



ESTIMATING THE RELIABILITY & ROBUSTNESS OF RESEARCH

ERROR REPORT

Wessel, J. R. (2018). Prepotent motor activity and inhibitory control demands in different variants of the go/no-go paradigm. *Psychophysiology*, 55(3), e12871.
<https://doi.org/10.1111/psyp.12871>

reviewed by

Russ Poldrack, Stanford University

Mar 27, 2024

I. METHODS, MEASUREMENT, AND DESIGN

1. Design No errors found ▾

Are there errors in the conceptual design of the study? E.g., flawed randomisation technique

Note that this ERROR review only examined the literature analysis portion of the study. I did not identify any errors in the design of the literature analysis.

Time spent: 10 minutes

2. Measurement No errors found ▾

Are there any measures, techniques, or devices that were incorrectly applied or inappropriate for the specific task described in the paper?

Time spent: 10 minutes

3. Preregistration Consistency Not applicable ▾

Are there substantial deviations from the preregistration, particularly undisclosed ones?

Time spent: minutes/hours

4. Sampling Not applicable ▾

Is there an error in the sampling strategy? Is the power analysis reproducible? Does the model used for the power analysis match the model in the substantive analyses? Were separate power analyses conducted for all primary analyses?

Time spent: minutes/hours

5. Other Aspects Related to Methods and Measures Indeterminable ▾

The Pubmed query used for the literature search is not incorrect per se, but may not have been sufficient to identify all possibly relevant papers. Note that the following statements are based on the Pubmed query expansion engine as of March 2024; it is likely that the query expansion has changed since the original research was completed, and it is thus impossible to know exactly how the results might have differed at that point.

The authors report using the following Pubmed query: "inhibition AND human AND go-nogo task." When terms are included in a search without quotation marks around the individual terms (which I am assuming was the case here), then the query is expanded to

find related terms. As of today, Pubmed expands the authors' original query as follows (obtained by downloading the search history from the Pubmed Advanced Search Builder at <https://pubmed.ncbi.nlm.nih.gov/advanced/>):

```
("inhibit"[All Fields] OR "inhibitable"[All Fields] OR
"inhibite"[All Fields] OR "inhibited"[All Fields] OR
"inhibites"[All Fields] OR "inhibiting"[All Fields] OR
"inhibition, psychological"[MeSH Terms] OR ("inhibition"[All
Fields] AND "psychological"[All Fields]) OR "psychological
inhibition"[All Fields] OR "inhibition"[All Fields] OR
"inhibitions"[All Fields] OR "inhibitive"[All Fields] OR
"inhibits"[All Fields]) AND ("human s"[All Fields] OR
"humans"[MeSH Terms] OR "humans"[All Fields] OR "human"[All
Fields]) AND ("go-nogo"[All Fields] AND "task"[All Fields])
```

Because of the manual filtering that was performed by the authors, I am not concerned about the potential for irrelevant papers to be identified by this expanded search. However, there is a concern that some relevant papers may have been missed; for example, if the authors used the term “paradigm” instead of “task”. As of today, the authors' query returns 729 matching papers in Pubmed. As an example of how the authors might have expanded their query to identify additional publications, I tested the following query:

```
inhibition AND human AND (go-nogo task OR "go-nogo paradigm" OR
"go/no-go paradigm")
```

This search returned 826 results, more than 10% more than the original search. I did not examine the additional papers found by this search to determine how many of them might have survived the authors' filters, but this example does show that the recall of the search could have been increased.

I am not concerned about the effects of these different queries on the results of the analysis, as it seems highly unlikely that the additional papers would have differed systematically in a way that would have changed the result.

Time spent: 20 minutes

II. DATA, CODE, AND STATISTICAL ANALYSES

1. Code Functionality Not applicable ▾

Does the provided code run without the need to make any adjustments and without errors? If not, what steps were needed to get it to run (if it was eventually possible)?

Time spent: *minutes/hours*

2. Computational Reproducibility of Reported Statistics Not applicable ▾

Is there a clear traceability of reported stats to code? Does the code output match what's reported in the paper? Are all reported statistics findable within the analysis code?

Time spent: *minutes/hours*

3. Data Processing Errors Errors found ▾

Are there substantive errors during the preparation or cleaning of data (e.g. duplication of rows during a merge) prior to substantive analyses and hypothesis tests?

I checked the coding of the P(nogo) and minimum/maximum SOA values that were provided in the shared data file against the original publications. Because of the large number of papers, I selected 10% of the papers (24 papers) for checking. I first started by selecting an additional 5 papers at random and comparing the coded values to the publication, for practice. For the re-coding of the 24 papers I was blinded to the original coded values. I then compared my recoded values against the original values. Details are provided in the attached code, and my recoded values are included in a spreadsheet titled "recoded_values.xlsx".

For P(nogo), I identified errors in 4/25 of the papers checked, with a maximum absolute difference of 10%.

For SOA values I identified differences in the original coded values from my recoded values for 9 of the 24 papers. Further examination showed that in 4 of these cases, it was not possible to tell which of the values was correct. Coding errors were identified in the remaining five papers, with a maximum absolute difference of 4000 ms for minimum SOA and 1300 ms for maximum SOA.

Time spent: *2.5 hours*

4. Model Misspecification Not applicable ▾

Are there any consequential issues with the assumptions or the form of a statistical model (e.g., overfitting, wrong distribution assumption) used to describe data?

Time spent: *minutes/hours*

5. Erroneous/Impossible/Inconsistent Statistical Reporting Not applicable ▾

Are there inconsistencies between test statistics, degrees of freedom, and p-values? Are there implausible degrees of freedom between compared SEM models? Are there point estimates outside the confidence interval bounds?

Time spent: *minutes/hours*

6. Other Aspects Related to Data or Code Errors found ▾

I attempted to recreate the Figures 1 and 2 from the provided data; see the computational notebook provided with my review. I was able to qualitatively reproduce each of the figures; there were small differences in the details due to differences in histogram binning, which could not be exactly reproduced due to the lack of original code.

I also compared the results of my reanalysis with the statements made in the paper. I did identify two minor errors in the text. First, on page 8 (Section 3.1), the authors state that "The trial duration parameter had a wider range, spanning from 650 ms to 17,550 ms between successive stimuli." The min duration reported in the text matches the min SOA in the data (variable "SOA (ms)"). The max duration reported in the text doesn't seem to match the max SOA (variable "SOA (ms, max)"), which is 17,500 ms; it seems that the 50 ms buffer was added to these values, even though that was not a study with a fixed trial duration, which was the intended use case for adding 50 ms. This issue also appears in the statement on Page 9, "The mode for the maximum duration between two trials was 2,050 ms, which is 550 ms longer than the fast-paced condition we used in the current study." The mode of the maximum SOA values computed from the shared data (variable "SOA (ms, max)") is 2000; it appears that the mode was computed after adding 50 ms to some studies. These are minor errors that do not change the interpretation of the results.

Time spent: *3 hours*

III. CLAIMS, PRESENTATION, AND INTERPRETATION

1. Interpretation Issues No errors found ▾

Throughout the entire paper, is there an incorrect substantive interpretation of data or statistical tests, causal inference issues, etc.?

Note that my analyses only focused on the Literature Analysis portion of the paper. I did not identify any incorrect substantive interpretations of data, or causal inference issues.

Time spent: 15 minutes

2. Overclaiming Generalisability No errors found ▾

Does the paper overclaim the generalisability of the findings with regards to stimuli, situations, populations, etc.? Is there hyping or overselling of the importance or relevance of findings?

Time spent: 5 minutes

3. Citation Accuracy Didn't check ▾

Are there misrepresentations of substantive claims by cited sources? Inaccurate direct quotes? Incorrectly cited or interpreted estimates? Citations of retracted papers?

Time spent: minutes/hours

4. Other Aspects Related to Interpretation No errors found ▾

Time spent: minutes/hours