

ISSN: 2688-836X



***Corresponding author:** Walter R Schumm, Adjunct Instructor, Math Department, Highland Community College, Wamego, Kansas, and Emeritus Professor, Department of Applied Human Sciences, Kansas State University, Manhattan, USA

Submission:  July 13, 2022

Published:  July 27, 2022

Volume 11 - Issue 3

How to cite this article: Walter R Schumm. The Challenge of Statistically Detecting Authors Responsible for Publishing Fabricated Research Data or Results: A Concept Paper. *Nov Res Sci.* 11(3). NRS. 000765. 2022.
DOI: [10.31031/NRS.2022.11.000765](https://doi.org/10.31031/NRS.2022.11.000765)

Copyright@ Walter R Schumm, This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use and redistribution provided that the original author and source are credited.

The Challenge of Statistically Detecting Authors Responsible for Publishing Fabricated Research Data or Results: A Concept Paper

Walter R Schumm*

Adjunct Instructor, Math Department, Highland Community College, USA

Abstract

Although uncommon, some professionals have published research where either the data or the results were fabricated. If there are several authors involved in such research, how can the primary source of the research misconduct be identified from among multiple authors? Use of correlation or regression analysis is one possibility, along with analysis of variance, for predicting research anomalies from level of authorship across multiple authors and articles. The procedures are demonstrated with a simulated set of data. Results indicate that the proposed methodologies could help identify authors primarily responsible for fabrication of data or results, even for medium-to-large samples. Awareness that responsible authors could be identified from among several co-authors of published articles may help discourage research practices involving misconduct. Co-authors, even if they are not directly responsible for such misconduct, should be careful lest they become unexpectedly associated with unethical research behavior.

Keywords: Research misconduct; Fabricated research; Research ethics; Data anomalies; Multiple authorship; Statistics; Research methods

Introduction

Science is founded upon the concept of truth-or at least facts. If scientists do not report their findings accurately, then such science cannot be trusted and remains unsuitable for the development of scientific theory and future research. Sometimes scholars may not understand how to use statistics or research methods, often because of inadequate training in graduate school [1,2]. Some scholars may feel obligated to learn difficult statistical packages, which can be confusing and easy to misuse, even though free methods are readily available on the internet [3]. Some scholars may suffer from confirmation bias and massage their data or findings to more closely approximate the results they expect or believe are for a good cause [2,4]. Sometimes scholars or type setters for journals can introduce statistical errors into their reports [5]. However, here the discussion will focus on research misconduct involving the outright fabrication of data or results.

Literature Review

Research misconduct is not as uncommon as the public might think. Horton et al. [6] report an increase in academic dishonesty among scholars. Some scholars have had dozens of their articles retracted [6-10]. Some researchers may outright fabricate data or results, as incredible and risky as that may seem to both those who do it and to the academic community in general. However, on the other hand, "The potential benefits from such academic misconduct are high, whilst the likelihood of such behaviour being detected is extremely low" [6]. There can be severe consequences if scientific misconduct is attached to a specific scholar [11]. Even co-authors who had nothing to do with fraud may have their reputations tarnished

and become mistrusted by others [6,12]. Pickett [13] recently identified several anomalies in research published by Professor Eric A. Stewart of the University of Florida. Pickett seemed to believe that Stewart had fabricated at least part of his data and more certainly some of his alleged research results. Stewart, as Pickett noted, ended up retracting several of his published articles once some of the problems were exposed, including issues with binary variables where the standard deviations and mean scores did not follow the required mathematical patterns [14,15]. Pickett [13] noted that Stewart had been the first or second author on five retracted articles and that he had been, for every retracted article, the author responsible for the data and analyses. However, it's one thing to allege research fraud but another to provide statistical evidence for it.

Research Question

Horton et al. [6] noted with apparent dismay that it is difficult to detect fraudulent research, even stating that "there are no processes to scrutinize the authenticity of such data" (p.2). Thus, they conclude that "any process that can be used to uncover potentially fabricated data should be of benefit to, and welcomed by, the academic community" (p.2). Even if one could use statistics to prove that research was likely fabricated [6], such a finding might do little to advance the progress of science unless those responsible could be identified. It may be that fraudulent researchers "tend to work in larger teams" [6,16], possibly to help conceal their identity among those of their co-authors. In other words, even if the research were to be identified as fraudulent, it might be impossible to identify which of the co-authors were responsible for the scientific misconduct, enabling the guilty to escape detection. For example, suppose a series of articles were co-authored by five scholars. Suppose that several articles were found to have such a large number of improbable anomalies that they could be determined to have been fabricated. The authors could easily deny knowledge of any fraud and coordinate with their university to cover up any suggestion of fraud, especially if their university or other organization stood to lose millions of dollars in research funding. It might also be true that most of the co-authors were unaware of any scientific misconduct and were largely innocent bystanders [6].

Thus, the question would naturally occur-how could any author responsible for research misconduct be identified specifically? In academia, a common pattern is for authorship rank to be determined by relative contributions, with providing data and conducting analyses being very important parts of scholarly contributions. The American Psychological Association [17] has stated that "Principal authorship and the order of authorship credit should accurately reflect the relative contributions of persons involved {APA Ethics Code Standard 8.12b, Publication Credit}" and that "The general rule is that the name of the principal contributor should appear first, with subsequent names in order of decreasing contribution" (p.19). If an author was fabricating research, they might well continue to insist upon due "credit" by a higher ranking in the list of authors for

the articles since they were, at least in theory, responsible for the management of data and often the conduct of statistical analyses. If so, then anomalies in research might well be correlated more strongly with rank of authorship for the guilty party. Anomalies might include standard deviations that didn't match mean scores in binary analyses [14,15], nonrandom final digits in results, first digits in results that do not conform in frequency to Benford's Law [18], and other nonrandom patterns in tables of results [6,13]. For example, the third digit of a correlation or regression coefficient is likely random; thus, if out of 500 correlations, only four ended in zero, that would represent an anomaly since obtaining 4/50 ending zeroes, when $p=10$, would be extremely unlikely in terms of probability, a result almost seven standard deviations below the expected value of 50. Pickett [13] has provided other examples of anomalies.

Since a higher number of anomalies might be expected for higher rank of authorship for the person most responsible for fraudulent research (with higher rank meaning first or second authorship rather than fourth or fifth authorship), a negative correlation or regression coefficient would be expected between authorship rank and the number of anomalies for "responsible" authors while a zero, low, or positive correlation might be expected for co-authors unaware of the fabrications. In a hypothetical, simulated data set, would the following results be observed?

- a. For authors responsible for fabrication: negative correlations or regression coefficients.
- b. For authors not directly responsible for fabrication: zero, small, or positive correlations or regression coefficients.
- c. As a concept paper, the objective is not to analyze existing data in an attempt to identify an actual source of scientific misconduct but, rather, to demonstrate how such an analysis might be performed, using simulated data.

Methods

Data

Data were simulated on an assumption that one co-author was "responsible" while another coauthor, among a total of five co-authors per article, was not responsible for the fabrication of data and/or results and that the co-author most responsible for data and results would have a higher author ranking in articles published. Data simulation is common when developing methods for detecting anomalies that might indicate cheating or fraud [19]. The dependent variable was the number of anomalies associated with each of 30 research articles. The number of anomalies was set to range between 0 and 40 for each article. The responsible co-author was assigned a higher number of anomalies for higher rank of authorship while the opposite was done for the other co-author. The simulated data used for the analyses are presented in Table 1. The question is whether data that these types of patterns had might show significant statistical results that might help identify responsible and non-responsible authors regarding fabrication of data or results. Table 1 assumes that each of 30 articles had five

co-authors, of which one co-author was primarily responsible for fraudulent data associated with anomalies. Row one of the table presents hypothetical data when the first author was responsible for fabricating data/results while the other co-authors were not directly responsible. As such, it would be expected that those six articles would feature a higher number of anomalies and that the other co-authors would have lower rankings in terms of author contributions/importance. The fifth row would represent when the author guilty of fabrication in other articles was a minor contributor to these six articles and thus the articles had few anomalies and other co-authors had higher author rankings in terms of importance. Rank of authorship would run from one (greatest contributing author) to five (least contributing author) so that low numbers represent higher rankings in terms of contributions/importance.

Analyses

Pearson zero-order correlations and nonparametric Spearman rho correlations were both used to predict anomalies from authorship levels for responsible and non-responsible authors, that

is, for those authors who were responsible or not for the anomalies. One-way analyses of variance were used to detect patterns of mean scores for anomalies as a function of level of authorship, using five levels of authorship from 1 (highest rank) to 5 (lowest rank). SPSS was used for all computations [20]. The anomaly scores were assumed to have been derived with equal weighting but in practice different weights might be assigned to those anomalies deemed more or less important. If actual data were available, one might predict anomalies not only from authorship levels but from other factors (age, gender, race, academic rank, etc.) in a multiple regression model.

Results

Correlations

From the data in Table 1, the zero-order Pearson correlation between number of anomalies and responsible authorship was -.814 (p<.001) while the Spearman rho was -.848 (p<.001). For the non-responsible author, the respective correlations were .555 (p<.001) and .616 (p<.001).

Table 1: Simulated data from 30 research reports for detecting co-authors responsible for fabricated data or results.

Rank of Responsible Co-Author	Rank of Other Co-Author	Number of Anomalies
1	3/4/5/5/1/3	15/21/27/32/40/22
2	1/3/4/4/4/4	6/13/22/29/31/16
3	1/4/4/5/2/1	2/5/23/17/19/10
4	1/2/2/3/5/2	1/3/5/2/3/5
5	3/4/1/2/2/2	1/2/1/3/2/1

Analyses of Variance

Table 2 presents the results for the analyses of variance. While both analyses of variance presented significant results, results were stronger for the responsible author, which may suggest that the process is better for identifying the “guilty” rather than clearing the “innocent”. While the assumption of homogeneity of variance

was violated in the analysis for the responsible authors, the robust test of equality of means remained significant (p<.001). The overall mean score and standard deviation for the anomalies for both responsible and non-responsible authors were 12.63 and 11.54, respectively, indicating that differences in anomalies would not have been identifiable without evaluating the rank of authorship.

Table 2: One-way analyses of variance predicting anomalies from rank of authorship for responsible and non-responsible authors. aFour tests of homogeneity of variance (based on mean, median, median with adjusted degrees of freedom, and trimmed mean) were all significant (p<.01); the Welch robust test of equality of means was 16.29 with 4, 10.97 df (p<.001). bFour tests of homogeneity of variance (based on mean, median, median with adjusted degrees of freedom, and trimmed mean) were all non-significant (p>.175); the Welch robust test of equality of means was 4.55 with 4, 11.34 df (p=.020).

Type of Author	Rank of Author	Mean Anomalies	SD	F	df	p
Responsible	1	26.17	8.89	13.51 ^a	4, 25	< .001
	2	19.50	9.65			
	3	12.67	8.31			
	4	3.17	1.6			
	5	1.67	0.82			
Not Responsible	1	3.20	3.83	3.86 ^b	4, 25	< .001
	2	5.67	6.77			
	3	11.00	8.46			
	4	21.00	12.1			
	5	17.00	11.54			

a. Four tests of homogeneity of variance (based on mean, median, median with adjusted degrees of freedom, and trimmed mean) were all significant ($p < .01$); the Welch robust test of equality of means was 16.29 with 4, 10.97 df ($p < .001$).

b. Four tests of homogeneity of variance (based on mean, median, median with adjusted degrees of freedom, and trimmed mean) were all non-significant ($p > .175$); the Welch robust test of equality of means was 4.55 with 4, 11.34 df ($p = .020$).

Further analysis

Since the data here were fabricated as part of this concept paper, could that fabrication be detected by the Benford's Law method [6]? Using the MAD method [6,21], with the data from Table 1, the percentages for the first through ninth digits were .333, .333, .167, .033, .100, .033, 0, 0, 0 compared to .301, .176, .125, .097, .079, .067, .058, .051, and .046 as predicted by Benford's Law, resulting in a total score of .056, well above the .02 criterion suggested by Banks [21]. Thus, there is external evidence that, as known a priori, the data in Table 1 were fabricated.

Discussion

It does not appear that this method has been used before in the literature on fraud detection; if by oversight it has been used elsewhere, it would bear repeating, given the importance of the issue. If scholars who fabricate their data continue to seek more credit for their "work" through a higher rank of authorship, then it should be possible to identify them using the correlation and analysis of variance techniques presented here if a measure of anomalies are available across a series of published articles or papers. However, if authorship rank is not correlated with levels of responsibility for preparation of articles, then this methodology might be less useful. In some areas of science, the last author may be the most responsible for the preparation of research reports. Even so, for larger number of articles, responsibility might be assessed as a binary variable—was a given scholar assigned or not as an author for each of the articles under review? If there are fewer articles under investigation, detection will be more difficult; correlations might be substantial but not significant because of small sample sizes. It might be more difficult to "clear" non-responsible authors because their roles might vary more than for the responsible author and because not every journal article authored by the responsible author will have the same coauthors (thus, the sample size for non-responsible authors will likely be smaller). This methodology will not work if authors publish their results without reporting numerical results that could be evaluated for anomalies. For example, authors should report means and standard deviations for their variables; but if they do not report standard deviations, then inconsistencies between binary means and standard deviations cannot be used to assess anomalies [14,15]. This methodology might be more limited if more than one author was primarily responsible for the scientific misconduct; however, if that were the situation, it would be more difficult for the authors to keep the others from eventually acknowledging their involvement. Another issue for consideration is that the religious faith of scholars could cut both ways; it might

mean a life of integrity for some scholars while for other scholars, it might be taken to mean that the end justifies the means, regardless of harm to others [22].

Conclusion

It may be possible to use statistics to identify which co-authors associated with published research found to have been fraudulent or fabricated were more responsible for the research misconduct. This by itself may deter future researchers from engaging in such research misconduct, realizing that now they might be exposed in a scientifically testable way. The use of statistics to investigate potential research misconduct is preferable to unsupported ad hominem attacks [23,24]. The penalties for publishing fabricated research might be severe for academics, including loss of promotion, position, or grant funding, even restrictions from being hired later by other universities. However, to be fair, penalties should be just and not based on anecdotes or rumors. Techniques such as those described here may form a more just basis for investigating and responding to alleged cases of scientific misconduct, especially with respect to fraudulent or fabricated research.

References

- Schumm WR, Pratt KK, Hartenstein JL, Jenkins BA, Johnson GA (2013) Determining statistical significance (alpha) and reporting statistical trends: controversies, issues, and facts. *Comprehensive Psychology* 2(10): 1-6.
- Schumm WR (2021) Confirmation bias and methodology in social science: an editorial. *Marriage & Family Review* 57(4): 285-293.
- Schumm WR, Dugan M, Nauman W, Sack B, Maldonado J, et al. (2021) Using free websites to perform statistical calculations in basic statistics courses at high school or college levels. *Sun Krist Sociology and Research Journal* 2(1): 1-7.
- Schumm WR, Palakuk CR, Crawford DW (2020) Forty years of confirmation bias in social sciences: two case studies of selective citations. *Internal Medicine Review* 6(4): 1-14.
- Schumm WR, Crawford DW, Fawver MM, Gray NK, Neiss ZM, et al. (2019) Statistical errors in major journals: two case studies used in a basic statistics class to assess understanding of applied statistics. *Psychology and Education-An Interdisciplinary Journal* 56(1-2): 35-42.
- Horton J, Kumar DK, Wood A (2020) Detecting academic fraud using benford's law: the case of professor james hunton. *Research Policy* 49(8): 104084.
- Fanelli D (2009) How many scientists fabricate and falsify research? a systematic review and meta-analysis of survey data. *Plos One* 4(5): e5738.
- Fanelli D (2013) Why growing retractions are (mostly) a good sign. *Plos Medicine* 10(12): e1001563.
- Steen RG, Casadevall A, Fang FC (2013) Why has the number of scientific retractions increased? *Plos One* 24(10): e68397.
- Stroebe W, Postmes T, Spears R (2012) Scientific misconduct and the myth of self-correction in science. *Perspect Psychol Sci* 7(6): 670-688.
- Stern AM, Casadevall A, Steen RG, Fang FC (2014) Financial costs and personal consequences of research misconduct resulting in retracted publications. *Elife* 3: e02956.
- Hussinger K, Pellens M (2019) Guilt by association: how scientific misconduct harms prior collaborators. *Research Policy* 48(2): 516-530.

13. Pickett JT (2020) The Stewart retractions: a quantitative and qualitative analysis. *Econ Journal Watch* 17(1): 152-190.
14. Schumm WR, Crawford DW, Lockett L (2019) Patterns of means and standard deviations with binary variables: a key to detecting fraudulent research. *Biomedical Journal of Scientific and Technical Research* 23(1): 171351-171353.
15. Schumm WR, Crawford DW, Lockett L (2019) Using statistics from binary variables to detect data anomalies, even Possibly fraudulent research. *Psychology Research and Applications* 1(4): 112-118.
16. Foo JYA, Tan XJA (2014) Analysis and implications of retraction period and coauthorship of fraudulent publications. *Account Res* 21(3): 198-210.
17. American Psychological Association (2010) *Publication Manual of the American Psychological Association*. (6th edn), American Psychological Association: Washington, DC, USA.
18. Benford F (1938) The law of anomalous numbers. *Proceedings of the American Philosophical Society* 78(4): 551-572.
19. Tendeiro Jn, Meijer Rr (2014) Detection of invalid test scores: the usefulness of simple nonparametric statistics. *Journal of Educational Measurement* 51(3): 239-259.
20. Aldrich JO (2018) *Using IBM SPSS Statistics: An Integrative Hands-On Approach*. Sage: Thousand Oaks, CA, USA.
21. Banks D (2000) Get M A D with Numbers! Moving Benford's Law from Art to Science.
22. Levin J (2022) Human flourishing in the era of covid-19: how spirituality and the faith sector help and hinder our collective response. *Challenges* 13(12): 1-8.
23. Schumm WR, Crawford DW, Lockett L, AlRashed A, Ateeq AB (2021) Nine ways to detect possible scientific misconduct in research with small (n<200) samples. *Psychology Research and Applications* 3(2): 29-40.
24. Holcombe AO (2021) Ad hominem rhetoric in scientific psychology. *Br J Psychol* 113(2): 434-454.