Inflated Co-authorship Introduces Bias to Current Scientometric Indices

Izet Masic¹, Slobodan M. Jankovic¹,²

ABSTRACT

Background: Although scientometry gradually became prevalent way of measuring one's research output, there are many inherent drawbacks in main indices that are used: impact factor, number of citations, number of published papers and Hirsch's index. Objective: The aim of this study was to analyze effects of inflated co-authorship on values of scientometric indices among authors in biomedicine who participated in published papers with more than 30 co-authors. Methods: The study was of cross-sectional type, based on 100 publications randomly extracted from the MEDLINE database. The inclusion criterion was publication with more than 30 authors. The studies with topics not related to humans were excluded from further analysis. Results: On average about 10% of papers published by the surveyed authors had more than 30 co-authors, but these papers brought more than 40% of all citations and more than 40% of Hirsch's index attributed to these authors. The duration of scientific activity was well correlated to number of citations, Hirsch's index and the number of publications themselves with 30 or less co-authors, while the correlation did not exist with number of citations, Hirsch's index and the number of publications with more than 30 authors. In summary, publications with > 30 authors carry more scientometric points than publications with less co-authors, and the researchers with shorter scientific activity had larger scientometric benefit from publications with more than 30 authors than senior researchers. Conclusion: Unjustified and prolific co-authorship is one of methods for inflation of scientometric indices that are not further reflecting true quality of research output of an individual. Further improvement of scientometric indicators may prevent unjustified co-authorship if it reflects the work invested in a research result.

Keywords: co-authorship; scientometric indices; inflation.

1. BACKGROUND

Writing a paper is a tedious job, however, following the established rules that work not only becomes much simpler, but also more accessible, which often results in the birth of the desire of researchers to write an article (1-4). Knowledge of the principles established by the process of scientific research demystifies the process (2, 3).

Emphasizes the importance of pursuing the following five steps (1): identification of the main research questions, the selection of a scientific approach, study design, data collection, and data analysis and presentation of the work (1).

a) The first step in the process is exploring a variety of themes as the focus of research and has multiple segments, such as: choice of the major topics of research, literature review, focusing on the question of research, drafting support team;

b) The second step in the research is to select the main access study. Access can be: review or meta-analysis, correlation (ecological) studies, case series, cross-sectional studies, case control studies, cohort studies, experimental studies or qualitative studies;

c) The third step of the process of scientific research is the development and implementation of a detailed study plan. It is necessary to know how to create a protocol for primary, secondary and tertiary studies. Overview of developing proposals and flow. Primary studies require: a sample of the population, determination of the sample size, the development of questionnaires, surveys and interviews, additional assessment, ethical issues, ethical review and authorization. Secondary studies include existing data sets, and tertiary studies include a systematic review and meta-analysis;
The fourth step in the research is the collection and analysis of data collected in the third step. Most researches require descriptive or comparative statistics. This step includes: management of data, descriptive statistics, comparative statistics and advanced biostatistics; e) The fifth and final step in the process is writing a research report and preparation for presentation and publication. In this step is described the structure of the article, quote, writing strategies, critical review, posters and presentations, choosing journals for publication, the process of teaching, examination and publication of the work and why publish?

In order to discover something new, it should be: the average intelligence, the ability to analyze and synthesis, power of perception, desire, determination, creativity, ethics, responsibility and, most importantly, a pure intention to achieve a desired goal (5-7).

Scientific impact measures are increasingly being used for academic promotions, grant evaluations and evaluation of job vacancy of candidates (they are also being used for the evaluations of university departments and research centres). Based on the cited literature scientometric indicators can be used to analyze and evaluate the work of researchers, institutions, regions and countries. Scientometric indicators of work of an author, in addition to the number of citations which is a priority in the modern scientific community (1, 8-10).

Although scientometry gradually became prevalent way of measuring one's research output all over the world, there are many inherent drawbacks in main indices that are used: impact factor, number of citations, number of published papers and Hirsch's index (11). Various ways of inflating scientometric indices were revealed, including unjustified self-citation, black-mailing researchers submitting their work to a journal to cite papers of the reviewers, forming alliances with other researchers to cite their papers in return, auto-plagiarism, divided (“salami”) publications, unjustified co-authorship, and many others.

In countries without strong scientific tradition, relying only on scientometry when building educational system and trying to use science for development of the society, could have devastating consequences, creating spurious institutions and networks of individuals that will actually block true scientific activities (12).

One of the ways of abusing scientometry for personal promotion, that was increasingly used in medicine during the last decade, is inflating co-authorship to previously un-imaginable limits. In the most prestigious medical journals papers with several hundreds, and even a couple of thousands of co-authors, are increasingly published, rocketing their personal scientometric indices (8, 9, 13). If each of numerous co-authors cites such a paper in another publication at least once, in a couple of months the paper will receive several hundreds of citations just by auto citing. Over a few years, young authors become highly cited, with high Hirsch’s index, and their scientometric performance outweighs those of others who worked hard their whole professional life, without participating in such “endeavours” (14). This becomes a problem when the authors with inflated scientometric indices compete for research grants with colleagues that work alone or have limited number of co-authors who substantially contributed to the published work: the latter would be handicapped in the very beginning of the competition, since their scientometric indices, being the only method of judging scientific competence, will be lower (15). However, although there is a number of publications dealing with the problem of inflated co-authorship in the literature, true extent of this phenomenon and its consequences remain unknown.

Three widely used bibliometric databases for analysis and evaluations of citations and the h index are Web of Science (Thomson Reuters), Scopus (Elsevier), and Google Scholar (2). Although Google Scholar and Scopus seem to provide higher numbers of citations there is mixed information on the h index. Portal webometrics.info regularly provides lists in which they rank the authors based on the h index (taking into account the authors whose h index is over five).

All analysis are conducted on the basis of the profile on Google Scholar platform, where the profile must be public (the condition that the profile is public is that the profile is verified via mail of the institution of the author). Google Scholar project or platform Google Scholar enabled a comprehensive approach to the list of articles of an author, and allowed access to the number of citations of one article, and based on the information found on Google Scholar creation of many list is enabled, in order to rank authors in a certain field. The work of the authors of this article is also analyzed, and different values are obtained, and the fact is that Scopus is quite selective, but it is still not clear whether it is more valid.

The first of the major problems in the scientometric analysis of published papers in indexed biomedical journals is authorship. Details about authorship malversations we described in another articles deposited on Pubmed Central (5-7). The second one is method of verification of Google Scholar, for whose activation of profile is required mail address of the institution where the author works (3). However, one should take into account a number of researchers in later years, who are not interested in this kind of self-promotion, and lists that are made without them, are not valid and are not a realistic indicator of the development of a certain field in one state.

Manipulation are often possible when creating the profile, so a lot of inconsistencies in the information can be seen. Google Scholar collects information on internet, and from address that are not most credible, and as such, is accessible to manipulation of content. Emilio Delgado Lopez-Cozar, Nicolas Robinson-Garcia and Daniel Torres-Salinas made an experiment, they created six documents authored by a faked author and uploaded them to a researchers personal website under the University of Granadas domain.

The result of the experiment meant an increase of 774 citations in 129 papers, increasing the authors and journals H index. These experiments have proven the numerous shortcomings of the Google Scholar and
Inflated Co-authorship Introduces Bias to Current Scientometric Indices

displayed the possibilities of artificial increasing of H index and i10 index. Hyland found that self-citation is 12% of all references in biology, engineering and physics, compared to 4% in sociology, philosophy, linguistics, or marketing. The fact is that it is needed to consider self-citation when making various list according to scientometric data.

2. OBJECTIVE
The aim of this study was to analyze effects of inflated co-authorship on values of scientometric indices among authors in biomedicine who participated in published papers with more than 30 co-authors.

3. METHODS
The study was of cross-sectional type, based on publications indexed in MEDLINE database. The inclusion criteria were publications with more than 30 authors, and having among them the authors with the following surnames: Smith, Wang and Li (these surnames were used in the search strategy as a tool for identification of studies with large number of authors, because these are the most frequent surnames in U.S.A. and China, respectively). The studies with topics not related to humans were excluded from further analysis by means of pre-set filters.

With the following search strategy “Smith [author] AND Wang [author] AND Li [author]” in total 1632 papers were retrieved; 374 papers were excluded for having 30 or less authors, and further 394 papers were excluded due to their topics unrelated to humans. From the remaining 864 studies a simple random sample of 100 studies was chosen, using tabulation and random number generator in Excel. Furthermore, from each of the sampled studies one author was randomly chosen, again by tabulation and random number generator, creating final study sample of 100 researchers. If a chosen author

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value (frequency and percentage, or mean ± standard deviation, median, range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (m/f, %/%)</td>
<td>74/26 (74%/26%)</td>
</tr>
<tr>
<td>Duration of scientific activity (years)</td>
<td>24.1 ± 9.7; 24 (5 – 60)</td>
</tr>
<tr>
<td>Total number of publications per participant</td>
<td>434.0 ± 441.0; 310 (4 – 2141)</td>
</tr>
<tr>
<td>Number of publications per participant with more than 30 authors each</td>
<td>47.1 ± 103.1; 21.5 (1 – 954)</td>
</tr>
<tr>
<td>Total number of citations per participant</td>
<td>39,091.9 ± 46,930.8; 19,210.5 (750 – 247,234)</td>
</tr>
<tr>
<td>Number of citations of papers with more than 30 authors per participant</td>
<td>18,234.4 ± 27,265.7; 4,317.0 (0 – 122,544)</td>
</tr>
<tr>
<td>Hirsh’s index based on all publications of an author</td>
<td>68.4 ± 43.1; 60.0 (1 – 212)</td>
</tr>
<tr>
<td>Hirsh’s index based only on publications of a participant with less than 30 authors each</td>
<td>33.1 ± 29.1; 27.5 (1 – 106)</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of the study sample (n=100).

Figure 1. The relationship between the duration of scientific activity and number of publications (total [a] and only of those with ≤ 30 authors [•]).

Figure 2. The relationship between the duration of scientific activity and number of citations (total [a] and only of those with ≤ 30 authors [•]).

Figure 3. The relationship between the duration of scientific activity and Hirsh’s index (based on all publications [a] and on only of those with ≤ 30 authors [•]).
did not have her/his profile at Google Scholar, she or he was excluded and replaced with another author from the same study (again randomly chosen) who had the profile. From the Google Scholar profiles of the sampled authors the following data were extracted: sex of an author, duration of scientific activity (abstracting the year of the earliest publication from 2021), total number of publications, number of publications with more than 30 authors, total number of citations, number of citations of papers with more than 30 authors, Hirsh’s index based on all publications, Hirsh’s index based only on publications with ≤ 30 authors and number of authors in the sampled publications.

Minimal acceptable size of the study sample was calculated on the basis of following parameters: targeted statistical power of 0.8, 95% confidence intervals of ± 5, and standard deviation of 11, taken from the study of Thompson et al which analyzed publication metrics of pharmacy practice chairs in U.S.A. (16). The formula

\[ n = \frac{Z_{1-\alpha/2}^2 \cdot CI^2}{\sigma^2} \]

was used for calculation (17), where CI was target confidence interval and SD standard deviation observed in the study by Thompson et al. Minimal sample size sufficient for giving statistical power of 0.8 to this study was 74 researchers.

The extracted data were first tabulated, and then checked for normality of distribution by Kolmogorov-Smirnov’s test. The data were then described with measures of central tendency (mean) and variability (standard deviation and range). The differences between values of bibliographic measures taken under different circumstances from the same study participants were tested by the Wilcoxon Signed Rank test. Correlation between duration of scientific activity and various bibliographic measures were tested by calculation of Spearman’s non-parametric correlation coefficient. All calculations were made using Statistical Program for Social Sciences (SPSS) version 18.

4. RESULTS

Average number of authors per published paper used for identification of the study participants was 214.6 ± 289.3, median 162.0, range 38 – 2,135. In total 100 authors participating in at least one paper with more than 30 authors were included in the study. Detailed characteristics of the study sample are shown in Table 1.

Since none of the characteristics of the study sample followed normal distribution, they were compared among themselves by non-parametric Wilcoxon Signed Rank test (WSR), and existence of correlation was tested by non-parametric Spearman’s coefficient. While per study participant number of publications with ≤ 30 authors was significantly different from number of publications with > 30 authors (WSR = 8.576, \( p = 0.000 \)), and also number of citations of publications with more and less than (or equal to) 30 authors (WSR = -3.700, \( p = 0.003 \)), the Hirsh’s indices based on publications with more and less than (or equal to) 30 authors were not significantly different among themselves (WSR = 1.608, \( p = 0.453 \)).

The duration of scientific activity (in years) was strongly correlated with total number of publications (Spearman’s \( \rho = 0.534 \), \( p = 0.000 \)), but it was not correlated well with the number of publications with > 30 authors (Spearman’s \( \rho = 0.084 \), \( p = 0.408 \)). While the duration of scientific activity was correlated well with Hirsh’s index based only on papers with ≤ 30 authors (Spearman’s \( \rho = 0.476 \), \( p = 0.000 \)), the correlation was weak with the Hirsh’s index based only on papers with > 30 authors (Spearman’s \( \rho = 0.097 \), \( p = 0.336 \)). In accordance with these findings, the duration of scientific activity was strongly correlated with the number of citations based only on papers with ≤ 30 authors (Spearman’s \( \rho = 0.600 \), \( p = 0.000 \)), but the correlation was weak with the number of citations based only on papers with > 30 authors (Spearman’s \( \rho = 0.171 \), \( p = 0.089 \)).

The relationships between the duration of scientific activity and number of publications, Hirsh’s index and number of citations are shown in Figures 1 to 3, respectively.

5. DISCUSSION

The issue of authorship is an extremely complex issue, and present a place where many malversations and violations of ethical principles can be found. In order to make rapid progress and sometimes because of goals that are extremely difficult to fulfill, authors have precise agreement with their colleagues about signing each other, also they add their colleagues in authors list, and sometimes it is simply assumed that employees of the same department sign each other on articles. In scientific literature has described several of inappropriate types of authorship (7): a) guest authorship; b) honorary or gift authorship; c) ghost authorship; d) anonymous authorship; group authorship, etc. Also, we have a violation of all ethical norms of scientific publication. Falsely signing of statements on authors’ contributions of one author in journal, produce false information about the participation in certain stages of the research itself, which results are published, violates the rating of both authors and journals (7). Growing problem today are articles in the form of Guidelines for various diseases written by various professional or scientific associations in which as co-authors appear national representatives in these associations, somewhere even more than 100 names of co-authors (Figure 5), many of whom are not wrote even a sentence in that article. These articles provide co-authors the special benefits of citing these articles, which artificially inflate the values of scientific indexes that validate the authors and coauthors of the article in question. Group authorship may be appropriate when a group of reserachers has colaborated on a project, such as a multicenter trial, a concensus document, or an expert panel (7). But, it can be inaccurate and impossible to list all colaborators, and all called “co-authors” need to think about how to communicinate credit and responsibility for content of the article.
Inflated Co-authorship Introduces Bias to Current Scientometric Indices

Authors Statement for Publication

<table>
<thead>
<tr>
<th>No.</th>
<th>Author’s full name (in capital letters)</th>
<th>Author’s signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Noni Novisari Soeroso</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Tengku Kemala Intan</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Jery</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Fannie Rizki Ananda</td>
<td></td>
</tr>
</tbody>
</table>

Authorship Statement:

According to International Committee of Medical Journal Editors (ICMJE): "An author is considered to be someone who has made substantive intellectual contributions to a published study. An author must take responsibility for at least one component of the work, should be able to identify who is responsible for each other component, and should ideally be confident in their co-authors’ ability and integrity." (available at: http://www.icmje.org/ethical_1author.html)

Please fill the table with numerical code of each author (from the before signed table) regarding authors contribution to specific component of research.

<table>
<thead>
<tr>
<th>Component of the research</th>
<th>Author’s number</th>
</tr>
</thead>
<tbody>
<tr>
<td>substantial contribution to conception and design</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>substantial contribution to acquisition of data</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>substantial contribution to analysis and interpretation of data</td>
<td>4</td>
</tr>
<tr>
<td>drafting the article</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td>critically revising the article for important intellectual content</td>
<td>1, 2, 4</td>
</tr>
<tr>
<td>final approval of the version to be published</td>
<td>1, 2, 4</td>
</tr>
</tbody>
</table>

Date (dd/mm/yyyy) 08-07-2020

Figure 4. Authors Contribition Statement Form signed by authors of paper published in Medical Archives journal
Our study showed that on average about 10% of papers published by the surveyed authors had more than 30 co-authors, but these papers brought more than 40% of all citations and more than 40% of Hirsch’s index attributed to these authors. The duration of scientific activity was well correlated to number of citations, Hirsch’s index and the number of publications themselves with 30 or less co-authors, while the correlation did not exist with number of citations. In summary, publications with > 30 authors carry more scientometric points than publications with less co-authors, and the researchers with shorter scientific activity had larger scientometric benefit from publications with more than 30 authors than senior researchers.

Although high number of co-authors in biomedical papers could be justified in certain circumstances, when research topics are so complex that each of the co-authors has to make substantial contribution (18), more often the efforts of co-authors are limited to collection of a few regional or national data from already published primary research (19). A survey of declarations of authorship when submitting a manuscript to a medical journal revealed that 45.8% of co-authors did not satisfy the International Committee of Medical Journal Editors (ICMJE) authorship criteria (20). However, no matter whether co-authorship is justified or not, scientometric benefit from publications with high number of co-authors is disproportionally large. Since researchers are well aware of it, majority readily participate in projects that will result with such publications. An analysis of publishing activity in science in modern Russia (21) showed that submitting and publishing papers with enormous number of co-authors (even 3000 per paper) was perceived and used by many as a tool to inflate scientometric performance and survive in the race for academic promotion and funding. Such publications with a couple of hundreds or thousands co-authors even received an unofficial name: “collider” publications (21). Indiscriminate and bureaucratic use of scientometry actually make control of quality of scientific work impossible, and opens doors of scientific institutions to individuals with unethical behavior ready to manipulate with formal measures of research output, that are both not sensitive and not specific enough. Such situation, if prolonged, may have long term deleterious consequence on research in any field, including medicine, as already noted in some countries (12, 21). One of possible ways out is construction of more sensitive and specific scientometric indicators that would capture the work invested in a research result, and how it really influenced advancement in knowledge, rather than just measuring how often something was mentioned (22).

Regardless of how the new indicators will be constructed, attributing to a co-author only a fraction of citations of his/her paper would (e.g., by dividing number of citations of a paper with number of the co-authors) surely decrease inflation of co-authorship in modern biomedical scientific literature. The accuracy of citation of a certain article is the key to perform a proper scientometric analysis, and editors of the magazine have to take into account very strict rules when quoting references (23-26).

Based on the cited literature scientometric indicators can be used to analyze and evaluate the work of researchers, institutions, regions and countries (3, 9, 10). Scientometric indicators of work of an author, in addition to the number of citations which is a priority in the modern scientific community. H index Jorge Hirsch (in 2005) introduced a new indicator for quantifying the research output of scientists. Hirsch’s so-called H index was proposed as an alternative to other bibliometric indicators - such as the number of publications, the average number of citations and the sum of all citations - “a scientist has index h if h of his or her Np papers have at least h citations each and the other (Np - h) papers have ≤ h citations each”. Also, i10-index - the number of publications with at least 10 citations. G-index - articles ranked in decreasing order of the number of citations that they received, the g-index is the (unique) largest number such that the top g articles received together at least g2 citations - an academic has a g-index of 30 if the top 30 most cited of his/her papers combined have at least 900 citations (3). A way to boost the h index is by self-citation (3, 8, 10). The problem arises also in the fact that many software work on the basis of information that offers Google Scholar (Publish and Perish), so they also give the wrong information (23). Sometimes the same names of the authors, are an additional problem, so we come to the conclusion that the identification number of an author (The Open Researcher and Contributor ID (ORCID)), should become a requirement when publishing a paper, because it is the only way to make a distinction between authors, and to conduct analysis of the work of one author in the right way. ORCID number arises as an imperative in the modern scientific digital world, and is also an essential thing in scientometric analysis of the work of one author. ORCID is aimed at registering scholarly contributors and averting the persistent ambiguity of recorded author names (registry is growing fast and integrating with other ID-generating platforms, thereby increasing the functionality of the integrated systems) (3, 10).

Our experiences are that the authors easily sign the Author Contribution Statement, (Figure 4) and definitely his/her co-authors must confirm the approval with publication of the manuscript in some journal. The first author must a sign agreement on behalf of all co-authors of the manuscript that all of them participated in the writing of manuscript to take public responsibility for it. It is therefore very important to know the criteria for the (co) authorship. Authorship should be based on substantial contribution to the researchers. Corresponding author is not only person who will put his/her ORCID ID or E-mail address, he or she in every situation could be or should be prepared to explain the presence and order of these individuals” (3, 8). Authors must be aware that after the article is accepted for publishing, there is no room for changing the list of authors, adding new authors, which is also a direct violation of all ethical codes of scientific publication (8-10).
Finally, we recommend to experts of Web of Science, Scopus, Medline Pubmed Central and other index databases to check (randomly) papers in published articles “covered” with signed “Author’s Contribution Form” (ACF) by all co-authors who personally handwrite their contribution in involving themselves in the written results of the study (like we presented ACF of this journal, following ICJME rule, COPE and Sarajevo Declaration on Integrity and Visibility of Scholarly Journals) (4, 23-26).

Limitations of the study
Our study has several limitations. Limited number of scientometric indices was taken into account, and only one database was used for selection of the study sample, which may have introduced certain degree of non-reporting bias. The search strategy limited enrollment of the studies to those with co-authors having specific surnames, introducing selection bias.

6. CONCLUSIONS
- Unjustified prolific co-authorship is one of methods for inflation of scientometric indices that are not further reflecting true quality of research output of an individual.
- Further improvement of scientometric indicators may prevent unjustified co-authorship if it reflects the work invested in a research result. However, one should take into account a number of researchers in later years, who are not interested in this kind of self-promotion, and lists that are made without them, are not valid and are not a realistic indicator of the development of a certain field in one state.
- Authorship guidelines are not sufficient and need to be upgraded. They are not widely known and may even be ignored by many authors.
- Knowledge about formal authorship criteria is highly variable and majority of scientist are not familiar with existing criteria or do not consider formal criteria necessary.
- The relationship between the author, the mentor, the data processing person, the person providing the moral support etc. must be established, and not all of them has a place in the list of authors, they should be given special places at the end of the article, a space for acknowledgements, where these people may be mentioned.
Inflated Co-authorship Introduces Bias to Current Scientometric Indices

• The role of Editors-in-Chiefs of the scientific journals in this case is very important - to follow current criteria, established and proposed by COPE and ICMJE and avoid and prevent of publishing papers with listed co-authors in the article without strictly described what every co-author participated in submitted article which eventually will be accepted for publishing

REFERENCES


