Reanalyzing Castner et al. Confirms Most Conclusions, Reveals Additional Findings, and Detects Some Statistical Errors

Olivia C. Robertson[1], Erik Parker[2], Beate Henschel[2], Anna L. M. Macagno[2], Daniel E. Kpormegbey[2], Deependra K. Thapa[2], David B. Allison[1,2]

1 Baylor College of Medicine, Houston, TX, USA
2 Indiana University, Bloomington, IN, USA
⋆*Corresponding author:* David B. Allison (David.Allison@bcm.edu)

Abstract

Patterns of findings generally held with discrepancies in *p*-values and additional significant findings revealed after reanalyzing with correction of suboptimal statistical practices

Received Date
Accepted Date
Published Date

Target article

Castner, S. A., Gupta, S., Wang, D., Moreno, A. J., Park, C., Chen, C., Poon, Y., Groen, A., Greenberg, K., David, N., Boone, T., Baxter, M. G., Williams, G. V., & Dubal, D. B. (2023). Longevity factor klotho enhances cognition in aged nonhuman primates. *Nature aging*, *3*(8), 931–937. https://doi.org/10.1038/s43587-023-00441-x

1 Goal

The goal of this robustness report was to evaluate the computational reproducibility and statistical methods used in Castner et al.'s study, "Longevity factor klotho enhances cognition in aged nonhuman primates,"[1] and assess whether better substantiated analytic approaches would yield different conclusions.

2 Methods

We first attempted to fully computationally reproduce the analyses and figures using the original methods that were reported in the paper. The methods consisted of linear mixed models as well as two-tailed and one-tailed t-tests. However, we required additional

Formatted: Indent: First line: 0"

Deleted:

Deleted:

information from the authors to reproduce the paper. We additionally re-analyzed any part of the paper's analyses, that in our view, were sub-optimal or constitute statistical errors, applying what we believe to be best practices.

3 Results

The original paper reported statistically significant findings that only low dose of klotho (KL) administration enhanced cognitive performance, while medium and high doses did not. Broadly, we were able to reproduce the analyses for each test and figure in the paper and confirmed our work by matching *p*-values. However, there were exceptions as small differences in *p*-values at the hundredth decimal point or smaller were found (e.g., see Figure 1 c, g, and i in the technical report).

In our re-analysis, we uncovered statistically significant between-group differences that were not tested in the original study using post-hoc comparisons for the effect of KL doses on cognitive tasks. Specifically, we reanalyzed the data using linear mixed models that treated klotho dose (10, 20, and 30 μg kg^{-1}) as a within-subject factor, in contrast to the original analysis, which ran separate models per dose (which inflates the type 1 error rate) (See Figure 2 c, d, g, and h) and then performed post-hoc comparison tests. The original paper reported that only the 10 μg kg^{-1} dose produced higher cognition scores than the vehicle on either the normal (NML) and high memory load (HML) tasks. In our post-hoc comparisons (which tested differences between doses) we additionally found that the 10 μg kg^{-1} dose group outperformed the 30 μg kg^{-1} group on the HML task and outperformed the 20 and 30 μg kg^{-1} group on the NML task. These findings reinforce the original conclusion regarding the superiority of the 10 μg kg^{-1} dose over the vehicle and further demonstrate its advantage over the higher doses. Otherwise, the pattern of findings held, often with our analysis yielding smaller *p*-values suggesting greater efficiency in modeling strategy.

We also reanalyzed a t-test comparison where hippocampal samples belonging to the same animal were treated as independent. We used a mixed model including a random intercept for animal identifier to account for the non-independence (Please see Figure 1 c). The *p* value from our analysis matched the authors'.

Finally, we re-analyzed one-tailed t-tests as two-tailed tests to compare KL serum levels between the 10 and 20 μg kg^{-1} and between the 10 and 30 μg kg^{-1} doses. We did so given the absence of a strong a priori directional hypothesis, allowing for the possibility that higher doses could perform better or worse than the 10 μg/kg dose. One-tailed t-tests can upwardly bias type 1 error rates if investigators do not unequivocally and scrupulously eschew the possibility of declaring a statistically significant result in the opposite predicted direction [2]. The authors mentioned that the one-tailed t-test was suggested by a reviewer and was not part of the authors' original analytic plan. Our re-analysis produced smaller *p* values using a mixed model with post hoc comparisons (See Figure 1 i).

4 Conclusion

While the main findings of the original study were generally reproducible, our reanalysis suggests that the original authors' statistical approach was suboptimal and biased toward the null. Moreover, reviewer-influenced decisions may compromise the integrity of published results. This reanalysis supports the authors' core conclusions while revealing that more rigorous methods would have strengthened the evidence.

Deleted: detected
Deleted: s
Deleted: both
Deleted: groups
Deleted:
Deleted: but
Deleted: ed
Deleted: our
Deleted: O
Formatted: Font: Italic
Deleted: a
Deleted:
Deleted: a
Deleted:
Deleted: used
Deleted: test
Deleted: isons
Deleted: between
Deleted: dose
Deleted: for
Deleted: This adjustment was made

Acknowledgments and Disclosures

We thank the authors for their collegiality in sharing their data with us to make this reproducible report possible.

Reproducibility

We were able to computationally reproduce the original analysis and results with assistance and additional information from the authors not reported in the original study.

Code and Data Availability The code for this reproducibility report is available here (https://osf.io/49736/). The data required to reproduce the original and secondary analyses conducted for this report is available upon request from the original study authors.

Author Contributions

The authors confirm contribution to the paper as follows: Conceptualization: Allison, Parker; Analysis and interpretation of results: Parker, Henschel, Macagno; Manuscript draft preparation: Robertson, Parker; Reviewed the manuscript critically for important intellectual content; Robertson, Parker, Henschel, Macagno, Kpormegbey, Thapa, Allison; All Authors approved the final version of the manuscript.

Funding
DBA (National Institutes of Health: R25DK099080, R25HL124208, R25GM141507; Nathan Shock Center: P30AG050886, U24AG056053)

Conflicts of Interest Collectively, the authors and their institutions have received payments for consultation, grants, contracts, in-kind donations, and contributions from multiple for-profit and not-for-profit entities interested in statistical design and analysis of experiments but not directly related to the research questions addressed in the paper in question.

References

1 S. A. Castner *et al.* Longevity factor klotho enhances cognition in aged nonhuman primates. *Nature Aging* **3**, 931-937 (2023). https://doi.org/10.1038/s43587-023-00441-x

2 A. W. Brown *et al.* Childhood obesity intervention studies: A narrative review and guide for investigators, authors, editors, reviewers, journalists, and readers to guard against exaggerated effectiveness claims. *Obesity Reviews* **20**, 1523-1541 (2019). https://doi.org/10.1111/obr.12923