**Editorial statement**

This article offers a large-scale empirical analysis of spelling errors in scientific titles and abstracts, evaluating their prevalence and potential as a proxy for research quality. The article has been reviewed by two reviewers. The reviewers emphasize the paper’s novelty and scope but also identify a number of weaknesses. They inter alia recommend stronger grounding in existing literature and caution against overinterpreting spelling errors as standalone indicators of research quality. They also highlight the need for more clearly distinguishing between correlation and causation, for example, when it comes to the reported link between error frequency and reduced citation rates. Similarly, they caution against speculative attributions, such as the role of LLMs in reducing spelling errors since 2017. The reviewers also recommend some clarification in regard to data and methods used. For example, the author should consider adding more details about types of articles that were analyzed, as well as an explicit justification for the choice of the databases. Overall, the reviewers feel that the study provides an important foundation for further research into quality markers in scholarly communication, but would benefit from more refinement of its theoretical framing and methodological rigor.

Response: I agree with the reviewers about the speculative suggestion of the role of LLMs in the trend and have removed this suggestion. I found both reviewers’ comments to be useful and have responded to them below.

**Review 1**

The author has explored the number of spelling errors in titles and abstracts of indexed papers, with the view that this may indicate quality of the publication.  In this nicely constructed paper, the argument is introduced and tested using two searchable indexing systems. A low but measurable amount of spelling errors was observed, with incidence trends apparent over time, with different publishers and from different countries.

Publish or perish as a driver of hasty publication is perhaps not as well validated as we would like. The mantra has been around for many decades, consistent with a longstanding theme in academic fields (e.g. see Garfield E., The Scientist, Vol:10, #12, p.11 , June 10, 1996), but there seems little in the way of credible evidence to support it. This paper may provide the first support with a significant reduction in citations.

Given the relatively low incidence of papers with an error (approximately 2 per 1,000), the author needs to consider how useful is this marker in deciding on quality of a specific publication, or an individual author's track record? Perhaps the paper would benefit from a discussion around whether an 8% reduction in citations on those papers with 2 errors would be of practical significance in scenarios where track records are being assessed. It needs to be clear what was the incidence of papers with two or more errors (from Figure 6, n=668 from 21 million abstracts)?

Response: Thanks for your interesting points. I think any single measure of publication quality is unlikely, given the complexity of most scientific articles. However, spelling errors could be part of a wider assessment. For example, spelling errors were considered in an early version of the INSPECT-SR tool, but not used in the final tool (DOI: 10.1016/j.jclinepi.2025.111824).

I have now added the practical significance using the absolute scale for the 8% reduction in citations, and this impact is small at just 1.7 fewer citations for an abstract with two spelling errors. Hence, this is a statistically strong finding, but of relatively small practical significance. I have now noted the small absolute effect in the discussion.

Speculation that LLMs may be responsible for the recent improvement in spelling needs a closer comparison of the timeline of LLM availability/usage by authors to support the association. It is a logical concept and may be a useful practice for those prone to spelling errors, but I am not convinced that the widespread adoption of LLMs for copy editing is early enough to explain the decline in spelling errors from 2017.

Response: Talking only of LLMs was speculative and I have now removed this claim. A more likely reason is the growing use of reviews over time, and these tend to have fewer spelling errors as shown in Figure 2.

Minor issues

A justification for use of Pubmed and Openalex is required and explanation of how the two databases were managed to avoid duplicate counts.

Response: I used PubMed largely because my experience has been in health and medical research and the database is very large and easy to work with. I used OpenAlex because it is also easy to work with and because it combines multiple other databases and has good coverage (see DOI: 10.1007/s11192-025-05293-3). I’ve added some details to the methods.

The possibility of duplicates was minimised as both databases used unique PMIDs for each article. I also regularly checked for duplicates in the coding.

Consistency edits in bibliography format are required, there are many styles in title capitalization, italicization, full name or first initials, etc.

Response: Hopefully fixed now.

I am uncertain whether track-record should be hyphenated in this usage?  In papers on spelling it seems important to avoid errors of usage. Can the authors please review this?

Response: Agreed and changed to unhyphenated.

**Review 2**

The author presents a large-scale analysis of spelling errors in scientific abstracts, emphasizing that they can be seen as indicators of poor research quality. I have several comments that the author could use to improve their manuscript. Most importantly, I believe it is pertinent that the author is more explicit in which expectations were confirmatory and which were exploratory. The paper reads as if it is a descriptive study, for example because standard statistical results are omitted, but does toss in language referring to hypotheses like “As expected”. As such, it does not become entirely clear what the goals of the study were. Please find below some other comments that the authors could use to improve the paper.

Response: Thanks for pointing this out, I’ve now provided more background to these statements as described below.

p. 2 🡪 “However, if the relative rate of errors is increasing, then that would be a sign that the average quality of research is declining.” 🡪 This is a pretty strong claim so I would use “would” instead of “could” and probably add some references that link spelling errors to research quality. This would add weight to the findings in the paper. In general, the paper is lacking a (short) literature review on the prevalence of spelling errors in science, and the factors that influence them. The author could start by elaborating on the findings in reference 17.

Response: I agree that ‘would’ is better than ‘could’ and this section has been changed in response to comments below. There is no previous literature that I am aware of that links spelling errors to quality, apart from our blog (reference 17) and the cited paper on “myocardial infraction” (reference 50).

p. 2 🡪 The choice of 2008 and the associated 88% rate seem to be rather arbitrary. Why was this choice made instead of a year corresponding to a slightly higher or lower percentage?

Response: Yes, this was an arbitrary choice and was mostly based on having information available from 2008 thanks to the cited report. PubMed did not publish annual updates of this percentage.

p. 2 🡪 What did the verification checks show? Did all automatically detected spelling errors constitute an actual error after the manual check? It appears that the sample for these checks is rather low when compared to the total sample. Could you explain why you think these samples are sufficient to draw a conclusion about the validity of the algorithm?

Response: The sample sizes for the random checks are relatively small, but they still provide a reasonably accurate estimate of the true error because the error rates are close to zero, so the binomial variance is small. I am comfortable in using these samples to verify the validity of the algorithm because they were random samples and the upper 90% error is relatively small. Any biases from these small errors should be minimised by the study’s large sample size.

The verification checks are detailed in supplement S.4 and showed a low false positive rate (all automatically detected errors were true errors) but relatively high false negative rate. Hence, some of the control abstracts will include spelling errors.

p. 3 🡪 “We examined 200 randomly selected abstracts with no errors detected by our algorithm to check for false positives.” 🡪 This reads like it could also be about false negatives, so you may want to rewrite.

Response: Thanks for spotting this slip, it has been corrected.

p. 3 🡪 “We suspected that the association between the number of authors and spelling errors would be non-linear.” 🡪 Why did you suspect this, and was this a formal hypothesis?

Response: I expected a diminishing return with more authors. I have added this now.

p. 3 🡪 “We used random effects for country and publisher and used horseshoe priors as we expected the effects for some countries and publishers to be close to zero“ 🡪 Why did you expect this?

Response: This was because some countries and publishers would have small sample sizes. I have added this now.

p. 6 🡪 “We predicted that spelling errors would increase over time due to an average decline in research quality.” 🡪 This seems to be the main prediction but it only surfaces clearly in the discussion and not much evidence is given for the decline in research quality that is assumed to determine this effect. Moreover, in the introduction the relationship between spelling errors and research quality is formulated in a non-directional way “if the relative rate of errors is increasing, then that would be a sign that the average quality of research is declining”. It is vital that it is clear whether there was an a priori hypothesis, and whether the cause (a decline in research quality) was also part of the hypothesis or merely an explanator.

Response: Thanks for pointing this out. I have now made my *a priori* expectation clearer in the introduction.

p. 4 🡪 “We included the years since publication as a predictor, as papers generally accumulate citations over time. We included the potential confounders of article type, number of authors, first author’s country, and publisher.” 🡪 For confounders, please explain why they were incorporated in the model and how any omitted variable bias would look like.

Response: I drew a DAG to determine the confounders, this is available here: <https://github.com/agbarnett/spelling_wikipedia/blob/main/figures/dag.jpg>. The key omitted variables I identified were editorial effort, native English speaker, spell check software, and research quality. For native English speaker, I have the imperfect proxy of country. I have no proxy for spell-check software but cannot envisage a serious bias from not having this variable. Research quality will have many predictors, but my hope is that most would be independent of spelling errors after accounting for trends over time and country.

p. 4 🡪 “There were 1,916 unique spelling errors that were detected 48,420 times in over 21 million abstracts (0.2%).” 🡪 Please make it clear to what number the percentage refers to

Response: Thanks for spotting this. I’ve moved the percentage to the next sentence.

p. 5 🡪 “The solid lines are the mean and shaded areas are 95% confidence intervals for the mean.” 🡪 I believe this should be “means” instead of “mean” as there are multiple.

Response: Agreed and changed.

p. 5 🡪 It would be good to already mention which article types you looked at in the methods section. This is especially useful given the double meaning of the word “letters”.

Response: Agreed and changed.

p. 7 🡪 “so an increased use of LLMs may explain the decreasing trend from 2016 (Figure 1).” 🡪 I don’t find this convincing since massive uptake of LLMs came later. Does the author have any alternative explanations for the decreasing trend?

Response: This was also highlighted by the other reviewer and I’ve reflected on this and read a recently published paper that indicates that LLM use is far more recent (DOI: 10.1126/sciadv.adt3813). Hence, I’ve removed this argument and talked instead about the growing use of reviews which tend to have fewer spelling errors.

p. 8 🡪 “As expected, we found a large variance in spelling error rates by the author’s first country (Figure 3).” 🡪 This expectation was not explicated earlier in the paper.

Response: I have changed the wording for this sentence to start with “Unsurprisingly”.

p. 9 🡪 “We found that more spelling errors were associated with fewer citations (Figure 6).” 🡪 Is this association statistically significant? In general, the paper provides convincing descriptive statistics but to make claims about associations I think statistical tests are in order, especially for any hypotheses the author had a priori.

Response: The statistical significance can be derived from the confidence intervals in Figure 6 which are far from including zero, so the p-value will be well below the standard 0.05 threshold. I prefer not to give p-values as many readers then put too much weight on the statistical significance ahead of the practical significance. I have added the 95% credible interval to the text to help clarify the statistical strength of the association.

p. 9 🡪 “Some researchers may not cite a paper after reading a spelling error in the abstract, as this may undermine their confidence in the paper’s robustness. Researchers who make spelling errors may be more likely to make other more serious errors, such as incomplete reporting of the results, and hence the reduction in citations could be due to wider quality concerns” 🡪 This result seems to be the most important one in the paper so more discussion about the implications seems warranted. Maybe the author can discuss whether and how future research would provide more information on these trends.

Response: This was an interesting result, and the other reviewer also highlighted this finding. The other reviewer suggested adding practical significance, which I have done by adding the absolute reduction in citations which was small.

I have speculated that looking for an association between spelling errors and the quality of reporting would be interesting, as this is another important marker of research quality and carefully conducted research.

p.10 🡪 The small section with related work could perhaps be better placed in the introduction to give the readers more context on the research area.

Response: Agreed and moved.

p. 10 🡪 “We only examined spelling errors in the titles and abstracts, as the full text of all papers is not freely available.” 🡪 Why would doing the analysis on a subset of full text paper be less informative? Would there be too much bias? In what direction?

Response: Authors potentially spend more time on the abstract, as many understand that some editors use the abstract to decide whether to desk-reject a paper. Hence, I think it likely that full texts have more errors. I have added a comment on this to the discussion.

p. 10 🡪 “We did not check grammar because we could not automate grammar checks on millions of papers.” 🡪 Why was this not possible, even using AI tools?

Response: Good point, it is likely now possible to use large language models to provide a grammar check. I have added this to the paper.

p. 10 🡪 “We only examined the health and medical literature, meaning our results are not generalisable to all scientific fields” 🡪 What was the reason for not looking at other literatures? Any speculation as to how social sciences or the natural sciences would fare?

Response: The main reason for only examining medicine is that the PubMed database is very large and excellent to work with. I imagine that the broad patterns found in this study would be like other areas of science. However, I am not willing to make a strong bet on this and it needs to be examined by new studies.

p. 10 🡪 “However, we believe that these errors are relatively small (Table S.1) and do not impact our general conclusions.” 🡪 What do you mean by small errors?

Response: I estimated that the errors are likely (over 90% probability) to be smaller than 1.0% for false positive and 5.8% for false negatives. I believe that these are small enough to not cause bias in the results given the large sample size.