

Response to: ‘Correspondence on ‘Statistical review: frequently given comments’ by Waki *et al*

Waki *et al*¹ generally agree with my recommendations in ‘Statistical review: frequently given comments’.² But they question my recommendation to generally report mean and SD rather than median and quartiles, also for non-normally continuous distributions. I am thankful for this opportunity to elaborate this issue more in depth.

The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline checklist does not specify recommendations regarding mean versus median for descriptive data. Waki *et al* refer to the STROBE guidelines as reported in Ref. 3, which comments the guidelines with more details, including the following recommendation on page 822: “We advise authors to summarize continuous variables for each study group by giving the mean and standard deviation, or when the data have an asymmetrical distribution, as is often the case, the median and percentile range (eg, 25th and 75th percentiles).” Similar advice is found in other literature including two other references given in Ref. 1. But I have not seen any argument why mean and SD should not be relevant in this context.

We must distinguish between the two purposes of presenting purely descriptive statistics, on one side, and the assumptions on which our analysis methods rely, on the other side. For descriptive statistics, the choice between mean and median may rely on the aim. For example, consider the length of stay in hospital for patients with a certain diagnosis. In order to estimate cost or need for personnel, the mean is the relevant quantity. On the other hand, to a single patient, the median may be more interesting.⁴ Some researchers claim that it is generally wrong to report mean and SD unless data are normally distributed. I cannot see any good arguments for that view. The mean and SD are well defined for all kinds of distributions, see Ref. 4.

Another point is that the median and quartiles can be directly unsuited for ordinal categorical data, especially when there are few categories. But the mean can have an interpretation in terms of excess probability for ordinal data, and can be a useful measure in some contexts.⁵ As a statistical reviewer, I regularly see authors who are not aware of this fact.

I do not assert that mean (SD) is always preferable over median (quartiles) for descriptive purposes. For example, in survival analysis, there are usually individuals who have not experienced the event (such as death). Then, the descriptive mean is undefined, but some percentiles, possibly the median, can be computed.

Waki *et al* write¹: “Moreover, if a meta-analysis is performed including papers that present significantly skewed data as the mean and SD, the results of the meta-analysis may be distorted. (Ref Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration, 2011).”

I do not have access to the Cochrane handbook version 5.1.0 to which Waki *et al* refer.¹ But I have access to version 6.2 (2021) as e-book.⁶ I have searched the book electronically and I have not found any support for preferring median over mean for summarising data in single studies. On the contrary, the authors state in Section 6.6.1: “Difficulties will be encountered if studies have summarized their results using medians.” However, in a meta-analysis, it can be relevant to report the median effect estimate across studies. But a median estimate across studies is a completely different issue than a summary statistic for raw data.

One of the favourable properties of the mean and SD is the possibility to calculate the overall mean and SD based on aggregated data consisting of sample size, mean and SD. The preference over mean and SD as input to meta-analyses is closely related to this property. The median and quartiles do not have such properties.

Finally, I use the opportunity to clarify some misunderstanding in the correspondence¹:

“We agree that the median is very close to the mean in data with sufficiently large sample sizes according to the central limit theorem.” This is incorrect. The difference between the mean and the median does not vary systematically with sample size. And the central limit theorem states that the mean value is approximately normally distributed in large sample sizes, regardless of the distribution of the data itself.

“However, for summary statistics of continuous data with an asymmetrical distribution, the median has been found to reflect the distribution more accurately than the mean.” This is an illogical statement. The mean and median quantify different properties of the distribution if data are not symmetrically distributed.

“Moreover, many studies on rheumatic diseases have reported significant results despite sample sizes being too small to statistically satisfy the condition of the central limit theorem because of their rarity.” Indeed, many analysis methods are based on the assumption that the data are, at least approximately, normally distributed. If data deviate substantially from normality, one must use methods accounting for this. For example, instead of a two-sample t-test, one can use a non-parametric Wilcoxon test or a bootstrapped t-test.²

“In conclusion, we believe that the appropriateness of the mean or median for nonparametric continuous variables should be considered by including the central limit theorem.” A variable is not parametric or non-parametric. A method can be parametric, such as the t-test, or non-parametric, such as the Wilcoxon test. And the central limit theorem gives the theoretic basis for some statistical methods, but is not relevant for choosing the mean (and SD) versus the median (and quartiles) for descriptive statistics.

As already stated, I am thankful for this opportunity to elaborate this issue more in depth. My viewpoint on this matter remains: “Mean (SD) is also relevant for non-normally distributed data”.²

Stian Lydersen

Regional Centre for Child and Youth Mental Health and Child Welfare, Department of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence to Professor Stian Lydersen, Regional Centre for Child and Youth Mental Health and Child Welfare, Norwegian University of Science and technology, Trondheim, Norway; stian.lydersen@ntnu.no

Handling editor Josef S Smolen

Contributors Stian Lydersen is the sole author and is responsible for the whole manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

© Author(s) (or their employer(s)) 2023. No commercial re-use. See rights and permissions. Published by BMJ.



To cite Lydersen S. *Ann Rheum Dis* 2023;**82**:e129.

Received 15 March 2021
Revised 16 March 2021
Accepted 16 March 2021
Published Online First 24 March 2021



► <http://dx.doi.org/10.1136/annrheumdis-2021-220332>

Ann Rheum Dis 2023;**82**:e129. doi:10.1136/annrheumdis-2021-220336

ORCID iD

Stian Lydersen <http://orcid.org/0000-0001-6613-8596>

REFERENCES

- 1 Waki D, Banno M, Fujiwara T. Correspondence on 'Statistical review: frequently given comments'. *Ann Rheum Dis* 2023;**82**:e128.
- 2 Lydersen S. Statistical review: frequently given comments. *Ann Rheum Dis* 2015;**74**:323–5.
- 3 Vandembroucke JP, von Elm E, Altman DG, *et al*. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Epidemiology* 2007;**18**:805–35.
- 4 Lydersen S. Mean and standard deviation or median and quartiles? *Journal of the Norwegian medical Association* 2020;140.
- 5 Lydersen S. How to summarise ordinal data. *Tidsskr Nor Laegeforen* 2020;140. doi:10.4045/tidsskr.20.0033
- 6 Cochrane Collaboration. *Cochrane Handbook for systematic reviews of interventions, version 6.2*. 2nd edn. Hoboken, NJ: Wiley-Blackwell, 2021.