Summary statistics for narrator analysis

Adrian Barnett

02 September 2022

This document contains summary statistics for our analysis of language over time in papers published on *PubMed*. In total there were 20,876,733 abstracts.

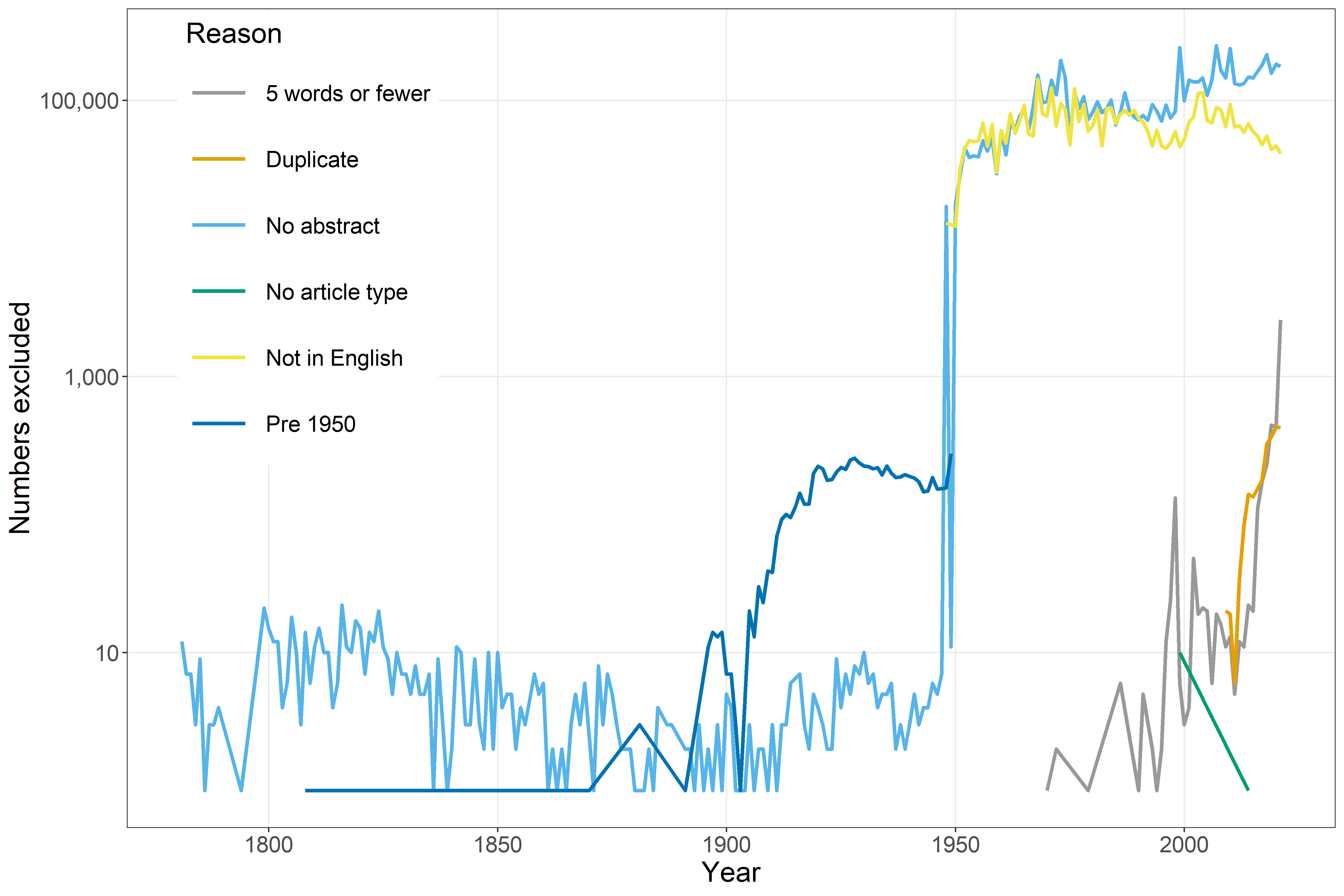
## Excluded abstracts

### Reasons for exclusion

| **reason** | **n** |
| --- | --- |
| No abstract | 7,601,688 |
| Not in English | 4,866,073 |
| Pre 1950 | 7,284 |
| 5 words or fewer | 4,397 |
| Duplicate | 2,311 |
| No article type | 11 |

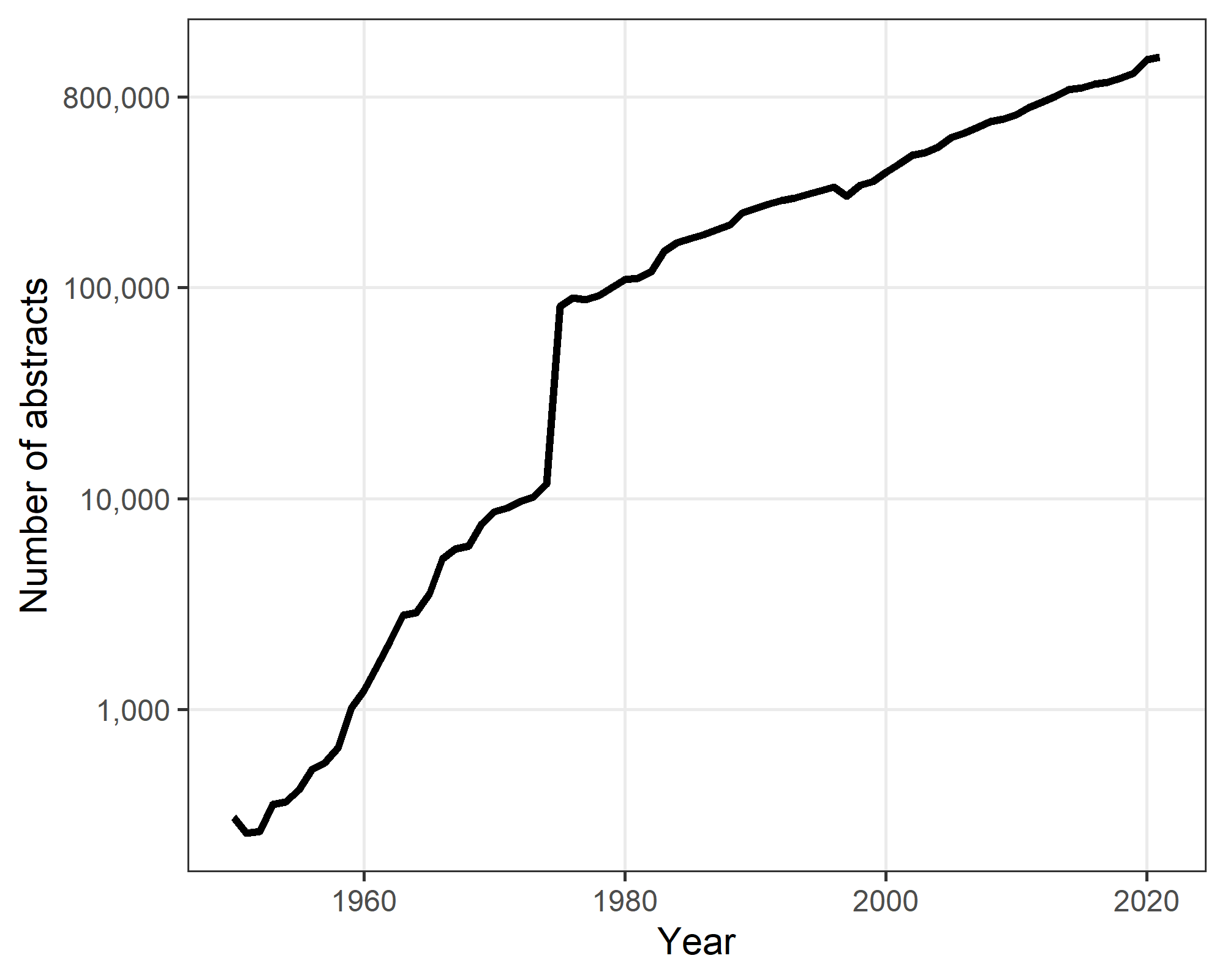
Most abstracts were excluded because there was no abstract or they were not in English.

### Excluded numbers by year



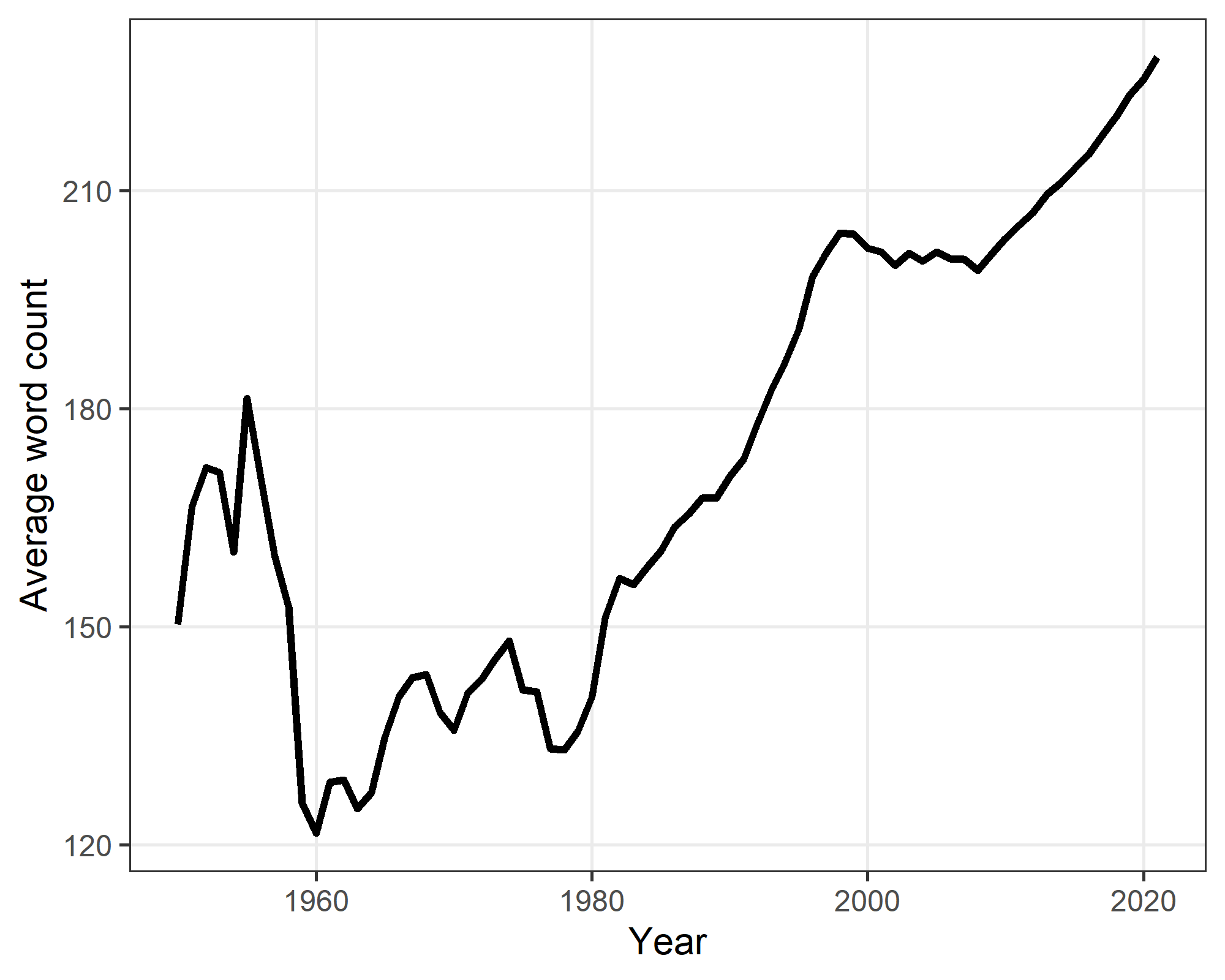
The y-axis is on a log (base 10) scale.

## Included abstracts over time

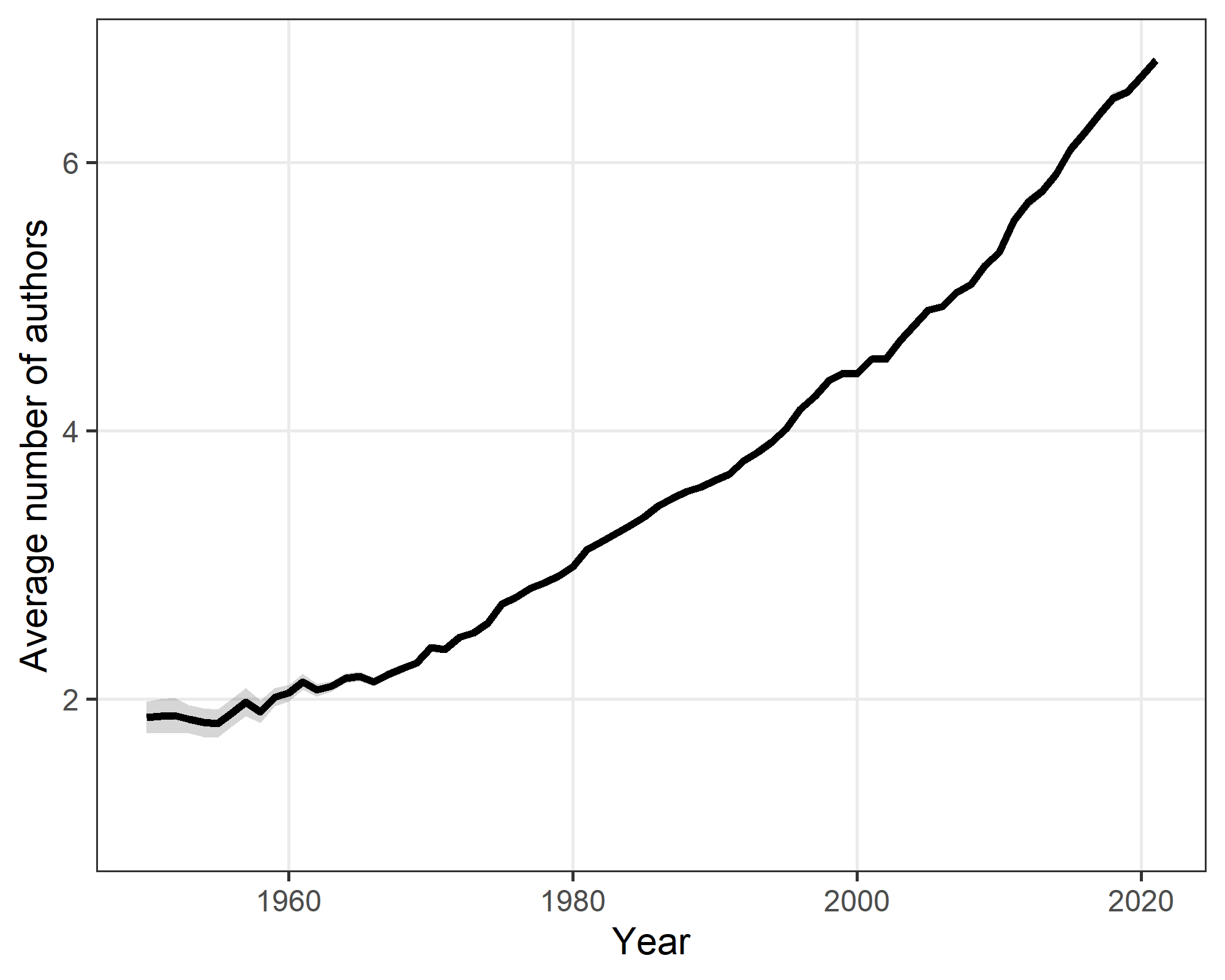


The plot shows the annual number of included abstracts per year. The y-axis is on a log (base 10) scale.

## Abstract word counts over time



## Author numbers over time



The plot shows the average number of authors per paper.

## Article type

| **type** | **n** | **percent** |
| --- | --- | --- |
| Journal Article | 17,485,081 | 83.8 |
| Other | 1,708,340 | 8.2 |
| Case Reports | 1,130,318 | 5.4 |
| Clinical Trial | 502,943 | 2.4 |
| Editorial | 29,487 | 0.1 |
| Comment | 17,713 | 0.1 |
| Preprint | 2,851 | <0.1 |

Most abstracts were journal articles.

## Number of papers per journal

| **Papers per journal** | **n** |
| --- | --- |
| 1 to 10 | 9,953 |
| 11 to 50 | 2,880 |
| 51 to 100 | 1,122 |
| 100+ | 8,996 |

## Top ten journals

These are the top ten journals included in the analysis.

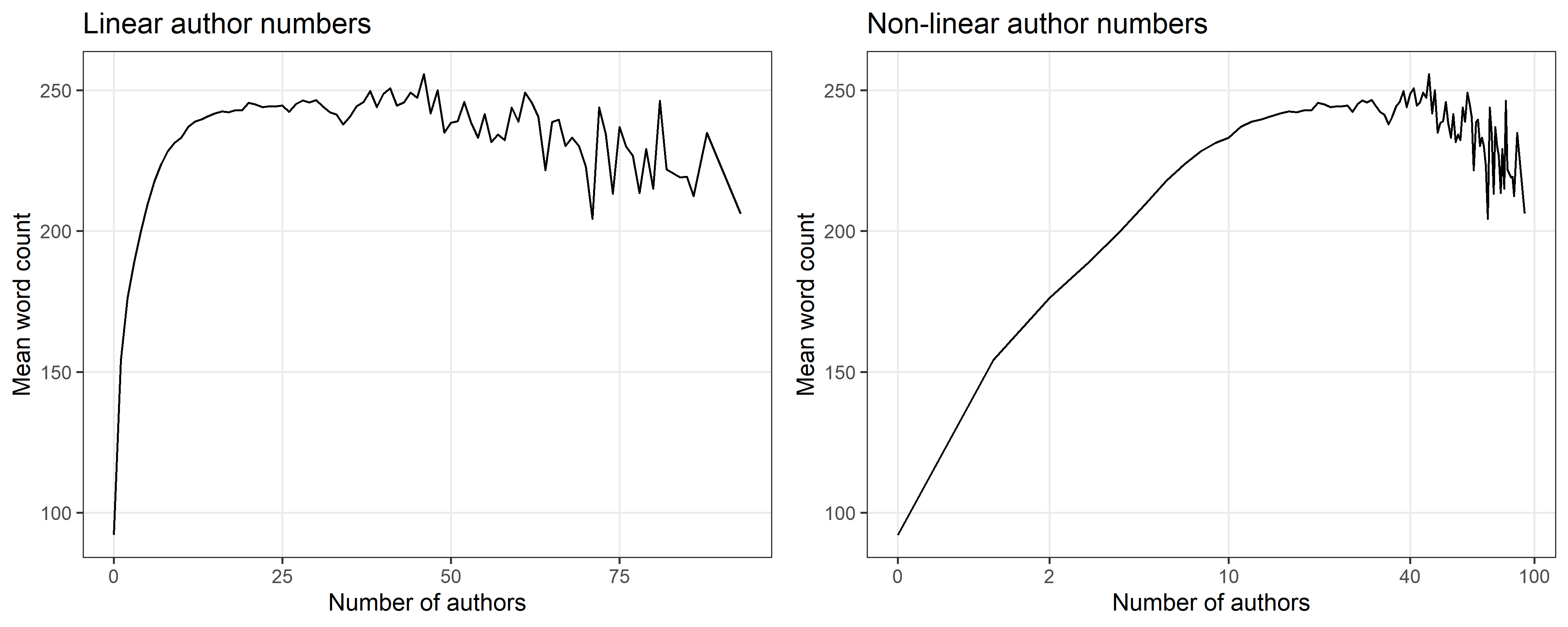
| **Journal name** | **n** |
| --- | --- |
| PLoS One | 249,097 |
| J Biol Chem | 156,260 |
| Sci Rep | 146,726 |
| Proc Natl Acad Sci U S A | 127,334 |
| Biochem Biophys Res Commun | 71,663 |
| Biochim Biophys Acta | 68,458 |
| Phys Rev Lett | 67,354 |
| J Immunol | 57,453 |
| Biochemistry | 56,630 |
| J Am Chem Soc | 56,629 |

## Top ten countries

| **country** | **n** | **percent** |
| --- | --- | --- |
| United States | 5,430,365 | 26.0 |
| \_Missing\_ | 3,190,046 | 15.3 |
| China | 1,655,815 | 7.9 |
| Japan | 1,102,283 | 5.3 |
| United Kingdom | 1,023,553 | 4.9 |
| Germany | 861,268 | 4.1 |
| Italy | 660,656 | 3.2 |
| Canada | 650,520 | 3.1 |
| France | 563,405 | 2.7 |
| Australia | 490,650 | 2.4 |

Using the country of the first author.

# Word count by author number



The plots show the mean word count by the number of authors. The left plot uses authors on a linear scale and the right plot uses a non-linear transform of author numbers. There is a strong non-linear association with an initial large growth in word count followed by a decline for papers with a large number of authors.