Check for updates

Editorial: Big Data analytics to advance stroke and cerebrovascular disease: a tool to bridge translational and clinical research

Alexis Nétis Simpkins[1,2]*, Hari Kishan Reddy Indupuru[3] and Sean Isaac Savitz[4]

[1]Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, United States, [2]Department of Neurology, University of Florida, Gainesville, FL, United States, [3]Department of Neurology, University of Texas Health Science Center, Houston, TX, United States, [4]Institute for Stroke and Cerebrovascular Disease, University of Texas Health Science Center, Houston, TX, United States

KEYWORDS

Big Data, stroke, machine learning, translational research, intracerebral hemorrhage, stroke risk, ischemic stroke

Editorial on the Research Topic
Big Data analytics to advance stroke and cerebrovascular disease: a tool to bridge translational and clinical research

Big Data analysis has the potential to enhance the high through put processing required to better phenotype patient outcomes post treatment, select potential therapeutic targets, and refine biomarker selection for risk assessment and disease monitoring (1). With data registries, more advanced imaging, data storage tools, and more detailed electronic clinical documentation, robust analysis can be conducted with large datasets with very granular individual patient level data (1–3). Analysis of large datasets requires special considerations to ensure that the significant associations or findings are clinically meaningful and without bias (1).

Use of a Big Data approach can aid in the discovery of pertinent biomarkers for diagnosis and assessment of stroke risk. Wu et al., used regression modeling to determine which factors were associated with patients with brain infarction detected on magnetic resonance imaging (MRI) in a cohort of 1.4 million patients living in China, demonstrating that there were geographic, sex-related, and metabolic disease risk factors for having infarction detected on brain MRI. Efficacy of anticoagulant type was compared by Lee et al., demonstrating a lower risk of stroke and bleeding complications associated with non-oral vitamin K antagonists. Liao et al. conducted a study including over 5 million patients to confirm the increased risk of stroke in association with markers of insulin resistance. Shu et al. demonstrated that altitude has an increased risk of the development of ischemic changes on MRI and an inverse relationship with risk of clinical events of acute stroke. Yang W.-X. et al. studied the efficacy of several machine learning models to predict genetic stroke risk (LASSO, artificial neural network, random forest, and support vector machine - recursive feature elimination model), showing that there are limitations to using these approaches as their models were limited in their accuracy and specificity. Another approach that can be useful are Mendelian randomization models. Ma et al. were able to demonstrate that genetic variants previously demonstrated

to be associated with elevated homocysteine levels were not associated with an increased risk of intracranial aneurysm detection by using several Mendelian randomization models. Zhou et al. were able to use random forest models to better predict risk of subarachnoid hemorrhage in patients with middle cerebral artery aneurysms. Combining imaging and clinical variables can improve patient phenotyping. Guo et al. investigated machine learning models as a diagnostic tool to diagnosis stroke by automation. For example, Li Y. et al. demonstrated that CT imaging features and markers of small vessel disease are predictive of the presence of >10 cerebral microbleeds on MRI. More research is needed before Big Data analysis such as artificial intelligence and machine learning can be more ubiquitously applied to clinical care (2–6).

Having practical models that allow for quick assessment of risk for hemorrhagic conversion and risk factors for hemorrhagic conversion have the potential to help with stratifying risk of revascularization therapies such as thrombolysis as there are still risks even after special considerations for eligibility for thrombolysis are made based on clinical factors such as duration of symptoms (7, 8), medications, imaging, and clinical comorbidities within 4.5 h window and in the extended time window per the American Heart Association Guidelines on acute ischemic stroke management (8). Ren et al. used modeling and area under the curve receiver operation characteristic curve analysis to develop a score for predicting risk of hemorrhagic conversion with thrombolysis with an area under the curve value of 0.82. Yang M. et al. created a nomogram that predicts stroke risk with thrombolysis using a combination of imaging, clinical, and blood biomarkers. Risk of ischemic hemorrhagic conversion associated with thrombolysis is further reviewed by Shao et al..

Machine learning can also be used to parse areas of cerebral hypoperfusion and areas of normal cerebral perfusion, which is information that has been useful in thrombectomy clinical trials and was incorporated into clinical guidelines for patient selection for thrombectomy (8) Machine learning has been investigated for its diagnostic utility. Lin X. et al. demonstrated that early patient characteristics available within the first 24 h of hospital admission can be predictive of early outcomes post thrombectomy. They were able to fine tune those predictions using different models such as a the SHapley Additive exPlanations approach (Lin X. et al.). Modeling can also be useful in investigations on posterior circulation infarction such as basilar artery occlusion. Zhao C. et al. confirmed that risk factors such as atrial fibrillation increase risk of recurrent stroke but do not influence basilar artery thrombectomy outcomes. While, Lin S. et al. developed a nomogram to help predict in which patient's with basilar artery occlusion recanalization would be futile. Zeng et al., also looked at futility, but they focused on thrombectomy outcomes in the anterior circulation using a combination of machine learning models combined with the stacking method. Currently, the American Heart Association only endorses volumetric analysis for thrombectomy patients in the extended 24 h window (8). However, several large core endovascular trials have subsequently demonstrated that even patients with large cores may still have some benefit from thrombectomy (9–11). More research is needed to optimize prediction tools for patient selection for thrombolysis and thrombectomy.

Cost of stroke care is projected to be over $90 billion dollars by 2035 (1, 12). Part of those costs are attributed to extra healthcare costs related to stroke associated morbidity (1, 12). Determining who is at risk of medical complications after a stroke and tailoring a post stroke recovery plan could be quite impactful (1). Ji et al. used modeling to develop a risk score to predict the risk of being diagnosed with a deep vein thrombosis in patients that were hospitalized with an intracerebral hemorrhage, and optimized their score using external cohort validation. Comparison of machine learning models can demonstrate which model provides the best sensitivity and specificity to predict the clinical outcome of interest. For example, Zheng et al. compared several machine learning models to determine which model would be most specific and sensitive for predicting which patients admitted with an intracerebral hemorrhage would have a post stroke course complicated by the development of pneumonia, showing that the Gaussian naïve Bayes and logistic regression models both performed well depending on whether the internal or external validation cohorts were used. Feng et al. demonstrated similar proteins were elevated during thrombotic events (acute myocardial infarction and acute ischemic stroke), identifying markers of inflammation. In a study including over 100,000 intracerebral hemorrhage patients, Zhao J. et al. combined regression analysis with causal mediation analysis to determine driving factors behind sex-related outcomes, showing the hemorrhage location and clinical severity were the strongest driving factors of mortality and morbidity. Gu et al. demonstrated that mortality rates are higher in critically ill patients with intracerebral hemorrhage and low calcium levels. Others have used Big Data analytic approaches to study length of stay, healthcare utilization, and healthcare costs (3). Currently, there are no widely accepted models for predicting morbidity and mortality for clinical purposes.

Big Data analysis has been studied to provide prediction models to improve management and coordination of post-acute care. Resource utilization post stroke and needs can vary in patients after hospital discharge, and best practices for managing stroke recovery can change over time (13). Prediction models can be used to determine which patient characteristics are the most associated with likelihood of hospital re-admission within 30 days (Chen Y.-C. et al.), which can used to better allocate resources and services for patients. Chen Y.-C. et al. compared multiple models and assessed the sensitivity and specificity of machine learning models to select the best machine learning model that predicted readmission within 30 days of hospital discharge. Yarfi et al. propose using mixed methods models and qualitative analysis to assess post stroke rehabilitation outcomes. Boutros et al. demonstrated that depression was associated with recurrent stroke and mortality 1 year after stroke. Another model that can be useful in analyzing clinical trial data is Bayesian Network Meta-analysis. Li Z. et al. evaluated several approaches for addressing post stroke cognitive dysfunction, and found that transmagnetic stimulation and acupuncture could be helpful. Chen R. et al. demonstrated that machine learning can be used to differentiate responses to transcranial magnetic stimulation between patient's during the post stroke recovery phase by using unsupervised hierarchical clustering, which could have utility in tracking post stroke recovery. In addition, several studies have used a Big Data approach for assessing quality of life indices (3).

Big Data analytics is a rapidly evolving field and there are important considerations and pauses that should be factored into data interpretation and application. It is important to be aware of biases that may be present in datasets as a result of patient recruitment (1–6). Even within large datasets, there may be unknown missing confounders. It is important to consider validation of results in different datasets (1–6).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Simpkins AN, Janowski M, Oz HS, Roberts J, Bix G, Doré S, et al. Biomarker application for precision medicine in stroke. *Transl Stroke Res.* (2020) 11:615–27. doi: 10.1007/s12975-019-00762-3

2. Liu Y, Luo Y, Naidech AM. Big data in stroke: how to use big data to make the next management decision. *Neurotherapeutics.* (2023) 20:744–57. doi: 10.1007/s13311-023-01358-4

3. Olaiya MT, Sodhi-Berry N, Dalli LL, Bam K, Thrift AG, Katzenellenbogen JM, et al. The allure of big data to improve stroke outcomes: review of current literature. *Curr Neurol Neurosci Rep.* (2022) 22:151–60. doi: 10.1007/s11910-022-01180-z

4. Chavva IR, Crawford AL, Mazurek MH, Yuen MM, Prabhat AM, Payabvash S, et al. Deep learning applications for acute stroke management. *Ann Neurol.* (2022) 92:574–87. doi: 10.1002/ana.26435

5. Nenning KH, Langs G. Machine learning in neuroimaging: from research to clinical practice. *Radiologie.* (2022) 62:1–10. doi: 10.1007/s00117-022-01051-1

6. Dipietro L, Gonzalez-Mego P, Ramos-Estebanez C, Zukowski LH, Mikkilineni R, Rushmore RJ, et al. The evolution of Big Data in neuroscience and neurology. *J Big Data.* (2023) 10:116. doi: 10.1186/s40537-023-00751-2

7. Simpkins AN, Tahsili-Fahadan P, Buchwald N, De Prey J, Farooqui A, Mugge LA, et al. Adapting clinical practice of thrombolysis for acute ischemic stroke beyond 4.5 hours: a review of the literature. *J Stroke Cerebrovasc Dis.* (2021) 30:106059. doi: 10.1016/j.jstrokecerebrovasdis.2021.106059

8. Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the american heart association/american stroke association. *Stroke.* (2019) 50:e344–418. doi: 10.1161/STR.0000000000000211

9. Derraz I, Moulin S, Gory B, Kyheng M, Arquizan C, Costalat V, et al. Endovascular thrombectomy outcomes with and without intravenous thrombolysis for large ischemic cores identified with CT or MRI. *Radiology.* (2023) 309:e230440. doi: 10.1148/radiol.230440

10. Sarraj A, Hassan AE, Abraham MG, Ortega-Gutierrez S, Kasner SE, Hussain MS, et al. Trial of endovascular thrombectomy for large ischemic strokes. *N Engl J Med.* (2023) 388:1259–71. doi: 10.1056/NEJMoa2214403

11. Huo X, Ma G, Tong X, Zhang X, Pan Y, Nguyen TN, et al. Trial of endovascular therapy for acute ischemic stroke with large infarct. *N Engl J Med.* (2023) 388:1272–83. doi: 10.1056/NEJMoa2213379

12. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, et al. Heart disease and stroke statistics-2018 update: a report from the american heart association. *Circulation.* (2018) 137:e67–e492. doi: 10.1161/CIR.0000000000000573

13. Olasoji EB, Uhm DK, Awosika OO, Doré S, Geis C, Simpkins AN. Trends in outpatient rehabilitation use for stroke survivors. *J Neurol Sci.* (2022) 442:120383. doi: 10.1016/j.jns.2022.120383