

Reliability of Pathologic Anterior Instability Presence on Shoulder Imaging – Methodological Issues: Letter to the Editor

Dear Editor:

We were interested to read an article published by Beason and colleagues in the August 2019 issue of the *Orthopaedic Journal of Sports Medicine*.¹ The authors aimed to determine the reliability of pathological diagnostic indices associated with anterior shoulder instability using plain radiography and magnetic resonance imaging (MRI).¹ To gain the inter- and intrarater reliability, shoulder/sports medicine surgeons reviewed 40 sets of images (20 radiograph sets, 20 MRI series) over 2 points in time. Finally, analysis and interpretation were performed with the kappa coefficient. The authors reported that kappa values for the interrater agreement of shoulder radiographs were 0.49, 0.59, 0.35, and 0.50 for the presence of glenoid lesions, the estimate of glenoid lesion surface area, the presence of Hill-Sachs lesion, and the estimate of Hill-Sachs surface area, respectively. For the intrarater agreement of radiographs, it was 0.48 to 0.57. Also, kappa values for the intrarater reliability of shoulder MRI were 0.59, 0.52, 0.50, 0.51, 0.53, and 0.63 for the presence of glenoid lesions, the presence of a Hill-Sachs lesion, the estimate of Hill-Sachs surface area, humeral head edema, the presence of a capsulolabral injury, and glenoid lesion surface area, respectively, while the intrarater agreement for determining the specific type of capsulolabral injury was fair ($\kappa = 0.48$).¹

There are methodological issues in the reliability assessment that can affect the outcome or main message of the study. One of the drawbacks is the use of kappa, which in certain circumstances can affect the results of the study for the following reasons. First, the amount of kappa depends on the prevalence in each group. Second, it depends on the number of categories.²⁻⁶ It should be noted that when a variable with >2 categories or an ordinal scale (arranged in ≥ 3 categories) is used, a weighted kappa would be a good choice. Finally, the third problem is when 2 voters have uneven marginal distributions in their responses.²⁻⁶ Table 1 shows agreement with different interpretations and conclusions based on kappa (0.43 as moderate) and weighted kappa values (0.63 as good). In this

Final revision submitted September 11, 2019; accepted October 22, 2019.

The Orthopaedic Journal of Sports Medicine, 8(1), 2325967119893970
DOI: 10.1177/2325967119893970
© The Author(s) 2020

This open-access article is published and distributed under the Creative Commons Attribution - NonCommercial - No Derivatives License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits the noncommercial use, distribution, and reproduction of the article in any medium, provided the original author and source are credited. You may not alter, transform, or build upon this article without the permission of the Author(s). For article reuse guidelines, please visit SAGE's website at <http://www.sagepub.com/journals-permissions>.

TABLE 1
Kappa and Weighted Kappa Values for Calculating Agreement Between Surgeons for >2 Categories

Surgeon 2	Surgeon 1			Sum
	Grade 1	Grade 2	Grade 3	
Grade 1	60	20	1	81
Grade 2	2	12	4	18
Grade 3	3	11	11	25
Sum	65	43	16	124
Estimate				
Kappa	0.43			
Weighted kappa	0.63			

table, the number of the category (>2) and the marginal distribution of the first category (grade 1) are different from the other categories.^{2,3}

Poor to moderate agreement was reported by surgeons to evaluate imaging studies of anterior shoulder instability. Also, agreement on identifying pathologic features on radiography and MRI was similar, while agreement on the presence of glenoid injury tended to improve but was low for specific capsular lesions.¹ In this letter, we discuss important limitations of applying the Cohen kappa coefficient to assess reliability.²⁻⁶ Any conclusion on reliability analyses needs to be supported by the methodological and statistical issues mentioned here. Otherwise, misinterpretation cannot be avoided.

Mehdi Naderi, MSc
Kermanshah, Iran

Siamak Sabour, MD, PhD
Tehran, Iran

Address correspondence to Siamak Sabour, MD, PhD (email: s.sabour@sbm.ac.ir).

The authors declared that there are no conflicts of interest in the authorship and publication of this contribution. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

REFERENCES

1. Beason AM, Koehler RJ, Sanders RA. Surgeon agreement on the presence of pathologic anterior instability on shoulder imaging studies. *Orthop J Sports Med*. 2019;7(8):2325967119862501.
2. Naderi M, Sabour S. Reproducibility of the Bethesda System for reporting thyroid cytopathology: a methodological issue. *J Cytol*. 2019;36(3):185-186.

3. Naderi M, Sabour S. Reproducibility of diagnostic criteria associated with atypical breast cytology: a methodological issue. *Cytopathology*. 2018;29(4):396.
4. Sabour S. Reliability of the ASA physical status scale in clinical practice: methodological issues. *Br J Anaesth*. 2015;114(1):162-163.
5. Sabour S, Dastjerdi EV. Reliability of four different computerized cephalometric analysis programs: a methodological error. *Eur J Orthod*. 2013;35(6):848.
6. Szklo M, Nieto FJ. *Epidemiology Beyond the Basics*. 3rd ed. Manhattan, NY: Jones & Bartlett; 2014.