Iran J Public Health, Vol. 45, No.7, Jul 2016, pp.897-904

Original Article

Assessing the Prevalence of Publication Misconduct among Iranian Authors Using a Double List Experiment

Maryam HADJI [1], ****Fariba ASGHARI*** [2], ***Masoud YUNESIAN*** [3], ***Payam KABIRI*** [1], ***Akbar FOTOUHI*** [1]

1. *Dept. of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran*
2. *Medical Ethics and History of Medicine Research Center, Tehran University of Medical Sciences, Tehran, Iran*
3. *Dept. of Environmental Health Engineering, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran*

***Corresponding Author:** Email: fasghari@tums.ac.ir

(Received 14 Sep 2015; accepted 11 Jan 2016)

Abstract

Background: This study was done to determine the prevalence of publication misconduct among Iranian authors.

Methods: Data were collected through an email survey of corresponding authors of papers published in Iranian journals indexed in Scopus during 2009-2011. Using the double list experiment, these individuals were indirectly questioned about committing one of the five misconducts including duplicate publication, falsification, guest authorship, plagiarism, and fabrication over the past year.

Results: The survey was sent to 2321 individuals; 100 emails bounced, and of the remaining, 813 (36.60%) people responded to the questions. The prevalence rates were 4.15% for fabrication, 4.90% for plagiarism, 18.10% for guest authorship, 12.65% for falsification of the study methods, and -5.40% for duplicate publication. Among respondent 56.50% trusted the method and confidentiality of the survey and 6.50% did not trust the method or confidentiality at all.

Conclusion: We found that the double list experiment method is simple and reliable for use in the academic community, and it can be conducted easily in an e-survey. According to our results, the most common misconducts among Iranian authors are guest authorship and falsification of the methodology. In light of the negative and maleficent impact of publication misconduct in the scientific society, we recommend raising awareness and educating authors and investigators in this regard. To determine the accuracy of the method used in this study, further studies on publication misconduct using a control group and direct questioning, as well as other indirect methods are suggested.

Keywords: Publication misconduct, Double list experiment, Unmatched count technique, Iran

Introduction

Research misconduct, including fabrication, falsification, guest authorship, and plagiarism is one of the unethical behaviors in scientific research and has been of interest to the scientific society in recent years (1, 2). The increased rate of various types of publication misconduct since the late 20th century has been a cause of raised concern (3, 4). Studies show that the percentage of scientific paper retracted due to fraud has been increasing (5). Increased publication misconduct and lack of academic honesty can encumber the production of new knowledge, waste resources, weaken public trust, and reduce motivation to participate in research and surveys (6). Publication misconduct not only jeopardizes the author's reputation but can also create issues for the affiliated organization (7). An accurate estimate of publication misconduct is necessary to understand the problem and plan solutions. To estimate publication misconduct, studies use indirect methods to reduce

bias caused by withholding the truth. Most studies have estimated professional misconduct through anonymous questionnaires. Self-administrated questionnaires tend to be less costly and their data collection is less time-consuming. More importantly, their anonymity helps avoid response bias, but they are associated with an increased non-response rate, especially to sensitive questions. In addition to lower response rates, respondents may withhold the truth or provide false responses (8). A possible solution to this issue is using indirect methods in which trust is increased through randomization and aggregation, and chances of inaccurate or missing responses to sensitive questions are reduced. With aggregation, the sensitive question is placed in a list of non-sensitive questions, and respondents are asked how many of the items on the list apply to them. Since not all items apply to each individual, the respondents can rest assured that they will not be identified through their response to the sensitive question. Block total response is a common aggregation technique (9), and one of the special methods is the item count technique, the unmatched count technique (UCT), or the list experiment (10).

In the list experiment method, there are two lists of questions. The first one is a list of non-sensitive questions and the second one includes one sensitive question in addition to non-sensitive questions. These two lists are presented to two groups of participants; the baseline group (control) and the treatment group. Respondents only need to state the number of items that are true about them. If these two groups provide truthful responses to the questions, the rate of the sensitive question can be estimated. This estimate is the product of the difference between the average responses in the treatment and baseline groups. If the sample respondents are randomly chosen from the target population, and other sources of bias are negligible, an accurate and bias-free estimate can be achieved. To increase the sample size and reduce estimate error, the technique can be applied twice to the same sample using a different list of non-sensitive questions each time. This is known as the double list experiment (11).

The present study was conducted with the assumption that using anonymous research tools can increase participants' trust, reduce public acceptance bias, and thus reduce underestimation, and that indirect methods such as double list experiment can be an appropriate alternative to direct questionnaires. To the best of our knowledge, this is the first study of its kind to use the double list experiment, except for the study by Jann et al. who examined the prevalence of plagiarism among students (12). Our study objective was to use the double list experiment to assess the prevalence of publication misconduct among Iranian authors who have published in Iranian journals indexed in Scopus.

Materials and Methods

The target population of this investigation was Iranian authors who have published an article in a Scopus indexed Iranian journal between 2009 and 2011. The study was approved by the Institutional Review Board of Tehran University of Medical Sciences. Individuals' participation in the study and responding to questions was regarded as consent to participate in the study.

Questionnaire

In the double list experiment, two different lists of questions (lists A and B) are randomly distributed between two groups of participants. The first group acts as the Treatment group for list A and the Baseline group for list B. The second group acts as the Treatment group for list B and the Baseline group for list A (11). If the two groups have the same sample size, and samples are allocated to groups by random, the variance of a given estimate only depends on the variance of the two Baseline lists and the variance of the sensitive item. When the correlation between non-sensitive items in one list is negative, variance decreases, and when the correlation between non-sensitive items in two lists is positive, the inter-list covariance is increased.

The research tool in this study was anonymous questionnaires, which were developed based on the assumptions for the double list experiment model, the sensitivity of the issue, and the target population. First, a series of non-sensitive questions with an average "Yes" response of 0.5 and some sensitive questions concerning publication misconduct were suggested. In the questionnaires, each sensitive question was embedded within four non-sensitive questions, totally apart from the sensitive response choice item. The five sensitive questions used in this study were:

1. In the past year, I have submitted/published a paper I had published before exactly as it was, after translating it, or after minimal revisions, without asking permission from the first publisher.
2. In the past year, I have copied exact sentences (at least 5 sentences) from others' publications to write my paper.
3. In the past year, I have falsified or fabricated part of the research data/results for a paper I was writing.
4. In the past year, the research methods I described in at least one of my papers were not exactly true.
5. In the past year, my name has been listed, with my consent or to my request, as a co-author of a paper to which I did not contribute significantly (neither the study execution nor the preparation of the article).

The validity and reliability of the questions were tested. To determine the validity, 20 students and staff were asked to reply to non-sensitive questions. In addition, we consulted five experts in the field of ethics about the clarity of the questions, and used their feedback to make the final amendments. According to the respondents of this section, the clarity of the questions was 85.03%.

Next, we determined the reliability of the questions. For this purpose, we ran a Test Retest. Questions were directed at 40 epidemiology doctoral students twice, with a two week interval, and the intra class correlation (ICC) was determined as the reliability between them. The ICC was calculated for answers to each list (lists A and B), and they ranged between 0.988 (95% CI: 0.951, 0.997) to 0.656 (95% CI: -0.194, 0.901) (Table 1).

Table 1: Intra class correlation (ICC) and 95% confidence intervals (CI) of each series of questions on questionnaires 1 and 2

Series	Questionnaire 1	ICC (95% CI)	Questionnaire 2	ICC (95% CI)
1	List A1, including question on duplicate publication	0.917 (0.711, 0.976)	List A1, not including question on duplicate publication	0.757 (0.156, 0.930)
2	List A2, not including question on falsifying study methods	0.820 (0.330, 0.951)	List A2, including question on falsifying study methods	0.749 (0.130, 0.928)
3	List A3, including question on plagiarism	0.855 (0.462, 0.961)	List A3, not including question on plagiarism	0.942 (0.798, 0.983)
4	List B3, not including question on plagiarism	0.895 (0.609, 0.972)	List B3, including question on plagiarism	0.656 (- 0.194, 0.901)
5	List B2, including question on falsifying study methods	0.866 (0.504, 0.964)	List B2, not including question on falsifying study methods	0.685 (- 0.094, 0.909)
6	List A4, not including question on guest authorship	0.865 (0.498, 0.964)	List A4, including question on quest authorship	0.797 (0.296, 0.942)
7	List A5, including question on data fabrication	0.824 (0.291, 0.956)	List A5, not including question on data fabrication	0.905 (0.671, 0.973)
8	List B1, not including question on duplicate publication	0.942 (0.766, 0.986)	List B1, including question on duplicate publication	0.685 (- 0.094, 0.909)
9	List B4, including question on quest authorship	0.846 (0.381, 0.962)	List B4, not including question on quest authorship	0.904 (0.665, 0.972)
10	List B5, not including question on data fabrication	0.988 (0.951, 0.997)	List B5, including question on data fabrication	0.921 (0.726, 0.977)

Since any question regarding participants' demographics was likely to reduce their trust, no such questions were included. To measure the level of trust in the study method regarding confidentiality and difficulty, we included the following two questions at the conclusion of the questionnaire: "How did you find the instructions for completing the questionnaire?" and "How trustworthy do you find this method in terms of confidentiality of your responses?" The study questionnaire was used in a pilot through electronic survey services affiliated with the Tehran University of Medical Sciences.

Implementation: We searched for papers published between 2009 and 2011 in biomedical Iranian journals indexed in Scopus (approximately 50 journal titles). All those with a non-Iranian corresponding author were excluded from the list, and the final study population was 2321 corresponding authors with an address in Iran. After retrieving their email addresses, a description of the study and its objectives were sent to them, and they were invited to participate in the web-based study. Two reminders were sent at two-week intervals. Corresponding authors were randomly divided into two groups and one of the two study questionnaires was sent to them.

Statistical Analysis: In the double list experiment method, there are two question lists for each sensitive question (for instance, lists 1 and 8 for assessing duplicate publication), which we shall refer to as lists A and B from now on (Table 1). For each list, one group of the participants is the Baseline and the other is the Treatment group.

By selecting an equal sample size for each group, and using similar sensitive questions in both lists, $y_{Aik} = y_{Bik}$ for all participants, and the total of sensitive and non-sensitive questions in these two groups would be:

$$\tilde{y}Aik = \begin{Bmatrix} 1 \text{ if individual } i \text{ would honestly say yes to item } \mathcal{K} \text{ on the A list} \\ 0 \text{ if individual } i \text{ would honestly say no to item } \mathcal{K} \text{ on the A list} \end{Bmatrix}$$

$$\tilde{y}Bik = \begin{Bmatrix} 1 \text{ if individual } i \text{ would honestly say yes to item } \mathcal{K} \text{ on the B list} \\ 0 \text{ if individual } i \text{ would honestly say no to item } \mathcal{K} \text{ on the B list} \end{Bmatrix}$$

$$\tilde{y}^{K}_{Ai+} = \sum_{K=1}^{K} \tilde{y}\, Aik \qquad \tilde{y}^{K-1}_{Ai+} = \sum_{K=1}^{K-1} \tilde{y}\, Aik$$

$$\tilde{y}^{K}_{Bi+} = \sum_{K=1}^{K} \tilde{y}\, Bik \qquad \tilde{y}^{K-1}_{Bi+} = \sum_{K=1}^{K-1} \tilde{y}\, Bik$$

Thus, one group of participants responds to list A, not including the sensitive question, and forms the Baseline group for list A, and the same group responds to list B, including the sensitive question, and forms the Treatment group for list B. Questions for the other group are the reverse, i.e. for list B, they are the Baseline group, and for list A, they form the Treatment group. Assuming that participants understand the method and respond honestly, the difference in the proportion of "yes" responses for each list is the prevalence estimate of the sensitive question based on each given list. Since the number of respondents in both treatment groups A and B was the same, the total prevalence of each sensitive question is the mean of the two estimated prevalence rates calculated for lists A and B.

To determine the variance of the estimate, we added the variances of lists A and B in both groups, subtracted their covariance from the sum, and divided the results by the sample size. The square root of the calculated variance was used as the standard error to determine the 95% confidence intervals (13).

Results

The questionnaire was sent to 2321 corresponding authors in two groups (n=1160 and 1161); 100 emails were undeliverable, and 813 (36.60) of the remaining responded to study questions.

First, we determined the prevalence, average, and variance of publication misconduct based on lists A and B, and then, the prevalence rates of different types of publication misconduct were determined based on the double list experiment method. The prevalence of data fabrication was 4.15%, the rate for plagiarism was 4.90%, and rates for guest authorship, falsifying methods, and duplicate publication were 18.10%, 12.65%, and -5.40%, respectively (Table 2).

The instructions were very easy for 64.30% of respondents, 29.70% stated they were quite easy, 5.60% had some trouble responding, and 0.40% found it difficult to complete the survey. In terms of confidentiality, 56.50% of respondents trusted the method and confidentiality of their responses, 34.50% stated their level of trust was average, and 6.50% did not trust the method at all. This is while 2.50% had little trust in the confidentiality of their responses.

Table 2: Prevalence of publication misconduct using the double list experiment method

Variable	Prevalence	Standard Error	95% Confidence Interval
Duplicate Publication	-5.40	0.061	-5.81, 4.90
Data Fabrication	4.15	0.067	3.69, 4.60
Plagiarism	4.90	0.064	4.46, 5.33
Guest Authorship	18.10	0.056	17.72, 18.44
Falsified Methods	12.65	0.062	12.23, 13.06

Discussion

In this study, the prevalence of duplicate publication was -5.40%, the rate of data fabrication was 4.15%, and rates of plagiarism, guest authorship, and falsification of methods were 4.90%, 18.10%, and 12.65%, respectively. The most common publication misconduct was guest authorship and the least common one was duplicate publication.

The rate of plagiarism in this study was 4.90%. In a study concerning academic honesty among medical students of Tehran University of Medical Sciences using the Randomized Response Technique, the prevalence of plagiarism was reported 31% with a confidence interval from 18% to 44% (14). In another study the prevalence of plagiarism among UK academics engaged in biologic research was 4.2% (SE=10.8%) using UCT (15). In another study on 474 Swiss and German students using the double list experiment, the prevalence rates of relative and extensive plagiarism were 9% and -4%, respectively (12). McFarlin et al. published a report concerning the International Journal of Exercise (IJES) stating that 46% of submitted papers contained plagiarism; this was seen mostly in the introduction and methodology of the manuscript, and only 3% in the discussion and conclusion (16). In addition to true inter-population differences in the incidence of publication misconduct, such differences in the reported rates of plagiarism are due to differences in research methodology, target populations, and definitions of plagiarism.

Other misconducts we studied included fabrication and falsification, and the rates were 4.15% and 12.65%, respectively. According to a systematic review, 1.97% of authors (95% confidence interval: 0.86 to 4.45) had committed fabrication or falsification at least once (17). In a study, 27% of participating scholars and authors stated they had encountered papers that were suspicious in terms of falsification of methodology, data fabrication, or plagiarism (18). The rate of fabrication or falsification was 5.20%, and 22% of them were professors or senior scholars (19). In a study, the rates of fabrication and falsification were 20% (95% confidence interval: 8 to 32) and 39% (95% confidence interval: 26 to 52), respectively (14).

Based on a comparison between our study and theirs, the prevalence rates of 3 publication misconducts - plagiarism, data fabrication, and falsifying the methods – were higher using the Randomized Response Technique. The confidence intervals of these estimates were all greater than our results using the double list experiment. Considering the larger sample size and different methods, smaller confidence intervals were quite expected. The difference between our results and those by Mortaz Hejri et al. can be attributed to different target populations. In their study, the target population was medical students, but ours was Iranian authors. In addition, we only looked at publication misconduct and not all academic performances of the participants.

Guest authorship was the most common publication misconduct in this study. Wislar et al. studied the rate of guest authorship among corresponding authors publishing original articles, reviews and editorials in high impact factor journals such as Annals of Internal Medicine, JAMA, Lancet, Nature Medicine, New England Journal of Medicine, and PLoS Medicine in 2008 and compared results with the prevalence rate in 1996; the comparison indicated that the 2008 rate of 17.6% (95% confidence interval: 14.60 to 21.00) was not significantly different from that in 1996. Rates were 25% in original research articles, 15% in review articles, and 11.20% in editorials (20), which are quite comparable to our results.

The rate of duplicate publication, which is a type of redundant publication, was -5.4% (95% confidence interval: -4.90 to -5.81). While a negative prevalence is meaningless, the computation method involved subtraction to arrive at the final estimate, and when the prevalence rate for a sensitive question is close to zero, the prevalence can turn out negative. Another reason could be providing false responses; when respondents do not trust the study method, feel uncomfortable about sensitive questions, and fear being identified, underestimation occurs and results in a negative prevalence rate (12). However, the question regarding confidentiality indicated that the respondents trusted that their responses would remain confidential. One reason for the low prevalence rate could be the definition we presented for duplicate publication. Questions were designed based on a definition of extensive duplication, i.e. publishing the entire article again, while the true definition is serious overlapping material without citing the previous paper. In a Cochrane review, the rate of redundant publication in biomedical sciences, or publishing the same material with different data was 10-25% (21). To determine the prevalence of redundant publication, Susser et al. examined various journals and found the rate had increased from 5% in 1988 to about 12% in 1990 (22).

The present study demonstrated that the double list experiment is easy to comprehend, because it does not need additional resources for random selection. Furthermore, it can simply be implemented electronically, and thus, data collection can be done faster. In this study, 64.3% of our participants stated the questionnaire guidelines were very easy and 29.7% stated completing the questionnaire was quite simple. This is while in the study by Mortaz Hejri et al, 10% of the participants thought the guidelines were difficult, 45% thought they were average, and 44% stated they were simple (14). In addition, 56.5% expressed complete trust in the confidentiality of UCT, and 34.5% had average trust in the method, while only 29% of the participants had trusted the Randomized Response Technique in answering the questions (14).

One of the strong points of our study technique was that respondents did not have to respond to sensitive questions directly which increased their trust, the questionnaire was very easy to complete and no special tool or skill was needed, and it took only 10 minutes of the participants' time. A large sample size of 813 participants is another advantage of our study. Additionally, we tried to assess different types of publication misconduct to determine prevalence differences, while other studies addressed only one or two types of misconduct.

Our study had certain limitations as well. Firstly, the participation rate was 35%. Since 64% stated they had understood the methods, non-participation could be because they did not understand

how to complete the survey, and since 57% stated they trusted the study, another reason for non-participation could be lack of their trust in keeping their responses confidential. In any event, the high non-response rate limits our ability to generalize findings to the general population. In addition, considering the nature of sensitive questions and limitations of the study, we did not collect authors' demographics, and thus, we were not able to do a more detailed analysis.
Further studies are suggested using other indirect techniques such as the Crosswise Model, or the direct method with the same target population, so that the accuracy of the method is compared against other methods and differences attributable to double list experiment are determined.

Conclusion

Double list experiment is well trustable and simple method for the academic society and can be conducted through e-surveys. Other studies using indirect methods and direct methods with control groups are suggested to assess the accuracy of this method in estimating publication misconduct. The most common publication misconducts among Iranian authors are guest authorship and falsification of the study methods. In light of the negative and maleficent impact of publication misconduct in the scientific society, we recommend raising awareness and educating authors and investigators in this regard, or even adding a study unit in this regard to the curriculum.

Ethical considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Acknowledgments

This paper is the result of the Maryam Hadji's dissertation for MS degree in Epidemiology funded by Tehran University of Medical Sciences.

The authors declare that there is no conflict of interest.

References

1. Mavrinac M, Brumini G, Bilić-Zulle L, et al. (2010). Construction and Validation of Attitudes Toward Plagiarism Questionnare. *Croat Med J,* 51(3):195-201.
2. Mason PR (2009). Plagiarism in scientific publications. *J Infect Dev Ctries,* 3(1):1-4.
3. Brice J, Bligh J (2005). Author misconduct: not just the editors' responsibility. *Med Educ,* 39(1):83-89.
4. Mundt LA (2008). Perceptions of scientific misconduct among graduate allied health students relative to ethics education and gender. *J Allied Health,* 37(4):221-224.
5. Fang FC, Steen RG, Casadevall A (2012). Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci U S A,*109(42):17028-17033.
6. DuBois JM (2004). Is compliance a professional virtue of researchers? Reflections on promoting the responsible conduct of research. *Ethics Behav,* 14(4):383-395.
7. Kenkre JE, Semple M (2003). Research fraud: Why nurses should become aware. *J Res Nursing,* 8(1):47-56.
8. Dalton DR, Wimbush JC, Daily CM (1994). Using the unmatched count technique (UCT) to estimate base rates for sensitive behavior. *Pers Psychol,* 47(4):817-829.
9. Raghavarao D, Federer WT (1979). Block total response as an alternative to the randomized response method in surveys. *J R Stat Soc,* B 41(1):40-45.
10. Miller JD (1984). *A new survey technique for studying deviant behavior*. George Washington University.
11. Berinsky AJ (1999). The two faces of public opinion. *Am J Pol Sci*, pp.: 1209-1230.
12. Jann B, Krumpal I, Naeher AF (2011). Plagiarism in student papers: prevalence estimates using special techniques for sensitive questions. *Journal of Economics and Statistics,* 231(5-6):749-760.
13. Glynn AN. (2010). What can we learn with statistical truth serum? Design and analysis of the list experiment. available from: *http://isites.*

harvard. edu/fs/docs/icb. topic646669. files/StatTruthSerum. pdf [June 23 2015].

14. Mortaz Hejri S, Zendehdel K, Asghari F, et al. (2013). Academic disintegrity among medical students: a randomised response technique study. *Med Educ,* 47(2):144-153.
15. Roberts DL, St John FA (2014). Estimating the prevalence of researcher misconduct: a study of UK academics within biological sciences. *PeerJ,* 2:e562.
16. McFarlin BK, Lyons TS, Navalta JW (2010). Prevalence of Plagiarism in Manuscript Submissions and Solutions. *Int J Exerc Sci,* 3(3): 68–69.
17. Fanelli D (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics,* 90(3):891-904.
18. Dingell JD (1993). Misconduct in medical research. *N Engl J Med,* 328(22):1610-1615.
19. Orton C (2010). Concerns of Editors and Publishers: Plagiarism, Rights of Authors, Open Access, etc. Paper presented at: *World Congress on Medical Physics and Biomedical Engineering, September* 7-12, Munich, Germany.
20. Wislar JS, Flanagin A, Fontanarosa PB, et al. (2011). Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. *BMJ*, 25;343:d6128.
21. Jefferson T (1998). Redundant publication in biomedical sciences: Scientific misconduct or necessity? *Sci Eng Ethics,* 4(2):135-140.
22. Susser M, Yankauer A (1993). Prior, duplicate, repetitive, fragmented, and redundant publication and editorial decisions. *Am J Public Health,* 83(6):792-793.