MEDICAL SCHOOL
UNIVERSITY OF MICHIGAN

Capstone for Impact Submission | GY2021

Project Title: Increasing false positive diagnoses may lead to overestimation of stroke incidence in the young.

Student Name(s): Abhinav Appukutty

Advisor Names(s): Dr. James Burke

If this project can be continued by another UMMS student, please include your contact information or any other details you would like to share here:

---

*The material in this CFI submission form has been modified from a first-author manuscript I recently wrote and submitted on this work.

Summary (~250-500 words):

Stroke incidence is reportedly increasing in younger populations, although the reasons for this are not clear. [1-5] We explored possible reasons by quantifying trends in neurologically focused emergency department (ED) visits, classification of stroke vs. TIA, and imaging use. We performed a retrospective, serial, cross-sectional study using the National Hospital Ambulatory Medical Care Survey to examine time trends in age-stratified primary reasons for visit, stroke/TIA diagnoses, and MRI utilization from 1995-2000 and 2005-2015. 5.8 million ED visits with a primary diagnosis of stroke (CI 5.3M-6.4M) were represented in the data. The incidence of neurologically focused reason for visits (Neuro RFVs) increased over time in both the young and in older adults (young: +111 Neuro RFVs/100,000 population/year, CI +94–+130; older adults: +70 Neuro RFVs/100,000 population/year, CI +34–+108). The proportion of combined stroke and TIA diagnoses decreased over time amongst older adults with a Neuro RFV (OR 0.95 per year, $p < 0.01$, CI 0.94–0.96) but did not change in the young (OR 1.00 per year, $p = 0.88$, CI 0.95–1.04). Within the stroke/TIA population, no changes in the proportion of stroke or TIA were identified. MRI utilization rates amongst patients with a Neuro RFV increased for both age groups. In summary, we found, but did not anticipate, increased incidence of neurologically focused ED visits in both age groups. Given the lower pre-test probability of a stroke in younger adults, this suggests that false positive stroke diagnoses may be increasing and may be increasing more rapidly in the young than in older adults.

Methodology:

We performed a retrospective, serial, cross-sectional study on a nationally representative sample of all ED visits in the United States using National Hospital Ambulatory Medical Care Survey (NHAMCS) data from 1995-2000 and 2005-2015. NHAMCS is a set of annual, national probability sample data on utilization and provision of ambulatory care services in hospital emergency and outpatient departments and in ambulatory surgery centers. NHAMCS includes data such as demographics, visit characteristics such as patient's reason for visit (RFV), procedural utilization, and provider's diagnoses. This dataset was chosen based on its national representation and its inclusion of long-term time trends, RFV coding, and all ages of patients. Because this study relies on publicly available data without personal identifiers, it was exempt from review by the Institutional Review Board.

Determining Neurologic Primary Reason for ED Visit (Neuro RFV) – A strength of NHAMCS is inclusion of RFV which represents the patient's complaint, symptom, or other reason for the visit. RFV data were coded in NHAMCS according to "A Reason for Visit Classification for Ambulatory Care". [6] Our final Neuro RFV population was defined as visits by patients with a primary RFV of neurologically focused symptoms or concerns.

Study Populations – Our primary study population was the neurologically focused ED visit (Neuro RFV) population, defined as any patient visit to the ED with a neurologic symptom as the primary reason for visit. Our secondary study population was specifically the Stroke/TIA Population, defined as any patient visit to the ED that receives a primary diagnosis of stroke or TIA by the ED physician. We used International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) to determine visits by patients in whom the ED physician's primary diagnosis was TIA (435.XX) or ischemic stroke (433.x1, 434.x1, 436.xx). [7,8]

Our primary analysis was performed on these two different population subsets from 1995-2015, excluding the years 2001-2004 based on the unavailability of an individual MRI flag for these years. For our primary analysis we stratified population into four broader age groups (< 18 years, 18 – 44 years, 45 – 64 years, and > 65 years). "Young" was defined as 18-44 years old and "older adults" were defined as >65 years old.

We first summarized the characteristics of both study populations (Neuro RFV and Stroke/TIA) with percentages or means and standard deviations (SDs). We then explored the most common RFVs among the stroke/TIA population on the whole and then stratified across age categories. We also examined trends in the most common RFVs in the stroke/TIA population over time, examining top RFVs in three time periods (1995-2000, 2005-2009, and 2010-2015). Additionally, we examined the most common primary diagnoses that were assigned to the Neuro RFV population and how these varied by age group and time period.

Trends in Neuro RFVs – To understand whether the Neuro RFV population itself was changing over time, we also examined trends in Neuro RFV incidence by reporting survey estimates of the number of Neuro RFVs per 100,000 population per year and stratified by age category. We also examined the changes in the proportion of Neuro RFVs out of all ED visits, stratifying by year and age category. We then built a logistic regression model to estimate how the proportion of Neuro RFVs in the total population changed over time using Neuro RFV as the dependent variable and time (year) as the independent variable. We repeated these analyses adjusting for race, sex, and insurance status (private vs. Medicare vs. Medicaid vs. other) to assess whether other factors that may influence Neuro RFVs and stroke/TIA diagnoses over time are contributing to differences in time trends. Finally, we repeated this analysis including an age category-time interaction term to assess whether time trends vary by age group.

We tested our first hypothesis, that stroke/TIA diagnoses are increasing amongst Neuro RFVs using a similar approach. Within the Neuro RFV population, we calculated the proportion that received a stroke or TIA diagnosis, assessed for adjusted and unadjusted time trends using survey-weighted logistic regression and estimated population incidence. To explore the relationship of these trends to MRI utilization, we examined how often MRIs were ordered/performed on the Neuro RFV population using a logistic model. This was examined over each year, stratified by age group, and adjusted for race, sex, and insurance status.

Trends in Stroke/TIA Diagnoses – To test our second hypothesis, we calculated the percentage of strokes and TIAs separately within the stroke/TIA population. To assess for adjusted time trends, we used a logistic model, examined diagnoses over each year, stratified by age group, and adjusted for race, sex, and insurance status. We also examined trends in incidence as explained above.

Guidelines for Statistical Analyses – All statistical analyses were conducted using Stata 14 as to consider the clustered nature of the sample.

Results:

Over the 17-year study period (1995-2000; 2005-2015), 189 million ED visits (95% CI 174M-204M) with a neurologically focused primary RFV (Neuro RFV population) were identified (**Table 1**). Mean age was 46 ± 23 years, and 59% were female (95% CI 58%-60%). 5.8 million ED visits with a primary stroke diagnosis (95% CI 5.3M-6.4M) were identified compared to 3.7 million primary TIA diagnosis (95% CI 3.3M-4.2M). Stroke and TIA patients were older, more likely to be insured by Medicare, and more likely to have comorbidities such as hypertension, diabetes, cerebrovascular disease, and hyperlipidemia. Compared to white patients, black patients had lower representation in the stroke/TIA population (13% in black patients) compared to the broader Neuro RFV population (20% black patients). The Neuro RFV population, compared to the total NHAMCS population, was older and had higher rates of comorbidities.

Characterizing Primary RFVs and Primary Diagnoses – Within the entire Neuro RFV population, the most common primary diagnoses were headache (12%), migraine (7%), dizziness and giddiness (5%), other convulsions (4%), and syncope and collapse (3%). There was little variation in these diagnoses over time. However, there were differences amongst age groups, with older age groups having lower proportions of headache diagnoses. Within the stroke/TIA population, the most common primary RFVs were cerebrovascular disease (14%), (neurologic) weakness (11%), (anesthesia) loss of feeling (9%), general weakness (9%), and vertigo-dizziness (5%). There was little variation in these RFVs over time and between age groups, with cerebrovascular disease being the top primary RFV regardless of time period or age group. However, for older age groups, cerebrovascular disease made up a lower proportion of the primary RFVs.

Temporal Trends in Neuro RFVs – The Neuro RFV incidence (**Figure 1**) was higher overall in absolute terms in older adults, with a significantly increasing trend over time in the young and older adults (young: +111 Neuro RFVs/100,000 population/year, 95% CI +94 – +130; older adults: +70 Neuro RFVs/100,000 population/year, 95% CI +34 – +108). Neuro RFV incidence rose faster in the young compared to older adults (p=0.022). This finding was consistent in subgroup analyses.

Evaluation of Hypothesis 1: Temporal Trends of Stroke/TIA diagnoses – The probability of a stroke/TIA diagnosis within the Neuro RFV population is shown in **Figure 2**. There is a slight downward trend over time in both unadjusted (OR 0.962, 95% CI 0.951 – 0.973, $p < 0.001$) and adjusted (OR 0.960, 95% CI 0.948 – 0.971, $p < 0.001$) analyses. This downward trend was driven primarily by the older adult population (adjusted OR 0.952, 95% CI 0.939 – 0.965, $p < 0.001$) compared to the young (adjusted OR 0.997, 95% CI 0.954 – 1.042, $p = 0.883$).

Evaluation of Hypothesis 2: Variation in the Proportion of Strokes and TIAs – Within the stroke/TIA population, there is a general trend over time towards a higher proportion of strokes compared to TIAs, particularly after 2009 (**Figure 3**). However, there was no clear evidence of a linear temporal trend (adjusted OR 1.001, 95% CI 0.982 – 1.021, $p = 0.896$) or trends over time in the young (adjusted OR 1.018, 95% CI 0.933 – 1.109, $p = 0.692$) or older adult populations (adjusted OR 1.005, 95% CI 0.982 – 1.029, $p = 0.648$). Stroke and TIA incidence were decreasing in older adults and stable in the young.

MRI Utilization – MRI utilization rates for visits with a neurological primary RFV have increased throughout the 17-year period studied for all age groups. Overall, MRI utilization increased over time (adjusted OR 1.078, 95% CI 1.057 – 1.099, $p < 0.001$). This temporal trend was more evident in older adults (adjusted OR 1.090, 95% CI 1.060 – 1.121, $p < 0.001$) compared to the young (adjusted OR 1.059, 95% CI 1.028 – 1.091, $p < 0.001$).

Conclusion (~250-500 words):

In a nationally representative sample of ED visits over 17 years, we did not find evidence to support our primary hypotheses. There was neither a differential increase in the proportion of young people with strokes amongst those presenting with neurologic complaints compared to older adults nor evidence of differential classification of TIA to stroke over time in any age group. We also did not find a disproportionate rise in

MRI use for Neuro RFVs in older adults compared to young adults. Our analysis did identify, however, a robust trend that may play an important role in the apparent increase in stroke in the young.

Specifically, our data support two premises: (1) the prior probability of stroke is considerably higher amongst older adults than in young adults presenting with neurologic symptoms; and (2) the overall incidence of Neuro RFVs is increasing over time and increasing faster in the young compared to older adults. Using these premises, we can estimate the number of false positives is rising, faster in the younger population. Based on those assumptions, it is credible that an increase in false positive strokes in the young may lead to a disproportionate overestimate of the incidence of stroke in the young.

To be clear, these data do not directly measure trends in either true or false positive strokes, rather they measure population-level parameters that enable inferences about the rate of true and false positive strokes/TIAs. Importantly, this logic would apply even if stroke were assessed via any algorithm with less than 100% specificity (even a gold-standard algorithm) as opposed to a claims-based definition as was applied here.

In conclusion, we found, but did not anticipate, increased incidence of neurologically focused ED visits in young and older adults. Given the lower pre-test probability of a stroke in younger adults, our data suggest that false positive stroke diagnoses may be increasing and may be increasing more rapidly in the young than in older adults. Thus, increasing false positive diagnoses in the young might be a contributing factor to the observed increases in stroke incidence in this population. If stroke in the young is truly rising, then this represents a failure of our healthcare system to understand and serve the needs of a large segment of our population. However, we present an alternate theory that merits further exploration.

Reflection/Impact Statement:

My objectives when I started this project were multifaceted: (1) to become familiar with health services research, (2) to learn the statistical analysis skills required to do this research, (3) to learn about stroke and the epidemiology, risk factors, and healthcare problems associated with it, and (4) to design and execute a research project independently. Through the process of conducting this research, I faced many challenges that required confronting my limitations in each of these objectives. I became familiar with the concept of health services research. I also learned how critical understanding large data sets is today to be able to identify problems in health care, analyze current situations, and develop solutions to address these solutions. Moreover, only though the experience of independently leading this project did I encounter and learn to overcome challenges associated with academic research (i.e., writing, editing, and dissemination of work).

This CFI project is initially beneficial to stroke epidemiology researchers as it provides a new perspective and challenges our current understanding of stroke trends in the young adult population. While it would not provide direct clinical benefit to patients on the short term, by encouraging future investigation into this topic, it may ultimately lead public health experts and advisors to reassess our current stroke education and infrastructure strategies to maximize our impact on the age groups that are most affected by stroke. While I have already presented this work at the International Stroke Conference in Los Angeles, CA in February 2020, I will continue to work to disseminate this work in the form of a journal article. I am currently in the submission and revision process for a first-author manuscript, which has itself been challenging and educational.

This experience has been filled with setback, excitement, and growth. I would encourage any student interested in research to begin an independent project early in their career. I believe it was through my experience of "jumping into the deep end", teaching myself new skills, and having a mentor that trusted me and gave me room to experiment and fail that enabled me to develop as a researcher, and ultimately, achieve success in my project.

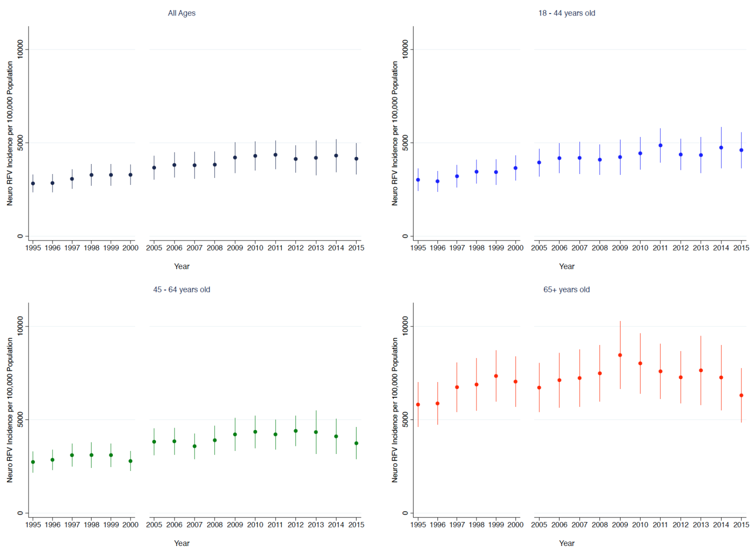
Tables and Figures

TABLE 1: Study Population Baselines Characteristics					
Demographics n (95% CI)	**Neurological RFV n = 189M (174M - 204M)**	**Stroke or TIA n = 9.55M (8.68M – 10.4M)**	**Stroke n = 5.82M (5.25M - 6.39M)**	**TIA n = 3.73M (3.30M - 4.17M)**	**All Visits n = 2.01B (1.85B – 2.17B)**
Age, mean yr (SD)	46 (23)	70 (15)	70 (15)	70 (15)	36 (24)
Female	59%	56%	55%	59%	54%
Race/ethnicity					
White	62%	73%	70%	78%	59%
Black	20%	13%	15%	9%	21%
Hispanic	11%	6%	6%	6%	13%
Other	7%	8%	9%	7%	7%
Insurance					
Private	30%	21%	20%	24%	32%
Medicare	26%	60%	60%	62%	17%
Medicaid	19%	7%	8%	6%	24%
Other	24%	11%	12%	9%	27%
MRI	2%	10%	10%	8%	< 1%
Age Distribution					
< 18	11%	< 1%	1%	< 1%	24%
18 - 44	40%	6%	5%	6%	41%
45 - 64	24%	26%	27%	25%	20%
65 +	25%	68%	67%	69%	15%
Comorbidities†					
Hypertension	32%	66%	65%	68%	22%
Diabetes	13%	27%	33%	19%	9%
CEBVD	7%	60%	60%	61%	3%
Hyperlipidemia	11%	37%	37%	37%	7%

MRI = magnetic resonance imaging; CEBVD = cerebrovascular disease; SD = standard deviation.

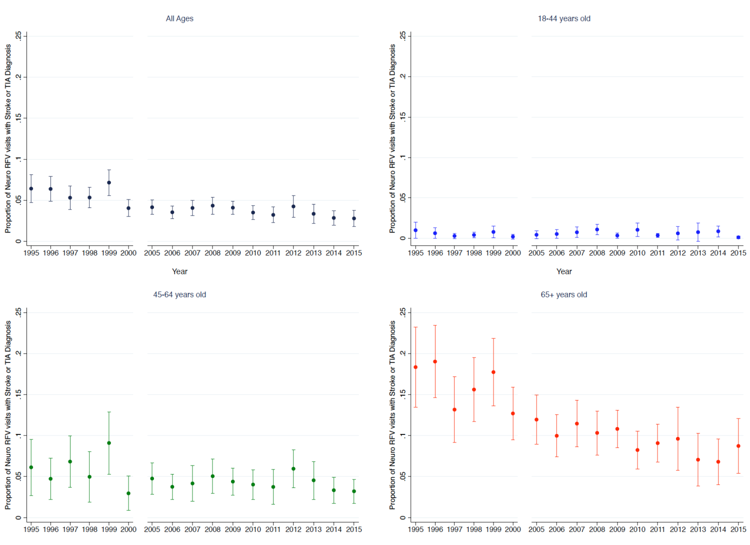
†Diabetes and CEBVD data were only available from 2009 and beyond. Hypertension and Hyperlipidemia data were only available from 2014 and beyond.

Figure 1: Neuro RFV Incidence by Age Group



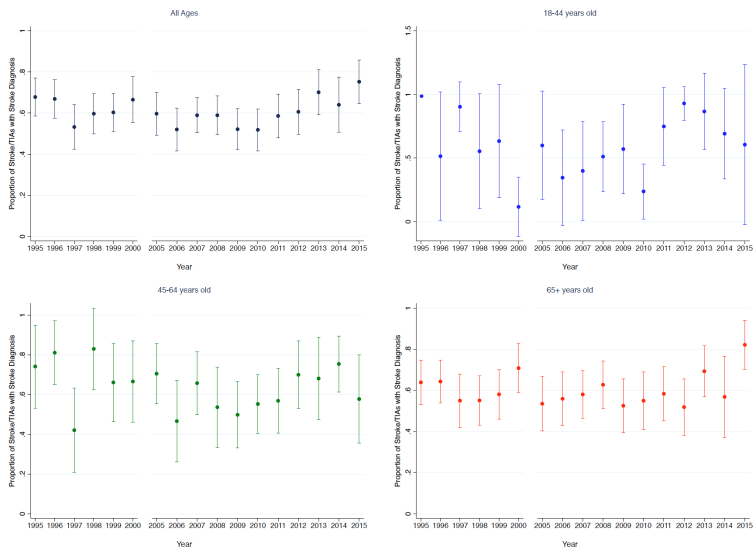
The incidence of Neuro RFVs per 100,000 population by year and by age group (all ages, 18-44 years old, 45-64 years old, and 65+ years old). Error bars represent 95% confidence intervals. In the young (18-44 years, the incidence was +111 Neuro RFVs/100,000 population/year (95% CI +94 – +130), whereas in older adults, incidence was +70 Neuro RFVs/100,000 population/year (95% CI +34 – +108).

Figure 2: Stroke/TIA Diagnoses within Neuro RFV Population



The probability of a stroke/TIA diagnosis within the Neuro RFV population by year and by age group (all ages, 18-44 years old, 45-64 years old, and 65+ years old). Error bars represent 95% confidence intervals. For unadjusted analysis of all ages: OR = 0.962 (95% CI 0.951 – 0.973, $p < 0.001$), while for adjusted: OR = 0.960 (95% CI 0.948 – 0.971, $p < 0.001$). This downward trend was driven primarily by the older adult population (adjusted OR 0.952, 95% CI 0.939 – 0.965, $p < 0.001$) compared to the young (adjusted OR 0.997, 95% CI 0.954 – 1.042, $p = 0.883$).

Figure 3: Proportion of Stroke/TIA Population with Stroke Diagnosis



The percent of stroke or TIA diagnoses with a primary stroke diagnosis by year and by age group (all ages, 18-44 years old, 45-64 years old, and 65+ years old). Error bars represent 95% confidence intervals. There is no evidence of a linear temporal trend in the all ages group (adjusted OR 1.001, 95% CI 0.982 – 1.021, p = 0.896) or trends in the young (adjusted OR 1.018, 95% CI 0.933 – 1.109, p = 0.692) or older adult populations (adjusted OR 1.005, 95% CI 0.982 – 1.029, p = 0.648). For the 18-44 years old group, the proportion of stroke/TIAs with stroke diagnosis was 1 in 1995, so this year was not included in the logistic model, and thus, lacks error bars.

References

1. George MG, Tong X, Bowman BA. Prevalence of cardiovascular risk factors and strokes in younger adults. JAMA Neurol 2017;74:695.

2. Tong X, George MG, Gillespie C, Merritt R. Trends in hospitalizations and cost associated with stroke by age, United States 2003–2012. Int J Stroke 2016;11:874-81.

3. Béjot Y, Delpont B, Giroud M. Rising stroke incidence in young adults: more epidemiological evidence, more questions to be answered. J Am Heart Assoc 2016;5:e003661.

4. Casagrande SS, Menke A, Cowie CC. Cardiovascular risk factors of adults age 20–49 years in the United States, 1971–2012: a series of cross-sectional studies. PLoS One 2016;11:e0161770

5. Leppert MH, Poisson SN, Sillau SH, Campbell JD, Ho PM, Burke JF. Is prevalence of atherosclerotic risk factors increasing among young adults? It depends on how you ask. J Am Heart Assoc 2019;8:e010883.

6. D. Schneider, L. Appleton, T. McLemore. A reason for visit classification for ambulatory care. Vital Health Stat 2 1979;(78):i-vi, 1-63.

7. Goldstein LB. Accuracy of ICD-9-CM coding for the identification of patients with acute ischemic stroke. Stroke 1998;29:1602-4.

8. Tirschwell DL, Longstreth W. Validating administrative data in stroke research. Stroke 2002;33:2465-70.