Results for predictors of citations using the Bayesian horseshoe model

Adrian Barnett

23 March, 2025

The sample size for this analysis was 157,731.

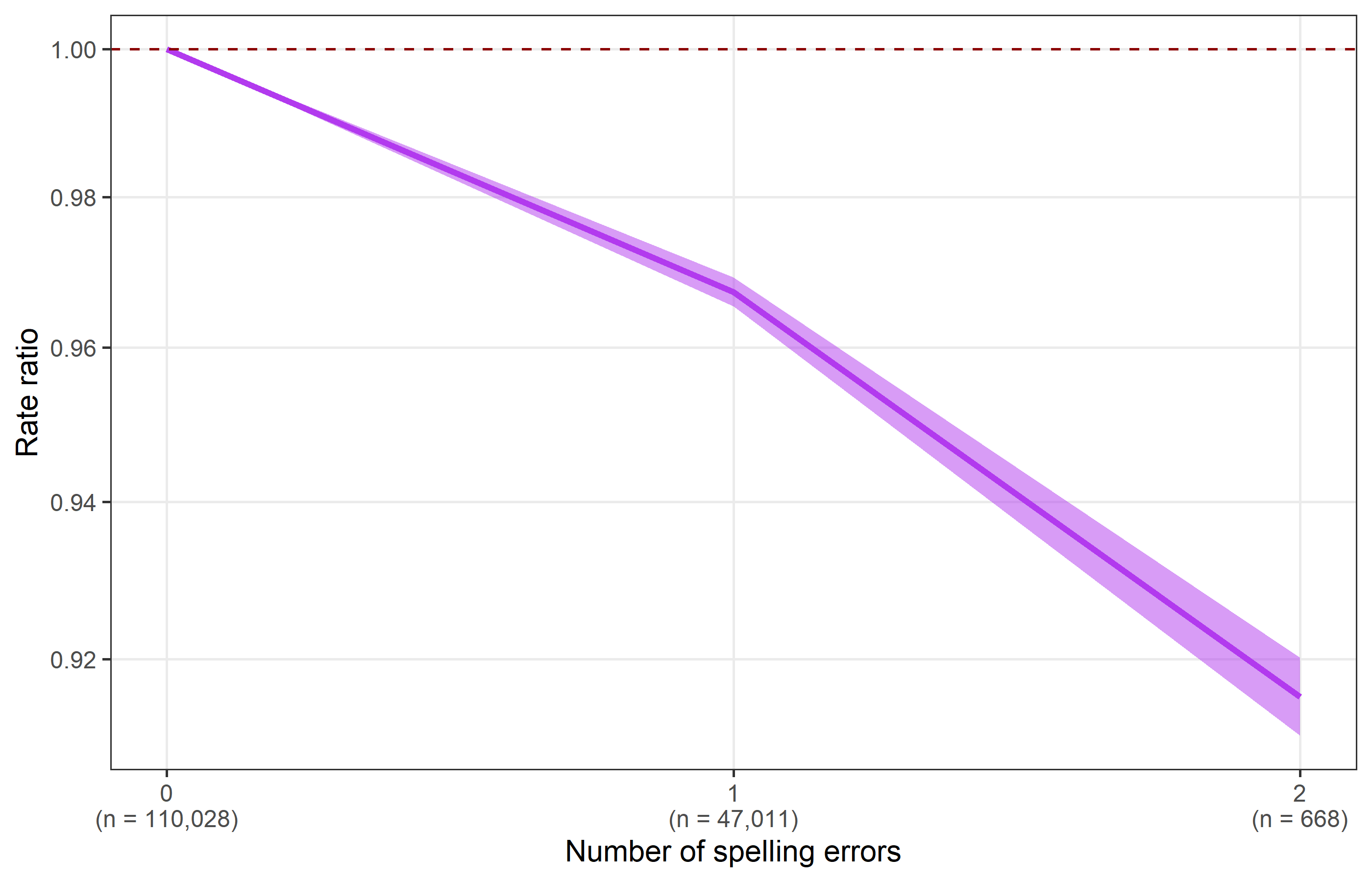
## Table of results

| **Variable** | **mean** | **lower** | **upper** | **posterior probability** |
| --- | --- | --- | --- | --- |
| Authors +5 | 1.16 | 1.16 | 1.16 | < 0.001 |
| Article | 1.00 |  |  |  |
| Editorial | 0.34 | 0.34 | 0.35 | < 0.001 |
| Erratum | 0.13 | 0.12 | 0.13 | < 0.001 |
| Letter | 0.25 | 0.25 | 0.26 | < 0.001 |
| Other | 0.56 | 0.53 | 0.60 | < 0.001 |
| Preprint | 0.89 | 0.85 | 0.93 | < 0.001 |
| Retraction | 0.08 | 0.07 | 0.09 | < 0.001 |
| Review | 2.24 | 2.23 | 2.24 | < 0.001 |

The table shows the mean rate ratio, 95% credible interval, and posterior probability that the estimate is zero.

## Estimate of spelling errors

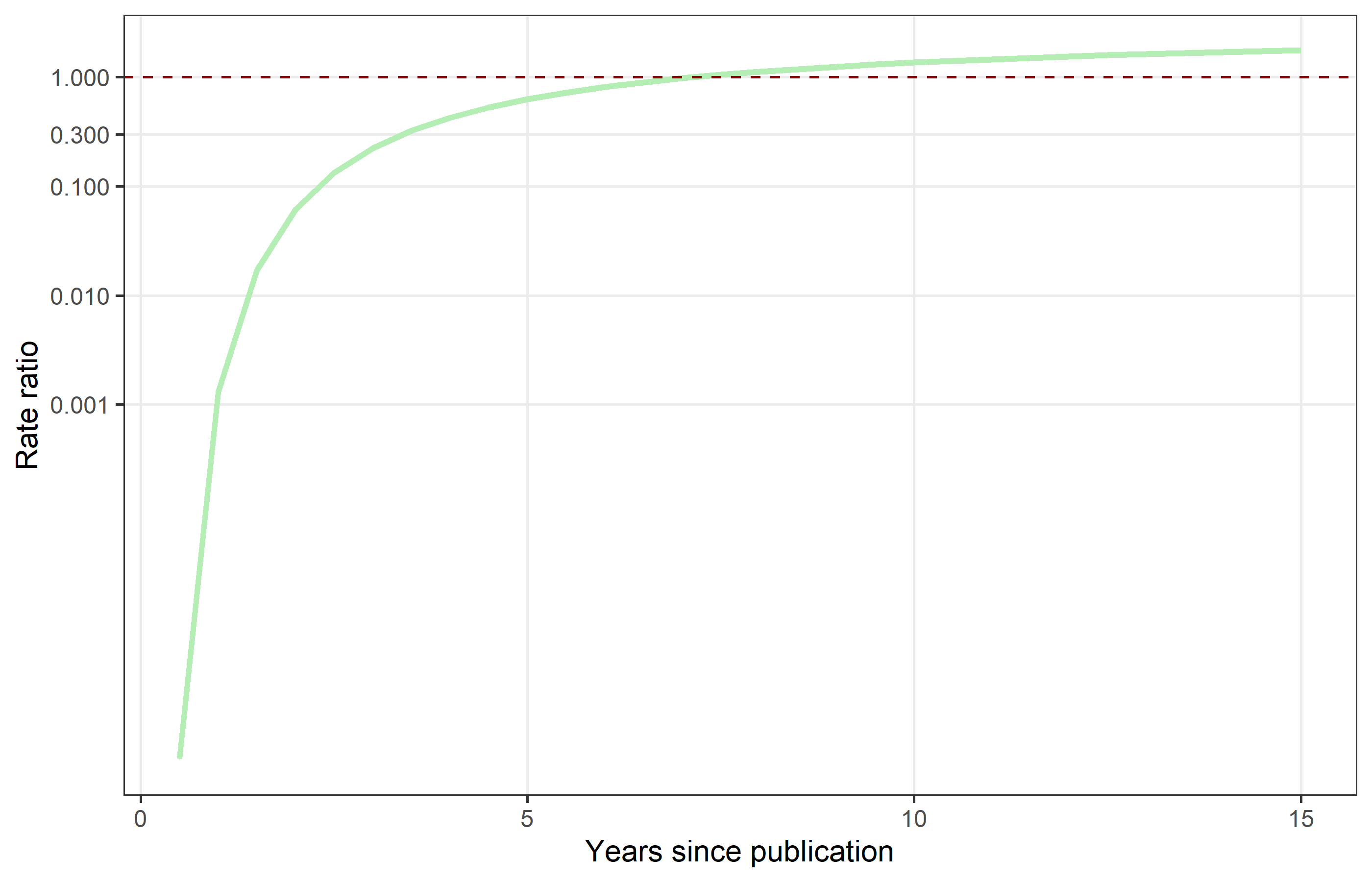
We used the fractional polynomial approach to allow a potentially non-linear association between spelling errors and citation counts. The best transformation was the squared, and we plot the estimated effects below.



The maximum number of errors was 4, but we only plot up to 2 as there were very few abstracts with 3 or 4 errors.

## Estimate of time

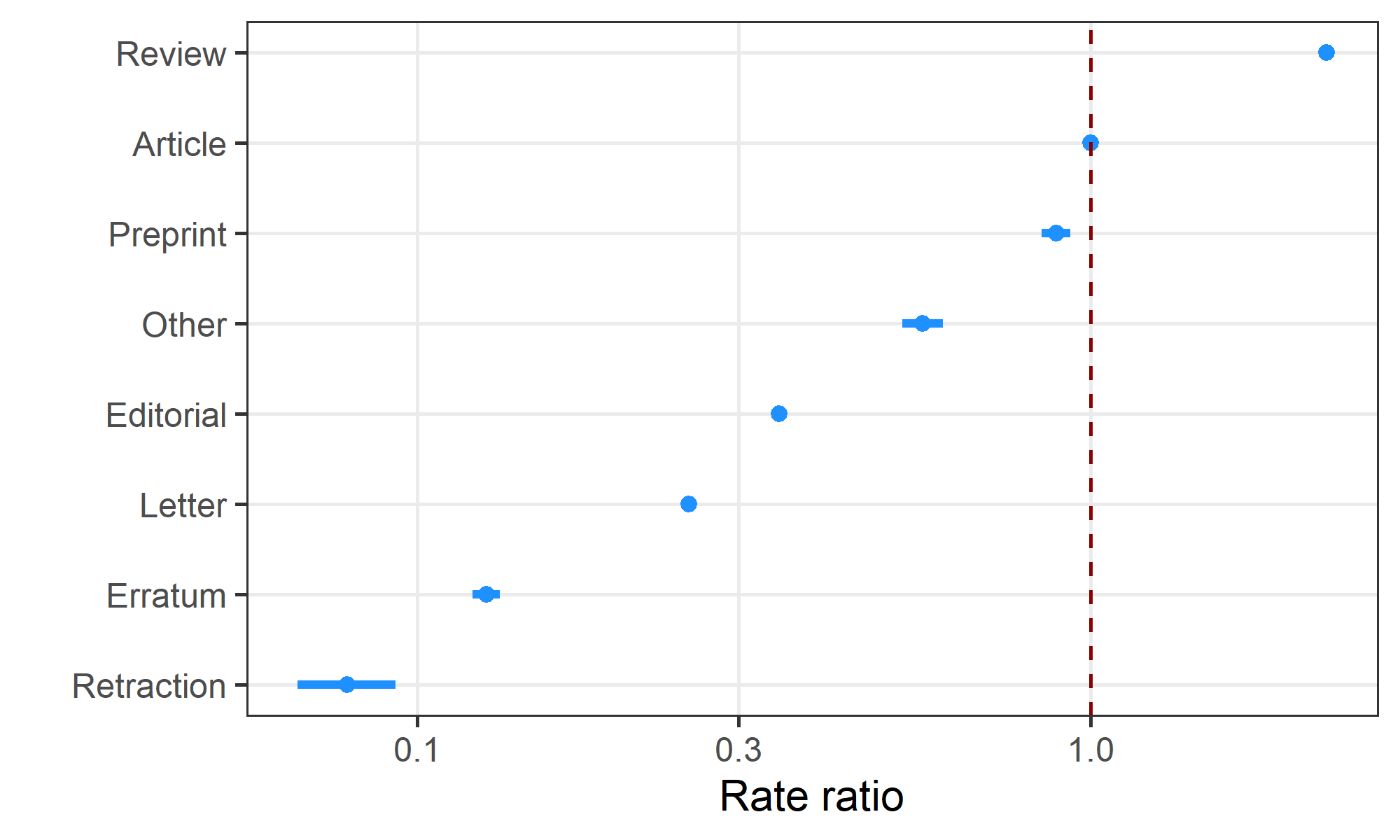
Time since publication was modelled to account for a greater number of citations for papers that have been published for longer. We used the fractional polynomial approach to allow a potentially non-linear association between time and citation counts. The best transformation was the inverse, and we plot the estimated effects below.



The reference time was 7 years.

The plot shows that, on average, papers accrue citations over time, but the gain in citations slows over time.

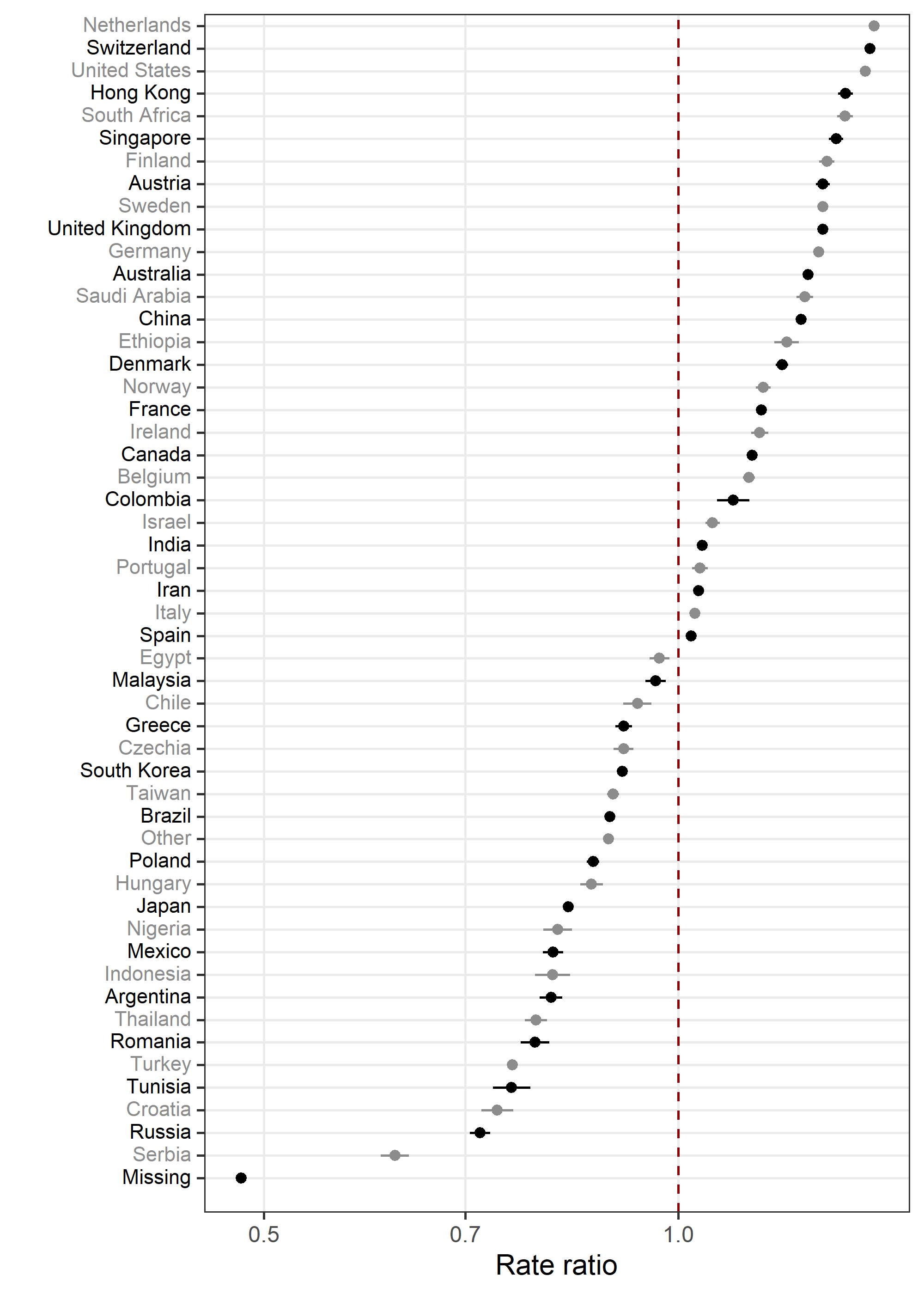
## Article types



A type of “Article” was the reference category. Article type had a very strong effect on citation counts.

The credible intervals around most of the estimates are very narrow.

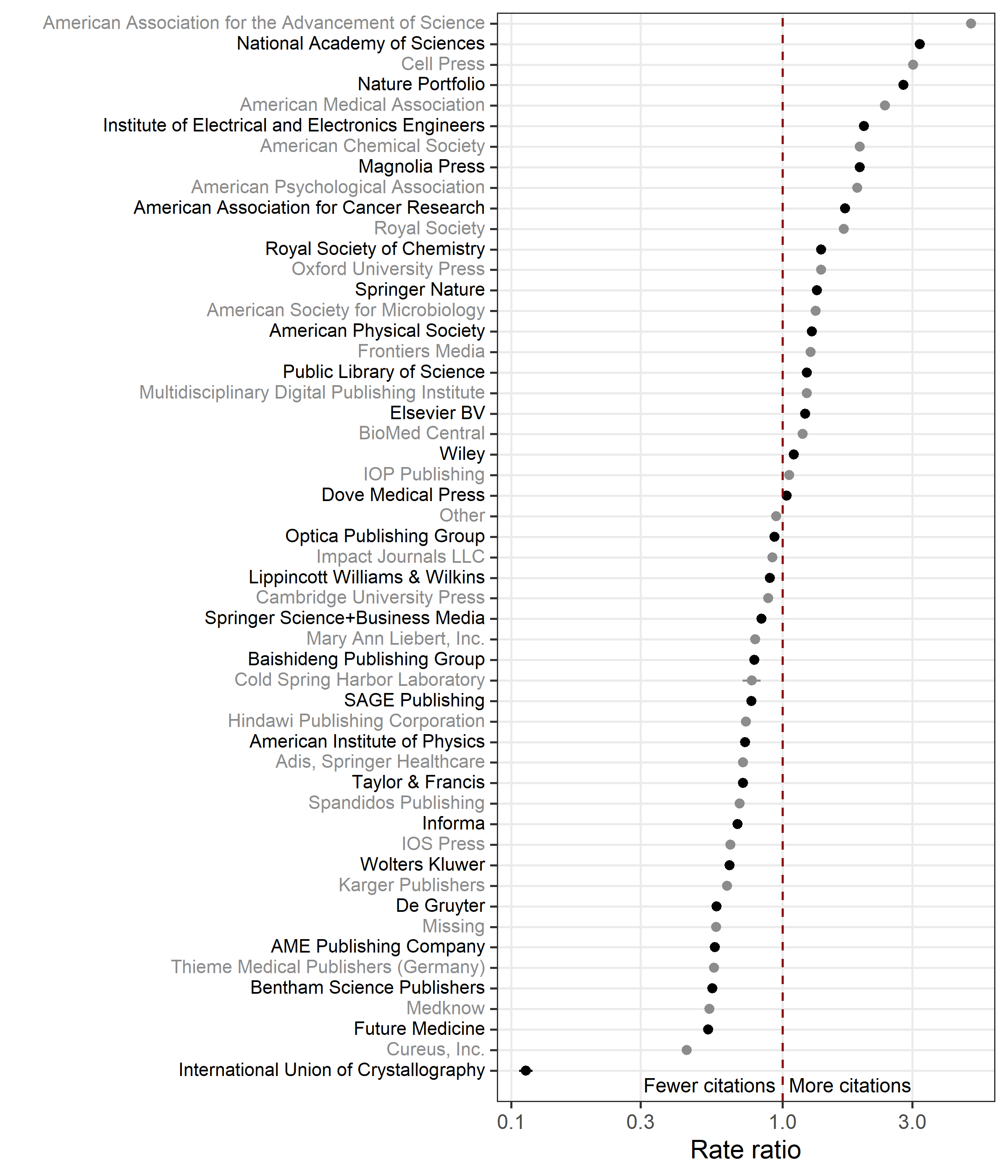
## Countries



The plot only shows those countries where there is an over 0.99 probability that the mean is not zero.

The credible intervals for most countries are very narrow and so not visible.

## Publisher



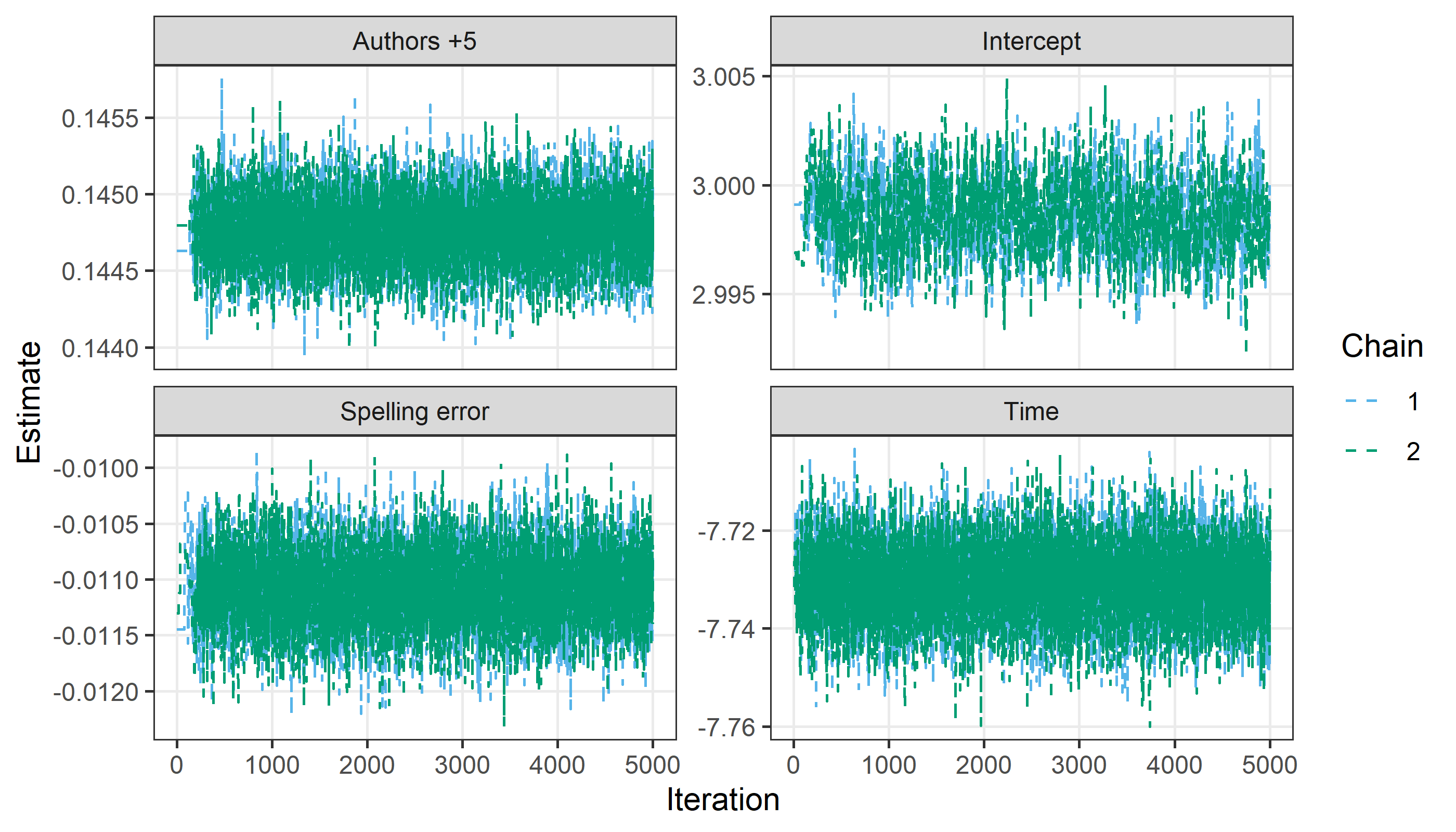
The plot only shows those publishers where there is an over 0.99 probability that the mean is not zero.

The credible intervals for most publishers are very narrow and so not visible.

### Check model convergence

We used 5,000 samples thinned by 5 with two chains.

#### Chains for the intercept and word count estimates



The chains appear to have converged with good mixing.

#### Densities for the intercept and word count estimates

