RESPONSE TO "INTER-RATER RELIABILITY IN THE CLINICAL EVALUATION OF MYOFASCIAL TRIGGER POINTS IN THREE ANKLE MUSCLES"

To the Editor:

I was interested to read the paper by Sanz et al.[1] The purpose of the study was to evaluate interrater reliability in the diagnosis of myofascial trigger points in the tibialis anterior, peroneus brevis, and extensor digitorum longus muscles.[1] Three physical therapists with clinical experience in myofascial pain functioned as raters randomly and bilaterally evaluated the ankles of 40 participants in the Madrid public health care system. The absence or presence of myofascial trigger points, nodules in taut bands, patterns of referred pain, local twitch response, and jump sign were evaluated.

They used κ statistics as a measure of inter-rater reliability.[1] Based on their results, pairwise inter-rater agreement and κ-value concordance were as follows: presence or absence of trigger points (55%-85%; κ = 0.12-0.60), palpable nodules in taut bands (63%-90%; κ = 0.24-0.60), referred pain (63%-85%; κ = 0.20-0.54), and jump sign (62%-89%; κ = 0.15-0.72) in the 3 studied muscles. The local twitch response could only be evaluated in the tibialis anterior (43%-70%; κ = 0.05-0.21), and evaluation was not possible for the other muscles.[1]

I propose that using kappa is a mistake in reliability analysis for qualitative outcomes. Two important weaknesses of κ statistics to assess reproducibility are as follows. It depends upon the prevalence in each category, which means it can be possible to have a different κ value with the same percentage for both concordant and discordant cells.

Table 1 shows that in both (a) and (b) situations the prevalence of concordant cells are 80% and discordant cells are 20%; however, we get a different κ value (0.38 and 0.60, respectively). More importantly, κ value depends upon the number of categories.[2-7] They had 3 physical therapists as raters. Therefore, using weighted κ in such situations should be considered.[2-7]

They concluded that inter-rater agreement is acceptable and seems to be muscle dependent, especially regarding the depth of the muscle. Such a conclusion should be supported by the aforementioned statistical and methodological issues.

Siamak Sabour, MD, MSc, DSc, PhD
Safety Promotion and Injury Prevention Research Center, Department of Clinical Epidemiology, Shahid Beheshti University of Medical Sciences, Tehran, Iran
E-mail address: s.sabour@sbmu.ac.ir

REFERENCES

1. Sanz DR, Lobo CC, López DL, Morales CR, Marín CS, Corbalán IS. Interrater reliability in the clinical evaluation of myofascial trigger points in three ankle muscles. *J Manipulative Physiol Ther.* 2016;39(9):623-634.
2. Szklo M, Nieto FJ. *Epidemiology Beyond the Basics.* 2nd ed. New York, NY: Jones and Bartlett Publisher; 2007.
3. Sabour S. Reliability of lumbar movement dysfunction tests for chronic low back pain patients; methodological concerns to avoid misinterpretation. *Man Ther.* 2016;26:e4.

Table 1. *Comparison of 2 Observers' Diagnoses With Different Prevalence in the 2 Categories*

Situation		Observer 1		Total
		Positive	Negative	
Situation a				
Observer 2	Positive	70	10	80
	Negative	10	10	20
	Total	80	20	100
κ = 0.38				
Situation b				
Observer 2	Positive	40	10	50
	Negative	10	40	50
	Total	50	50	100
κ = 0.60				

4. Sabour S. Reliability Assurance of EML4-ALK rearrangement detection in non-small cell lung cancer: a methodological and statistical issue. *J Thorac Oncol.* 2016;11(7):e92-e93.
5. Sabour S, Ghassemi F. Submaximal step tests to estimate maximal oxygen uptake in healthy adults: methodological issues about validity and reliability. *Sports Med.* 2016;46(9): 1381-1382.
6. Sabour S, Farzaneh F, Peymani P. Evaluation of the sensitivity and reliability of primary rainbow trout hepatocyte vitellogenin expression as a screening assay for estrogen mimics: methodological issues. *Aquat Toxicol.* 2015;164:175-176.
7. Sabour S. Reliability of the ASA physical status scale in clinical practice: methodological issues. *Br J Anaesth.* 2015;114(1): 162-163.

RESPONSE TO LETTER TO EDITOR: "INTER-RATER RELIABILITY IN THE CLINICAL EVALUATION OF MYOFASCIAL TRIGGER POINTS IN THREE ANKLE MUSCLES"

To the Editor:

The main aim of our study[1] was to investigate the inter-rater reliability to diagnose the presence of myofascial trigger points (MTrPs) and to analyze each criterion proposed by Bron et al[2] in 3 lower limb muscles (tibialis anterior, peroneus brevis, and extensor digitorum longus). These were selected owing to their referred pain pattern to the ankle.

To illustrate the complexity of this methodological procedure,[3] physical therapists with clinical experience in myofascial pain management functioned as raters who randomly and bilaterally evaluated the ankles of 40 participants. The absence or presence of MTrPs, nodules in taut bands, patterns of referred pain, local twitch response, and jump sign were assessed. Before beginning the study, both raters and observers followed the training protocol proposed by Bron et al.[2] This training protocol was a program with 8 hours for practicing their skills, the palpation technique, participant positioning, and localization of MTrPs.[2,3] We based our study in the research model of Bron et al[2] to analyze different diagnosis criteria for lower limb muscles in patients with ankle pain. The pairwise percentage of agreement (PA) and pairwise Cohen κ values were calculated for the reliability analysis. The PA is the ratio of the number of concordances, both positives and negatives, to the total ratings.[4,5]

Our results showed the pairwise inter-rater agreement and κ-value concordance for the presence or absence of trigger points (55%-85%; κ = 0.12-0.60), palpable nodules in taut bands (63%-90%; κ = 0.24-0.60), referred pain (63%-85%; κ = 0.20-0.54), and jump sign (62%-89%; κ = 0.15-0.72) in the 3 studied muscles. The local twitch response could only be evaluated in the tibialis anterior (43%-70%; κ = 0.05-0.21) because it was not possible to measure in the other muscles.[1]

Regarding statistical analysis, we support our methodological development according to Landis and Koch,[6] who established different ranges of κ values: poor (0.00), slight (0.001/0.20), fair (0.21/0.40), moderate (0.41/0.60), good (0.61/0.80), and very good (0.81/1.00) reliability.[1] In our study, we considered a PA value ≥70% to be indicative of acceptable inter-rater reliability, given that a PA value ≥70% leads to a κ ≥0.40 as proposed by Bron et al.[2] Therefore, PA values were prioritized over κ values, as in other studies, because the κ statistic is probably inappropriate for studies with an irregular distribution of positive and negative findings.[1] This was the main reason to protect the acquired data with a high-priority pairwise PA over κ value as we reported in the manuscript.

Our conclusions determined that inter-rater agreement was acceptable and seemed to be muscle dependent, especially regarding the muscle depth. Such a conclusion was supported by both statistics, PA and κ value, to enforce these conclusions, not only based on the obtained κ value.[1] The palpation area of the peroneus brevis muscle is located under the peroneus longus tendon, which may be considered deeper than the tibialis anterior and extensor digitorum longus.[3] Considering these limitations, further studies are necessary to improve the inter-rater reliability knowledge of myofascial pain.

David Rodríguez-Sanz, PhD, MSc, PT, DP
Department of Nursing, Physical Therapy and Podiatry
Universidad Complutense de Madrid, Madrid, Spain

Carlos Romero-Morales, PhD, MSc, PT
Carlos Sosa-Marín, MSc, PT
Faculty of Sport Sciences, Universidad Europea de Madrid
Villaviciosa de Odón, Madrid, Spain

Irene Sanz-Corbalán, PhD, MSc, DP
Department of Nursing, Physical Therapy and Podiatry, Universidad Complutense de Madrid, Madrid, Spain

César Calvo-Lobo, PhD, MSc, PT
Nursing and Physical Therapy Department
Faculty of Health Sciences, University of León
Ponferrada, León, Spain
E-mail address: *cecalvo19@hotmail.com*

Daniel López-López, PhD, MSc, DP
Research, Health and Podiatry Unit, Department of Health Sciences, Faculty of Nursing and Podiatry, Universidade da Coruña, Ferrol, Spain