Semantic Grouping and Publication Trends in Neurosurgery over the Past 50 Years

Rikhil Makwana, Rosalind Franklin University, Chicago Medical School

Supplementary/ Citations

Introduction

- Neurosurgery is a field deeply rooted in scientific progress, with ongoing research critical to advancing patient care and safety.
- For neurosurgeon-scientists to remain at the forefront, it is essential to understand evolving research trends.

Objective

To conduct a bibliometric analysis of neurosurgical research from 1994 to 2024, identifying publication trends and guiding future inquiry.

c-TF-IDF

For a term **x** within class **c**:

$$W_{x,c} = \| tf_{x,c} \| \times \log\left(1 + \frac{A}{f_x}\right)$$

$tf_{x,c}$ = frequency of word **x** in class **c**

f_x = frequency of word **x** across all classes

A = average number of words per class

Fig 1) c-TF-IDF Score. Formula for calculating the class-based Term Frequency–Inverse Document Frequency (c-TF-IDF) score. Groups manuscripts based on the relative number of occurrences of certain phrases across documents. Common words such as “the” have relatively low scores. Rare but co-occurring words have higher scores.

Methods

- A SCOPUS[1] search using the term “Neurosurgery” and related keywords **(Supplemental (Supp) 1)** identified English-language articles and reviews published between 1994 and 2024, conducted on May 1, 2025.
- Citations and abstracts were imported into Python for analysis.
- Semantic topic modeling was performed using BERTopic[2] and the *all-MiniLM-L6-v2*[3] transformer on article abstracts with c-TF-IDF scores being calculated per topic generated **(Figure (Fig) 1)**
- Articles were further classified by neurosurgical subspecialty using keyword filtering **(Supp 2)**
- Linear regression models assessed trends across article types and subspecialties[4]

Results

- N = 69,828 citations were identified (n = 52,052 articles; n = 17,776 reviews) **(Fig 2)**
- The most prolific journal was World Neurosurgery (n = 5,763); 5, 895 journals were identified
- BERTopic identified 865 semantic groups, with COVID-19 emerging as the most frequent topic **(Fig 3)**
- Among subspecialties, Spine accounted for the largest number of publications overall (n = 18,059), and from 2001 to 2024 **(Supp 3)**
- Publication trends were best modeled using a quadratic fit (R^2 = 0.967; F = 487) **(Fig 4)**, revealing exponential growth in Spine and Tumor publications **(Supp 4)**

Topic (Most → Least Prevalent)	Keywords (c-TF-IDF Score)				
COVID	covid19 (.04)	pandemic (.025)	coronavirus (.013)	sarscov2 (.008)	2020 (.008)
Aneurysm	avm (0.02)	Avms (0.02)	Arteriovenous (.017)	malformations (0.012)	embolization (0.011)
Stroke/Hemorrhage	ich (0.023)	intracerebral (0.017)	hemorrhage (0.011)	evacuation (0.011)	hematoma (0.009)
Intraoperative Neuromonitoring	evoked (0.015)	Ionm (0.014)	Potentials (0.013)	Monitoring (0.012)	mep (0.01)
Chiari I	Chiari (0.034)	Malformation (0.021)	Cmi (0.014)	syringomyelia (0.011)	decompression (0.01)

Fig 3) Most Common Topics in Neurosurgical Research over a 50 Year Period. COVID was the most common semantic group identified by the BERTopic model. There were a total of 865 total groups identified.

Conclusions

1. Neurosurgical literature has expanded exponentially over the past five decades.
2. Spine, Functional/Stereotactic, and Tumor have published the most across subspecialties.
3. The prominence of COVID-19 underscores the field's responsiveness to global health events.
4. These findings may inform strategic research planning and highlight areas of emerging interest.

Fig 2) Neurosurgical Research Published over a 50-year Period. There were a total of 69,828 titles discovered (Articles = 52,052; Reviews = 17,776). Reviews were classified based automated keyword-based title/abstract screening.

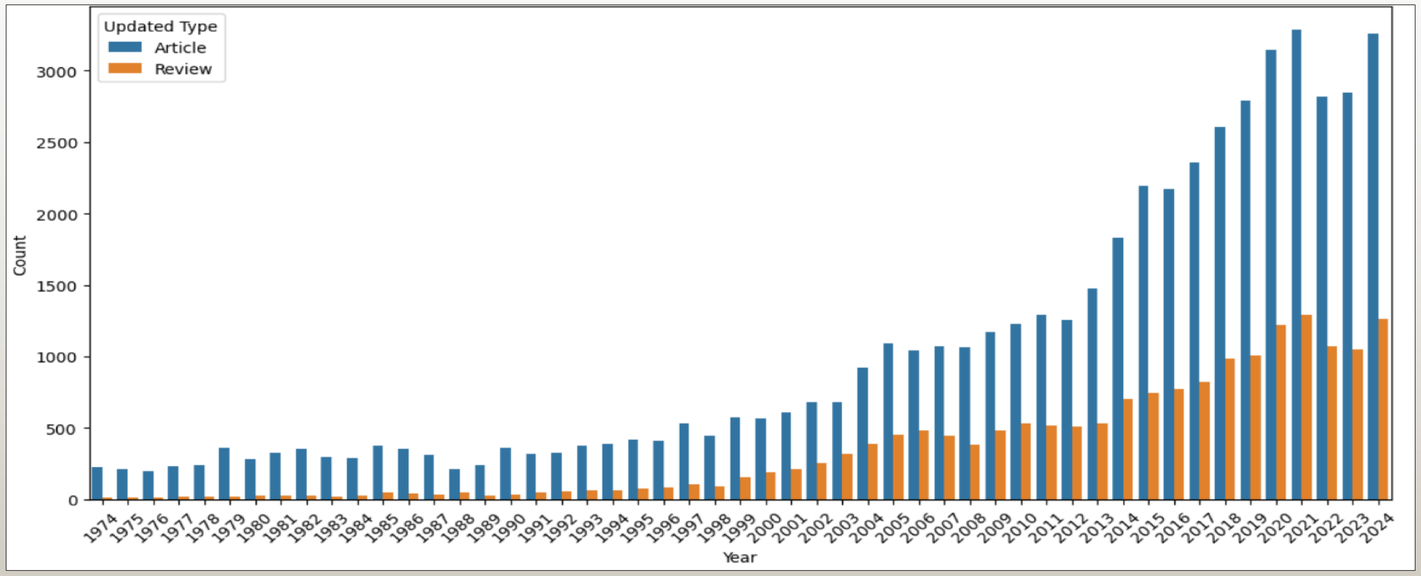

Fig 4) Neurosurgical Research Output Count by Subspecialty over a 50 Year Period. Spine accounts for the largest number of articles published overall (n = 18,059; 30.3%) and from the period of 2001 – 2024. Peripheral Never published the least (n = 304; 0.4%). A quadratic equation was used to model publication trends with shaded areas representing 95% confidence intervals after performing least squares analysis.