Summary statistics on the baseline tables for the trials published on Pubmed Central

Adrian Barnett

10 June 2022

# Exclusions

## Exclusions prior to reading data

From the 10,000 original papers.

| **reason** | **n** | **percent** |
| --- | --- | --- |
| No baseline table | 2,826 | 37.4% |
| Full text page not available | 2,123 | 28.1% |
| Not a randomised trial | 1,549 | 20.5% |
| No sample size | 405 | 5.4% |
| Just one column in table | 306 | 4.1% |
| Follow-up results in baseline table | 162 | 2.1% |
| Just one sample size | 74 | 1.0% |
| Single-arm study | 43 | 0.6% |
| Pre-post comparison | 42 | 0.6% |
| Difficult layout | 14 | 0.2% |
| Could not detect statistics | 9 | 0.1% |
| Total | 7,553 | 100.0% |

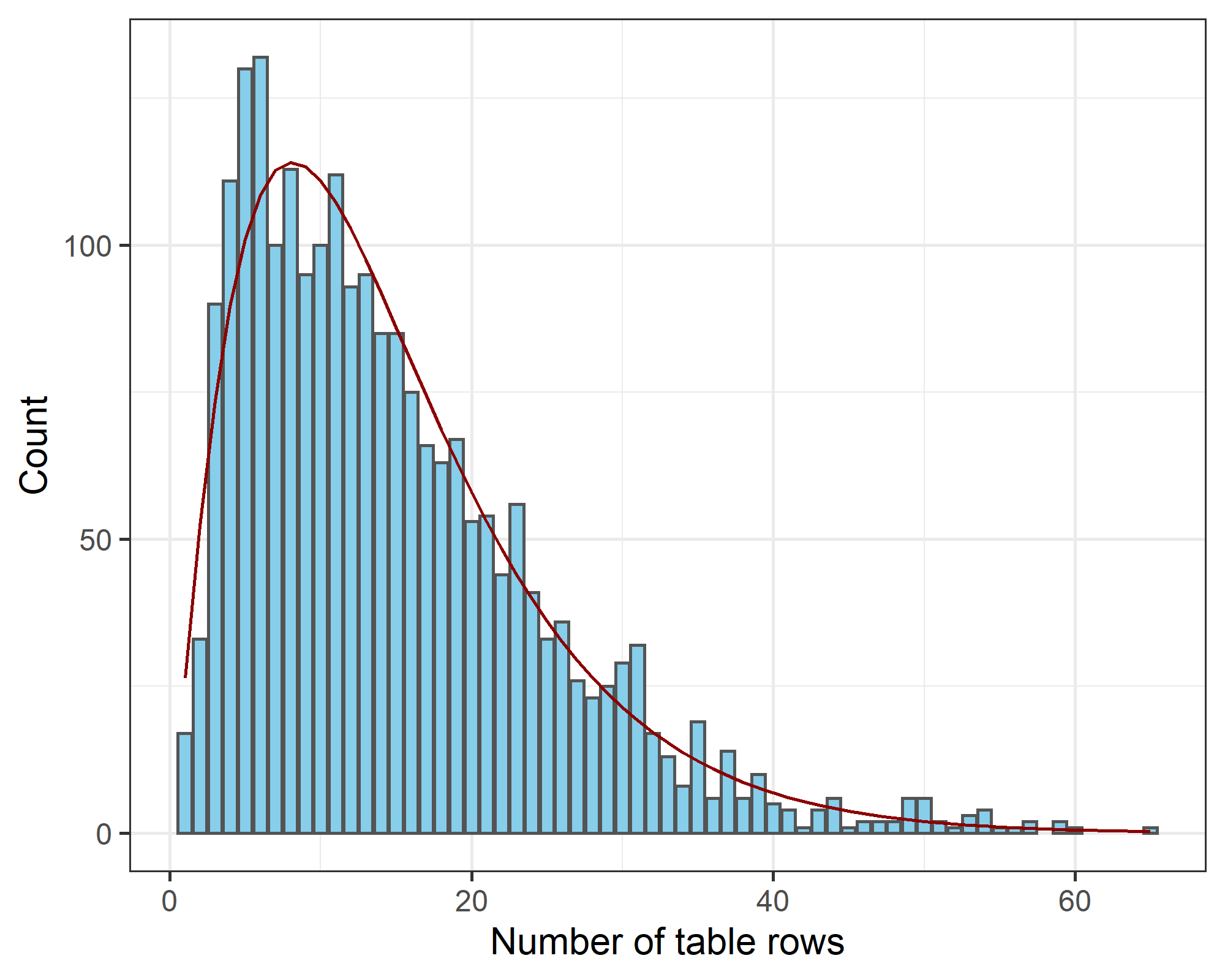
## Exclusions from data that could be read in

| **reason** | **count** |
| --- | --- |
| No second statistic for continuous/median (results) | 3,108 |
| Not a percent | 1,551 |
| Numbers that are not integers | 637 |
| Guessed percent but n > sample size (rows) | 570 |
| Only one result per row | 507 |
| Algorithm error | 161 |
| Trials with sample size under three | 27 |
| Trials with eight or more columns | 20 |
| Trials with non-integer sample size | 14 |
| Upper CI interval less than lower (results) | 10 |
| Rows with negative standard deviation | 3 |

The table shows the number of excluded rows of data, apart from the “Algorithm error” row which is the total number of excluded studies.

The total number of included studies was 2264 with 72,090 rows.

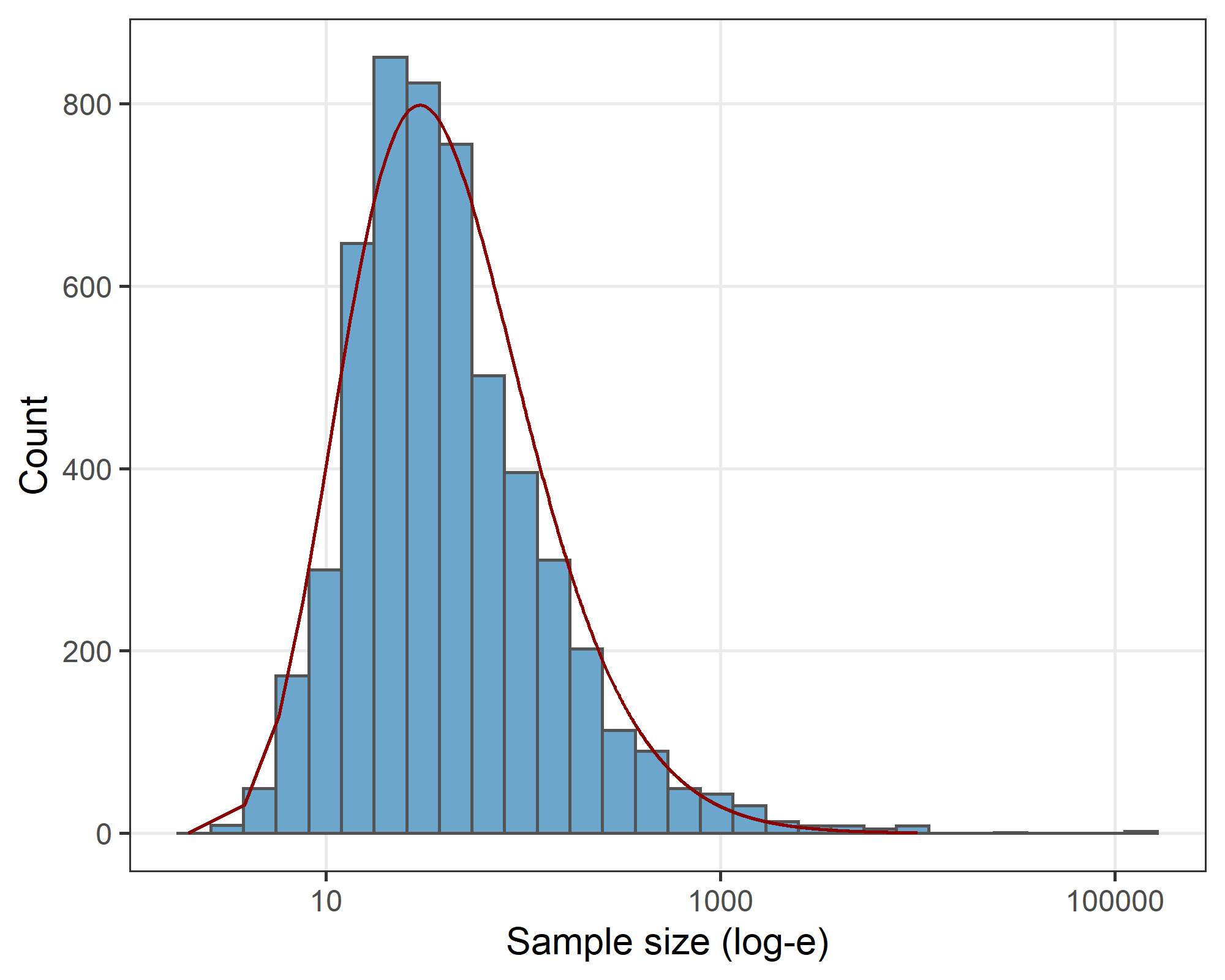
# Number of summary statistics (rows) per table



The mean number of rows per table was 15. The red line shows a gamma distribution with a shape parameter of 2.2 and rate of 0.15.

# Sample sizes

### Histogram of sample size per group



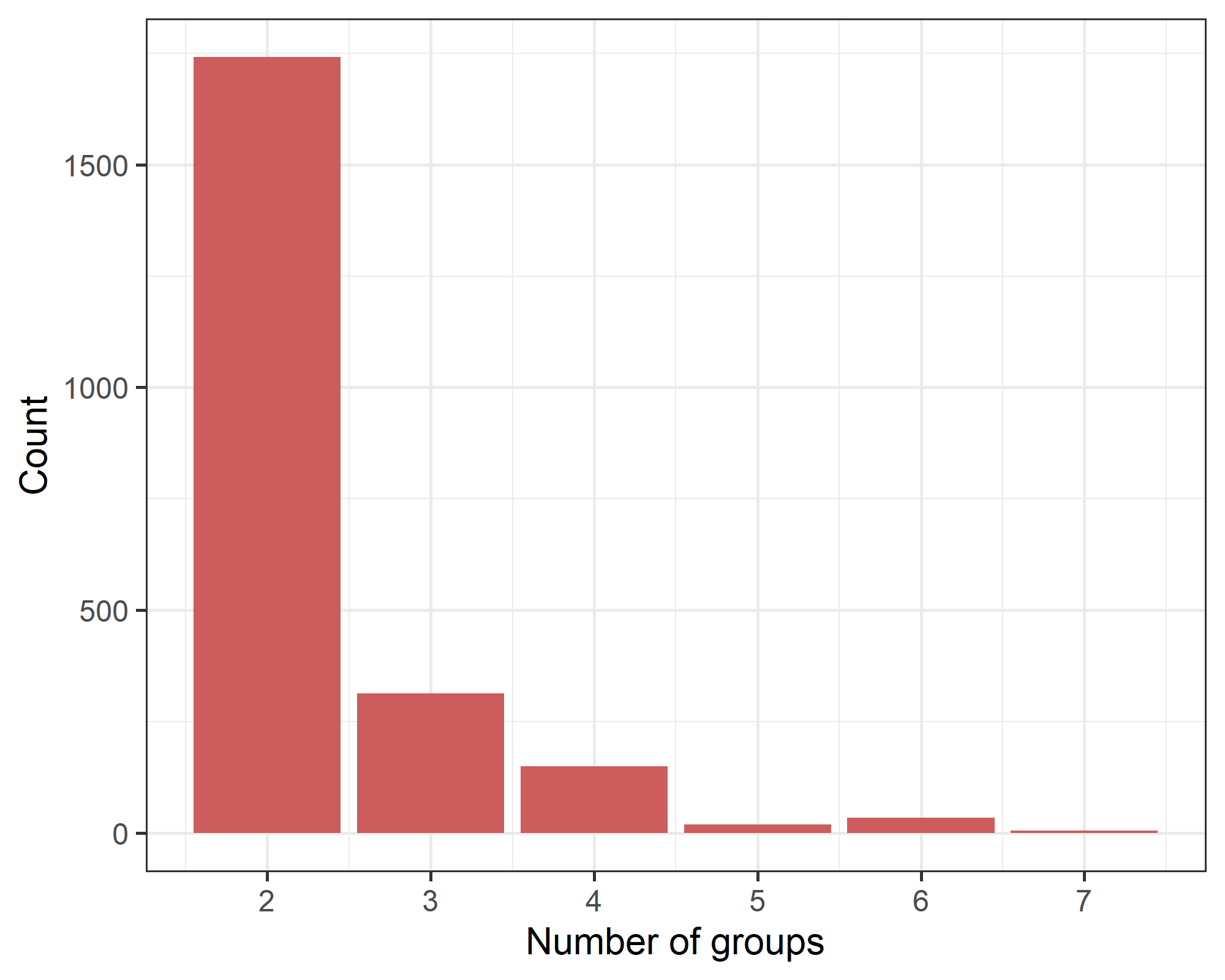
The gamma distribution (red line) fitted to the log-transformed (base e) sample size has a shape parameter of 11.2 and rate of 3.00.

### Summary statistics for sample size per group

| **percentile** | **sample** |
| --- | --- |
| 0.05 | 9 |
| 0.25 | 20 |
| 0.50 | 34 |
| 0.75 | 75 |
| 0.95 | 349 |

The table shows the sample size for five percentiles. This is for the sample size per group.

# Number of groups (columns) per table



# Frequency table of statistics

| **Included** | **Statistic** | **n** | **percent** |
| --- | --- | --- | --- |
| Yes | Percent | 37,229 | 52 |
| Continuous | 27,539 | 38 |
| Numbers | 2,816 | 4 |
| Confidence interval | 38 | 0 |
| No | Median | 4,212 | 6 |
| P-values | 206 | 0 |
| Range | 50 | 0 |
|  | Total | 72,090 | 100 |

# About the papers

## Dates

Q1 median Q3  
1 2016-06-15 2019-05-17 2020-06-15

## Top ten journals

| **journal** | **n** |
| --- | --- |
| Evidence-based Complementary and Alternative Medicine : eCAM | 85 |
| Journal of Physical Therapy Science | 84 |
| Journal of Clinical Medicine | 44 |
| PLoS ONE | 32 |
| Experimental and Therapeutic Medicine | 26 |
| Anesthesiology and Pain Medicine | 25 |
| Journal of Pain Research | 25 |
| Nutrients | 25 |
| Scientific Reports | 24 |
| Frontiers in Psychology | 23 |

In total there were 843 journals.

## Top ten countries

| **affiliation** | **n** |
| --- | --- |
| USA | 329 |
| China | 287 |
| Iran | 240 |
| Japan | 140 |
| South Korea | 137 |
| India | 99 |
| UK | 95 |
| Germany | 71 |
| Republic of Korea | 70 |
| Australia | 67 |

In total there were 85 countries. The country was based on the affiliation of the first author.

## Study design features

| **Study design** | **N** | **Yes** | **percent** |
| --- | --- | --- | --- |
| rct | 2,336 | 593 | 25 |
| block\_randomisation | 2,336 | 345 | 15 |
| pilot | 2,336 | 140 | 6 |
| cluster | 2,336 | 58 | 2 |
| sem | 2,336 | 40 | 2 |
| adaptive\_randomisation | 2,336 | 1 | 0 |

The variables are based on mentions in the text as follows:

* rct = RCT or versions of “randomised controlled trial” in the title or abstract
* block\_randomisation = “block” and versions of “randomisation” within 5 words of each other anywhere in paper
* adaptive randomisation = “adaptive randomisation” anywhere in paper
* cluster = “cluster\*” in title or abstract
* pilot = “pilot” in title
* sem = “SE”, “SEM” or “standard error” in baseline table or footnote

# P-values

Here we examine the number of trials that included p-values in the table or text.

| **In table** | **In text** | **n (%)** |
| --- | --- | --- |
| No | No | 769 (27) |
| Yes | 219 (8) |
| Missing | 234 (8) |
| Yes | No | 692 (25) |
| Yes | 744 (27) |
| Missing | 148 (5) |

The results are missing for the table check where there was no text that refered to the table.

The overall percentage of trials where p-values were mentioned in the table only was 34% with a 95% CI from 33% to 36%.

The overall percentage of trials where p-values were mentioned in the text or table was 64% with a 95% CI from 62% to 66%.

### Excluding cluster-randomised trials

We repeat the above table after excluding cluster-randomised trials.

| **In table** | **In text** | **n (%)** |
| --- | --- | --- |
| No | No | 610 (27) |
| Yes | 178 (8) |
| Missing | 173 (8) |
| Yes | No | 579 (25) |
| Yes | 613 (27) |
| Missing | 125 (5) |

Excluding the 58 cluster-randomised trials made little differences to the percentages.

# Decimal places

Below we examine the number of decimal places for continuous variables. When researchers use few decimal places then the comparison between groups can appear under-dispersed. The results are stratified based on the size of the mean as larger means required fewer decimal places. The cells show the decimal place quantiles at 10, 25, 50 (median), 75 and 90.

| **Range for mean** | **n** | **q10** | **q25** | **q50** | **q75** | **q90** |
| --- | --- | --- | --- | --- | --- | --- |
| [0,10) | 7,782 | 0 | 1 | 1 | 2 | 2 |
| [10,100) | 15,968 | 0 | 1 | 1 | 2 | 2 |
| [100,1e+03) | 3,377 | 0 | 0 | 1 | 2 | 2 |
| [1e+03,Inf) | 412 | 0 | 0 | 0 | 1 | 2 |