAs performed increases, the absolute number of patients presenting with these complications will undoubtedly rise, along with associated acetabular and femoral bone loss [15]. In order to appropriately manage acetabular and femoral bone loss at the time of surgery, surgeons

Level of Evidence: Level III: Systematic Review of Level II and III Studies

A. Gu, BS · M. Adriani, MD · M.-. Malahias, MD, PhD ·
A. A. Nocon, MPH · M. P. Bostrom, MD · P. K. Sculco, MD (✉)
Stavros Niarchos Foundation Complex Joint Reconstruction Center,
Department of Orthopedic Surgery, Hospital for Special Surgery,
535 East 70th St.,
New York, NY 10021, USA
e-mail: Sculcop@hss.edu

A. Gu, BS · M. Adriani, MD · S. C. Fassihi, MD
Department of Orthopedic Surgery, George Washington School of
Medicine and Health Sciences,
Washington, DC, USA

increasingly rely on classification systems to help guide their surgical intervention and conduct strong evidence-based comparative studies.

A variety of surgical reconstruction options are available to treat acetabular and femoral bone loss at the time of revision THA. In general, the choice of treatment is guided by the severity of bone loss [25, 26]. For severe acetabular bone deficits, earlier, and now less frequently used, options include cemented hemispherical acetabular components, cages, and bulk structural allografting [26]. More recently, the trend has been toward bony ingrowth and durable biologic fixation through uncemented hemispherical acetabular components, modular highly porous augments, cup–cage constructs, and custom triflange implants [26]. Earlier and less frequently used options for significant femoral bone loss include a cement-in-cement technique, femoral impaction grafting, allograft–prosthetic composite, monoblock proximally coated femoral stems, and extensively porous-coated cylindrical stems. Currently, the more commonly used options include proximally modular stems; modular or monolithic tapered, fluted stems; and proximal femoral replacement when indicated in severe cases [3, 25].

There is currently no “gold standard” for the treatment of severe bone loss in the hip arthroplasty setting, and the complexity of the condition necessitates careful pre-operative planning. The classification systems used by the majority of surgeons are based on the assessment of plain radiographs. The ideal classification system would, naturally, have the ability to both indicate prognosis and guide treatment, yet classification schemes that are only descriptive in nature or that classify patients on the basis of incomplete or limited information are unlikely to be of optimal clinical utility. With such a heavy reliance on classification systems to determine surgical treatment, it is also crucial that such systems be reproducible (have high reliability) and accurate (have high validity). In addition, an effective classification system is not only integral to achieving desired patient outcomes; it can also be used for comparative research on the effectiveness of different reconstruction options. However, a systematic review of the validity and reliability of the currently available classification systems has not been performed.

The aim of our study was to determine the validity and reliability of the three most commonly used of these classification systems: the American Academy of Orthopaedic Surgeons (AAOS) [6], Paprosky [18], and Saleh and Gross [22] systems.

Methods

We analyzed the three systems according to their acetabular and femoral components; not all studies involved all three systems or both components of each system. We used the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement to conduct the systematic review (www.prisma-statement.org). We searched the US National Library of Medicine (PubMed/MEDLINE), Embase, and the Cochrane Database of Systematic Reviews

for studies published from January 1980 to August 2018 using keywords pertinent to revision THA, acetabular or femoral classification systems, and reliability and validity. Only abstracts that evaluated the validity or reliability (or both) of a recognized acetabular or femoral bone loss classification system were reviewed. The specific search terms we used are detailed further in Table 1.

Studies using human subjects of any age and sex, using a population comprising at least 15 patients who underwent primary or revision THA in the presence of acetabular or femoral bone loss, and measured reliability or validity of any recognized classification system for acetabular or femoral bone loss with respect to revision THA studies were included. Case reports, review articles, letters to the editor, technical papers, abstracts, book chapters, in vitro studies, and non-English language publications were excluded. The reference lists of articles meeting these criteria were screened for additional studies not captured in the initial search.

Two authors (A.G., M.A.) independently conducted the search. Each compiled a list of papers not excluded after application of the inclusion and exclusion criteria. Discrepancies between the two lists were resolved by discussion. During initial review of the data, the following information was collected from each study: title, first author, year of publication, study design, number of patients, system used to classify acetabular or femoral bone loss, number of hips, and intraobserver reliability (defined as the level of agreement in multiple assessments by the same rater), interobserver reliability (defined as the agreement among multiple observers), or both when available, and validity. In general, reliability was based upon κ value and validity was based upon pre-operative and then surgical observation. The level of agreement among raters was determined, in most instances, using the criteria of Landis and Koch [16], in which a kappa (κ) value of less than 0 indicates no agreement, 0 to 0.20 indicates poor agreement, 0.21 to 0.40 is fair, 0.41 to 0.60 is moderate, 0.61 to 0.80 is good, and 0.81 to 1 is excellent. In most of the reviewed studies, a classification system was considered reliable if a κ score of 0.61 or higher (good or excellent) was reached. Reliability and validity in each study were determined on the basis of the criteria used in that study.

The level of evidence in each study was determined using the Oxford Centre for Evidence-Based Medicine Levels of Evidence [13]. Further assessment of the quality of the studies was conducted using the Coleman criteria, adapted in a manner similar to that used by Sambandam et al. [23]. Selective reporting bias was not included in this analysis.

Results

The database search resulted in 312 abstracts that discussed the reliability or validity (or both) of acetabular and femoral bone loss classification systems (Fig. 1). After the elimination of duplicate articles, we applied the predetermined inclusion and exclusion criteria. In total, seven articles met the inclusion criteria and were included in our analysis

Table 1 Search strategy in the three databases (PubMed/MEDLINE, the Cochrane Database of Systematic Reviews, and Embase)

Date of search	October 2018
Strategy	No. 1 AND no. 2 AND no. 3
Limits	Human AND English
No. 1	(“arthroplasty, replacement, hip”[mh] OR (“total hip arthroplasty”)[tw] OR (tha)[tw] OR (“total hip replacement”)[tw] OR (thr)[tw]) OR “hip arthroplasty”[All Fields]) AND (“acetabular”[All Fields] OR “femoral”[All Fields])
No. 2	(“acetabular”[All Fields] OR “femoral”[All Fields])
No. 3	(“validity”[All Fields] OR “reliability”[All Fields])

[mh] MeSH headings, [tw] text words, (tha) total hip arthroplasty, (thr) total hip replacement

(Table 2) [2, 4, 7, 10, 19, 21, 28]. Four studies (57%) [2, 4, 19, 29] were retrospective cohort studies, and three (43%) were prospective cohort studies [7, 10, 21].

All of the studies included patients who had a failed total hip prosthesis in need of revision surgery. The bone loss classification was determined using pre-operative plain radiographs. Reviewers were blinded in all seven studies. Three studies assessed pre-operative anteroposterior (AP) radiographs [2, 7, 29], three studies assessed pre-operative AP and lateral radiographs [10, 19, 21], and one study assessed pre-operative AP and Judet oblique radiographs [4]. Pre-operative radiographs were assessed once in two studies [7, 21], four studies involved two evaluations [2, 4, 10, 19], and one study analyzed pre-operative radiographs on three separate occasions [29]. Six of the studies used the criteria of Landis and Koch to interpret the κ scores and inter- and intraobserver reliability, and one analysis [4] used the criteria of Svanholm et al. With respect to level of evidence, three studies were level II [7, 10, 21] and four studies were level III [2, 4, 19, 29]. The total mean modified Coleman score in our review was 42 out of 100, ranging from 34 [4, 19] to 51 [7].

In total, 499 patients were included in this review; 364 femur and 286 acetabular radiographs were analyzed. Table 3 shows inter- and intraobserver reliability and validity scores according to classification system used. Table 4 shows results of the studies individually.

Only one of the studies [7] reported the age and sex of the cohort. In this study, the mean age was 68.5, with 35 male and 46 female patients.

Acetabular Bone Loss Classification

Four studies [4, 10, 19, 29] reported on the reliability of the Paprosky acetabular classification system [18]. The mean patient sample size was 44 (range, 25 to 83). Three of these studies [4, 10, 19] used the unweighted κ score to measure the level of agreement, whereas Yu et al. [29] used the weighted κ coefficient [21]. The mean intraobserver reliability was moderate ($\kappa = 0.45$). Campbell et al. [4] reported that the intraobserver reliability of the classification's creator [18] was good ($\kappa = 0.75$). However, the mean interobserver reliability κ value was moderate ($\kappa = 0.46$).

Two studies evaluated the validity of the Paprosky acetabular bone loss classification system [10, 29]. Yu et al. [29] observed a significant correlation between defects in the anterior, medial, and superior acetabular walls and their respective radiographic landmarks. No relationship was seen between posterior acetabular defect size and ischial defect grade noted on radiographs. However, only AP radiographs were used for classification. Gozzard et al. compared the pre-operative bone loss grading with the bone loss classified intra-operatively and reported good validity ($\kappa = 0.65$) [10]. In that study, both AP and lateral radiographs were used in the pre-operative bone loss classification [10]. Finally, Yu et al. evaluated whether reliability was dependent on teachability. They found that teaching sessions combined with the use of a visual scoring guide substantially improved reliability, from moderate (0.56) to good (0.79) agreement [29].

Three studies reported data on the reliability of AAOS acetabular classification system [4, 10, 19]. The mean number of acetabular radiographs evaluated was 30 (range 25 to 33). All of them [4, 10, 19] used an unweighted κ score to analyze agreement between observers. Overall, the mean intraobserver reliability was moderate (0.51), and the mean interobserver reliability was fair (0.24). Gozzard and Parry used AP and lateral radiographs for pre-operative classification, whereas Campbell used AP and Judet views [4, 10, 19]. Campbell et al. also asked the creators of the classifications to give their evaluation, and D'Antonio, co-creator of the AAOS system, had an intraobserver reliability of moderate (0.57) [4]. Overall, none of the studies found this classification system to have good intra- or interobserver reliability. Among the studies included in our analysis, none assessed the validity of the AAOS acetabular bone loss classification system.

Two studies [7, 21] reported on the reliability and validity of the acetabular Saleh and Gross classification system [22]. In both, a weighted κ score was used to evaluate the level of agreement between the observers; Saleh et al. (which was written by some of the creators of the classification) also used unweighted scores. A total of 126 cases were evaluated. Saleh et al. included 45 acetabular bone loss cases and reported good mean interobserver reliability ($\kappa > 0.75$) and validity ($\kappa > 0.75$) [22]. The second study [7] had a cohort of 81 hips and reported an interobserver reliability of moderate (0.54) and an intra-operative validity of fair-moderate (0.41).

Femoral Bone Loss Classification

Three articles [2, 10, 19] reported on the reliability of the Paprosky femoral bone loss classification system [8]. Pre-operative AP and lateral femoral radiographs size were evaluated in all studies. The mean total number of cases graded was 84 (range, 23 to 205). In all studies, the unweighted κ score was used to measure the level of agreement for both intraobserver and interobserver reliability. Overall, the intraobserver reliability was good (mean 0.63; range, 0.34 to 0.78), whereas the interobserver reliability was moderate (mean 0.57; range, 0.39 to 0.72). Only one of the

Fig. 1. Systematic review flow diagram

studies [10] reported on the validity of the femoral Paprosky system, comparing pre-operative classification to intra-operative findings in 25 cases and found moderate agreement ($\kappa = 0.54$).

Two articles [10, 19] analyzed reliability of AAOS femoral classification system [5]. The total number of femurs evaluated was 48. Both of the studies used unweighted κ score to measure the level of agreement between the observers. The mean intraobserver reliability was good ($\kappa = 0.64$; range, 0.46 to 0.81), and the mean interobserver reliability was fair ($\kappa = 0.38$; range, 0.09 to 0.66). We were unable to identify any studies evaluating the validity of AAOS femoral bone loss classification.

Two articles [7, 21] reported data on reliability and validity of the Saleh and Gross femoral bone loss classification system [11]. They used a weighted κ score to assess the degree of agreement between surgeons; Saleh et al. (the authors of which included creators of the classification system) also used unweighted κ scores. They had a cohort of 45 patients, with an interobserver reliability of good ($\kappa \geq 0.75$) and a mean validity of good ($\kappa > 0.75$) [22]. However, in the evaluation of 81 patients by Davis et al., interobserver

reliability was moderate ($\kappa = 0.56$) and validity was fair ($\kappa = 0.39$) [7].

Discussion

Multiple classification systems have been developed to categorize acetabular and femoral bone loss and ideally are designed to help guide surgical management. An optimal classification system should be reproducible, as reflected in acceptable levels of intra- and interobserver correlation (reliability), and reflective of what is encountered intra-operatively (validity). For acetabular and femoral bone loss, the three most commonly used classification systems are those of Paprosky, the AAOS, and Saleh and Gross systems [11, 14]. The purpose of this paper was to perform a systematic review of the literature and evaluate the currently available evidence on the reliability and validity of these classification systems. The results of this review demonstrate that, overall, these three acetabular and femoral bone loss classification systems are not consistently reproducible and have limited validity (although accuracy improves when grading is performed by evaluators who created the system themselves, an admittedly rare and impractical circumstance).

There are several limitations to this study. We were limited by the methodological quality of the studies, the small numbers of articles on each classification system, and the relatively small sample sizes. Also, four out of seven of the reviewed papers were retrospective cohort studies. Future studies assessing the reliability and validity of both acetabular and femoral bone loss classification systems using larger sample sizes and higher levels of evidence will allow for more evidence-based conclusions on this topic.

Table 2 Included studies with year of publication and study design

Study	Year	Type
Campbell et al. [4]	2001	Retrospective cohort
Saleh et al. [21]	2001	Prospective cohort
Davis et al. [7]	2003	Prospective cohort
Gozzard et al. [10]	2003	Prospective cohort
Parry et al. [19]	2010	Retrospective cohort
Yu et al. [29]	2013	Retrospective cohort
Brown et al. [2]	2014	Retrospective cohort

Table 3 Interobserver and intraobserver reliability and validity scores according to systems used

Mean sample size	Mean interobserver reliability: κ (range)	Number of studies	Mean intraobserver reliability (range)	Number of studies	Mean validity	Number of studies	
Acetabular Paprosky	44	0.46 (0.25–0.60)	4	0.45 (0.30–0.62)	3	0.65	1
Acetabular AAOS	30	0.24 (0.09–0.41)	3	0.51 (0.37–0.67)	3	— ^a	—
Acetabular Saleh	56	— ^a	2	—	—	—	2
Femoral Paprosky	84	0.57 (0.39–0.72)	3	0.63 (0.34–0.78)	3	0.54	1
Femoral AAOS	24	0.38 (0.09–0.66)	2	0.64 (0.46–0.81)	2	— ^a	—
Femoral Saleh	56	— ^a	2	—	—	—	2

Paprosky Paprosky classification for [femoral or acetabular] bone loss, AAOS American Academy of Orthopaedic Surgeons classification for [femoral or acetabular] bone loss, Saleh Saleh and Gross classification for [femoral or acetabular] bone loss

^aThe papers used different scales to grade interobserver reliability, and it is therefore not possible to report a mean value

The Paprosky classification remains the most commonly used system in the literature and in our practice [26, 27]. It is based on the presence or absence of the supporting structures of the hip (acetabular rim, superior dome, medial wall, anterior and posterior columns) on pre-operative AP radiographs [18]. In addition, this classification system is designed to guide treatment. Generally, type I acetabular defects have minimal deformity with an intact rim and can usually be treated with traditional uncemented hemispherical acetabular components [18]. Type II defects have varying degrees of superior (types IIA, IIB) or medial (type IIC) bone loss [18]. In such cases, augments, bone grafting, or both may be required, but there is often enough bone to achieve a rim fit. Type III defects have a higher degree of bone loss and encompass cases of pelvic discontinuity (type IIIB) [18]. Historically, cages or large structural bone grafts were used in these patients, whereas modern implant options include cup–cage constructs and custom triflanges. In the analysis of the Paprosky acetabular bone loss classification system, our study found overall moderate mean intraobserver and interobserver reliability. As a result, the Paprosky acetabular system is considered relatively unreliable. The AAOS recommends the AAOS classification proposed by D’Antonio et al. [6]. However, the available studies analyzing the validity and reliability of AAOS classification system found it, too, to be unreliable. The AAOS classification is based on the presence of segmental, cavitary, or combined defects. It is a descriptive method that does not relate directly to the magnitude of the defect or specific reconstructive options.

The two studies that evaluated the reliability of the third most commonly used classification system, Saleh and Gross, had conflicting results. The study by the system’s designers (Saleh et al.) showed good interobserver reliability and validity [22], whereas the non-creator evaluators reported poorer (moderate) validity and reliability [7]. Because of this discrepancy, we recommend additional studies to further determine validity and reliability of the Saleh classification system.

The Paprosky femoral bone loss classification system was shown to have moderate interobserver reliability and good intraobserver reliability, with the studies concluding that the system is reliable. The Paprosky classification system is based on the pattern and extent of metaphyseal and diaphyseal bone loss, which guide treatment. Within this system, type I femurs have minimal metaphyseal bone loss, type II femurs have extensive metaphyseal bone loss with an intact diaphysis, type IIIa femurs have extensive metadiaphyseal bone loss with at least 4 cm of intact diaphyseal cortex, type IIIB femurs have extensive metadiaphyseal bone loss with less than 4 cm of intact diaphyseal cortex, and type IV femurs have severe metadiaphyseal bone loss with a non-supportive diaphysis [14]. In the past, this system was able to effectively guide treatment because types I, II, and IIIa could all be treated with extensively porous-coated cylindrical stems [25]. However, as the use of monolithic and modular tapered fluted stems for femoral component revisions becomes more common, the ability of this classification scheme to guide treatment has diminished.

Table 4 Individual study results

Study	Systems analyzed	Sex	Mean age	Number of femurs/acetabula	Number and type of raters (intraobserver reliability)
Gozzard et al. [10]; (N = 25)	Acetabular Paprosky	—	—	25	4 (2 experts and 2 nonexperts)
	Femoral Paprosky	—	—	25	4 (2 experts and 2 nonexperts)
	Acetabular AAOS	—	—	25	4 (2 experts and 2 nonexperts)
Campbell et al. [4] (N = 30)	Femoral AAOS	—	—	25	4 (2 experts and 2 nonexperts)
	Acetabular Paprosky	—	—	33	6 (3 experts and 3 nonexperts)
	Acetabular AAOS	—	—	33	6 (3 experts and 3 nonexperts)
Yu et al. [28] (N = 83)	Acetabular Paprosky	—	—	85	2 (1 expert and 1 nonexpert)
	Acetabular AAOS	—	—	85	—
	Acetabular Paprosky	—	—	32	3 experts
Parry et al. [19] (N = 30)	Acetabular Paprosky	—	—	23	3 experts
	Acetabular AAOS	—	—	32	3 experts
	Femoral AAOS	—	—	23	3 experts
Saleh et al. [21] (N = 45)	Acetabular Saleh	—	—	45	—
	Femoral Saleh	—	—	45	—
	Acetabular Saleh	35 m, 46 f	68.5	81	—
Davis et al. [7] (N = 81)	Acetabular Saleh	35 m, 46 f	68.5	81	—
	Femoral Paprosky	—	—	205	4 experts
	Brown et al. [2] (N not reported)				
Study	Number and type of raters (interobserver reliability)	Mean intraobserver reliability (range)	Mean interobserver reliability (range)	Mean expert and nonexpert validity	
Gozzard et al. [10]; (N = 25)	4 (2 experts and 2 nonexperts)	0.30 (0.14–0.41)	0.42	0.65	
	4 (2 experts and 2 nonexperts)	0.34 (0.09–0.64)	0.39	0.54	
	4 (2 experts and 2 nonexperts)	0.48 (–0.01 to 0.74)	0.41	—	
	4 (2 experts and 2 nonexperts)	0.46 (0.08–0.64)	0.09	—	
Campbell et al. [4] (N = 30)	6 (3 experts and 3 nonexperts)	0.42 (0.26–0.60)	0.25 (0.02–0.41)	—	
	6 (3 experts and 3 nonexperts)	0.37 (0.05–0.75)	0.23 (0.08–0.45)	—	
Yu et al. [28] (N = 83)	2 (1 expert and 1 nonexpert)	^a	0.56	—	
	2 experts	—	0.49	—	
Parry et al. [19] (N = 30)	3 experts	0.62 (0.26–0.96)	0.60	—	
	3 experts	0.77 (0.60–0.88)	0.72	—	
	3 experts	0.67 (0.48–0.90)	0.09	—	
	3 experts	0.81 (0.71–0.90)	0.66	—	
Saleh et al. [21] (N = 45)	2 experts	—	> 0.75	> 0.75	
	2 experts	—	> 0.75	> 0.75	
Davis et al. [7] (N = 81)	3 experts	—	0.54	0.41	
	3 experts	—	0.56	0.39	
Brown et al. [2] (N not reported)	4 experts	0.78 (0.75–0.81)	0.61	—	

Paprosky Paprosky classification for [femoral or acetabular] bone loss, AAOS American Academy of Orthopaedic Surgeons classification for [femoral or acetabular] bone loss, Saleh Saleh and Gross classification for [femoral or acetabular] bone loss, m male, f female

^a Authors did not report intraobserver reliability before teaching lessons, so we excluded it from the analysis

The studies analyzing the validity and reliability of the AAOS femoral bone loss classification system reported a mean κ score that was fair for interobserver reliability and good for intraobserver reliability. The conclusion of these reports is that the AAOS femoral bone loss classification system is unreliable. The AAOS system divides the femoral bone defects into segmental and cavitary. Segmental defects include loss of supporting cortical bone, whereas cavitary defects are defined as any bony loss of the cancellous medullary bone [1]. Between the two studies observing the Saleh and Gross femoral bone loss classification, conflicting results were once again seen. Saleh et al. [21] showed good reliability and validity, whereas Davis et al. [7] showed moderate reliability and validity. Therefore, we recommend additional study to determine the validity and reliability of Saleh and femoral bone loss classification, as compared with other currently available scores [22].

Overall, the results of this systematic review of the literature evaluating the three most commonly used acetabular and femoral bone loss classification systems found variable reliability and validity. This may be related to several factors, such as the complexity of the classification system, the inadequacy of pre-operative radiographs, the experience of the graders, or a lack of education on their use. Yu et al. studied the impact of education and a visual grading sheet on the validity and reproducibility of the Paprosky classification system. After a teaching session that included interpretation and quantification of bone loss using standardized radiographic landmarks, a significant improvement in interobserver reliability was observed [29]. In addition, although numerous attempts have been made to quantify acetabular bone loss through cross-sectional imaging [12, 24], all classification systems currently in use are still based on plain radiographs [6, 11, 14, 19, 21, 22]. Although conventional AP and lateral radiographs remain the most common radiographic modalities used in revision THA, the reliability and validity of currently available grading systems might be improved if the recent improvements in and standardization of cross-sectional imaging and three-dimensional (3D) reconstruction with computed tomography were added into the equation. Integration of additional modalities, such as 3D printing, into the current classification systems might also increase their reliability and validity.

Although none of the studies analyzed in this systematic review employed computed tomographic (CT) scanning or 3D reconstructions, previous work has found computed tomography-based models to be superior to plain radiographs for assessing acetabular bone loss [17, 20]. In a series by Robertson et al., bone loss in 20 hips was measured using radiographs, CT scans, and CT-based 3D models. These measurements were compared with those found intraoperatively [20]. Radiographs were found to significantly underestimate bone loss and had poor interobserver reliability ($\kappa = 0.06$), whereas model-based planning had far higher interobserver reliability ($\kappa = 0.73$) and resulted in the correct selection of prosthesis in all cases [20]. Consequently, CT scanning and 3D reconstruction are often used in the pre-operative evaluation of patients with acetabular bone loss, especially in cases of more severe bone loss or suspected pelvic discontinuity [9, 26].

In conclusion, surgical techniques, implants, treatments, and imaging technologies have all evolved over the past decade. However, acetabular and femoral bone loss classification systems have remained largely unchanged since they were first developed in the 1990s. Our results indicate that improvements to these systems may be needed to keep pace with imaging technologies and implant designs and improve the reliability, validity, and accuracy of treatment recommendations.

Compliance with Ethical Standards

Conflict of Interest: Alex Gu, BS, Marco Adriani, MD, Michael-Alexander Malahias, MD, PhD, Allina A. Nocon, MPH, declare that they have no conflicts of interest. Safa C. Fassihi, MD, reports receiving research support from Intellijoint Surgical and being an unpaid consultant for Lima Corporate; Mathias P. Bostrom, MD, reports receiving royalties, consulting fees, and research support from Smith and Nephew, research support from the Mandl Research Foundation and the National Institutes of Health (the National Institute of Arthritis and Musculoskeletal and Skin Diseases and the Eunice Kennedy Shriver National Institute of Child Health and Human Development), serving on the editorial board of *HSS Journal*, and having board membership or committee appointments with the American Austrian Foundation and the Hip Society; Peter K. Sculco, MD, reports receiving research support from Intellijoint Surgical and being an unpaid consultant for Lima Corporate, outside the submitted work.

Human/Animal Rights: N/A

Informed Consent: N/A

Required Author Forms Disclosure forms provided by the authors are available with the online version of this article.

References

1. Anagnostakos K, Schmid NV, Kelm J, Grün U, Jung J. Classification of hip joint infections. *Int J Med Sci*. 2009;6(5):227–233.
2. Brown NM, Foran JR, Valle CJ, et al. The inter-observer and intra-observer reliability of the Paprosky femoral bone loss classification system. *J Arthroplasty*. 2014;29(7):1482–1484.
3. Callaghan JJ. *The Adult Knee*. Lippincott Williams and Wilkins. 2003.
4. Campbell DG, Garbuz DS, Masri BA, Duncan CP. Reliability of acetabular bone defect classification systems in revision total hip arthroplasty. *J Arthroplasty*. 2001;16(1):83–86.
5. D'Antonio JA. Periprosthetic bone loss of the acetabulum. Classification and management. *Orthop Clin North Am*. 1992;23(2):279–290.
6. D'Antonio J, Capello W, Borden L, et al. Classification and management of acetabular abnormalities in total hip arthroplasty. *Clin Orthop Relat Res*. 1989;(243):126–137.
7. Davis AM, Schemitsch EH, Gollish JD, et al. Classifying failed hip arthroplasty: generalizability of reliability and validity. *Clin Orthop Relat Res*. 2003;(415):171–179.
8. Della Valle CJ, Paprosky WG. The femur in revision total hip arthroplasty evaluation and classification. *Clin Orthop Relat Res*. 2004;(420):55–62.
9. Fehring K, Howe B, Martin J, Taunton M, Berry D. Preoperative evaluation for pelvic discontinuity using a new reformatted computed tomography scan protocol. *J Arthroplasty*. 2016;31(10):2247–2251.

10. Gozzard C, Blom A, Taylor A, Smith E, Learmonth I. A comparison of the reliability and validity of bone stock loss classification systems used for revision hip surgery. *J Arthroplasty*. 2003;18(5):638–642.
11. Gross AE, Duncan CP, Garbuz D, Mohamed EM. Revision arthroplasty of the acetabulum in association with loss of bone stock. *Instr Course Lect*. 1999;48:57–66.
12. Hettich G, Schierjott RA, Ramm H. Method for quantitative assessment of acetabular bone defects. *J Orthop Res*. 2019;37(1):181–189.
13. Howick J, Chalmers I, Galsziou P, et al. Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document). Oxford Centre for Evidence-Based Medicine. <https://www.cebm.net/index.aspx?o=5653>.
14. Ibrahim DA, Fernando ND. Classifications in brief: the Paprosky classification of femoral bone loss. *Clin Orthop Relat Res*. 2017;475(3):917–921.
15. Kurtz S, Lau E, Ong K, Zhao K, Kelly M, and Bozic KJ. Future young patient demand for primary and revision joint replacement: national projection from 2010 to 2030. *Clin Orthop Relat Res*. 2009;467: 2606–2612.
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1): 159–174.
17. Munjal S, Leopold SS, Kornreich D, Shott S, Finn HA. CT-generated 3-dimensional models for complex acetabular reconstruction. *J Arthroplasty*. 2000;15(5):644–653.
18. Paprosky WG, Perona PG, Lawrence JM. Acetabular defect classification and surgical reconstruction in revision arthroplasty. A 6-year follow-up evaluation. *J Arthroplasty*. 1994;9(1):33–44.
19. Parry MC, Whitehouse MR, Mehendale SA, et al. A comparison of the validity and reliability of established bone stock loss classification systems and the proposal of a novel classification system. *Hip Int*. 2010;20(1):50–55.
20. Robertson DD, Sutherland CJ, Lopes T, Yuan J. Preoperative description of severe acetabular defects caused by failed total hip replacement. *J Comput Assist Tomogr*. 1998;22(3):444–449.
21. Saleh KJ, Holtzman J, Gafni A, et al. Reliability and intraoperative validity of preoperative assessment of standardized plain radiographs in predicting bone loss at revision hip surgery. *J Bone Joint Surg Am*. 2001;83-A(7):1040–1046.
22. Saleh KJ, Holtzman J, Gafni A, et al. Development, test reliability and validation of a classification for revision hip arthroplasty. *J Orthop Res*. 2001;19(1):50–56.
23. Sambandam SN, Gul A, Priyanka P. Analysis of methodological deficiencies of studies reporting surgical outcome following cemented total-joint arthroplasty of trapezio-metacarpal joint of the thumb. *Int Orthop*. 2007;31(5):639–645.
24. Schierjott RA, Hettich G, Graichen H, et al. Quantitative assessment of acetabular bone defects: a study of 50 computed tomography data sets. *PLoS One*. 2019;14(10):e0222511.
25. Sheth NP, Nelson CL, Paprosky WG. Femoral bone loss in revision total hip arthroplasty: evaluation and management. *J Am Acad Orthop Surg*. 2013;21(10):601–612.
26. Sheth NP, Nelson CL, Springer BD, Fehring TK, Paprosky WG. Acetabular bone loss in revision total hip arthroplasty: evaluation and management. *J Am Acad Orthop Surg* 2013;21(3):128–139.
27. Telleria JJ, Gee AO. Classifications in brief: Paprosky classification of acetabular bone loss. *Clin Orthop Relat Res*. 2013;471(11):3725–3730.
28. Ulrich SD, Seyler TM, Bennett D, et al. Total hip arthroplasties: what are the reasons for revision? *Int Orthop*. 2008;32(5):597–604.
29. Yu R, Hofstaetter JG, Sullivan T, Costi K, Howie DW, Solomon LB. Validity and reliability of the Paprosky acetabular defect classification. *Clin Orthop Relat Res*. 2013;471(7):2259–2265.