Assessing last decimal digits in retracted research articles: accounting for missing zeroes

Walter R. Schumm

Kansas State University, United States

Abstract

The cumulative nature of knowledge in science in general and for social science in particular depends on research being conducted with integrity, with fraudulent research being one of the most harmful of questionable research practices. It has been observed that, in some fabricated research, there has been a deficit of zeroes in the last digits of regression results. In research, authored by Dr. Eric Stewart, that has been retracted, there appears to have been a paucity of zeroes as last digits in his regression results. Eight hypotheses were tested to re-examine the relative lack of zeroes and possible increases in ones or nines as last digits. Some support was found for all hypotheses. It appears that in Stewart's retracted research, not only were zeroes underrepresented as terminal digits in regression coefficients, but that ones and nines were more frequently detected as terminal digits, in roughly equal proportions. Furthermore, the increased proportions of ones and nines was found to be highly correlated with other anomalies in the same retracted articles, possibly indicative of fabrication of the results in the retracted articles. Further research in this area might use symbolic interaction theory or justification theory to guide our developing knowledge about this type of questionable research practice.

Keywords: Scientific misconduct, fraud, regression digits, zeroes, Eric Stewart, fabrication of results

1. Introduction

Increases in knowledge, in science and in subdisciplines such as social science, depend on the accumulation of research results that have been obtained through ethical research practices. However, retractions of scientific journal articles – due to scientific misconduct or questionable research practices (QRPs) - appear to have increased over the past two decades [1, 2, 3]. Scientific misconduct may include data fabrication, data falsification, image manipulation, and plagiarism" [2: 1]. Reisig, Holtfreter, and Berzofsky [4] found that data fabrication was seen by a sample of academics as the least common form of such misconduct, but possibly as among the most consequential. It may be fortunate that authentic and ethical scientific results are a complex mixture of systematic variance and random error, a mixture that is hard to fabricate. Too much in the way of convenient results or too little random error may both be clues regarding data fabrication. With respect to random error, it has been argued that the last decimal digits in regression results are essentially random [5, 6, 7] with zeroes often being omitted as last digits in fabricated research.

In particular, suppose we have a regression coefficient of 0.140. By itself, the last number zero (or any other number as a last digit) is essentially a random number. Therefore, the relative proportion of the numbers zero through nine should be approximately ten percent. By itself, the number zero in that regression coefficient means little substantively. But if in a whole ensemble of last digits of regression coefficients, one were to find a statistically significant surplus or deficit of zeroes (or any other number), that might be an indicator of data fabrication. Benford's law can be used to evaluate the validity of the first digit numbers in regression coefficients because those patterns are *not* random [8, 9]. Those who would commit scientific fraud by reporting fabricated numbers have the nearly impossible challenge of conforming their *first* digits to the *non-random* patterns of Benford's law and, at the same time, their *last* digits to *random* patterns.

Pickett [7] found that several articles, eventually retracted, by Professor Eric A. Stewart, seemed to have non-uniform distributions for their last digits, with fewer zeroes and more one's or nine's. When asked about this, Dr. Stewart argued that he had rounded up or down when he found a last digit of zero [7]. Pickett [7: 157] asserted that Dr. Stewart explained to him that he had taken regression coefficients such as .000007 and rounded them up to .007, a substantial change, multiplication by one thousand! However, rounding up a last digit of zero to one nine would be easier than multiplying last digits by thousands. Pickett [7: 158] presented figures from five of Stewart's

retracted articles that showed low percentages of terminal digit zeroes and – higher than for any other digits – high percentages of one's and nine's as terminal digits. His chi-squared test with nine degrees of freedom for variation across all ten terminal digits was 252.07 (p < .00001). However, Pickett [7] did not test the 1s and 9s as last digits specifically.

A variety of anomalies, including violations of Benford's law, have been detected in research authored by Dr. Stewart [7, 10]. Both Pickett [7] and Bresnau [11] have accused Dr. Stewart of serial fabrication of results. Savolainen [12] has described the Stewart situation as "the most embarrassing scandal in the history of American criminology" [12: 10] and that "Although the retraction notes published in the journals suggest the errors were due to mistakes, fabrication remains the most parsimonious explanation of the documented irregularities" [12: 10]. Savolainen remarks with insight that "Studies that are not based on reality as it was actually observed should be retracted because they do not contribute to the growth of knowledge but detract from it" [12: 2]. In March 2023, Florida State University removed Dr. Stewart from his position due to numerous allegations of scientific misconduct, but an appeal process appears to be continuing.

2. Hypotheses

Here, the focus will be on the "last digit" problem in research that some have claimed to have been fabricated [7, 11]. Breznau [11: 2] asserted that Dr. Stewart was "sociology's first serial data and results faker that we know of." Pickett's [7: 158] report presented diagrams that featured elevated percentages of last digits of ones and nines, but he did not statistically test that particular pattern of data (ones and nines versus zeroes) in the same way done here.

If Dr. Stewart did avoid zeroes as last digits and rounded up/down to one's or nine's, those patterns should be detectable statistically. First, the proportion of zeroes as last digits should be significantly lower than 10% while the proportions of one's and nine's should be higher than 10%. Since some one and nines will occur by chance, addition of other ones and nines as substitutes for zeroes, should increase the total proportions of ones and nines beyond the expected 10% levels. Furthermore, the proportion of ones or nines should be greater than 20% while the proportions of zeroes, ones, or nines should be approximately 30%. If the normal proportion of zeroes, ones, and nines as last digits is about 30%, even if some zeroes are converted to ones or nines, the total proportions of all three last

digits should remain at about 30%. If the proportion of ones and nines, each, is expected to exceed

10%, then the total proportion of ones and nines should exceed 20%. If the substitution of ones or nines

for zeroes is random, then the proportion of ones or nines should be equivalent. Finally, the

proportions of ones and of nines, each, should be greater than that of zeroes.

 Eight hypotheses can be derived from the above expected patterns.

 H1: The proportion of last digit zeroes will be significantly lower than the otherwise expected

 level of 10%.

 H2: The proportion of last digit ones will be significantly higher than the otherwise expected

 level of 10%.

 H3: The proportion of last digit nines will be significantly higher than the otherwise expected

 level of 10%.

 H4: The proportion of both last digits one and nine will be significantly higher than the

 otherwise expected level of 20%.

 H5: The proportion of last digits of zero, one, and nine will not be significantly different than

 the expected level of 30%.   In other words, the null hypothesis will not be rejected.

 H6:   The proportions of last digits of one and nine will not be significantly different, if they

 have been randomly assigned in place of last digits of zero. The null hypothesis will not

 be rejected.

 H7: The proportion of one's as last digits will be significantly greater than the proportion of

 zeroes as last digits.

 H8:   The proportion of nine's as last digits will be significantly greater than the proportion of

 zeroes as last digits.

3. Methods

 Data from seven articles by Dr. Stewart and others that have been either corrected [13] or retracted

[14-19] were analyzed. The dates of these published articles range from 2003 to 2019. Regression

tables were examined for both regression coefficients and standard errors (or equivalent results). The

last digits were assessed for the number of zeroes, ones, and nines and the proportions of those last

digits as a function of the total number of last digits available for each article (Table 1).

Table 1. Presentation of Raw Data and Test Results for Eight Hypotheses Related to Last Digits from Regression Tables in Seven Retracted Articles

Article	0	1	9	H1: 0 < 10% (a)	H2: 1 > 10% (a)	H3: 9 > 10% (a)	H4: 1 or 9 > 20% (a)	H5: 0, 1, or 9 9 vs. 30% (a)	H6: 1 = 9 (b)	H7: 1 > 0 (b)	H8: 9 > 0 (b)
Stewart, 2003	0/75 (.000)	9/75 (.120)	21/75 (.280)	-2.89 (c) .0039 .00 - .00	0.58 .5637 .05 - .19	5.20 < .0001 .18 - .38	4.33 < .0001 .29 - .51	1.89 .059 .29 - .51	2.45 .0143	3.09 .0020	4.94 < .0001
Johnson et al., 2011	3/150 (.020)	30/150 (.200)	13/150 (.087)	-3.27 .0011 .00 - .04	4.08 < .0001 .14 - .26	-0.54 .5862 .04 - .13	2.65 .0079 .21 - .36	0.179 .8579 .23 - .38	2.80 .0051	4.98 < .0001	2.57 .0102
Stewart et al., 2015	2/114 (.018)	17/114 (.149)	24/114 (.211)	-2.93 .0033 .00 - .04	1.75 .0806 .08 - .21	3.93 < .0001 .14 - .29	4.26 < .0001 .27 - .45	1.79 .0721 .29 - .47	-1.21 .2263	3.59 < .0004	4.58 < .0001
Mears et al., 2017	3/108 (.028)	11/108 (.102)	6/108 (.056)	-2.50 .0124 .00 - .06	0.06 .9489 .04 - .16	-1.54 .1237 .01 - .10	-1.11 .2685 .09 - .23	-2.60 .0092 .11 - .26	1.26 .2077	2.21 .0271	1.02 .3077
Stewart et al., 2018	1/524 (.002)	66/524 (.126)	79/524 (.151)	-7.48 < .0001 .00 - .01	1.98 .0477 .10 - .15	3.87 .0001 .12 - .18	4.39 < .0001 .24 - .32	-1.07 .2857 .24 - .32	-1.16 .2460	8.21 < .0001	9.07 < .0001
Stewart et al., 2019	0/364 (.000)	58/364 (.159)	54/364 (.148)	-6.36 < .0001 .00 - .00	3.77 < .0002 .12 - .20	3.07 .0021 .11 - .18	5.14 < .0001 .26 - .36	0.32 .7488 .26 - .36	0.41 .6818	7.94 < .0001	7.64 < .0001
Total of 6	9/1335 (.007)	191/1335 (.143)	197/ 1335 (.148)	-11.36 < .0001 .00 - .01	5.25 < .0001 .12 - .16	5.79 < .0001 .13 - .17	8.27 < .0001 .27 - .32	-0.21 .8344 .27 - .32	-0.33 .7414	13.38 < .0001	13.64 < .0001
Mears et al., 2019	40/332 (.120)	44/332 (.133)	23/332 (.069)	1.25 .2131 .09 - .16	1.98 .0482 .10 - .17	-1.87 .0620 .04 - .10	0.08 .9343 .16 - .25	0.87 .3755 .27 - .37	2.71 .0067	0.47 .6384	-2.25 .0244
Total of 7	49/1667 (.029)	235/1667 (.141)	220/ 1667 (.132)	-9.61 < .0001 .02 - .04	5.58 < .0001 .12 - .16	4.35 < .0001 .12 - .15	7.45 < .0001 .25 - .29	0.21 .8344 .27 - .32	-0.76 .4473	11.54 < .0001	10.87 < .0001

(a)   One sample proportion z- test [www.statology.org/one-proportion-z-test-calculator]

(b)   Z score for two population proportions [www.socscistatistics.com/tests/ztest/default2.aspx]

(c)   Results within each cell represent z score, two-tailed significance level, and 95% Confidence Interval

One sample proportion z tests, with significance levels and 95% confidence intervals, were used to

compare raw data to expected proportions for testing hypotheses one through five. Z scores, with

significance levels, for two population proportions were used to test hypotheses six through eight. Results were assessed for each article separately and for totals from the articles. One article, by visual inspection, did not feature an unusually low percentage of zeros as last digits. However, this article was included in the analysis as evidence that not all of the retracted articles featured that type of anomaly and to show how the tests would perform in that type of more conventional situation. However, totals were assessed for both the six articles with an apparent zeroes anomaly and for all seven articles together (Table 1). Two-tailed tests of significance were used, even though only hypotheses five and six were two-tailed. Previous analyses of last digits have used chi-squared tests [7] or binomial tests [10], supplementing previous research with a new approach for this type of problem. Seven of Stewart's articles were assessed rather than five as used by Pickett [7].

4 Results

Results are presented in Table 1. With respect to hypothesis 1, for the six articles with apparent deficits of zeroes as last digits, the percentage of zeroes ranged between 0% and 2.8%; the z tests consistently yielded significant results, indicating that the results – of less than ten percent - were not likely due to chance. The same results occurred when combining the data for either six or seven articles. For hypothesis 2, for all of the articles and the totals, the percentages of ones as last digits always exceeded ten percent; however, the results were significant only for four of the articles and both totals, although for a fifth article the results were significant by a one-tailed test (which would be appropriate because our hypothesis was directional, expecting a result greater than ten percent). For hypothesis 3, the results were significant only for four of the articles and both totals. For a majority of the data, the percentages of last digits with either a one or a nine exceeded ten percent, as predicted. With respect to hypothesis 4, the results were significant for five of the articles and both totals, indicating that for the majority of the data, the combined result for last digits of either one or nine exceeded twenty percent. For hypothesis 5, where the expected result was "not different than 30%", and two-tailed tests were appropriate, only one result was significant, including for the two totals, indicating that the shortfall in the percentages for zeroes as last digits seemed to be "made up" for by an excess of last digits of one or nine.

For hypothesis 6, four of the results for the articles and for both totals, the results were not

significant, as predicted. Thus, in general, the increases in last digits of one or nine were approximately equal, as expected. For hypothesis 7, results for six of the articles and for both totals were significant. For hypothesis 8, results for five of the articles and for both totals were significant, indicating that the proportions of ones as last digits exceeded that of zeroes as last digits.

After testing the eight hypotheses, we used the anomaly severity scores for the anomalies associated with the retracted articles [10], minus the contribution of the zeroes anomalies, and correlated the sum of the proportions for the ones and nines, obtaining r = .79 (d = 2.58, p = .035, two-tailed). Despite the very small sample size, there was a significant zero-order correlation, with a very large effect size, between higher rates of reporting ones or nines in the last digits of regression tables and other anomalies in the same articles. Thus, the anomaly of higher excesses of ones or nines in the last digits of regression table results appears to be associated with other anomalies in the same articles. While one could argue that artificially manipulating last digits is only a small type of research misconduct, it does appear to be correlated substantially and significantly with other anomalies associated with more substantial questionable research practices.

5. Discussion

Results clearly supported our expectations for the first hypothesis. For six of the seven retracted articles, results indicated a deficit of zeroes as last digits in the regression tables. This result suggests that, for reasons that remain unclear, there was a tendency to avoid the appearance of zeroes as last digits in regression tables for six of the retracted articles. The seventh of Dr. Stewart's articles indicates that this tendency was not automatic nor found in all of Dr. Stewart's published research.

For the second hypothesis, the percentage of ones as last digits always exceeded ten percent, though the results were significant only for five of the articles and for both totals. For the third hypothesis, the percentage of nines as last digits exceeded ten percent and were significant statistically for only four of the articles and for both totals. For the fourth hypothesis, the combination of ones and nines as last digits significantly exceeded twenty percent only for five of the articles and both totals. Only for one article was the combination of zeroes, ones, and nines significantly different from thirty percent.

For four of the articles and for both totals, the percentages of ones and nines as last digits were equivalent, with respect to hypothesis six. With respect to hypothesis seven, for six of the articles and

for both totals, the percentage of ones as last digits exceeded the percentage of zeroes as last digits. For hypothesis eight, five of the articles and both totals featured significantly greater percentages of nines as last digits compared to the percentage of zeroes.

For six of the retracted papers, it seems clear that zeroes as last digits were underrepresented in general (versus 10%) and with respect to the percentages for ones and nines as last digits. In general, the overrepresentation of ones and nines led to an excess beyond twenty percent, as expected but results were stronger for the combination of all three last digits being approximately thirty percent as expected. That last result suggests that the substitution of digits for zeroes was generally restricted to ones and nines rather than other possible numbers (e.g., two through eight).

The strong and significant correlation observed between the summed proportion of ones and nines as last digits and other anomalies (unrelated to low levels of zeroes as last digits) in the seven retracted articles would seem to indicate that the rates of ones and nines are part of the larger picture of unusual anomalies in the retracted articles. The question that remains unanswered is whether the results were simply made up without any data analysis or whether some type of data was used and then the results were modified, including changing last digit zeroes to ones or nines much of the time. Perhaps the results were made up and then observed coefficients that ended in zero were modified to ones or nines most of the time, a combination of different approaches to creating results. It is possible that the data were fabricated and once an excess of ones and nines was reported [7], Stewart made up a reason for how that had happened, on the theory that confession of a smaller error was preferable to confessing a larger error (i.e., wholesale fabrication of data or results). Even if genuine data were used, modifying results by rounding up or down from zero to one or nine would represent data manipulation by itself and not be a valid scientific practice.

It does appear that in the retracted articles studied here there was a tendency to avoid zeroes as last digits in regression tables and to replace those zeroes, not with just any numbers, but with ones or nines. Overrepresentation of ones or nines as last digits was associated with greater numbers of other types of anomalies detected previously in the seven retracted articles.

It is not clear why this apparent approach might have been used, although the unusual characteristics of the number zero may, perhaps subconsciously, bother some of those who commit academic fraud. If the data were fraudulent, made up out of thin air, so to speak, was it feared that

zeroes would look unusual and might give away the fraud?   If it was thought that replacing zeroes as last digits with ones or nines would obscure the fraud, that approach would seem to have worked until fraud was detected and reported. The results confirm the difficulty of faking data and provide further evidence of ways to compare data to expected patterns to help detect potential scientific fraud.

6. Limitations and Directions for Future Research

We examined only articles that were retracted or corrected and only seven articles in total. However, a majority of the results were confirmed on an article-by-article basis, despite the small sample of articles. Some of the results obtained might have been due to random chance, even when the results were statistically significant, as remains true for all statistical testing. Future research might well examine the entire corpus of research by Dr. Stewart or other co-authors to determine if similar patterns would be found [20]. Other types of anomalies seen in the retracted articles [7, 10] should be investigated further. According to some scholars, further emphasis should be placed on appropriate consequences for scholars found responsible for scientific misconduct [2], but such consequences might backfire by discouraging authors from acknowledging needed corrections or retractions.

Further work needs to be done to reduce scientific misconduct in general and with respect to sociology in particular [11]. Symbolic interaction theory [21] or justification theory [22] might be useful theoretical approaches to understanding the thinking of academics who fabricate research and/or how they justify their actions to themselves or others. Accurate and useful knowledge will continue to be gained in social sciences only if research is conducted with appropriate ethical and methodologically correct standards; fabrication of either data or results may severely hinder scientific development of knowledge about topics and issues in social science; as noted "Studies that are not based on reality…. do not contribute to the growth of knowledge but detract from it" [12: 2]. It can be hoped that awareness that fabricated results can more easily be detected with modern techniques will deter future academics from engaging in questionable research practices of all kinds, including but not inclusive of fabrication of data or results. If so, the ability of science to accumulate accurate and valid knowledge in social science will be enhanced greatly.

References

[1] Fanelli, D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS ONE, 2009, 4, e5738.

[2] Hu, G., & Li, S. B. Why research retraction due to misconduct should be stigmatized. Publications, 2023, 11, 18, 1-2.

[3] Fanelli, D. Why growing retractions are (mostly) a good sign. PLoS Medicine, 2013, 10, e1001563.

[4] Reisig, M. D., Holtfreter, K., & Bersofsky, M. E. Assessing the perceived prevalence of research fraud among faculty at research-intensive universities in the USA. Accountability in Research, 2020, 27, 457-475.

[5] Mosimann, J. E., Wiseman, C. V., & Edelman, R. E. Data fabrication: Can people generate random digits? Accountability in Research, 1995, 4, 31-55.

[6] Mosimann, J. E., Dahlberg, J. E., Davidian, N. M., & Krueger, J. W. Terminal digits and the examination of questioned data. Accountability in Research, 2002, 9, 75-92.

[7] Pickett, J. T. The Stewart retractions: A quantitative and qualitative analysis. Econ Watch Journal, 2020, 17, 152-190.

[8] Eckhartt, G. M., & Ruxton, G. D. Investigating and preventing scientific misconduct using Benford's Law. Research Integrity & Peer Review, 2023, 8, 1, 1-10.

[9] Horton, J., Kumar, D. K., Wood, A. Detecting academic fraud using Benford's law: The case of professor James Hunton. Research Policy, 2020, 49(8), 104084.

[10] Schumm, W. R., Crawford, D. W., Lockett, L., Ateeq, B. A., & AlRashed, A. Can retracted social science articles be distinguished from non-retracted articles by some of the same authors, using Benford's Law or other statistical methods? Publications, 2023, 11, 14, 1-13.

[11] Breznau, N. Does sociology need open science? Societies, 2021, 11, 9, 1-25.

[12] Savolainen, J. Unequal treatment under the flaw: race, crime, and retractions. Current Psychology, 2023, advance online.

[13] Mears, D. P., Stewart, E. A., Warren, P. Y., & Simons, R. L. Culture and formal social control: The effect of the code of the street on police and court decision-making. Justice Quarterly, 2017, 34, 217-247.

[14] Johnson, B. D., Stewart, E. A., Pickett, J., & Gertz, M. Ethnic threat and social control: Examining public support for judicial use of ethnicity in punishment. Criminology, 2011, 49, 401-441.

[15] Mears, D. P., Stewart, E. A., Warren, P. Y., Craig, M. O., & Arnio, A. N. A legacy of lynchings: perceived criminal threat among Whites. Law & Society Review, 2019, 53, 487-517.

[16] Stewart, E. A. School social bonds, school climate, and school misbehavior: A multilevel analysis. Justice Quarterly, 2003, 20, 575-604.

[17] Stewart, E. A., Johnson, B. D., Warren, P. Y., Rosario, J. L., & Hughes, C.   The social context of criminal threat, victim race, and Punitive Black and Latino sentiment. Social Problems, 2019, 66, 194-221.

[18] Stewart, E. A., Martinez, R., Jr., Bamer, E. P., & Gertz, M. The social context of Latino threat and punitive Latino sentiment. Social Problems, 2015, 62, 68-92.

[19] Stewart, E. A., Mears, D. P., Warren, P. Y., Baumer, E. P., & Arnio, A. N. Lynchings, racial threat, and Whites' punitive views toward Blacks. Criminology, 2018, 56, 455-480.

[20] Schumm, W. R., Crawford, D. W., Lockett, L., AlRashed, A., & bin Ateeq, A. Research anomalies in criminology: How serious? How extensive over time? And who was responsible? Accountability in Research, 2023, in press. Doi: 10.1080/08989621.2023.2241127

[21] Carter, M. J., & Fuller, C. Symbolic interactionism. Sociopedia, 2015, 1(1), 1-17.

[22] Warner, C. T., & Olson, T. D. Another view of family conflict and family wholeness. Family Relations, 1981, 30(4), 493-503.