Are Short Research Titles More Attractive? Cross-Sectional Analysis of 9,830 Titles from PubMed

George Choueiry*[1,2]

[1]*Department of Epidemiology and Biostatistics, School of Public Health, Lebanese University, Fanar, Lebanon*
[2]*School of Pharmacy, Lebanese University, Hadath, Lebanon*

Abstract

Attractive titles are expected to drive more reads and thus more citations to a research article, so studying the effect of title length on its attractiveness can be reduced to analyzing its influence on the citation count. Previous studies on the subject showed conflicting results that are probably attributable to bias and confounding, since they mostly focused on predicting citation count based on title length instead of using a causal model to explain the relationship between the two. The present study aims to quantify the direct influence of title length on its attractiveness guided by a causal diagram to identify and eliminate alternative explanations such as indirect effects and confounding. The study used data on 9,830 biomedical research articles from PubMed Central, downloaded through an API created by Comeau and colleagues. Poisson regression modeled the citation rate as a function of title length, adjusting for mediators of indirect effects—such as the use of a comma and a colon in the title—and confounders—such as the journal impact factor and the mention of study design in the title. The model shows that each word removed from the title increases the citation rate by 2.5%. This means that, for the median article that receives 2.2 citations per year, each word removed from the title causes a gain of 0.055 citations per year, equivalent to 1 citation every 19 years. Although statistically significant, this effect is practically negligible—so shortening a research title is not an effective strategy for earning more citations.

Keywords *Title length; Title attractiveness; Citation count; Causal diagram.*

1 Introduction

The role of a research title is to draw the reader's attention while providing an overview of the article's content. Finding a way to engage readers is important since only 18% of those who read the title proceed to read the abstract (Mabe and Amin, 2002).

*Corresponding author. Email: georgecymail@gmail.com

Title attractiveness may be affected by its length; but studies on this subject have been inconsistent and sometimes contradictory (Subotic and Mukherjee, 2014; Letchford et al., 2015; Guo et al., 2018; Jacques and Sebire, 2010; Habibzadeh and Yadollahie, 2010; Stremersch et al., 2007; Falahati Qadimi Fumani et al., 2015). This may be due to bias and confounding since these studies did not follow a causal model to eliminate alternative explanations and indirect effects.

The confusion over the effect of title length led to a gap between what professional writers recommend and what researchers do in practice: while professionals recommend keeping titles as short as possible (Zeiger, 1999; Neill, 2007), in practice, titles are getting longer (Milojević, 2017; Whissell, 2012) and more descriptive (mentioning the study objective, the variables involved, the main result, and the study design).

To help resolve this issue, the present study aims to quantify the direct influence of title length on its attractiveness by analyzing data on 9,830 biomedical research papers from PubMed and adjusting for confounding and indirect effects through the use of a causal diagram.

2 Methods

For this cross-sectional study, data were downloaded from PubMed Central in March 2021 using a web API created by Comeau et al. (2019). From a collection of about 3 million biomedical research articles from various journals, 105,984 were chosen at random from those uploaded between the years 2016 and 2021.

From these 105,984 articles, a total of 96,154 were discarded for incomplete data, leaving 9,830 articles ready for analysis (Figure 1). Reasons for discarding articles included: unavailable full text, unmentioned study design, missing impact factor of the journal in which the article was published, missing article DOI, and unavailable citation count.

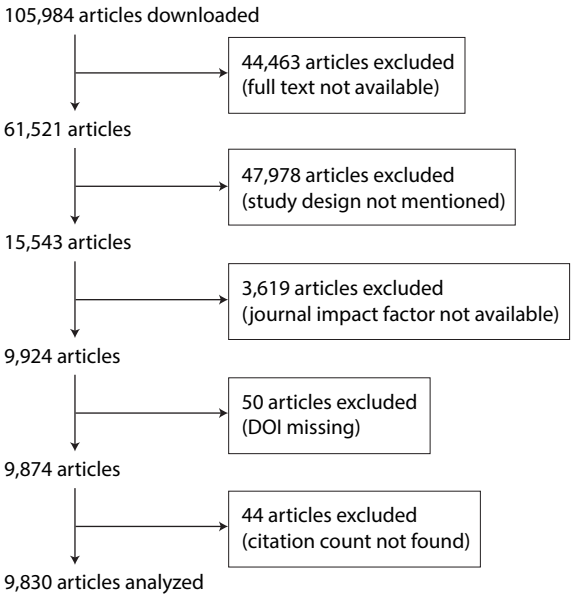

Figure 1: Flow of articles through a 5-step exclusion process.

To study the influence of title length on its attractiveness, and in order to avoid defining and measuring ***Title attractiveness***, I substituted this variable with another closely related one: the ***Citation count*** for a given article; this can work provided that we block all alternative paths other than the direct effect of ***Title attractiveness*** on ***Citation count***. Looking at the causal diagram in figure 2, we notice that there is only one alternative path, and it can be blocked by adjusting for the ***Journal*** in which the article was published. Since the data contained articles from 1,040 different journals (and to avoid complicating the analysis by creating 1,039 dummy variables), I ended up adjusting for the ***Journal impact factor***, a direct descendent of the deconfounding variable ***Journal***, thus representing most of its effect.

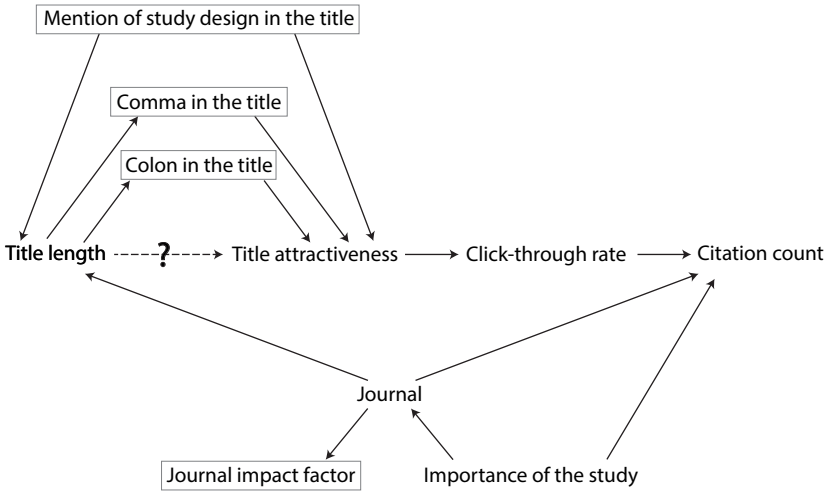

Figure 2: Causal diagram representing, based on background knowledge, the relationship between ***Title length***, ***Title attractiveness***, and other related factors. Boxes were placed around variables that were adjusted for in the analysis, and the question mark reflects the main question of the study: does ***Title length*** have a direct influence on ***Title attractiveness***? Arrows represent the direction of causal effects, for instance: since a more attractive title receives more clicks and therefore more citations, arrows were drawn from ***Title attractiveness*** to ***Click-through rate*** to ***Citation count***; since some journals impose an upper limit on title length, an arrow was drawn from ***Journal*** to ***Title length***, and since articles in more influential journals are more likely to be read and cited, another arrow was drawn from ***Journal*** to ***Citation count***; since longer titles tend to use a colon to indicate a pause, an arrow was drawn from ***Title length*** to ***Colon in the title***, and since the use of a colon in the title may be one of the factors that draw the reader's attention, an arrow was drawn from ***Colon in the title*** to ***Title attractiveness***; since the mention of the study design increases the length of the title, an arrow was drawn from ***Mention of study design in the title*** to ***Title length***, and since the mention of a meta-analysis or a randomized controlled trial may make a title more attractive to readers, another arrow was drawn from ***Mention of study design in the title*** to ***Title attractiveness***.

To compute the direct causal effect of ***Title length*** on ***Title attractiveness***, alternative explanations of the association between these two such as confounding and indirect effects must also be eliminated. From figure 2, we see that this can be accomplished by adjusting for the ***Mention of study design in the title*** (a confounder) and the use of ***Comma in the title*** and ***Colon in the title*** (indirect

effects).

After determining the variables that we want to adjust for, Poisson regression was used to compute the effect of ***Title length*** on ***Citation count***. In our case, a Poisson model has 2 major advantages over linear regression: (1) it fits the data better, since counts follow a Poisson rather than a normal distribution, and (2) it accounts for different publication dates of different articles, which is important to offset the advantage of older articles regarding the time they had to collect citations (this can be accomplished by including ***Years since publication*** as an offset in the model).

The Poisson model described above can be summarized with the following equation:

$$\begin{aligned}\log(\text{Citation count}) =& \beta_0 + \beta_1 \times \text{Title length} + \beta_2 \times \text{Journal impact factor} \\ &+ \beta_3 \times \text{Mention of study design in the title} + \beta_4 \times \text{Comma in the title} \\ &+ \beta_5 \times \text{Colon in the title} + \log(\text{Years since publication})\end{aligned}$$

Variables in the model, such as ***Citation count***, ***Title length***, and ***Journal impact factor***, were summarized using the median and the interquartile range (IQR), since they follow either a Poisson or a skewed non-normal distribution.

3 Results

In our sample of 9,830 articles, the median title composed of 16 words (IQR = 6), had 2.2 yearly citations (IQR = 3.33), and was published in a journal with an impact factor of 2.74 (IQR = 1.67). Also, 4,317 (43.9%) of titles contained at least one colon, 1,442 (14.7%) contained at least one comma, and 2,794 (28.4%) mentioned the study design.

The Poisson model shows a significant negative effect of longer titles on citation count (Table 1). Specifically, each additional word in the title causes a drop of 2.5% in the citation rate (95% confidence interval: [-2.7%, -2.3%]; $p < 0.001$). Equivalently, we can say that removing one word from the title causes an increase of 2.5% in the citation rate. To put that into perspective, removing one word from the title of the median article (that has 2.2 citations per year) causes a gain of 0.055 ($= 2.2 \times 0.025$) citations per year, equivalent to 1 citation every 19 years.

Table 1: Output of the Poisson model obtained by regressing ***Citation count*** on ***Title length*** (title word count), adjusting for confounders and indirect effects. The asterisk (*) indicates a statistically significant coefficient ($p < 0.05$).

Variable	Coefficient	Standard error
(Intercept)	1.172*	0.009
Title word count	-0.025*	0.001
Journal impact factor	0.123*	<0.001
Colon in the title	0.091*	0.006
Comma in the title	0.085*	0.007
Mention of study design in the title	0.264*	0.006

4 Discussion

This study shows that shorter research titles are more engaging by proving that they attract more citations. However, this effect, although statistically significant, is practically negligible since removing one word from a title will attract, on average, a single additional citation every 19 years—so I would not recommend shortening research titles as a strategy for increasing the citation count.

Previous studies on the subject reported conflicting results for articles in different disciplines since they did not use a causal approach to control bias and confounding. For instance, they found that shorter titles attracted more citations in psychology (Subotic and Mukherjee, 2014) and general scientific research (Letchford et al., 2015), but less in economics (Guo et al., 2018) and medicine (Jacques and Sebire, 2010; Habibzadeh and Yadollahie, 2010), and had no effect in marketing research (Stremersch et al., 2007) and scientometrics (Falahati Qadimi Fumani et al., 2015). What distinguishes the present study was the use of a causal diagram to identify and block alternative paths between title length and citation count, removing all but the causal explanation of any association between the two.

However, there are some limitations: (1) the 3 million biomedical research articles that are freely available on PubMed Central from which our sample was drawn may not accurately represent all published articles—thus introducing selection bias; (2) adjusting for the journal impact factor instead of the journal itself (to reduce model complexity) may have resulted in some residual confounding; and (3) the general approach taken to adjust for bias and confounding using a causal diagram (figure 2) created based on my understanding of the subject may have incorporated an element of subjectivity into the analysis. Future studies can address these issues by: (1) collecting data on articles from different disciplines (to increase the result's generalizability), (2) including a larger number of articles from each journal (to enable adjusting for ***Journal*** instead of ***Journal impact factor***), and (3) validating,

either theoretically or analytically, the structure of the causal diagram (to reduce subjectivity).

Finally, this study proves that shortening a research title is not an effective strategy for earning more citations. Yet, writing shorter titles may still have other benefits, such as: getting more reads on Mendeley (Zahedi and Haustein, 2018; Didegah and Thelwall, 2013), tweets (Haustein et al., 2015), appearances in social media in general (Zagovora et al., 2018), and avoiding truncation when they appear on the results page of an online search engine like Google.

References

Comeau D, Wei C, Islamaj Doğan R, Lu Z (2019). PMC text mining subset in BioC: about three million full-text articles and growing. *Bioinformatics (Oxford, England)*, 35(18): 3533–3535.

Didegah F, Thelwall M (2013). Which factors help authors produce the highest impact research? collaboration, journal and document properties. *Journal of Informetrics*, 7(4): 861–873.

Falahati Qadimi Fumani M, Goltaji M, Parto P (2015). The impact of title length and punctuation marks on article citations. *Annals of Library and Information Studies*, 62(3): 126–132.

Guo F, Ma C, Shi Q, Zong Q (2018). Succinct effect or informative effect: the relationship between title length and the number of citations. *Scientometrics*, 116(3): 1531–1539.

Habibzadeh F, Yadollahie M (2010). Are shorter article titles more attractive for citations? cross-sectional study of 22 scientific journals. *Croatian Medical Journal*, 51(2): 165–170.

Haustein S, Costas R, Larivière V (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLOS ONE*, 10(3): e0120495.

Jacques T, Sebire N (2010). The impact of article titles on citation hits: an analysis of general and specialist medical journals. *JRSM short reports*, 1(1): 2.

Letchford A, Moat H, Preis T (2015). The advantage of short paper titles. *Royal Society Open Science*, 2(8): 150266.

Mabe M, Amin M (2002). Dr Jekyll and Dr Hyde: author-reader asymmetries in scholarly publishing. *Aslib Proceedings*, 54(3): 149–157.

Milojević S (2017). The length and semantic structure of article titles—evolving disciplinary practices and correlations with impact. *Frontiers in Research Metrics and Analytics*, 2.

Neill U (2007). How to write a scientific masterpiece. *The Journal of Clinical Investigation*, 117(12): 3599–3602.

Stremersch S, Verniers I, Verhoef P (2007). The quest for citations: Drivers of article impact. *Journal of Marketing*, 71(3): 171–193.

Subotic S, Mukherjee B (2014). Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *Journal of Information Science*, 40(1): 115–124.

Whissell C (2012). The trend toward more attractive and informative titles: American psychologist 1946-2010. *Psychological Reports*, 110(2): 427–444.

Zagovora O, Weller K, Janosov M, Wagner C, Peters I (2018). What increases (social) media attention: Research impact, author prominence or title attractiveness? In: *International Conference on Science and Technology Indicators (STI) 2018*. arXiv.org, Leiden, The Netherlands.

Zahedi Z, Haustein S (2018). On the relationships between bibliographic characteristics of scientific documents and citation and mendeley readership counts: A large-scale analysis of web of science publications. *Journal of Informetrics*, 12(1): 191–202.

Zeiger M (1999). *Essentials of Writing Biomedical Research Papers*. McGraw Hill / Medical, 2nd edition.