MRI Findings at the Bone-Component Interface in Symptomatic Unicompartmental Knee Arthroplasty and the Relationship to Radiographic Findings

Mehdi Naderi, MSc · Siamak Sabour, MD, PhD

Received: 6 October 2018/Accepted: 9 January 2019/Published online: 3 April 2019
© Hospital for Special Surgery 2019

Keywords MRI · unicompartmental knee arthroplasty · multiacquisition variable-resonance image combination (MAVRIC) · bone–component interface · inter-rater reliability · methodological issues

Dear Editor,

We were interested to read the article by Kleeblad and colleagues published in the October 2018 issue of HSS Journal [1]. One of this study's valuable goals was to determine the inter-rater agreement of individual magnetic resonance imaging (MRI) findings at the bone–component interface, which was assessed with Cohen's kappa statistic (κ). In this study, 55 patients with cemented unicompartmental knee arthroplasty (UKA) were examined. Using MRI with multiacquisition variable-

resonance image combination (MAVRIC) sequences, the authors reported excellent inter-rater agreement for assessment of the bone–component interface along the femoral component for all findings ($\kappa > 0.830$). Also, the inter-rater agreement of bone marrow edema and fibrous membrane at the tibial interface were substantial ($\kappa = 0.703$ and $\kappa = 0.740$, respectively), and none of the tibial components appeared loose. Therefore, no inter-rater agreement could be determined.

To assess agreement of a qualitative variable, using κ coefficient is not always appropriate and is not reliable in some circumstances. First, κ depends on the prevalence in each category. Second, it also depends on the number of categories [2–6]. When a variable with more than two categories or an ordinal scale is used (with three or more ordered categories), the weighted κ would be a good choice for assessing agreement. Finally, when the two raters have unequal marginal distributions of their responses, applying a κ value can easily cause misleading messages [2–6]. Table 1 shows a hypothetical agreement by applying κ (0.43 as moderate) and weighted κ (0.63 as good) that have different values and consequently different interpretations. In this table, the marginal distribution in the first category (grade 1) is different from the other categories, and the number of categories is more than two. Kleeblad and colleagues concluded that MRI with the addition of MAVRIC sequences could be a complementary tool for assessing symptomatic UKA and for quantifying appearances at the bone–component interface. Also, this technique showed good reproducibility for analysis of the bone–component interface after cemented UKA. Our opinion is that this conclusion may be misleading due to limitations of applying simple κ to assess reproducibility or agreement.

Because of the limitations of using κ to assess reliability [2–6], any conclusion in reliability analyses needs to be supported by the methodological and statistical issues mentioned above.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11420-019-09672-4>) contains supplementary material, which is available to authorized users.

M. Naderi, MSc
Department of Operating Room, School of Paramedical,
Kermanshah University of Medical Sciences,
Kermanshah, Iran

S. Sabour, MD, PhD (✉)
Department of Clinical Epidemiology, School of Public Health,
Shahid Beheshti University of Medical Sciences,
Tehran, Iran
e-mail: s.sabour@sbmu.ac.ir

S. Sabour, MD, PhD
Safety Promotion and Injury Prevention Research Center, Shahid
Beheshti University of Medical Sciences,
Tehran, Iran

Table 1 κ and weighted κ values for calculating agreement between two radiologists for more than two categories and depending on prevalence

		Radiologist 1			Sum
Radiologist 2	Grade	1	2	3	
	1	60	20	1	81
	2	2	12	4	18
	3	3	11	11	25
Sum	3	65	43	16	124
		Estimate	Lower	Upper	
κ		0.43	0.31	0.55	
Weighted κ		0.63	0.52	0.74	

Compliance with Ethical Standards

Conflict of Interest: Mehdi Naderi, MSc, and Siamak Sabour, MD, PhD, declare that they have no conflicts of interest.

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. Kleeblad LJ, Zuiderbaan HA, Burge AJ, Amirtharaj MJ, Potter HG, Pearle AD. MRI findings at the bone-component interface in symptomatic unicompartmental knee arthroplasty and the relationship to radiographic findings. *HSS J.* 2018;14(3):286–293.
2. Naderi M, Sabour S. Reproducibility of diagnostic criteria associated with atypical breast cytology: a methodological issue. [Letter to the editor.] *Cytopathology.* 2018;29(4):396.
3. Sabour S. Reliability of automatic vibratory equipment for ultrasonic strain measurement of the median nerve: common mistake. [Letter to the editor.]. *Ultrasound Med Biol.* 2015;41(4):1119–1120.
4. Sabour S. Reliability assurance of EML4-ALK rearrangement detection in non-small cell lung cancer: a methodological and statistical issue. [Letter to the editor.]. *J Thorac Oncol.* 2016;11(7):e92–93.
5. Sabour S, Rutten A, van der Schouw YT, et al. Inter-scan reproducibility of coronary calcium measurement using multi detector-row computed tomography (MDCT). *Eur J Epidemiol.* 2007;22(4):235–243.
6. Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics.* 3rd ed. New York, NY: Jones & Bartlett; 2012.