

East Tennessee State University

## Digital Commons @ East Tennessee State University

---

Undergraduate Honors Theses

Student Works

---

5-2021

### Investigating the Misrepresentation of Statistical Significance in Empirical Articles

Blythe Lybrand

Ginette Blackhart

*East Tennessee State University*

Amanda Parish

*East Tennessee State University*

Hannah Lowe

*East Tennessee State University*

Follow this and additional works at: <https://dc.etsu.edu/honors>



Part of the [Social Psychology Commons](#)

---

#### Recommended Citation

Lybrand, Blythe; Blackhart, Ginette; Parish, Amanda; and Lowe, Hannah, "Investigating the Misrepresentation of Statistical Significance in Empirical Articles" (2021). *Undergraduate Honors Theses*. Paper 646. <https://dc.etsu.edu/honors/646>

This Honors Thesis - Withheld is brought to you for free and open access by the Student Works at Digital Commons @ East Tennessee State University. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact [digilib@etsu.edu](mailto:digilib@etsu.edu).

**Investigating the Misrepresentation of Statistical Significance in Empirical Articles**

By

Blythe Battle Lybrand

Co-Authors: Ginette C. Blackhart, Amanda Parish, & Hannah Lowe

An Undergraduate Thesis Submitted in Partial Fulfillment  
of the Requirements for the  
Honors College  
and the  
Fine and Performing Arts Scholars Program  
College of Arts and Sciences  
Self and Relationships Lab  
East Tennessee State University



---

Blythe B. Lybrand 5/5/2021  
Date



---

Dr. Ginette C. Blackhart, Thesis Mentor 5/5/2021  
Date

*Samantha Castelblanco* 5-5-21

---

Samantha A. Castelblanco, M.A., Reader Date

**Table of Contents**

|  |       |
|--|-------|
| Acknowledgements.....                              | 4     |
| Abstract.....                                      | 5     |
| Misrepresentation of Statistical Significance..... | 6     |
| Inferential Hypothesis Testing.....                | 7-8   |
| “Marginal” Significance.....                       | 8-10  |
| Method.....  | 10-12 |
| Study Design.....                                  | 10    |
| Data Collection.....                               | 11    |
| Data Analysis.....                                 | 12    |
| Results.....                                       | 12-14 |
| Discussion.....                                    | 15-17 |
| Limitations.....                                   | 15    |
| Future Research.....                               | 16    |
| Conclusion.....                                    | 16-17 |
| References.....                                    | 18    |

**Tables and Figures**

Table 1, Excel Spreadsheet Example.....12

Table 2, Descriptive Statistics.....13

Figure 1, Descriptive Graph.....14

### **Acknowledgements**

I would like to personally thank those in the research lab who assisted with the collection and coding of the data for this study. In particular, Dr. Ginette C. Blackhart played an instrumental role in ensuring that the process was smooth by uploading all of the necessary files and making them readily available to myself as well as other research assistants. Thank you all for your involvement with this research.

### Abstract

In an attempt to preserve research integrity, the aim of this study is to examine how often statistical results are being misrepresented in empirical studies by using terms such as “marginally significant,” “approached significance,” or “trend toward significance” when interpreting findings. The use of these terms gives ambiguous significance to results that are in fact nonsignificant, which threatens future research by contributing to issues such as the replication crisis. For this study, data were coded from 437 empirical articles published online in *The Journal of Personality and Social Psychology (JPSP)* over a 4-year period between 2017 and 2020. According to our findings, although misrepresentation of statistical results are prevalent within *JPSP* articles, rates decreased significantly over the four-year time period examined. Additionally, as the number of studies published in *JPSP* increased each year during the four-year period examined, there may be a potential rise in representatively sound studies and decrease of misrepresentation within this discipline.

*Keywords:* margin\*, approach\*, trend\*, marginally significant, misrepresentation, JPSP, empirical articles

## Investigating the Misrepresentation of Statistical Significance in Empirical Articles

The misrepresentation of statistical significance in empirical articles is an issue that has been infecting empirical studies and journals within the scientific community for decades (Nuijten, Hartgerink, Van Assen, Epskamp, & Wicherts, 2015). While  $p$ -hacking and falsely creating results are more well-known acts of scientific fraud, there is a lesser-known issue that contributes to the replication crisis in the psychological sciences, which is using misrepresentative terms to attribute statistical significance to nonsignificant results. Phrases commonly used to attribute statistical significance to nonsignificant results include “marginally significant,” “approaching significance,” “a trend toward significance,” and similar phrases (for a complete list of ways in which researchers may misrepresent the statistical significance of their results, see Hankins, 2013). It is important to note that marginal significance is not the same as true statistical significance. Using the term “marginally significant” is an indirect way of admitting that the resulting  $p$ -value was not statistically significant while attempting to give it the appearance of statistical significance. This is more commonly seen when the  $p$ -value is very close to being less than 0.05 but is ultimately greater than 0.05. Researchers commit this fraudulent act because the more significant their results are (or in this case, seem), the more likely the study is to be published; in turn, these publications may reward researchers with grants or financial aid to continue their research or begin new studies. This is not only dangerous to the validity and reliability of present and future research, but it also takes financial aid away from potential researchers and studies who are being honest with their results and finding significance worth researching and expanding further.

## Inferential Hypothesis Testing

Within inferential hypothesis testing, the null hypothesis assumes that there is no significant results within the data (e.g., no relationship between the variables, no difference between groups). On the other hand, the alternative hypothesis (or research hypothesis) is the outcome researchers hope their data support. If the results are significant, the null hypothesis will be rejected because the predictions of the researcher were supported. Oppositely, if the results are insignificant, the null hypothesis fails to be rejected and thus indicates that the predictions of the researchers were unsupported.

*P*-value is the probability of evidence that the null hypothesis should be rejected in a study; a smaller *p*-value indicates stronger evidential support for the alternative hypothesis, and is ultimately what researchers hope to find. As a standard in the scientific community, confidence levels are typically set at 95% and alpha is set at .05, which translates to researchers hypothesizing that their chances of making a Type I error is less than 5%. Thus, *p*-values with alpha set at .05 are only significant when resulting in less than .05, or less than 5%.

To recap, in inferential hypothesis testing, statistical significance is indicative of the resulting *p*-value; if research indicates a *p*-value of less than .05, the results are revealed as being statistically significant, thus the null hypothesis gets rejected, allowing for studies to be replicated and investigated further. However, if the *p*-value is greater than .05, the results are ultimately insignificant, and the null hypothesis fails to be rejected and researchers either move on or find new measures to study.

It is important to note that in some instances, while a null hypothesis may fail to be rejected, it can still be an important finding to research and should not be discredited or



overlooked. Often in psychology, researchers are expected to have results which reject their null hypotheses in order to get published, but results that fail to reject these hypotheses are important because we learn something from them too. However, this becomes negated when researchers claim marginal significance or phrases of the like.

### **“Marginal” Significance**

As stated above, marginal significance is not the same as true statistical significance because it distorts the actual results and suggests significance for nonsignificant findings (Pritschet, Powell, & Horne, 2016). This misrepresentation and the use of terms such as “marginal” to describe significance for nonsignificant statistical results can cause several problems, the predominant issue being that it contributes to the replication crisis.

In the psychological sciences, the replication crisis, also referred to as the reproducibility or replicability crisis, refers to the increasing belief that the results of numerous scientific studies are incapable of being reproduced or replicated by other researchers (Nosek, Cohoon, Kidwell, & Spies, 2016). This belief often leads to the assumption that those results are insignificant, that the research is wrong, and that it should not be trusted which is a growing concern for many researchers and experts in the field. Some issues that directly contribute to this crisis are *p*-hacking or falsifying data, a lack of uniformity pertaining to acceptable statistical significance within the field, researcher bias and the incentive to publish significant results, and inconsistencies within the peer review system. These issues negatively impact psychological research as a whole by confirming the negative biases of those questioning the research, falsely influencing the findings of future studies, wasting resources and finances, and causing researchers to overlook relationships that may in fact be significant. All of these issues threaten

the integrity of psychological research, as well as the integrity of the researchers and journals to which these problematic studies are published.

In 2019, research conducted by Olsson-Collentine, van Assen, and Hartgerink on the misrepresentation of statistical significance was published in *Psychological Science*. The study, titled "The Prevalence of Marginally Significant Results in Psychology Over Time," investigated the percentage of  $p$ -values ( $.05 < p\text{-value} \leq .10$ ) that were being reported as marginally significant in published psychological journals spanning from 1985 to 2016. Researchers assessed a total of 74,489 online articles and coded 42,504  $p$ -values falling between .05 and .10.

Their findings indicated that roughly 40% of  $p$ -values within this range were recorded by researchers as being marginally significant, and additionally indicated that there was significant variation between nine psychological disciplines. Specifically, the practice was observed the most in organizational psychology (45.4%) and social psychology (44.5%), and observed the least in clinical psychology (30.1%). Furthermore, results showed that the percentage of  $p$ -values being reported as marginally significant either remained constant or decreased over time across disciplines. It was noted that, of the two journals assessed, *The Journal of Personality and Social Psychology (JPSP)* in particular showed an increase of nonsignificant  $p$ -values being reported as marginally significant, unjustly representing the social psychology discipline. As a result, the purpose of this study is to examine whether or not more recent publications to *JPSP* will replicate these former trends, as well as to hopefully discover that this prevalence is decreasing as the years progress.

The specific goal of this study is to assess the prevalence of misrepresented statistical results within empirical articles published to *JPSP* over a four-year period (2017-2020), as well as to determine whether that prevalence is increasing or decreasing over time. The hypotheses

for this study are as follows: **H<sub>1</sub>**: From 2017-2020, social and personality psychology researchers continue to use the terms "margin\*," "approach\*," and "trend\*" as they relate to statistical significance in *JPSP* articles, misrepresenting their statistical results. **H<sub>2</sub>**: The prevalence of misrepresentation of statistical results in *JPSP* articles will decrease from year to year from 2017-2020.

It is important to note that the second research hypothesis somewhat contradicts the findings of Olsson-Collentine et al. (2019) which claims the percentage of *p*-values being recorded as marginally significant in *JPSP* is increasing annually. The reason for this opposition stems from the idea that researchers within the psychological field, particularly personality and social psychological disciplines, have likely heard of or read these results seeing as this research was published in 2019 and subjected to peer review. Thus, the publication of this study likely caught the attention of researchers who regularly or sometimes use terms such as marginal significance and resulted in an individualized realization that they need to make a mindful effort to negate these terms from their vocabulary, being especially particular when publishing to *JPSP*.

## Method

### Study Design

Seeing as this study did not use participants, but rather examined and collected preexisting data, it is considered to be a correlational design due to the nature and desire of this research to evaluate instances of misrepresentation when interpreting statistical significance. Four years' worth of articles from *JPSP* were assessed and compared for their inherent

relationship with misrepresentation of statistical significance by using the key terms “marginally significant\*,” “approached significance\*,” and/or “trended significance\*.”

### **Data Collection**

To test the study hypotheses, three research assistants and I coded original empirical articles published to the *Journal of Personality and Social Psychology (JPSP)* from 2017 – 2020 ( $n = 437$ ). *JPSP* was looked at in particular for the purpose of replicating and furthering the previous findings of Olsson-Collentine et al. (2019), and additionally because it is one of the most commonly accessed social psychological journals. Truncation was used to manually search for three key words in each original empirical article as they relate to statistical results: "margin\*," "approach\*," and "trend\*." We then examined whether those terms were used in relation to presenting statistical results. The number of instances those key terms were used in relation to presenting statistical results were then recorded. For example, the procedure would begin with an article being opened online and searched for the word "margin\*." If a certain article highlighted 5 instances where the word "margin" occurred, research assistants would investigate whether or not it was related to the representation of statistical significance, such as saying they had found "marginally significant" results rather than statistically significant results. For every time this was the case, it was counted and the total number of times the term “margin\*” was used per article was documented in an Excel file as shown in Table 1. This process was then repeated using the words "approach\*" and "trend\*" for the same document before moving on to the next article and starting the process over.

| Journal | Year | Volume | Issue | Author            | Coded by: | # "Margin" | # "Approach" | # "Trend" | Total # |
|---------|------|--------|-------|-------------------|-----------|------------|--------------|-----------|---------|
| JPSP    | 2017 | 112    | 3     | Crawford et al.   | Lybrand   | 19         | 1            | 1         | 21      |
| JPSP    | 2017 | 112    | 3     | Chester & DeWall  | Lybrand   | 0          | 0            | 0         | 0       |
| JPSP    | 2017 | 112    | 3     | Bitterly et al.   | Lybrand   | 2          | 0            | 0         | 2       |
| JPSP    | 2017 | 112    | 3     | Rogers et al.     | Lybrand   | 1          | 0            | 0         | 1       |
| JPSP    | 2017 | 112    | 3     | Mottus et al.     | Lybrand   | 0          | 0            | 0         | 0       |
| JPSP    | 2017 | 112    | 3     | Lipnevich et al.  | Lybrand   | 0          | 0            | 0         | 0       |
| JPSP    | 2017 | 112    | 3     | Milojev & Sibley  | Lybrand   | 0          | 0            | 7         | 7       |
| JPSP    | 2017 | 112    | 4     | Zwebner et al.    | Lybrand   | 2          | 0            | 0         | 2       |
| JPSP    | 2017 | 112    | 4     | Snyder & Tormala  | Lybrand   | 2          | 0            | 0         | 2       |
| JPSP    | 2017 | 112    | 4     | Mooijman et al.   | Lybrand   | 1          | 0            | 0         | 1       |
| JPSP    | 2017 | 112    | 4     | Boothby et al.    | Lybrand   | 0          | 0            | 0         | 0       |
| JPSP    | 2017 | 112    | 4     | MacGregor et al.  | Lybrand   | 0          | 0            | 0         | 0       |
| JPSP    | 2017 | 112    | 4     | Webb et al.       | Lybrand   | 5          | 1            | 0         | 6       |
| JPSP    | 2017 | 112    | 4     | Parrigon et al.   | Lybrand   | 0          | 0            | 1         | 1       |
| JPSP    | 2017 | 112    | 5     | Twenge et al.     | Lybrand   | 0          | 0            | 0         | 0       |
| JPSP    | 2017 | 112    | 5     | Erle & Topolinski | Lybrand   | 0          | 0            | 0         | 0       |
| JPSP    | 2017 | 112    | 5     | Zou & Cheryan     | Lybrand   | 2          | 0            | 0         | 2       |

**Table 1.** This image depicts the layout and coded results of a section from the Excel spreadsheet for visual aid.

## Data Analysis

All data analyses were conducted in JASP 11.1. Descriptive statistics were used to assess prevalence of the use of misrepresentative terms of statistical significance in empirical articles published in *JPSP*. Additionally, a one-way ANOVA and a Tukey post-hoc test were conducted to assess variation between years regarding the use of such terms.

## Results

Descriptive statistics revealed a significant use of the truncated terms "margin\*," "approach\*," and "trend\*" in *JPSP* articles over time when reporting statistical results as they are presented in Table 2. A One-Way ANOVA was conducted to determine whether articles submitted to *JPSP* from 2017 to 2020 would vary between years regarding the misrepresentation of statistical significance. Results indicated there was a significant difference of misrepresentation between years,  $F(3,430) = 6.022, p < .001$ . A Tukey post-hoc comparisons showed that these differences are significant between the years 2017 and 2019 ( $M = 1.227, p = .049$ ), 2017 and 2020 ( $M = 1.863, p < .001$ ) and 2018 and 2020 ( $M = 1.196, p = .037$ ).

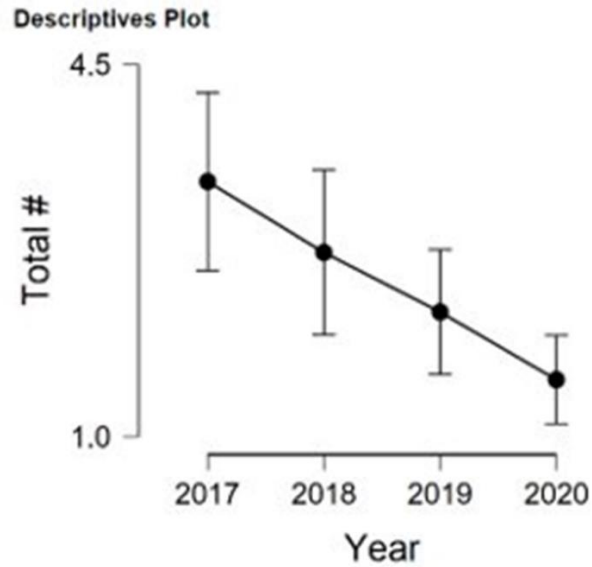
Seeing as the descriptive statistics indicated a significant use of misrepresentation as it relates marginal significance, the first research hypothesis was ultimately supported. This also supports the findings of Olsson-Collentine et al. (2019), suggesting that misrepresentation of statistical results in the psychological sciences is prevalent. The second research hypothesis predicting a decrease in misrepresentation was also supported by the indicated variance between years.

Additionally, the descriptive statistics revealed that there was an increased number of new articles published in *JPSP* each year, which is promising in terms of research integrity because if more and more articles are being published with true significance, there's a greater change that new standards will develop to prevent the replication crisis from continuing any further. Since the psychological and scientific communities want to see more truly significant articles with no misrepresentation, this significant decrease of misrepresentation of statistical results in articles published in *JPSP* is encouraging for social and personality psychology.

Descriptive Statistics ▼

|                | Total # |        |        |        |
|----------------|---------|--------|--------|--------|
|                | 2017    | 2018   | 2019   | 2020   |
| Valid          | 93      | 104    | 111    | 129    |
| Missing        | 0       | 0      | 0      | 0      |
| Mean           | 3.398   | 2.731  | 2.171  | 1.535  |
| Std. Deviation | 4.068   | 3.974  | 3.110  | 2.385  |
| Minimum        | 0.000   | 0.000  | 0.000  | 0.000  |
| Maximum        | 21.000  | 22.000 | 16.000 | 15.000 |

**Table 2.** Descriptive statistics show an increase of articles submitted annually and a decreased use of ambiguous terms within those articles.



**Fig. 1.** This graph depicts a decreasing linear trend of misrepresentation using terms from 2017 to 2020. Additionally, confidence intervals can be seen getting smaller as the years progress.

Figure 1 above depicts a graphing of these statistics which resulted in a beautiful decreasing linear trend, suggesting that misrepresentation as it relates to marginal significance is decreasing annually within *JPSP* articles. Additionally, it can be observed that confidence intervals are decreasing, indicating that researchers are more confident that their chances of committing a type II error are small. Seeing as many researchers and individuals within the psychological and scientific communities are pushing for the confidence interval to be shifted from 95% to 99%, this is a positive direction for the future validity in psychological research.

## Discussion

The goals of this research were to (1) examine how often statistical results are being misrepresented in empirical studies by using terms such as “marginally significant” to interpret results, and (2) to determine whether there is variation in such misrepresentation between the years spanning 2017-2020. To an extent, our findings replicated those of Olsson-Collentine et al. (2019), such as demonstrating the continued prevalent use of misrepresentation within *JPSP*. Independently, our findings indicated that misrepresentation using these terms is in fact occurring within *JPSP*, however decreasing annually. This information supported both research hypotheses and gave additional insight to the growing number of articles being submitted without misrepresentation annually. While it is unfortunate that misrepresentation takes place at all within the psychological field, it is positive that its prevalence is decreasing in *JPSP* articles as the years progress.

## Limitations

There were a few limitations to this particular study, the biggest one being that only articles from *JPSP* were analyzed. Although *JPSP* is the flagship journal of the Society for Personality and Social Psychology and is the highest-impact social and personality psychology journal within the U.S., it may not be representative of how social and personality psychology researchers are presenting statistical results in other social and personality psychology journals. Additionally, although the words "margin," "approach," and "trend" were the key terms searched for, other terms may have been used to misrepresent statistical significance of results. Although we made a conscious effort to find the top three most commonly used terms in psychology to misrepresent data, there is always a chance that other variations or synonyms went undetected. Lastly, there were instances where these words were used in conjunction with one another, such



as using the phrase "marginal trend" or "approaching trend," so when this occurred, research assistants only counted the instance once under one term or the other. While this was not a common occurrence and is unlikely to impact the study, it is important to recognize for those hoping to replicate or further this study in the future.

### **Future Research**

This study will be furthered by continuing with the current method and looking at articles from the *Journal of Experimental Social Psychology (JESP)*, *Personality and Social Psychology Bulletin (PSPB)*, and *Social Psychological and Personality Science (SPPS)* journals. This research is already being conducted within the Self and Relationships Lab and will hopefully render more results not long after the completion of this portion of the study.

Additionally, in terms of misrepresentation consistency, it would be interesting if future research were to look into whether authors who misrepresent their statistical results have their articles published to these journals repeatedly publish studies over the years, as well as whether they do so to the same or different journals. Assuming there are cases where researchers have repeatedly published their results but also repeatedly misrepresented their findings in said results, it would be especially interesting to discover whether those researchers are increasing, decreasing, or maintaining this behavior.

### **Conclusion**

Within the scientific and psychological communities, dishonesty and misrepresentation, whether intentional or unintentional, is a dominating issue that disadvantages other researchers, the research itself, publishers, and consumers alike. Because this study provides evidence that this misrepresentation is still prevalent, it is revealed as an ongoing issue that will hopefully one

day be resolved. However, seeing as the evidence also suggests significant decreases in this prevalence, one can confidently assume that it will soon become a less common issue within the field.

## References

- Hankins, M. C. (2013). Still Not Significant. Retrieved April 13, 2021, from <https://mchankins.wordpress.com/2013/04/21/still-not-significant-2/>
- Nosek, B. A., Cohoon, J., Kidwell, M., & Spies, J. R. (2016). Estimating the reproducibility of Psychological Science. *Science*, *349*(6251). doi:10.31219/osf.io/447b3
- Nuijten, M. B., Hartgerink, C. H., Van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2015). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, *48*(4), 1205-1226. doi:10.3758/s13428-015-0664-2
- Olsson-Collentine, A., Van Assen, M. A., & Hartgerink, C. H. (2019). The prevalence of marginally significant results in psychology over time. *Psychological Science*, *30*(4), 576-586. doi:10.1177/0956797619830326
- Pritschet, L., Powell, D., & Horne, Z. (2016). Marginally significant effects as evidence for hypotheses. *Psychological Science*, *27*(7), 1036-1042. doi:10.1177/0956797616645672