Summary statistics for cases and controls

Adrian Barnett

24 March 2025

## Missing data

| **variable** | **n\_miss** | **pct\_miss** |
| --- | --- | --- |
| publisher | 9,504 | 6.00 |
| country | 5,477 | 3.46 |
| open\_access | 42 | 0.0265 |
| typeo | 0 | 0 |
| n\_authors | 0 | 0 |
| word\_count | 0 | 0 |

This table shows the number and percentage of missing data for the independent variables.

## Case and control numbers

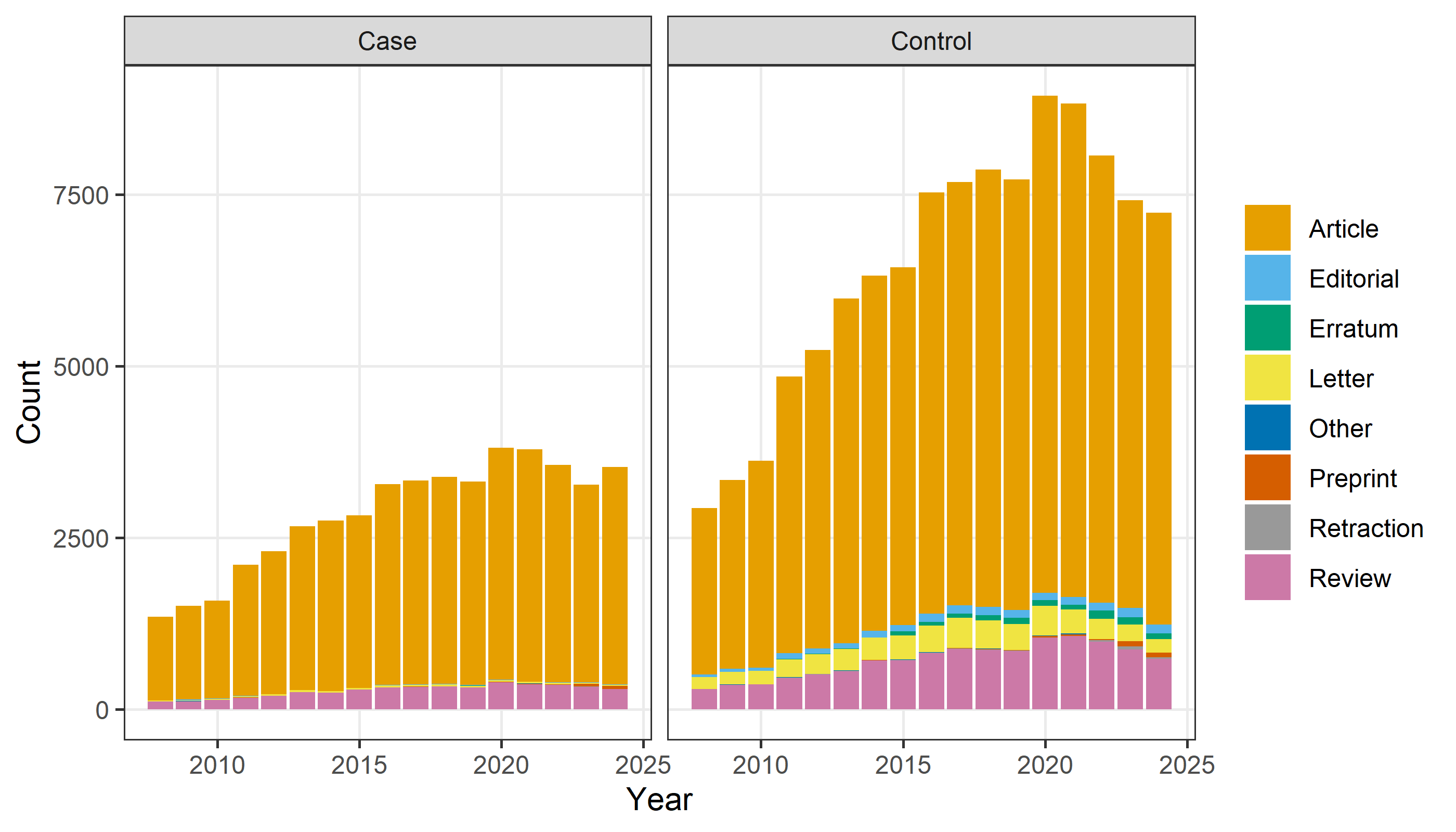
| **case** | **n** | **percent** |
| --- | --- | --- |
| 0 | 110,028 | 69.4 |
| 1 | 48,420 | 30.6 |

The percentages are as close to expected 33:67 split with two controls per case. However, some cases were accidentally double-counted and hence some additional controls were added.

## Article types

| **Type** | **n** | **percent** |
| --- | --- | --- |
| Article | 133,009 | 84 |
| Review | 16,711 | 11 |
| Letter | 5,557 | 4 |
| Editorial | 1,736 | 1 |
| Erratum | 823 | 1 |
| Preprint | 352 | <1 |
| Retraction | 161 | <1 |
| Other | 99 | <1 |

#### Article types by year and case



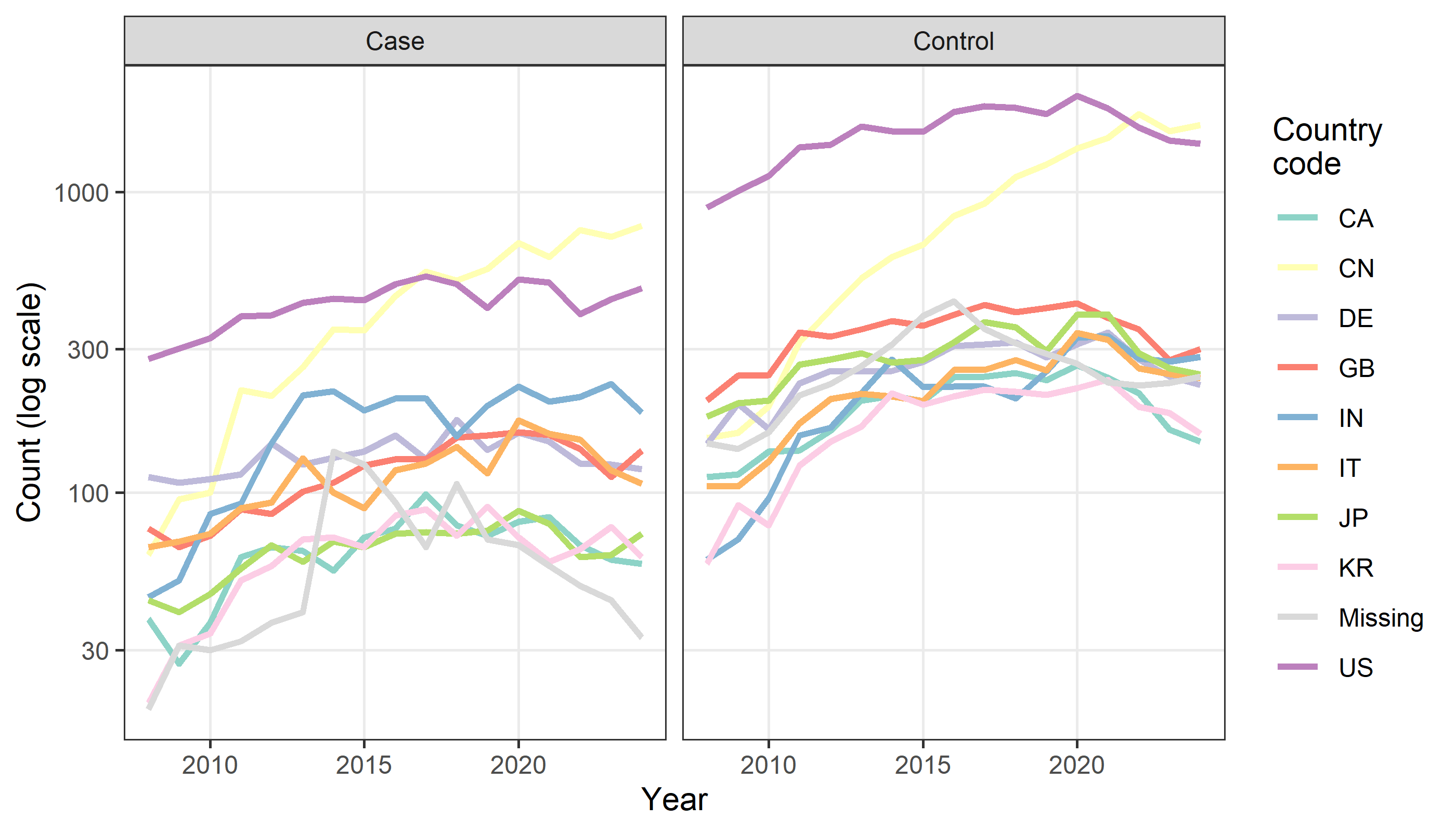
## Publishers (top 20)

| **publisher** | **n** | **percent** |
| --- | --- | --- |
| Elsevier BV | 29,829 | 19 |
| Wiley | 14,778 | 9 |
| Springer Science+Business Media | 12,268 | 8 |
| Missing | 9,504 | 6 |
| BioMed Central | 6,202 | 4 |
| Multidisciplinary Digital Publishing Institute | 6,002 | 4 |
| Lippincott Williams & Wilkins | 5,527 | 3 |
| Oxford University Press | 4,739 | 3 |
| Taylor & Francis | 4,471 | 3 |
| American Chemical Society | 4,299 | 3 |
| Frontiers Media | 4,183 | 3 |
| Nature Portfolio | 4,078 | 3 |
| Public Library of Science | 4,045 | 3 |
| SAGE Publishing | 3,858 | 2 |
| Springer Nature | 2,581 | 2 |
| Royal Society of Chemistry | 2,369 | 1 |
| Medknow | 1,950 | 1 |
| BMJ | 1,660 | 1 |
| Hindawi Publishing Corporation | 1,457 | 1 |
| Wolters Kluwer | 977 | 1 |

## Countries (top 20)

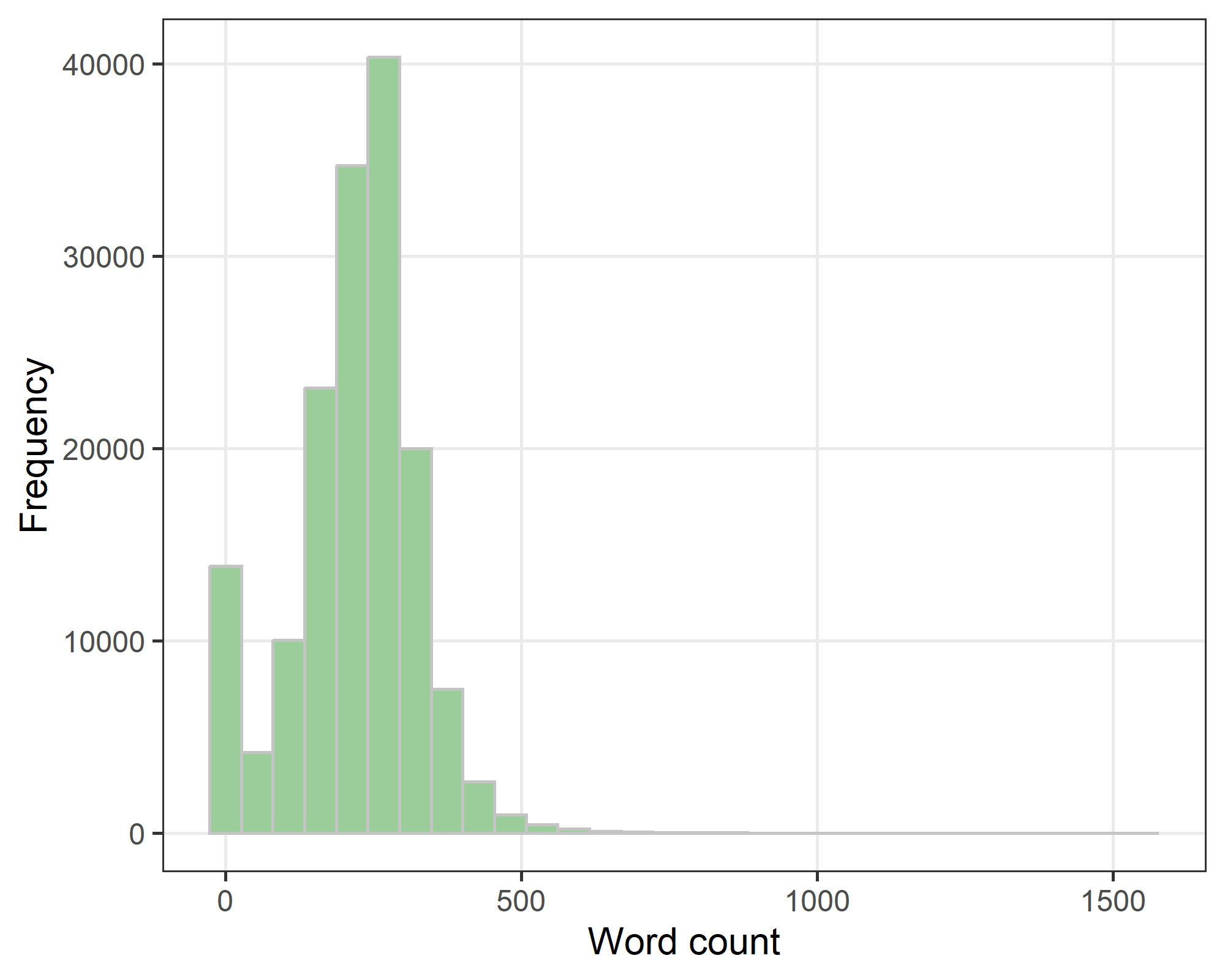
| **country** | **n** | **percent** |
| --- | --- | --- |
| United States | 34,013 | 21 |
| China | 22,330 | 14 |
| United Kingdom | 7,799 | 5 |
| Germany | 6,646 | 4 |
| India | 6,517 | 4 |
| Japan | 5,984 | 4 |
| Italy | 5,707 | 4 |
| Missing | 5,477 | 3 |
| Canada | 4,374 | 3 |
| South Korea | 3,994 | 3 |
| France | 3,892 | 2 |
| Australia | 3,856 | 2 |
| Turkey | 3,854 | 2 |
| Spain | 3,539 | 2 |
| Brazil | 3,353 | 2 |
| Iran | 2,622 | 2 |
| Netherlands | 2,592 | 2 |
| Poland | 1,811 | 1 |
| Taiwan | 1,683 | 1 |
| Switzerland | 1,640 | 1 |

#### Countries (top 10) over time and by case



We show the top ten over time. We use the top 10 to reduce clutter. We use the log-scale on the y-axis as two countries are dominant.

## Word count

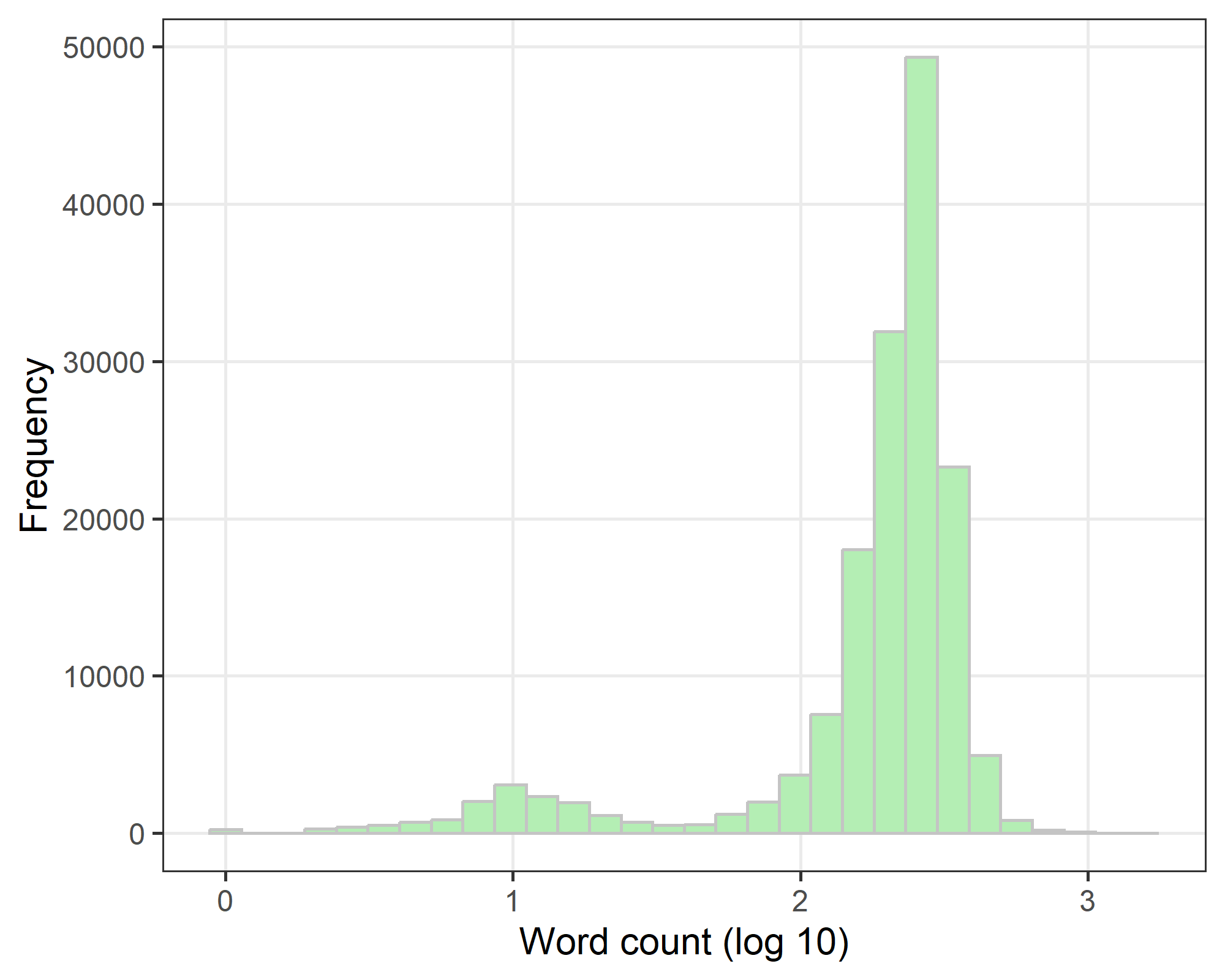


The word count distribution is bimodal because of short abstracts such as letters and retractions, and those abstracts that only have a title.

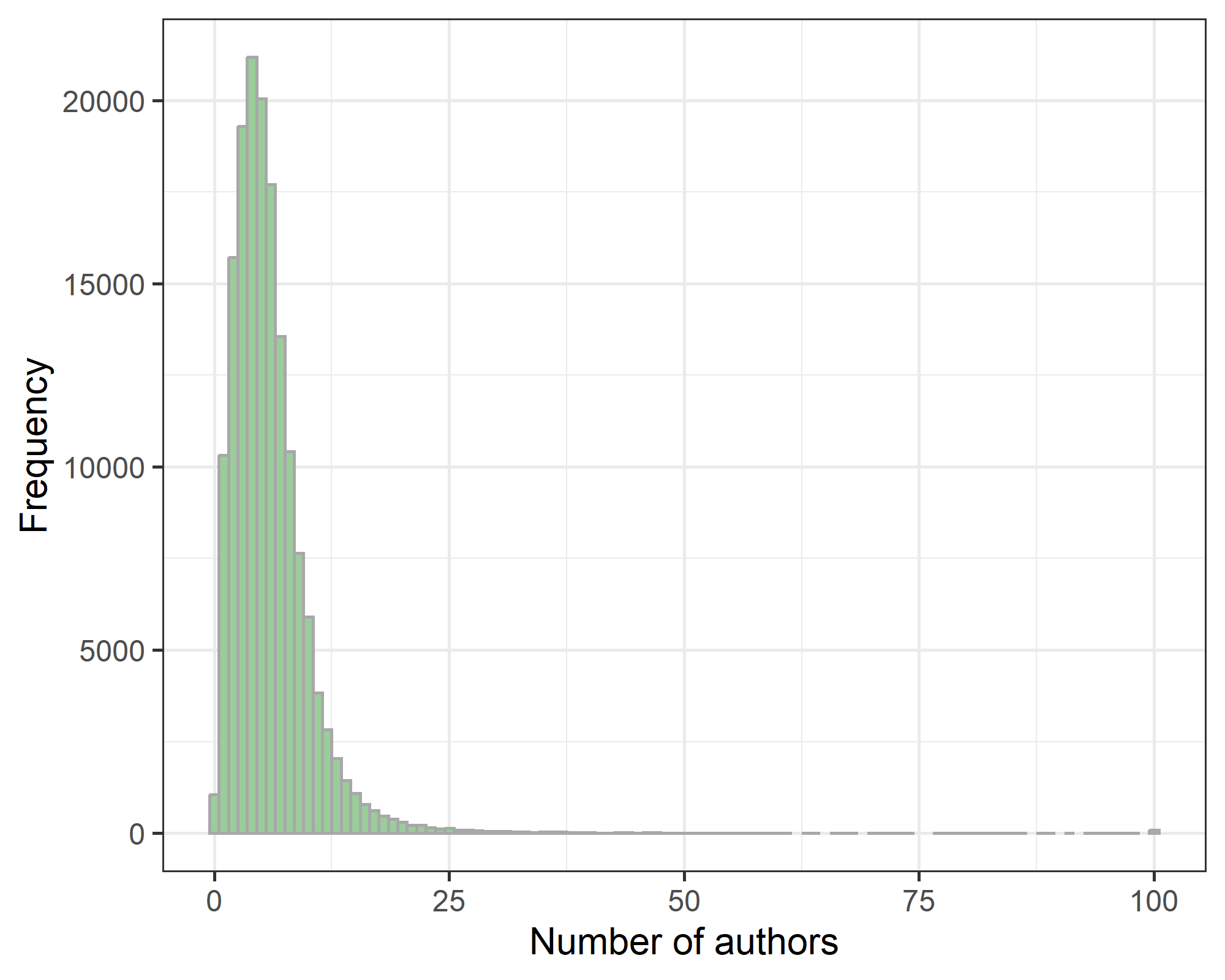
#### Word count - Summary statistics

| **median** | **q1** | **q3** |
| --- | --- | --- |
| 231 | 165 | 282 |

## Word count - log-transformed



## Author numbers



#### Author numbers - Summary statistics

| **median** | **q1** | **q3** |
| --- | --- | --- |
| 5 | 3 | 7 |

## Open access

| **Open access** | **n** | **percent** |
| --- | --- | --- |
| FALSE | 79,710 | 50 |
| TRUE | 78,696 | 50 |
| Missing | 42 | <1 |

## Retracted

| **retracted** | **n** | **percent** |
| --- | --- | --- |
| FALSE | 158,264 | 99.9 |
| TRUE | 184 | 0.1 |

## Overall summary table

This table is at the abstract level to avoid double-counting abstracts with more than one error.

Stratified by case  
 level Case   
 n 47703   
 typeo (%) Article 42573 (89.2)   
 Editorial 84 ( 0.2)   
 Erratum 26 ( 0.1)   
 Letter 339 ( 0.7)   
 Other 19 ( 0.0)   
 Preprint 115 ( 0.2)   
 Retraction 9 ( 0.0)   
 Review 4538 ( 9.5)   
 n\_authors (median [IQR]) 5.0 [4.0, 8.0]   
 word\_count (median [IQR]) 258.0 [205.0, 302.0]  
 Stratified by case  
 Control   
 n 110028   
 typeo (%) 89784 (81.6)   
 1649 ( 1.5)   
 797 ( 0.7)   
 5216 ( 4.7)   
 80 ( 0.1)   
 235 ( 0.2)   
 152 ( 0.1)   
 12115 (11.0)   
 n\_authors (median [IQR]) 5.0 [3.0, 7.0]   
 word\_count (median [IQR]) 219.0 [147.0, 274.0]