Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1335977

Open Access Journal
e-ISSN: 2619 – 8991

Research Article
Volume 6 - Issue 4: 414-420 / October 2023

BIBLIOMETRIC ANALYSIS OF THE STUDIES DETERMINED BY THE MONTE CARLO SIMULATION TECHNIQUE OF THE POWER OF THE TEST

Malik ERGIN[1]*, Rabia ALBAYRAK DELIALIOĞLU[2], Yasin ALTAY[3], Özgür KOŞKAN[1]

[1]*Isparta University of Applied Sciences, Faculty of Agriculture, Department of Animal Science, 32000, Isparta, Türkiye*
[2]*Ankara University, Faculty of Agriculture, Department of Animal Science, 06000, Ankara, Türkiye*
[3]*Eskisehir Osmangazi University, Faculty of Agriculture, Department of Animal Science, 26000, Eskişehir, Türkiye*

Abstract: In order to give detailed information about the characteristics of scientific studies related to the power of the test, a bibliometric analysis of studies in which the power of the tests used in statistical data analysis was determined with the aid of Monte Carlo simulation techniques was carried out. The Web of Science (WoS) database's 1309 scientific studies in which the power of the test was determined by using Monte Carlo simulation techniques served as the study's material of data. The analysis includes the number of studies by year, average citation count by year, the number of articles published in scientific journals, countries of responsible authors, and the most relevant words in the studies. It was observed that Communications in Statistics-Simulation and Computation and Journal of Statistical Computation and Simulation are the journals with the highest number of published articles. The United States of America (USA) takes the lead when considering the countries of corresponding authors, while Türkiye is in 8th place. The most used keywords in the ten-year time period of scientific studies were respectively "power", "test", "skewness", "model" and "inference". As a result, it can be concluded that test power studies can be obtained safely using the Monte Carlo simulation technique.

Keywords: Type I error probability, Test power, Bibliometric analysis, Monte Carlo, Simulation

***Corresponding author:** Isparta University of Applied Sciences, Faculty of Agriculture, Department of Animal Science, 32000, Isparta, Türkiye
E mail: malikergin@isparta.edu.tr (Malik ERGIN)

Malik ERGIN https://orcid.org/0000-0003-1810-6754
Rabia ALBAYRAK DELIALIOĞLU https://orcid.org/0000-0002-1969-4319
Yasin ALTAY https://orcid.org/0000-0003-4049-8301
Özgür KOŞKAN https://orcid.org/0000-0002-5089-6250

Received: August 01, 2023
Accepted: September 05, 2023
Published: October 15, 2023

Cite as: Ergin M, Albayrak Delialioğlu R, Altay Y, Koşkan Ö. 2023. Bibliometric analysis of the studies determined by the monte carlo simulation technique of the power of the test. BSJ Eng Sci, 6(4): 414-420.

1. Introduction

Thanks to the progress in computer technology, statistical simulation techniques have also rapidly developed. With the rapid advancement of programming power in computer technology, the behaviors of various statistical methods can be evaluated using the Monte Carlo simulation method by generating random numbers from specific statistical distributions (Kroese et al., 2014; Morris et al., 2019). Depending on the purpose of the simulation study, various combinations can be designed when comparing means between two or more groups (Burton et al., 2006). These combinations include variance ratios between groups, sample sizes, distributions of populations from which observations are taken, and standard deviation differences between or among group means. These simulation combinations help in investigating the results shown by variance analysis techniques (ANOVA) or alternative methods more clearly. The advantage provided by the simulation technique is the ability to determine the parameters of populations in advance. When evaluating many statistical methods through simulation experiments, the literature mostly considers Type I error and power values of the test. In recent years, various statistical methods are still being examined using the Monte Carlo simulation technique (Patric, 2009; Koşkan ve Gürbüz, 2008; Ferreira et al., 2012; Lantz, 2013). In the field of medicine, through the Monte Carlo simulation technique, a newly developed mathematical technique can be evaluated by creating various simulation designs. In addition, when designing a study based on the assumptions required by a statistical method to be used in real-life data (such as variance analysis, t-test, etc.), comparing two or more methods, or calculating values such as sample size or test power can be used (Morris et al., 2019).

When evaluating the performance of any test, the criterion we focus on can be the test power. The test power or empirical power is the probability of being able to detect a difference when there is a difference between group means. If there is a 0.75 standard deviation difference between or among group means, the ratio at which the null hypothesis (H_0) is rejected at a significance level of 5.0% after 100000 simulation trials is expressed as the test power. There is a generally accepted measure in the literature regarding the level of

power that is considered sufficient for test performance. Murphy and Myors (2014) stated that if the power of the test is 80% or higher, it can be considered adequate. There are also mathematical approaches developed to assess the test power. Studies have suggested that it is not appropriate to compare the test power when the probabilities of Type I errors are not equal (Zhang and Boos, 1994; Lloyd, 2005). While a test maintains a Type I error rate of 5%, it can have lower power values compared to other tests (Çavuş, 2020). In Monte Carlo simulation studies, the test power is investigated by many researchers. Mahapoonyanont et al. (2010) compared the power of the test using original and transformed data in analysis of variance. Öztuna et al. (2006) compared the power of Lilliefors corrected Kolmogorov-Smirnov, D'Agostino Pearson, Jarque-Bera, and Shapiro-Wilk tests, where normality assumption can be checked. In conclusion, it was reported that the Shapiro-Wilk test yielded better results in terms of power for small sample sizes. In addition, Başpınar and Gürbüz (2000) evaluated the power of variance analysis using randomly generated numbers from symmetric, right-skewed, and left-skewed distributions. The researchers emphasized that the distribution shapes of populations do not have significance in achieving the desired power.

Evaluating the results and some bibliometric data of previous studies in the literature regarding the subject at hand enables research to progress faster and in a more interconnected manner. Therefore, a bibliometric study helps the researcher assess the literature related to the research subject of interest by employing quantitative approaches on the bibliometric data (Büyükkıdık, 2022). Donthu et al. (2021) emphasized the use of bibliometric analysis in cases where the scope of evaluation is broad and when manually reviewing data from studies becomes inconvenient. The most common databases where bibliometric data can be generated include Web of Science (WOS), Scopus, Google Scholar, Microsoft Academic, and Dimensions. Thanks to libraries developed in open-source programming languages such as Python or R, such as Bibliometrix, BiblioTools, Citan, Metaknowledge, Scientopy, and scientoText, bibliometric analyses can now be easily conducted (Moral-Munoz et al., 2020).

The lack of sufficient knowledge about the progress made despite years of studying the power of any test using Monte Carlo simulation techniques is an important factor in determining the status of research conducted. In this study, a bibliometric analysis was conducted on studies examining the power of tests using Monte Carlo simulation technique in the Web of Science (WOS) database.

2. Materials and Methods

Bibliometric analysis is a technique supported by computer technology to review the literature on a specific area and to network these studies (Han et al. 2020). The simulation studies conducted between 1980 and 2023 were used in the literature review stage by utilizing the Web of Science database. Due to the various interpretations of the simulation technique in different disciplines, a literature search was conducted focusing on "test power" and "Monte Carlo simulation" keywords. After the WOS search, Monte Carlo Simulation and Power test words were review, and 1309 scientific studies were reached by selecting the research areas related to statistics. To construct the list of reviewed documents, each paper included in a single dataset and repeated papers were eliminated. The collected data set included the number of studies by year, average citation count by year, number of articles published in scientific journals, countries of corresponding authors, and commonly used keywords in the studies.

All analyses conducted in the following steps:

(1) Extracting data on statistical studies related to the test power and Monte Carlo simulation from the WOS database

(2) Uniting and reorganizing the dataset

(3) Determining a relationship among these studies reviewed in a certain year period and visualizing this relationship with different ways

(4) Interpretation of the results obtained.

Bibliometric analysis was conducted with a total of 1309 papers via Bibliometric package developed in the R programming language (Aria and Cuccurullo, 2017).

3. Results and Discussion

Based on the literature available in WOS, a total of 1309 papers constitute the body of simulation studies published between 1980 and 2023. Within these 1309 publications, there are subcategories such as research articles, book chapters, full-text conference papers, and reviews. The findings obtained from bibliometric analysis in this study were evaluated in terms of the number of studies by year, average citation count by year, number of articles published in scientific journals, countries of responsible authors, and the commonly used keywords.

As seen in Figure 1, according to the studies in the WOS database, the first years in which the Monte Carlo simulation technique was used date back to 1980. It is noteworthy that any study was not conducted until the early 1990s, and the number of studies gradually increased in the subsequent years. In 2008, there was a significant increase with 42 studies, and the highest number of studies occurred in 2017 with 87 studies. Although the number of studies did not decrease rapidly in the following years, it remained limited to 29 studies within the year 2023. As a result, it is clearly observed that simulation studies still have a widespread presence and are of interest to researchers nowadays.

The average citation counts of simulation studies in the WOS database are presented by year in Figure 2. The highest average citation count per year is 4.76 in 2007. If we observe the number of studies by year shown in Figure 1, the rapid increase in 2007 can be considered as

the reason for the increase in citation count. However, the average citation counts significantly decreased in the subsequent years.

The knowledge of which journals published studies conducted using the Monte Carlo simulation technique can significantly impact researchers' future work plans. Figure 3 displays the number of articles published in various journals for simulation studies conducted within the specified years in the WOS database. A total of 357 scientific journals in the WOS database have published simulation studies. Among them, the journal named as "Communications in Statistics-Simulation and Computation" takes the lead with 140 articles. Following closely is the "Journal of Statistical Computation and Simulation" with 111 articles.

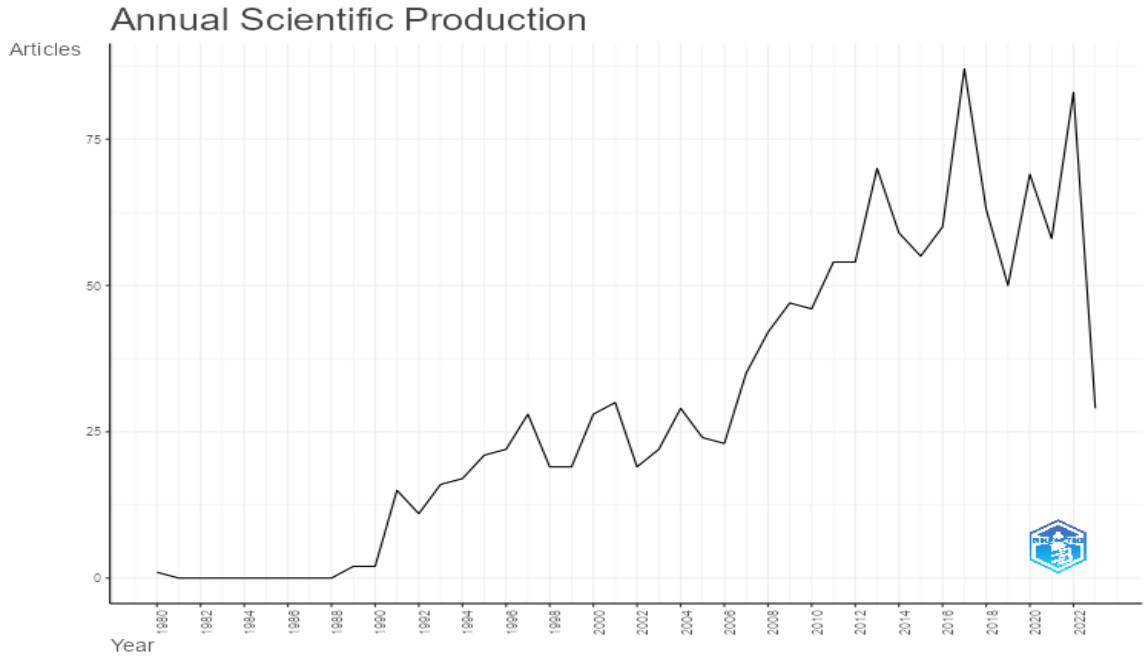

Figure 1. Annual Scientific Production (between 1980 – 2023).

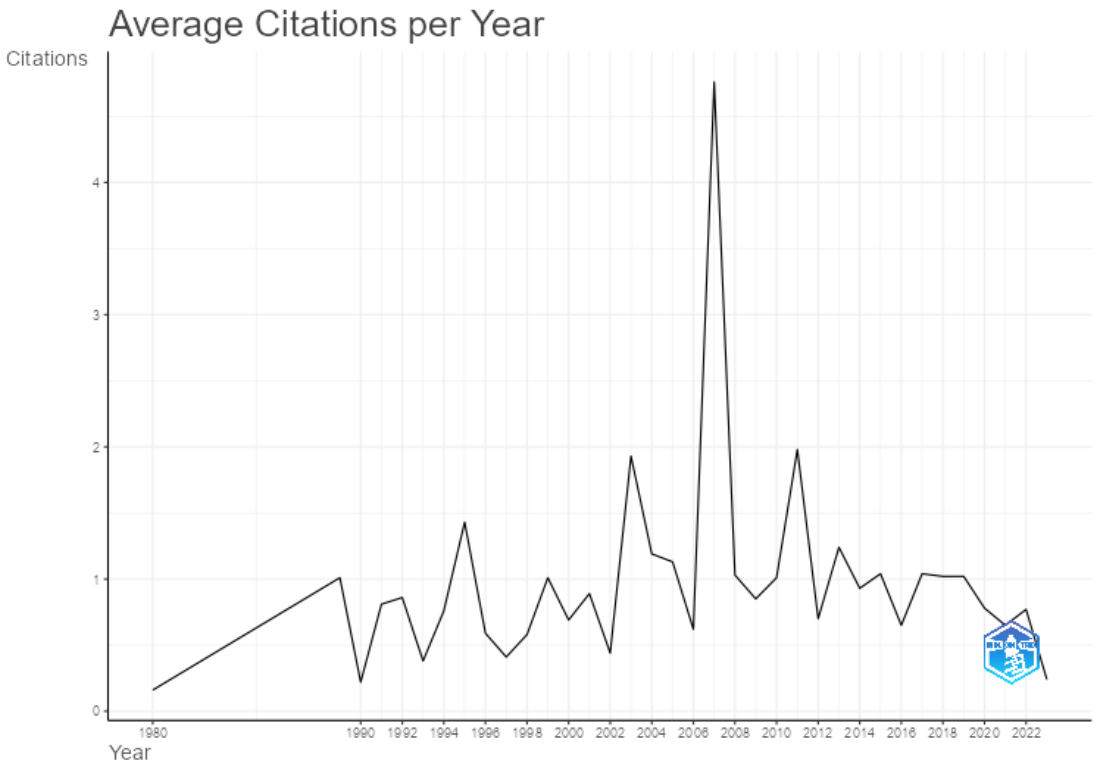

Figure 2. Average citations per year.

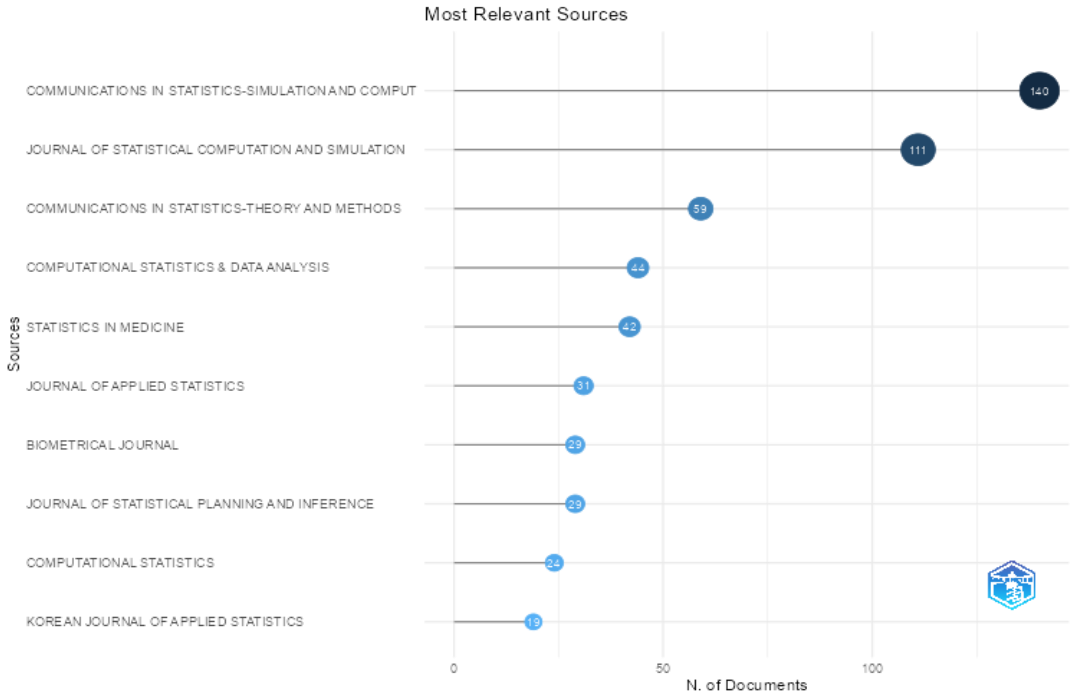

Figure 3. Most relevant journals.

The countries of the corresponding authors for the papers are shown in Figure 4. The parts highlighted with Single Country Publications (SCP) and Multiple Country Publications (MCP) in the graph respectively represent the number of publications made by researchers from the same country and the number of publications made by researchers from multiple countries. Considering the countries of the corresponding authors, the United States of America (USA) ranks first with 332 publications (SCP: 276, MCP: 56), followed by China with 170 publications (SCP: 129, MCP: 41). Türkiye ranks 8th among 66 countries with 50 publications (SCP: 48, MCP: 2).

The most relevant words within the scope of 1309 studies included in the data are presented in Figure 5. Accordingly, the words tests (f = 116), power (f = 86), statistics (f = 82), models (f = 74) are widely used.

In Figure 6, the frequency of the most used words by years is presented in Figure 6. While there are years between 1980-2023 on the X axis, the cumulative frequencies are on the Y axis. The cumulative frequencies of the words tests, powers, statistics, and model resulted as 116, 86, 82 and 74 in 2023, respectively.

In Figure 7, the most used words are visualized with a Sankey diagram according to specific year periods. Accordingly, the keywords generally vary within each year interval. It can be said that words became more prevalent between 2001-2010 and 2011-2020 periods.

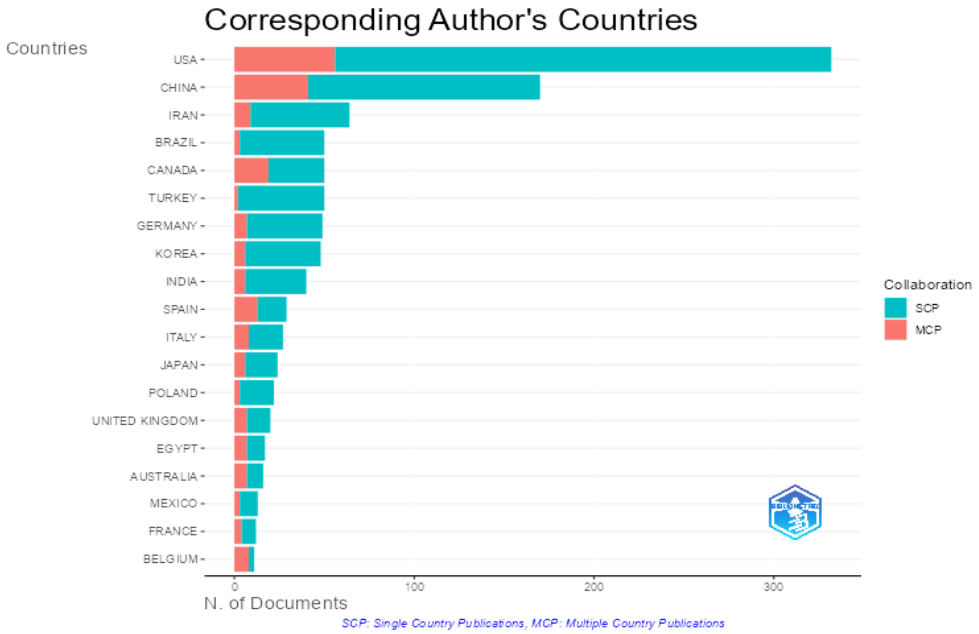

Figure 4. Corresponding authors's countries.

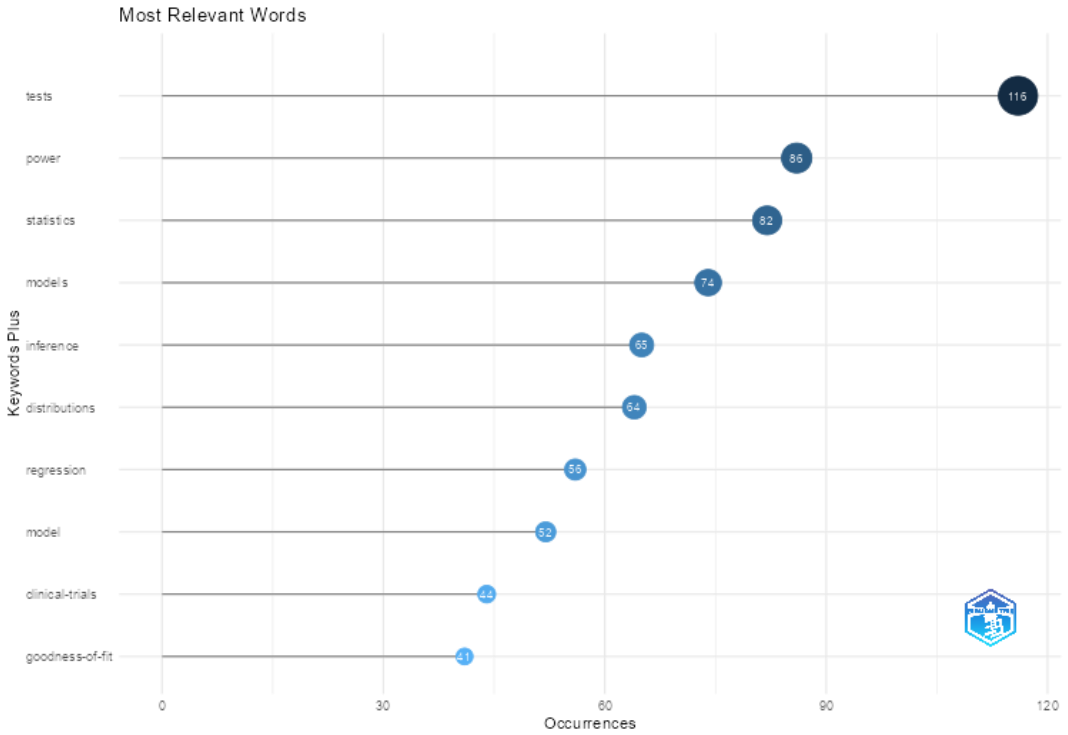

Figure 5. Most relevant words.

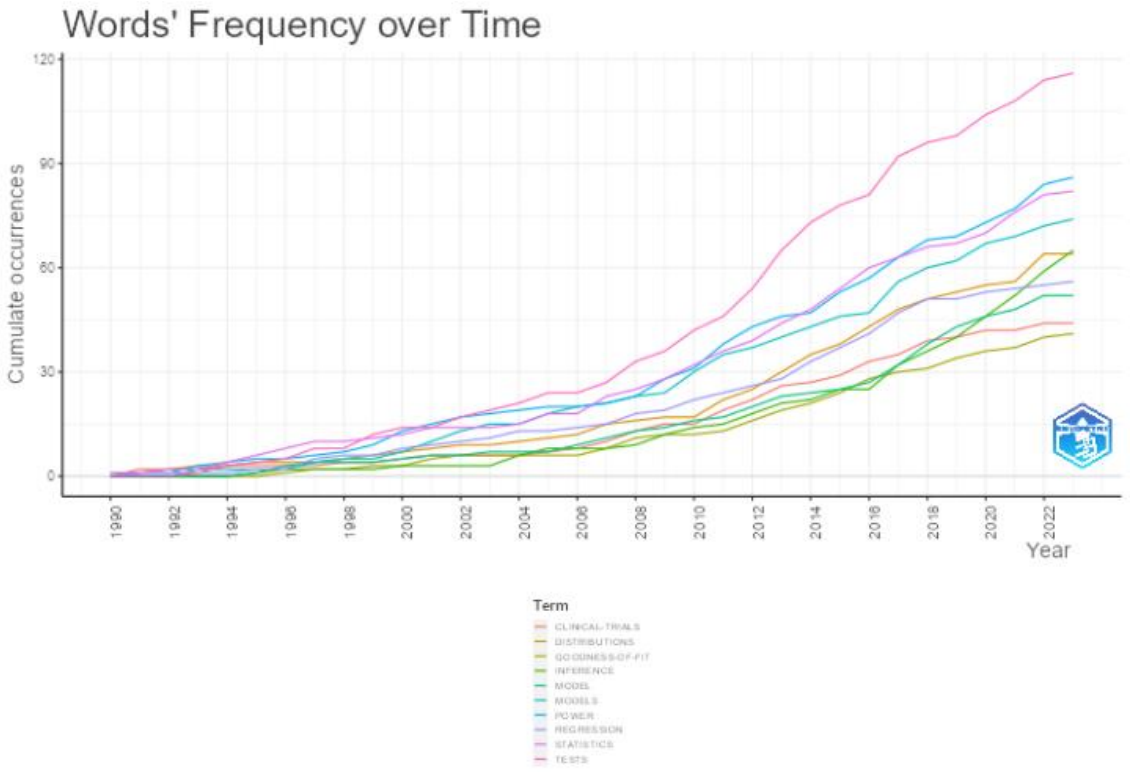

Figure 6. Word's frequency over time.

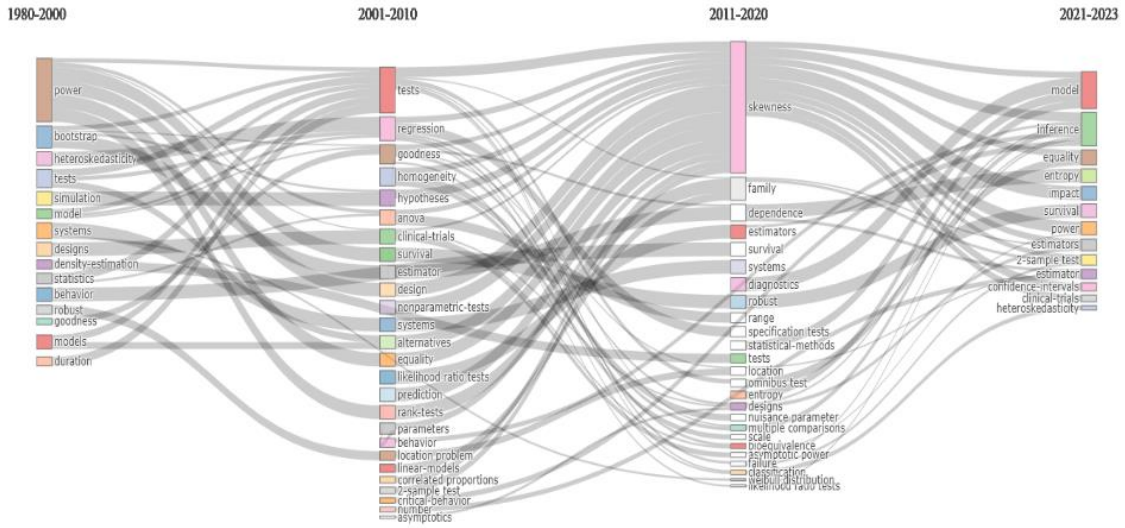

Figure 7. Sankey diagram of most relevant words within certain years.

5. Conclusion

The indication that the findings obtained through bibliometric analysis can be validated by determining the power of tests using the Monte Carlo simulation technique is recognized by the scientific community. Furthermore, when examining the distribution of publications in this field, it is evident that the significance of the subject has never diminished. The reason for this is that the reliability of newly developed tests may be established through studies on the power of the test. It is believed that this study can serve as a guide for researchers in identifying journals where they can publish their simulation studies. Although it is noteworthy that the number of simulation studies has slightly decreased compared to previous years in the present era, they still maintain their relevance even in 2023.

Author Contributions

The percentage of the author(s) contributions is presented below. All authors reviewed and approved the final version of the manuscript.

	M.E.	R.A.D.	Y.A.	Ö.K.
C	30	20	30	20
D	30	20	30	20
S				100
DCP			100	
DAI	40	10	40	10
L	30	20	30	20
W	30	20	30	20
CR	25	25	25	25
SR	50		50	
PM	25	25	25	25
FA	25	25	25	25

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans. The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to.

Acknowledgements

This study is the expanded full text of the abstract presentation titled "Bibliometric Analysis of The Studies Determined by the Monte Carlo Simulation Technique of The Power of The Test" that was presented at the 7th International Conference on Medical & Health Sciences. The previous abstract summarized the main findings of this study; however, this study presents the existing findings in more depth.

References

Aria M, Cuccurullo C. 2017. Bibliometrix: An R-tool for comprehensive science mapping analysis. J Informetr, 11(4): 959-975.

Baspinar E, Gurbuz F. 2000. The power of the test in the samples of various sample sizes were taken from the binary combinations of the normal, beta, gamma and Weibull distributions. J Agric Sci, 6(1): 116-127.

Burton A, Altman DG, Royston P, Holder RL. 2006. The design of simulation studies in medical statistics. Stat Med, 25(24): 4279-4292.

Büyükkidik S. 2022. A Bibliometric Analysis: A tutorial for the bibliometrix package in r using IRT literature. J Meas Eval, 13(3): 164-193.

Çavuş M. 2020. The Computational approach tests for testing the equality of population means under two-parameter exponential distribution. PhD thesis, Eskisehir Technical University, Institute of Graduate Programs, Eskisehir, Türkiye, pp: 105.

Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. 2021. How to conduct a bibliometric analysis: An overview and

guidelines. J Bus Res, 133: 285-296.

Ferreira EB, Rocha MC, Mequelino DB. 2012. Monte Carlo evaluation of the ANOVA's F and Kruskal-Wallis tests under binomial distribution. Sigmae, 1(1): 126-139.

Han J, Kang HJ, Kim M, Kwon GH. 2020. Mapping the intellectual structure of research on surgery with mixed reality: Bibliometric network analysis (2000–2019). J Biomed Inform, 109: 103516.

Koşkan Ö, Gürbüz F. 2009. Comparison of F test and resampling approach for type i error rate and test power by simulation method. J Agric Sci, 15(1): 105-111.

Kroese DP, Brereton T, Taimre T, Botev ZI. 2014. Why the Monte Carlo method is so important today. Interdiscip Rev Comput Stat, 6: 386-392.

Lantz B. 2013. The impact of sample nonnormality on ANOVA and alternative methods. Br J Math Stat Psychol, 66(2): 224–244.

Lloyd CJ. 2005. Estimating test power adjusted for size. J Stat Comput Simul, 75(11): 921-933.

Mahapoonyanont N, Mahapoonyanont T, Pengkaew N, Kamhangkit R. 2010. Power of the test of one-way anova after transforming with large sample size data. Procedia Soc, 9: 993-937.

Moral-Muñoz JA, Herrera-Viedma E, Santisteban-Espejo A, Cobo MJ. 2020. Software tools for conducting bibliometric analysis in science: An up-to-date review. Prof de la Inf, 29(1).

Morris TP, White IR., Crowther MJ. 2019. Using simulation studies to evaluate statistical methods. Stat Med, 38: 2074–2102.

Murphy KR, Myors B, Wolach A. 2014. Statistical power analysis: a simple and general model for traditional and modern hypothesis tests. Routledge, New York, USA, pp: 244.

Öztuna D, Elhan AH, Tüccar E. 2006. Investigation of four different normality tests in terms of type I error rate and power under different distributions. Turk J Med Sci, 36(3): 171-176.

Patric JD. 2009. Simulations to analyze type I error and power in the anova F test and nonparametric alternatives. Master Thesis, University of West Florida, Florida, USA, pp: 80.

Zhang J, Boos DD. 1994. Adjusted power estimates in Monte Carlo experiments. Commun Stat Simul, 23(1): 165-173.