

Multiple authorship trends in prestigious journals from 1950 to 2005

Sami Shaban, PhD.

ABSTRACT

Objectives: Editors of many established journals have noticed throughout the years an increase in the number of authors per article. I wish to put this theory to rest by calculating this linear increase for several prestigious journals.

Methods: I chose several prestigious journals and obtained authorship information for all articles published throughout the existence of these journals from the PubMed website. The data collection was performed on January 22, 2006, and was performed for the years from 1/1/1950 to 12/31/2005 for each journal. With the use of several self-developed software programs and database queries, I was able to calculate the average number of authors per article per year and the percent of single-author articles per year for these journals.

Results: The result confirms the increase in the number of authors per article in a linear fashion for all examined journals. The result also shows a decrease in the number of single-author articles in a sporadic fashion for all examined journals.

Conclusion: The data collection and calculation method can be used for any journal indexed in PubMed. The resulting trends are likely due to the increasing complexity of research, the increasing pressure on researchers to publish or perish, among others. This paper, however, focuses less on the reasons for this trend and more on the computerized methods used to obtain the results.

Saudi Med J 2007; Vol. 28 (6): 927-932

From the Health Information Management and Technology Program, College of Applied Medical Sciences, King Faisal University, Dammam, Kingdom of Saudi Arabia.

Received 14th October 2006. Accepted 5th February 2007.

Address correspondence and reprint request to: Dr. Sami Shaban, Health Information Management and Technology Program, College of Applied Medical Sciences, King Faisal University, PO Box 2435, Dammam 31451, Kingdom of Saudi Arabia. E-mail: shabansf@kfu.edu.sa

The increase in the number of authors per article is well noticed in scientific journals and has been the case for many years. Since the early 1970s,^{1,2} through the 1980s,^{3,4} and 1990s,^{5,6} and recently during the new millennium,^{7,8} this trend has been well documented. The decrease in the number of single-author articles, while less noticed, has also been discussed.⁸ Reasons for this trend range from the increase in complexity of research, the sudden emergence of multi-disciplinary research topics, the increased ability for researchers to communicate in the Internet era, the increasingly common practice of including honorary or guest authors, and therefore collaborate, and finally the ever increasing pressures on researchers to publish or perish.⁹ To this end, editors of several journals have established authorship guidelines and criteria for authorship made available as editorials or articles.¹⁰⁻¹⁶ Such guidelines can be summarized by the statement that each author must contribute to the manuscript in each of the following areas: 1. Substantial contribution to conception and design, or acquisition of data, or analysis and interpretation. 2. Drafting or revising the article. 3. Final approval. All 3 criteria must be met for an individual to be listed as an author or co-author on a manuscript. Although these guidelines are well known by researchers, the increase in the number of authors per articles continues, and the decrease in the single-author articles also continues. The past studies of authorship trends¹⁻⁸ restricted themselves to a sample of articles and a limited range of years. The reason for this is obvious: they studied the articles manually. In this study, I obtained authorship information (number of authors per article) for all articles and all years of the chosen journals. This information was obtained from downloading the authorship information from the PubMed website. It is possible to perform this analysis for any journal provided that it is

indexed in PubMed. This article is mainly concerned with the methodology of collecting, organizing, and analyzing authorship trend information obtained from PubMed. The results confirm trends cited in previous studies.

Methods. Data collection and transformation was performed in 3 phases: 1) Collection of raw data from PubMed, 2). Extraction of wanted data, and 3) Transformation of wanted data into analysis form. Six prestigious journals were chosen for multiple authorship analysis. They are: New England Journal of Medicine (N Engl J Med), Nature, Science, Lancet, British Medical Journal (BMJ), and Journal of American Medical Association (JAMA).

1. Collection of raw data from PubMed. The PubMed website www.ncbi.nlm.nih.gov/PubMed was used to collect the journal data for the 6 chosen journals. The abbreviated journal name is entered in the "for" field, the "limits" link is clicked, and "Journal" is chosen in the "Limited to:" pick list. Then the "Go" button is clicked and the results are displayed on the screen. The results are presented in "Summary" form, which contains author names, article title, journal abbreviation, date of publication, volume, and page numbers, and a

unique ID for the article in PubMed. This is not an easy form for automated extraction of wanted data. Rather, the "Medline" form is ideal for automated extraction. It is chosen rather than the "Summary" form by clicking on the "Display" pick list and choosing "Medline". In the same fashion, the sort order "Pub Date" is chosen so that articles are sorted by publication date and the "Send to" pick list is changed to "File". The complete journal's data are then downloaded and saved in a text file on the local computer. This process is performed for each journal. The result is 6 text files ranging between 60-120 Megabytes in size, which are then used in the extraction phase. It is important to note that this data collection was performed on January 22, 2006, and was performed for the years from 1/1/1950 to 12/31/2005.

2. Extraction of wanted data. The 6 downloaded text files now contain each journal's data in "Medline" form. The following is an example of the data of one article in the "Medline" form (***Appendix 1**).

The form is suitable for easy extraction of the items in bold into another data file in tabular form. Using the following Visual Basic program (**Table 1**), all 6 files are processed and one data file in tabular form is outputted as a result.

Table 1 - Visual Basic Program used to extract needed data items from Medline downloaded text files.

```

InputFile = Me.txtPath & Me.InputFileName
JN = Me.InputFileName
OutputFile = Me.txtPath.Text & Me.OutputFileName
Open InputFile For Input As #1
Open OutputFile For Append As #2
Input #1, InputLine
While Not EOF(1)
    If Left(InputLine, 5) = «PMID-» Then
        PMIDprev = PMID
        PMID = Trim(Mid(InputLine, 6))
        If Y <> "" Then
            Print #2, JN & "," & PMIDprev & "," & Y & "," & AN & "," & PT
            Me.lblFinish = "Processing " & JN & " " & Str(UISum)
            DoEvents
            AN = 0 : mypt = "" : Y = ""
        End If
        UISum = UISum + 1
    End If
    If Left(InputLine, 5) = "DP -" Then Y = Mid(InputLine, 7, 4)
    If Left(InputLine, 5) = "AU -" Then AN = AN + 1
    If Left(InputLine, 5) = "PT -" Then PT = PT + Mid(InputLine, 6) + "|"
    Input #1, InputLine
Wend
Print #2, JN & "," & PMID & "," & Y & "," & AN & "," & PT
Me.lblFinish = "Done " & JN & " " & Str(UISum) & vbNewLine
Close #2
Close #1

```

*The full text including Appendix is available in PDF format on Saudi Medical Journal website (www.smj.org.sa)

The output of this program is one data file in tabular form with 3 columns. They are: Journal Abbreviation, Year of Article Publication, and Number of Authors. Each line in this table represents data from one article.

3. Transformation of wanted data into analysis form. The third methodology phase was to transform the above data file into 2 formats suitable for analysis. The first is to perform regression analysis per journal on average articles per year with more than one author. The data must be transformed into tabular form with the following column headings. I will call this the final average table:

Journal Name	Publication Year	Total Authors	Total Articles	Average Number of Authors per Multiple-Author Articles
--------------	------------------	---------------	----------------	--

Each line in this table will represent the average authors and total articles published per year per journal for articles with more than one author. The second is to perform regression analysis per journal on the percent of articles per year with only one author. This requires the data to be transformed into tabular form with the following column headings. I will call this the final percent table:

Journal Name	Publication Year	Total One Author Articles	Total Articles	Percent of Single-Author Articles
--------------	------------------	---------------------------	----------------	-----------------------------------

The first format is achieved by importing the downloaded table into a database and running an "average" query to find the average authors per year. The query is as follows and is performed for articles with more than one author only: SELECT Journal, Year, Sum(Authors), Sum(ArticleID), Avg(Authors) FROM Articles GROUP BY Journal, Year;

The result of the query is stored as the final average table. Each line in this table represents the average authors, and total articles published per year per journal, whereas each line in the downloaded table represents one article. The second format is achieved by importing the downloaded table into a database and running a "percent" query to find the percent of single-author articles per year. The query is as follows and is performed for one author articles only: SELECT Journal, Year, Sum(OneAuthorArticles) AS TOAA, Sum(ArticleID) AS TA, TOAA*100/TA AS [Percent One Author Articles] FROM Articles GROUP BY Journal, Year;

The result of the query is stored as the final percent table. These 2 resulting tables are graphed and shown in the results section. Finally, this method can be used for any journal indexed in PubMed.

Results. Publications were categorized into "Articles" and "Letters" using the Publication Type (PT) field in the "Medline" form of the articles downloaded from PubMed. "Articles" include the following publication types: original research, review articles, case report, and clinical trial; and "Letters" include: letter to the editor and comment. For each journal, the number of "Articles" and "Letters" are further categorized by number of authors: "0" no authors listed, "1" single author articles, and ">1" more than one authors per article.

Table 2 shows the percent single author articles and average authors for multiple author articles by decade from 1950-2005 for all journals except JAMA, which started in 1960. It is noteworthy that for some journals a large percentage of articles were indexed in Medline with no author information. These articles were excluded from all calculations. The percentage of articles with no author information for each journal is as follows:

Table 2 - Percent single author articles and average authors for multiple author articles by decade from 1950-2005 as indexed in Medline.

Journal	Authors	50's	60's	70's	80's	90's	00-05	
New England Journal Medicine	1	40	41	48	21	20	25	%
	>1	2.8	3.0	3.5	4.9	5.6	5.9	Avg
Nature	1	46	32	24	12	11	25	%
	>1	2.5	2.6	3.0	4.0	5.1	5.8	Avg
Science	1	40	35	25	16	16	20	%
	>1	2.6	2.7	3.1	4.0	5.0	5.7	Avg
Lancet	1	52	41	37	19	30	35	%
	>1	2.6	3.0	3.3	4.5	4.7	5.3	Avg
British Medical Journal	1	60	51	44	33	39	34	%
	>1	2.5	2.8	3.2	3.6	3.9	4.2	Avg
Journal of American Medical Association	1		42	49	31	35	26	%
	>1		3.0	3.1	3.8	4.9	6.4	Avg

NEJM 12%, Nature 3%, Science 0.5%, Lancet 13%, BMJ 18%, and JAMA 13%.

From the methodology section, the average number of authors per journal per article per year was calculated for journals with more than one author. Regression lines are shown, and R^2 values are calculated to be very high signaling a good fit to the linear model. The slope is positive indicating an increase in the average number of authors per article throughout the years. Single-author articles have also been calculated in the methodology section. Regression lines are shown, and R^2 values are calculated to be fairly low signaling a bad fit to the linear model. The slope is positive indicating an increase in the average number of authors per article throughout the years. Both the percent of single-author articles and the average authors per multiple-author articles for 4 of the 6 journals are plotted in Figures 1 to 4. Nature and Science Journals followed a similar pattern to the New England Journal of Medicine so they are not shown here. The values for Nature were: Number of articles with single authors: 15,890 from 1950 to 2005. Range of percent of single author articles: ~55% in 1950 to ~25% in 2005. Linear regression line: $y = -0.68x + 1363$. R-Squared value: 0.66. Number of articles with multiple authors: 42,005 from 1950 to 2005. Range of average authors per article:

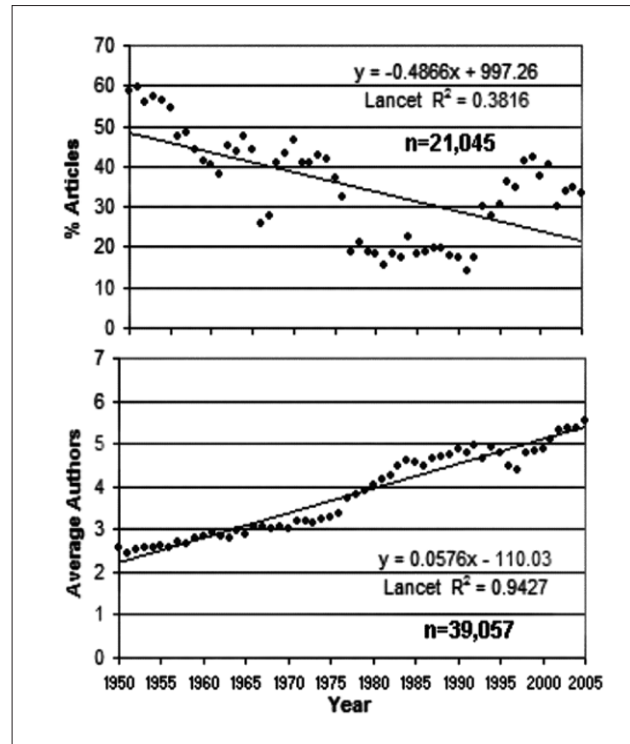


Figure 2 - Percent single-author articles and average authors per multiple-author articles for Lancet.

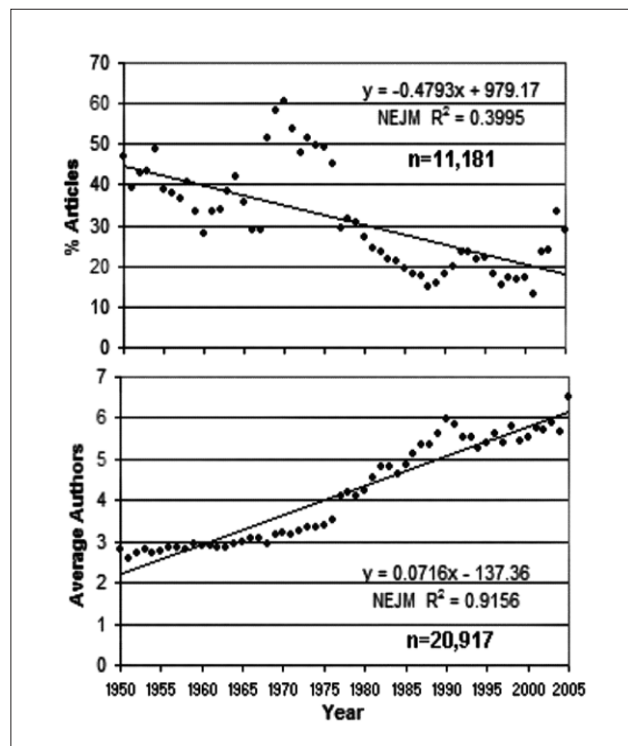


Figure 1 - Percent single-author articles and average authors per multiple-author articles for New England Journal of Medicine.

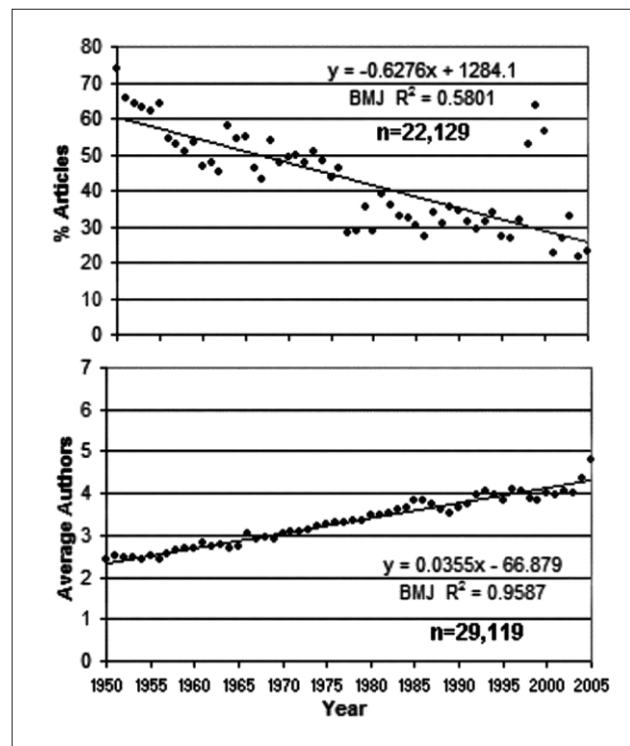


Figure 3 - Percent single-author articles and average authors per multiple-author articles for British Medical Journal.

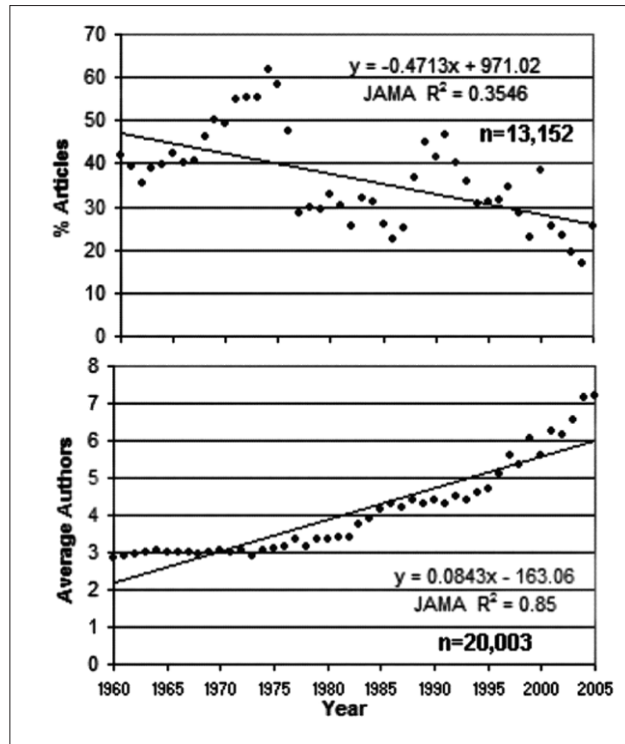


Figure 4 - Percent single-author articles and average authors per multiple-author articles for the Journal of American Medical Association.

~2.5 in 1950 to ~6.0 in 2005. Linear regression line: $y = 0.07x - 137$. R-Squared value: 0.89.

The values for Science were: Number of articles with single authors: 9,295 from 1950 to 2005. Range of percent of single author articles: ~42% in 1950 to ~20% in 2005. Linear regression line: $y = -0.52x + 1033$. R-Squared value: 0.71. Number of articles with multiple authors: 28,617 from 1950 to 2005. Range of average authors per article: ~2.5 in 1950 to ~6.0 in 2005. Linear regression line: $y = 0.064x - 123$. R-Squared value: 0.89

Discussion. As far as I know, this is the only study that uses a completely automated method for calculating the authorship trends. Previous articles depended on the manual study of publications and their authorship information. This significant method allows trend analysis to be performed on any journal indexed in PubMed for all or any part of the journal's publication life. The PubMed database is considered to be a cornerstone of scientific research, and it is important to note that these results depend heavily on the publication data that exists in PubMed. If it is incorrect or inaccurate then these results will be the same. For example, the categorization of publications in PubMed was used in

this study to categorize publications into "Articles" with publication types: original research, review articles, case report, and clinical trial; and "Letters" with publication types: letter to the editor and comment. Only those categorized as Articles were used in this study. Furthermore, some articles were found with no author information and were not used in this study. A study by Weeks et al in 2004 manually examined a sample of articles' authorship information from 4 medical journals including JAMA and NEJM from 1980 to 2000. They found the trend to be increasing very similarly to this study. However, they found that the average number of authors per article in these 2 journals was consistently above our calculated averages by one to 2 authors. For NEJM, this study found that the average authors per article per year between 1980 and 2000 to be between 4.1 and 5.4 as compared to 5.3 to 8.2 found by the previous study. For JAMA, this study found the average to be between 3.4 and 5.6 as compared to 3.9 to 7.3 found by the previous study. Finally, the larger than expected number of articles with no authors may be due to missing data in PubMed or because a portion of those articles are written using group authorship with no indication in PubMed regarding that. This could explain the higher average in the previous study.

In conclusion, the result confirms the increase in the number of authors per article in a linear fashion. The result also shows a decrease in the number of single-author articles in a sporadic fashion. These trends are likely due to the increasing complexity of research, the multidisciplinary nature of research today resulting in a need for collaboration, the increased ability of researchers to communicate in the Internet era, the increasingly common practice of including honorary or guest authors, as well as because of the increasing pressure on researchers to publish or perish.

References

1. Diamond D. Multi-authorship explosion. *N Engl J Med* 1969; 280: 1484-1485.
2. Roland C. Trends in authorship. *Arch Inter Med* 1970; 125: 771-772.
3. Burman K. "Hanging from the masthead": Reflections on authorship. *Ann Inter Med* 1982; 97: 602-605.
4. Powers R. Multiple authorship, basic research, and other trends in the emergency medicine literature (1975-1986). *Am J Emerg Med* 1988; 6: 647-650.
5. Epstein R. Six authors in search of a citation: Villains or victims of the Vancouver convention? *Brit Med J* 1993; 306: 765-767.
6. Khan, K et al. A controlled analysis of authorship trends over two decades. *Am J Obs Gynec* 1999; 181: 503-507.
7. Weeks WB, Wallace AE, Kimberly BCS. Changes in Authorship Patterns in Prestigious US Medical Journals. *Social Sciences and Medicine* 2004; 59: 1949-1954.

8. Slafer GA. Multiple authorship of crop science papers: Are there too many co-authors? *Field Crops Research* 2005; 94: 272-276.
9. Rennie, D, Yank, V, and Emmanuel, L. When authorship fails: A proposal to make contributors accountable. *Journal of the American Medical Association* 1997; 20:579-585.
10. Lundberg, G and Flanigan, A. New requirements for authors: Signed statements of authorship responsibility and financial disclosure. *J Am Med Assoc* 1989; 262: 2003-2004.
11. Kassirer, J and Angell, M. On authorship and acknowledgments. *N Engl J Med* 1991; 325:1510-1512.
12. Rennie, D and Flanagin, A. Authorship! authorship! guests, ghosts, grafters, and the two-sided coin. *J Am Med Assoc* 1994; 271:469-471.
13. Rennie, D. Who did what? Authorship and contribution in 2001. *Muscle Nerve* 2001; 24:1274-1277.
14. Bailey BJ. What is an author? *Otolaryngol Head Neck Surg* 2001; 124: 2-3.
15. Hozack WJ, Ranawat C, and Rothman RH. Guidelines for authorship. *J Arthroplasty* 2002; 17: 527.
16. Cronenwett, JL and Seeger, JM. Criteria for authorship. *J Vasc Surg* 2005; 42: 599.

New Peer Reviewers

Join our team of expert peer reviewers for Saudi Medical Journal by registering through the website at http://www.smj.org.sa/_Authors/ and select "register now" or sending an enquiry and summarized CV to info@smj.org.sa. Note that SMJ reviewers, whose reviews are returned on time and are judged satisfactory by the Editors, may receive 1 CME credit per review, with a maximum of 5 credit per year, from the Saudi Council for Health Specialties.