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

Enrico Toffalini, Margherita Calderan, Tommaso Feraco, Filippo Gambarota

2025-03-05

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, in our experience, psychology researchers frequently test interaction effects to identify moderators of known main effects, but also because it is often just a standard routine to test interactions after testing main effects. Typically, classical linear models or equivalent methods (e.g., ANOVA) are employed. By default, these methods use identity link functions, despite many psychological response variables being inherently non-normally distributed due to natural bounds (e.g., accuracies, response times, sum or average scores, error counts, proportions). Moreover, psychometric variables have been systematically shown to deviate from normal distributions (Micceri 1989). Failing to employ appropriate link functions can lead to the detection of pseudo-interactions, as incorrect link functions distort interval scales, transforming equal intervals into unequal ones and *vice versa*. Consequently, using inappropriate link functions places researchers at risk of identifying statistically significant interactions that do not reflect the true data-generating processes.

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.

---

2. Methods

2.1 Sample

A sample of 200 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, 40 articles will be randomly sampled, resulting in a total sample of 200 articles.

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 Scopus database of Elsevier, selecting all articles published in these journals during 2024. For each journal, 40 articles will be randomly selected using R software (by applying the `sample()` function on the list of records, preceded by `set.seed(0)` to ensure reproducibility). Sampled articles that do not meet eligibility criteria during screening will be excluded and replaced by additional articles selected through subsequent rounds of random sampling, continuing until 40 eligible articles per journal are obtained. Should a journal have fewer than 40 eligible articles published in 2024, the sampling procedure will be extended to articles published by that journal in 2023.

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.

2.2 Coding procedure

Table 1 reports the variables that will be coded for each included study. Each study will be independently coded by two researchers. Any discrepancies will be resolved through discussion or, if necessary, consultation with the other two researcher.

Table 1: : Data Dictionary

Variable Name	Description	Value Type
Authors	Full list of authors' names.	Text
Title	Title of the article.	Text

Variable Name	Description	Value Type
Year	Year of publication.	Integer
Journal	Name of the journal in which the article was published.	Text
Tests_interactions	Whether the article tests at least one interaction term in a model or using an ANOVA.	Boolean
Uses_non_identity_link_function	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
Explicit_link_function	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
Incorrect_identity_link_function	Whether at least one tested interaction is analyzed using an identity link function when another, non-identity link function would be more appropriate.	Boolean
Finds_significant_interactions	Whether at least one interaction (where the link function was incorrectly specified as identity) yields a statistically significant result.	Boolean

Variable Name	Description	Value Type
Response_variable_types	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

2.3 Data analysis

References

Hardwicke, Tom E, Robert T Thibault, Beth Clarke, Nicholas Moodie, Sophia Crüwell, Sarah R Schiavone, Sarah A Handcock, 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.

Micceri, Theodore. 1989. "The Unicorn, the Normal Curve, and Other Improbable Creatures." *Psychological Bulletin* 105 (1): 156.