

Review protocol: Link Functions and Generalized Linear Models in Interaction Testing within Psychological Research

Enrico Toffalini, Margherita Calderan, Tommaso Feraco, Filippo Gambarota

2025-03-26

1. Objectives and rationale

The planned review will analyze a sample of recently published psychological research articles targeting a selection of reputable journals across diverse subfields. The primary objective is to describe and quantify how frequently interaction effects are tested in empirical research articles, how often these articles explicitly set non-identity link functions (e.g., by appropriately specifying generalized linear models), and how frequently the use of an identity link function may be problematic, given the type of variables analyzed. Additionally, we aim to determine how often statistically significant interaction effects are detected, particularly when identity link functions are employed. Finally, we aim to document the most commonly analyzed response variables in studies testing interactions (e.g., accuracies, sum scores).

The rationale for this review is that psychology researchers frequently test interaction effects, typically to identify moderators of known main effects, but also because testing interactions often becomes a standard routine following the analysis of main effects. Classical linear models or equivalent methods (e.g., ANOVA) are typically employed, and by default, these methods use identity link functions. However, many psychological response variables are inherently non-normally distributed and present natural bounds (e.g., accuracies, response times, sum or average scores, error counts, proportions). Psychometric variables, more generally, have been systematically shown to deviate from normal distributions (Micceri, 1989). Failing to account for the appropriate distribution can lead to false positives in interaction effects, biased estimates, or even reversed effects (Domingue et al., 2022; Liddell & Kruschke, 2018). In particular, we argue that the core issue lies in the use of inappropriate link functions, as the link function transforms equal intervals into unequal ones, and *vice versa*. Consequently, using the wrong link function places researchers at risk of identifying statistically significant interactions that do not reflect the true data-generating processes. Importantly, this issue goes beyond just choosing a model that matches the distribution of the response variable.

For example, binomial regression is appropriate for Bernoulli outcomes, but different choices of link function, such as *logit* versus *probit*, can still yield different results, underscoring the importance of specifying the link function explicitly and appropriately.

We believe it is important to quantify how often this methodological oversight occurs within psychological research to raise awareness and caution in the field. The current review does not aim for an exhaustive systematic review covering the entire psychological literature. Rather, the goal is to provide a general overview and preliminary estimates regarding the prevalence and magnitude of this potentially widespread issue. As a subsequent step after the review, we plan to conduct simulation studies to illustrate how neglecting appropriate link functions when testing interactions can frequently lead to spurious findings in realistic research scenarios, and what diagnostics can be performed in typical cases.

2. Methods

2.1 Sample

A sample of recently published articles will be analyzed, drawn from five reputable journals representing diverse research areas within psychology. Specifically, we will select one journal from each of the following categories: general psychology, experimental psychology, social psychology, clinical psychology, and developmental psychology. From each journal, we will randomly select articles to include 40 that test interaction effects on empirical data, resulting in a total sample of 200 articles. A larger number of articles will be screened to exclude non-eligible ones (see criteria below), and those that analyze empirical data but do not test interactions (one of our goals is also to estimate the proportion of articles that include interaction tests among those that analyze quantitative data).

The chosen sample size ($N = 200$) and the level of detail in data coding reflect considerations of feasibility and available resources. Given that our primary objective is to obtain a general estimate of how frequently link functions are incorrectly used in testing interactions, we intentionally selected articles only from high-impact journals, ensuring a sufficient representation to approximate the magnitude of the issue. Following a strategy similar to that adopted by Hardwicke et al. (2024), our sample size determination was guided by considerations of precision. For binomial responses (e.g., whether an article reports at least one statistically significant interaction), the estimated proportion is most variable at 50%. With a sample size of $N = 200$, the corresponding 95% uncertainty bound is approximately $\pm 6.9\%$ (computed using standard methods such as the Wilson score interval, probit and logit transformations, or likelihood-ratio test-based intervals.)

2.1.1 Sources and Search strategy

Articles will be sampled from five journals, each corresponding to the categories previously listed. Journals are selected based on a discretionary combination of criteria: high Impact Factor (IF, based on the latest available Clarivate Journal Citation Reports from 2023; all selected journals rank within the first quartile in 2023), not primarily publishing review articles, not excessively specialized in a narrow empirical subfield, and a strong general reputation in their respective fields, as jointly evaluated by the coauthors of this review.

The selected journals are:

- *Psychological Science* (for general psychology);
- *Journal of Experimental Psychology-General* (for experimental psychology);
- *Journal of Personality and Social Psychology* (for social psychology);
- *Developmental Science* (for developmental psychology);
- *Psychological Medicine* (for clinical psychology).

Article records will be retrieved from Elsevier’s Scopus database, including all articles published in each of these journals in 2024. For each journal, a separate list of articles will be created. Within each list, exactly 40 eligible articles will be randomly selected for inclusion in the meta-analysis. To ensure randomization, R will be used: the `sample()` function will be applied separately to each journal’s full list, preceded by `set.seed(0)` for reproducibility. Articles will then be screened in order until 40 eligible articles per journal are included. If a journal has fewer than 40 eligible articles in 2024, the selection will be extended to 2023 using the same criteria.

2.1.2 Eligibility Criteria

The review will include only empirical articles presenting original psychological research. Articles will be excluded if they:

- Are reviews or meta-analyses;
- Are editorials, commentaries, opinion papers, or theoretical articles without empirical data;
- Exclusively employ qualitative research methods;
- Consist solely of replication studies;
- Focus primarily on the validation of psychometric instruments.

Replication studies and validation studies are excluded because the former are typically constrained to replicate existing methods from the original studies, while the latter are often focused on psychometric properties rather than hypothesis testing.

Following this initial screening, articles will be further examined, and detailed coding (as specified below in Table 1) will be performed only for those that :

- Test at least one interaction effect, either through statistical modeling or ANOVA.

2.2 Coding procedure

Table 1 reports the variables that will be coded for each included study, that is, each article identified as testing at least one interaction term in a statistical model or using ANOVA. Each study will be independently coded by two researchers. Any discrepancies will be resolved through discussion or, if needed, by consulting the other two researchers.

Table 1: Data Dictionary

| Variable Name | Description | Value Type |
|---|---|------------|
| <i>Authors</i> | Full list of authors' names. | Text |
| <i>Title</i> | Title of the article. | Text |
| <i>Year</i> | Year of publication. | Integer |
| <i>Journal</i> | Name of the journal in which the article was published. | Text |
| <i>Uses_non_identity_link_function</i> | Whether at least one of the tested interaction is presumably analyzed with a non-identity link function (e.g., because a generalized linear model was used). | Boolean |
| <i>Explicit_link_function</i> | Whether the link function is explicitly defined for at least one statistical model in the article (that is, in this case just indicating which generalized linear model or family was used is not enough; we code whether at least one link function is explicitly declared). | Boolean |
| <i>Incorrect_identity_link_function</i> | Whether at least one tested interaction is analyzed using an identity link function when another, non-identity link function would be more appropriate. | Boolean |

| Variable Name | Description | Value Type |
|---------------------------------------|---|------------|
| <i>Finds_significant_interactions</i> | Whether at least one interaction (where the link function was incorrectly specified as identity) yields a statistically significant result, or is considered and discussed in an analogous way if non-frequentist (e.g., Bayesian) analyses were carried out. | Boolean |
| <i>Response_variable_types</i> | A list of of the types of response variables on which interactions are tested. Possible values include: Sum score, Accuracy, Error count, Response times, Ordinal/Likert, This list may be expanded. | Category |

Use of AI tools

Researchers may optionally use generative AI tools (e.g., OpenAI’s GPT models) as assistants to help locate relevant information within articles. However, all relevant coding will be made by human researchers based on their own judgment. As a further exploration of AI potentiality, however, we will also have a parallel coding entirely conducted by AI tools using OpenAI’s (or others) API. Prompts worded in different ways will be used to extract the same variables without human intervention. The results of this AI-only coding will not be used in the review, but will be reported separately to explore the accuracy of AI language models in extracting categorical methodological information from scholarly articles.

2.3 Data analysis

The analysis will primarily focus on descriptive statistics. No hypothesis testing will be conducted, but 95% confidence intervals (CIs) will be computed for the estimated proportions of the Boolean variables. For each coded variable listed in Table 1, excluding *Authors*, *Title*, *Year*, and *Journal* (as they are not relevant to the analysis—summary statistics) the following will be computed:

- Boolean variables: Proportions will be calculated along with 95% CIs, using the Wilson score interval to estimate uncertainty Proportions will also be reported separately for each journal, but without CIs.
- *Response_variable_types*: The proportion of studies testing interactions on each type of response variable will be computed.

All analyses will be conducted in R, and results will be presented in summary tables. Graphical representations may also be included to facilitate interpretation.

References

- Domingue, B. W., Kanopka, K., Trejo, S., Rhemtulla, M., & Tucker-Drob, E. M. (2022). Ubiquitous bias and false discovery due to model misspecification in analysis of statistical interactions: The role of the outcome's distribution and metric properties. *Psychological Methods*, 29(6), 1164–1179.
- Hardwicke, T. E., Thibault, R. T., Clarke, B., Moodie, N., Crüwell, S., Schiavone, S. R., Handcock, S. A., Nghiem, K. A., Mody, F., Eerola, T., et al. (2024). Prevalence of transparent research practices in psychology: A cross-sectional study of empirical articles published in 2022. *Advances in Methods and Practices in Psychological Science*, 7(4), 25152459241283477.
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79, 328–348.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105(1), 156.