

# Big Data Research in Neurosurgery: A Critical Look at this Popular New Study Design

Chesney Oravec; Kevin Reed BS; Mustafa Motiwala; Andrew J. Gianapp; Paul Klimo MD MPH Department of Neurosurgery, University of Tennessee Health Science Center, Memphis, Tennessee Semmes Murphey Clinic, Memphis, Tennessee

THE UNIVERSITY of TENNESSEE 🔰



## Publications by Journal Per Year



	Artic	les Sorte	d by Gene	ral Researc	h Topic
		Vascular		77	23.8%
		Neuro-on	ncology	66	20.4%
		General		61	18.8%
		Pediatrics		56	17.3%
		Spine		55	17.0%
		Functiona	I	9	2.8%
		TOTAL		324	
Articles Sorted b	y Object	ive			
Outcomes	154	47.5%			
Cost	65	20.1%			
Other	62	19.1%			
Complications	43	13.3%			
TOTAL	324				

# Objectives

- To evalutate the trend in usage of "big data" in neurosurgical research looking back to 2000.
- To understand how these databases are being used for research purposes.
- To assess which institutions are producing the majority of these articles.

#### Introduction

As healthcare has transitioned from fee-for-service to qualitybased, the need to capture outcomes data has been met by creation of information repositories called administrative databases, what we are calling "big data". These databases are state or nationwide collections of demographic information, diagnostic and procedural codes, outcome information, and much more. Federal and state governments and private insurance companies use these data for administrative and billing purposes, but clinicians now use these data for research purposes. However, utilizing these databases for research comes with limitations.

#### Methods

- Three major neurosurgical journals (Neurosurgery; Journal of Neurosurgery; World Neurosurgery) were searched for any article, published from 2000 to 2016, that used a nonneurosurgical or an administrative database to answer questions about a neurosurgical disease.
- Information collected from each article included journal title, publication date and publishing journal, database(s) used, sample size, study topic, study objective, and the institutional affiliation of the primary and senior authors.
- Study topics were classified according to a general topic, as well as a specific topic.
- Study objective was defined broadly into either Outcomes, Cost, Complications, or Other.

	By Primary Author			By Senior Author		
1	Harvard Medical School	31	9.6%	Harvard Medical School	28	8.6%
2	Stanford University	17	5.2%	Stanford University	16	4.9%
3	Johns Hopkins University	14	4.3%	University of California San Francisco	16	4.9%
4	University of California San Francisco	13	4.0%	Columbia University		4.3%
5	Columbia University	13	4.0%	Johns Hopkins University	13	4.0%
6	Duke University	12	3.7%	Louisiana State University	12	3.7%
7	Louisiana State University	12	3.7%	University of Minnesota	12	3.7%
8	University of Minnesota	11	3.4%	Mayo Clinic	12	3.7%
9	Mayo Clinic	11	3.4%	University of Southern California	12	3.7%
10	Cedars-Sinai Medical Center	11	3.4%	Duke University	11	3.4%
				University of Florida	11	3.4%
	TOTAL	145	44.8%	TOTAL	157	48.59

#### **Publications by Institution for 324 Articles**

# Results

A total of 324 articles were identified since 2000 with an exponential increase since 2011 (257/324, 79%). The Journal of Neurosurgery Publishing Group published the greatest total number (n=200). The National Inpatient Sample (NIS) was the most commonly used database (n=136). The average study size was 114,841 subjects (range, 30-4,146,777). The most prevalent topics were vascular (n=77) and neuro-oncology (n=66). When categorizing study objective (recognizing that many papers reported more than one type of study objective), "Outcomes" was the most common (n=154). The top 10 institutions by primary or senior author accounted for 45–50% of all publications. Harvard Medical School was the top institution, using this research technique with 59 representations (31 by primary author and 28 by senior).

## Conclusions

Publications in the neurosurgical literature using non-neurosurgery -specific, ready-made databases have dramatically increased over the last 6 years. Many of these articles are looking at procedural outcomes for patients. Two of the major issues with research using big data is the quality and integrity of the data itself. Although beyond the scope this study, the statistical analyses and conclusions drawn from these tests need to be inspected for validity. Thus the value of such studies remains to be determined.

Database Utilized			Mean Sample	Range
HCUP	164	50.6%		
HCUP NIS (Adult)	136	82.9%	161,772	170-1,507,336
HCUP KID (Pediatric)	24	14.6%	42,426	205-443,194
HCUP SID (Adult+Pediatric)	7	4.3%	130,559	936-717,379
NSQIP	40	12.3%		
ACS NSQIP (Adult+Pediatric)	38	95.0%	108,251	114-177,035
VA NSQIP (Adult)	1	2.5%	1,560	N/A
ACS NSQIP (Pediatric)	1	2.5%	21,56	N/A
Medicare	11	3.4%		
Medicare SAF	4	36.4%	9,715	688-22,177
MEDPAR	4	36.4%	8,084	3,210-11,716
Medicare National Claims History 5% sample	2	18.2%	8,371	1,672-15,069
Medicare Physician/Supplier Procedure Summary Master File	1	9.1%	181,421	N/A
Other	112	34.6%		
SEER	35	31.3%	7,477	30-51,125
National Trauma Databank	10	8.9%	144,473	301-839,210