



## How to assess response

Steffen Zitzmann<sup>1</sup> · Christoph Lindner<sup>2</sup>

Received: 29 February 2024 / Accepted: 27 May 2024  
© The Author(s) 2024

The term response has been established as a key concept in psychiatry. It is an indispensable part of the jargon and frequently applied in practice and research. By a patient who showed a response, we mean someone whose symptoms improved to an extent that is practically important. Often, it is implied that this improvement has occurred in response to how the patient was treated, such as by taking a drug. In clinical research, however, the term response often refers to the improvement alone without a clear reference to the specific cause for this improvement. The reason is that we can hardly know how this patient would have improved if they had not been treated or had received a placebo.

To assess whether a patient has responded, their percentage symptom reduction from baseline has to be obtained first. This value can be computed as  $(S_2/S_1 - 1) \cdot 100\%$ , where  $S_1$  and  $S_2$  stand for the patient's symptoms at baseline and after the treatment, respectively [1]. Then, the patient can be classified into a category of improvement. These categories are based on cutoffs, for example, on steps of 25% [2]. However, there is an issue with this procedure, which arises because interviews and questionnaires are often used. Such measures are blurred by noise called measurement error [3, 4]. The amount of error is typically expressed by the measure's reliability, which is defined as the proportion of the variability of the measured score that can be attributed to true differences between patients [5].

To illustrate, consider the Positive and Negative Syndrome Scale (PANSS) [6]—a measure of symptom severity in schizophrenia. Recent research found that the internal consistency of the PANSS (i.e., an estimate of its reliability) ranged up to 0.94 [7]. Before we continue, note that to compute a patient's percentage PANSS reduction from baseline, their total scores need to be corrected by subtracting 30 points [9, 10]. Now, suppose the patient got a corrected

total score of  $S_1^* = 60$  points at baseline and  $S_2^* = 40$  after treatment. Using the equation for computing percentage symptom reduction from baseline, their symptoms seemingly decreased by 33.3%. Thus, if measurement error was ignored, the patient would be classified into the 25 to 50% improvement category, meaning the patient would be described as having improved. However, we may indicate uncertainty due to measurement error by a confidence interval. The interval can be computed as:

$$(S_2^*/S_1^* - 1) \cdot 100\% \pm 1.96 \cdot \sqrt{(1 - \rho) \cdot (1 + S_2^{*2}/S_1^{*2} \cdot \sigma_1^2/\sigma_2^2) \cdot \sigma_2^2/S_1^2} \cdot 100\%$$

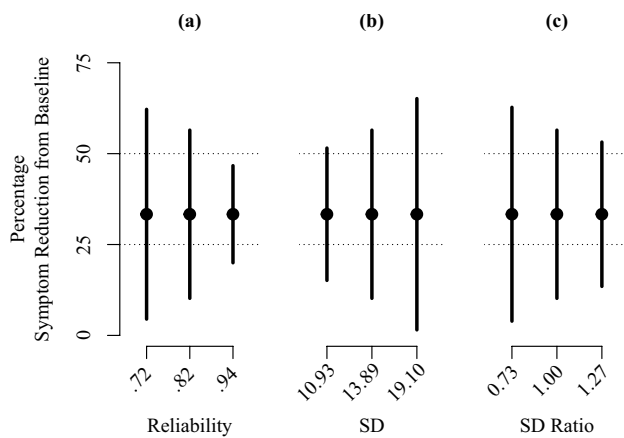
where  $\rho$  is the reliability of the PANSS.  $\sigma_1$  and  $\sigma_2$  are the standard deviations at baseline and after treatment [1, 11]. Consider the patient whose symptoms reduced from 60 to 40 points. If, for example, the reliability of  $\rho = .94$  is used in the calculation of the patient's confidence interval, and the same standard deviation of  $\sigma_1 (= \sigma_2) = 10.93$  is assumed at baseline and after treatment, we yield  $-33.3\% \pm 10.5\%$ . Thus, we cannot be certain that their actual improvement was not 23% instead of 33.3%, for example. More importantly, this interval includes 25%, indicating that this patient could fall into the 0 to 25% category (i.e., no practically important improvement). In other words, it cannot be decided whether this patient belongs to the 0 to 25% or 25 to 50% category.

It should be noted that the confidence interval involves making assumptions about the measure, particularly regarding quality and distributional aspects. Thus, depending on the context, one may use different values for the reliability and the standard deviations in the calculation. To illustrate how they impact the patient's confidence interval, see Fig. 1. Panel (a) shows that when the level of reliability  $\rho$  is increased, the interval will be narrowed. By contrast, increasing the standard deviation  $\sigma_1 (= \sigma_2)$  will increase the width [Panel (b)]. Finally, given unequal standard deviations, a greater ratio  $\sigma_1/\sigma_2$  will lead to a decrease in width [Panel (c)].

✉ Steffen Zitzmann  
steffen.zitzmann@medschool-hamburg.de

<sup>1</sup> Medical School Hamburg, Hamburg, Germany

<sup>2</sup> Max Planck Institute of Psychiatry, Munich, Germany



**Fig. 1** Dependence of the confidence interval on reliability  $\rho$  [Panel (a)], standard deviation  $\sigma_1 (= \sigma_2)$  [Panel (b)], and standard deviation ratio  $\sigma_1/\sigma_2$  [Panel (c)]. *SD* standard deviation

To conclude, the patient whose confidence interval includes a cutoff illustrates the need for a refinement of the assessment of response. Interviews and questionnaires are not error-free measures. As psychiatrists, we should take this message serious and “embrace” the uncertainty. To decide whether a patient has responded, we should not focus solely on their percentage symptom reduction from baseline. In addition, a confidence interval should be placed around this value so that it can be better evaluated into which improvement category the patient falls. If the interval includes a cutoff, a definite categorization cannot be made, and further information has to be gathered.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source,

provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Zitzmann S, Lindner C, Leucht C, Leucht S (2023) A potential issue with PANSS responder analysis. *Schizophr Res* 261:287–290. <https://doi.org/10.1016/j.schres.2023.10.009>
- Leucht S, Davis JM, Engel RR, Kissling W, Kane JM (2009) Definitions of response and remission in schizophrenia: recommendations for their use and their presentation. *Acta Psychiatr Scand* 119:7–14. <https://doi.org/10.1111/j.1600-0447.2008.01308.x>
- Novick MR (1966) The axioms and principal results of classical test theory. *J Math Psychol* 3:1–18. [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)
- Zitzmann S (2023) A cautionary note regarding multilevel factor score estimates from lavaan. *Psych* 5:38–49. <https://doi.org/10.3390/psych5010004>
- Norman G (2014) When i say ... reliability. *Med Educ* 48:946–947. <https://doi.org/10.1111/medu.12511>
- Kay SR, Fiszbein A, Opler LA (1987) The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 13:261–276. <https://doi.org/10.1093/schbul/13.2.261>
- Lin C-H, Lin H-S, Lin S-C, Kuo C-C, Wang F-C, Huang Y-H (2018) Early improvement in PANSS-30, PANSS-8, and PANSS-6 scores predicts ultimate response and remission during acute treatment of schizophrenia. *Acta Psychiatr Scand* 137:98–108. <https://doi.org/10.1111/acps.12849>
- Li L, Ma H, Wang X, Meng E (2021) Validation of Chinese version of positive and negative syndrome scale-6 in clinical setting: a preliminary study. *Psychiatry Clin Psychopharmacol* 31:386–391. <https://doi.org/10.5152/pcp.2021.21060>
- Leucht S, Kissling W, Davis JM (2010) The PANSS should be rescaled. *Schizophr Bull* 36:461–462. <https://doi.org/10.1093/schbul/sbq016>
- Obermeier M, Mayr A, Schennach-Wolff R, Seemüller F, Möller H-J, Riedel M (2010) Should the PANSS be rescaled? *Schizophr Bull* 36:455–460. <https://doi.org/10.1093/schbul/sbp124>
- Zitzmann S (2023) Einzelfallbezogene Veränderungsdiagnostik [Diagnostics of individual change]. In: Dohrenbusch R (ed) *Psychologische Begutachtung*. Springer, Wiesbaden, pp 1–9