Misuse of nonparametric tests with heteroscedasticity: A semi-automated review of obesity research

Kroeger CM[1,2], Hannon BA[3], Halliday TM[4], Teran-Garcia M[3], Allison DB[2]

Charles Perkins Centre, Faculty of Medicine and Health, School of Pharmacy, University of Sydney, Camperdown, NSW, Australia[1]

Department of Epidemiology and Biostatistics, Indiana University School of Public Health-Bloomington, Bloomington, IN, USA[2]

Division of Nutritional Sciences, University of Illinois at Urbana-Champaign, Champaign, IL, USA[3]

Division of Endocrinology, Metabolism and Diabetes, University of Colorado Anschutz Medical Campus, Aurora, CO, USA[4]

Background

Use of classic nonparametric tests (NPTs) (i.e., rank-based tests, such as Kruskal-Wallis, Wilcoxon Rank, and others) in the presence of variance heterogeneity (heteroscedasticity) as a test of differences in central tendency (e.g., mean or median) is an invalid approach [1-4]. However, the extent to which this practice occurs within obesity research is unknown. Further, the systematic reviews needed to gather evidence on such topics are time consuming and often difficult to replicate.

We aimed to (1) determine the prevalence and impact of this error within nutrition and obesity articles indexed in PubMed Central in 2017 and (2) improve efficiency and replicability review methods by utilizing R packages to semi-automate search and data extraction methods.

Methods

We used the Open Science R package “rEntrez” to identify nutrition and obesity articles published within PubMed Central in 2017 and containing heteroscedasticity and nonparametric terms. Search terms were developed with a librarian at the University of Illinois at Urbana-Champaign.

Reason for the use of a heteroscedasticity term within statistical methods sections of included studies and study characteristics were independently, manually extracted by two investigators. Use of heteroscedasticity term was categorized within one of the following four categories:

1. Use of a NPT when heteroscedasticity was explicitly tested for and described as not present (**Correct**)
2. Use of a NPT to correct for or explicitly in the presence of heteroscedasticity (**Error**)
3. Mentions heteroscedasticity term with no link to rationale for use of NPT (**No Link**)
4. States heteroscedasticity assumptions were tested or met but does not explicitly state whether heteroscedasticity was present or whether assumptions were met for all NPTs (**Ambiguous**)

The R packages “rCrossref” and “rAltmetric” were used to automatically extract data on scientific and public impact, respectively. We archived our data and code to allow for automatic replication of our literature search, data extraction on scientific and public impact, and figure generation [5].

Results

Out of 38 studies that met our inclusion criteria, NPTs were used within 9 as a mistaken solution for or in the presence of variance heterogeneity, 25 without directly stating whether heteroscedasticity was present, and 4 after appropriately declaring that heteroscedasticity was not present. The papers containing clear errors and ambiguities already have been cited within scientific papers (n = 19 and n = 38, respectively) and public posts (n = 40 and n = 286, respectively), such as via Twitter and news outlets and by the following cohorts: the public, scientists, science communicators, and practitioners.

Figure 1. Flowchart of inclusion procedure

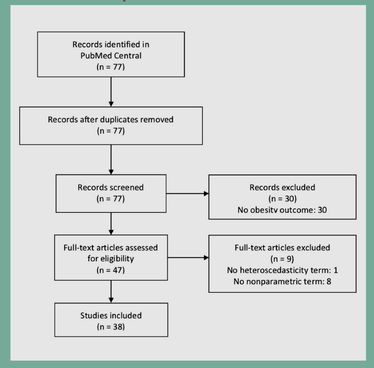


Figure 2. Appropriateness of NPT use in included studies

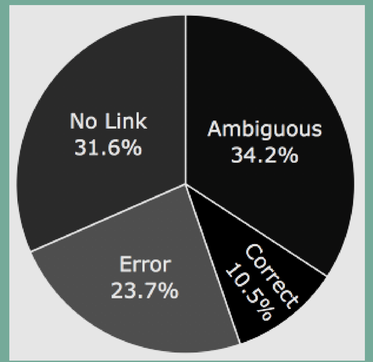
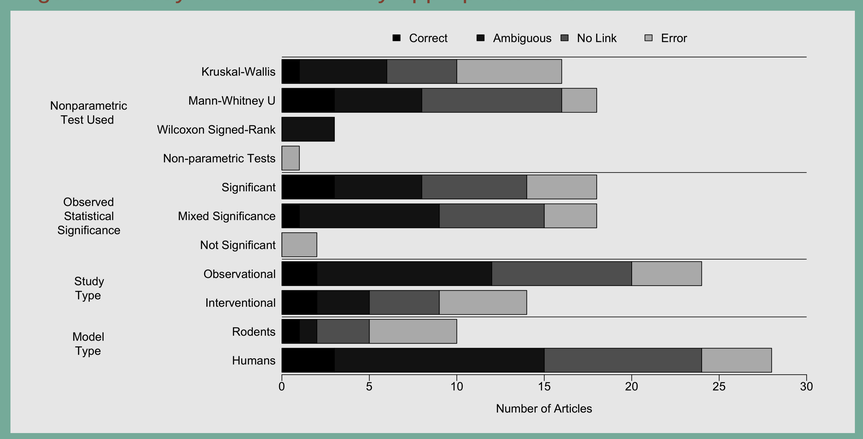
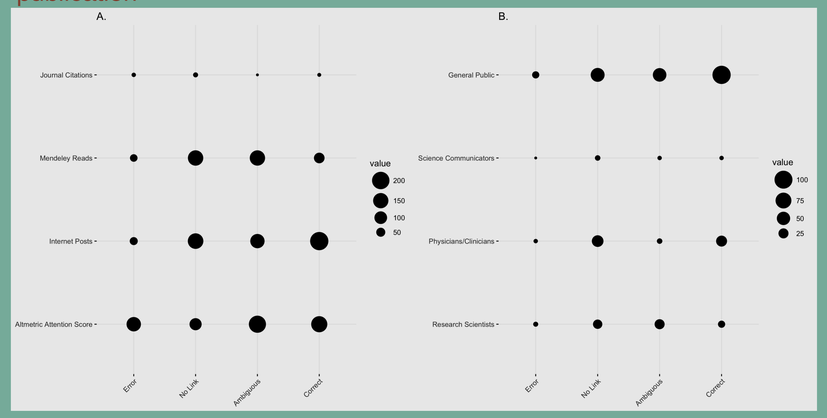


Figure 3. Study characteristics by appropriateness of NPT use



Non-parametric Tests: Description of the test used was “non-parametric tests.” Observed statistical significance: Number of articles that report results obtained from a nonparametric test as statistically significant (significant), mixed (mixed significance), or not significant.

Figure 4. Study dissemination after an average of 314 days since publication



A. Total number of journal citations, Mendeley reads, and internet posts, and total Altmetric Attention Score by appropriateness of nonparametric test use. B. Total number of internet posts (e.g., news, blogs, twitter, Wikipedia, LinkedIn, Facebook, etc.) by cohort and appropriateness of NPT use.

Conclusions

The majority of obesity papers indexed in PubMed Central in 2017 that mention heteroscedasticity and nonparametric terms either utilized a NPT due to the erroneous rationale that such choice is appropriate when data are heteroscedastic or without stating that heteroscedasticity was not present. Because these articles have been widely disseminated within scientific and public domains, future work that assesses the impact of this error on statistical inference may be useful.

References

1. George BJ, Beasley TM, Brown AW, Dawson J, Dimova R, Divers J, Goldsby TU, Heo M, Kaiser KA, Keith SW, et al. Common scientific and statistical errors in obesity research. Obesity (Silver Spring) 2016;24(4):781-90. doi: 10.1002/oby.21449.
2. Wilcox RR. Introduction to Robust Estimation and Hypothesis Testing: Statistical Modeling and Decision Science. 3 ed: Academic Press, 2011.
3. Zimmerman DW. A Note on Homogeneity of Variance of Scores and Ranks. The Journal of Experimental Education 2014;64(4):351-62. doi: 10.1080/00220973.1996.10806603.
4. Kaiser KA, Gadbury GL. Estimating the range of obesity treatment response variability in humans: methods and illustrations. Hum Hered 2013;75(2-4):127-35. doi: 10.1159/000351738.
5. Kroeger CM. Data and analysis code for conference abstract: Misuse of nonparametric tests with heteroscedasticity: A semi-automated review of obesity research. Zenodo. 2018. DOI: 10.5281/zenodo.1480042

Funding

Supported in part by NIH grant, F32DK107157. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or any other organization.