Investigating the rise of meta-analysis articles

by Chase An, Karanpreet Benipal, Derek Ha, Remi Jiang

December 5, 2023

Executive summary:

It has been claimed that the number of articles publishing meta-analysis result based on human participants is increasing over recent years, and our team collected several samples from Journals Statistics in Medicine, Statistical Methods in Medical Research, and Journal of Clinical Epidemiology to investigate this claim. During the investigation, we applied stratified sampling using journals as strata and made use of both proportional and neyman allocation using the absolute error approach to determine our sample sizes. After randomly choosing our samples, using simple random sampling without replacement, we calculated their estimated totals, proportions with their variances, standard errors, and confidence intervals. Based on these results, we found that the number of these articles have been indeed increasing over recent years, but their proportion has been steady.

Objectives:

Our goal is to determine whether the number of peer-reviewed articles publishing meta-analysis results has increased. In the analysis, data from three specific journals have been collected, to assess any trends in the publication of meta-analysis articles over time. To provide insights into the prevalence of meta-analysis publications in these journals, a statistical analysis was conducted using proportional and optimal (Neyman) allocation methods. Finally, we will consider the limitations of the methods used, such as the sample size used and any potential sampling bias, to ensure a comprehensive interpretation of the results.

Statistical Methods:

Our target population is all peer-reviewed articles written in English from 1992 to 2023. We sampled some articles from the three journals mentioned in the Executive Summary during the same period. Our sampling units are randomly selected articles from the list of all articles published on these journals during 1992 and 2023. Our observation unit is the articles published each individual year. We defined our response variables to be the proportion and total of articles relating to meta-analysis using human participants, and use

year as our explanatory variable. Our sampling frames were created by downloading all the citations of articles of three journals during 1992 and 2023 from PUBMED, and then we added filter/indicator variable "Meta-analysis" to the citations we downloaded.

As for our sampling method, we created three strata, namely Statistics in Medicine strata(first stratum), Statistical Methods in Medical Research(second stratum), and Journal of Clinical Epidemiology(third stratum) with strata sizes $N_1 = 434$, $N_2 = 229$, and $N_3 = 1439$ respectively. Note $N = N_1 + N_2 + N_3 = 2102$. The variances of strata are $\sigma_1^2 = 0.1402$, $\sigma_2^2 = 0.0801$, and $\sigma_3^2 = 0.0495$ respectively. The corresponding strata standard deviations are $\sigma_1 = 0.3745$, $\sigma_2 = 0.2829$, $\sigma_3 = 0.2225$, and the strata totals are $T_1 = 73$, $T_2 = 20$, $T_3 = 79$ respectively.(calculated using R)

Since we use simple random sampling without replacement to choose strata samples , we calculated the sample size using absolute tolerable error approach by choosing the margin of error, e, to be 5%, and significance level, α, to be 5% as well. We choose our precision, σ_y, to be 0.5 using the maximum variance of $\sigma_y^2 \leq \frac{1}{4}$ assuming P = $\frac{1}{2}$ under our sampling design. Then we calculated our sample size as below

$$n_0 = z_{\frac{\alpha}{2}}^2 \cdot \frac{\sigma_y^2}{e^2} = 1.96^2 \cdot \frac{0.25}{0.05^2} \approx 384.16$$

. Now we are able to find the total sample size n by

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{384.16}{1 + \frac{384.16}{2102}} \approx 324.79$$

, which rounds up to 325. Finally, we could calculate the strata sample sizes as:

-Under proportional allocation(using the formula $n_h = n \cdot \frac{N_h}{N}$):

$$n_1 = n \cdot \frac{N_1}{N} = 325 \times \frac{434}{2102} = 67.1 \approx 67$$

$$n_2 = n \cdot \frac{N_2}{N} = 325 \times \frac{229}{2102} = 35.4 \approx 35$$

$$n_3 = n \cdot \frac{N_3}{N} = 325 \times \frac{1439}{2102} = 222.5 \approx 223$$

-Under Neyman allocation(using the formula $n_h = n \cdot \frac{W_h\sigma_h}{\sum_{h=1}^{H} W_h\sigma_h}$, H = 3):

$$\sum_{h=1}^{H} W_h\sigma_h = \frac{434}{2102} \cdot 0.3745 + \frac{229}{2102} \cdot 0.2829 + \frac{1439}{2102} \cdot 0.2225 = 0.2605$$

$$n_1 = n \cdot \frac{W_1\sigma_1}{\sum_{h=1}^{H} W_h\sigma_h} = 325 \cdot \frac{\frac{434}{2102}0.3745}{0.2605} = 96.48 \approx 97$$

$$n_2 = n \cdot \frac{W_2\sigma_2}{\sum_{h=1}^{H} W_h\sigma_h} = 325 \cdot \frac{\frac{229}{2102}0.2829}{0.2605} = 38.45 \approx 38$$

$$n_3 = n \cdot \frac{W_3\sigma_3}{\sum_{h=1}^{H} W_h\sigma_h} = 325 \cdot \frac{\frac{1439}{2102}0.2225}{0.2605} = 190.04 \approx 190$$

We then randomly chose samples of corresponding size from each stratum using a randomizer in Google Sheets. After picking the samples, we calculated the summaries of the samples under proportional and Neyman allocation respectively.

-Under proportional allocation, the sample data is summarized as:

Table 1: Proportional allocated sample summaries

Stratum	**Mean/Proportion**	**Sample Error**	**Sample Total**
1	0.1642	0.1393	11
2	0.0286	0.0286	1
3	0.0718	0.0669	16

-Under Neyman allocation, the sample data is summarized as:

Table 2: Neyman allocated sample summaries

Stratum	**Mean/Proportion**	**Sample Error**	**Sample Total**
1	0.2165	0.1714	21
2	0.1053	0.0967	4
3	0.0579	0.0548	11

Note confidence intervals can be calculated by $estimate \pm z_{\frac{\alpha}{2}} \cdot se(estimate)$. Using the data summaries above, we are able to find the estimated totals and proportions under proportional and Neyman allocation respectively.

Under stratified random sampling, we used the sample total and proportion formulas:

$$y_{st} = N \cdot \sum_{h=1}^{H} W_h \bar{y_h}$$

$$p_{st} = \sum_{h=1}^{H} W_h p_h$$

where $p_h = \frac{m_h}{n_h}$, where m_h is the sample total of sample h.

-Under proportional allocation, the estimated proportion is calculated as:

$$p_{st} = \sum_{h=1}^{H} W_h p_h = \frac{434}{2102} \times 0.16418 + \frac{229}{2102} \times 0.02857 + \frac{1439}{2102} \times 0.07175 = 0.08613$$

$$var_{prop}(p_{st}) = (1 - \frac{n}{N}) \cdot \frac{1}{n} \sum_{h=1}^{H} W_h \sigma_h^2 = 0.00020$$

the estimated total is calculated as:

$$y_{st} = N \cdot \sum_{h=1}^{H} W_h \bar{y_h} = 2102 \cdot ((\frac{434}{2102} \times 0.16418) + (\frac{229}{2102} \times 0.02857) + (\frac{1439}{2102} \times 0.07175)) = 181.0432$$

$$var_{prop}(y_{st}) = N^2 var_{prop}(p_{st}) = 2102^2 \times 0.000202 = 892.716$$

the confidence interval for y_{st} is $[122.4827, 239.6037]$ and that for p_{st} is $[0.05827, 0.11399]$.

-Under Neyman allocation, the estimated proportion is calculated as:

$$p_{st} = \sum_{h=1}^{H} W_h p_h = \frac{434}{2102} \times 0.21649 + \frac{229}{2102} \times 0.10526 + \frac{1439}{2102} \times 0.05789 = 0.09580$$

$$var_{min}(p_{st}) = \frac{1}{n} \cdot (\sum_{h=1}^{H} W_h \sigma_h)^2 - \frac{1}{N} \sum_{h=1}^{H} W_h \sigma_h^2 = 0.00017$$

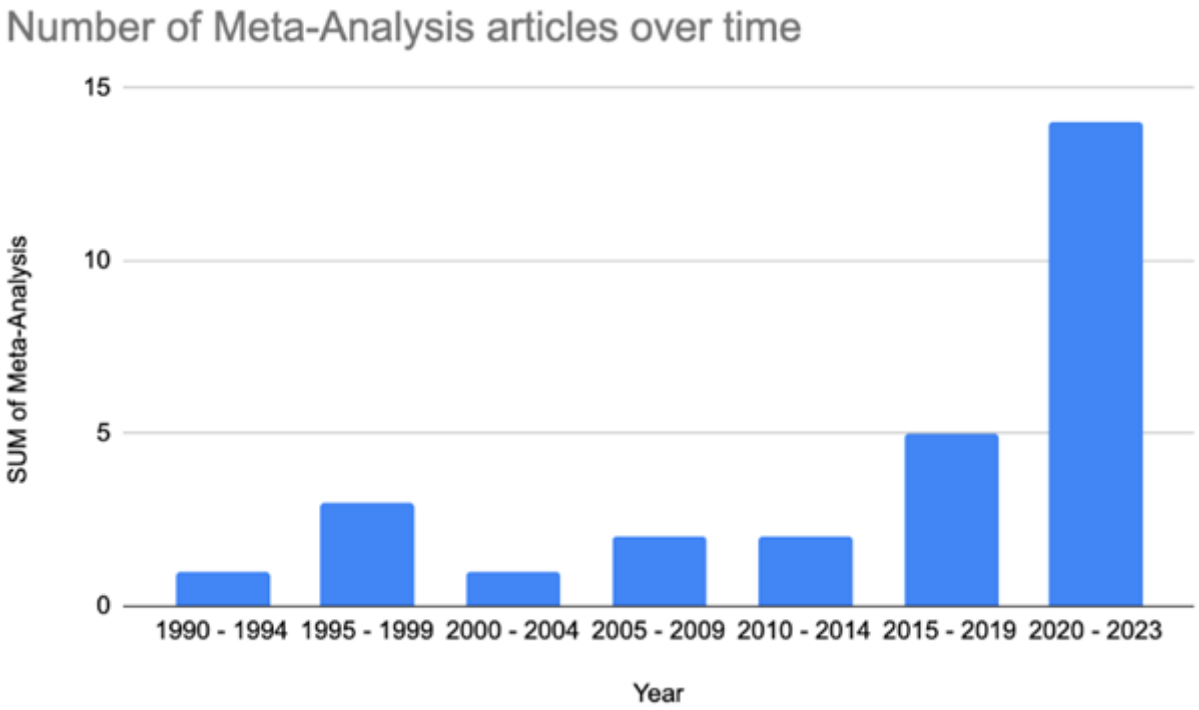
the estimated total is calculated as:

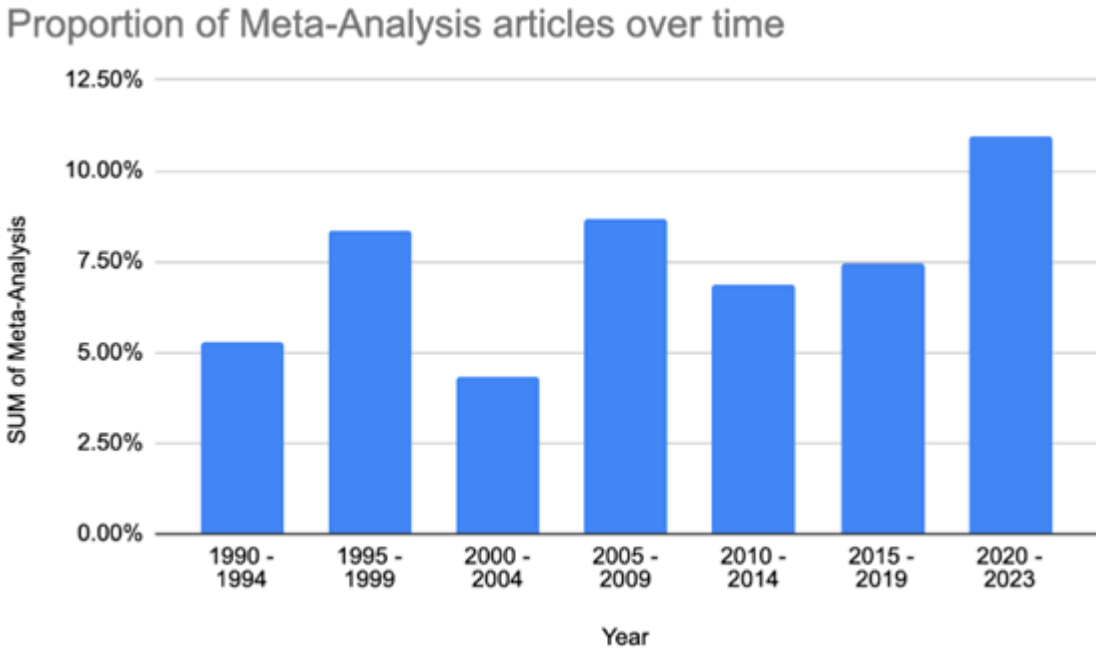
$$y_{st} = N \cdot \sum_{h=1}^{H} W_h \bar{y_h} = 2102 \cdot ((\frac{434}{2102} \times 0.21649) + (\frac{229}{2102} \times 0.10526) + (\frac{1439}{2102} \times 0.05789)) = 201.37$$

$$var_{min}(y_{st}) = N^2 var_{min}(p_{st}) = 2102^2 \times 0.00017 = 771.720$$

the confidence interval for y_{st} is $[146.9270, 255.8221]$ and that for p_{st} is $[0.06989, 0.12170]$.

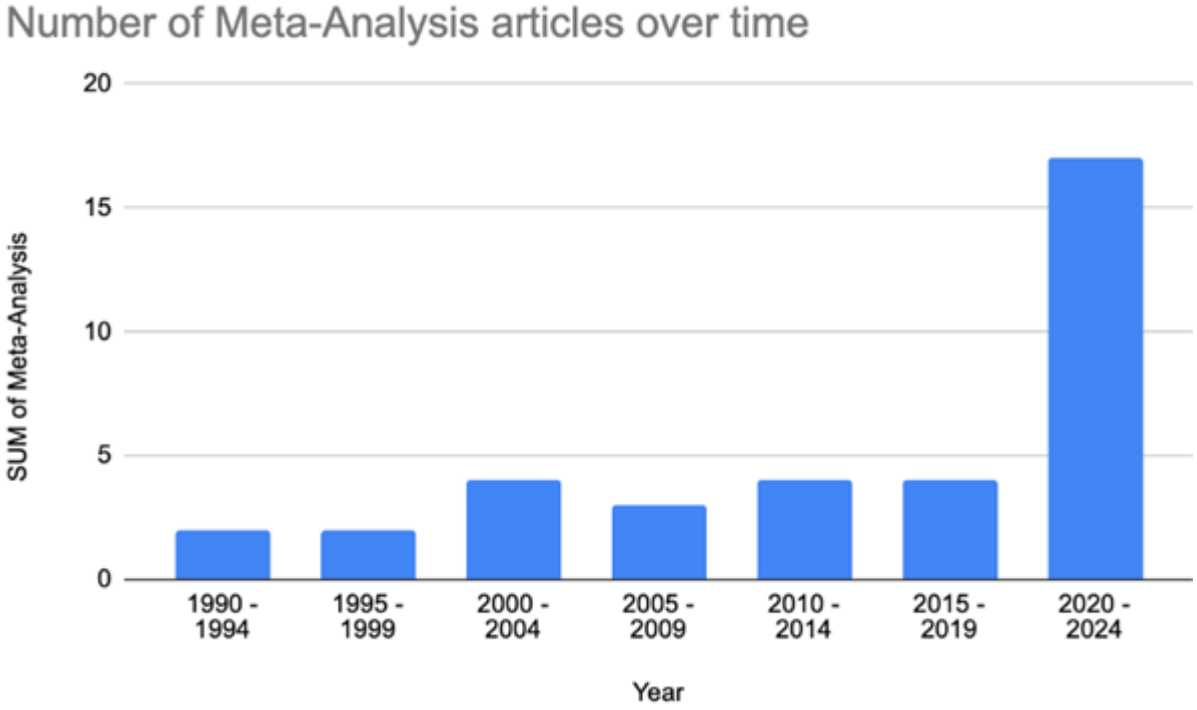
Results:

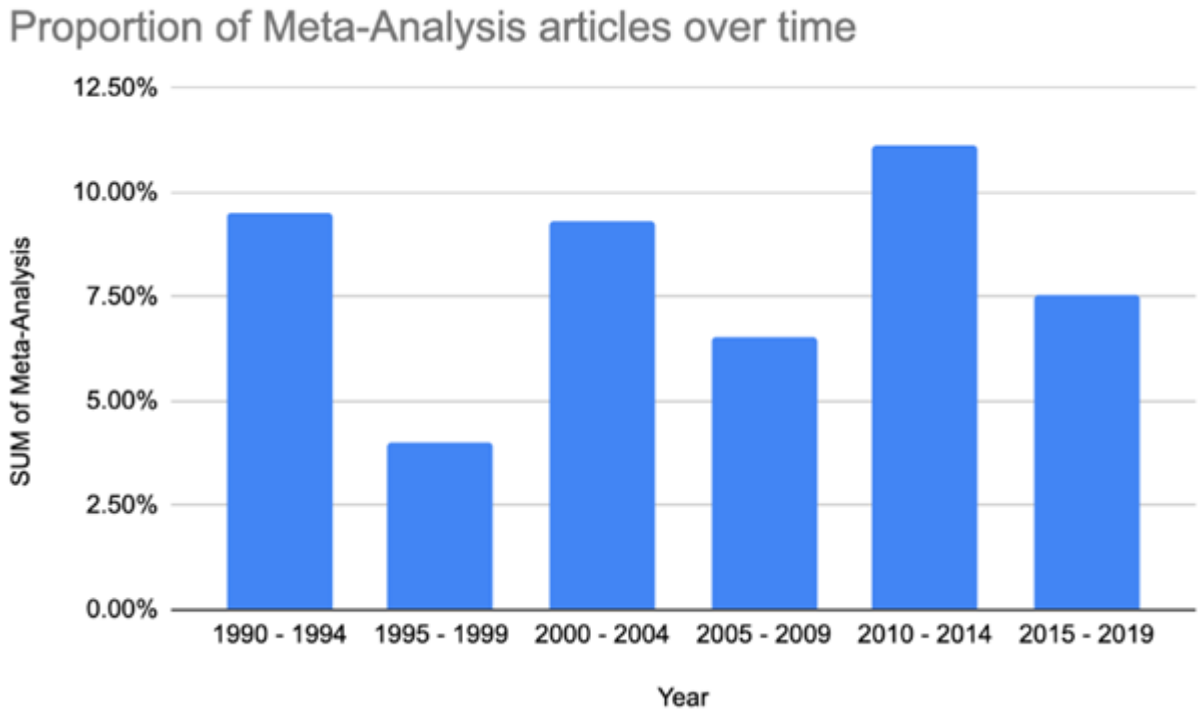




Under Proportional Allocation, we can see there is an increasing trend of the total number of meta-analysis articles over years. The overall trend exhibits a gradual increase from 1990 to 2019, followed by a significant surge after 2020.

But if we look at the second chart which focuses on the trend of population proportion, the proportion fluctuates from 4% to 11%. It does not show a noticeable positive trend.





Under Neyman Allocation, there is an increasing trend in the number of meta-analyses, mirroring the proportional allocation approach. Due to the fluctuations in proportional allocation, it is inconclusive to assert a clear increasing trend in the proportion of

meta-analyses.

In summary, there is a discernible increasing trend in the number of meta-analysis articles. However, there is insufficient evidence to support the claim that the proportion of meta-analyses has seen a similar increase.

Discussions:

Our analysis suggests a possible increase in the number of peer-reviewed articles presenting meta-analysis results. However, there are some limitations we noticed through the sampling process.

-Firstly, the relatively small sample size we used might impact the credibility of our results. We noticed that we have more data for recent years, however we do not have enough data for earlier years which may explain the sudden spike in population totals. To ensure more reliable outcomes in future studies, we recommend incorporating a larger sample size to capture more representative data.

-Secondly, our focus on articles from three specific journals could introduce bias, possibly excluding relevant publications from other sources. To address this, future research should adopt a broader sampling strategy that encompasses a diverse range of journals. We sampled from PubMed and not directly from the journal databases, so there may be some bias in terms of what types of articles PudMed includes.

-Additionally, our streamlined search process, involving the exclusion of articles without the keyword "meta-analysis" in the title, might have overlooked relevant studies. While this approach expedited the identification process, it's worth exploring alternative methods for a more comprehensive identification of meta-analysis articles.

-Furthermore, limiting our analysis to English-language articles may restrict the generalizability of our findings. Including articles in other languages could offer a more accurate portrayal of overall trends in meta-analysis publications.

Bibliography:

The rest of the pdf contains our citations.