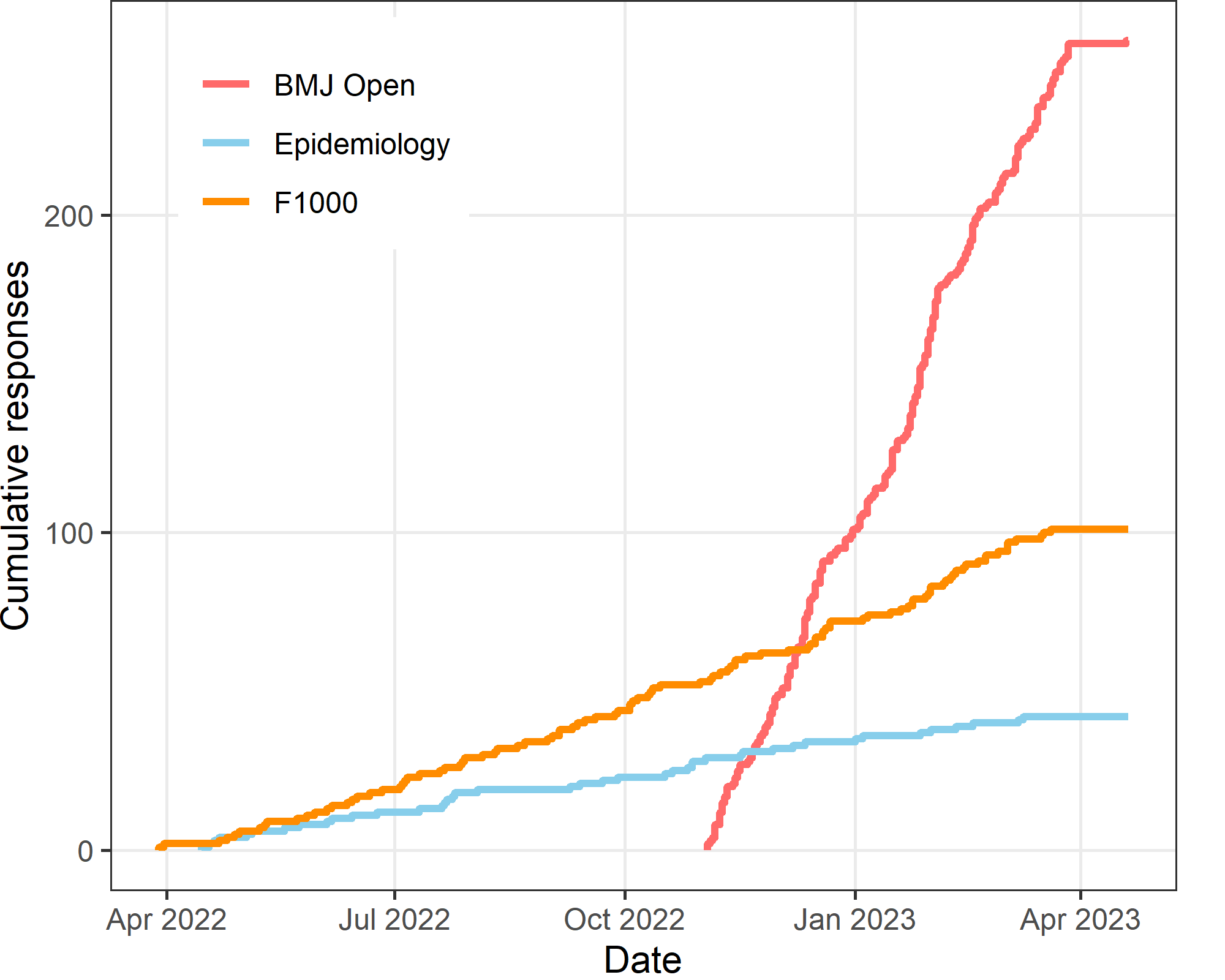
Summary statistics for peer review uncertainty study

Adrian Barnett

19 March, 2024

# Recruitment

Plot of recruitment over time by journal.



The total number recruited is 398. The first participant was recruited on 29 Mar 2022 and the last on 19 Apr 2023, which is 386 days.

### Number of responses by journal

| **Journal** | **n** | **percent** |
| --- | --- | --- |
| BMJ Open | 255 | 64 |
| Epidemiology | 42 | 11 |
| F1000 | 101 | 25 |

## Time taken to answer questions

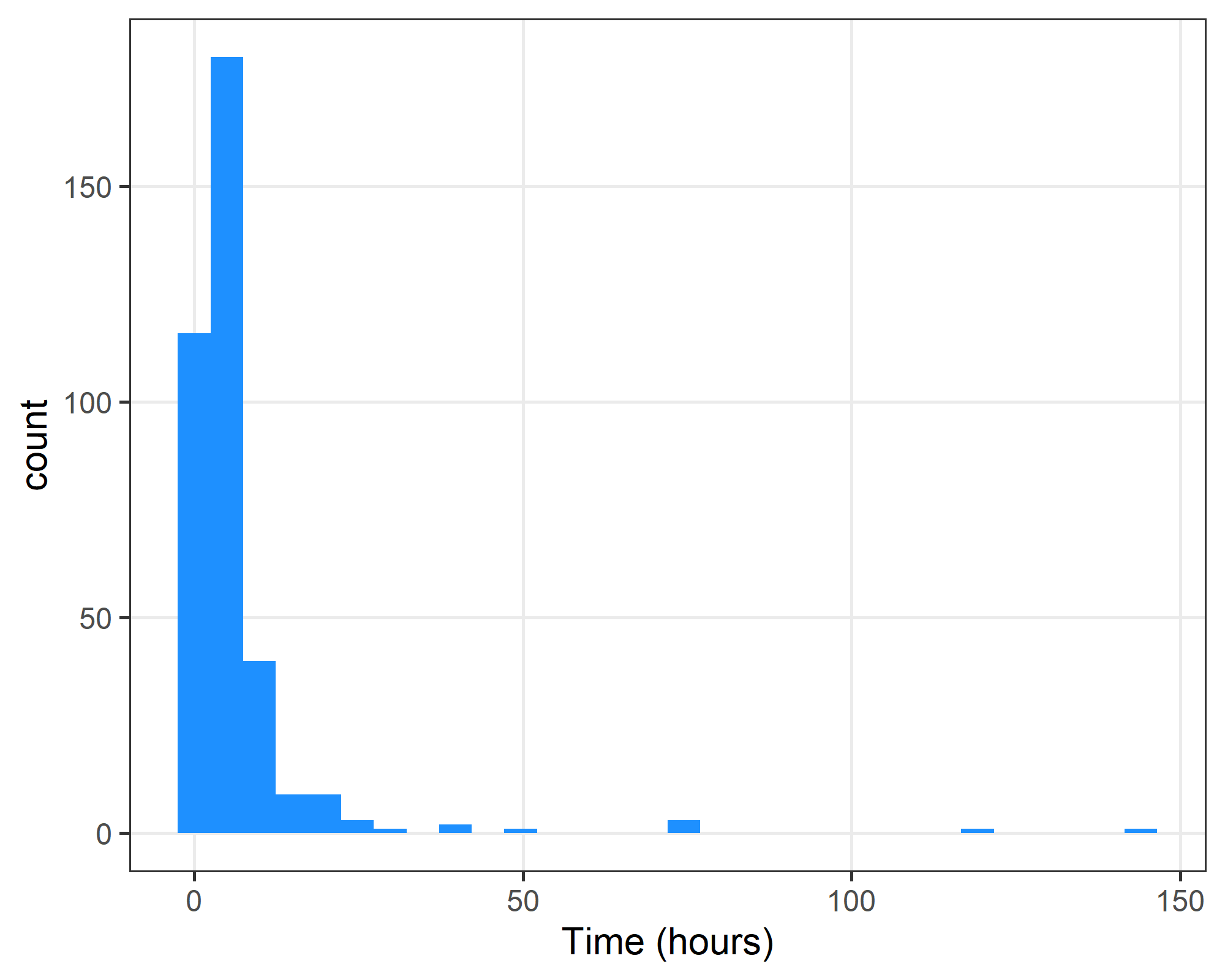
| **Q1** | **median** | **Q3** |
| --- | --- | --- |
| 2 | 3 | 4 |

The summary statistics show the time in minutes to complete the online questions. Q1 is the lower quartile and Q3 is the upper quartile.

# Reviewers’ characteristics

### Time taken for review

How much time did you spend on your review in hours? Include your time reading the paper and providing feedback.



There are a few large positive outliers where reviewers spent many weeks on the paper. It is possible that these reviewers wrongly answered in terms of minutes rather than hours.

##### Summary statistics on time taken (hours)

| **missing** | **Q1** | **Median** | **Q3** |
| --- | --- | --- | --- |
| 32 | 2 | 3 | 6 |

### Reviewers’ gender

| **Gender** | **n** | **percent** |
| --- | --- | --- |
| Female | 177 | 44 |
| Male | 204 | 51 |
| Missing | 11 | 3 |
| Non-binary / third gender | 1 | 0 |
| Prefer not to say | 5 | 1 |

### Reviewers’ experience

How many years have you worked in research? Answer in terms of working years, e.g., if 6 years of working half-time then answer 3 years (“Less than 5 years”).

| **Experience** | **n** | **percent** |
| --- | --- | --- |
| Less than 5 years | 46 | 12 |
| 6 to 10 years | 108 | 27 |
| 11 to 15 years | 84 | 21 |
| 16 to 20 years | 45 | 11 |
| 21 years or more | 104 | 26 |
| Missing | 11 | 3 |

### Reviewers’ country

| **Country** | **n** | **percent** |
| --- | --- | --- |
| USA | 78 | 20 |
| Australia | 46 | 12 |
| UK | 36 | 9 |
| Canada | 19 | 5 |
| Missing | 16 | 4 |
| Germany | 14 | 4 |
| Italy | 14 | 4 |
| India | 13 | 3 |
| Netherlands | 9 | 2 |
| Ethiopia | 7 | 2 |
| France | 7 | 2 |
| Ireland | 7 | 2 |
| Poland | 7 | 2 |
| Spain | 7 | 2 |
| Sweden | 7 | 2 |
| Malaysia | 6 | 2 |
| South Africa | 6 | 2 |
| Other | 99 | 24 |

To reduce the size of the table, we combine the countries that were given 5 times or fewer into “Other”.

# Paper characteristics

### Paper type

| **Paper type** | **n** | **percent** |
| --- | --- | --- |
| Original research | 208 | 52 |
| Protocol | 76 | 19 |
| Missing | 36 | 9 |
| Original article | 30 | 8 |
| Brief report | 8 | 2 |
| Review | 5 | 1 |
| Systematic review | 5 | 1 |
| Software tool article | 4 | 1 |
| Cohort profile | 3 | 1 |
| Method article | 3 | 1 |
| Opinion article | 3 | 1 |
| Case report | 2 | 1 |
| Communication | 2 | 1 |
| Data note | 2 | 1 |
| Genome note | 2 | 1 |
| Policy brief | 2 | 1 |
| Research letter | 2 | 1 |
| Review article | 2 | 1 |
| Case study | 1 | 0 |
| Commentary | 1 | 0 |
| Validation study | 1 | 0 |

The rows are ordered by frequency. Those with a missing paper type were the papers that could not be matched to the reviews.

We combined “Study protocol” and “Protocol”, and “Research article” and “Original research”.

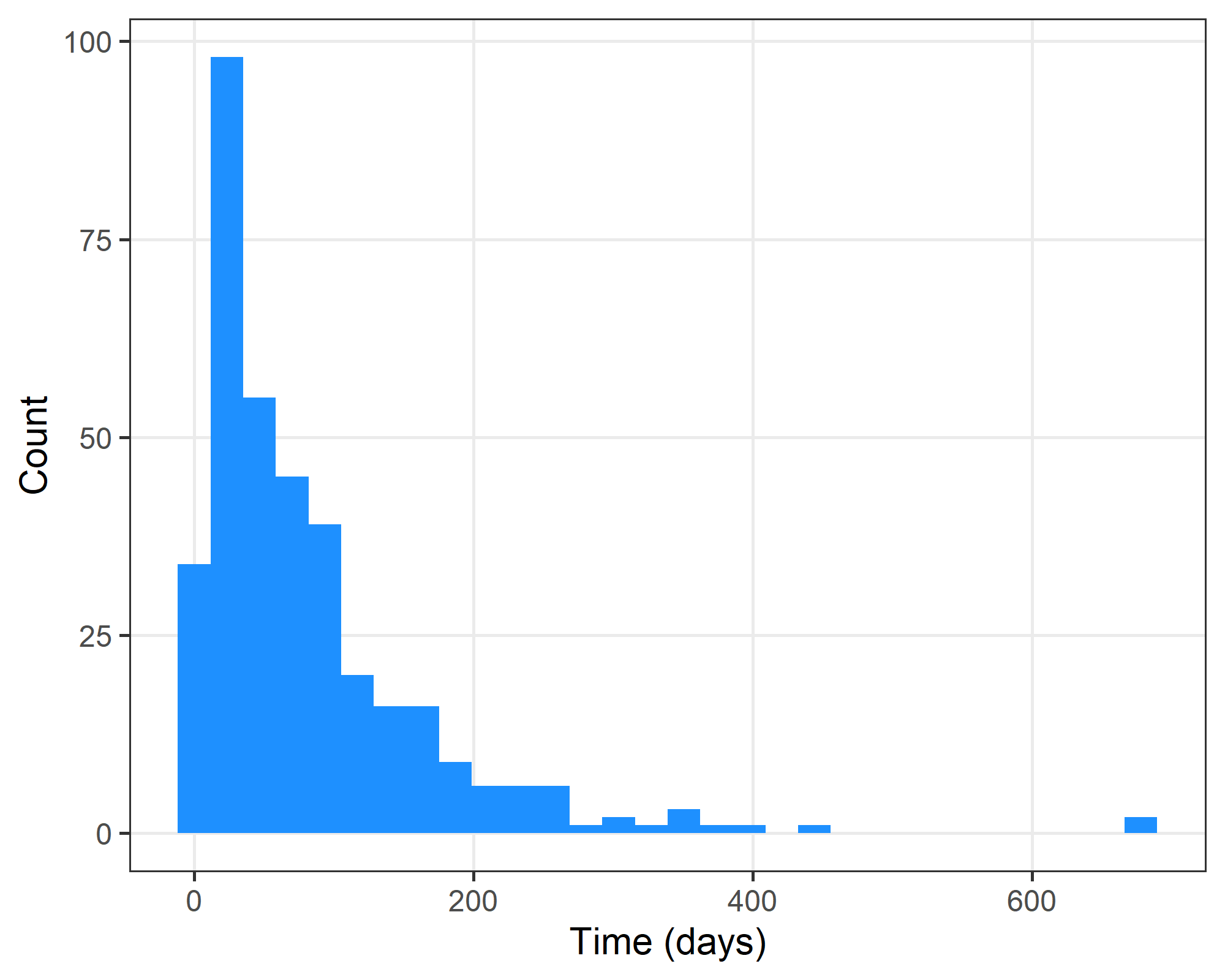
### Version number

| **Version number** | **n** | **percent** |
| --- | --- | --- |
| 1 | 304 | 76 |
| 2 | 52 | 13 |
| 3 | 6 | 2 |
| Missing | 36 | 9 |

The paper’s version number. Those with a missing version number were the papers that could not be matched to the reviews.

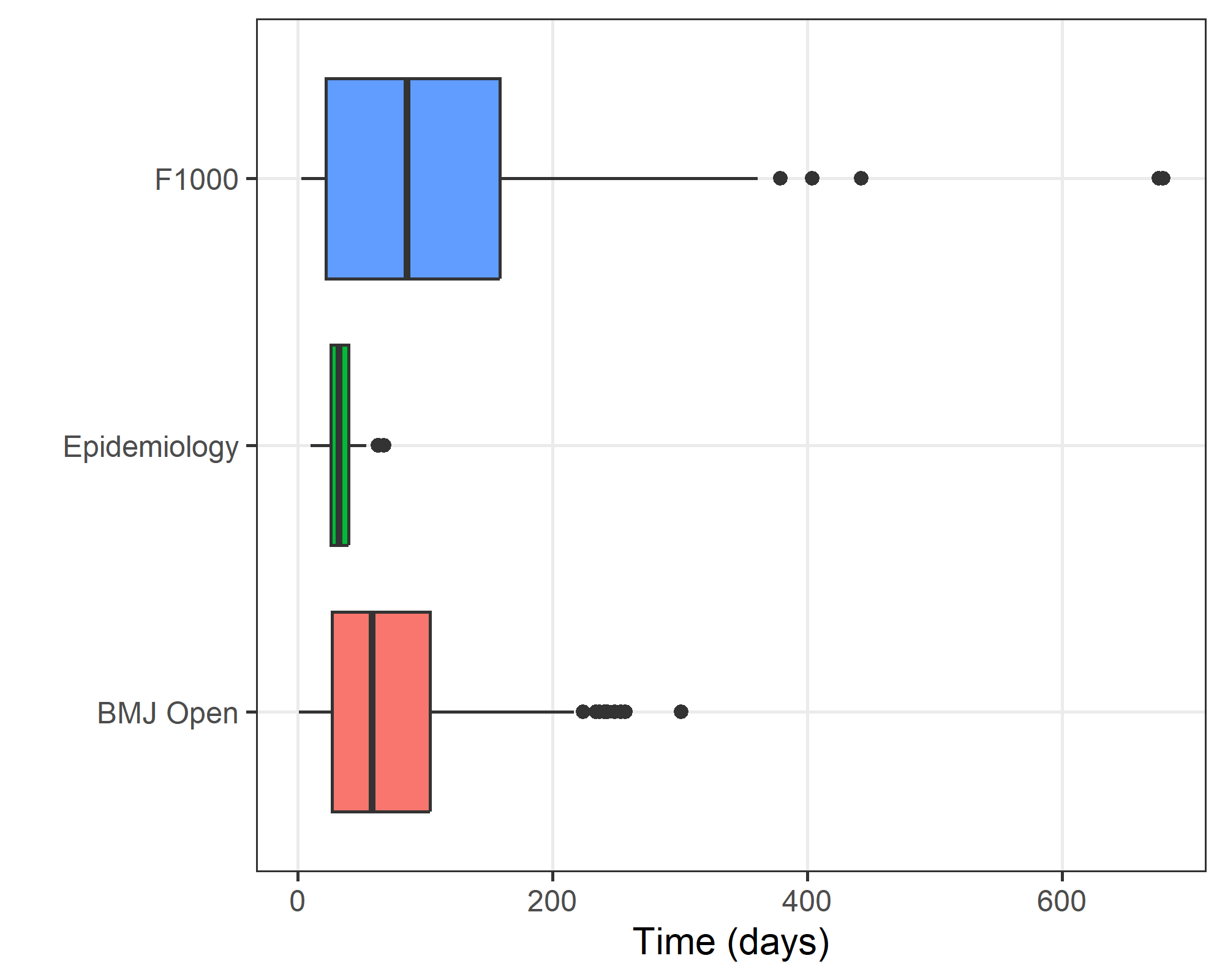
### Time between journal submission and review

This is the time (in days) between the date the paper was submitted to the journal and the date of the review.



The median time was 56 with a first to third quartile of 26 to 108.

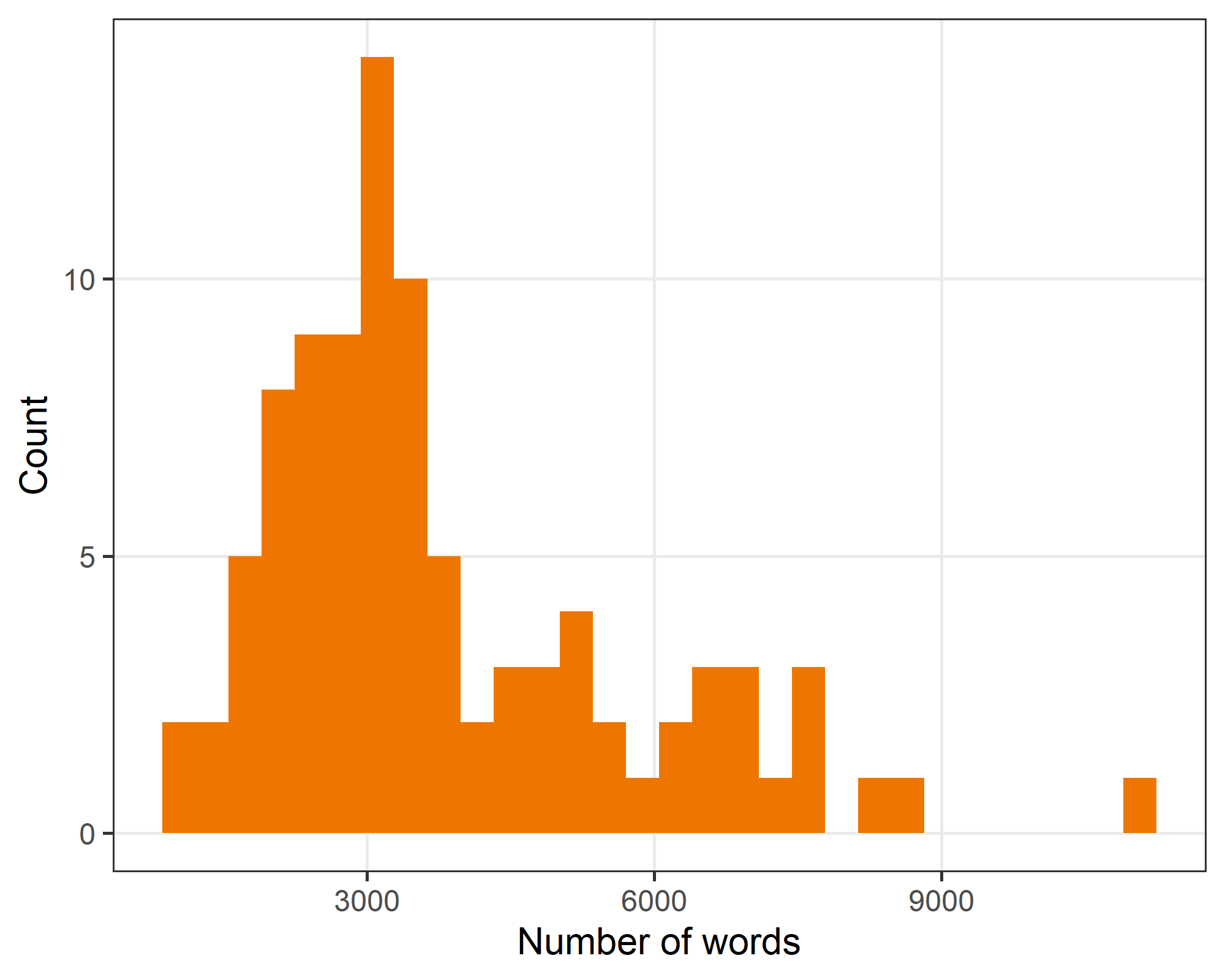
### Time between journal submission and review (by journal)



There are two outliers for F1000 which were checked.

### Number of words

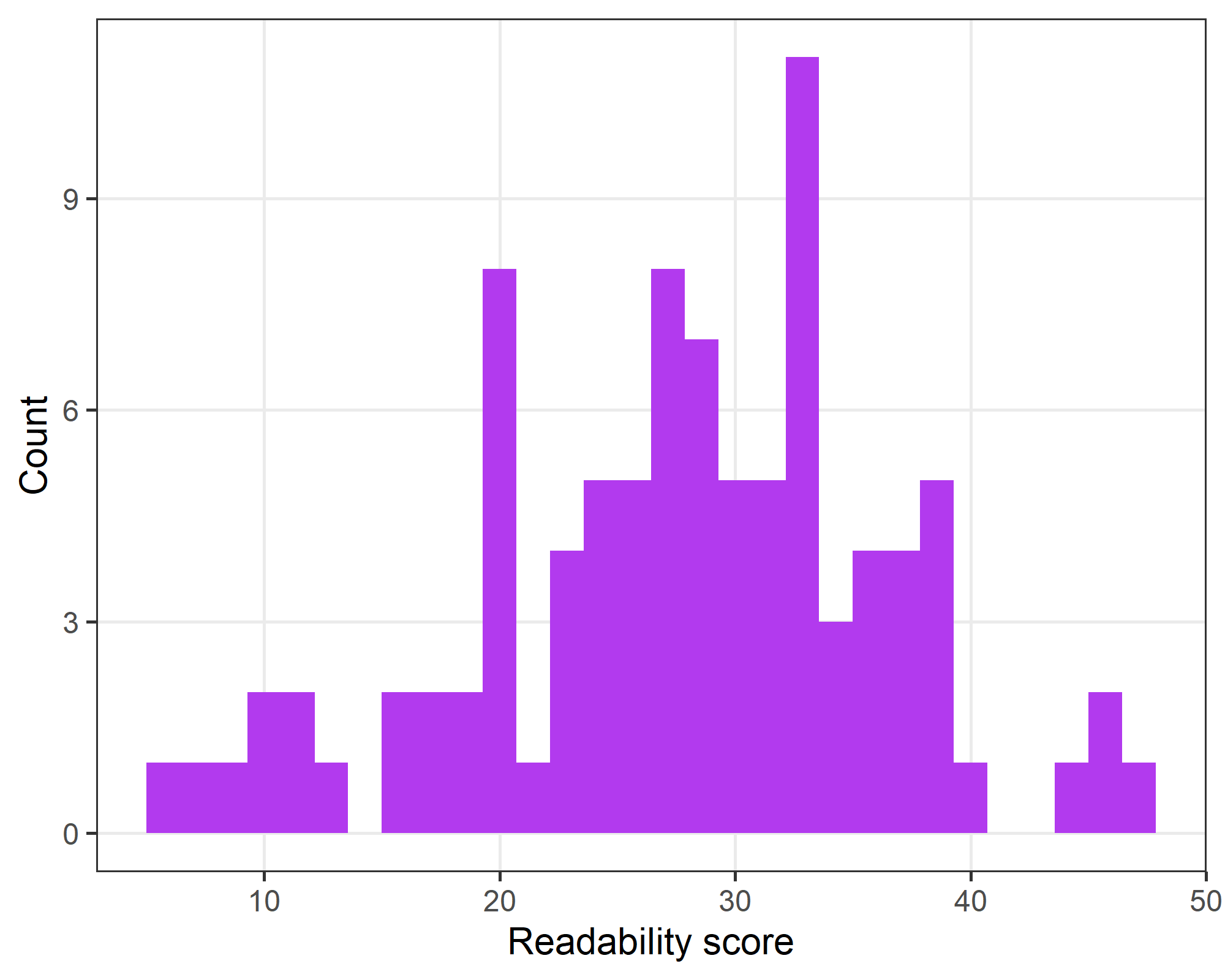
The number of words was only available for F1000. We looked at the words in the main text, so excluding the abstract, references, tables and figures.



| **Q1** | **median** | **Q3** |
| --- | --- | --- |
| 2,450 | 3,246 | 4,920 |

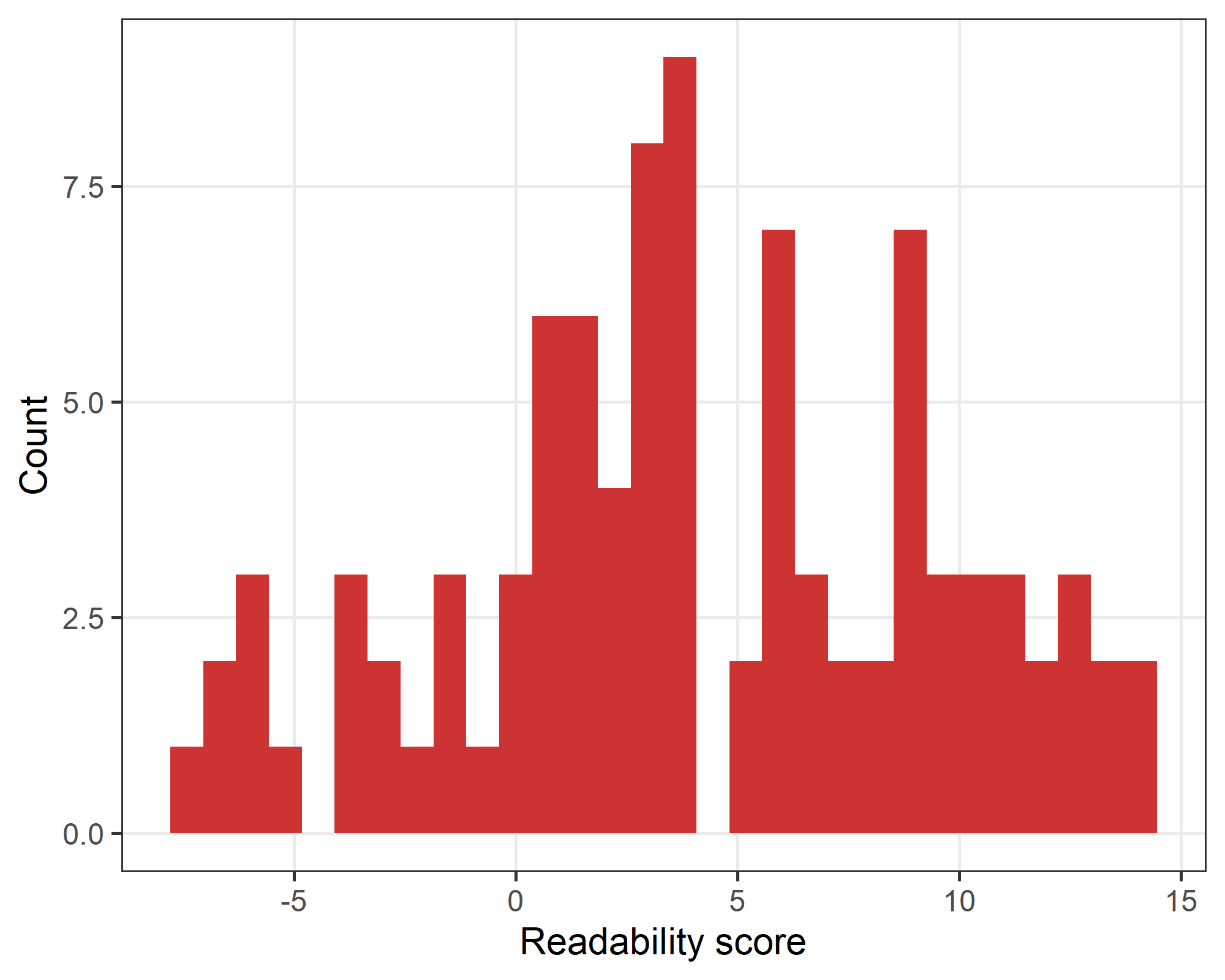
### Flesch’s Reading Ease Score

This score is on a scale of 1 to 100. The higher the score, the easier it is to understand the document.



| **Q1** | **median** | **Q3** |
| --- | --- | --- |
| 22 | 28 | 33 |

### The New Dale-Chall Readability formula



| **Q1** | **median** | **Q3** |
| --- | --- | --- |
| 1 | 4 | 9 |

# Uncertainty in peer review

### Number who could not answer the question

As eliciting uncertainty is a relatively unusual concept, we gave reviewers the option of saying they could not answer. Below the question on uncertainty, we included the text: “If you cannot answer the question then please click this radio button.” The table below shows the number who clicked this button.

| **Could not answer** | **n** | **percent** |
| --- | --- | --- |
| Not ticked | 390 | 98 |
| Ticked | 8 | 2 |

Only a small percentage of reviewers could not answer the question.

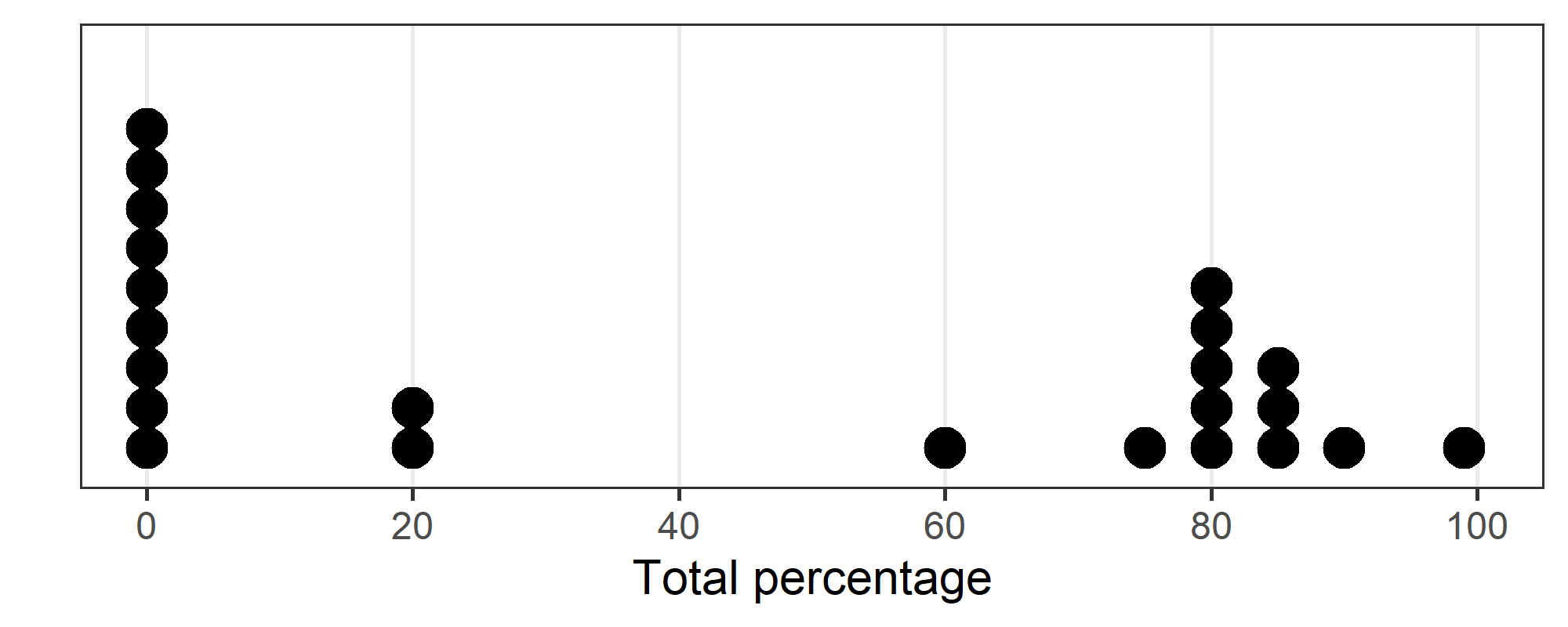
6 reviewers clicked the button, but still gave reasonable percentages. We therefore only exclude reviewers from the analyses below if only gave only zero percentages or clicked the radio button and gave no responses.

### Number where percentages do not add to 100%

Here we examine if the uncertainty percentages given by reviewers added to the expected 100%. If reviewers gave an answer under or over 100% it may be a simple arithmetic error or a misunderstanding of the question.

| **Adds to 100%** | **n** | **percent** |
| --- | --- | --- |
| No | 23 | 6 |
| Yes | 375 | 94 |

The plot below shows the percentage totals for the responses that did not add to 100%.



We exclude the reviewers whose total percentages were 0%, assuming this meant they could not answer the question. For those whose answers were above zero but not 100%, we redistributed the percentages. For example, if a reviewer answered 80% Approved, 10% Approved with Reservations, and 0% Not Approved (which adds to 90%), we changed this to: 89%, 11%, 0%.

The results in the following sections exclude the 9 reviewers who gave zero responses.

### Number and percent with no uncertainty

The table below shows the number of times reviewers reported no uncertainty. This is regardless of their recommendation, so includes 100% certainty of approve or reject.

| **Uncertainty** | **n** | **percent** |
| --- | --- | --- |
| Some | 301 | 77 |
| None | 88 | 23 |

The percentage with some uncertainty is 77% with a 95% confidence interval from 73% to 81%. This confidence interval was made using a logit parameterisation, using the normal approximation gave almost the same intervals.

The percentage with no uncertainty was 23% with a 95% confidence interval from 19% to 27%.

##### No uncertainty by decision and journal

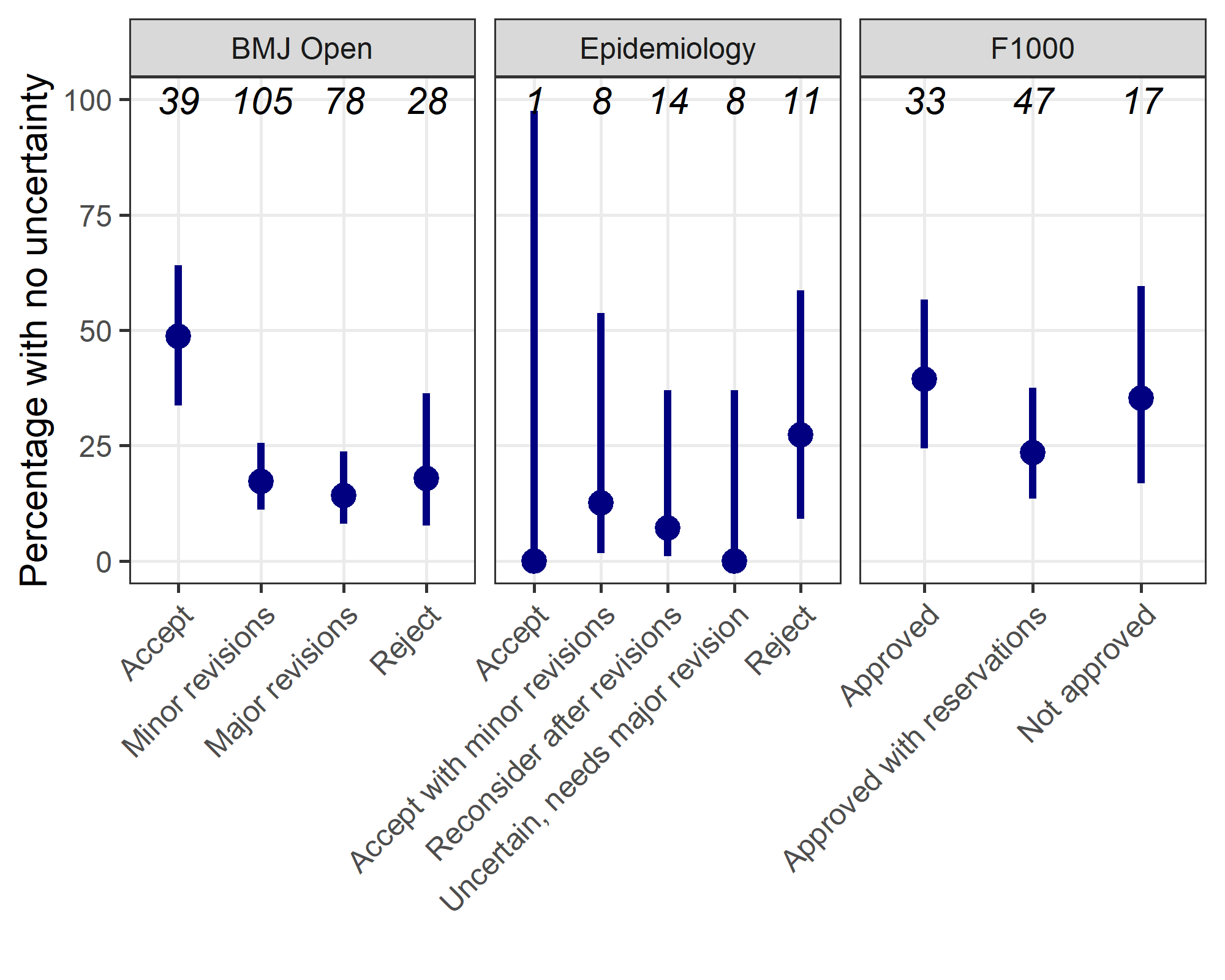
| **journal** | **Decision** | **r / n (%)** | **CI** |
| --- | --- | --- | --- |
| BMJ Open | Accept | 19/39 (49) | 34 to 64 |
| Minor revisions | 18/105 (17) | 11 to 26 |
| Major revisions | 11/78 (14) | 8 to 24 |
| Reject | 5/28 (18) | 8 to 36 |
| Epidemiology | Accept | 0/1 (0) | 0 to 98 |
| Accept with minor revisions | 1/8 (12) | 2 to 54 |
| Reconsider after revisions | 1/14 (7) | 1 to 37 |
| Uncertain, needs major revision | 0/8 (0) | 0 to 37 |
| Reject | 3/11 (27) | 9 to 59 |
| F1000 | Approved | 13/33 (39) | 24 to 57 |
| Approved with reservations | 11/47 (23) | 13 to 38 |
| Not approved | 6/17 (35) | 17 to 60 |

The table shows the number with no uncertainty (r), the denominator (n), the percent, and a 95% confidence interval for the percent. We excluded the responses with missing decisions.

There was generally more certainty (higher percentage) for the accept/reject decisions, and more uncertainty for the middle-ground decisions.

##### Plot

The plot shows the results presented in the table above, with the mean and 95% confidence interval for the percentage with no uncertainty.



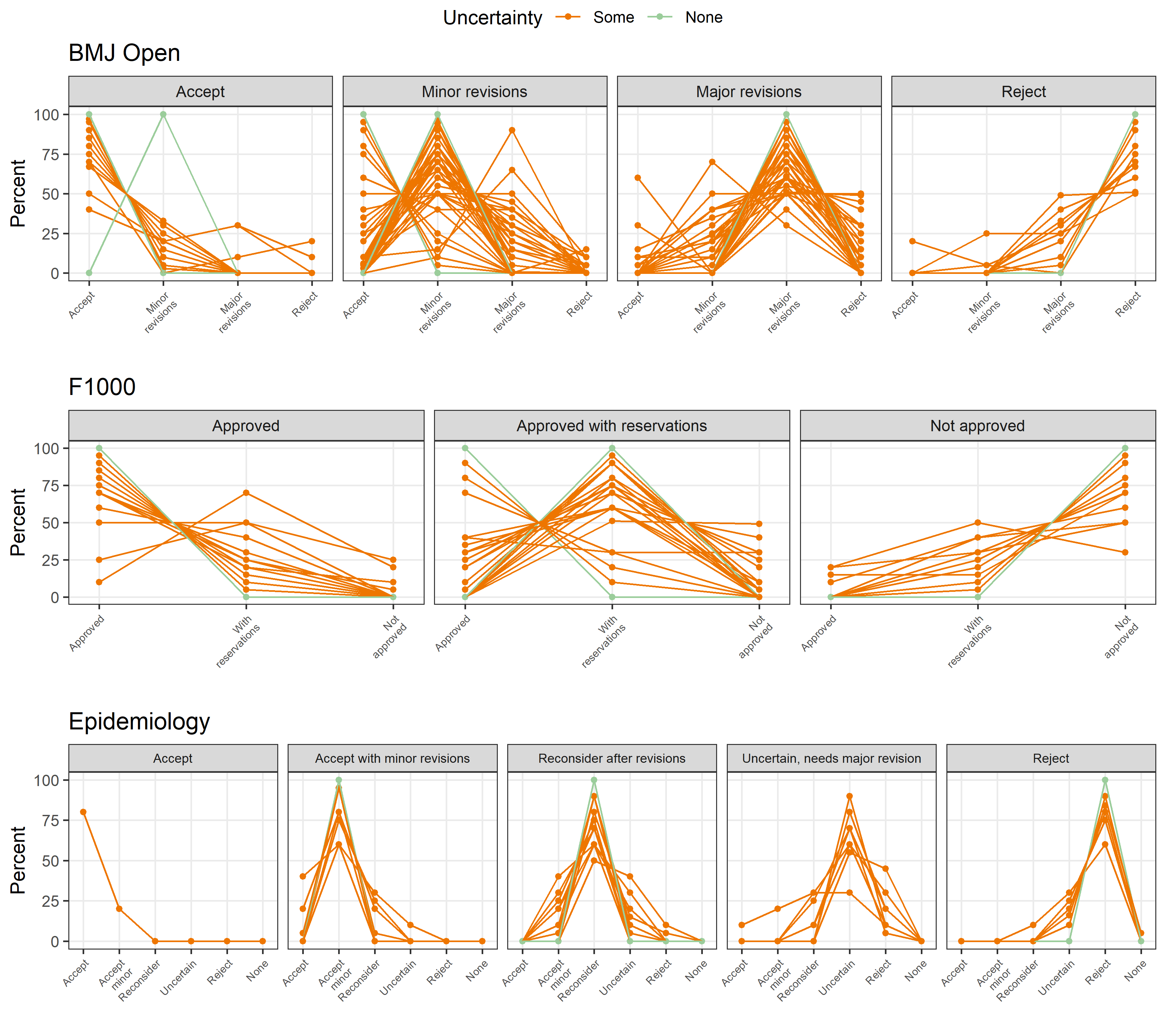
##### “No recommendation” at Epidemiology

Epidemiology had an option of “No recommendation” as a recommendation. Here we check if this was ever selected in the uncertainty results. The table below shows the summary statistics for the “No recommendation” category.

| **n** | **Not zero for None** | **Highest percent for None** |
| --- | --- | --- |
| 42 | 1 | 5 |

There was only 1 response that included “None” and this allocated 5%. For the statistical models that assumed an ordinal response, this single response was ignored and the respondents remaining 95% were scaled to add to 100%.

## Uncertainty visualisation



The plot shows the percentages given to each recommendation by the reviewers. There is a separate a panel for each recommendation and journal. The recommendation categories are different for each journal.

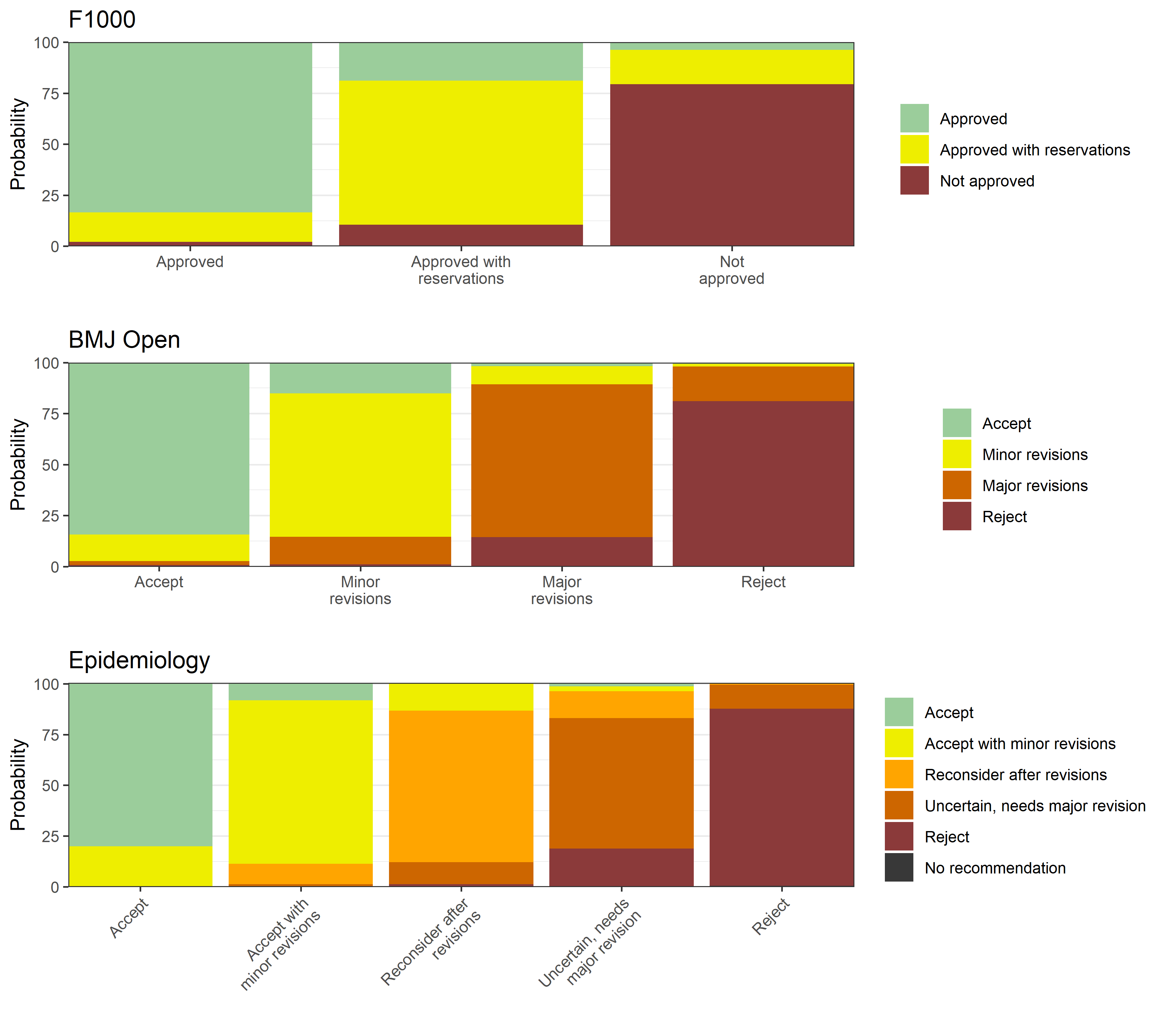
The lines in the plot are coloured by whether there was any uncertainty or not.

Four of the responses in the “Approved with reservations” panel for F1000 have a modal probability of “Approved”. It is possible that these respondents made an error when answering our question about their final recommendation.

The plot shows that any uncertainty tends to be in the neighbouring categories.

## Uncertainty visualisation using averages

The plot below shows the average probabilities by journal and recommendation. The x-axis is the reviewers’ decision, and the bars are coloured according to their probabilities. More colours in a bar indicates more uncertainty.



# Results by reviewer characteristics

## Results by gender

Here we plot the average probabilities by journal and gender.

