A review of Big Data analytics and potential for implementation in the delivery of global neurosurgery

James L. West, MD, Kyle M. Fargen, MD, MPH, Wesley Hsu, MD, Charles L. Branch Jr., MD, and Daniel E. Couture, MD

Department of Neurosurgery, Wake Forest Baptist Health, Winston-Salem, North Carolina

Global access to neurosurgical care is still a work in progress, with many patients in low-income countries not able to access potentially lifesaving neurosurgical procedures. "Big Data" is an increasingly popular data collection and analytical technique predicated on collecting large amounts of data across multiple data sources and types for future analysis. The potential applications of Big Data to global outreach neurosurgery are myriad: from assessing the overall burden of neurological disease to planning cost-effective improvements in access to neurosurgical care, and collecting data on conditions which are rare in developed countries. Although some global neurosurgical outreach programs have intelligently implemented Big Data principles in their global neurosurgery initiatives already, there is still significant progress that remains to be made. Big Data has the potential to drive the efficient improvement of access to neurosurgical care across low- and medium-income countries.

KEYWORDS Big Data; global neurosurgery; global outreach neurosurgery; data science; low-income country; public health; epidemiology

Recent work in global health estimates that as many as 90% of the people who reside in low-income countries (LICs) may lack access to basic surgical care. Although there have been multiple attempts to increase access to basic neurosurgical care across the spectrum of LICs, no detailed data exist describing the neurosurgical need in these countries. Additionally, no large-scale studies exist addressing a method of population-level implementation of neurosurgical care across LICs or development of research infrastructure to characterize these endeavors. “Big Data,” a discipline within data science, involves methodologies of data collection and handling with the potential to help not only answer the question of how and where neurosurgery is needed across the spectrum of LICs, but also assess these interventions and the impact they have already had.

Background

Global Neurosurgical Need

In recent years, there has been increasing emphasis on improving access to surgical care in the developing world. The Lancet Commission report emphasized a previously identified but under-appreciated lack of access to surgical care for almost 5 billion people in the developing world. It highlighted the potential impact of access to timely, quality surgical care on quality of life and patient survival. Of note, about 150 million “necessary” surgeries are not performed each year. It is estimated that 47 million deaths were preventable had there been adequate access to surgical care for these patients.

Although it is difficult to fully quantify the potential life-saving utility of neurosurgery for the developing world, it is clear that access to neurosurgery is essential for a successful trauma program to prevent death and disability in young patients who typically experience trauma. Additionally, neurosurgical care is important in combating hydrocephalus, which afflicts primarily the young, as well as in treating both ischemic and hemorrhagic strokes affecting the enlarging elderly population. Further, spine and spinal degenerative disease account for significant numbers of lost work days in high-income countries (HICs) and have a major impact in the developing world, where

ABBREVIATIONS HIC = high-income country; LIC = low-income country; TBI = traumatic brain injury.
SUBMITTED June 1, 2018. ACCEPTED July 12, 2018.
INCLUDE WHEN CITING DOI: 10.3171/2018.7.FOCUS18278.
the value it adds are all components that describe the 5 Vs of Big Data. In addition to the volume of data being processed and assessed, the rate at which it is created (velocity), the different types of media the data are stored in (variety), the truthfulness of the data (veracity), and the value it adds are all components that describe the 5 Vs of Big Data.

patients who cannot work may not be able to afford the basic necessities of life.9

Big Data Defined

The term “Big Data” is difficult to accurately define, in part because it is often used interchangeably to reference a number of independent concepts. Initially, the first references to Big Data evoked associations with large, curated datasets or databases. In the 1990s John Mashey popularized the term to describe not only the size of the datasets, but also the variety of information sources and media types, as well as the ever-increasing rate at which the data were being generated.11 By this description, “Big Data” is a broad term applied to a methodology of data analysis rather than just a large collection of homogenous data.

The initial characteristics used to define Big Data included the 3 Vs of Volume, Variety, and Velocity, which were subsequently expanded to the 5 Vs, including Value and Veracity10 (Fig. 1). Chief among the driving forces in the development of Big Data was the explosion of the Internet and social media. The use of these new media by consumers produced orders of magnitude more data than could be collected by traditional focus groups and stored in standard databases. The implementation of this first-generation of Big Data analytics was primarily in the business and manufacturing sector, with examples including improvement of manufacturing productivity by optimizing energy expenditures as well as targeted marketing based on social media posting and search engine queries. In addition to this first wave of Big Data, the combined application of “machine learning” or “deep learning” with Big Data infrastructure for collection and storage of discrete information has been transformative. Within the past decade, computing power has advanced to enable development of deep learning and machine learning algorithms capable of processing and analyzing the vast amounts of data that were being collected and stored. Big Data in its current, most efficient form is the collection of vast amounts of data with subsequent pattern analysis by machine learning algorithms. This model of data analysis is particularly enticing because it does not depend on a priori assumptions of outcome or covariance among the data; it simply analyzes the data for the potential relationships that exist among the myriad variables.

Big Data in Medicine

Although large datasets such as the National (Nationwide) Inpatient Sample (NIS) or the NeuroPoint Alliance Quality Outcomes Database (QOD) are an integral part of data science and driving improvement in neurosurgery, these are not complete examples of Big Data, but rather one component of it. These databases represent the accumulation of a large volume of data about specific patient populations, disease states, and treatments; however, there is little variety to the data being analyzed. Complete application of Big Data techniques occurs not only across large volumes of organized data, but also through the integration of multiple varieties of data, such as social media, vital sign collection, and hospital admission demographics. Big Data techniques were used to assess the progression of cholera in Haiti following the earthquake, when posts to social media were cataloged and stored." Retrospective analysis of these data found that social media trends not only accurately predicted rates of cholera cases, but they were available in near-real time, in contrast to official health department reports, which lagged almost 2 weeks behind. An additional application of Big Data methodologies involved following search-engine queries for dengue fever–related searches regionally. Again, it was found that models generated based off the search engine data accurately predicted the official time course of dengue cases with a correlation coefficient of 0.82–0.99.4

Big Data and Global Neurosurgery

With the established need for surgical services in LICs of almost 4–5 billion people, the potential for data collection and further analysis is huge.12 In addition, the types of data being collected across populations will likely be highly variable, which harnesses a potential strength of Big Data to better tolerate variable data types compared to traditional analytics. In addition to this, a number of problems to be addressed when it comes to the implementation of global neurosurgery across LICs may not be thought of at the time of initial data collection and the beginning of program implementation, which again speaks to a strength of Big Data analysis paradigms. By collecting the data as it is generated and storing it for future analysis, one does not necessarily have to “begin with the end in mind.”

Potential Applications of Big Data for the Delivery of Global Neurosurgery

Defining Global Neurosurgical Need

Perhaps the most clear and pressing objective for global
neurosurgery at the present is establishing and quantifying the prevalence of neurosurgical diseases in LICs. This is a task that Big Data is uniquely suited for. The first step with any endeavor, establishing the need for an intervention, could be completed by compiling multiple national health registries and databases with variable data collection across LICs. Although not specific to LICs, the combination of aggregate databases to assess epidemiological health questions is not new. Previous reports have described the pooling of health registries among the Scandinavian countries to assess for patient outcome and procedural complication data. Assessing aggregate databases established from public health data for rates of neurosurgically treatable diseases such as hydrocephalus, traumatic brain injury (TBI), and stroke or tumor incidence would follow out of this. These data could then potentially be combined with regional social media posts or Internet activity related to key terms defined for each neurosurgical disease process. Combining traditionally slow-adjusting input types such as epidemiological databases with different, more fluid media such as Internet activity and social media presence may allow a more accurate calculation of a composite prevalence of diseases amenable to neurosurgical intervention across LICs. This potential application of Big Data is an important step in assessing the neurosurgical needs of LICs, as the population composition can vary significantly from that of HICs. Demographic reports describing LICs routinely estimate that almost half of the population is made up of children younger than 18 years. Such a radically different population histogram has significant implications for planning neurosurgical efforts when neurosurgeons typically treat patients closer to 50–70 years of age in HICs. This accurate, timely estimate of the prevalence of neurosurgically treatable conditions allows the prediction of the potential economic as well as quality-of-life impact of these diseases which could then be used to appropriate public health funds efficiently to begin combating them. Additionally, defining the need for neurosurgeons and neurosurgical care up front provides an estimated goal for programs to strive for when it comes to raising revenue or setting goals for ratios of neurosurgeons to population in LICs.

Planning for Efficient Neurosurgical Care

Once the overall need has been established, the effective planning of global neurosurgical initiatives is imperative. Locations and populations for intervention should be selected strategically as the beginning battles of a war. Big Data promises to assist in determining the most efficient implementation of global neurosurgery across LICs. By assessing overall prevalence and geospatial distance up front and in concert with epidemiological disease data, efficient points can be selected for the development and implementation of neurosurgical infrastructure. The Duke East Africa neurosurgery program has already demonstrated the power of these types of interventions. In multiple studies they assessed the epidemiology of road traffic injuries across 4 LICs in Africa and also assessed the access to surgical care for pediatric patients stratified by geographic distance. These studies represent informed use of Big Data principles to assess and plan surgical and neurosurgical needs of LICs to plan for global neurosurgical initiatives for maximum impact. In addition to these examples, this same group also demonstrated the potential expansions of their neurosurgical program as well as assessing the cost of their most commonly performed neurosurgical procedures in LICs.

In addition to assisting in the efficient implementation of global neurosurgical initiatives, tracking of geospatial location of injuries as well as site of treatment in LICs can be particularly useful. In LICs with limited access to neurosurgical care, the patient may have to be transported significant distances in order to receive the definitive neurosurgical care they need. By “weighting” each injury with the distance it occurred from the eventual treatment site, the impact of patient transport can be better assessed and mitigated when found to be a significant factor in poor outcomes. Aside from the distance traveled, the time delay in receiving prompt neurosurgical care may also alter complication profiles from those typically seen in HICs. In patients with open skull fractures or cerebrospinal fluid leaks, this time delay could result in a significant rise in associated rates of osteomyelitis or meningitis.

Big Data for Uncommon Clinical Scenarios

Big Data is uniquely suited for the definition of an epidemiological problem as well as planning efficient interventions to address identified disparities. In addition to these methodological advantages, Big Data has unique potential to help us understand more about neurosurgical disease when applied in LICs. The reality of the healthcare in LICs is that, in a number of instances, the ability to diagnose is preserved while the ability to treat may not be readily available, as it would be in a developed HIC. In these cases, observation of natural history, especially of a disorder in which it would be unethical to observe if a treatment were readily available, could help broaden scientific knowledge. Specifically, the natural history of intracranial hypertension secondary to a number of causes, if recorded and collated in a Big Data registry, could help us understand more about the acute changes associated with a myriad of TBI lesions. In addition to TBI, the natural history of a variety of infectious diseases (tuberculosis, rabies, chronic untreated human immunodeficiency virus [HIV] infection, for example) or tumors could be monitored.

Most important when it comes to learning more about these less commonly encountered pathologies is the establishment of a registry or data storage repository. In order to enable this, the establishment of research collaborations among neurosurgical foundations and research programs from high- and low-income countries is necessary. Through partnering and subsidizing of data collection in these LICs, information inaccessible to those from HICs can be collected and analyzed using Big Data analytical paradigms to assess for previously unrecognized relationships or patterns. This analysis could lead to improved care of not only patients in LICs but also patients who present with advanced disease in HICs.

TBI and Intracranial Pressure Monitoring

Although it did not use Big Data methodology, the
study examining intracranial pressure monitoring by Chesnut et al. is a powerful example of the potential for global neurosurgical research to answer questions of importance in both high- and low-income countries. By thoughtfully assessing the problem, the authors designed and implemented a study that was ethical while assessing the efficacy of clinical assessment with imaging interpretation versus serial intracranial pressure monitoring in TBI patients. This innovative melding of neurotrauma and global neurosurgery research serves as an example of a study design concept that could be further employed to answer more neurosurgical questions in a manner that is ethically and methodologically sound.

Duke East Africa Program

This program was initiated in Uganda to develop an infrastructure for the delivery of quality neurosurgical care to the people of east Africa. A number of previously mentioned studies by this program serve as blueprints for the successful integration of Big Data analytics with global neurosurgery. Specifically, their cost analyses of neurosurgical care delivery in LICs as well as their analysis of neurological needs across Uganda and associated referral hospitals typifies application of Big Data principles to global neurosurgical problems. These principles were further utilized during projects assessing the impact of traffic injuries across a number of LICs as well as geographic isolation of pediatric surgical services. All of these aforementioned programs demonstrate the potential power of Big Data for global neurosurgery, although there is still significant room for improvement.

Future Directions

One pillar in the development of Big Data initiatives across the realm of global neurosurgery is the development of new research collaborations between HICs and LICs. A centralized neurosurgical data storage resource would significantly aid in the collection of the requisite data for future application of Big Data principles. Prospective data collected about diseases and the epidemiology of trauma across multiple LICs will not only inform our knowledge of disease, but it will also allow for rapid, intelligent interventions in new LICs by identifying the areas that would benefit the greatest population with the most cost-effective neurosurgical interventions.

In addition to these infrastructure-related improvements and goals, research incorporating the concepts of Big Data more completely is needed. The combination of Big Data multimedia such as Internet search history by location and regional social media trends with traditional data types, such as epidemiological or public health surveys, would maximize the utility of the data generated and could improve upon existing methods. For example, assessing public health records for TBI by location and combining this data with search-engine history for signs and symptoms of postconcussive syndrome could aid in accurately assessing the burden of TBI in specific geographic locales within LICs. By marrying real-time Internet-derived data with more methodically collected records, accurate estimations of neurosurgical disease prevalence and geographic distribution could be generated; these estimations would likely be more accurate than official public health records which typically lag behind emigration and immigration trends. This would aid in targeted application of limited resources to address neurosurgical disease in LICs.

Limitations

A potential limitation of the application of Big Data to global neurosurgery lies in the type of conclusions that may be drawn from this type of data analysis. As in any other research design, bias can be imparted by incomplete data collection leading to a selection-type bias. Additionally, combined data of varying qualities into an aggregate database makes analysis of outcomes and complications difficult. Because of these limitations of Big Data analytics in healthcare, the most appropriate utility of Big Data in global neurosurgery is likely up front, in aiding the characterization of the epidemiology of neurological disease in LICs. In addition to technical limitations, the primary potential logistic limitation is funding and data storage. Meaningful use of Big Data requires significant information storage and analysis capacity, which may not be initially available in LICs, and this would likely necessitate seed funding to set up a viable, long-term venture.

Conclusions

Global neurosurgery is a burgeoning field with increasingly recognized importance, as the need for access to surgical care continues to increase in low-income countries (LICs). Big Data is a promising data analysis technique which, although still in its infancy, has the potential to help enable the efficient growth and implementation of global neurosurgery across LICs. Efficiently improving access to neurosurgical care in LICs promises to reduce mortality and improve overall quality of life for these populations.

References

1. Abdelgadir J, Smith ER, Punchak M, Vissoci JR, Staton C, Muhindo A, et al: Epidemiology and characteristics of neurosurgical conditions at Mbarara regional referral hospital. World Neurosurg 102:526–532, 2017
2. Abdelgadir J, Tran T, Muhindo A, Obiga D, Mukasa J, Ssenyonjo H, et al: Estimating the cost of neurosurgical procedures in a low-income setting: an observational economic analysis. World Neurosurg 101:651–657, 2017
3. Central Intelligence Agency: The World Factbook 2018. Washington, DC: Central Intelligence Agency, 2018
4. Chan EH, Sahai V, Conrad C, Brownstein JS: Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. PLoS Negl Trop Dis 5:e1206, 2011
5. Chesnut RM, Temkin N, Carney N, Dikmen S, Rondina C, Videtta W, et al: A trial of intracranial-pressure monitoring in traumatic brain injury. N Engl J Med 367:2471–2481, 2012
6. Chunara R, Andrews JR, Brownstein JS: Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg 86:39–45, 2012
7. Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L: Clinical epidemiology in the era of big data: new opportunities, familiar challenges. Clin Epidemiol 9:245–250, 2017
8. Fuller AT, Corley J, Tran TM, Butler EK, Vissoci JR, Andrade L, et al: Prevalence of surgically untreated face, head, and neck conditions in Uganda: a cross-sectional nationwide household survey. *World Neurosurg* **110**:e747–e754, 2018

9. Guo HR, Tanaka S, Halperin WE, Cameron LL: Back pain prevalence in US industry and estimates of lost workdays. *Am J Public Health* **89**:1029–1035, 1999

10. Jain A: The 5 Vs of big data. *Watson Health Perspectives*. September 17, 2016. (https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/) [Accessed August 8, 2018]

11. Lohr S: The origins of “Big Data”: an etymological detective story. *New York Times*. February 1, 2013. (https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/) [Accessed August 8, 2018]

12. Meara JG, Hagander L, Leather AJM: Surgery and global health: a Lancet Commission. *Lancet* **383**:12–13, 2014

13. Mooney SJ, Westreich DJ, El-Sayed AM: Commentary: epidemiology in the era of big data. *Epidemiology* **26**:390–394, 2015

14. Oravec CS, Motiwala M, Reed K, Kondziolka D, Barker FG II, Michael LM II, et al: Big data research in neurosurgery: a critical look at this popular new study design. *Neurosurgery* **82**:728–746, 2018

15. Ploss B, Abdelgadir J, Smith ER, Fuller A, Nickenig Vissoci JR, Muhindo A, et al: Pilot use of a novel tool to assess neurosurgical capacity in Uganda. *World Neurosurg* **108**:844–849, 849.e1–849.e4, 2017

16. Smith ER, Vissoci JRN, Rocha TAH, Tran TM, Fuller AT, Butler EK, et al: Geospatial analysis of unmet pediatric surgical need in Uganda. *J Pediatr Surg* **52**:1691–1698, 2017

17. Zafar SN, Canner JK, Nagarajan N, Kushner AL: Road traffic injuries: cross-sectional cluster randomized countrywide population data from 4 low-income countries. *Int J Surg* **52**:237–242, 2018

**Disclosures**

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

**Author Contributions**

Conception and design: West, Couture. Acquisition of data: West. Analysis and interpretation of data: West, Couture. Drafting the article: all authors. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: West. Study supervision: West, Couture.

**Correspondence**

James L. West: Wake Forest Baptist Health, Winston-Salem, NC. jlw@wakehealth.edu.