



Correspondence

Allergy skin testing Methodologic issues on reliability and validity



I was interested to read the article by Shtessel and colleagues¹ published in the January 2018 issue of the *Annals*. Percutaneous allergen skin testing remains an established benchmark for diagnosing atopic disease. The reliability of skin testing depends greatly on the performance of allergen extracts used, methods used, and the presence of antihistamine medications. The authors aimed to determine the differential effect of cetirizine on 2 different concentrations of histamine control solution and 5 common allergens used for percutaneous skin testing. A 2-tailed, matched *t* test on mean wheal and flare measurements for the 2 concentrations of histamine on each day was performed. A Fisher exact test was used to compare the allergen false-negative rate between the 1-mg/mL and 6-mg/mL histamine control reagents. It is crucial to know that neither the aim of the authors nor the statistical methods used were consistent with the title of the article. Actually, applied methods have nothing to do with reliability.^{2–6}

Reliability (precision, repeatability) and validity (accuracy) are 2 different methodologic issues.² False-positive and false-negative rates, as well as sensitivity, specificity, positive predictive value, negative predictive value, and positive and negative likelihood ratios, are among the estimates to assess validity (accuracy) of a diagnostic test and have nothing to do with reliability.^{2–6} Reliability (precision) as a different methodologic issue should be assessed using appropriate tests. For qualitative variables, weighted κ can be applied with caution. Regarding quantitative variables, intraclass correlation coefficient and Bland-Altman plot are among the well-known approaches.^{2–6}

They reported that twice as many individuals had false-negative allergen responses using 6 mg/mL of histamine vs 1 mg/mL as a

positive control, although this finding did not reach statistical significance. Clinical importance and statistical significance are 2 different issues. In reliability analysis, our approach should be based on the individual instead of the global average. Therefore, applying the *t* test and emphasizing statistical significance are among common mistakes in reliability analyses.^{2–6} The authors' conclusion should be supported by the previously mentioned methodologic and statistical issues.

Siamak Sabour, MD, PhD^{*,†}

^{*}Safety Promotion and Injury Prevention Research Center
Shahid Beheshti University of Medical Sciences
Tehran, Iran

[†]Department of Clinical Epidemiology
Shahid Beheshti University of Medical Sciences
Tehran, Iran

s.sabour@sbmu.ac.ir

References

- [1] Shtessel M, Tversky J. Reliability of allergy skin testing. *Ann Allergy Asthma Immunol*. 2018;120:80–83.
- [2] Szklo M, Nieto FJ. *Epidemiology Beyond the Basics*. 2nd ed. New York, NY: Jones & Bartlett Publisher; 2007.
- [3] Sabour S. Reliability of a new modified tear breakup time method: methodological and statistical issues. *Graefes Arch Clin Exp Ophthalmol*. 2016;254:595–596.
- [4] Sabour S, Ghassemi F. Accuracy and reproducibility of the ETDRS visual acuity chart: methodological issues. *Graefes Arch Clin Exp Ophthalmol*. 2016;254:2073–2074.
- [5] Sabour S, Ghassemi F. The validity and reliability of a signal impact assessment tool: statistical issue to avoid misinterpretation. *Pharmacoevid Drug Saf*. 2016;25:1215–1216.
- [6] Sabour S. Reproducibility of dynamic Scheimpflug-based pneumotonometer and its correlation with a dynamic bidirectional pneumotonometer device: methodological issues. *Cornea*. 2015;34:e14–e15.

Disclosures: Author has nothing to disclose.

Authors' response

We appreciate the comments related to our recent publication, "Reliability of allergy skin testing." The author writes, "Reliability (precision, repeatability) and validity (accuracy) are 2 completely different methodologic issues." We agree that other titles, such as "Accuracy of allergy skin testing" may have fit more pedantic

epidemiologic definitions. However, this would not have emphasized the clinical concerns any better. We chose to use the lay term *reliability* as defined in the *Oxford English Dictionary* as "the degree to which the result of a measurement, calculation, or specification can be depended on to be *accurate*."¹

Regardless, we would not want discourse over title semantics to detract from the overarching point of the article, which relates to *clinical reliability*, a term that is not infrequently used by clinicians in lay terms to describe the trustworthiness of allergy skin prick testing.² From a clinical standpoint, correct interpretation

Disclosures: Authors have nothing to disclose.

Funding Sources: Dr Tversky received funding for this study from Lincoln Diagnostics.

<https://doi.org/10.1016/j.ana.2018.01.021>

1081-1206/© 2018 American College of Allergy, Asthma & Immunology. Published by Elsevier Inc. All rights reserved.

