Article Type: Letter to the Editor

Corresponding Author Email ID: emilycf@med.umich.edu

Authors' reply to the letter-to the-editor by Sabour

Author Manuscript

We appreciate the response to our article (Foxen-Craft et al., 2018) by Pordanjani, Iranpour, & Sabour for opening the discussion about use of statistical techniques in psychometric validation. Overall, the authors' critiques regarding the limits of the kappa statistic are fair, but there is also justification for its use in this study.

In our research, we aimed to explore the reliability and validity of a body map tool for pain location in children undergoing orthopedic surgeries. We used percent agreement and kappa statistics to examine intrarater agreement, and Spearman's rho to examine descriptive and associative validity. Specifically, participants completed a body map by checking boxes where they had experienced pain, and this was coded into a binary variable (yes/no). Kappa statistics were then used to test the agreement of youths' identification of different pain location sites at two time points.

The editorial letter writers aimed to explore the appropriateness of different statistical techniques of assessing reliability, suggesting the use of intra-class correlation coefficient or Bland Altman Plot instead of kappa statistics, consistent with their previous letters to the editor in other journals (Sabour, 2015; Sabour, 2016). Main concerns regarded dependence of the kappa statistic on prevalence in each category and the number of categories.

Though aspects of the authors' letter offered reasonable comments, the authors' concerns may not be entirely relevant in the current research. First, in our study, reliability was assessed regarding binary data (i.e. children indicated yes/no to different body parts), and kappa is recommended for binary data in contrast to continuous data (Mandrekar, 2011). In contrast, Sabour's previous critiques in favor of intra-class correlation coefficient regarded continuous data (Sabour, 2015; Sabour, 2016). It is possible for an intra-class correlation coefficient to be

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ejp.1345

This article is protected by copyright. All rights reserved

calculated with a mixed model if the data are assumed to be numerical, but in our analyses, the children's ratings were binary, not numerical, making kappa the appropriate statistic.

Second, the authors argue that kappa is influenced by prevalence. Though not mentioned by the authors, it may also be influenced by bias (Sim and Wright, 2005). Generally, kappa is proportional to prevalence and inversely related to bias. We now present our kappa statistics and percent agreement in accordance with guidelines McHugh, 2012, as well as Sim and Wright, 2005 in Table 1 (previously Table 2 in the original article).

We note that overall, our prevalence rates were high and bias was low, indicating that kappa was accurate and may have actually been conservatively estimated. However, there seemed to be higher variability in responses among the back regions with lower prevalence, and therefore those kappa statistics should be interpreted with caution. It is very possible that as expected, children's pain did truly change in that area related to their back diagnosis.

In summary, it is possible that future body maps will incorporate use of continuous pain ratings of different pain location sites, in which case reliability analysis using intra-class correlation coefficient would be important to examining reliability. Continued exploration of appropriate matching of statistical techniques to data and study questions is certainly encouraged and may lead to improved standards in clinical research.

Correseponding Author:

Emily Foxen-Craft, Ph.D.[1]

Aleda M. Thompson, M.S.[2]

Eric L. Scott, Ph.D.[1,2]

Kristin A. Kullgren, Ph.D.[1]

Reilly Philliben[3)]

Caroline Hyman[3]

Author Manuscript

This article is protected by copyright. All rights reserved

Marianna Dorta[3]

Anne Murphy, M.D.[4]

Terri Voepel-Lewish, Ph.D., R.N.[2]

[1]Division of Pediatric Psychology, Department of Pediatrics, CS Mott Children's Hospital, University of Michigan, Ann Arbor, MI, USA

[2]Department of Anesthesiology, University of Michigan, Ann Arbor, MI, USA

[3]University of Michigan, Ann Arbor, MI, USA

[4]Department of Pediatrics, CS Mott Children's Hospital, University of Michigan

Foxen-Craft, E., Scott, E. L., Kullgren, K. A., Philliben, R., Hyman, C., Dorta, M., ... & Voepel-Lewis, T. (2018). Pain location and widespread pain in youth with orthopaedic conditions: Exploration of the reliability and validity of a body map. *European Journal of Pain*.

Mandrekar, J. N. (2011). Measures of interrater agreement. *Journal of Thoracic Oncology*, *6*(1), 6-7.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica*, *22*(3), 276-282.

Pordanjani, S.R., Iranpour, S., Sabour, S. (2019). Pain location and widespread pain in youth with orthopedic conditions: Methodological and statistical issues on reliability analysis. *EurJPain,* this issue.

Sabour, S. (2015). Reliability of automatic vibratory equipment for ultrasonic strain measurement of the median nerve: common mistake. *Ultrasound in Medicine and Biology*, *41*(4), 1119-1120.

Sabour, S. (2016). Adherence to guidelines strongly improves reproducibility of brachial artery flow-mediated dilation. Common mistakes and methodological issue. *Atherosclerosis*, *251*, 490.

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy*, *85*(3), 257-268.

Author Manuscript

This article is protected by copyright. All rights reserved

Table 1. Body map agreement by site with prevalence and bias.

Author Manuscript

This article is protected by copyright. All rights reserved

Table 2. Agreement between baseline and follow-up body map reports by site

Body Site	Agreed: Pain absent at both times (n)	Agreed: Pain present at both times (n)	Percent agreement	κ
L Shoulder Girdle	71	7	95.12%	.75**
R Shoulder Girdle	62	9	86.59%	.54**
L Arm Upper	78	1	96.34%	.39**
R Arm Upper	77	1	95.12%	.31*
L Arm Lower	76	2	95.12%	.48**
R Arm Lower	77	2	96.34%	.55**
L hip	72	4	92.68%	.54**
R hip	76	1	93.90%	.26*
L leg upper	74	2	92.68%	.37**
R leg upper	76	4	97.56%	.79**
L leg lower	73	4	93.90%	.58**
R leg lower	74	4	95.12%	.64**
L Knee	73	3	92.68%	.46**
R Knee	71	8	96.34%	.82**
L Jaw	80	1	98.78%	.66**
R Jaw	80	1	98.78%	.66**
Chest	77	0	93.90%	-.03
Abdomen	67	9	92.68%	.71**
Neck	63	12	91.46%	.72**
Back Upper	30	34	78.05%	.56**
Back Middle	20	47	81.71%	.59**
Back Lower	34	30	78.05%	.56**
Head	62	12	90.24%	.69**

Note: L = Left, R = Right, *p < .05, **p < .001

Author Manuscript

This article is protected by copyright. All rights reserved