J Shoulder Elbow Surg (2009) ■, 1-6

ELSEVIER

Journal of Shoulder and Elbow Surgery

www.elsevier.com/locate/ymse

Management of proximal humeral fractures: Surgeons don't agree

Charles J. Petit, MD[a,b], Peter J. Millett, MD MSc[b,*], Nathan K. Endres, MD[a], David Diller, BA[a], Mitchel B. Harris, MD[c], Jon J.P. Warner, MD[a]

[a]*Massachusetts General Hospital, Boston, MA, USA*
[b]*Steadman Hawkins Clinic and Steadman Hawkins Research Foundation, Vail, CO, USA*
[c]*Brigham and Women's Hospital, Boston, MA, USA*

Hypothesis: The management options for proximal humeral fractures have expanded in recent years. Patients with displaced, unstable proximal humeral fractures may have improved outcomes if managed operatively. We investigated the decision making of fellowship-trained orthopedic surgeons when presented with the same group of cases. We hypothesized that interobserver and intraobserver agreement for surgical management would be poor and independent of fellowship training.

Method: Eight fellowship-trained orthopedic surgeons (3 shoulder, 5 trauma) viewed the preoperative plain radiographs of patients with proximal humeral fractures. All surgeons viewed the same 38 radiographs in a blinded fashion. Surgeons chose from 1 of 6 management options. Interobserver variability was calculated by using the weighted κ coefficient. Intraobserver variability was calculated by comparing each surgeon's survey results with the operation they originally performed.

Results: Overall interobserver agreement on management was moderate (weighted $\kappa = 0.41$) and did not differ significantly between trauma surgeons and shoulder surgeons. Reducing the number of management choices increased agreement between all surgeons. Testing for intraobserver agreement showed that surgeons picked the same operation in the survey as in the actual clinical setting only 56% of the time.

Conclusion: Interobserver agreement was moderate overall and improved when the number of management choices was reduced. Intraobserver agreement was less frequent, however, raising the question about consistent decision making by a given surgeon. Although surgeons agree in the method of treatment only to a modest degree, it remains for further outcomes research to establish if the choice of treatment actually influences the clinical outcome.

Level of evidence: Level 4, case series.

Keywords: Proximal humeral fractures; surgeon decision making; intraobserver and intraobserver agreement; surgical management

Proximal humeral fractures account for 4% to 5% of all fractures in adults, with an incidence of 6.6/1000 person-years.[1] In patients aged older than 65, they represent the third most common fracture, after hip and distal radius fractures.[1] Proximal humeral fractures are increasing in incidence as the population ages.[10] Historically, up to 80% of proximal humeral fractures have been considered appropriate for nonoperative treatment.[17] A wide variety of options are available for the remaining 20% in which operative fixation is

*Reprint requests: Peter J. Millett MD, MSc, Steadman Hawkins Research Foundation, 181 West Meadow Drive, Suite 400, Vail, CO 81657.

E-mail address: Marilee.horan@shsmf.org (P.J. Millett).

1058-2746/2009/$36.00 - see front matter © 2009 Journal of Shoulder and Elbow Surgery Board of Trustees.
doi:10.1016/j.jse.2009.06.012

indicated. In recent years the management options for proximal humeral fractures have expanded to include closed reduction alone, closed reduction and percutaneous pinning, open reduction and transosseous suture fixation, intramedullary nailing (IM), open reduction and internal fixation (ORIF) with conventional plating techniques, ORIF with fixed angle devices, and arthroplasty.[6,7,14,19,20,27] The current recommendation is that fractures of the proximal humerus with parts that are displaced greater than 45° or 1 cm be managed with closed or open reduction and operative fixation, as determined by the stability of the fragments and patient-specific factors.[17,18]

Despite these recommendations, the treatment rendered often varies significantly between surgeons. There is currently no literature clearly illustrating the variability of surgical decision making between surgeons with regard to proximal humeral fractures. The purpose of our study was to investigate the decision making of shoulder and trauma fellowship-trained orthopedic surgeons when presented with the same group of patients managed at 2 affiliated level 1 trauma centers. We hypothesized that the interobserver agreement for surgical management based on plain radiographs would be poor and independent of fellowship training, and further hypothesized that intraobserver agreement would also be poor.

Materials and methods

This study was approved by the Institutional Review Board of Partners Human Research Committee (IRB# 2005-P-002450/1). This investigation was performed at Massachusetts General Hospital and Brigham and Women's Hospital, Boston, Massachusetts.

Eight attending orthopedic surgeons, including 3 authors (M.B.H., P.J.M., J.P.W.), consented to participate in the study. Among the 8 surgeons, 3 were fellowship-trained in shoulder surgery, 4 in orthopedic traumatology, and 1 completed an upper extremity surgery fellowship specializing in upper extremity trauma and was included with the other 4 trauma surgeons for statistical purposes. All surgeons were well experienced in both the operative and nonoperative treatment of these fractures and had an average of 12.6 years of experience after fellowship, as summarized in Table I. Furthermore, all reviewers were experienced in operative fixation methods discussed in the survey, each reporting treating approximately 20 cases operatively each year.

Excluding isolated tuberosity fractures, all proximal humeral fractures treated operatively at 2 affiliated level 1 trauma centers between 2000 and 2005 by these surgeons were gathered into a database for a total of 185 patients. Patients that possessed adequate and available preoperative radiographs were included in the survey (94 operative cases plus 6 randomly chosen nonoperative cases). Computed tomography (CT) scans were not included for review.

Two authors who did not participate in the survey constructed a PowerPoint (Microsoft Corp, Redmond, WA) file for each surgeon to review. Each case presentation included the pertinent history, physical examination findings, and medical comorbidities along with at least 3 radiographs (scapular anteroposterior, axillary, scapular Y views). Each survey included a common pool of 32 operative cases (4 from each surgeon, chosen to represent a wide variety of Neer fracture types) and 6 nonoperative cases, which were reviewed by all surgeons. In addition, each surgeon's survey contained all of his own patients during that time period treated operatively (with adequate and available preoperative radiographs) that were not included in the common pool (range, 3-11 additional cases). Patients were given a random number and presented in a blinded fashion to the surgeons.

For each patient, surgeons were asked to choose 1 of 6 management options (Table II): (1) nonoperative, (2) closed manipulation under anesthesia, (3) closed or open reduction and percutaneous pinning; (4) ORIF with a fixed-angle device (proximal humeral locking plate or blade plate); (5) ORIF with an alternative fixation device (such as suture fixation, IM nail, or tension band); or (6) hemiarthroplasty.

Statistical analysis

Interobserver variability was calculated using the κ coefficient generated by SAS 9.1 software (SAS Institute Inc, Cary, NC). As described by Landis,[13] the κ coefficient measures the percentage of instances of agreement, taking into account agreement by chance alone. A value of $\kappa = 1.00$ indicates perfect agreement, whereas $\kappa = 0.00$ indicates no more agreement than expected by chance alone. According to the guidelines presented by Landis, values of less than 0.20 represent poor or slight agreement, 0.21 to 0.40 represent fair agreement, 0.41 to 0.60 represent moderate agreement, 0.61 to 0.80 represent substantial agreement, and values of 0.81 or greater represent excellent agreement. Intraobserver variability was calculated comparing the percentage of time surgeons chose the same option in the survey as in the original clinical setting.

The χ^2 test was used to test for differences between groups. The correlation coefficient was used to test for correlation between years of experience and intraobserver variability. The statistics for interobserver and intraobserver variability were run again, decreasing the heterogeneity into 3 armed responses: closed treatment (choices 1 and 2) vs operative fixation (choices 3, 4, and 5) vs hemiarthroplasty (choice 6). This was again repeated, further decreasing heterogeneity into 2 armed responses: nonoperative management (choices 1 and 2) vs operative management (choices 3, 4, 5, and 6).

Results

The common pool of 38 patients reviewed by each surgeon contained 32 operative and 6 nonoperative cases. The original treatment performed on the 32 operative patients was as follows: closed manipulations under anesthesia in 2; closed reduction percutaneous pinning in 5; proximal humeral locking plates in 17; and hemiarthroplasty in 8. Of the 38 patients, there was perfect agreement among all 8 surgeons in only 4 patients: 2 nonoperative, 1 proximal humeral locking plate, and 1 hemiarthroplasty.

Interobserver agreement between all surgeons by κ coefficient was 0.41 (95% confidence interval [CI],

Table I Intraobserver agreement

Surgeon	Individual								
	Specialty	Years of exp	Cases	Perfect match	%	Non-op vs fix vs hemi	%	Non-op vs op	%
1	Trauma	4	8	4	50.0	4	50.0	5	62.5
2	Trauma	22	8	5	62.5	5	62.5	5	62.5
3	Shoulder	6	7	3	42.9	6	85.7	6	85.7
4	Shoulder	6	11	5	45.5	9	81.8	10	90.9
5	Trauma	8	15	11	73.3	11	73.3	14	93.3
6	Trauma	20	13	8	61.5	8	61.5	8	61.5
7	Trauma	17	9	4	44.4	4	44.4	7	77.8
8	Shoulder	18	14	8	57.1	13	92.9	14	100.0
	Weighted average								
	Overall		85	48	56.5	60	70.6	69	81.2
	Trauma		53	32	60.4	32	60.4	39	73.6
	Shoulder		32	16	50.0	28	87.5	30	93.8

Table II Management choices presented for each case

Choice	Description
1	Nonoperative
2	Closed manipulation under anesthesia
3	Closed reduction and percutaneous pinning
4	ORIF with fixed-angle device (proximal humeral locking plate)
5	ORIF with alternative method (suture fixation)
6	Hemiarthroplasty

ORIF, open reduction, internal fixation.

0.38-0.44). The interobserver agreement between trauma surgeons only was 0.44 (95% CI, 0.37-0.51), and the interobserver agreement between shoulder surgeons only was 0.50 (95% CI, 0.42-0.59). Decreasing the heterogeneity of responses to 3 possibilities (nonoperative: 1 and 2; internal fixation: 3, 4, and 5; hemiarthroplasty: 6) resulted in a slight increase in the overall κ coefficient to 0.45 (95% CI, 0.42-0.48), the κ coefficient among trauma surgeons to 0.44 (95% CI, 0.36-0.51), and the κ coefficient among shoulder surgeons to 0.52 (95% CI, 0.42-0.63).

Further decreasing the heterogeneity of responses to 2 possibilities (nonoperative: 1 and 2; operative: 3, 4, 5, 6) showed a higher level of agreement among surgeons. Overall, the κ coefficient was 0.48 (95% CI, 0.44-0.52) for the 2-armed response. The agreement between trauma surgeons increased to 0.47 (95% CI, 0.38-0.55), and the agreement between shoulder surgeons increased to 0.71 (95% CI, 0.44-0.71). There was no statistically significant difference between subspecialties.

The number of cases used for intraobserver agreement ranged from 7 to 15 (Table I). An evaluation of each surgeon's intraobserver agreement showed that surgeons chose the same management option in the survey as in the original clinical setting only 56.5% (range, 44.9%-73.3%) of the time. Intraobserver agreement was 60.4% (range, 44.4%-73.3%) for trauma surgeons and 50.0% (range, 42.9%-57.1%) for shoulder surgeons. This difference was not significantly different.

When the heterogeneity of responses was decreased to 3 armed responses (nonoperative: 1, 2; internal fixation: 3, 4, 5; and hemiarthroplasty: 6), the overall intraobserver agreement increased to 70.6% (range, 44.4%-92.9%). Shoulder surgeons showed an increase in agreement to 87.5% (range, 81.8%-92.9%), whereas the intraobserver agreement for trauma surgeons was unchanged at 60.4% (range, 44.4%-73.3%). For the 3-armed response, shoulder surgeons were more consistent in their decision making than trauma surgeons ($P = .008$). Further decreasing the heterogeneity of responses to nonoperative (1 and 2) and operative (3-6) resulted in an increase in overall intraobserver agreement to 81.2% (range, 61.5%-100%). For the 2-armed responses, intraobserver agreement among shoulder surgeons increased to 93.8% (range, 85.7%-100%) and increased among trauma surgeons to 73.6% (range, 61.5%-93.3%). Again, for the 2-armed responses, shoulder surgeons were more consistent in their decision making than trauma surgeons ($P = .02$).

When intraobserver agreement was evaluated with respect to the surgeon's experience, no significant correlation was found between years of experience and intraobserver agreement. Further, when surgeons were divided into those with more than 10 years experience (4 surgeons) and less than 10 years experience (4 surgeons), no difference was found between groups using χ^2 analysis.

Discussion

This study illustrates an overall lack of agreement regarding management among 8 orthopedic surgeons who actively treat proximal humeral fractures at 2 affiliated

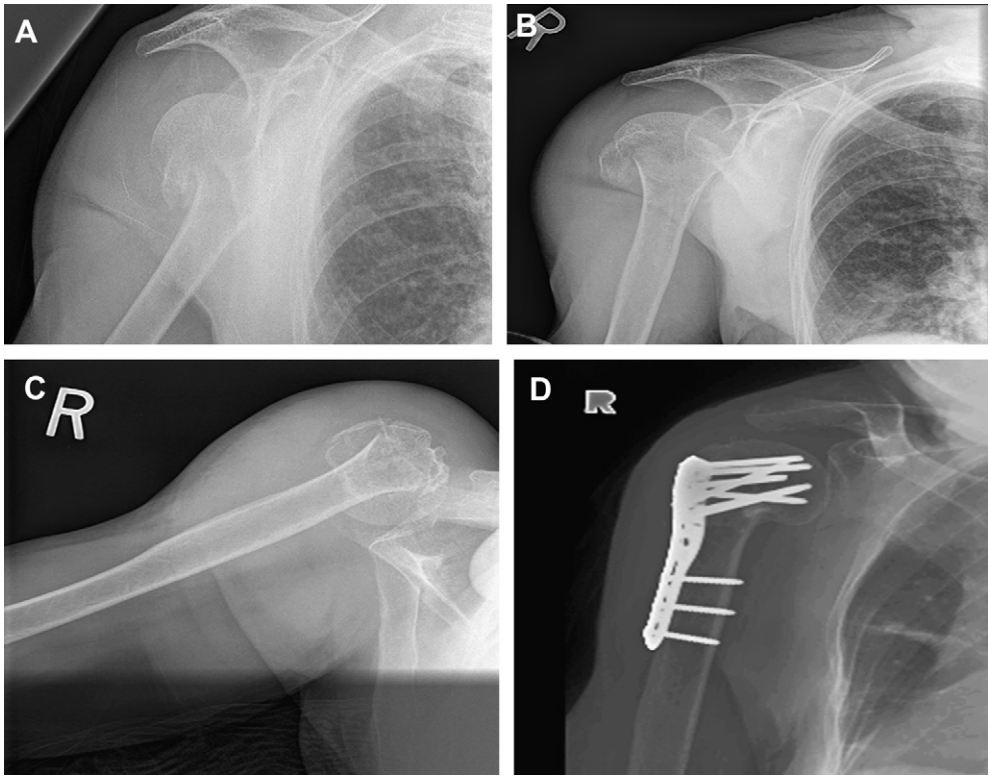

Figure 1 **(A-D)** Preoperative and postoperative images of a 79-year-old retired woman without significant medical comorbidities. Only 2 of 8 surgeons agreed with the initial management. The treating surgeon chose nonoperative management on the survey.

hospitals. Further, this study shows a concerning lack of consistency among surgeons in their own decision making.

Despite the increase in frequency of these fractures and the variety of treatment options available, few studies have directly compared one treatment method with another. A recent systematic review of treatment modalities for proximal humeral fractures found that less than 15% of the studies included in the review established an appropriate study design.[14] Randomized studies are lacking, and outcome measures are often not uniform.[3,6]

Most studies present only the outcomes of a specific intervention. Therefore, when a surgeon chooses one method over another, there is little to guide decision making other than individual surgeon experience and comfort level with particular techniques. As technology and understanding of these injuries has improved over time, treatment algorithms have evolved.[13] Recent emphasis has been on the results of locked plating for displaced proximal humeral fractures.[4,7,9,11,15,21] However, there is still considerable disagreement about the optimal treatment method for different fracture patterns among surgeons who treat these injuries.

One reason for the lack of consensus among surgeons treating proximal humeral fractures is the lack of agreement regarding classification. The 2 most recognized classification systems for proximal humeral fractures are the Neer[17] classification and the AO/OTA classification.[16] The Neer classification is based on the recognition of displacement of various anatomic components of the fracture. The AO/OTA classification includes information about the energy and severity of the fracture as well as the likelihood of vascular disruption. Both systems are based on plain radiographs and have relatively limited intraobserver reliability and interobserver reproducibility.[2,12,22,24,25]

In a study of 100 anteroposterior and lateral radiographs of surgical neck fractures, Kristiansen et al[12] found a low degree of agreement using the Neer classification. Siebenrock and Gerber[24] also reported that the interobserver and intraobserver reliabilities for both the Neer and AO/OTA classifications systems were fair or poor. Sidor et al[23] demonstrated a slightly improved mean interobserver reliability coefficient as well as mean intraobserver reproducibility for the Neer classification.

One could argue that the addition of a CT scan might have improved consensus in our study, and indeed, many of these patients did undergo a CT scan. In a study assessing the Neer classification system with the use of radiographs, Bernstein et al[2] determined the intraobserver reliability was substantial and interobserver reproducibility was moderate. The addition of a CT scan was associated with a slight increase in intraobserver reliability, but no increase in interobserver reproducibility. Sjoden et al[25] also concluded that the addition of a CT scan did not improve the reproducibility of either the Neer or AO/OTA classifications.

Sallay et al[22] divided 2 groups of observers into experts and nonexperts in shoulder surgery. Each group was asked

to review the radiographs and 3-dimensional CT scans of 12 patients with proximal humeral fractures. Both groups demonstrated low reliability for the identification of displaced fracture fragments. The addition of a CT scan did not improve reliability or reproducibility.[22]

Recently, Hertel et al[8] published radiographic criteria for determining perfusion of the humeral head. Although we have found these guidelines to be extremely useful when approaching these fractures, the implications in changing treatment guidelines and outcomes are unknown. This lack of consensus for the utility of a CT scan for classification of fractures led us to choose not to include a CT scan in our study.

Our study highlights the lack of agreement and inconsistency in decision making among surgeons who treat proximal humeral fractures. Our findings did not differ significantly among fellowship-trained shoulder surgeons and trauma surgeons. We did find that agreement improved when decision making was limited to broader categories (nonoperative vs operative), suggesting that surgeons do tend to agree on the general indications for surgery. There was much less agreement when surgeons were asked to determine a specific method of operative treatment (Figure 1).

Even when presented with actual previous cases, surgeons chose a different treatment option nearly 50% of the time. As with interobserver agreement, the intraobserver agreement increased when decision making was limited to broader categories. Shoulder surgeons did show more consistent decision making than trauma surgeons when broader categories were considered; however, the significance of this is unknown. Potential explanations for the poor intraobserver agreement demonstrated in this study include a differing interpretation of the radiographs or an evolution in the surgeon's approach to the treatment of these injuries. However, this observation clearly shows a lack of consistent decision making for a given surgeon.

Previous publications in the field of arthroplasty have shown improved outcomes and decreased hospital charges for higher-volume surgeons than lower-volume surgeons.[5,26] Because each of our surgeons reported operatively treating about 20 proximal humeral fractures per year, we were not able to test interobserver or intraobserver agreement vs volume. When years of experience was taken into consideration, no correlation between experience and intraobserver agreement was found. When surgeons were grouped into those with less than and greater than 10 years' experience, no statistical difference was found. One explanation for this difference is that the treatment of proximal humeral fractures has evolved significantly in recent years. The proximal humeral locking plate is one new device that has gained popularity in recent years. However, the current literature lacks prospective data to guide surgeons' use of this device, and as a result, a surgeon's experience before the widespread use of this device likely did not add to consistent decision making.

One limitation of this study is that in presenting cases to surgeons using this method, we cannot provide all available information for each patient. For many of these patients, a CT scan was available for review, which might or might not have changed the treatment choice. Obviously, a preoperative physician-patient discussion about the risks and benefits of treatment methods is impossible. In addition, the information about each case was gathered in a retrospective fashion.

In conclusion, given the current availability of numerous treatment options for proximal humeral fractures, this study highlights the need for prospective randomized studies with consistent outcome measures to determine the optimal treatment of specific patterns of proximal humeral fractures. As these studies become available, we expect to see more uniformity in the treatment of these challenging fractures. Finally, no conclusion can be made about the effect of decision making on outcome of treatment. More than one method of treatment for a given fracture may result in a satisfactory outcome, even if surgeons do not have uniformity of opinion about the method of treatment. This question would be an important subsequent consideration in future studies.

Disclaimer

Drs. Diller, Endres, Harris, and Warner have not received nor will receive something of value from or own stock in a commercial company or institution related directly or indirectly to the subject of this manuscript. Dr Petit and Dr Millett receive research support from the Steadman-Hawkins Research Foundation that receives financial support from Arthrex, Ossur, Smith & Nephew, Arthrocare, Orthorehab, and Siemens. IRB providing approval: Partners Human Research Committee, IRB# 2005-P-002450/1.

References

1. Baron JA, Barrett JA, Karagas MR. The epidemiology of peripheral fractures. Bone 1996;18(3 suppl):209-13S.
2. Bernstein J, Adler LM, Blank JE, Dalsey RM, Williams GR, Iannotti JP. Evaluation of the neer system of classification of proximal humeral fractures with computerized tomographic scans and plain radiographs. J Bone Joint Surg Am 1996;78:1371-5.
3. Bhandari M, Matthys G, McKee MD. Evidence-Based Orthopaedic Trauma Working Group. Four part fractures of the proximal humerus. J Orthop Trauma 2004;18:126-7.
4. Fankhauser F, Boldin C, Schippinger G, Haunschmid C, Szyszkowitz R. A new locking plate for unstable fractures of the proximal humerus. Clin Orthop Relat Res 2005:176-81.
5. Hammond JW, Queale WS, Kim TK, McFarland EG. Surgeon experience and clinical and economic outcomes for shoulder arthroplasty. J Bone Joint Surg Am 2003;85:2318-24.
6. Handoll H, Madhok R. Interventions for treating proximal humeral fractures in adults. The Cochrane Library 2007:1-50.
7. Helmy N, Hintermann B. New trends in the treatment of proximal humerus fractures. Clin Orthop Relat Res 2006;442:100-8.

8. Hertel R, Hempfing A, Stiehler M, Leunig M. Predictors of humeral head ischemia after intracapsular fracture of the proximal humerus. J Shoulder Elbow Surg 2004;13:427-33.
9. Hirschmann MT, Quarz V, Audige L, et al. Internal fixation of unstable proximal humerus fractures with an anatomically preshaped interlocking plate: a clinical and radiologic evaluation. J Trauma 2007;63:1314-23.
10. Kannus P, Palvanen M, Niemi S, Parkkari J, Jarvinen M, Vuori I. Increasing number and incidence of osteoporotic fractures of the proximal humerus in elderly people. BMJ 1996;313:1051-2.
11. Koukakis A, Apostolou CD, Taneja T, Korres DS, Amini A. Fixation of proximal humerus fractures using the PHILOS plate: early experience. Clin Orthop Relat Res 2006;442:115-20.
12. Kristiansen B, Andersen UL, Olsen CA, Varmarken JE. The neer classification of fractures of the proximal humerus. an assessment of interobserver variation. Skeletal Radiol 1988;17:420-2.
13. Landis JR, Koch GG. The measurement of observeragreement for categorical data Biometrics 1977;33:159−174.
14. Lanting B, MacDermid J, Drosdowech D, Faber KJ. Proximal humeral fractures: a systematic review of treatment modalities. J Shoulder Elbow Surg 2008;17:42-54.
15. Moonot P, Ashwood N, Hamlet M. Early results for treatment of three- and four-part fractures of the proximal humerus using the PHILOS plate system. J Bone Joint Surg Br 2007;89:1206-9.
16. Müller ME, Nazarus S, Koch P, Schatzker J. The comprehensive classification of fractures of long bones. 1st ed. Berlin: Springer-Verlag; 1990.
17. Neer CS 2nd. Displaced proximal humeral fractures. I. Classification and evaluation. J Bone Joint Surg Am 1970;52:1077-89.
18. Neer CS 2nd. Displaced proximal humeral fractures. II. Treatment of three-part and four-part displacement. J Bone Joint Surg Am 1970;52: 1090-103.
19. Nho SJ, Brophy RH, Barker JU, Cornell CN, MacGillivray JD. Management of proximal humeral fractures based on current literature. J Bone Joint Surg Am 2007;89(suppl 3):44-58.
20. Nho SJ, Brophy RH, Barker JU, Cornell CN, MacGillivray JD. Innovations in the management of displaced proximal humerus fractures. J Am Acad Orthop Surg 2007;15:12-26.
21. Owsley KC, Gorczyca JT. Displacement/screw cutout after open reduction and locked plate fixation of humeral fractures. J Bone Joint Surg Am 2008;90:233-40.
22. Sallay PI, Pedowitz RA, Mallon WJ, Vandemark RM, Dalton JD, Speer KP. Reliability and reproducibility of radiographic interpretation of proximal humeral fracture pathoanatomy. J Shoulder Elbow Surg 1997;6:60-9.
23. Sidor ML, Zuckerman JD, Lyon T, Koval K, Cuomo F, Schoenberg N. The Neer classification system for proximal humeral fractures. An assessment of interobserver reliability and intraobserver reproducibility. J Bone Joint Surg Am 1993;75:1745-50.
24. Siebenrock KA, Gerber C. The reproducibility of classification of fractures of the proximal end of the humerus. J Bone Joint Surg Am 1993;75:1751-5.
25. Sjoden GO, Movin T, Guntner P, et al. Poor reproducibility of classification of proximal humeral fractures. additional CT of minor value. Acta Orthop Scand 1997;68:239-42.
26. Taylor HD, Dennis DA, Crane HS. Relationship between mortality rates and hospital patient volume for Medicare patients undergoing major orthopaedic surgery of the hip, knee, spine, and femura. J Arthroplasty 1997;12:235-42.
27. Vallier HA. Treatment of proximal humerus fractures. J Orthop Trauma 2007;21:469-76.