

DETERMINANTS OF LONG-TERM OUTCOMES AFTER STROKE: INSIGHTS FROM
LINKED MICHIGAN COVERDELL ACUTE STROKE REGISTRY

By

Ra'ed S. Hailat

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Epidemiology – Doctor of Philosophy

2024

ABSTRACT

Data needed to provide a comprehensive assessment of long-term recovery of stroke survivors is lacking for the Michigan Acute Stroke Registry, referred to as MiSP (Michigan Stroke Program), as it is for many other US-based stroke registries. The overall objective of this dissertation is to bridge this knowledge gap by linking stroke registry data with administrative claims data to obtain follow-up data on patient outcomes. The administrative data source is the Michigan Value Collaborative (MVC) a comprehensive, statewide, claims database that includes claims from Blue Cross Blue Shield of Michigan (BCBSM) (private and Medicare Advantage plans) and Medicare fee-for-service (FFS).

In the first aim we generated a linked database by combining a 5-year retrospective cohort (2016-2020) of all acute stroke discharges entered into MiSP registry from 31 stroke certified hospitals (n=46,330) with MVC (n= 30,685) claims data using both deterministic and probabilistic matching techniques. We evaluated the accuracy, completeness, and representativeness of the linkage results using pre linkage qualitative and post linkage quantitative methods. We then generated descriptive data on 30-day, 90-day and 1-year outcome event rates including mortality, all-cause hospital readmissions, stroke recurrence, use of post-acute care services (i.e., inpatient rehabilitation facility (IRF), skilled nursing facility (SNF), and home health), use of out-patient visits, and home time. We showed that probabilistic linkage between MiSP acute stroke registry and MVC claims data using indirect identifiers produced slightly higher linkage rates compared to deterministic linkage and that our linkage is feasible and resulted in a valid linked dataset that has acceptable representation of Medicare FFS and BCBSM insured population in Michigan.

In the second aim we developed 30-day and 1-year all-cause readmission prediction models using 3 different machine learning (ML) methods: LASSO logistic regression, XGBoost

and ANN, and compared the predictive performance of these methods. After identifying the optimal performing model, we report the most important predictors. Our findings demonstrated that prediction of all cause readmission can be achieved with relatively modest accuracy (AUC range 0.67 - 0.68), that LASSO regression was able to predict readmission after stroke with similar accuracy to more advanced ML methods, and that clinical features of stroke (e.g., NIHSS, stroke etiology) were less important than the burden of existing comorbidities (e.g., chronic renal failure, atrial fibrillation, heart failure) or the hospitalization (e.g., admission duration, discharge destination) in predicting post-stroke readmission, especially over longer periods of time (1-year).

The third aim was to estimate the comparative effectiveness of IRF versus SNF rehabilitation care to improve functional recovery 90 days and 1 year following discharge using inverse probability of treatment weighting analysis of differences in home time (number of days alive and outside of inpatient care) and mortality. This analysis was limited to Medicare FFS stroke patients. Our findings provided further evidence that discharge to IRF versus SNF was associated with longer home time and lower mortality over 1-year post discharge. However, our sensitivity analysis illustrated that home time is heavily impacted by rehabilitation length of stay especially over 90-days and hence future studies should avoid using home time and rely on more stable measures (less prone to bias) like mRS or successful community discharge (home for >30 consecutive days).

In conclusion, we illustrated that data linkage can provide needed information to describe patient recovery up to 1 year after acute stroke discharge. Future studies should expand on the generalizability of the linkage by including data from more hospitals and claims data from Medicaid and other insurance providers.

Copyright by
RA'ED S. HAILAT
2024

I dedicate this work to my family especially my parents for their unconditional support and teaching me the vitality of knowledge and humbleness. To my teachers and mentors who saw the potential in me. Thank you.

ACKNOWLEDGMENTS

I would like to acknowledge and thank Dr. Mathew J. Reeves – my committee chair and primary advisor. Dr. Reeves’s dedicated mentorship has been invaluable to my professional, scholarly, and personal development over the past five years. It was through his skilled guidance that I have been able to conduct this interdisciplinary work, and successfully pursue an AHA-funded award (predoctoral fellowship) that facilitated my training. Additionally, before obtaining my own funding, I benefited from multiple Research and Teaching Assistant positions at the Department of Epidemiology and Biostatistics, College of Human Medicine, and Michigan State University. These funding opportunities made my PhD possible, and I hope to do the same forward when I have mentees of my own.

I would also like to acknowledge and thank my dissertation committee – Dr. Michael Thompson especially for his help in accessing Michigan Value Collaborative data, Dr. Gustavo De Los Campos and Dr. Adam Oostema – for generously sharing their expertise and wisdom as I made my way through this dissertation. Also, I would like to thank Adrienne Nickles at Michigan Department of Health and Human Services for helping in obtaining access to the Michigan Stroke Program data.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1 Objective and Specific Aims.....	1
1.2 Significance.....	3
1.3 Dissertation Organization and Overview	7
1.4 Dissertation Funding	7
1.5 Dissertation Institutional Review Board (IRB) and Data Usage Agreement (DUA).....	7
BIBLIOGRAPHY	9
APPENDIX	16
CHAPTER 2: BACKGROUND LITERATURE REVIEW	20
2.1 Historical Discovery of Brain Vasculature and Stroke Pathology From the Ancient Egyptians to the Early 20th Century	20
2.2 What is a Stroke? Case Definitions of Acute Stroke for Clinical, Research, and Administrative Purposes	24
2.3 Stroke Registries and The Importance of Data Linkage	26
2.4 The Michigan Stroke Registry and Michigan Value Collaborative Claims Database	28
2.5 Summary of Published Stroke Outcome Statistics in the US.....	31
BIBLIOGRAPHY	36
CHAPTER 3: MANUSCRIPT 1 – ACCURACY AND REPRESENTATIVENESS OF PATIENT-LEVEL OUTCOMES DATA FOLLOWING LINKAGE OF A STATEWIDE STROKE REGISTRY TO AN ADMINISTRATIVE CLAIMS DATABASE.....	48
3.1 Abstract	48
3.2 Introduction	50
3.3 Methods.....	52
3.4 Results	73
3.5 Discussion	85
3.6 Conclusions	94
BIBLIOGRAPHY	96
APPENDIX	103
CHAPTER 4: MANUSCRIPT 2 – PREDICTION OF HOSPITAL READMISSION AFTER STROKE USING MACHINE LEARNING IN A 5-YEAR LINKED COHORT FROM THE MICHIGAN STROKE REGISTRY.....	117
4.1 Abstract	117
4.2 Introduction	119
4.3 Methods.....	123
4.4 Results	139
4.5 Discussion	151
4.6 Conclusions	158
BIBLIOGRAPHY	159
APPENDIX	167

CHAPTER 5: MANUSCRIPT 3 – THE COMPARATIVE EFFECTIVENESS OF INPATIENT REHABILITATION FACILITY VERSUS SKILLED NURSING FACILITY IN PATIENTS DISCHARGED FOLLOWING ACUTE STROKE IN A MICHIGAN COHORT.....	187
5.1 Abstract	187
5.2 Introduction	190
5.3 Methods.....	194
5.4 Results	208
5.5 Discussion	218
5.6 Future Directions and Conclusions	227
BIBLIOGRAPHY	229
APPENDIX	236
CHAPTER 6: SUMMARY AND DISCUSSION	248
6.1 Summary of Findings and Limitations.....	248
6.2 Direction of Future Research	253
6.3 Implications for Public Health, Clinical Practice, and Public Policy.....	255
6.4 Conclusions	256

CHAPTER 1: INTRODUCTION

1.1 Objective and Specific Aims

Data needed to provide a comprehensive assessment of long-term recovery of stroke survivors is lacking for the Michigan Acute Stroke Registry, referred to as MiSP (Michigan Stroke Program), as it is for many other US-based stroke registries.^{1, 2} The overall objective of this dissertation is to bridge this knowledge gap by linking clinical and administrative claims data sources to obtain comprehensive data on patient outcomes. The administrative data source is Michigan Value Collaborative (MVC)- a comprehensive, statewide, claims database that includes data from Medicare fee-or-service (FFS) and Blue Cross Blue Shield of Michigan (BCBSM) private and Medicare Advantage insured populations.³ The linked dataset will enable us to report on a wide range of stroke outcome measures including mortality, recurrence, and readmission, and admission to rehabilitation. The specific aims of this research are:

Specific aim 1:

1a: Generate a unique database by linking a 5-year retrospective cohort of all acute stroke discharges entered into MiSP registry between 2016-2020 with MVC registry data using both deterministic and probabilistic matching techniques.

1b: Evaluate the accuracy, completeness, and representativeness of the linkage results using pre linkage qualitative and post linkage quantitative methods.

1c: Use the linked data to generate descriptive data on 30-day, 90-day and 1-year outcome event rates including use of post-acute care services (i.e., inpatient rehabilitation facility, skilled nursing facility, and home health), use of out-patient visits, all-cause hospital readmissions, stroke recurrence, and mortality and home time (these latter two outcomes are available only for Medicare FFS beneficiaries).

Specific aim 2:

2a: Develop 30-day and 1-year all-cause readmission prediction models using simple machine learning (ML) LASSO logistic regression, and two non-linear ML based methods (XGBoost and ANN), compare the predictive performance of these three methods when applied to stroke registry data, and report the most important predictors from the best performing prediction method.

2b: To examine the impact of using different combinations of data sources (i.e., MISP registry, MVC administrative data, and American Hospital Association hospital survey) on the predictive performance of the three methods.

Specific aim 3: Use the linked dataset to estimate the comparative effectiveness of inpatient rehabilitation facility versus skilled nursing facility institutional rehabilitation care to improve functional recovery of Medicare FFS acute stroke patients using home time calculated at 90-days and 1-year following discharge from an index stroke hospitalization. As a secondary outcome we compared 90-day and 1-year all-cause mortality between the two settings.

Through this work, we will contribute to the scientific body of literature in several important ways. First, by generating a statewide linked dataset that will permit assessment of long-term (up to 1-year) outcomes following hospitalization for acute stroke from Medicare FFS and BCBSM beneficiaries. Although up to 10 prior papers have reported results from data linkages between GWTG-S registry data and claims data, a linked dataset using MiSP data has not been generated before.⁴⁻¹³ In addition, only one of these prior linked GWTG-S studies included claims data from commercial health plans and Medicare Advantage members.¹³ Second, providing a comprehensive assessment of the accuracy of the linkage results using pre

linkage qualitative and post linkage quantitative methods will be novel. Prior stroke registries including GWTG-Stroke that have created linked datasets using claims data⁴⁻¹³ have not previously conducted assessments of linkage accuracy. Third, comparing the performance of a simple ML method (i.e., LASSO logistic regression) and 2 advanced ML (i.e., XGBoost, and ANN) methods in predicting readmission at 30-days and 1-year post discharge using linked data will be novel because only two previous US-based stroke readmission prediction studies relied on ML models and both utilized electronic medical records data.^{14, 15} Also we note that prior stroke registry based studies have mostly reported on 30-day readmissions¹⁴⁻¹⁸. Fourth, examining the impact of using different combinations of data sources (i.e., registry, administrative, and hospital survey data) on the predictive performance of the ML methods has been performed by only one study.¹⁴ Finally, investigating the comparative effectiveness of inpatient rehabilitation facility vs skilled nursing facility institutional rehabilitation care on functional recovery post-acute stroke discharge using home time has only been done by one study previously.¹⁹ Each of these 3 specific aims will be presented in Chapters 3, 4, and 5 of this dissertation, respectively.

1.2 Significance

1.2.1 Registries and Data Linkage

Registries collect a uniform body of data to evaluate population characteristics and outcomes for a defined disease, condition, or exposure with an ultimate aim of improving quality of care.^{20, 21} In the last 20 years the development of national-level²² and state-level² hospital-based acute stroke registries have provided data needed to facilitate important improvements in the quality of stroke care,²³⁻²⁶ reduced treatment gaps^{27, 28} and disparities in stroke care,²³ and have contributed to improved patient outcomes for acute stroke patients.^{11, 23, 28-30} The large

volume of data collected, which for the national Get With The Guideline-Stroke (GWTG-S) program, now exceeds 9 million stroke discharges from more than 2,000 hospitals, have allowed for the detailed examination of the associations between patient- and hospital- level characteristics and improvements in quality of care and outcomes for stroke patients up to the point of hospital discharge.^{26, 30} However, these studies are limited by the fact that patient outcomes are not collected following hospital discharge. Although patients who survive stroke can continue to recover and improve for many months if not years after the index event, the majority of the recovery of function and community participation occur within the first three to six-months following hospital discharge.³¹ Collection of patient-level outcomes data addressing survival, recurrence, utilization, function, and quality of life has been a challenge for stroke registries because of the substantial investment of resources, both human and financial, required to follow-up and interview stroke survivors post discharge.³² A more feasible and sustainable alternative to tracking each individual patient is to obtain data through data linkage between stroke registries and other large-scale databases including administrative (claims) data, electronic medical records (eMR), vital records and census data which can provide a rich source of patient-level information on outcomes,^{5, 33-35} including post discharge mortality,^{4, 6, 8-12, 36} readmissions,^{4, 6, 8, 9, 11} stroke reoccurrence,^{4, 6, 9-11} use of post-acute care services,³⁶ home time,^{4, 6, 7, 9} and cost.¹⁰ The lack of post discharge data means that stroke registries cannot achieve one of their central aims which is to improve not just the acute, but also longer-term outcomes for acute stroke patients through the delivery of high-quality stroke care.

Different data sources (i.e., electronic medical records, administrative or claims, registry, and hospital survey data) are designed to serve different purposes hence each will have different strengths and limitations.^{20, 21, 37, 38} Registries collect a uniform body of data to evaluate the

quality of medical care as well as specific patient outcomes for a defined disease, condition, or exposure.^{20, 21} Administrative or claims-based data are generated at every encounter with the health care provider including but not limited to physician visits, procedures, hospital or facility admissions, and prescription fillings. However, claims data includes limited clinical information because it is collected for insurance billing and reimbursement purposes.³⁸ Hospital surveys such as those administered by the American Hospital Association are annual surveys designed to collect quantitative and qualitative system- and hospital- level information related to operations, utilization, service lines, staffing, system structure, and other data points. However, hospital surveys do not collect any patient-level data.³⁷ Combining the above data sources through data linkage can bridge gaps in data limitations from a single data source, providing a richer source of detailed patient-, hospital, and system- level data. These data sources once linked together can identify associations that would be difficult if not impossible to determine otherwise.^{34, 35, 38}

1.2.2 Hospital Readmission and Patient Recovery Following Rehabilitation

Two important patient outcomes following acute stroke are hospital readmission and functional recovery following rehabilitation care. A meta-analysis published in 2016 of 10 reports published between 2006 and 2015, estimated the pooled 30-day and 1-year all-cause stroke readmission rates in the US as 17.4% (95% CI, 12.7–23.5%) and 42.5% (95% CI, 34.1–51.3%), respectively.³⁹ Medicare insured patients (>=65 years) have a high all-cause 30-day readmission rate of 16.9% with an estimated total cost of about \$26 billion annually.^{15, 40, 41} The Center for Medicare and Medicaid Services (CMS) regards reducing readmissions as one of the central goals of national healthcare reforms.^{15, 42} In 2012, CMS identified readmissions as a measure of hospital quality and developed the Hospital Readmissions Reduction Program (HRRP) with the goal of reducing readmissions nationwide.^{40, 42-44} However, hospital

readmission rates are driven by a myriad of patient, hospital- and system- level factors,⁴⁵⁻⁴⁸ and identifying these factors is important as they can guide the development of potential interventions.⁴⁵⁻⁴⁸ For hospitals using prediction models to identify patients at high risk of readmission before discharge can be helpful to identify patients that can be targeted to receive specific interventions such as enhanced transitional care management.⁴⁹

To promote recovery following stroke approximately two thirds of stroke survivors discharged from hospital receive post-acute care that typically includes rehabilitation care.^{50, 51} Nationally representative data from GWTG-S registry data reported that 25.4% and 19.5% of acute stroke patients were discharged to inpatient rehabilitation facility and skilled nursing facility rehabilitation services, respectively.⁵¹ Inpatient rehabilitation facilities provide intensive, interdisciplinary rehabilitation care under the direct supervision of a physician,⁵² whereas skilled nursing facilities provide less intensive rehabilitation care to stroke survivors who need both nursing or rehabilitation care.⁵² The clinical decision to discharge a given patient to one of these two types of facilities is complex, in part because there remains considerable uncertainty regarding the comparative effectiveness of the two settings on the functional recovery for individual stroke patients.^{53, 54}

Although there is general consensus among researchers that discharge to inpatient facility is associated with better functional outcomes compared to skilled nursing facility,^{50, 55} all of the comparative studies conducted to date in the US (total of 10) mostly relied on observational designs.^{19, 52, 56-59} The limited number of studies that utilized follow up data is attributed to the lack of data on functional recovery following institutional-based rehabilitation care and the fragmentation of health services in the US.⁵⁰ Obtaining functional recovery metrics (e.g., modified Rankin Scale, and mobility scores) relies on individual patient follow-up that is costly

and hard to achieve at scale and over the long term.⁶⁰ Home time which is defined as the total post discharge time spent alive and out of an inpatient care setting (i.e., hospital, inpatient rehabilitation, skilled nursing facility, and long-term care hospitals), is an alternative approach to quantifying functional recovery that can be generated from administrative data.¹⁹

1.3 Dissertation Organization and Overview

This Dissertation has been organized in 6 chapters. In Chapter 1, we provided an overview of the overall dissertation objective, three principal specific aims, and the significance of the work. In Chapter 2, we will provide relevant background information and literature review regarding stroke registries, data linkage, stroke outcomes including stroke readmission, and post-acute rehabilitation care in the US. In Chapters 3, 4 and 5, we will present publishable work that describes the the results of the three specific aims. Finally, Chapter 6 will provide a discussion summarizing the findings and implications of this dissertation.

1.4 Dissertation Funding

This dissertation was supported by the American Heart Association through a Predoctoral Fellowship Grant Number 909423 (PI: Raed S Hailat). The content is solely the responsibility of the authors and does not necessarily represent the official views of the American Heart Association.

1.5 Dissertation Institutional Review Board (IRB) and Data Usage Agreement (DUA)

This research was approved by Michigan State University (MSU), University of Michigan (UM), and Michigan Department of Health and Human Services (MDHHS) Institutional Review Boards (IRB) (See Appendix). Data Usage Agreements (DUA) between MSU-UM, MSU-MDHHS and UM-PI were signed to transfer data to a secured server at UM and gain access to the data through a secured VPN connection. MiSP and MVC datasets are both

classified as limited data sets according to The Health Insurance Portability and Accountability Act (HIPPA) hence patient consent was not required.⁶¹

BIBLIOGRAPHY

1. Michigan Department of Health and Human Services (MDHHS), Michigan Stroke Program (MiSP). <https://www.michigan.gov/mdhhs/keep-mi-healthy/communicablediseases/epidemiology/chronicepi/stroke> (accessed 2023).
2. Center of Disease Control and Prevention, Paul Coverdell National Acute Stroke Program. https://www.cdc.gov/dhdsp/programs/stroke_registry.htm (accessed 2023).
3. Michigan Value Collaborative, MVC Data Resources. <https://michiganvalue.org/resources-2/> (accessed 2023).
4. Xian, Y.; Wu, J.; O'Brien, E. C.; Fonarow, G. C.; Olson, D. M.; Schwamm, L. H.; Bhatt, D. L.; Smith, E. E.; Suter, R. E.; Hannah, D.; Lindholm, B.; Maisch, L.; Greiner, M. A.; Lytle, B. L.; Pencina, M. J.; Peterson, E. D.; Hernandez, A. F., Real world effectiveness of warfarin among ischemic stroke patients with atrial fibrillation: observational analysis from Patient-Centered Research into Outcomes Stroke Patients Prefer and Effectiveness Research (PROSPER) study. *BMJ* **2015**, *351*, h3786. <https://doi.org/10.1136/bmj.h3786>.
5. Reeves, M. J.; Fonarow, G. C.; Smith, E. E.; Pan, W.; Olson, D.; Hernandez, A. F.; Peterson, E. D.; Schwamm, L. H., Representativeness of the Get With The Guidelines-Stroke Registry: comparison of patient and hospital characteristics among Medicare beneficiaries hospitalized with ischemic stroke. *Stroke* **2012**, *43* (1), 44-9. <https://doi.org/10.1161/STROKEAHA.111.626978>.
6. O'Brien, E. C.; Greiner, M. A.; Xian, Y.; Fonarow, G. C.; Olson, D. M.; Schwamm, L. H.; Bhatt, D. L.; Smith, E. E.; Maisch, L.; Hannah, D.; Lindholm, B.; Peterson, E. D.; Pencina, M. J.; Hernandez, A. F., Clinical Effectiveness of Statin Therapy After Ischemic Stroke: Primary Results From the Statin Therapeutic Area of the Patient-Centered Research Into Outcomes Stroke Patients Prefer and Effectiveness Research (PROSPER) Study. *Circulation* **2015**, *132* (15), 1404-13. <https://doi.org/10.1161/CIRCULATIONAHA.115.016183>.
7. Fonarow, G. C.; Liang, L.; Thomas, L.; Xian, Y.; Saver, J. L.; Smith, E. E.; Schwamm, L. H.; Peterson, E. D.; Hernandez, A. F.; Duncan, P. W.; O'Brien, E. C.; Bushnell, C.; Prvu Bettger, J., Assessment of Home-Time After Acute Ischemic Stroke in Medicare Beneficiaries. *Stroke* **2016**, *47* (3), 836-42. <https://doi.org/10.1161/STROKEAHA.115.011599>.
8. Fonarow, G. C.; Smith, E. E.; Reeves, M. J.; Pan, W.; Olson, D.; Hernandez, A. F.; Peterson, E. D.; Schwamm, L. H.; Get With The Guidelines Steering, C.; Hospitals, Hospital-level variation in mortality and rehospitalization for medicare beneficiaries with acute ischemic stroke. *Stroke* **2011**, *42* (1), 159-66. <https://doi.org/10.1161/STROKEAHA.110.601831>.
9. Kaufman, B. G.; O'Brien, E. C.; Stearns, S. C.; Matsouaka, R.; Holmes, G. M.; Weinberger, M.; Song, P. H.; Schwamm, L. H.; Smith, E. E.; Fonarow, G. C.; Xian, Y., The

Medicare Shared Savings Program and Outcomes for Ischemic Stroke Patients: a Retrospective Cohort Study. *J Gen Intern Med* **2019**, *34* (12), 2740-2748. <https://doi.org/10.1007/s11606-019-05283-1>.

10. Kaufman, B. G.; Shah, S.; Hellkamp, A. S.; Lytle, B. L.; Fonarow, G. C.; Schwamm, L. H.; Lesen, E.; Hedberg, J.; Tank, A.; Fita, E.; Bhalla, N.; Atreja, N.; Bettger, J. P., Disease Burden Following Non-Cardioembolic Minor Ischemic Stroke or High-Risk TIA: A GWTC-Stroke Study. *J Stroke Cerebrovasc Dis* **2020**, *29* (12), 105399. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105399>.

11. Song, S.; Fonarow, G. C.; Olson, D. M.; Liang, L.; Schulte, P. J.; Hernandez, A. F.; Peterson, E. D.; Reeves, M. J.; Smith, E. E.; Schwamm, L. H.; Saver, J. L., Association of Get With The Guidelines-Stroke Program Participation and Clinical Outcomes for Medicare Beneficiaries With Ischemic Stroke. *Stroke* **2016**, *47* (5), 1294-302. <https://doi.org/10.1161/STROKEAHA.115.011874>.

12. Reeves, M. J.; Fonarow, G. C.; Xu, H.; Matsouaka, R. A.; Xian, Y.; Saver, J.; Schwamm, L.; Smith, E. E., Is Risk-Standardized In-Hospital Stroke Mortality an Adequate Proxy for Risk-Standardized 30-Day Stroke Mortality Data? Findings From Get With The Guidelines-Stroke. *Circ Cardiovasc Qual Outcomes* **2017**, *10* (10). <https://doi.org/10.1161/CIRCOUTCOMES.117.003748>.

13. Patorno, E.; Schneeweiss, S.; George, M. G.; Tong, X.; Franklin, J. M.; Pawar, A.; Mogun, H.; Moura, L.; Schwamm, L. H., Linking the Paul Coverdell National Acute Stroke Program to commercial claims to establish a framework for real-world longitudinal stroke research. *Stroke Vasc Neurol* **2022**, *7* (2), 114-123. <https://doi.org/10.1136/svn-2021-001134>.

14. Lineback, C. M.; Garg, R.; Oh, E.; Naidech, A. M.; Holl, J. L.; Prabhakaran, S., Prediction of 30-Day Readmission After Stroke Using Machine Learning and Natural Language Processing. *Front Neurol* **2021**, *12*, 649521. <https://doi.org/10.3389/fneur.2021.649521>.

15. Darabi, N.; Hosseinichimeh, N.; Noto, A.; Zand, R.; Abedi, V., Machine Learning-Enabled 30-Day Readmission Model for Stroke Patients. *Front Neurol* **2021**, *12*, 638267. <https://doi.org/10.3389/fneur.2021.638267>.

16. Chen, Y. C.; Chung, J. H.; Yeh, Y. J.; Lou, S. J.; Lin, H. F.; Lin, C. H.; Hsien, H. H.; Hung, K. W.; Yeh, S. J.; Shi, H. Y., Predicting 30-Day Readmission for Stroke Using Machine Learning Algorithms: A Prospective Cohort Study. *Front Neurol* **2022**, *13*, 875491. <https://doi.org/10.3389/fneur.2022.875491>.

17. Lv, J.; Zhang, M.; Fu, Y.; Chen, M.; Chen, B.; Xu, Z.; Yan, X.; Hu, S.; Zhao, N., An interpretable machine learning approach for predicting 30-day readmission after stroke. *Int J Med Inform* **2023**, *174*, 105050. <https://doi.org/10.1016/j.ijmedinf.2023.105050>.

18. Xu, Y.; Yang, X.; Huang, H.; Peng, C.; Ge, Y.; Wu, H.; Wang, J.; Xiong, G.; Yi, Y., Extreme Gradient Boosting Model Has a Better Performance in Predicting the Risk of 90-Day Readmissions in Patients with Ischaemic Stroke. *J Stroke Cerebrovasc Dis* **2019**, *28* (12), 104441. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.104441>.
19. Prvu Bettger, J.; Thomas, L.; Liang, L. *Comparing Recovery Options for Stroke Patients*; Patient-Centered Outcomes Research Institute (PCORI): Washington (DC), 2019.
20. In *Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition, Addendum 2*, Gliklich, R. E.; Leavy, M. B.; Dreyer, N. A., Eds. Rockville (MD), 2019.
21. In *Registries for Evaluating Patient Outcomes: A User's Guide*, 3rd ed.; Gliklich, R. E.; Dreyer, N. A.; Leavy, M. B., Eds. Rockville (MD), 2014.
22. American Heart Association, Get With The Guide Line - Stroke. <https://www.heart.org/en/professional/quality-improvement/get-with-the-guidelines/get-with-the-guidelines-stroke/get-with-the-guidelines-stroke-overview> (accessed 2023).
23. Schwamm, L. H.; Reeves, M. J.; Pan, W.; Smith, E. E.; Frankel, M. R.; Olson, D.; Zhao, X.; Peterson, E.; Fonarow, G. C., Race/ethnicity, quality of care, and outcomes in ischemic stroke. *Circulation* **2010**, *121* (13), 1492-501. <https://doi.org/10.1161/CIRCULATIONAHA.109.881490>.
24. George, M. G.; Tong, X.; McGruder, H.; Yoon, P.; Rosamond, W.; Winquist, A.; Hinchey, J.; Wall, H. K.; Pandey, D. K.; Centers for Disease, C.; Prevention, Paul Coverdell National Acute Stroke Registry Surveillance - four states, 2005-2007. *MMWR Surveill Summ* **2009**, *58* (7), 1-23.
25. Parker, C.; Schwamm, L. H.; Fonarow, G. C.; Smith, E. E.; Reeves, M. J., Stroke quality metrics: systematic reviews of the relationships to patient-centered outcomes and impact of public reporting. *Stroke* **2012**, *43* (1), 155-62. <https://doi.org/10.1161/STROKEAHA.111.635011>.
26. Howard, G.; Schwamm, L. H.; Donnelly, J. P.; Howard, V. J.; Jasne, A.; Smith, E. E.; Rhodes, J. D.; Kissela, B. M.; Fonarow, G. C.; Kleindorfer, D. O.; Albright, K. C., Participation in Get With The Guidelines-Stroke and Its Association With Quality of Care for Stroke. *JAMA Neurol* **2018**, *75* (11), 1331-1337. <https://doi.org/10.1001/jamaneurol.2018.2101>.
27. Fonarow, G. C.; Smith, E. E.; Saver, J. L.; Reeves, M. J.; Hernandez, A. F.; Peterson, E. D.; Sacco, R. L.; Schwamm, L. H., Improving door-to-needle times in acute ischemic stroke: the design and rationale for the American Heart Association/American Stroke Association's Target: Stroke initiative. *Stroke* **2011**, *42* (10), 2983-9. <https://doi.org/10.1161/STROKEAHA.111.621342>.

28. Heidenreich, P. A.; Hernandez, A. F.; Yancy, C. W.; Liang, L.; Peterson, E. D.; Fonarow, G. C., Get With The Guidelines program participation, process of care, and outcome for Medicare patients hospitalized with heart failure. *Circ Cardiovasc Qual Outcomes* **2012**, 5 (1), 37-43. <https://doi.org/10.1161/CIRCOUTCOMES.110.959122>.
29. American Heart Association, Get With The Guidelines® - Stroke Patient Management Tool. <https://www.heart.org/en/professional/quality-improvement/get-with-the-guidelines/get-with-the-guidelines-stroke/get-with-the-guidelines-stroke-patient-management-tool>.
30. Ormseth, C. H.; Sheth, K. N.; Saver, J. L.; Fonarow, G. C.; Schwamm, L. H., The American Heart Association's Get With the Guidelines (GWTG)-Stroke development and impact on stroke care. *Stroke Vasc Neurol* **2017**, 2 (2), 94-105. <https://doi.org/10.1136/svn-2017-000092>.
31. Lee, K. B.; Lim, S. H.; Kim, K. H.; Kim, K. J.; Kim, Y. R.; Chang, W. N.; Yeom, J. W.; Kim, Y. D.; Hwang, B. Y., Six-month functional recovery of stroke patients: a multi-time-point study. *Int J Rehabil Res* **2015**, 38 (2), 173-80. <https://doi.org/10.1097/MRR.0000000000000108>.
32. Reeves, M.; Lisabeth, L.; Williams, L.; Katzan, I.; Kapral, M.; Deutsch, A.; Prvu-Bettger, J., Patient-Reported Outcome Measures (PROMs) for Acute Stroke: Rationale, Methods and Future Directions. *Stroke* **2018**, 49 (6), 1549-1556. <https://doi.org/10.1161/STROKEAHA.117.018912>.
33. Yu, A. Y.; Holodinsky, J. K.; Zerna, C.; Svenson, L. W.; Jette, N.; Quan, H.; Hill, M. D., Use and Utility of Administrative Health Data for Stroke Research and Surveillance. *Stroke* **2016**, 47 (7), 1946-52. <https://doi.org/10.1161/STROKEAHA.116.012390>.
34. Bradley, C. J.; Penberthy, L.; Devers, K. J.; Holden, D. J., Health services research and data linkages: issues, methods, and directions for the future. *Health Serv Res* **2010**, 45 (5 Pt 2), 1468-88. <https://doi.org/10.1111/j.1475-6773.2010.01142.x>.
35. Dusetzina, S. B.; Tyree, S.; Meyer, A. M.; Meyer, A.; Green, L.; Carpenter, W. R., In *Linking Data for Health Services Research: A Framework and Instructional Guide*, Rockville (MD), 2014.
36. Cadilhac, D. A.; Kim, J.; Lannin, N. A.; Kapral, M. K.; Schwamm, L. H.; Dennis, M. S.; Norrving, B.; Meretoja, A., National stroke registries for monitoring and improving the quality of hospital care: A systematic review. *Int J Stroke* **2016**, 11 (1), 28-40. <https://doi.org/10.1177/1747493015607523>.
37. American Hospital Association, AHA Annual Survey Database. <https://www.ahadata.com/> (accessed 2023).

38. Cadarette, S. M.; Wong, L., An Introduction to Health Care Administrative Data. *Can J Hosp Pharm* **2015**, *68* (3), 232-7. <https://doi.org/10.4212/cjhp.v68i3.1457>.
39. Zhong, W.; Geng, N.; Wang, P.; Li, Z.; Cao, L., Prevalence, causes and risk factors of hospital readmissions after acute stroke and transient ischemic attack: a systematic review and meta-analysis. *Neurol Sci* **2016**, *37* (8), 1195-202. <https://doi.org/10.1007/s10072-016-2570-5>.
40. Beauvais, B.; Whitaker, Z.; Kim, F.; Anderson, B., Is the Hospital Value-Based Purchasing Program Associated with Reduced Hospital Readmissions? *J Multidiscip Healthc* **2022**, *15*, 1089-1099. <https://doi.org/10.2147/JMDH.S358733>.
41. Weiss, A. J.; Jiang, H. J., Overview of Clinical Conditions With Frequent and Costly Hospital Readmissions by Payer, 2018. In *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*, Rockville (MD), 2021.
42. Lichtman, J. H.; Leifheit-Limson, E. C.; Jones, S. B.; Wang, Y.; Goldstein, L. B., Preventable readmissions within 30 days of ischemic stroke among Medicare beneficiaries. *Stroke* **2013**, *44* (12), 3429-35. <https://doi.org/10.1161/STROKEAHA.113.003165>.
43. Bambhroliya, A. B.; Donnelly, J. P.; Thomas, E. J.; Tyson, J. E.; Miller, C. C.; McCullough, L. D.; Savitz, S. I.; Vahidy, F. S., Estimates and Temporal Trend for US Nationwide 30-Day Hospital Readmission Among Patients With Ischemic and Hemorrhagic Stroke. *JAMA Netw Open* **2018**, *1* (4), e181190. <https://doi.org/10.1001/jamanetworkopen.2018.1190>.
44. Fischer, C.; Lingsma, H. F.; Marang-van de Mheen, P. J.; Kringos, D. S.; Klazinga, N. S.; Steyerberg, E. W., Is the readmission rate a valid quality indicator? A review of the evidence. *PLoS One* **2014**, *9* (11), e112282. <https://doi.org/10.1371/journal.pone.0112282>.
45. Leppin, A. L.; Gionfriddo, M. R.; Kessler, M.; Brito, J. P.; Mair, F. S.; Gallacher, K.; Wang, Z.; Erwin, P. J.; Sylvester, T.; Boehmer, K.; Ting, H. H.; Murad, M. H.; Shippee, N. D.; Montori, V. M., Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. *JAMA Intern Med* **2014**, *174* (7), 1095-107. <https://doi.org/10.1001/jamainternmed.2014.1608>.
46. Hansen, L. O.; Young, R. S.; Hinami, K.; Leung, A.; Williams, M. V., Interventions to reduce 30-day rehospitalization: a systematic review. *Ann Intern Med* **2011**, *155* (8), 520-8. <https://doi.org/10.7326/0003-4819-155-8-201110180-00008>.
47. Finkelstein, A.; Taubman, S.; Doyle, J., Health Care Hotspotting - A Randomized, Controlled Trial. Reply. *N Engl J Med* **2020**, *382* (22), 2173-2174. <https://doi.org/10.1056/NEJMc2001920>.

48. Kansagara, D.; Chiovaro, J. C.; Kagen, D.; Jencks, S.; Rhyne, K.; O'Neil, M.; Kondo, K.; Relevo, R.; Motu'apuaka, M.; Freeman, M.; Englander, H., So many options, where do we start? An overview of the care transitions literature. *J Hosp Med* **2016**, *11* (3), 221-30. <https://doi.org/10.1002/jhm.2502>.
49. Marafino, B. J.; Escobar, G. J.; Baiocchi, M. T.; Liu, V. X.; Plimier, C. C.; Schuler, A., Evaluation of an intervention targeted with predictive analytics to prevent readmissions in an integrated health system: observational study. *BMJ* **2021**, *374*, n1747. <https://doi.org/10.1136/bmj.n1747>.
50. Winstein, C. J.; Stein, J.; Arena, R.; Bates, B.; Cherney, L. R.; Cramer, S. C.; Deruyter, F.; Eng, J. J.; Fisher, B.; Harvey, R. L.; Lang, C. E.; MacKay-Lyons, M.; Ottenbacher, K. J.; Pugh, S.; Reeves, M. J.; Richards, L. G.; Stiers, W.; Zorowitz, R. D.; American Heart Association Stroke Council, C. o. C.; Stroke Nursing, C. o. C. C.; Council on Quality of, C.; Outcomes, R., Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke* **2016**, *47* (6), e98-e169. <https://doi.org/10.1161/STR.000000000000098>.
51. Prvu Bettger, J.; McCoy, L.; Smith, E. E.; Fonarow, G. C.; Schwamm, L. H.; Peterson, E. D., Contemporary trends and predictors of postacute service use and routine discharge home after stroke. *J Am Heart Assoc* **2015**, *4* (2). <https://doi.org/10.1161/JAHA.114.001038>.
52. Deutsch, A.; Granger, C. V.; Heinemann, A. W.; Fiedler, R. C.; DeJong, G.; Kane, R. L.; Ottenbacher, K. J.; Naughton, J. P.; Trevisan, M., Poststroke rehabilitation: outcomes and reimbursement of inpatient rehabilitation facilities and subacute rehabilitation programs. *Stroke* **2006**, *37* (6), 1477-82. <https://doi.org/10.1161/01.STR.0000221172.99375.5a>.
53. Ottenbacher, K. J.; Graham, J. E., The state-of-the-science: access to postacute care rehabilitation services. A review. *Arch Phys Med Rehabil* **2007**, *88* (11), 1513-21. <https://doi.org/10.1016/j.apmr.2007.06.761>.
54. Hayes, H. A.; Mor, V.; Wei, G.; Presson, A.; McDonough, C., Medicare Advantage Patterns of Poststroke Discharge to an Inpatient Rehabilitation or Skilled Nursing Facility: A Consideration of Demographic, Functional, and Payer Factors. *Phys Ther* **2023**, *103* (4). <https://doi.org/10.1093/ptj/pzad009>.
55. Alcusky, M.; Ulbricht, C. M.; Lapane, K. L., Postacute Care Setting, Facility Characteristics, and Poststroke Outcomes: A Systematic Review. *Arch Phys Med Rehabil* **2018**, *99* (6), 1124-1140 e9. <https://doi.org/10.1016/j.apmr.2017.09.005>.
56. Hoenig, H.; Sloane, R.; Horner, R. D.; Zolkewitz, M.; Reker, D., Differences in rehabilitation services and outcomes among stroke patients cared for in veterans hospitals. *Health Serv Res* **2001**, *35* (6), 1293-318.

57. Kind, A. J.; Smith, M. A.; Liou, J. I.; Pandhi, N.; Frytak, J. R.; Finch, M. D., Discharge destination's effect on bounce-back risk in Black, White, and Hispanic acute ischemic stroke patients. *Arch Phys Med Rehabil* **2010**, *91* (2), 189-95. <https://doi.org/10.1016/j.apmr.2009.10.015>.
58. Wang, H.; Sandel, M. E.; Terdiman, J.; Armstrong, M. A.; Klatsky, A.; Camicia, M.; Sidney, S., Postacute care and ischemic stroke mortality: findings from an integrated health care system in northern California. *PM R* **2011**, *3* (8), 686-94. <https://doi.org/10.1016/j.pmrj.2011.04.028>.
59. Hong, I.; Goodwin, J. S.; Reistetter, T. A.; Kuo, Y. F.; Mallinson, T.; Karmarkar, A.; Lin, Y. L.; Ottenbacher, K. J., Comparison of Functional Status Improvements Among Patients With Stroke Receiving Postacute Care in Inpatient Rehabilitation vs Skilled Nursing Facilities. *JAMA Netw Open* **2019**, *2* (12), e1916646. <https://doi.org/10.1001/jamanetworkopen.2019.16646>.
60. ElHabr, A. K.; Katz, J. M.; Wang, J.; Bastani, M.; Martinez, G.; Gribko, M.; Hughes, D. R.; Sanelli, P., Predicting 90-day modified Rankin Scale score with discharge information in acute ischaemic stroke patients following treatment. *BMJ Neurol Open* **2021**, *3* (1), e000177. <https://doi.org/10.1136/bmjno-2021-000177>.
61. National Institute of Health HIPPA privacy rule. https://privacyruleandresearch.nih.gov/pr_08.asp (accessed 2023).

APPENDIX

Michigan Department of Health and Human Services
Institutional Review Board for the Protection of Human Research Subjects
South Grand Building, 5th Floor, 333 S. Grand Ave., P.O. Box 30195, Lansing, MI 48909
E-mail: MDHHS-IRB@michigan.gov Phone: (517) 241-1928 Fax: (517) 241-1200

DETERMINATION NOTICE – EXPEDITED REVIEW PROCEDURES

To: Adrienne Nickles Responsible Department Employee
From: Ian A. Horste Institutional Review Board Chair
CC: Beth Anderson Authorizing Bureau/Office Director

MDHHS IRB Log #: 202106-03-EA **Date Received:** 05/26/22

Study Title: Determinants of long-term outcomes after ischemic stroke: insights from linked Michigan Coverdell Acute Stroke Registry

Primary Investigator(s): Matthew Reeves

Funding Source(s): Non-Federal - Unspecified

Determination Type:

- Not human subjects research
 Exempt human subjects research
 Non-exempt human subjects research approved by expedited review

Comments: This research remains eligible for continued approval under the 2018 Common Rule. Established waivers of informed consent requirements remain in effect as previously documented.

Chair Signature: 

Determination Date: 06/07/22

Status Update Due*: 06/20/23

**For exempt human subjects research and non-exempt human subjects research approved by expedited review, the MDHHS IRB requests an update on the status of the research before this date. Conduct of human subjects research after this date without acknowledgement by the MDHHS IRB of a status update is noncompliant with MDHHS policy.*

The MDHHS IRB must approve any change to non-exempt human subjects research protocols and associated study documents. Approval of changes must precede implementation, unless a change is necessary to eliminate an apparent immediate hazard to research subjects. The Primary Investigator and Responsible Department Employee must see that any unanticipated problem or severe adverse event in approved research (exempt or non-exempt) is reported as soon as possible (usually within 48 hours of discovery) to the MDHHS IRB administrative office at (517) 241-1928 or MDHHS-IRB@michigan.gov.

Michigan Department of Health and Human Services FWA00007331, IRB00000421

**MICHIGAN STATE
UNIVERSITY**

**Initial Study APPROVAL
Revised Common Rule**

April 29, 2021

To: Mathew J Reeves

Re: **MSU Study ID:** STUDY00006017
IRB: Biomedical and Health Institutional Review Board
Principal Investigator: Mathew J Reeves
Category: Expedited 5
Submission: Initial Study STUDY00006017
Submission Approval Date: 4/29/2021
Effective Date: 4/29/2021
Study Expiration Date: **None; however modification and closure submissions are required (see below).**

Title: Determinants of long-term outcomes after ischemic stroke: insights from linked Michigan Coverdell Acute Stroke Registry



**Office of
Regulatory
Affairs
Human Research
Protection Program**

4000 Collins Road
Suite 136
Lansing, MI 48910

517-355-2180
Fax: 517-432-4503
Email: irb@msu.edu
www.hrpp.msu.edu

This submission has been approved by the Michigan State University (MSU) BIRB. The submission was reviewed by the Institutional Review Board (IRB) through the Non-Committee Review procedure. The IRB has found that this study protects the rights and welfare of human subjects and meets the requirements of MSU's Federal Wide Assurance (FWA00004556) and the federal regulations for the protection of human subjects in research (e.g., 2018 45 CFR 46, 21 CFR 50, 56, other applicable regulations).

Institutional restrictions to in-person human subject research activities conducted by MSU employees, MSU students, or agents of MSU are in place, but MSU is phasing in human research that has the potential for in-person interactions with participants, using a Tier approach. Restrictions to in-person interactions with human research participants by MSU employees, MSU students, or agents of MSU are in place until the activity is permitted under a Tier and a Human Research Plan for a Safe Return is approved. Visit <http://hrpp.msu.edu/COVID-19/index.html> for the restrictions, Tiers, forms, and the process.

How to Access Final Documents

To access the study's final materials, including those approved by the IRB such as consent forms, recruitment materials, and the approved protocol, if applicable, please log into the Click™ Research Compliance System, open the study's workspace, and view the "Documents" tab. To obtain consent form(s) stamped with the IRB watermark, select the "Final" PDF version of your consent form(s) as applicable in the "Documents" tab. Please note that the consent form(s) stamped with the IRB watermark must typically be used.



Medical School Institutional Review Board (IRB MED) • 2800 Plymouth Road, Building 52D, Suite 3214, Ann Arbor, MI 48109-2800 • phone (734) 763 4768 • fax (734) 763 9603 • irbmed@umich.edu

To: [Michael Thompson](#)

From:

Michael	Geisser
Alan	Sugar
Robertson	Davenport

Cc:

Michael	Thompson
Kathryn	Ashbaugh

Subject: Notice of Determination of "Not Regulated" Status for [HUM00204773]

SUBMISSION INFORMATION:

Title: Determinants of Long-Term Outcomes After Ischemic Stroke
 Full Study Title (if applicable):
 Study eResearch ID: [HUM00204773](#)
 Date of this Notification from IRB: 8/19/2021
 Date of IRB Not Regulated Determination: 8/19/2021

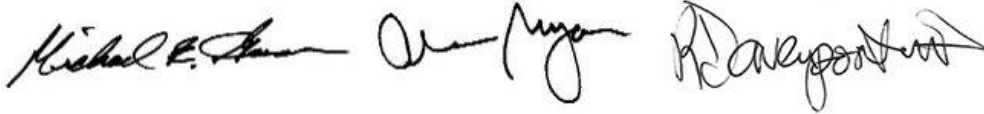
IRB NOT REGULATED STATUS:

Category	Outcome Letter Text
Research Involving Coded Private Information	Based on the information provided, this study involves only coded private information or biological specimens that cannot be linked to a specific individual by the investigator(s) directly or indirectly through a coding system. In accordance with OHRP guidance on this subject (See http://www.hhs.gov/ohrp/policy/cdebiol.html), IRB approval is not required as the data cannot be tracked to a human subject.

HIPAA REVIEW:

The IRB has reviewed the project referenced above and determined that the proposed project conforms with applicable regulations and policies. This project must be conducted in accordance with the description and information provided in the application and associated documents.

Note: This project is regulated under the HIPAA Privacy Rule, which requires you to account for certain disclosures of Protected Health Information (PHI).



Michael Geisser
Co-chair, IRBMED

Alan Sugar
Co-chair, IRBMED

Robertson Davenport
Co-chair, IRBMED

CHAPTER 2: BACKGROUND LITERATURE REVIEW

This chapter provides an overview of the history of stroke pathology from ancient to modern times (Section 1), presents the most recent case definition of stroke (Section 2), and discuss the importance of clinical stroke registries, their limitations, and the potential for data linkage to bridge the registries data collection gaps (Section 3). In Section 4, we provide background information on the data sources that will be used in this dissertation. Finally in Section 5, we introduce a summary of published stroke outcomes statistics in the US including recurrence, readmission, mortality, rehabilitation, outpatient visits, and functional recovery. Expanded background and further relevant details of stroke epidemiology will be covered in the introduction sections of Chapters 3, 4 and 5.

The references used to produce the review presented in this chapter were obtained by conducting a PubMed search utilizing the most relevant terminology followed by a citation search of key references related to the topics discussed in this chapter. The majority of the references correspond to research conducted in the US and published in the last 15 years. In the event a relevant peer reviewed references could not be located, other types of references were chosen including, websites, books, and reports.

2.1 Historical Discovery of Brain Vasculature and Stroke Pathology From the Ancient Egyptians to the Early 20th Century

Every year, more than 800,000 people in the United States have a stroke. About 600,000 of these are first or incident strokes.¹ Stroke has been studied extensively in the last fifty years, and most of the therapeutic and diagnosis advancements have taken place after the end of World War II.² However, these advancements could not have occurred without the wealth of knowledge that was discovered in previous historical eras.

The first mention and description of the brain dates back to ancient Egyptian records. A 3,500 BC surgical papyrus called *The Edwin Smith Surgical Papyrus* contained the word brain with description of its coverings.² Stroke was first recognized and described over 2,400 years ago by Hippocrates (460–370 BC), the father of medicine. Back then, stroke was called ‘apoplexy,’ which in Greek means ‘struck down by violence.’ Hippocrates described signs of apoplexy as “unaccustomed attacks of numbness and anesthesia.” Those signs can be interpreted in modern medicine as a transient ischemic attack.² During Hippocrates’s era, little was known about brain anatomy. Apoplexy was described as an accumulation of black bile in the brain arteries, obstructing the passage of animated spirits from the heart ventricles. The heart was considered as the thinking organ while the brain was an organ devoid of blood.³

Galen (131–201 AD), a roman physician, surgeon, and philosopher who worked in Pergamon (present day Bergama, Turkey) was the first to research the vascular anatomy of the brain. He described the apoplectic attack as a sudden, simultaneous, and complete loss of motion and sensation, which includes a sleep-like trouble of consciousness and severe respiratory failure.⁴ Galen was aware that hemiplegia resulted from a lesion in the opposite side of the brain but did not know that hemorrhage was a likely cause of apoplexy.² Galen believed in the humoral theory; that the body contained four important liquids called humors termed phlegm, blood, yellow bile, and black bile. It was believed that these humors must remain in balance for a person to remain healthy; if there was too much of one humor, illness occurred.⁵

After Galen’s era, neuroscience research was blocked by the church for nearly 12 centuries until time of Vesalius- an anatomist and physician in 1543 coinciding with the early renaissance, the era where major neuroanatomical discoveries took place.² Muslims preserved Galen’s work by translating it to Arabic and this was later translated back to European languages

during the 1500s.² Avicenna (980–1037), an Islamic scientist, reconciled Galenic beliefs with the Aristotelian view of the heart as the seat of the mind.⁶

Vesalius, the father of modern anatomy, greatly modified Galens' findings in 1543 by publishing *De humani corporis fabrica* (Latin for *On the Fabric of the Human Body*). In his work, comprised of seven books, he produced drawings that are far more accurate than Galen's descriptions with many corrections to Galen's beliefs.² However, the decisive blow to Galen's humoral theory came in 1628 through William Harvey's work entitled *Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus* (Latin for *An Anatomical Exercise on the Motion of the Heart and Blood in Living Beings*).³ Harvey- an English physician at the Royal College of Physicians was the first to correctly describe in exact detail the circulatory system as a closed system where the blood is pumped around the body by the heart, returning back to it to be recirculated. This description formed the foundation for the recognition of the role of blood vessels in the pathogenesis of stroke.³ Harvey's description was completed after the development of the microscope, when M. Malpighi (1628-1694) discovered the connective capillaries.²

In the mid-1600s J. Wepfer, a physician who worked in Schaffhausen, Switzerland, carried out examinations on the cerebral blood vessels and the brains of patients who had suffered apoplexy.² He found that a few patients who died with apoplexy had bleeding in the brain. He described the vascular anatomy of patients that suffered from stroke in his work entitled *Observationes anatomicae, ex cadaveribus eorum, quos sustulit apoplexia cum exercitatione de ejus loco affect* (Latin for *Anatomical observations from the corpses of those who sustained apoplexy, with a discussion of its localization*). His discovery opened a door to recognize that apoplexy can be caused by bleeding in the brain.³

During the same period Sir Thomas Willis took the work of William Harvey, Vesalius, and J. Wepfer, and added new contributions by describing the vascular arrangement of the base of the brain and making exquisite drawings of the brain vasculature. This work was described in the book entitled: *De Cerebri Anatome* (Latin for *The Anatomy of the Brain*) that was published in 1664.⁷ His work represented the most complete and correct description of the nervous system at that time.⁷ In honor of his work the arterial circle at the base of the brain was named the “circle of Willis”.⁷

In 1761, the Italian physician Morgagni (1682-1771), the father of modern pathology, published his observations that examined the correlation between the anatomical region of apoplexy and the patient’s symptoms in over 700 cases⁸. His work led to the classification of strokes to ischemic (serous apoplexy) and hemorrhagic (sanguineous apoplexy). Morgagni also confirmed that ischemic stroke is caused by blockage to the vessels.² His work led to the foundation of clinical pathology; hence, he was named the father of modern pathology.⁸

Rostan (1790-1866) a French internist who worked at Salpêtrière Hospital in Paris, carried on Morgagni’s work by examining the difference between brain infections and apoplexy. He established a link between the condition of the arteries (ossification) and brain parenchymatous lesions. He was the first one to stop using the term apoplexy.² Later, Virchow (1821–1902), a German pathologist, anthropologist, and statesman, described arterial thrombosis and embolism and recognized the important interaction between blood and the arterial wall⁹. Virchow recognized the consequences of stopping blood flow to an organ or tissue and used the term “ischemia” to denote this process.^{2,9} Further progress in the pathology of stroke included Rokitansky (1804-1878) who developed the link between heart dilation (heart failure) and

strokes. Also, Charcot (1825–1893) described that hemorrhagic stroke can result from a vascular malformation called aneurysm.²

In 1905, Chiari (1851–1916) was among the first to propose that occlusive disease of the extracranial blood vessels (e.g., carotid artery) could be responsible for neurological symptoms. Hunt (1872-1937) called attention to Chiari findings and proposed that the cerebral lesions in most stroke victims could be the effect and not the cause due to embolic materials that could break away from the plaques of carotid arteries.²

The science of stroke diagnosis began in earnest in 1923 when Foix (1882–1927), considered as the first vascular neurologist, investigated the clinical effects of blockage and its relation to the site of blockage.^{2, 10} The huge leap in stroke diagnosis happened in 7 July 1927, when E. Moniz (1874–1955) a Portuguese neurologist that worked at the University of Lisbon reported the first use of cerebral angiography at the Societe´ de Neurologie in Paris using sodium iodide as a contrast medium.¹¹ Moniz’s work introduced the radiological diagnostic measure that could confirm a physician’s clinical diagnosis of stroke.^{2, 11}

2.2 What is a Stroke? Case Definitions of Acute Stroke for Clinical, Research, and Administrative Purposes

In 1970, the World Health Organization defined stroke as “rapidly developed clinical signs of focal (or global) disturbance of cerebral function, lasting more than 24 hours or leading to death, with no apparent cause other than of vascular origin”.¹² This definition however is now considered outdated since stroke nature, timing, clinical recognition, and imaging findings have significantly advanced since 1970.¹² In 2013 the American Heart Association (AHA) and American Stroke Association (ASA) published an updated case definition that refers to stroke as “a central nervous system infarction in the brain, spinal cord, or retinal cell death attributable to

ischemia or hemorrhage, based on pathological, imaging, or other objective evidence of cerebral, spinal cord, or retinal focal ischemic or hemorrhagic injury in a defined vascular distribution, or clinical evidence of cerebral, spinal cord, or retinal focal ischemic or hemorrhagic injury based on symptoms persisting ≥ 24 hours or until death, and other etiologies excluded".¹³ The traditional clinical definition of stroke from the World Health Organization is still included in this revised definition; however, the revised definition includes the silent infarcts that lack clinically overt stroke-like symptoms but demonstrate tissue changes in radiological investigations.¹²⁻¹⁴ It is important to mention that a transient ischemic attack (TIA), a form of stroke, is separately defined by the AHA/ASA as "a transient episode of neurological dysfunction caused by focal brain, spinal cord or retinal ischemia, without acute infarction".^{12, 15} In the event that radiological confirmation of a brain infarct is achieved, a transient ischemic attack will fall under a central nervous system infarct.^{12, 15}

Confirmed clinical diagnosis of stroke have been documented in US health-based databases since October 2015 using the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10) codes.¹⁶ Ischemic strokes are defined using I630 - I639 ICD-10 codes,¹⁷ whereas hemorrhagic strokes are coded to subarachnoid or intracerebral hemorrhage using I600 - I609 and I610 - I619 ICD-10 codes, respectively.¹⁸ Several studies have assessed the validity of identifying strokes using ICD-10 codes,¹⁷⁻²⁴ of which three studies were conducted in the US.^{17, 19, 20} All the studies reported a high positive predictive value (probability of having a relevant stroke ICD-10 code in the event of a confirmed incident stroke diagnosis) between 85.0% and 99.4%. The three US-based studies explicitly examined ischemic strokes and reported a PPV of 93.0%,¹⁹ 97.6%,¹⁷ and 89.0%.²⁰

2.3 Stroke Registries and The Importance of Data Linkage

Clinical registries collect a uniform body of data to evaluate quality of care including specific patient characteristics, diagnostic evaluations, treatments and outcomes for a defined disease, condition, or exposure.^{25, 26} In the last 20 years, the development of national-level²⁷ and state-level²⁸ hospital-based acute stroke registries has generated the data needed to facilitate important improvements in the quality of stroke care,²⁹⁻³² reduced treatment gaps^{33, 34} and disparities in stroke care,²⁹ and contributed to improve patient outcomes for acute stroke patients.^{29, 34-37} The large volume of data collected, which for the national Get With The Guideline-Stroke (GWTG-S) program now exceeds 9 million stroke discharges from more than 2,000 hospitals. These data has enabled a detailed examination of the associations of patient- and hospital-level characteristics with the quality of health care and health outcomes assessed at the point of discharge from the hospital.^{32, 37} However, these studies are limited by the fact that patient outcomes are not collected following hospital discharge. A post discharge medical follow-up is critical to monitor a patient's progress, adjust care activities and medications, and reduce the risk of recurrence and re-hospitalization.^{38, 39} The lack of post discharge data imply that stroke registries cannot achieve one of their central aims which is to improve not just the acute, but also longer-term outcomes for stroke patients through the delivery of high-quality stroke care.

Although patients who survive stroke can continue to improve for many months if not years, the majority of the recovery of function and community participation occur within the first three to six-months following hospital discharge.⁴⁰ The collection of patient-level outcomes data addressing survival, utilization, function, and quality of life has been a challenge for stroke registries because of the substantial resources needed, both human and financial, to follow-up

and interview stroke survivors.⁴¹ A more feasible and sustainable alternative to tracking each patient is to obtain data through linkage to other databases.⁴²⁻⁴⁵ Data linkage between clinical disease registries including stroke registries and other large-scale databases including administrative (claims) data, electronic medical records (eMR), vital records and census data can provide a rich source of patient-level information on outcomes,^{42, 46-48} including post discharge data on mortality,^{36, 49-55} readmissions,^{36, 51, 53-55} stroke recurrence,^{36, 50, 51, 54, 55} use of post-acute care services,⁴⁹ home time,^{51, 54-56} and cost.⁵⁰

Different data sources (i.e., electronic medical records, administrative or claims, registry, and hospital survey data) are designed to serve different purposes; therefore, each source has different strengths and limitations.^{25, 26, 57, 58} Registries collect a uniform body of data to evaluate specific characteristics, management, treatment, and outcomes for a defined disease, condition, or exposure.^{25, 26, 59} Administrative or claims-based data in the US are generated at every encounter with the health care provider including but not limited to physician visits, procedures, hospital or facility admissions, and prescription fillings.⁶⁰ However, claims data includes limited clinical information (e.g. clinical diagnosis and procedures) and are collected solely for billing and insurance purposes.^{58, 60} Hospital surveys such as those administered by the American Hospital Association are annual surveys designed to collect quantitative and qualitative system- and hospital- level information related to operations, utilization, service lines, staffing, system structure, and other data points. Hospital surveys do not collect any patient-level data.⁵⁷ Combining the above three data sources through data linkage can bridge gaps in data limitations from a single data source, providing a richer source of detailed patient-, hospital-, and system-level data. These data sources once linked together can identify associations that would be difficult if not impossible to determine otherwise.^{46, 47, 58}

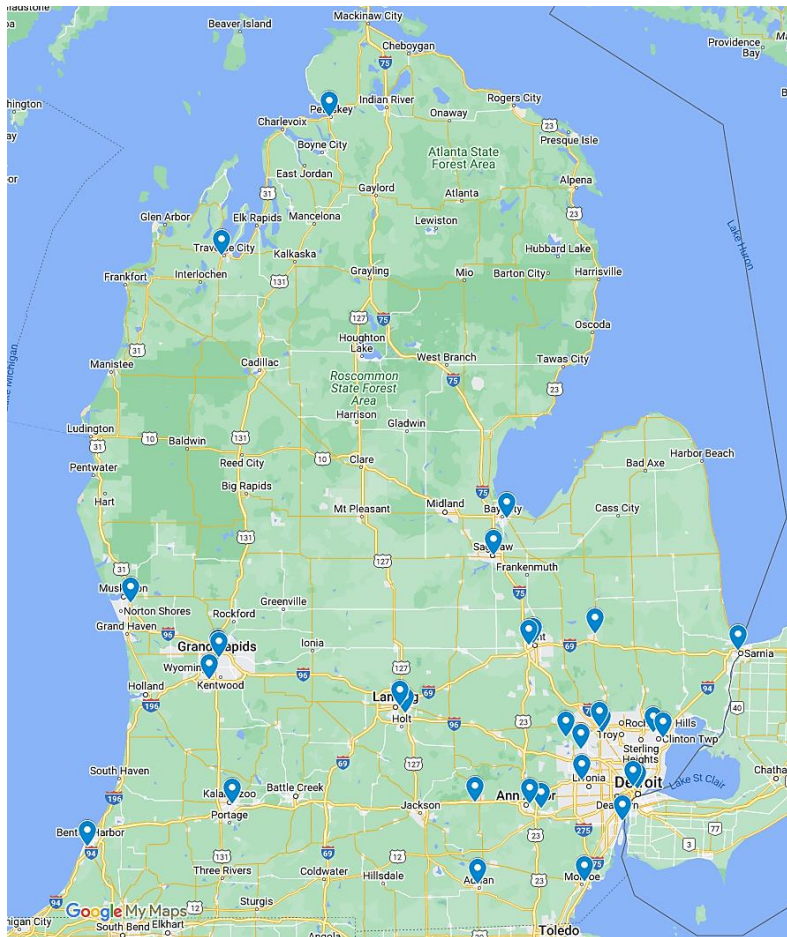
On a national level, post-discharge follow-up data has been obtained by linking the GWTG-S registry to Medicare fee-for-service (FFS) data.^{36, 48, 50-56} Findings from these linkage studies concluded that stroke patients treated at hospitals participating in the GWTG-S program had improved post-discharge functional outcomes and reduced post-discharge mortality and readmissions, compared to patients treated at non-GWTG-S hospitals.³⁶ Other Salient findings from these studies include an increased risk of stroke recurrence and death due to unmet needs of prolonged rehabilitation care and preventative therapies among TIA and minor ischemic stroke patients,⁵⁰ disparities in acute stroke care according to the hospital participation in Medicare Shared Saving Program,⁵¹ and the effect of warfarin and statins treatment effects on reducing cardiovascular events.^{54, 55} While these reports help illustrate the value of data linkage, they are limited to studying outcomes in the Medicare FFS population; thus, there are few studies on stroke patients younger than 65 years or among Medicare Advantage beneficiaries.^{45, 61, 62} Stroke among those aged < 65 (which constitute about 1/3rd of all stroke events⁶³) remain an understudied population despite evidence of increasing rates of stroke in younger adults,⁶⁴ high health-care costs, and loss of labor productivity.^{62, 65} In addition, recently published data by Centers for Medicare and Medicaid Services (CMS) shows that Medicare Advantage population has been growing steadily from 25% in 2010 to 42% in 2020 of Medicare beneficiaries. Therefore, the reliance on only Medicare FFS data for data linkage risks becoming less and less representative of the US population of individuals 65 years of age or older.⁶⁶

2.4 The Michigan Stroke Registry and Michigan Value Collaborative Claims Database

The Michigan Stroke Registry (MiSP) is a representative statewide, hospital-based acute-stroke registry which is part of the Centers of Disease Control and Prevention (CDC) Paul Coverdell National Acute Stroke Program (PCNASP) that has continuously collected data

between 2016-2020 from 31 participating certified stroke centers in Michigan (Figure 2.1). Of the 31 hospitals, 20 were primary stroke centers, 3 were thrombectomy capable stroke centers, and 8 were comprehensive stroke centers. These 31 hospitals include the majority of the 49 certified stroke centers in MI that represents an estimated ~64% of all stroke admissions in the state.^{28, 67} MiSP aims to track and improve stroke care and outcomes through the implementation of quality improvement programs.^{28, 67} Twenty nine additional participating hospitals joined MiSP after 2020. The program was first established in 2001 as the Michigan Acute Stroke Care Overview and Treatment Surveillance System (MASCOTS). Through many years of CDC fundings and collaboration with the American Heart Association (AHA), the Michigan Stroke Registry evolved into Michigan Ongoing Stroke Registry to Accelerate Improvement of Care (MOSAIC).⁶⁷ MOSAIC expanded the scope of its work by developing a statewide comprehensive integrated stroke system of care focused on quality improvement across pre-hospital, in-hospital, and post-hospital settings. In 2021, the registry was renamed as MiSP after renewed CDC funding was secured. The registry expanded its data collection efforts to include data from emergency medical services care and focused evidence-based quality improvement and risk-factor reduction strategies in underserved areas that experience disparities in stroke burden, incidence, and outcomes.⁶⁷ The principle aim of MiSP continues to track and improve stroke care and patient outcomes through the implementation of quality improvement programs.^{28, 67} MiSP identifies stroke discharges using a clinical case definition.⁶⁷ For each discharge detailed clinical data are entered into the American Heart Association's Get-with-the-Guidelines-Stroke (GWTG-S) comprehensive case record form (CRF).⁶⁸

Figure 2.1: Distribution of the 31 Michigan Stroke Program hospitals included in the dissertation research.



The Michigan Value Collaborative (MVC) is a comprehensive, statewide, claims-based database that covers 101 participating hospitals and 40 physician organizations in the state.⁶⁹ The MVC data registry covers 71% of Michigan's 143 hospitals including all the major non-federal acute care hospitals.⁶⁹ MVC contains claims data for Michigan residents insured by Medicare FFS, Medicaid, and all insurance plans covered by Blue Cross Blue Shield of Michigan (BCBSM), including, BCBSM Preferred Provider Organization (PPO), BCBSM Blue Care Network health maintenance organization (HMO), Medicare Advantage PPO and HMO plans. All told, MVC data covers approximately 84% of Michigan's insured population.⁶⁹ The MVC

registry data are organized according to individual episodes of care that begin with an index event and includes all post-discharge claims that occur over the following 90-days.⁶⁹ These index events are grouped into 39 individual medical and surgical conditions (including stroke).⁶⁹ The post discharge facility and professional claims (i.e., readmissions, admission to in-patient rehabilitation facility (IRF), skilled nursing facility (SNF), outpatient visits, emergency department visits, and prescription fillings) enable tracking of healthcare utilization, expenditures, and other patient outcomes over time. Tracking individual subjects in MVC data is done through an assigned unique member ID where all the claims fall under during the follow up period. The member ID remains the same throughout the follow up period unless a change of insurance coverage takes place. Due to restrictions in MVC's data usage agreement, Medicaid data were not available to be used for this study. Mortality data was only available for Medicare FFS beneficiaries.

2.5 Summary of Published Stroke Outcome Statistics in the US

2.5.1 Stroke Recurrence

In the US, recurrent stroke make up almost 25% of the nearly 800,000 estimated acute stroke events every year.⁷⁰ Nationally recurrent ischemic stroke rates among Medicare FFS beneficiaries have been declining between 2001 and 2017 with an adjusted annual decrease of 2.3% (95% CI, 2.2%-2.4%).⁷¹ The 30-day, 90-day, and 1-year ischemic stroke recurrence rates in 2017 among Medicare FFS ischemic stroke population were 2.4%, 4.0%, and 7.6%, respectively.⁷¹

2.5.2 Readmission

Compared to other medical conditions, stroke is associated with high all-cause readmission, ischemic stroke ranks among the top 20 conditions with respect to readmission rates

in the US.⁷²⁻⁷⁵ Published stroke readmission rates in the US vary widely in large part due to different population inclusion criteria including age, payer, single vs multi center, planned vs unplanned readmission, and stroke type.⁷⁶⁻⁸¹ For example, in the US the reported 30-day readmission rates range from 8.9% to 15.4%, 90-day readmission rates range from 13.7% to 19.9%, and 1-year readmission rates range from 27.2% to 48.7%.⁷⁶⁻⁸¹ However, a 2016 meta-analysis of 10 reports (7 of which are US-based) published between 2006 and 2015, estimated the pooled 30-day and 1-year all-cause stroke readmission rates as 17.4% (95% CI, 12.7–23.5%) and 42.5% (95% CI, 34.1–51.3%), respectively.⁷⁶

2.5.3 Mortality

In the US, stroke accounts for approximately one of every 19 deaths, and ranks fifth among all causes of deaths after heart disease, cancer, COVID-19, and accidents.⁷⁰ Nationally, 1-year all-cause mortality post ischemic stroke among Medicare (FFS) beneficiaries increased from 18.7% in 2001 to 21.8% in 2017.⁷¹ Nationally representative data (that include all age groups) on 30-day and 90-day post stroke mortality rates are scarce.^{50, 82} Published reports often include only specific subgroups; therefore, the reported mortality rates tend to be highly variable. For example, a study of Medicare FFS population linked to the GWTG-S registry from 1,471 participating hospitals with minor stroke (non-cardioembolic ischemic stroke with NIHSS ≤ 5 or high-risk TIA) reported 30-day, 90-day and 1-year all-cause mortality of 1.6%, 4.3%, and 11.5%, respectively.⁵⁰ Among Medicare FFS beneficiaries with acute ischemic stroke that were linked to GWTG-S data from 366 participating hospitals, 30-day and 1-year all cause-mortality rates were 15.5% and 28.5%, respectively.³⁶ A study conducted at Kaiser Permanente Northern California among surviving hospitalized ischemic stroke patients with atrial fibrillation reported that 30-day

mortality and 1-year was 24.7% and 40.1%, respectively.⁸³ We could not explain the substantially high mortality rates in the Kaiser study compared to the previous two studies.

2.5.4 Rehabilitation

To promote recovery following stroke approximately two thirds of stroke survivors receive post-acute care after hospital discharge that typically includes rehabilitation care.^{84, 85} Across the US- about 19% of Medicare FFS stroke patients are discharged to inpatient rehabilitation facility (IRF), 25% to skilled nursing facility (SNF), and another 12% receive home health (HH) care.^{70, 86} Data from the GWTG-S registry that included patients from all age groups reported that 25.4%, 19.5%, 11.5% were discharged to IRF, SNF, and HH care services, respectively.⁸⁵ IRFs provide intensive, interdisciplinary rehabilitation care under the direct supervision of a physician,⁴⁴ whereas SNFs provide less intensive rehabilitation care to stroke survivors who need both nursing or rehabilitation care.⁴⁴

Although a 2015 report by the CDC on the use of outpatient rehabilitation among stroke survivors found that around a third of stroke discharges participated in outpatient rehabilitation or utilization of home health rehabilitation services, data on outpatient rehabilitation following stroke in the US is limited and difficult to interpret given the conflation of home- and office-based rehabilitation care settings.^{87, 88}

2.5.5 Post Acute Care Outpatient Visit Follow Up

Follow up visits to a generalist and a specialist physician after hospital discharge are crucial for secondary stroke prevention.⁸⁹ However, data from Medicare FFS stroke survivors report long delays in obtaining post-hospital follow up visits with neurology specialists or primary care.⁹⁰ A retrospective cohort study of all Medicare FFS patients discharged home after an acute ischemic stroke reported that 61% and 16% of patients had a primary care and

neurology visit within 30 days of discharge, respectively.⁹¹ A study conducted at a single primary stroke center with 416 ischemic stroke patients in Pennsylvania in 2013 found that only 47% had a neurology follow-up visit within 21 days post discharge, and 36.3% never had any follow-up in 2.5 years.⁹² A nationally representative claims database of commercially insured Americans aged between 18 and 89 years old and discharged with stroke from 2009 to 2015 reported that 59.3% and 70.8% had a primary care visit within 30- and 90-days post discharge, respectively, and 24.4% and 41.8% had a neurology visit.⁸¹ The outpatient utilization rates presented by the earlier studies^{81, 90-92} indicate a low overall utilization rate of post-acute primary and specialized follow up services but these rates also are likely affected by a myriad of factors including but not limited to demographics, inpatient care, discharge plan, and social factors. Therefore, there is a need to investigate the drivers that affect patient follow up post discharge using comprehensive data sources.

2.5.6 Functional Recovery

More than 47 different instruments or scales have been developed and/or used to measure the functional recovery of patients post stroke, but the most commonly used scales include Barthel Index (BI), modified Rankin Scale (mRS), activities of daily living (ADL), as well as various instruments that measure quality of life (QOL).⁹³ The modified Rankin Scale at 90-days post discharge is widely accepted as the standard measure to assess functional recovery after stroke especially in clinical trials.⁹⁴⁻⁹⁷ This ordinal clinician-rated scale grades patient's global disability from 0 (no deficit) to 6 (death).^{96, 97} For this scale, a single-point reduction in mRS scores is considered as a clinically meaningful change in functional recovery.^{96, 97} A secondary analysis of a large clinical trial conducted at 60 stroke-receiving hospitals in two large California counties of patients with acute ischemic stroke and intracranial hemorrhage that examined the

effect of intravenous magnesium sulfate versus placebo beginning within two hours after symptoms onset compared the functional recovery between day 4 and day 90 post discharge and reported that mRS improved (decreased by 1 or more grades) in 72.6% and 77.3% of acute ischemic and hemorrhagic stroke patients, respectively.⁹⁸ In addition, acute ischemic stroke patients mRS scores improved from a mean of 4.17 (± 0.7) to 2.84 (± 1.5) and intracranial hemorrhage patients improved from a mean of 4.35 (± 0.7) to 2.75 (± 1.3).⁹⁸

Conducting population-based studies to assess post stroke recovery is very difficult in the US because obtaining functional recovery measures relies on individual patient follow-up data which is costly and hard to achieve post discharge (e.g., 90-days).⁹⁴ A practical solution for this limitation is through calculation of home time, a validated outcome measure of functional recovery post stroke using claims data.^{99, 100} Home time is defined as the number of days in the year post discharge that is spent alive and out of an inpatient care setting including hospital, inpatient rehabilitation, skilled nursing facility, and long term care hospital.⁵⁶ Two GWTG-S studies that used linked Medicare FFS data reported a median 90-day and 1-year post ischemic stroke home time of 59.5 and 79.0 days and 270.2 and 349.0 days, respectively.^{56, 101}

BIBLIOGRAPHY

1. Virani, S. S.; Alonso, A.; Benjamin, E. J.; Bittencourt, M. S.; Callaway, C. W.; Carson, A. P.; Chamberlain, A. M.; Chang, A. R.; Cheng, S.; Delling, F. N.; Djousse, L.; Elkind, M. S. V.; Ferguson, J. F.; Fornage, M.; Khan, S. S.; Kissela, B. M.; Knutson, K. L.; Kwan, T. W.; Lackland, D. T.; Lewis, T. T.; Lichtman, J. H.; Longenecker, C. T.; Loop, M. S.; Lutsey, P. L.; Martin, S. S.; Matsushita, K.; Moran, A. E.; Mussolino, M. E.; Perak, A. M.; Rosamond, W. D.; Roth, G. A.; Sampson, U. K. A.; Satou, G. M.; Schroeder, E. B.; Shah, S. H.; Shay, C. M.; Spartano, N. L.; Stokes, A.; Tirschwell, D. L.; VanWagner, L. B.; Tsao, C. W.; American Heart Association Council on, E.; Prevention Statistics, C.; Stroke Statistics, S., Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. *Circulation* **2020**, *141* (9), e139-e596. <https://doi.org/10.1161/CIR.0000000000000757>.
2. Paciaroni, M.; Bogousslavsky, J., The history of stroke and cerebrovascular disease. *Handb Clin Neurol* **2009**, *92*, 3-28. [https://doi.org/10.1016/S0072-9752\(08\)01901-5](https://doi.org/10.1016/S0072-9752(08)01901-5).
3. Tatu, L.; Moulin, T.; Monnier, G., The discovery of encephalic arteries. From Johann Jacob Wepfer to Charles Foix. *Cerebrovasc Dis* **2005**, *20* (6), 427-32. <https://doi.org/10.1159/000088980>.
4. Karenberg, A.; Hort, I., Medieval descriptions and doctrines of stroke: preliminary analysis of select sources. Part I: The struggle for terms and theories - late antiquity and early Middle Ages. *J Hist Neurosci* **1998**, *7* (113 Pt 2), 162-73. <https://doi.org/10.1076/jhin.7.3.162.1849>.
5. Hajar, R., Medicine from Galen to the Present: A Short History. *Heart Views* **2021**, *22* (4), 307-308. https://doi.org/10.4103/heartviews.heartviews_125_21.
6. Karenberg, A.; Hort, I., Medieval descriptions and doctrines of stroke: preliminary analysis of select sources. Part II: between Galenism and Aristotelism - Islamic theories of apoplexy (800-1200). *J Hist Neurosci* **1998**, *7* (3), 174-85. <https://doi.org/10.1076/jhin.7.3.174.1858>.
7. Dumitrescu, A. M.; Costea, C. F.; Cucu, A. I.; Dumitrescu, G. F.; Turliuc, M. D.; Scripcariu, D. V.; Ciocoiu, M.; Tanase, D. M.; Turliuc, S.; Bogdanici, C. M.; Nicoara, S. D.; Carauleanu, A.; Schmitzer, S.; Sava, A., The discovery of the circle of Willis as a result of using the scientific method in anatomical dissection. *Rom J Morphol Embryol* **2020**, *61* (3), 959-965. <https://doi.org/10.47162/RJME.61.3.38>.
8. Tedeschi, C. G., Giovanni Battista MORGAGNI, the founder of pathologic anatomy. A biographic sketh, on the occasion of the 200th anniversary of the publication of his "De sedibus et causis morborum per anatomen indagatis". *BMQ* **1961**, *12*, 112-25.
9. Safavi-Abbasi, S.; Reis, C.; Talley, M. C.; Theodore, N.; Nakaji, P.; Spetzler, R. F.; Preul, M. C., Rudolf Ludwig Karl Virchow: pathologist, physician, anthropologist, and politician.

Implications of his work for the understanding of cerebrovascular pathology and stroke. *Neurosurg Focus* **2006**, *20* (6), E1. <https://doi.org/10.3171/foc.2006.20.6.1>.

10. Caplan, L. R., Charles Foix--the first modern stroke neurologist. *Stroke* **1990**, *21* (2), 348-56. <https://doi.org/10.1161/01.str.21.2.348>.

11. Artico, M.; Spoletini, M.; Fumagalli, L.; Biagioni, F.; Ryskalin, L.; Fornai, F.; Salvati, M.; Frati, A.; Pastore, F. S.; Taurone, S., Egas Moniz: 90 Years (1927-2017) from Cerebral Angiography. *Front Neuroanat* **2017**, *11*, 81. <https://doi.org/10.3389/fnana.2017.00081>.

12. Coupland, A. P.; Thapar, A.; Qureshi, M. I.; Jenkins, H.; Davies, A. H., The definition of stroke. *J R Soc Med* **2017**, *110* (1), 9-12. <https://doi.org/10.1177/0141076816680121>.

13. Sacco, R. L.; Kasner, S. E.; Broderick, J. P.; Caplan, L. R.; Connors, J. J.; Culebras, A.; Elkind, M. S.; George, M. G.; Hamdan, A. D.; Higashida, R. T.; Hoh, B. L.; Janis, L. S.; Kase, C. S.; Kleindorfer, D. O.; Lee, J. M.; Moseley, M. E.; Peterson, E. D.; Turan, T. N.; Valderrama, A. L.; Vinters, H. V.; American Heart Association Stroke Council, C. o. C. S.; Anesthesia; Council on Cardiovascular, R.; Intervention; Council on, C.; Stroke, N.; Council on, E.; Prevention; Council on Peripheral Vascular, D.; Council on Nutrition, P. A.; Metabolism, An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* **2013**, *44* (7), 2064-89. <https://doi.org/10.1161/STR.0b013e318296aeca>.

14. Vermeer, S. E.; Longstreth, W. T., Jr.; Koudstaal, P. J., Silent brain infarcts: a systematic review. *Lancet Neurol* **2007**, *6* (7), 611-9. [https://doi.org/10.1016/S1474-4422\(07\)70170-9](https://doi.org/10.1016/S1474-4422(07)70170-9).

15. Easton, J. D.; Saver, J. L.; Albers, G. W.; Alberts, M. J.; Chaturvedi, S.; Feldmann, E.; Hatsukami, T. S.; Higashida, R. T.; Johnston, S. C.; Kidwell, C. S.; Lutsep, H. L.; Miller, E.; Sacco, R. L.; American Heart, A.; American Stroke Association Stroke, C.; Council on Cardiovascular, S.; Anesthesia; Council on Cardiovascular, R.; Intervention; Council on Cardiovascular, N.; Interdisciplinary Council on Peripheral Vascular, D., Definition and evaluation of transient ischemic attack: a scientific statement for healthcare professionals from the American Heart Association/American Stroke Association Stroke Council; Council on Cardiovascular Surgery and Anesthesia; Council on Cardiovascular Radiology and Intervention; Council on Cardiovascular Nursing; and the Interdisciplinary Council on Peripheral Vascular Disease. The American Academy of Neurology affirms the value of this statement as an educational tool for neurologists. *Stroke* **2009**, *40* (6), 2276-93. <https://doi.org/10.1161/STROKEAHA.108.192218>.

16. Hirsch, J. A.; Nicola, G.; McGinty, G.; Liu, R. W.; Barr, R. M.; Chittle, M. D.; Manchikanti, L., ICD-10: History and Context. *AJNR Am J Neuroradiol* **2016**, *37* (4), 596-9. <https://doi.org/10.3174/ajnr.A4696>.

17. Alhajji, M.; Kawsara, A.; Alkhouli, M., Validation of Acute Ischemic Stroke Codes Using the International Classification of Diseases Tenth Revision. *Am J Cardiol* **2020**, *125* (7), 1135. <https://doi.org/10.1016/j.amjcard.2020.01.004>.
18. Hsieh, M. T.; Huang, K. C.; Hsieh, C. Y.; Tsai, T. T.; Chen, L. C.; Sung, S. F., Validation of ICD-10-CM Diagnosis Codes for Identification of Patients with Acute Hemorrhagic Stroke in a National Health Insurance Claims Database. *Clin Epidemiol* **2021**, *13*, 43-51. <https://doi.org/10.2147/CLEP.S288518>.
19. Shirley, A. M.; Morrisette, K. L.; Choi, S. K.; Reynolds, K.; Zhou, H.; Zhou, M. M.; Wei, R.; Zhang, Y.; Cheng, P.; Wong, E.; Sangha, N.; An, J., Validation of ICD-10 hospital discharge diagnosis codes to identify incident and recurrent ischemic stroke from a US integrated healthcare system. *Pharmacoepidemiol Drug Saf* **2023**, *32* (12), 1439-1445. <https://doi.org/10.1002/pds.5675>.
20. Hirsch, J. L.; Burke, J. F.; Kerber, K. A., Validation of Vascular Location Subcodes for Acute Ischemic Stroke by the International Classification of Diseases-10. *J Stroke Cerebrovasc Dis* **2024**, *33* (4), 107590. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2024.107590>.
21. Kokotailo, R. A.; Hill, M. D., Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. *Stroke* **2005**, *36* (8), 1776-81. <https://doi.org/10.1161/01.STR.0000174293.17959.a1>.
22. Sedova, P.; Brown, R. D., Jr.; Zvolosky, M.; Kadlecova, P.; Bryndziar, T.; Volny, O.; Weiss, V.; Bednarik, J.; Mikulik, R., Validation of Stroke Diagnosis in the National Registry of Hospitalized Patients in the Czech Republic. *J Stroke Cerebrovasc Dis* **2015**, *24* (9), 2032-8. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2015.04.019>.
23. Hsieh, M. T.; Hsieh, C. Y.; Tsai, T. T.; Wang, Y. C.; Sung, S. F., Performance of ICD-10-CM Diagnosis Codes for Identifying Acute Ischemic Stroke in a National Health Insurance Claims Database. *Clin Epidemiol* **2020**, *12*, 1007-1013. <https://doi.org/10.2147/CLEP.S273853>.
24. McCormick, N.; Bhole, V.; Lacaille, D.; Avina-Zubieta, J. A., Validity of Diagnostic Codes for Acute Stroke in Administrative Databases: A Systematic Review. *PLoS One* **2015**, *10* (8), e0135834. <https://doi.org/10.1371/journal.pone.0135834>.
25. In *Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition, Addendum 2*, Gliklich, R. E.; Leavy, M. B.; Dreyer, N. A., Eds. Rockville (MD), 2019.
26. In *Registries for Evaluating Patient Outcomes: A User's Guide*, 3rd ed.; Gliklich, R. E.; Dreyer, N. A.; Leavy, M. B., Eds. Rockville (MD), 2014.

27. American Heart Association, Get With The Guide Line - Stroke. <https://www.heart.org/en/professional/quality-improvement/get-with-the-guidelines/get-with-the-guidelines-stroke/get-with-the-guidelines-stroke-overview> (accessed 2023).
28. Center of Disease Control and Prevention, Paul Coverdell National Acute Stroke Program. https://www.cdc.gov/dhdsp/programs/stroke_registry.htm (accessed 2023).
29. Schwamm, L. H.; Reeves, M. J.; Pan, W.; Smith, E. E.; Frankel, M. R.; Olson, D.; Zhao, X.; Peterson, E.; Fonarow, G. C., Race/ethnicity, quality of care, and outcomes in ischemic stroke. *Circulation* **2010**, *121* (13), 1492-501. <https://doi.org/10.1161/CIRCULATIONAHA.109.881490>.
30. George, M. G.; Tong, X.; McGruder, H.; Yoon, P.; Rosamond, W.; Winquist, A.; Hinchey, J.; Wall, H. K.; Pandey, D. K.; Centers for Disease, C.; Prevention, Paul Coverdell National Acute Stroke Registry Surveillance - four states, 2005-2007. *MMWR Surveill Summ* **2009**, *58* (7), 1-23.
31. Parker, C.; Schwamm, L. H.; Fonarow, G. C.; Smith, E. E.; Reeves, M. J., Stroke quality metrics: systematic reviews of the relationships to patient-centered outcomes and impact of public reporting. *Stroke* **2012**, *43* (1), 155-62. <https://doi.org/10.1161/STROKEAHA.111.635011>.
32. Howard, G.; Schwamm, L. H.; Donnelly, J. P.; Howard, V. J.; Jasne, A.; Smith, E. E.; Rhodes, J. D.; Kissela, B. M.; Fonarow, G. C.; Kleindorfer, D. O.; Albright, K. C., Participation in Get With The Guidelines-Stroke and Its Association With Quality of Care for Stroke. *JAMA Neurol* **2018**, *75* (11), 1331-1337. <https://doi.org/10.1001/jamaneurol.2018.2101>.
33. Fonarow, G. C.; Smith, E. E.; Saver, J. L.; Reeves, M. J.; Hernandez, A. F.; Peterson, E. D.; Sacco, R. L.; Schwamm, L. H., Improving door-to-needle times in acute ischemic stroke: the design and rationale for the American Heart Association/American Stroke Association's Target: Stroke initiative. *Stroke* **2011**, *42* (10), 2983-9. <https://doi.org/10.1161/STROKEAHA.111.621342>.
34. Heidenreich, P. A.; Hernandez, A. F.; Yancy, C. W.; Liang, L.; Peterson, E. D.; Fonarow, G. C., Get With The Guidelines program participation, process of care, and outcome for Medicare patients hospitalized with heart failure. *Circ Cardiovasc Qual Outcomes* **2012**, *5* (1), 37-43. <https://doi.org/10.1161/CIRCOUTCOMES.110.959122>.
35. American Heart Association, Get With The Guidelines® - Stroke Patient Management Tool. <https://www.heart.org/en/professional/quality-improvement/get-with-the-guidelines/get-with-the-guidelines-stroke/get-with-the-guidelines-stroke-patient-management-tool>.
36. Song, S.; Fonarow, G. C.; Olson, D. M.; Liang, L.; Schulte, P. J.; Hernandez, A. F.; Peterson, E. D.; Reeves, M. J.; Smith, E. E.; Schwamm, L. H.; Saver, J. L., Association of Get With The Guidelines-Stroke Program Participation and Clinical Outcomes for Medicare

Beneficiaries With Ischemic Stroke. *Stroke* **2016**, 47 (5), 1294-302. <https://doi.org/10.1161/STROKEAHA.115.011874>.

37. Ormseth, C. H.; Sheth, K. N.; Saver, J. L.; Fonarow, G. C.; Schwamm, L. H., The American Heart Association's Get With the Guidelines (GWTG)-Stroke development and impact on stroke care. *Stroke Vasc Neurol* **2017**, 2 (2), 94-105. <https://doi.org/10.1136/svn-2017-000092>.

38. In *Emergency and acute medical care in over 16s: service delivery and organisation*, London, 2018.

39. Coppa, K.; Kim, E. J.; Oppenheim, M. I.; Bock, K. R.; Conigliaro, J.; Hirsch, J. S., Examination of Post-discharge Follow-up Appointment Status and 30-Day Readmission. *J Gen Intern Med* **2021**, 36 (5), 1214-1221. <https://doi.org/10.1007/s11606-020-06569-5>.

40. Lee, K. B.; Lim, S. H.; Kim, K. H.; Kim, K. J.; Kim, Y. R.; Chang, W. N.; Yeom, J. W.; Kim, Y. D.; Hwang, B. Y., Six-month functional recovery of stroke patients: a multi-time-point study. *Int J Rehabil Res* **2015**, 38 (2), 173-80. <https://doi.org/10.1097/MRR.0000000000000108>.

41. Reeves, M.; Lisabeth, L.; Williams, L.; Katzan, I.; Kapral, M.; Deutsch, A.; Prvu-Bettger, J., Patient-Reported Outcome Measures (PROMs) for Acute Stroke: Rationale, Methods and Future Directions. *Stroke* **2018**, 49 (6), 1549-1556. <https://doi.org/10.1161/STROKEAHA.117.018912>.

42. Yu, A. Y.; Holodinsky, J. K.; Zerna, C.; Svenson, L. W.; Jette, N.; Quan, H.; Hill, M. D., Use and Utility of Administrative Health Data for Stroke Research and Surveillance. *Stroke* **2016**, 47 (7), 1946-52. <https://doi.org/10.1161/STROKEAHA.116.012390>.

43. Reker, D. M.; Reid, K.; Duncan, P. W.; Marshall, C.; Cowper, D.; Stansbury, J.; Warr-Wing, K. L., Development of an integrated stroke outcomes database within Veterans Health Administration. *J Rehabil Res Dev* **2005**, 42 (1), 77-91. <https://doi.org/10.1682/jrrd.2003.11.0164>.

44. Deutsch, A.; Granger, C. V.; Heinemann, A. W.; Fiedler, R. C.; DeJong, G.; Kane, R. L.; Ottenbacher, K. J.; Naughton, J. P.; Trevisan, M., Poststroke rehabilitation: outcomes and reimbursement of inpatient rehabilitation facilities and subacute rehabilitation programs. *Stroke* **2006**, 37 (6), 1477-82. <https://doi.org/10.1161/01.STR.0000221172.99375.5a>.

45. Patorno, E.; Schneeweiss, S.; George, M. G.; Tong, X.; Franklin, J. M.; Pawar, A.; Mogun, H.; Moura, L.; Schwamm, L. H., Linking the Paul Coverdell National Acute Stroke Program to commercial claims to establish a framework for real-world longitudinal stroke research. *Stroke Vasc Neurol* **2022**, 7 (2), 114-123. <https://doi.org/10.1136/svn-2021-001134>.

46. Bradley, C. J.; Penberthy, L.; Devers, K. J.; Holden, D. J., Health services research and data linkages: issues, methods, and directions for the future. *Health Serv Res* **2010**, *45* (5 Pt 2), 1468-88. <https://doi.org/10.1111/j.1475-6773.2010.01142.x>.
47. Dusetzina, S. B.; Tyree, S.; Meyer, A. M.; Meyer, A.; Green, L.; Carpenter, W. R., In *Linking Data for Health Services Research: A Framework and Instructional Guide*, Rockville (MD), 2014.
48. Reeves, M. J.; Fonarow, G. C.; Smith, E. E.; Pan, W.; Olson, D.; Hernandez, A. F.; Peterson, E. D.; Schwamm, L. H., Representativeness of the Get With The Guidelines-Stroke Registry: comparison of patient and hospital characteristics among Medicare beneficiaries hospitalized with ischemic stroke. *Stroke* **2012**, *43* (1), 44-9. <https://doi.org/10.1161/STROKEAHA.111.626978>.
49. Cadilhac, D. A.; Kim, J.; Lannin, N. A.; Kapral, M. K.; Schwamm, L. H.; Dennis, M. S.; Norrving, B.; Meretoja, A., National stroke registries for monitoring and improving the quality of hospital care: A systematic review. *Int J Stroke* **2016**, *11* (1), 28-40. <https://doi.org/10.1177/1747493015607523>.
50. Kaufman, B. G.; Shah, S.; Hellkamp, A. S.; Lytle, B. L.; Fonarow, G. C.; Schwamm, L. H.; Lesen, E.; Hedberg, J.; Tank, A.; Fita, E.; Bhalla, N.; Atreja, N.; Bettger, J. P., Disease Burden Following Non-Cardioembolic Minor Ischemic Stroke or High-Risk TIA: A GWTG-Stroke Study. *J Stroke Cerebrovasc Dis* **2020**, *29* (12), 105399. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105399>.
51. Kaufman, B. G.; O'Brien, E. C.; Stearns, S. C.; Matsouaka, R.; Holmes, G. M.; Weinberger, M.; Song, P. H.; Schwamm, L. H.; Smith, E. E.; Fonarow, G. C.; Xian, Y., The Medicare Shared Savings Program and Outcomes for Ischemic Stroke Patients: a Retrospective Cohort Study. *J Gen Intern Med* **2019**, *34* (12), 2740-2748. <https://doi.org/10.1007/s11606-019-05283-1>.
52. Reeves, M. J.; Fonarow, G. C.; Xu, H.; Matsouaka, R. A.; Xian, Y.; Saver, J.; Schwamm, L.; Smith, E. E., Is Risk-Standardized In-Hospital Stroke Mortality an Adequate Proxy for Risk-Standardized 30-Day Stroke Mortality Data? Findings From Get With The Guidelines-Stroke. *Circ Cardiovasc Qual Outcomes* **2017**, *10* (10). <https://doi.org/10.1161/CIRCOUTCOMES.117.003748>.
53. Fonarow, G. C.; Smith, E. E.; Reeves, M. J.; Pan, W.; Olson, D.; Hernandez, A. F.; Peterson, E. D.; Schwamm, L. H.; Get With The Guidelines Steering, C.; Hospitals, Hospital-level variation in mortality and rehospitalization for medicare beneficiaries with acute ischemic stroke. *Stroke* **2011**, *42* (1), 159-66. <https://doi.org/10.1161/STROKEAHA.110.601831>.
54. Xian, Y.; Wu, J.; O'Brien, E. C.; Fonarow, G. C.; Olson, D. M.; Schwamm, L. H.; Bhatt, D. L.; Smith, E. E.; Suter, R. E.; Hannah, D.; Lindholm, B.; Maisch, L.; Greiner, M. A.; Lytle,

- B. L.; Pencina, M. J.; Peterson, E. D.; Hernandez, A. F., Real world effectiveness of warfarin among ischemic stroke patients with atrial fibrillation: observational analysis from Patient-Centered Research into Outcomes Stroke Patients Prefer and Effectiveness Research (PROSPER) study. *BMJ* **2015**, *351*, h3786. <https://doi.org/10.1136/bmj.h3786>.
55. O'Brien, E. C.; Greiner, M. A.; Xian, Y.; Fonarow, G. C.; Olson, D. M.; Schwamm, L. H.; Bhatt, D. L.; Smith, E. E.; Maisch, L.; Hannah, D.; Lindholm, B.; Peterson, E. D.; Pencina, M. J.; Hernandez, A. F., Clinical Effectiveness of Statin Therapy After Ischemic Stroke: Primary Results From the Statin Therapeutic Area of the Patient-Centered Research Into Outcomes Stroke Patients Prefer and Effectiveness Research (PROSPER) Study. *Circulation* **2015**, *132* (15), 1404-13. <https://doi.org/10.1161/CIRCULATIONAHA.115.016183>.
56. Fonarow, G. C.; Liang, L.; Thomas, L.; Xian, Y.; Saver, J. L.; Smith, E. E.; Schwamm, L. H.; Peterson, E. D.; Hernandez, A. F.; Duncan, P. W.; O'Brien, E. C.; Bushnell, C.; Prvu Bettger, J., Assessment of Home-Time After Acute Ischemic Stroke in Medicare Beneficiaries. *Stroke* **2016**, *47* (3), 836-42. <https://doi.org/10.1161/STROKEAHA.115.011599>.
57. American Hospital Association, AHA Annual Survey Database. <https://www.ahadata.com/> (accessed 2023).
58. Cadarette, S. M.; Wong, L., An Introduction to Health Care Administrative Data. *Can J Hosp Pharm* **2015**, *68* (3), 232-7. <https://doi.org/10.4212/cjhp.v68i3.1457>.
59. In *Registries for Evaluating Patient Outcomes: A User's Guide*, 4th ed.; Gliklich, R. E.; Leavy, M. B.; Dreyer, N. A., Eds. Rockville (MD), 2020.
60. Mazzali, C.; Duca, P., Use of administrative data in healthcare research. *Intern Emerg Med* **2015**, *10* (4), 517-24. <https://doi.org/10.1007/s11739-015-1213-9>.
61. Lichtman, J. H.; Leifheit-Limson, E. C.; Goldstein, L. B., Centers for medicare and medicaid services medicare data and stroke research: goldmine or landmine? *Stroke* **2015**, *46* (2), 598-604. <https://doi.org/10.1161/STROKEAHA.114.003255>.
62. Ekker, M. S.; Boot, E. M.; Singhal, A. B.; Tan, K. S.; Debette, S.; Tuladhar, A. M.; de Leeuw, F. E., Epidemiology, aetiology, and management of ischaemic stroke in young adults. *Lancet Neurol* **2018**, *17* (9), 790-801. [https://doi.org/10.1016/S1474-4422\(18\)30233-3](https://doi.org/10.1016/S1474-4422(18)30233-3).
63. Hall, M. J.; Levant, S.; DeFrances, C. J., Hospitalization for stroke in U.S. hospitals, 1989-2009. *NCHS Data Brief* **2012**, (95), 1-8.
64. Bejot, 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* (5). <https://doi.org/10.1161/JAHA.116.003661>.

65. Maaijwee, N. A.; Rutten-Jacobs, L. C.; Arntz, R. M.; Schaapsmeeders, P.; Schoonderwaldt, H. C.; van Dijk, E. J.; de Leeuw, F. E., Long-term increased risk of unemployment after young stroke: a long-term follow-up study. *Neurology* **2014**, *83* (13), 1132-8. <https://doi.org/10.1212/WNL.0000000000000817>.
66. Services, C. f. M. a. M. Medicare and Medicaid enrollment reports. <https://data.cms.gov/summary-statistics-on-beneficiary-enrollment/medicare-and-medicaid-reports/medicare-monthly-enrollment>.
67. Michigan Department of Health and Human Services (MDHHS), Michigan Stroke Program (MiSP). <https://www.michigan.gov/mdhhs/keep-mi-healthy/communicablediseases/epidemiology/chronicepi/stroke> (accessed 2023).
68. American Heart Association, Get With The Guidelines® - Stroke Case Record Form. <https://www.heart.org/-/media/Files/Professional/Quality-Improvement/Get-With-the-Guidelines/Get-With-The-Guidelines-Stroke/Stroke--Diabetes-CRFJuly21.pdf>.
69. Michigan Value Collaborative, MVC Data Resources. <https://michiganvalue.org/resources-2/> (accessed 2023).
70. Tsao, C. W.; Aday, A. W.; Almarzooq, Z. I.; Anderson, C. A. M.; Arora, P.; Avery, C. L.; Baker-Smith, C. M.; Beaton, A. Z.; Boehme, A. K.; Buxton, A. E.; Commodore-Mensah, Y.; Elkind, M. S. V.; Evenson, K. R.; Eze-Nliam, C.; Fugar, S.; Generoso, G.; Heard, D. G.; Hiremath, S.; Ho, J. E.; Kalani, R.; Kazi, D. S.; Ko, D.; Levine, D. A.; Liu, J.; Ma, J.; Magnani, J. W.; Michos, E. D.; Mussolino, M. E.; Navaneethan, S. D.; Parikh, N. I.; Poudel, R.; Rezk-Hanna, M.; Roth, G. A.; Shah, N. S.; St-Onge, M. P.; Thacker, E. L.; Virani, S. S.; Voeks, J. H.; Wang, N. Y.; Wong, N. D.; Wong, S. S.; Yaffe, K.; Martin, S. S.; American Heart Association Council on, E.; Prevention Statistics, C.; Stroke Statistics, S., Heart Disease and Stroke Statistics-2023 Update: A Report From the American Heart Association. *Circulation* **2023**, *147* (8), e93-e621. <https://doi.org/10.1161/CIR.0000000000001123>.
71. Leifheit, E. C.; Wang, Y.; Goldstein, L. B.; Lichtman, J. H., Trends in 1-Year Recurrent Ischemic Stroke in the US Medicare Fee-for-Service Population. *Stroke* **2022**, *53* (11), 3338-3347. <https://doi.org/10.1161/STROKEAHA.122.039438>.
72. Weiss, A. J.; Jiang, H. J., Overview of Clinical Conditions With Frequent and Costly Hospital Readmissions by Payer, 2018. In *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*, Rockville (MD), 2021.
73. Bambhroliya, A. B.; Donnelly, J. P.; Thomas, E. J.; Tyson, J. E.; Miller, C. C.; McCullough, L. D.; Savitz, S. I.; Vahidy, F. S., Estimates and Temporal Trend for US Nationwide 30-Day Hospital Readmission Among Patients With Ischemic and Hemorrhagic Stroke. *JAMA Netw Open* **2018**, *1* (4), e181190. <https://doi.org/10.1001/jamanetworkopen.2018.1190>.

74. Bayliss, W. S.; Bushnell, C. D.; Halladay, J. R.; Duncan, P. W.; Freburger, J. K.; Kucharska-Newton, A. M.; Trogdon, J. G., The Cost of Implementing and Sustaining the COMprehensive Post-Acute Stroke Services Model. *Med Care* **2021**, *59* (2), 163-168. <https://doi.org/10.1097/MLR.0000000000001462>.
75. Darabi, N.; Hosseinichimeh, N.; Noto, A.; Zand, R.; Abedi, V., Machine Learning-Enabled 30-Day Readmission Model for Stroke Patients. *Front Neurol* **2021**, *12*, 638267. <https://doi.org/10.3389/fneur.2021.638267>.
76. Zhong, W.; Geng, N.; Wang, P.; Li, Z.; Cao, L., Prevalence, causes and risk factors of hospital readmissions after acute stroke and transient ischemic attack: a systematic review and meta-analysis. *Neurol Sci* **2016**, *37* (8), 1195-202. <https://doi.org/10.1007/s10072-016-2570-5>.
77. Zhou, L. W.; Lansberg, M. G.; de Havenon, A., Rates and reasons for hospital readmission after acute ischemic stroke in a US population-based cohort. *PLoS One* **2023**, *18* (8), e0289640. <https://doi.org/10.1371/journal.pone.0289640>.
78. Vahidy, F. S.; Donnelly, J. P.; McCullough, L. D.; Tyson, J. E.; Miller, C. C.; Boehme, A. K.; Savitz, S. I.; Albright, K. C., Nationwide Estimates of 30-Day Readmission in Patients With Ischemic Stroke. *Stroke* **2017**, *48* (5), 1386-1388. <https://doi.org/10.1161/STROKEAHA.116.016085>.
79. Lichtman, J. H.; Leifheit-Limson, E. C.; Jones, S. B.; Wang, Y.; Goldstein, L. B., Preventable readmissions within 30 days of ischemic stroke among Medicare beneficiaries. *Stroke* **2013**, *44* (12), 3429-35. <https://doi.org/10.1161/STROKEAHA.113.003165>.
80. Bushnell, C. D.; Kucharska-Newton, A. M.; Jones, S. B.; Psioda, M. A.; Johnson, A. M.; Daras, L. C.; Halladay, J. R.; Prvu Bettger, J.; Freburger, J. K.; Gesell, S. B.; Coleman, S. W.; Sissine, M. E.; Wen, F.; Hunt, G. P.; Rosamond, W. D.; Duncan, P. W., Hospital Readmissions and Mortality Among Fee-for-Service Medicare Patients With Minor Stroke or Transient Ischemic Attack: Findings From the COMPASS Cluster-Randomized Pragmatic Trial. *J Am Heart Assoc* **2021**, *10* (23), e023394. <https://doi.org/10.1161/JAHA.121.023394>.
81. Leppert, M. H.; Sillau, S.; Lindrooth, R. C.; Poisson, S. N.; Campbell, J. D.; Simpson, J. R., Relationship between early follow-up and readmission within 30 and 90 days after ischemic stroke. *Neurology* **2020**, *94* (12), e1249-e1258. <https://doi.org/10.1212/WNL.00000000000009135>.
82. Ariss, R. W.; Minhas, A. M. K.; Lang, J.; Ramanathan, P. K.; Khan, S. U.; Kassi, M.; Warraich, H. J.; Kolte, D.; Alkhouli, M.; Nazir, S., Demographic and Regional Trends in Stroke-Related Mortality in Young Adults in the United States, 1999 to 2019. *J Am Heart Assoc* **2022**, *11* (18), e025903. <https://doi.org/10.1161/JAHA.122.025903>.

83. Fang, M. C.; Go, A. S.; Chang, Y.; Borowsky, L. H.; Pomernacki, N. K.; Udaltsova, N.; Singer, D. E., Long-term survival after ischemic stroke in patients with atrial fibrillation. *Neurology* **2014**, *82* (12), 1033-7. <https://doi.org/10.1212/WNL.000000000000248>.
84. Winstein, C. J.; Stein, J.; Arena, R.; Bates, B.; Cherney, L. R.; Cramer, S. C.; Deruyter, F.; Eng, J. J.; Fisher, B.; Harvey, R. L.; Lang, C. E.; MacKay-Lyons, M.; Ottenbacher, K. J.; Pugh, S.; Reeves, M. J.; Richards, L. G.; Stiers, W.; Zorowitz, R. D.; American Heart Association Stroke Council, C. o. C.; Stroke Nursing, C. o. C. C.; Council on Quality of, C.; Outcomes, R., Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke* **2016**, *47* (6), e98-e169. <https://doi.org/10.1161/STR.000000000000098>.
85. Prvu Bettger, J.; McCoy, L.; Smith, E. E.; Fonarow, G. C.; Schwamm, L. H.; Peterson, E. D., Contemporary trends and predictors of postacute service use and routine discharge home after stroke. *J Am Heart Assoc* **2015**, *4* (2). <https://doi.org/10.1161/JAHA.114.001038>.
86. Skolarus, L. E.; Feng, C.; Burke, J. F., No Racial Difference in Rehabilitation Therapy Across All Post-Acute Care Settings in the Year Following a Stroke. *Stroke* **2017**, *48* (12), 3329-3335. <https://doi.org/10.1161/STROKEAHA.117.017290>.
87. Olasoji, E. B.; Uhm, D. K.; Awosika, O. O.; Dore, S.; Geis, C.; Simpkins, A. N., Trends in outpatient rehabilitation use for stroke survivors. *J Neurol Sci* **2022**, *442*, 120383. <https://doi.org/10.1016/j.jns.2022.120383>.
88. Ayala, C.; Fang, J.; Luncheon, C.; King, S. C.; Chang, T.; Ritchey, M.; Loustalot, F., Use of Outpatient Rehabilitation Among Adult Stroke Survivors - 20 States and the District of Columbia, 2013, and Four States, 2015. *MMWR Morb Mortal Wkly Rep* **2018**, *67* (20), 575-578. <https://doi.org/10.15585/mmwr.mm6720a2>.
89. Karve, S.; Balkrishnan, R.; Seiber, E.; Nahata, M.; Levine, D. A., Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. *J Stroke Cerebrovasc Dis* **2013**, *22* (7), 938-45. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2011.11.004>.
90. Duncan, P. W.; Bushnell, C.; Sissine, M.; Coleman, S.; Lutz, B. J.; Johnson, A. M.; Radman, M.; Pvr Bettger, J.; Zorowitz, R. D.; Stein, J., Comprehensive Stroke Care and Outcomes: Time for a Paradigm Shift. *Stroke* **2021**, *52* (1), 385-393. <https://doi.org/10.1161/STROKEAHA.120.029678>.
91. Terman, S. W.; Reeves, M. J.; Skolarus, L. E.; Burke, J. F., Association Between Early Outpatient Visits and Readmissions After Ischemic Stroke. *Circ Cardiovasc Qual Outcomes* **2018**, *11* (4), e004024. <https://doi.org/10.1161/CIRCOUTCOMES.117.004024>.

92. Allen, A.; Barron, T.; Mo, A.; Tangel, R.; Linde, R.; Grim, R.; Mingle, J.; Deibert, E., Impact of Neurological Follow-Up on Early Hospital Readmission Rates for Acute Ischemic Stroke. *Neurohospitalist* **2017**, *7* (3), 127-131. <https://doi.org/10.1177/1941874416684456>.
93. Balkaya, M.; Cho, S., Optimizing functional outcome endpoints for stroke recovery studies. *J Cereb Blood Flow Metab* **2019**, *39* (12), 2323-2342. <https://doi.org/10.1177/0271678X19875212>.
94. ElHabr, A. K.; Katz, J. M.; Wang, J.; Bastani, M.; Martinez, G.; Gribko, M.; Hughes, D. R.; Sanelli, P., Predicting 90-day modified Rankin Scale score with discharge information in acute ischaemic stroke patients following treatment. *BMJ Neurol Open* **2021**, *3* (1), e000177. <https://doi.org/10.1136/bmjno-2021-000177>.
95. Gardener, H.; Romano, L. A.; Smith, E. E.; Campo-Bustillo, I.; Khan, Y.; Tai, S.; Riley, N.; Sacco, R. L.; Khatri, P.; Alger, H. M.; Mac Grory, B.; Gulati, D.; Sangha, N. S.; Olds, K. E.; Benesch, C. G.; Kelly, A. G.; Brehaut, S. S.; Kansara, A. C.; Schwamm, L. H.; Romano, J. G., Functional status at 30 and 90 days after mild ischaemic stroke. *Stroke Vasc Neurol* **2022**, *7* (5), 375-80. <https://doi.org/10.1136/svn-2021-001333>.
96. Broderick, J. P.; Adeoye, O.; Elm, J., Evolution of the Modified Rankin Scale and Its Use in Future Stroke Trials. *Stroke* **2017**, *48* (7), 2007-2012. <https://doi.org/10.1161/STROKEAHA.117.017866>.
97. Chye, A.; Hackett, M. L.; Hankey, G. J.; Lundstrom, E.; Almeida, O. P.; Gommans, J.; Dennis, M.; Jan, S.; Mead, G. E.; Ford, A. H.; Beer, C. E.; Flicker, L.; Delcourt, C.; Billot, L.; Anderson, C. S.; Stibrant Sunnerhagen, K.; Yi, Q.; Bompont, S.; Nguyen, T. H.; Lung, T., Repeated Measures of Modified Rankin Scale Scores to Assess Functional Recovery From Stroke: AFFINITY Study Findings. *J Am Heart Assoc* **2022**, *11* (16), e025425. <https://doi.org/10.1161/JAHA.121.025425>.
98. Taleb, S.; Lee, J. J.; Duncan, P.; Cramer, S. C.; Bahr-Hosseini, M.; Su, M.; Starkman, S.; Avila, G.; Hochberg, A.; Hamilton, S.; Conwit, R. A.; Saver, J. L., Essential information for neurorecovery clinical trial design: trajectory of global disability in first 90 days post-stroke in patients discharged to acute rehabilitation facilities. *BMC Neurol* **2023**, *23* (1), 239. <https://doi.org/10.1186/s12883-023-03251-1>.
99. Quinn, T. J.; Dawson, J.; Lees, J. S.; Chang, T. P.; Walters, M. R.; Lees, K. R.; Gain; Investigators, V., Time spent at home poststroke: "home-time" a meaningful and robust outcome measure for stroke trials. *Stroke* **2008**, *39* (1), 231-3. <https://doi.org/10.1161/STROKEAHA.107.493320>.
100. McDermid, I.; Barber, M.; Dennis, M.; Langhorne, P.; Macleod, M. J.; McAlpine, C. H.; Quinn, T. J., Home-Time Is a Feasible and Valid Stroke Outcome Measure in National Datasets. *Stroke* **2019**, *50* (5), 1282-1285. <https://doi.org/10.1161/STROKEAHA.118.023916>.

101. O'Brien, E. C.; Xian, Y.; Xu, H.; Wu, J.; Saver, J. L.; Smith, E. E.; Schwamm, L. H.; Peterson, E. D.; Reeves, M. J.; Bhatt, D. L.; Maisch, L.; Hannah, D.; Lindholm, B.; Olson, D.; Prvu Bettger, J.; Pencina, M.; Hernandez, A. F.; Fonarow, G. C., Hospital Variation in Home-Time After Acute Ischemic Stroke: Insights From the PROSPER Study (Patient-Centered Research Into Outcomes Stroke Patients Prefer and Effectiveness Research). *Stroke* **2016**, *47* (10), 2627-33. <https://doi.org/10.1161/STROKEAHA.116.013563>.

CHAPTER 3: MANUSCRIPT 1 – ACCURACY AND REPRESENTATIVENESS OF PATIENT-LEVEL OUTCOMES DATA FOLLOWING LINKAGE OF A STATEWIDE STROKE REGISTRY TO AN ADMINISTRATIVE CLAIMS DATABASE

3.1 Abstract

Background and objectives: Collection of patient-level outcomes data addressing survival, function, utilization, and quality of life has been a challenge for stroke registries. Data linkage to administrative databases is a potential method to obtain outcomes data. We linked a cohort of acute stroke discharges entered into the Michigan Stroke Registry (MiSP) a statewide, hospital-based acute-stroke registry with the Michigan Value Collaborative (MVC) comprehensive, statewide, claims database that includes Medicare fee-or-service (FFS) and Blue Cross Blue Shield of Michigan private and Medicare advantage insured populations. We evaluated the accuracy, completeness, and representativeness of the linked data, and generated descriptive data on 30-day, 90-day and 1-year outcome events post stroke hospitalization.

Methods: Between 2016-2020, 46,330 confirmed acute-stroke discharges (ICD-10 I61-I63) from 31 MiSP hospitals were linked to 30,685 acute-stroke claims in MVC. Records were deterministically and probabilistically linked using five variables: date-of-birth, sex, admission-date, discharge-date, and hospital-ID using Match*Pro V2.3. Pre linkage qualitative and post linkage quantitative linkage evaluation was conducted to determine linkage completeness, accuracy, and representativeness. We used the linked index stroke claims data to generate descriptive data on 30-day, 90-day and 1-year outcome event rates including mortality (among Medicare FFS beneficiaries only), all-cause hospital readmissions, stroke recurrence, use of post-acute care services (i.e., IRF, SNF, and home health), use of out-patient visits, and home time (among Medicare FFS beneficiaries only).

Results: Of the 46,330 MiSP stroke events, 23,918 (51.6%) were linked to MVC claims database; these links represent 77.9% of the 30,685 MVC acute-stroke claims. Probabilistic linkage produced a higher number of unique linked pairs (n= 23,918) compared to deterministic linkage (n= 22,660). Substantially lower linkage rates (proportion of MiSP data that linked) were noted among the <65 age group compared to >=65 age group (29.2% vs 63.7%), yet there were fewer demographic differences between the linked and unlinked cases in the younger age group. Using 204 negative controls (observations that should not link) among the <65 year old age group, we found only 1 false positive link (0.5%). Outcome event rates were similar to previously published rates in the literature (Table 3.1).

Table 3.1: 30-day, 90-day, and 1-year post stroke discharge outcome event rates.

Outcome (N=19,382)*	30-day event rate % (n)	90-day event rate % (n)	1-year event rate % (n)
Inpatient rehabilitation facility utilization	24.9 (4,822)	25.5 (4,946)	26.7 (5,171)
Skilled nursing facility utilization	28.1 (5,449)	31.2 (6,049)	34.9 (6,765)
Home health utilization	27.5 (5,336)	38.4 (7,436)	44.7 (8,659)
Outpatient visit	46.4 (8,999)	70.8 (13,720)	85.3 (16,539)
All cause readmission	14.1 (2,724)	24.9 (4,833)	42.2 (8,169)
Stroke recurrence	3.3 (641)	5.1 (991)	8.3 (1,614)
Mortality**	4.0 (486)	9.1 (1,109)	19.8 (2,416)
Home time- Median (IQR) days**	22.0 (26.0)	79.0 (40.0)	347.0 (94.0)

* Numerator includes only the first occurrence for a given patient during follow up period for all outcomes but home time.

** Calculated only for Medicare FFS beneficiaries (n= 12,185).

Conclusions: Linkage between this acute stroke registry and claims data using indirect identifiers allowed for reporting of several stroke outcome metrics up to 1-year post discharge. Probabilistic linkage produced a marginally greater number of links compared to deterministic methods. These data provide important insights into patient outcomes that can be further studied by healthcare professionals, systems, and policy makers to develop interventions, evaluations, and incentives that can potentially lead to improvements in stroke care.

3.2 Introduction

3.2.1 Stroke Registries and the Promise of Data Linkage

Clinical registries collect a uniform body of data to evaluate population specific outcomes for a defined disease, condition, or exposure.^{1,2} In the last 20 years the development of national-level³ and state-level⁴ hospital-based acute stroke registries have provided data to facilitate important improvements in the quality of stroke care,⁵⁻⁸ reduce treatment gaps^{9,10} and identify disparities in stroke care,⁵ resulting in improved patient outcomes.^{5,10-13} The large volume of data collected, which for the national Get With The Guideline-Stroke (GWTG-S) program—now exceeds 9 million stroke discharges from more than 2,000 hospitals, has allowed for the detailed examination of the associations between patient- and hospital- level characteristics and improvements in quality of care and outcomes for stroke patients up to the point of hospital discharge.^{8,13}

However, US stroke registries are limited by the fact that patient outcomes are restricted to those that occur during the index hospitalization. Yet, the majority of patient recovery of function and community participation occur three to six-months following hospital discharge.¹⁴ Collection of patient-level outcomes data addressing survival, healthcare utilization, function, and quality of life has been a challenge for stroke registries because of the substantial investment of resources, both human and financial, required to follow-up and collect data on stroke survivors.¹⁵ A more feasible and sustainable alternative to tracking each patient is to obtain data through data linkage to other administrative, medical, or public health databases that can provide a more holistic assessment of patient outcomes including mortality, readmissions, and post-acute care utilization e.g., rehabilitation.¹⁶⁻¹⁹

The potential value of obtaining patient outcome data from data linkage is illustrated by previous studies that have linked the GWTG-S registry to Medicare FFS data.^{12, 20-27} Findings from these linkage studies have found that stroke patients treated at hospitals participating in the GWTG-S program had superior functional outcomes post-discharge and reduced post-discharge mortality and readmissions, compared to patients treated at non-GWTG-S hospitals.¹² Other studies that used linked GWTG-S data have identified post-acute treatment gaps among minor ischemic stroke and transient ischemic attack patients who are unable to ambulate independently at discharge that can be bridged by improved therapeutic options to reduce disability and the overall incidence of mortality and recurrence,²⁶ disparities in acute stroke care according to the hospital participation in Medicare health saving programs,²⁵ and identified predictors of major cardiovascular events post discharge.^{20, 22}

While these reports help illustrate the value of data linkage, they are limited to studying outcomes in the Medicare FFS population; thus, there are scarce data for stroke patients younger than 65 years or among Medicare Advantage beneficiaries.^{19, 28, 29} Stroke among those aged < 65 (which constitute about 1/3rd of all stroke events³⁰) remains an understudied population despite evidence of increasing rates of stroke in younger adults,³¹ high health-care costs, and loss of labor productivity.^{29, 32} In addition, recently published data by CMS shows that Medicare Advantage population has been growing steadily from 25% in 2010 to 42% in 2020 of eligible Medicare beneficiaries; thus reliance on Medicare FFS data for data linkage risks becoming less and less representative of the total >65 years old population.³³

3.2.2 Aims

In a cohort of acute stroke discharges from Michigan we aimed to (1) generate a unique database by linking a 5-year cohort of stroke discharges entered into the MiSP between 2016-

2020 with the MVC claims registry using both deterministic and probabilistic matching techniques, (2) evaluate the accuracy, completeness, and representativeness of the linkage results using pre linkage qualitative and post linkage quantitative methods, and (3) use the linked data to generate descriptive data on 30-day, 90-day and 1-year outcome event rates.

3.3 Methods

3.3.1 Linkage Databases

This study used three data sources to build a linked, longitudinal dataset defined by hospitalized index stroke events and all of the subsequent healthcare insurance claims for 1 year post discharge. These data sources include MiSP- an acute stroke registry, MVC- a statewide claims database, and the American Hospital Association- an annual hospital-based survey database. The initial starting cohort consists of acute ischemic and hemorrhagic stroke discharges (ICD-10 I61-I63) prospectively collected by 31 Michigan hospitals participating in the MiSP between January 2016 and December 2020. Statewide claims data are provided by the MVC registry. Both, MiSP and MVC datasets are deidentified and so do not contain any unique patient identifiers, hence linkage is achieved using non-unique identifiers i.e., date of birth, sex, admission date, discharge date, and hospital ID. In addition, data from the American Hospital Association's annual survey database was obtained and linked to the admitting hospital unique identification number and admission year.

The MiSP is a statewide, hospital-based acute-stroke registry which is part of the CDC Paul Coverdell National Acute Stroke Program (PCNASP). MiSP continuously collected data from 31 participating hospitals in Michigan during the target period of 2016-2020. These 31 hospitals include the majority of certified stroke centers in Michigan (totaling 49) that provide care for approximately 64% of all stroke admissions in the state.^{4, 34} MiSP tracks patient

demographic, clinical characteristics, diagnostic testing, treatment, and in-hospital outcomes (i.e., mortality, discharge destination, ambulatory status) for all stroke admissions to support hospital-based quality improvement programs.^{4, 34} MiSP identifies stroke discharges using a broad clinical case definition that is based on clinical characteristics as well as ICD-10 (ischemic and hemorrhagic stroke codes: I61-I63 and transient ischemic attack codes: G458-G459) discharge codes.³⁴ For each confirmed stroke discharge detailed clinical data are entered into the American Heart Association's GWTG-S comprehensive case record form (CRF)- a standardized data collection form by trained abstractors.³⁵ Stroke discharges are reported in MiSP as a standalone anonymized event and so there is no ability to link events related to the same patient. It is therefore, not possible to distinguish stroke discharges as either index stroke events versus stroke readmissions.

MVC is a comprehensive, statewide, claims-based database that includes data from 101 participating hospitals and 40 physician organizations in the state.³⁶ The MVC data registry covers 71% of Michigan's 143 hospitals.³⁶ MVC contains claims data for Michigan residents insured by Medicare fee-for-service or Medicaid, and all those covered by Blue Cross Blue Shield of Michigan (BCBSM) including, Preferred Provider Organization (PPO), Blue Care Network Health Maintenance Organization (HMO), and Medicare Advantage plans. The MVC database thus includes all insurance claims for approximately 84% of Michigan's insured population.³⁶ The MVC data are organized according to "episodes of care" that begin with an index hospitalization along with all post-discharge claims submitted over the following 90-days using a unique identifier (member ID).³⁶ The member ID remains the same throughout the follow up period unless a change of insurance coverage takes place.³⁶ Episodes of care are grouped into 39 individual medical and surgical conditions including stroke (MVC identifies stroke discharges

using primary diagnostic (ICD-10 I61-I63) code-based case definition).³⁶ The post discharge facility and professional claims (i.e., readmissions, admission to in-patient rehabilitation facility and skilled nursing facility, use of home health services, outpatient visits, emergency department visits, and prescription fillings) enable tracking of healthcare utilization, expenditures, and other patient outcomes (e.g. home time, successful discharge to community) over time. Due to restrictions in MVC's DUA, Medicaid data were not available to be used for this study.

The American Hospital Association's data represents the most reliable and comprehensive data about hospital facilities in the US.³⁷ The data is collected through a voluntary survey completed annually by nearly 6,300 hospitals and more than 400 health care systems. The survey collects extensive data on a wide variety of topics including hospital organizational structure, facilities and services, ownership/tax status, teaching status, utilization data (e.g., bed utilization rate, total number of emergency department visits, total number of admissions, total number of outpatient visits), physician arrangements (physician compensation and incentive systems), staffing, and community orientation.³⁷ The survey responses are supplemented with data from the US Census Bureau (e.g., county name, core-based statistical area name, statistical area type, area code), hospital accrediting bodies (e.g., joint commission), and other organizations.³⁷ The American Hospital Association reports an approximately 85% response rate to the survey each year. For those hospitals not responding, statistical imputation methods are used to estimate a number of key variables.³⁷

This research was approved by Michigan State University (MSU), University of Michigan (UM), and Michigan Department of Health and Human Services (MDHHS) Institutional Review Boards (IRB). MiSP and MVC datasets are both classified as limited data

sets according to The Health Insurance Portability and Accountability Act (HIPPA) hence patient consent was not required.³⁸

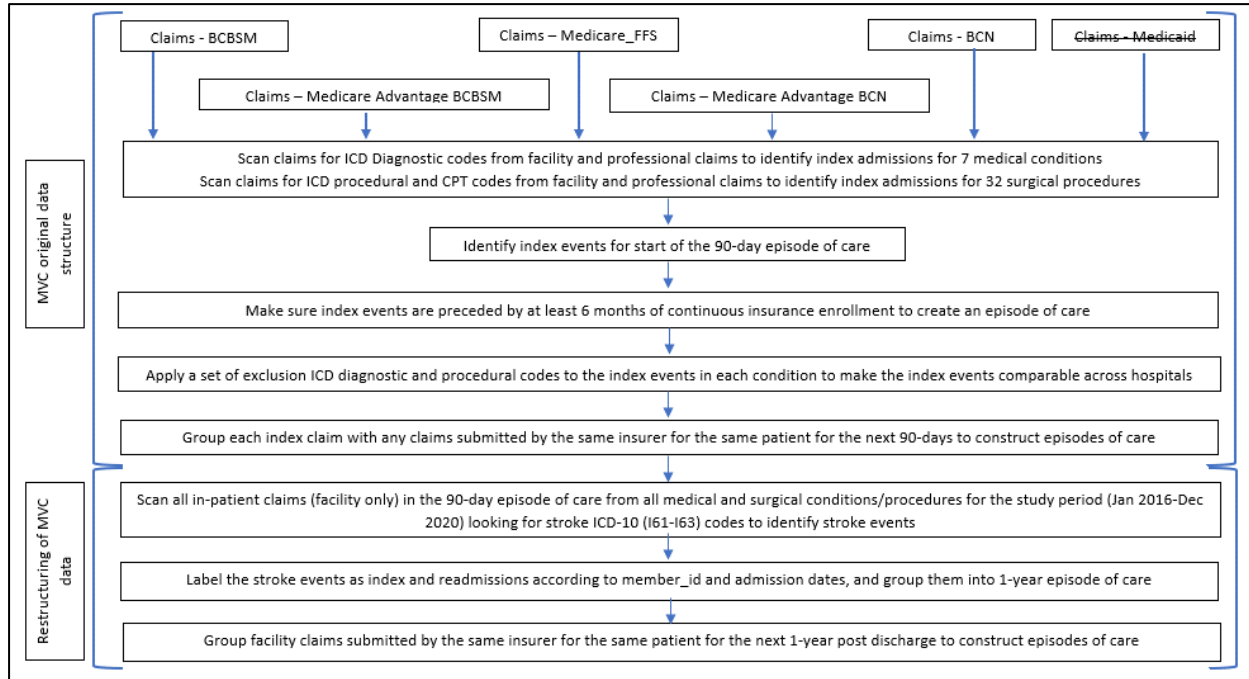
3.3.2 Data Cleaning

In this research, an *index stroke event* was defined as a patient's first stroke discharge during the 5-year study period, and a *recurrent stroke event* was defined as any subsequent stroke discharge occurring within one-year of the discharge date of an index stroke event. Recurrent stroke events that occur after one-year of the first index event for the same patient were ignored in the calculation of outcome event rates. Hence an individual patient with an index stroke will only appear one time in the dataset.

MVC's original organization of 39 medical and surgical conditions into 90-day episodes of care did not fully serve the purpose of this study. The original organization of index events into 39 conditions meant that some index stroke events would have been missed because stroke discharges can occur within an existing episode of care for conditions other than stroke. In addition, the aim of this research is to examine stroke outcomes up to 1-year rather than 90 days. To resolve these limitations, the data were restructured by identifying all stroke events for each unique member ID in the MVC data and labelling the first stroke event during the 5-year period as the index stroke (Figure 3.1). A stroke related admission was identified using primary ICD-10 I61-I63 discharge codes (Table 3A.1 – Appendix). For each index event, all subsequent medical claims reported within the 1-year period following discharge were identified (Figure 3.1). Hospitalizations occurring during the 1-year period following an index stroke admission were classified as either recurrent stroke (corresponding to ICD-10 primary discharge code ICD-10 I61-I63), or non-stroke readmissions (all other discharge codes). Following the identification and

episode of care building step, a comprehensive cleaning process took place to remove duplicate claims submitted for the same health service (Figure 3.2).

Figure 3.1: Reconstruction of MVC episode of care to identify index stroke events that occurred between 2016 and 2020 in the MVC data.

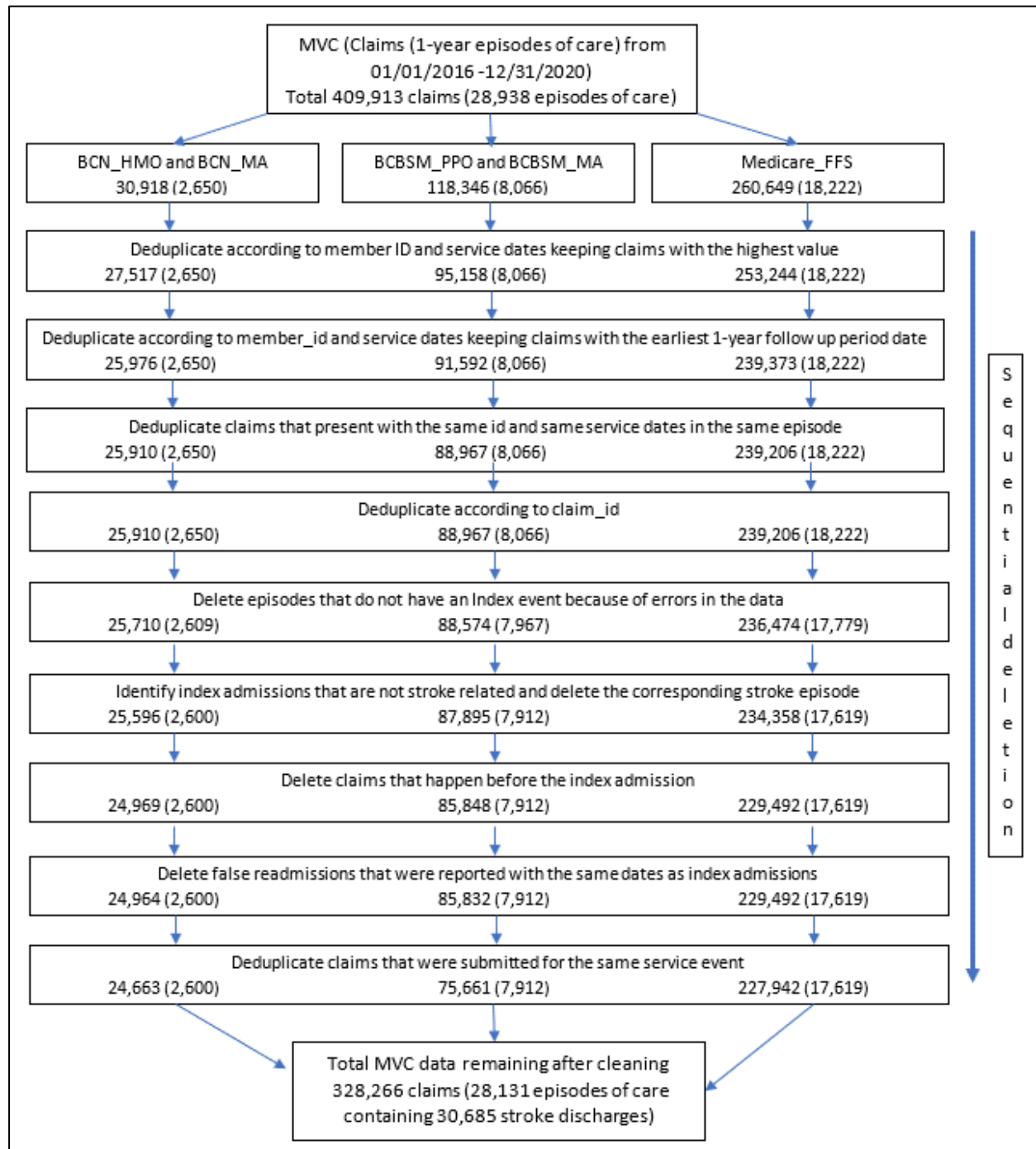


Due to the complex claims data cleaning procedures, only facility claims were included (professional and prescription related claims were not examined). Planned and unplanned admissions could not be distinguished because we did not have access in the MVC data to secondary ICD-10 diagnostic, ICD-10 procedural, or clinical classification software (CCS) codes that are needed to apply the Agency for Healthcare Research and Quality (AHRQ) algorithm to identify unplanned readmission events.³⁹

Compared to MiSP’s broad case definition of acute stroke that is based on clinical characteristics as well as ICD-10 (ischemic and hemorrhagic stroke codes: I61-I63 and transient ischemic attack codes: G458-G459) discharge codes, MVC’s stroke case definition is defined only by the primary discharge-based codes ICD-10 I61-I63 (Table 3A.1) but with the following exclusions: I63.013, I63.033, I63.113, I63.133, I63.213, I63.23, I63.233, I63.313, I63.323,

I63.333, I63.343, I63.41, I63.413, I63.423, I63.433, I63.443, I63.5, I63.513, I63.523, I63.533, I63.543, I63.81, I63.89, or G458-9. The specific reasons that these codes were excluded remain

Figure 3.2: MVC 1-year episode of care data cleaning procedure.



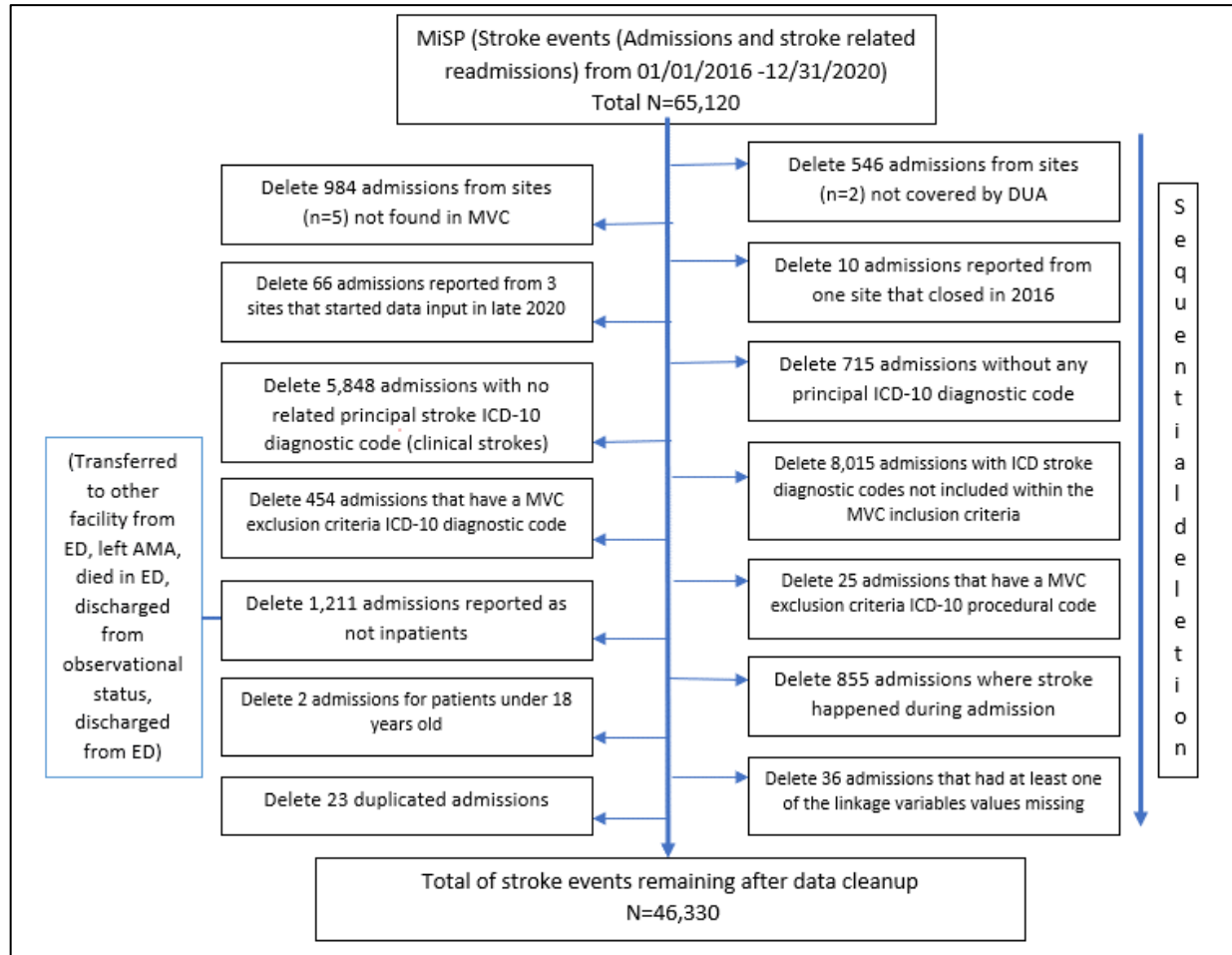
unclear to the researchers. Some of these codes are common including I6381 (Other cerebral infarction due to occlusion or stenosis of small artery), I6389 (Other cerebral infarction), and G458-9 (transient ischemic attacks). However, many of the excluded ischemic stroke codes (I63) represent relatively uncommon infarcts that occur in small or unspecified cerebral, cerebellar, or vertebral arteries or rare bilateral infarcts.

In addition to differences in codes used to define stroke between MVC and MiSP, there are several other differences between the two datasets including: 1) Medicaid data was excluded from MVC due to limitations of the current data use agreement, 2) MVC lacks data from non-BCBSM private insurance providers (such as plans provided by Henry Ford Health system, Priority Health, and United Health), and 3) MVC includes additional inclusion criteria (i.e., continuous insurance coverage 6 months prior to index event, in-patient admissions only) and exclusion criteria (i.e., selected diagnostic and procedural (CPT) codes (described in Figure 3.1)) which contributed to the net marked differences in the total number of reported acute stroke discharges between the registry and the MVC data in the 31 hospitals over the 5 year period (MVC:30,685 vs MiSP:63,514).

To improve the comparability of the two data sources prior to matching, cases that would not be present in the MVC data were removed from the MiSP data following MVC's inclusion and exclusion criteria (Figure 3.3). We were not able to completely implement all MVC's exclusion criteria due to unavailability of some data (i.e., type of insurance provider, and secondary ICD-10 diagnostic and procedural codes). After applying the MVC inclusion and exclusion criteria, the number of acute stroke discharges in the MiSP data was reduced from 63,514 to 46,330 (Figure 3.3). It is worth noting that 8,015 (12.6%) cases were excluded due to ICD-10 diagnostic codes not found in MVC stroke definition (Table 3A.1). Specifically, 90% of

these exclusions were due to 3 codes: 17.0% were coded to I6381 (Other cerebral infarction due to occlusion or stenosis of small artery), 15.1% were coded to I6389 (Other cerebral infarction), and 62.5% were coded to G458-G459 (transient ischemic attack). All data cleaning, merging, and linkage preparations were done using SAS software v9.4 (Cary, NC).

Figure 3.3: MiSP cleaning procedure to match MVC’s inclusion and exclusion criteria.



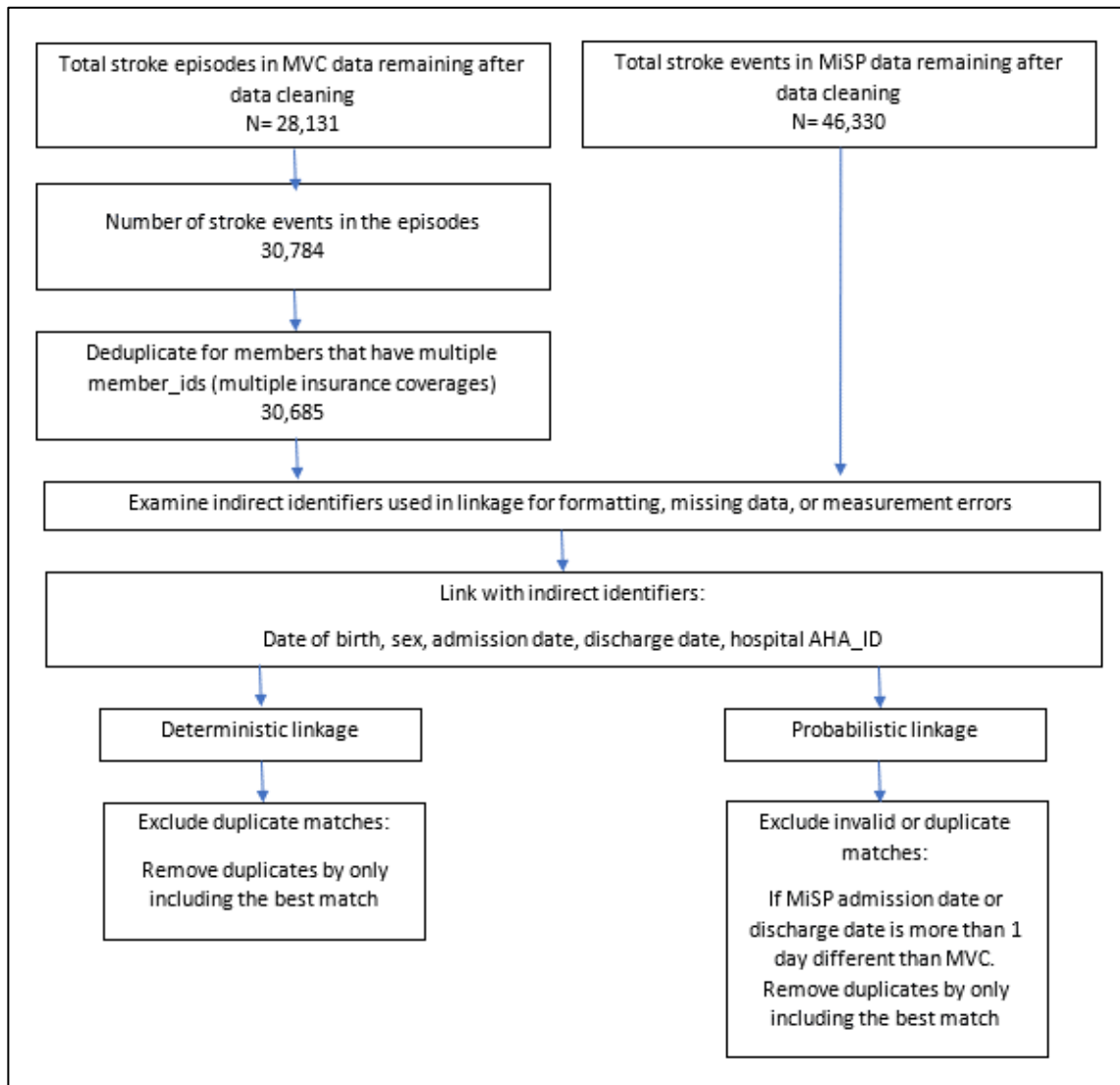
3.3.3 Data Linkage Process

Because the MiSP dataset is unable to distinguish between index events and recurrent stroke events, linkage with MVC must take place at the individual stroke event level. As described above (Figures 1 and 2), the MVC dataset was first reorganized to include stroke

events (identified as index or recurrent, n=30,685) recorded within the 1-year episode of care (n=28,131) with their corresponding linkage variables (Figure 3.4).

Because we lacked unique patient identifiers we used an indirect method^{40, 41} for data linkage based on five non unique identifiers (i.e., date of birth, sex, admission date, discharge

Figure 3.4: Deterministic and probabilistic linkage between MiSP and MVC.



date, and hospital identification number) that were recorded in each of the two datasets. Linkage variables in both datasets were examined to make sure that there were no formatting, missing data, or measurement errors (e.g., illogical values of age (e.g., <18 years old) or dates (e.g., discharge dates that don't follow admission dates)) (Figure 3.4).

Deterministic and probabilistic linkages were conducted between the 46,330 MiSP and 30,685 MVC acute stroke discharges using five linkage variables (i.e., date of birth, sex, admission date, discharge date, and hospital ID) (Table 3.2). Linkage rates were calculated as the proportion of MiSP stroke events that linked to MVC. To make the linkage process more efficient, we restricted the linkage process among subgroups of the data according to year of admission in a process referred to as blocking.^{42, 43}

Table 3.2: Baseline linkage criteria using deterministic and probabilistic linkage methods.

Linkage type	Linkage variable					
	DOB	Sex	Admission date	Discharge date	Hospital ID	Admission year
Deterministic*	X	X	X	X	X	Blocking
Probabilistic**	X	X	X (± 1 day)	X (± 1 day)	X	Blocking

*Linkage criteria employed exact matching on all linking variables.

**Linkage criteria used Fellegi-Sunter algorithm and employed exact matching on all linking variables except for admission and discharge dates where a linkage was allowed if dates differed by ± 1 day.

Under the deterministic approach, a link was considered valid only if all linkage variables between the two datasets exactly matched. Probabilistic linkage allowed us to relax some of the linkage conditions. Specifically, we allowed admission and discharge dates to vary by one day (+ or $- 1$ day) (Table 3.2). The decision to relax the admission and discharge dates was taken because recorded dates can vary especially for admission dates when the date a patient first seeks care in an emergency room may be different from the day they were admitted to the inpatient wards.

For probabilistic linkage a match weight is generated for each possible pairing and a threshold is set above which linked pairs are regarded as correct matches.⁴² The match weights are calculated to each field (linkage variable) for a potential match by means of two probabilities called the *m*- and *u*-probabilities.^{42, 43} The *m*-probability is the likelihood of two fields matching

if the records belong to the same individual. The u -probability is the likelihood of two fields matching by chance if the records do not belong to the same individual.

In our case since we don't know the real match status of our data and under the assumption that our fields are independent from each other, the m - and u -probabilities are estimated for each linkage field via Fellegi-Sunter algorithm.⁴³ The Fellegi-Sunter algorithm examines all the potential matching scenarios in a linkage field and determines the probability of matching and unmatching using information from other linkage fields.⁴³ After obtaining the m - and u -probabilities, a match or non-match weight will be calculated by taking the base 2 logarithm of the frequency ratio (Table 3.3).^{42, 43} A highly discriminative linkage field would be associated with a higher m -probability and a lower u -probability which corresponds to a higher weight.⁴³ The same process is implemented to calculate other linkage fields weights. A total linkage weight is simply the sum of the linkage weights from all fields.^{42, 43}

Table 3.3: Match and non-match field weight calculation.

Outcome	Proportion of links	Proportion of non-links	Frequency ratio	Weight
Match	m	u	m/u	$\text{Log}_2 m/u$
Non-match	$1-m$	$1-u$	$(1-m)/(1-u)$	$\text{Log}_2 (1-m)/(1-u)$

Many-to-many matching scheme was used within the blocked data, which allowed one stroke event in MiSP to link to many events in MVC and vice versa. Deduplication of potential links by including only the best match (highest linkage weight) of linked pairs took place to make sure that only unique events are linked in a one-to-one matching scheme.

After deduplication took place, we manually reviewed all the weights of the generated possible linkages to determine the lowest linkage weight of a plausible match (according to our linkage criteria) to serve as the linkage threshold. The selected threshold captured all the linked pairs that follow our linkage terms. After implementing the threshold, an additional manual review of the included linked pairs took place to make sure that their linkage fields followed our

linkage criteria. We reported on the probabilistic linkage weights and the manual review process in the Appendix (Table 3A.2). Linkage was done using Match*Pro v2.4.1.

3.3.4 Linkage Errors and Linkage Bias

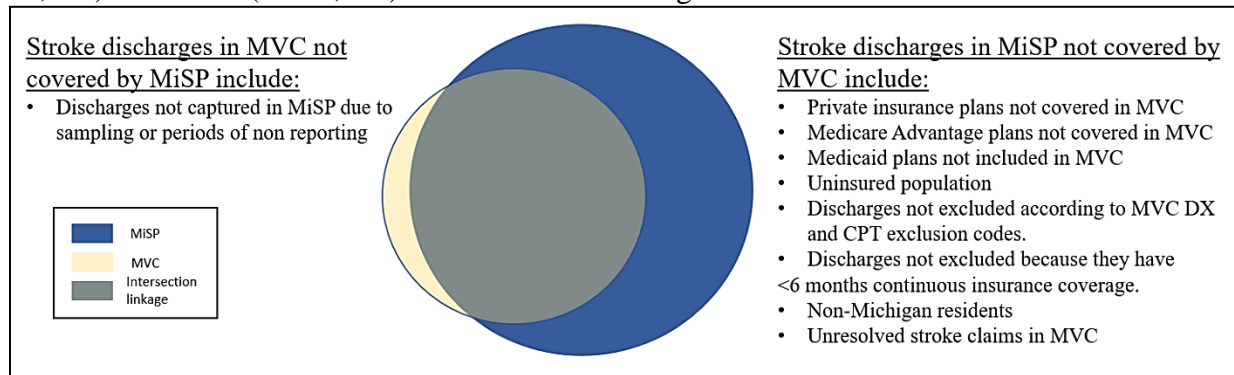
Linkage errors are introduced either as a missed link where records relating to the same entity do not link (referred to as a false negative) or a false link where records relating to different entities link (referred to as a false positive).^{44, 45} Although our linkage variables had no missing data (after dropping 36 stroke events in MiSP that had at least one missing linkage variable value) the inputted values of these variables may be inaccurately recorded (especially date variables) which may produce linkage errors and a potentially biased linked dataset that could affect study results.⁴⁶ This also could happen because of other reasons that are unknown to the researchers. Thus, conducting pre linkage and post linkage evaluation of the linked data through qualitative and quantitative methods is important to explore any potential linkage errors. The following two sections include detailed explanation of qualitative and quantitative linkage quality evaluation methods.

The linkage errors have different implications (i.e., selection, misclassification, and information biases) depending on the structure of the linkage between the datasets and how the linkage determines the value of a variable of interest (i.e., whether the data is measuring an outcome or exposure variable), or the inclusion or exclusion criteria of the final analysis dataset.^{44, 47}

In our case the structure of linkage is the intersection of MVC and MiSP stroke events, where a proportion of each dataset is expected to link. Figure 3.5 illustrates this linkage intersection for the 2 datasets and reasons why entities might not link. The final dataset that will

be used for our analysis (to explore stroke outcomes) will only include entities that linked (a complete case analysis approach) between the 2 datasets.

Figure 3.5: Expected intersection linkage structure and coverage limitations between MiSP (n= 46,330) and MVC (n= 30,685) datasets after cleaning.



3.3.5 Pre-Linkage Evaluation Using Qualitative Methods

Understanding the implications of the potential errors that may occur from a linkage before it takes place represents the qualitative phase of linkage evaluation. This step is also important to determine what quantitative approaches could be implemented. Given our linkage structure and inclusion criteria, selection bias which can result from missed links (false negative) and misclassification introduced by false links (false positive) were of greatest concern.^{44, 47} A false negative link occurs if a stroke event in MiSP and its corresponding claim in MVC fail to link. A false positive link occurs when a link between a stroke event in MiSP and a stroke claim in MVC occurs in error (they represent different events). Because our linkage is meaningfully interpreted (Outcomes are derived from the linkage), potential misclassification or measurement error in stroke outcomes can occur if the patient outcomes (e.g., mortality, readmission, recurrence) differ between false positive linked stroke events and false negative stroke events.^{44,}

⁴⁷

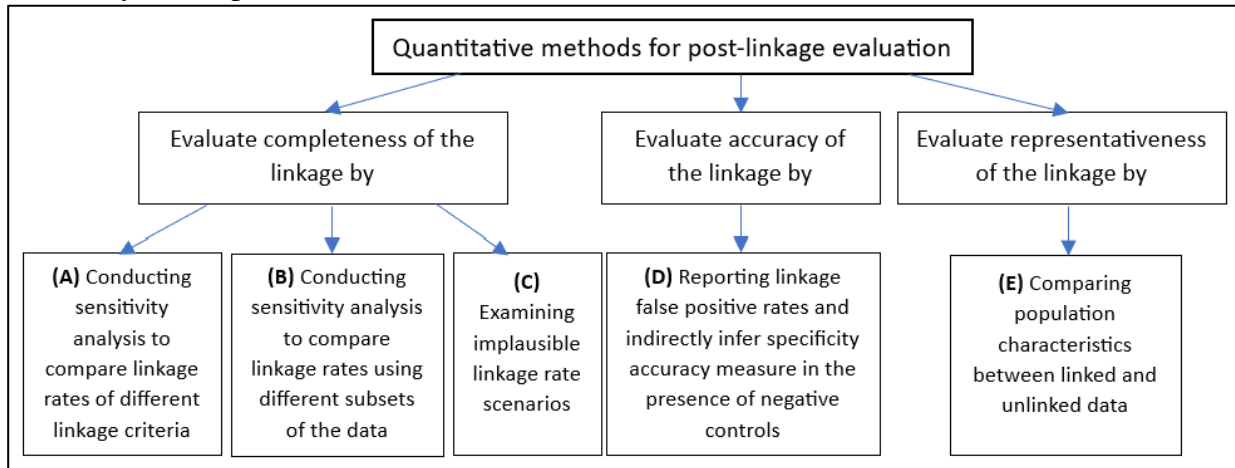
In addition to the above linkage errors, our linkage-based inclusion criteria will produce a dataset that will de facto suffer from selection bias because the MVC data does not contain data

from all relevant insurance coverages (i.e., Medicaid, Medicare advantage, and commercial plans other than BCBSM), compared to MiSP that includes stroke events with all types of insurance as well as the uninsured population. This selection bias is of concern because of the sizable proportion of the MiSP dataset that are not covered by these plans which will result in a linked dataset that is not representative of the total MiSP population. This will also reduce the external validity of the results and affect the representation of certain subgroups such as the younger than 65 years old population. Since this selection bias does not occur at random and because no gold standard linkage exists, valid multiple imputation or inverse probability weighting techniques to account for this missing data cannot be easily implemented.^{46, 48} A gold standard linkage is defined as a linkage that produces a reference dataset where true match status is known with certainty.⁴⁵ In our case such dataset is not available and cannot be produced due to the lack of direct identifiers (e.g., social security number, medical record number, patient name or address), and it is not possible to generate other comparable external data sources that could be used to validate the accuracy of linkage.

3.3.6 Post-Linkage Evaluation Using Quantitative Methods

Due to the absence of both direct identifiers and a gold standard dataset to compare to, the available options for quantitative methods to evaluate the linkage quality are somewhat limited and must rely on indirect measures of linkage quality.^{44, 45} Our evaluation assessed the completeness (how well did the linkage criteria capture all the potential linkage pairs as represented by linkage rates), accuracy (to the extent possible are the observed linkage results valid), and representativeness (how similar is the linked data to the original study population) of the linked dataset relative to MiSP dataset (Figure 3.6).

Figure 3.6: Post-linkage quantitative approaches implemented to evaluate completeness, accuracy, and representativeness of the linked dataset relative to MiSP dataset.*



* A-E represents different quantitative techniques performed on the available data.

3.3.6.1 Evaluating the Completeness of Linkage (See Boxes A, B, and C in Figure 3.6)

To evaluate the completeness of the linkage, we conducted several sensitivity analyses comparing linkage rates using different linkage criteria or methods (Figure 3.6 Box A).⁴⁵ Our evaluation process was built on the selected linkage variables: date of birth, sex, admission and discharge dates, and hospital ID in our deterministic or probabilistic linkage. Given our limited ability to quantify the accuracy of the match, we decided to use the linkage method that produces the highest linkage rate as the linkage method of choice. Our evaluation process was separately carried out for the deterministic and probabilistic linkage techniques.

The deterministic linkage technique was evaluated by comparing match rates achieved using 3 different linkage criteria applied to the baseline 5 linkage variables criteria, ranked from the least to most stringent: 1) linkage using date of birth, sex, admission and discharge dates with blocking on hospital ID (Linkage criteria 1 in Figure 3.7A), 2) linkage using date of birth, admission and discharge date, and hospital ID with blocking on admission year (Linkage criteria 2 in Figure 3.7A), 3) linkage using date of birth, sex, admission and discharge dates, hospital ID, and an additional linkage variable with blocking on admission year (the ICD-10 discharge

diagnostic code) (Linkage criteria 3 in Figure 3.7A). The latter (linkage criteria 3) was mainly used to evaluate the degree of discrimination of our linkage variables. Furthermore, we blocked on discriminating variables including hospital or admission year which in our case has the potential effect of decreasing the number of false links without affecting true matches (Figure 3.7A).⁴²

The completeness of baseline probabilistic linkage that utilized Fellegi-Sunter algorithm was assessed using the Expectation-Maximization (EM) algorithm, a widely used probabilistic algorithm for obtaining maximum likelihood estimates of agreement rates among true links (m-probability) and false links (u-probability) (Linkage method 4 in Figure 3.7A).⁴⁹ This method is most effective when linkage variables have missing data that can be replaced through imputation using information from other data fields.⁵⁰ In contrast to the human supervised probabilistic linkage that took place earlier, this unsupervised algorithm uses a preset sensitivity (defined by the user) determines the optimum linkage weight threshold where most of the true links are captured with minimum number of false links without a human review process.^{49, 50} Nevertheless, after applying the EM algorithm we undertook a manual review to exclude any false links (defined as implausible links that do not match the linkage criteria) (See Table 3A.2 in the Appendix).

Figure 3.7: Post-linkage quantitative approaches to evaluate the completeness of the linkage.

A) Comparing linked rates of different linkage criteria or methods.

Linkage criteria/method	Linkage type	Linkage variable*						
		Discharge ICD-10	DOB	Sex	Admission date	Discharge date	Hospital ID	Admission year
1	Deterministic		X	X	X	X	Blocking	-
2	Deterministic		X		X	X	X	Blocking
3	Deterministic	X	X	X	X	X	X	Blocking
4	Probabilistic**		X	X	X (± 1 day)	X (± 1 day)	X	Blocking

*Unless otherwise specified, linkage criteria employed exact matching of linking fields.
 **EM algorithm at 98% sensitivity.

B) Comparing linkage rates using different subsets of the data.

Included data in linkage process	MiSP (N)	MVC (N)
All data	46,330	30,685
All data except sampling hospitals (N=5)	40,774	24,349
All data except 2020 admissions	38,271	25,682
All data except sampling hospitals and 2020 admissions	33,624	20,404

The second set of sensitivity analyses were conducted by applying the linkage method of choice to different subsets of data to assess their influence on the linkage rates (Figure 3.6 Box B and Figure 3.7B).⁴⁴ This comparison took place between 1) the original data, 2) all data except MiSP hospitals that only provided a sample of their stroke admissions (these were identified as hospitals that had a higher number of MVC stroke claims compared to stroke events in MiSP (n=5 hospitals; number 2, 6, 17, 21, and 27 in Table 3A.3), 3) all data except stroke events that were recorded in 2020, and 4) all data except sampling hospitals and 2020 stroke events. These subsets were chosen because we found inconsistencies in reporting stroke events both by sampling hospitals and during the pandemic in 2020 which could contribute to higher levels of selection bias (Table 3A.3 and 3A.4- MiSP and MVC data stratified by admission year and hospital site).

To make sure that our linkage rates are within reasonable bounds (linkage rates did not exceed impossible thresholds), implausible linkage rate scenarios were identified by comparing

observed match rates to the maximum attainable linkage rate for each hospital and in the overall data (Figure 3.6 Box C and Table 3A.4). The maximum attainable linkage rate was calculated by dividing the number of stroke claims in MVC by the number of reported strokes in MiSP. For example, if a hospital has 200 stroke events reported in MiSP versus 101 in MVC, the maximum attainable linkage rate is 50% for that hospital. Sampling hospitals had a maximum linkage rate of 100%.

3.3.6.2 Linkage Accuracy Using Negative Controls (See Box D Figure 3.6)

Definitions of linkage accuracy measures are illustrated in Table 3.4. Given the absence of direct identifiers and a gold standard dataset that identifies true matches between the MiSP and MVC datasets, we were obligated to use alternative methods to report on the accuracy measures using a subset of the data with a known match status.⁴⁴ These subsets can be defined as negative (a subset of records that should definitely not link) and positive controls (a subset of records that should definitely link).⁴⁴ Positive and negative controls help assess the proportion of missed links and false links, respectively so that measures of linkage accuracy (i.e., sensitivity and specificity) can be calculated.⁴⁴

Table 3.4: 2x2 table representing accuracy in record linkage.*

		True match status		Total
		Match	Non-match	
Observed link status	Link	True match (True positive) <i>a</i>	False link (False positive) <i>b</i>	<i>a + b</i>
	Non-link	Missed link (False negative) <i>c</i>	True non-match (True negative) <i>d</i>	<i>c + d</i>
Total		<i>a + c</i>	214 (negative controls)	<i>a + b + c + d</i>

*Linkage accuracy measures: Sensitivity = $a/(a + c)$; specificity = $d/(b + d)$; positive predictive value = $a/(a + b)$; negative predictive value = $d/(c + d)$

In our case, only negative controls were available and consisted of 214 stroke events in MiSP that were identified as having Medicaid insurance (Table 3.4). The stroke events should not link with MVC stroke events since the MVC claims data does not include any Medicaid data.

Even though our pool of negative controls is small and is not representative of the total MiSP dataset (payer designation in MiSP is missing for 82.5% of stroke events), we utilized what data was available to generate an estimate of the rate of false positive matches with the aim of estimating the specificity of the overall match (the probability of not linking given that it is true non match) (Table 3.4).⁴⁴ Stratification by age group is important in our case because the absolute difference in the number of stroke events between MiSP and MVC is much higher in <65 years old compared to ≥ 65 years old group because we lack data on Medicaid as well as other private plans in the <65 years old group (Table 3A.5). This is demonstrated in the big difference in the distribution of negative controls among the <65 versus the ≥ 65 years old group (204 vs 10) which prompted us to report the linkage specificity only for the <65 years old age group. Due to the absence of positive controls, we did not report on the sensitivity (the probability of linking given that it is a true match) and the limited data on negative and positive controls meant that positive predictive value (PPV) (the probability of having a true match given that linkage takes place) or negative predictive value (the probability of having a true non-match given that linkage does not take place) was not reported because neither measures would have been meaningful.⁴⁴

3.3.6.3 The Representativeness of the Linkage (See Box E Figure 3.6)

Finally, to assess representativeness of the final linked data, a comparison of population characteristics between the linked and unlinked MiSP data was generated after stratifying according to age group (<65 and ≥ 65).⁴⁵ Characteristics included age, sex, race, ethnicity, 25 stroke related comorbidities, stroke type, NIHSS upon admission, duration of hospital stay, and discharge disposition. We quantified the differences between the linked and unlinked stroke events in MiSP by calculating absolute standardized differences (ASD), which present mean

differences in terms of the pooled standard deviation (Figure 3.8).^{51, 52} This method is preferred in large sample sizes and are interpreted as representing a meaningful difference at a value of 0.1 or higher.¹⁹ The representativeness of the linkage was also assessed in the linked versus unlinked MVC population after stratifying according to age group (<65 and >= 65) (Table 3A.6). The comparison was carried using age, sex, stroke type, and payer characteristics. Differences were quantified using ASD. Linkage evaluation was done using SAS software v9.4 (Cary, NC).

Figure 3.8: Absolute Standardized Difference equations for continuous and categorical variables among the linked and unlinked stroke events.

$$\text{Standardized difference for continuous variables} = \frac{(\bar{x}_{Linked} - \bar{x}_{Unlinked})}{\sqrt{\frac{s_{Linked}^2 + s_{Unlinked}^2}{2}}}$$

Where \bar{x}_{Linked} and $\bar{x}_{Unlinked}$ denotes mean of the variable in the linked and unlinked stroke events, where s_{Linked}^2 and $s_{Unlinked}^2$ denote the variance of the variable in linked and unlinked stroke event, respectively.

$$\text{Standardized difference for dichotomous variables}^* = \frac{(\hat{P}_{Linked} - \hat{P}_{Unlinked})}{\sqrt{\frac{\hat{P}_{Linked}(1 - \hat{P}_{Linked}) + \hat{P}_{Unlinked}(1 - \hat{P}_{Unlinked})}{2}}}$$

Where \hat{P}_{Linked} and $\hat{P}_{Unlinked}$ denotes the prevalence or mean of the dichotomous variable in the linked and unlinked stroke events. Multilevel categorical variables can be calculated using multivariate Mahalanobis distance method.

3.3.7 Descriptive Statistics of Outcome Variables

The MVC administrative dataset included all post-acute health services claims submitted to the insurance provider during the 12 months period post hospital discharge. The reported outcomes were calculated for the index strokes 1-year episode of care in the linked dataset (n=22,889). These services included in-patient rehabilitation (IPR), skilled nursing facility (SNF), long term acute care hospital (LTACH), home health care (HHC), outpatient medical visits (OP), and hospital readmissions. Mortality data were only available for Medicare FFS insured patients.

The above claims were used to calculate outcome variables including 30-day, 90-day and 1-year event rates of use of post-acute care services (i.e., IRF, SNF, and home health), outpatient visits, all-cause hospital readmissions, stroke recurrence, mortality (among Medicare FFS beneficiaries only), and home time (among Medicare FFS beneficiaries only). We reported outcome event rates for those discharged alive excluding those discharged to hospice care or who left against medical advice. For readmission, stroke recurrence and utilization of post-acute care services (i.e., IRF, SNF, home health, and outpatient visits) only the first occurrence in a given patient was reported during follow up period. Home time is defined as post discharge time spent alive and out of an inpatient care setting including hospital, inpatient rehabilitation, skilled nursing facility, and long-term care hospital. Home time was only calculated for FFS Medicare data because it was the only data source with reliable information on survival. In the event that the patient survived beyond a certain period, for example, 90 days, home time is calculated by subtracting total inpatient days from 90 days. In the event of post discharge death within 90 days home time was calculated by subtracting total inpatient days from the total number of days the subject lived post discharge. One year Kaplan-Meier curves for mortality (Medicare FFS only), all-cause readmission, and stroke recurrence were generated.

As describe earlier, for patients with multiple stroke episodes of care (i.e., another acute stroke admission that occurred at least 1 year apart); only the first episode was included in the analysis, the subsequent events were ignored. The final dataset included 19,382 1-year stroke episodes of care.

A sensitivity analysis was conducted to evaluate if our exclusion criteria (being discharged to hospice, leaving against medical advice, and excluding all stroke episodes of care after the first 1-year episode was recorded) resulted in unanticipated differences in outcome

event rates between the starting linked data 22,889 and the final dataset of 19,382 (Table 3A.7-differences in outcome event rates calculated using the 22,889 episodes sensitivity analysis dataset versus 19,382 episodes study data set).

Because stroke is a heterogeneous disease,⁵³ we stratified the outcomes according to stroke demographics and etiology, including age (Table 3A.8), race (Table 3A.9), stroke type (i.e., ischemic and hemorrhagic) (Table 3A.10), and stroke severity (Table 3A.11). Data analysis was done using SAS software v9.4 (Cary, NC) and R v4.2.3 in RStudio.

3.4 Results

3.4.1 Pre-Linkage Qualitative Evaluation Results

Based on the expected linkage structure as an intersection of MiSP and MVC datasets and the limited insurance coverage of MVC data compared to MiSP, our linked dataset will suffer from selection bias where stroke patients only insured by Medicare FFS and BCBSM are expected to link limiting the generalizability of the derived outcomes. Because we are meaningfully interpreting our linkage to populate outcome measures, false links might generate outcomes that are prone to measurement or misclassification bias. Post linkage quantitative evaluation will further clarify if these biases are of concern.

3.4.2 Linkage Rates and Post-Linkage Quantitative Evaluation Results

Between 2016 to 2020 there were 46,330 stroke events in the MiSP registry dataset and 30,685 stroke events in MVC claims dataset. Deterministic linkage using date of birth, sex, admission and discharge dates, and hospital resulted in 22,660 linked pairs (48.9% MiSP linkage rate) compared to 23,918 linked pairs (51.6% MiSP linkage rate) when using probabilistic linkage (Table 3.5). Because it produced higher number of unique linked pairs, data from the

probabilistic linkage was determined to be the best match and was used for the rest of the analysis.

Table 3.5: Baseline linkage rates using deterministic and probabilistic linkage methods and 5 linkage variables.

Linkage type	Linkage variable						Linked pairs	Linkage rate of n=46,330 in MiSP	Linkage rate of n=30,685 in MVC
	DOB	Sex	Admission date	Discharge date	Hospital ID	Admission year			
Deterministic*	X	X	X	X	X	Blocking	22,660	48.9%	73.8%
Probabilistic**	X	X	X (± 1 day)	X (± 1 day)	X	Blocking	23,918	51.6%	77.9%

*Linkage criteria employed exact matching on all linking variables.

**Linkage criteria employed exact matching on all linking variables except for admission and discharge dates where a linkage was allowed if dates differed by ± 1 day.

Our sensitivity analysis using 4 alternative linkage criteria or methods to assess the completeness of our linkage using deterministic (criteria 1, 2, and 3 in Table 3.6 versus our baseline deterministic linkage criteria in Table 3.5) and EM based probabilistic (method 4 in Table 3.6 versus our Fellegi-Sunter probabilistic method in Table 3.5) linkage methods did not result in major differences of linkage rates. In deterministic linkage, blocking on hospital ID (criteria 1 in Table 3.6) compared to using it as a linkage variable in our deterministic linkage criteria resulted in only 258 additional linked pairs. This result can be attributed to either missed duplicate matches of stroke events or stroke events that got recorded multiple times due to transfer between hospitals in either registry or claims datasets. Compared to our deterministic criteria, excluding sex as a linkage variable (criteria 2 in Table 3.6) resulted in only 44 additional linked pairs signaling that sex does not substantially add to the discriminative power of the other linkage variables. Adding discharge ICD-10 diagnostic codes on top of our 5 linkage variables (criteria 3 in Table 3.6) resulted in only 189 less linked pairs. This result can be explained by the occasional difference in recorded ICD-10 diagnostic code of the stroke event between the registry and the corresponding claim.

Table 3.6: Linkage rates in MiSP and MVC datasets using different linkage criteria/method.

Linkage criteria/method	Linkage type	Linkage variable*							Linked pairs	Linkage rate of n=46,330 in MiSP	Linkage rate of n=30,685 in MVC
		Discharge ICD-10	DOB	Sex	Admission date	Discharge date	Hospital ID	Admission year			
1	Deterministic		X	X	X	X	Blocking	-	22,918	49.5%	74.7%
2	Deterministic		X		X	X	X	Blocking	22,704	49.0%	74.0%
3	Deterministic	X	X	X	X	X	X	Blocking	22,471	48.5%	73.2%
4	Probabilistic**		X	X	X (± 1 day)	X (± 1 day)	X	Blocking	23,918	51.6%	77.9%

*Unless otherwise specified, linkage criteria employed exact matching of linking fields.

**EM algorithm at 98% sensitivity.

In probabilistic linkage, using the expectation maximization (EM) algorithm (method 4 in Table 3.6) initially resulted in generating many false positive additional linked pairs (see Table 3A.2) but after manual review these were eliminated and so the method did not yield any additional valid linked pairs compared to baseline probabilistic matching. Based on these results we believe that our baseline linkage criteria including date of birth, sex, admission and discharge dates, and hospital ID captured the highest number of potential linked pairs with high discrimination (Table 3.5).

Table 3.7 summarizes the results where baseline probabilistic linkage was conducted using different subsets of the data. We did not find significant differences in linkage rates, which suggests selection bias due to incomplete recording of stroke discharges either by sampling hospitals or during the 2020 COVID-19 pandemic year was not present to any important degree (Table 3.7). Our linkage rate of 51.6% did not surpass the maximum plausible linkage rate of 66.2% ($30,685/46,330 \times 100\%$) according to difference in stroke events recorded by MiSP and MVC by year of admission, and hospital site (Table 3A.3 and 3A.4). As for the linkage accuracy measures, we only found one false link among the 204 negative controls in the <65 age group (false positive rate of 0.5%). Thus, the specificity of the linkage was estimated to be 99.5% (Table 3.8).

Table 3.7: MiSP linkage rates using probabilistic linkage techniques on subsets of the data.

Included data in linkage process	MiSP (N)	MVC (N)	Linked pairs (N)	Linkage rate (% MiSP)	Linkage rate (% MVC)
All data	46,330	30,685	23,918	51.6%	77.9%
All data except sampling hospitals (N=5)	40,774	24,349	20,458	50.2%	84.0%
All data except 2020 admissions	38,271	25,682	20,078	52.5%	78.2%
All data except sampling hospitals and 2020 admissions	33,624	20,404	17,125	50.9%	83.9%

Table 3.8: Linkage specificity among the <65 years old age group based on 204 negative controls.*

MiSP (age <65)	True match	True non-match
Link	Unknown	1 (false link)
No link		203 (true non-link)
Total		204 (negative controls)

Specificity: 203/204 = 99.5%

*Sensitivity, positive predictive value, and negative predictive value accuracy measures were not calculated due to the absence of positive controls and the small pool of negative controls.

When comparing the population characteristics among the linked and unlinked populations we found that <65 years old age group linked and unlinked populations were more balanced compared to ≥65 years old age group. Among the <65 years old group, only 9 of 34 characteristics had an ASD of >0.1, whereas there were 30 characteristics with an ASD >0.1 among the ≥65 years old age group (Table 3.9). Among the <65 years old age group the linked and unlinked populations had a similar distribution of sex, ethnicity, stroke type, admission NIHSS, discharge disposition, and the majority of stroke related comorbidities. However, compared to the unlinked population, the linked population was older (mean age 55.1 vs 53.1 years), more white (72.2% vs 63.7%), had a shorter duration of admission, higher prevalence of atrial fibrillation, dyslipidemia, chronic renal insufficiency, and sleep apnea, but lower prevalence of smoking, and drug or alcohol abuse (Table 3.9). Among the ≥65 years old age group the linked and unlinked populations had

a similar distribution of only ethnicity, stroke type, admission NIHSS, and admission duration. Compared to the unlinked population, the linked population was older (mean age 79.0 vs 77.2 years), more white (82.4% vs 77.5%), had a higher proportion of females (55.4% vs 50.2%), more frequently discharge to rehabilitation (i.e., IRF and SNF) and hospice, and carried a higher burden of almost all comorbidities (Table 3.9). As similar comparison was carried out on the MVC data where we found that linked and unlinked populations were very similar to each other in terms of age and sex, but hemorrhagic stroke patients were more likely to link compared to ischemic stroke patients (Table 3A.6).

3.4.3 Stroke Outcome Event Rates

3.4.3.1 Stroke Outcomes in the Linked Population Without Stratification

Among the 19,382 linked 1-year stroke episodes of care, 24.9%, 28.1%, 27.5%, and 46.4% utilized IRF, SNF, home health, and outpatient care at least once within 30-days of hospital discharge, respectively (Table 3.10). Compared to 30-day event rates, IRF and SNF utilization rates increased slightly by 90-days to 25.5% and 31.2%, respectively. In contrast, home health utilization substantially increased from 27.5% at 30 days to 38.4% at 90-days and to 44.7% within 1-year (Table 3.10). At the end of the 30-days, 90-days, and 1-year follow-up period 46.4%, 70.8%, and 85.3% of the linked population had at least one outpatient visit, respectively. A total of 14.1%, 24.9%, and 42.2% of the linked population were readmitted at least once within 30-days, 90-days, and 1-year post discharge, respectively (Table 3.10 and Figure 3.9A). Only 3.3% of our linked population had a stroke recurrence (as defined by a hospital admission for stroke) within 30-days; this increased to 5.1% at 90-days and to 8.3% at 1-year post discharge (Table 3.10 and Figure 3.9B).

We only had accurate mortality data for the 12,185 Medicare FFS cases in the linked data; mortality rates were 4.0%, 9.1%, and 19.8% within 30-days, 90-days, and 1-year post discharge, respectively (Table 3.10 and Figure 3.9C). Median home time was found to be 22.0, 79.0, 347.0 days within 30-days, 90-days, and 1-year post discharge, respectively (Table 3.10). The sensitivity analysis conducted to compare outcome event rates between the 22,889 (the starting linked dataset) and 19,382 (the final linked dataset after excluding patients who were discharged to hospice, left against medical advice, and including only the first stroke episode) populations resulted in an anticipated modest decrease in events rates among the 22,889 episode of care population with the exception of mortality. For example, 30-day mortality rates among the starting linked dataset were 15.8% versus 4.0% in the final linked dataset after implementing the exclusions (Table 3A.7).

3.4.3.2 Stroke Outcomes in the Linked Population Stratified by Age Group

Compared to the linked <65 years old group (n= 4,167), the ≥65 years old group (n= 15,215) the utilization of IRF and HH over 1 year of follow up did not differ, whereas SNF utilization was as expected higher (Table 3A.8). The ≥65 years old group utilized outpatient services in a higher rate up to 90 days post discharge than the <65 years old group but in 1-year rates were not different. All cause readmission rate did not differ between the two-age group during the first 30 days of follow up but the ≥65 years old group had higher readmission rates 90 days and 1 years of follow up. Stroke recurrence among the ≥65 years old group was consistently lower during the 1 year of follow up compared to the <65 years old group. We did not report on the difference in mortality or home time due to limitations in mortality data among the <65 years old groups and MA beneficiaries.

3.4.2.3 Stroke Outcomes in the Linked Population Stratified by Race Group

Compared to the linked White (n= 15,457) or Other (n= 255) racial group, the Black racial group (n= 2,833) had a higher IRF utilization rate over 1-year of follow up (Table 3A.9). The same trend was observed for SNF and HH utilization but over 90 days and 1 year of follow up. On the other hand, the Black racial group utilized outpatient services at a lower rate over 1-year of follow up compared to White and Other racial groups. All cause readmission rates were higher in the black racial group compared to the other groups. This however was not the case in stroke recurrence where recurrence rates did not differ between the racial groups up to 90 days of follow up followed by higher recurrence rates in 1 year of follow up among the black group compared to the rest of the groups.

Mortality rates among the Medicare FFS beneficiaries did not differ between the racial groups over 1 year of follow up. However, the black racial group had a consistent lower home time over 1 year of follow up compared to the rest of the groups likely due to their higher readmission and rehabilitation utilization rates.

3.4.3.4 Stroke Outcomes in the Linked Population Stratified by Stroke Type

Compared to the linked hemorrhagic stroke patients (n= 2,461), ischemic stroke patients (n= 16, 921) had a lower IRF and SNF utilization rates over 1-year of follow up (Table 3A.10). However, HH and OP utilization rates were almost similar between the two stroke types over 1 year of follow up. All cause readmission and stroke recurrence rates were lower among ischemic stroke patients compared to hemorrhagic stroke patients over 1 year of follow up.

Among FFS beneficiaries, ischemic stroke patients had lower mortality rates compared to hemorrhagic stroke patients over 1 year of follow up. Ischemic stroke patients had a consistently

higher home time compared to hemorrhagic stroke patients over 1 year of follow up which is likely a reflection of their better survival.

3.4.3.5 Stroke Outcomes in the Linked Population Stratified by Stroke Severity

Among the mild (NIHSS 0-4) (n= 10,992), moderate (NIHSS 5-15) (n= 4,995), and severe (NIHSS 16-42) (n=1,563) stroke patients, IRF and HH utilization rates were the highest among the moderate stroke patient group, SNF utilization rates were the highest among the severe stroke patient group, and OP utilization rates were highest among the mild stroke patient group over 1 year of follow up (Table 3A.11). Patients with severe stroke had the highest all cause readmission rates over 1 year of follow up. However, stroke recurrence rates did not differ between the three severity groups over 1 year of follow up. Among FFS beneficiaries, patients with severe stroke had the highest mortality rates and lowest home time over 1 year of follow up.

Table 3.9: Comparison of demographics and stroke characteristics recorded in MiSP between linked and unlinked data stratified by age groups.

Variable		All MiSP (N= 46,330)		<65 years			Absolute standardized difference^	>=65 years			Absolute standardized difference^
		No.	%*	All MiSP (N= 16,192)*	Linked (N= 4,729, 29.2%)*	Unlinked (N= 11,463, 70.8%)*		All MiSP (N= 30,138)*	Linked (N= 19,189, 63.7%)*	Unlinked (N= 10,949, 36.3%)*	
Demographics											
Age	Mean (SD)	46,330	69.7 (14.6)	53.7 (9.1)	55.1 (8.2)	53.1 (9.4)	0.22	78.3 (8.5)	79.0 (8.5)	77.2 (8.4)	0.21
Sex	Female	22,882	49.4	41.7	43.5	41.0	0.05	53.5	55.4	50.2	0.10
	Male	23,448	50.6	58.3	56.5	59.0		46.5	44.6	49.8	
Race**	White	34,993	75.5	66.1	72.2	63.7	0.19	80.6	82.4	77.5	0.12
	Black	8,256	17.8	26.2	21.3	28.3		13.3	12.2	15.4	
	Other	694	1.5	1.8	1.6	1.8		1.3	1.2	1.5	
Ethnicity	Latino	1,881	4.1	5.3	5.8	5.2	0.03	3.4	3.1	3.9	0.04
Characteristics of stroke hospitalization											
Stroke type	Ischemic (%)	38,570	83.3	79.7	80.5	79.4	0.03	85.2	85.1	85.3	<0.01
	Hemorrhagic (%)	7,760	16.7	20.3	19.5	20.6		14.8	14.9	14.7	
Admission NIHSS	Median (IQR)	40,442	3.0 (8.0)	3.0 (7.0)	3.0 (6.0)	3.0 (7.0)	0.04	4.0 (9.0)	4.0 (9.0)	4.0 (9.0)	0.03
Admission duration	Mean (SD)	46,327	5.2 (5.8)	5.9 (7.2)	5.3 (6.0)	6.1 (7.7)	0.12	4.9 (4.8)	4.8 (4.5)	5.0 (5.2)	0.05
Discharge disposition	Home	21,508	46.4	58.2	58.6	58.0	0.06	40.1	38.7	42.6	0.11
	Hospice	2,594	5.6	1.4	1.2	1.4		7.9	8.4	7.0	
	Expired	2,917	6.3	5.3	5.1	5.4		6.8	6.9	6.7	
	Left against medical advice	374	0.8	1.7	1.4	1.8		0.3	0.3	0.3	
	Skilled nursing facility (SNF)	7,689	16.6	8.8	8.2	9.0		20.8	21.3	19.9	
	Inpatient rehabilitation facility (IRF)	9,609	20.7	20.2	20.8	19.9		21.0	21.6	20.1	
	Long term care hospital (LTACH)	661	1.4	1.9	2.1	1.8		1.2	1.2	1.1	
Other	136	0.5	0.4	0.3	0.4	0.5	0.5	0.6			
Past medical history											
Atrial fibrillation/flutter	Yes	9,006	19.4	7.1	8.9	6.3	0.10	26.1	27.3	24.0	0.12
	No	34,989	75.5	88.8	86.8	89.6		68.4	68.1	68.9	
Prosthetic heart valve	Yes	635	1.4	0.9	1.1	0.8	0.03	1.6	1.7	1.4	0.11
	No	43,360	93.6	94.9	94.7	95.0		92.9	93.7	91.5	

Table 3.9 (cont'd)

Coronary artery disease/ prior myocardial infarction	Yes	10,933	23.6	15.5	17.9	14.5	0.09	27.9	28.8	26.4	0.11
	No	33,062	71.4	80.3	77.9	81.3		66.5	66.6	66.5	
Carotid stenosis	Yes	2,176	4.7	2.7	2.9	2.7	0.01	5.8	5.8	5.7	0.11
	No	41,819	90.3	93.1	92.9	93.2		88.7	89.6	87.2	
Diabetes mellitus	Yes	14,318	30.9	29.4	31.6	28.5	0.07	31.7	31.6	31.9	0.11
	No	29,677	64.1	66.4	64.2	67.4		62.8	63.8	61.0	
Peripheral vascular disease	Yes	2,663	5.8	3.7	4.5	3.4	0.06	6.8	7.0	6.5	0.11
	No	41,332	89.2	92.1	91.3	92.4		87.7	88.4	86.5	
Hypertension	Yes	33,253	71.8	62.9	63.1	62.8	0.01	76.5	77.4	75.1	0.11
	No	10,742	23.2	32.9	32.6	33.0		18.0	18.0	17.8	
Smoking	Yes	10,105	21.8	39.5	34.2	41.7	0.16	12.3	11.4	13.8	0.13
	No	33,890	73.2	56.3	61.6	54.2		82.2	84.0	79.1	
Dyslipidemia	Yes	23,427	50.6	38.8	43.1	37.1	0.13	56.9	57.4	56.0	0.11
	No	20,568	44.4	57.0	52.7	58.8		37.6	38.0	37.0	
Heart failure	Yes	5,173	11.2	7.3	8.6	6.8	0.07	13.2	14.0	11.9	0.12
	No	38,822	83.8	88.5	87.2	89.0		81.3	81.4	81.1	
Sickle cell	Yes	31	0.1	0.1	0.1	0.1	0.01	<0.1	<0.1	<0.1	0.11
	No	43,964	94.9	95.7	95.6	95.7		94.5	95.3	92.9	
Previous stroke	Yes	11,630	25.1	22.5	21.7	22.9	0.03	26.5	25.7	27.8	0.12
	No	32,365	69.9	73.3	74.1	73.0		68.0	69.7	65.1	
Previous Transient ischemic attack	Yes	4,293	9.3	6.4	6.7	6.3	0.02	10.8	11.3	9.9	0.11
	No	39,702	85.7	89.4	89.1	89.6		83.7	84.1	83.0	
Drug/alcohol abuse	Yes	4,446	9.6	18.2	13.1	20.2	0.19	5.0	4.3	6.2	0.14
	No	39,549	85.4	77.7	82.7	75.6		89.5	91.1	86.7	
Family history of stroke	Yes	4,854	10.5	11.1	11.9	10.7	0.04	10.2	10.4	9.7	0.11
	No	39,141	84.5	84.8	83.8	85.2		84.3	85.0	83.2	
Hormonal replacement therapy	Yes	236	0.5	0.5	0.7	0.4	0.04	0.5	0.6	0.4	0.11
	No	43,759	94.5	95.3	95.1	95.4		94.0	94.8	92.5	
Migraine	Yes	1,853	4.0	6.7	6.9	6.7	0.01	2.5	2.7	2.3	0.11
	No	42,142	91.0	89.1	88.9	89.2		92.0	92.7	90.6	
Obesity overweight	Yes	19,663	42.4	47.7	50.1	46.7	0.07	39.6	41.2	36.9	0.13
	No	24,332	52.5	48.2	45.7	49.2		54.9	54.2	56.0	
Chronic renal insufficiency	Yes	5,309	11.5	7.6	9.9	6.7	0.12	13.5	14.4	11.9	0.12
	No	38,686	83.5	88.2	85.9	89.2		81.0	81.0	81.0	

Table 3.9 (cont'd)

Sleep apnea	Yes	3,711	8.0	8.2	10.4	7.4	0.11	7.9	8.2	7.4	0.11
	No	40,284	87.0	87.6	85.4	88.5		86.6	87.2	85.5	
Depression	Yes	8,376	18.1	19.4	21.9	18.3	0.09	17.4	17.8	16.7	0.11
	No	35,619	76.9	76.5	73.9	77.6		77.1	77.6	76.3	
Deep vein thrombosis/ pulmonary embolism	Yes	1,060	2.3	2.2	2.1	2.3	0.02	2.3	2.4	2.3	0.11
	No	42,935	92.7	93.6	93.7	93.6		92.2	93.0	90.7	
Familial hypercholesterolemia	Yes	97	0.2	0.2	0.3	0.2	0.02	0.2	0.2	0.2	0.11
	No	43,898	94.8	95.6	95.5	95.7		94.3	95.2	92.7	
Vaping	Yes	15	<0.1	0.1	0.1	0.1	0.02	<0.1	<0.1	<0.1	0.11
	No	43,980	94.9	95.8	95.7	95.8		94.5	95.4	92.9	
Emerging infectious diseases	Yes	54	0.1	0.1	0.1	0.1	0.01	0.1	0.1	0.1	0.11
	No	43,941	94.8	95.7	95.6	95.7		94.4	95.3	92.8	
Dementia	Yes	266	0.6	<0.1	<0.1	<0.1	<0.01	0.9	0.8	0.9	0.11
	No	43,729	94.4	95.8	95.7	95.8		93.6	94.5	92.0	

*Percentages might not add up to 100% due to missing data.

**Racial categories that were less than <1% of the total are not shown. About 5% of the population have missing race designation.

^ A value higher than 0.1 represents a meaningful difference.

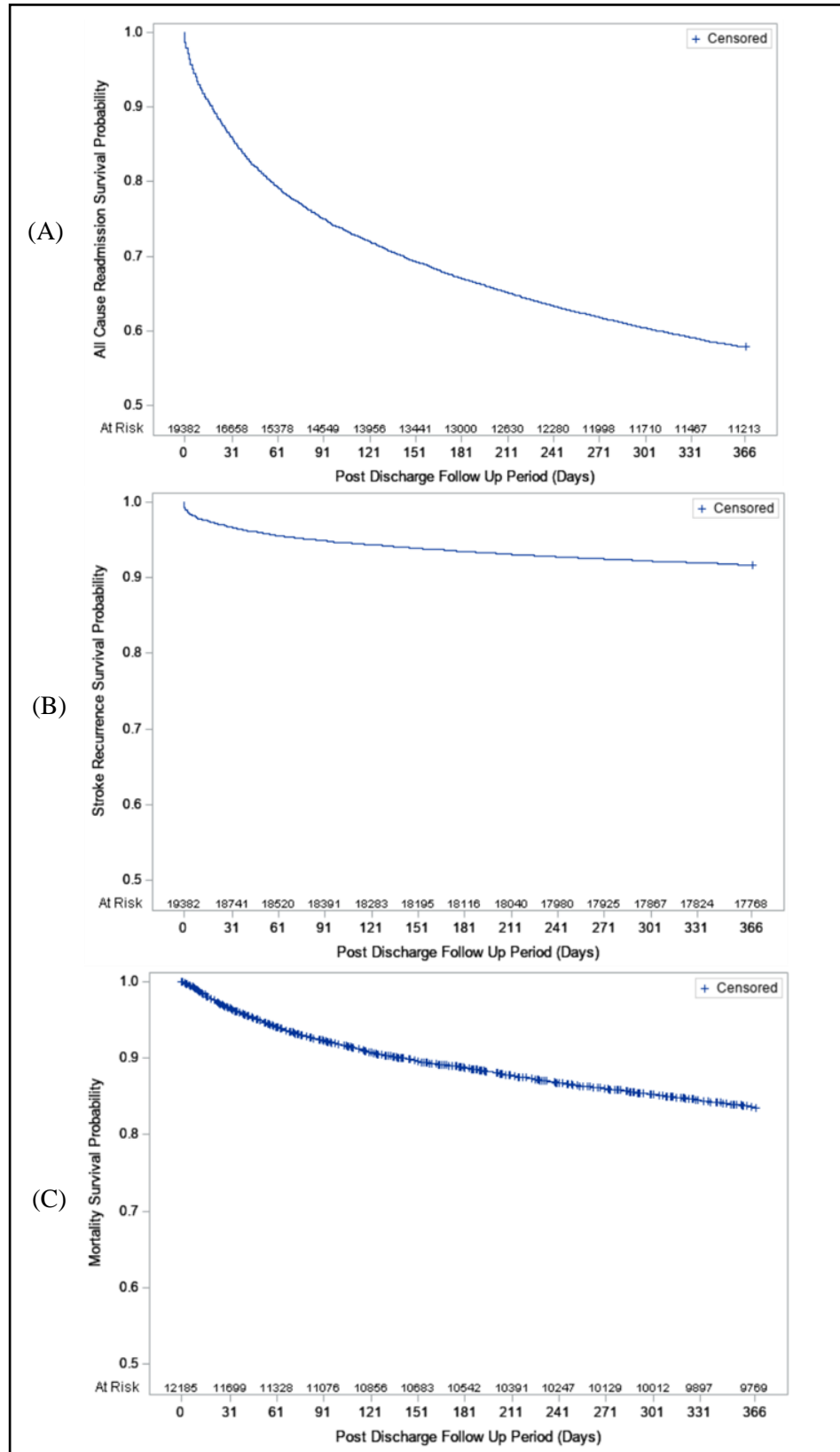
Table 3.10: Thirty-day, 90-day, and 1-year post stroke discharge outcome event rates among linked stroke patients.

Outcome (N=19,382)*	30-day event rate % (n)	90-day event rate % (n)	1-year event rate % (n)
Inpatient rehabilitation facility utilization	24.9 (4,822)	25.5 (4,946)	26.7 (5,171)
Skilled nursing facility utilization	28.1 (5,449)	31.2 (6,049)	34.9 (6,765)
Home health utilization	27.5 (5,336)	38.4 (7,436)	44.7 (8,659)
Outpatient visit	46.4 (8,999)	70.8 (13,720)	85.3 (16,539)
All cause readmission	14.1 (2,724)	24.9 (4,833)	42.2 (8,169)
Stroke recurrence	3.3 (641)	5.1 (991)	8.3 (1,614)
Mortality**	4.0 (486)	9.1 (1,109)	19.8 (2,416)
Home time- Median (IQR)**	22.0 (26.0)	79.0 (40.0)	347.0 (94.0)

* Numerator includes only the first occurrence for a given patient during follow up period for all outcomes but home time.

** Calculated only for Medicare FFS beneficiaries (n= 12,185).

Figure 3.9: Kaplan-Meier curves of time to all cause readmission (9A), stroke recurrence (9B), and mortality (9C).



3.5 Discussion

3.5.1 Evaluation of Linkage Rates and Assessment of Linkage Coverage

We linked a 5-year cohort of stroke discharges entered into the Michigan Stroke registry (MiSP) between 2016-2020 with the Michigan Value Collaborative (MVC) claims database using both deterministic and probabilistic matching techniques, and evaluated the completeness, accuracy, and representativeness of the linkage results by conducting pre linkage qualitative and post linkage quantitative linkage evaluation techniques. The baseline probabilistic linkage using the F-S algorithm resulted in the highest linkage rate, of 51.6% for the MiSP recorded acute stroke discharges and 77.9% of MVC recorded stroke discharges. Our probabilistic linkage rates did not change meaningfully from the deterministic linkage rate of 48.5% because we only made admission and discharge dates flexible by one day and left the rest of the linkage variables unchanged, the high degree of discrimination of the linkage variables when they are used together, and the extensive cleaning process that was conducted to match the inclusion criteria of MVC to MiSP. We could have relaxed the linkage criteria further, but we opted not to because we achieved a satisfactory high MVC linkage rate of 78% and in order to decrease the probability of linkage errors.

Our linkage method included claims data sources other than Medicare FFS. All but one of the previous 10 studies that utilized linked data between GWTG-S and claims data were conducted using only Medicare FFS data.^{12, 20-27} Only three of the nine studies that linked to Medicare FFS reported linkage rates (either directly or provided sufficient data to calculate the rate); the linkage rates were 61.3%,²⁵ 66.4%,²¹ and 69.0%²⁴ which are very similar to the 63.7% linkage rate for the above 65 years old population included in our study. The report by Patorno, et al. that included data between 2008 and 2015 from 11 states is the only study that linked

GWTG-S data with a combination of claims data from commercial health plans and Medicare Advantage members, in addition to Medicare FFS.¹⁹

Most of the prior linkage studies utilized either deterministic^{12, 20-25, 27} or probabilistic²⁶ linking methods with the exception of the Patorno, et al. study which explored the implementation of multiple linkage criteria using both deterministic and probabilistic methods but settled on utilizing a strict deterministic linkage criteria to generate their final dataset.¹⁹ Our study utilized more recent data recorded between 2016-2020, whereas these previous studies utilized GWTG-S data recorded between 2003 and 2015.^{12, 19-27} Patorno, et al. reported a linkage rate of only 5.4%, but this low rate was mainly explained by two phenomena which created a mis-match between the two data sources; the claims dataset used in the linkage had limited coverage in some of the states that mostly contributed to the GWTG-S registry during the study period, and second, there was limited participation of hospitals in the registry in some of the states that were strongly represented in the claims database.¹⁹ Due to scarcity of the published linkage studies that report rates, the differences in the stroke populations under study, the differences in the years of data collected, and the different linkage criteria used, made the comparison of linkage rates between our study and the published literature challenging, except for the Medicare FFS population where we found very similar linkage rates to the 3 prior studies that reported data for this population.^{21, 24, 25}

We conducted a thorough evaluation of our linkage results that included both pre linkage qualitative and post linkage quantitative approaches. Our qualitative approach identified that our linked data set will have selection bias toward patients who are insured by Medicare FFS and BCBSM which limited the generalizability of the generated outcomes to these insurance groups. The qualitative approach also identified several potential linkage errors (i.e., selection bias, and

misclassification or measurement errors) that could stem from our linkage structure, inclusion criteria, and nature of our non-unique linkage variables. These errors could in turn negatively affect the interpretation of the stroke outcomes data presented for the final linked data. Our post linkage quantitative methods evaluated the completeness, accuracy, and representativeness of the linked dataset relative to MiSP dataset. None of the previously published GWTG-S studies that included linked claims data conducted a comprehensive evaluation of the linkage results; specifically, none of the 10 studies conducted pre linkage qualitative evaluation,^{12, 20-27} and only 3 studies conducted post linkage quantitative evaluation by comparing the characteristics of the linked and unlinked populations.^{19, 21, 25} At the end of our qualitative and quantitative linkage evaluations we concluded that selection bias was unavoidable due to limitations in insurance coverage by MVC which resulted in restricting the generalizability of our findings to BCBSM and Medicare FFS population. However, we believe that misclassification bias ended up not being of concern because of the reassuring findings from the quantitative linkage evaluations mentioned in the next paragraph and due to the fact that the calculated outcomes event rates from the linked data were similar to rates in the published literature.

Our rigorous evaluation of the completeness of the data linkage included, first, a sensitivity analysis using multiple different combinations of linkage variables as part of the deterministic linkage found that the linkage rates did not change much when different linkage variables were used. Second, we conducted probabilistic linkage using different subsets of the data excluding hospitals that sampled stroke discharges (rather than collection a complete census) or excluding data from the 2020 COVID-19 pandemic year. Linkage results - as reflected by the overall MiSP linkage rate - were again robust to these different data manipulations. Third, we determined that our linkage rate of 51.6% did not surpass the

maximum plausible linkage rate of 66.2%. These results indicate that our linkage rates are within expected bounds, that our probabilistic data linkage criterion was highly discriminative, and that our linkage results are free of implausible links. Our evaluation of accuracy of linkage using quantitative techniques was limited to the calculation of the false positive linkage rate among the <65 age group using only 204 Medicaid cases, but we found only one false positive link for a false positive rate of 0.5% which if applied to the total match would imply a specificity of 99.5%. However, this calculated specificity figure should be interpreted with caution given the small pool of negative controls that could be used. This fact in addition to the fact that the dataset did not include any positive controls prevented the calculation of sensitivity, PPV and NPV accuracy measures.

Our evaluation of representativeness of the linkage results by comparing characteristics between linked and unlinked cases found that <65 years old age group linked and unlinked populations were more balanced compared to >=65 years old age group, despite a much lower overall match rate (29.2% vs 63.7%). The discrepancy among the linked and unlinked population among the >65 years old population in our study may be due to the limited representation of MA plans in the MVC claims data (27.4%) which naturally translates to a lower chance of linkage among MA insured MiSP population compared to FFS population. In 2020, about 48% of the above 65 years old population was covered by MA plans.⁵⁴ The lower representation of MA population among our linked population compared to the unlinked population is important because it is known that MA population tend to be healthier than FFS, have a higher proportion of non white racial groups, and have a higher proportion of women.⁵⁵ These differences that are likely driven by MA data were observed when we compared the characteristics of the linked to the unlinked >=65 years old population where the linked population was more white (82.4% vs

77.5%) and carried a higher burden of almost all comorbidities (Table 3.9). Compared to our findings, Patorno, et al also found a high number of characteristics with large differences (standardized difference >0.1) between linked and unlinked cases among the >=65 years old when compared to <65 years old.¹⁹ Of the 9 prior linkages that used Medicare FFS data, only Kaufman, et al. study compared characteristics between linked and unlinked cases finding almost perfect balance between the two populations.²⁵ This might be attributed to the fact that the Kaufman, et al. study, in contrast to Patorno, et al. and our study, only included Medicare FFS claims data and did not include Medicare Advantage data.^{19, 25}

3.5.2 Healthcare Utilization, Readmission, and Recurrent Stroke Outcomes

3.5.2.1 SNF, IRF, HH, and OP Post Acute Care Utilization

Overall, the stroke outcome event rates that were calculated after linkage were generally similar to previously published rates in the literature. We found that within 30-days post discharge 24.9%, 28.1%, 27.5% of stroke patients utilized IRF, SNF, and home health care at least once, respectively. However, these data are different from the data from a nationally representative GWTG-S population collected between 2003 and 2011 that reported that 25.4%, 19.5%, 11.5% were discharged to IRF, SNF, and HH post-acute care services, respectively.⁵⁶ This is not a fair comparison because this data is relatively old, does not reflect 30-day utilization, and only reports the initial discharge destination. However, this could indicate that these services are increasingly utilized post discharge and that many patients who are discharged without these services end up using them on the short term. Even though Medicare permits entering SNFs up to 30-days post discharge without the need for another inpatient hospital stay, we could not locate other nationally representative US-based studies that reported on 30-day utilization rates of post-acute care services after discharge for stroke.⁵⁷

Within 30-days, 90-days and 1-year post discharge, 46.4%, 70.8%, and 85.3% of the linked population, respectively had at least one outpatient visit. We find these rates are within expectation because in our population for example in the first 30-days more than half of patients were receiving IRF or SNF care, 15% were readmitted, and some patients died, all of which prevents patients from attending to their first follow up appointment. Our rates were close to utilization rates reported by a nationally representative claims database of commercially insured Americans from 2009 to 2015, where 59.3% and 70.8% of acute stroke patients had a primary care visit within 30 and 90 days post discharge, respectively, and 24.4% and 41.8% had a neurology visit within 30 and 90 days post discharge, respectively.⁵⁸

When we stratified the post-acute care utilization rates by age, race, stroke type and stroke severity, we found that IRF utilization was significantly higher among Black patients, patients who had a hemorrhagic stroke, and their stroke etiology was moderate. Whereas SNF utilization was significantly higher among older patients (≥ 65 years old), Black patients, hemorrhagic stroke patients, and patients with a severe stroke etiology. For HH, patients who were older (≥ 65 years old), Black, and had a moderate stroke etiology had the highest rate of HH utilization. We also found that patients who were older (≥ 65 years old), White, and with mild strokes utilized outpatient care at higher rate. These findings shed light at the importance of data linkage in the ability to stratify outcomes generated through administrative claims data using variables that are usually not available in claims data including race and stroke severity. It also emphasize the effect of stroke heterogeneity on the utilization rates of post-acute services, highlighting the importance of developing patient specific post discharge follow up plan by healthcare systems that ensures the delivery and compliance with the patient post-acute care.

3.5.2.2 Post Discharge Readmission and Recurrence

Our 30-day and 1-year readmission rates of 14.1% and 42.2% were similar to that reported by a recent meta-analysis that found pooled 30-day and 1-year all-cause stroke readmission rates of 17.4% and 42.5%, respectively.⁵⁹ We found that older patients (≥ 65 years old) versus younger (< 65 years old) patients, Black versus White patients, hemorrhagic versus ischemic stroke patients, and patient with severe versus mild or moderate strokes were readmitted at significant higher rates over 1 year of follow up.

Our 30-day, 90-day, and 1-year stroke recurrence rates of 3.3%, 5.1%, and 8.3%, respectively were similar to a report that examined 2017 Medicare FFS data on 30-day, 90-day, and 1-year ischemic stroke and found recurrence rates of 2.4%, 4.0%, and 7.6%, respectively.⁶⁰ We found that younger patients (< 65 years old) versus older (≥ 65 years old) patients, Black versus White patients, and hemorrhagic versus ischemic stroke patients had stroke recurrence at significantly higher rates over 1 year of follow up. However, we did not find difference in stroke recurrence rates according to stroke severity.

These findings show how readmissions and recurrence data can be generated by linked data and potentially be utilized to study risk factors for readmission and recurrence over long follow up times. The linked data could also be utilized to develop system or hospital specific prediction models that can be implemented to reduce stroke recurrence and preventable readmissions which ultimately leads to improvements in healthcare quality of care.

3.5.2.3 Mortality and Home Time

Our 30-day, 90-day, and 1-year post discharge mortality rates of 4.0%, 9.1%, and 19.8% were similar to rates reported by a study among Medicare FFS population with minor stroke, where 30-day, 90-day and 1-year all-cause mortality were reported as 3.7%, 7.6%, and 17.2%,

respectively.²⁶ We found that hemorrhagic versus ischemic stroke patients, and patients with severe versus mild or moderate strokes had a higher mortality rates over 1 year of follow up. We did not find difference in mortality rates according to race.

The calculated median 90-day and 1-year median home time estimates of 79.0 and 347.0 days were similar to data from Medicare FFS population that were reported in two studies: the median 90-day and 1-year post ischemic stroke home time value were 59.5 and 79.0 days and 270.2 and 349.0 days, in each of these two studies respectively.^{23, 61} We found that Black versus white patients, hemorrhagic versus ischemic stroke patients, and patients with severe versus mild or moderate strokes had lower median home time over 1 year of follow up.

Although these findings were calculated only for Medicare FFS beneficiaries, the calculation of home time could be used to track functional recovery post discharge over longer periods of follow up, and to compare the effectiveness of rehabilitation care in different settings.

3.5.3 Strengths and Limitations

One of the major strengths of this linkage study was the use of data recorded over a recent 5-year time period (2016-2020) from 31 hospitals in Michigan. Our linked population included patients that were insured by Medicare FFS and BCBS (the largest health insurer in the state) private and Medicare Advantage plans. This gave us the opportunity to include patients that are younger than 65 years- an understudied stroke patient population. We also reported on a comprehensive set of stroke outcome events that are rarely reported on in the literature (i.e., outpatient clinic utilization and home time) and that have been mostly reported on by Medicare FFS-based populations. In addition, most of our stroke outcome rates were similar to the published literature. Compared to our linkage study, none of the previously conducted linkage studies that utilized GWTG-S data included a thorough linkage evaluation using pre linkage

qualitative and post linkage quantitative techniques. Finally, our multiple sensitivity analysis ascertained the high discrimination power of our linkage and assessed the magnitude of linkage errors that could arise from sampling hospitals and due to the COVID-19 pandemic year.

Our study had several limitations. Due to limitations in data availability from MiSP registry we only included data from 31 of the 40 hospitals that were participating in the registry in 2016. All of the 31 included hospitals were stroke accredited and so we feel that our sample adequately represent the 49 total stroke accredited hospitals (i.e., PSC, TSR, and CSC) in Michigan. This representativeness however might not be extrapolated to non-stroke certified hospitals in Michigan due to well recognized differences between certified and non-certified stroke centers. The absence of Medicaid, private insurance plans other than BCBS, Medicare Advantage plans other than BCBS, and uninsured populations among MVC claims data insurance coverage limits extrapolation of findings to these other population subgroups. The MVC limitations of insurance coverage is likely the cause of the unbalanced nature of our linked and unlinked populations among the ≥ 65 years old group. Despite our desire to implement rigorous linkage evaluation techniques, we were limited to the availability of a small number of negative controls that provided. The lack of positive controls, a linkage gold standard made it impossible to report on a more comprehensive set of linkage accuracy measures. Finally, as mentioned earlier, the comparison of characteristics between the linked and unlinked populations was limited by the fact that neither dataset was expected to completely match (match rate =100%).

3.5.4 Future Directions

In this study, we included data from 31 hospitals with a patient population limited by insurance coverage to Medicare and BCBSM. Therefore, expansion to other hospitals and major

insurance providers including Medicaid and other private insurers (i.e., Health Alliance Plan (Henry Ford Health System), Priority Health, and United Health) would generate a more generalizable linked dataset. Additionally, to improve linkage rates, MVC data structure and inclusion criteria should be amended to include all stroke related ICD-10 codes (I61-I63). On the national level, to improve linkage rates future linkage attempts should use the same set of case inclusion and exclusion criteria for each dataset in order to avoid data coverage discrepancies. Further, external validation of the linkage should take place within Michigan in the event that unique personal identifiers or a gold standard MiSP and MVC linked datasets are available. Lastly, future studies should conduct detailed analysis of post discharge healthcare utilization, stroke recurrence, and home time to provide evidence of the effect of healthcare utilization on secondary prevention of strokes, study the drivers of the differences in utilization among different age, race, stroke type and stroke severity groups, predict post stroke readmission and recurrence, and investigate the comparative effectiveness of IRF vs SNF on functional outcomes post stroke using home time. All the previous points should overcome some of the limitations, reduce bias, produce an externally valid and generalizable linked data, and provide an important insight into inpatient and post discharge stroke care that can lead to improvements in stroke care and patient outcomes.

3.6 Conclusions

Probabilistic linkage between MiSP acute stroke registry and MVC claims database using indirect identifiers produced a valid linked dataset that has acceptable representation of Medicare FFS and BCBS insured population in Michigan. This linkage allowed acute stroke care data from a statewide registry to be combined with longitudinal claims data that permitted reporting on several stroke outcomes event rates up to 1-year post discharge. Generating data on long term

outcomes post discharge can provide important insight into inpatient and post discharge health care evaluations for health systems, health insurance providers, health policy makers, and hospital staff ultimately leading to improvements in stroke care and outcomes.

Because matching techniques contain inherent limitations that should be fully understood and disclosed in analysis of linked data, linkage evaluation techniques presented in this study can serve as an example to guide future linkages studies using GWTG-S data (or other stroke registries) with claims data from Medicare FFS and other insurance providers. We found that conducting qualitative pre linkage evaluation (e.g., identifying the linkage structure) was an important step that allows for identification of sources of bias before data linkage was undertaken. In addition, it is very important to determine data limitations and types of post linkage quantitative evaluation techniques (e.g., linking using different linkage criteria and on subsets of the data, calculating linkage accuracy measures, and comparing linked and unlinked populations) that could be implemented in evaluating completeness, accuracy, and representativeness of the linked data. Future studies should explore using claims data with broader coverage and externally validate linkage results preferably using personal identifiers as a gold standard.

BIBLIOGRAPHY

1. In *Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition, Addendum 2*, Gliklich, R. E.; Leavy, M. B.; Dreyer, N. A., Eds. Rockville (MD), 2019.
2. In *Registries for Evaluating Patient Outcomes: A User's Guide*, 3rd ed.; Gliklich, R. E.; Dreyer, N. A.; Leavy, M. B., Eds. Rockville (MD), 2014.
3. American Heart Association, Get With The Guide Line - Stroke. <https://www.heart.org/en/professional/quality-improvement/get-with-the-guidelines/get-with-the-guidelines-stroke/get-with-the-guidelines-stroke-overview> (accessed 2023).
4. Center of Disease Control and Prevention, Paul Coverdell National Acute Stroke Program. https://www.cdc.gov/dhdsp/programs/stroke_registry.htm (accessed 2023).
5. Schwamm, L. H.; Reeves, M. J.; Pan, W.; Smith, E. E.; Frankel, M. R.; Olson, D.; Zhao, X.; Peterson, E.; Fonarow, G. C., Race/ethnicity, quality of care, and outcomes in ischemic stroke. *Circulation* **2010**, *121* (13), 1492-501. <https://doi.org/10.1161/CIRCULATIONAHA.109.881490>.
6. George, M. G.; Tong, X.; McGruder, H.; Yoon, P.; Rosamond, W.; Winkvist, A.; Hinchey, J.; Wall, H. K.; Pandey, D. K.; Centers for Disease, C.; Prevention, Paul Coverdell National Acute Stroke Registry Surveillance - four states, 2005-2007. *MMWR Surveill Summ* **2009**, *58* (7), 1-23.
7. Parker, C.; Schwamm, L. H.; Fonarow, G. C.; Smith, E. E.; Reeves, M. J., Stroke quality metrics: systematic reviews of the relationships to patient-centered outcomes and impact of public reporting. *Stroke* **2012**, *43* (1), 155-62. <https://doi.org/10.1161/STROKEAHA.111.635011>.
8. Howard, G.; Schwamm, L. H.; Donnelly, J. P.; Howard, V. J.; Jasne, A.; Smith, E. E.; Rhodes, J. D.; Kissela, B. M.; Fonarow, G. C.; Kleindorfer, D. O.; Albright, K. C., Participation in Get With The Guidelines-Stroke and Its Association With Quality of Care for Stroke. *JAMA Neurol* **2018**, *75* (11), 1331-1337. <https://doi.org/10.1001/jamaneurol.2018.2101>.
9. Fonarow, G. C.; Smith, E. E.; Saver, J. L.; Reeves, M. J.; Hernandez, A. F.; Peterson, E. D.; Sacco, R. L.; Schwamm, L. H., Improving door-to-needle times in acute ischemic stroke: the design and rationale for the American Heart Association/American Stroke Association's Target: Stroke initiative. *Stroke* **2011**, *42* (10), 2983-9. <https://doi.org/10.1161/STROKEAHA.111.621342>.
10. Heidenreich, P. A.; Hernandez, A. F.; Yancy, C. W.; Liang, L.; Peterson, E. D.; Fonarow, G. C., Get With The Guidelines program participation, process of care, and outcome for Medicare patients hospitalized with heart failure. *Circ Cardiovasc Qual Outcomes* **2012**, *5* (1), 37-43. <https://doi.org/10.1161/CIRCOUTCOMES.110.959122>.

11. American Heart Association, Get With The Guidelines® - Stroke Patient Management Tool. <https://www.heart.org/en/professional/quality-improvement/get-with-the-guidelines/get-with-the-guidelines-stroke/get-with-the-guidelines-stroke-patient-management-tool>.
12. Song, S.; Fonarow, G. C.; Olson, D. M.; Liang, L.; Schulte, P. J.; Hernandez, A. F.; Peterson, E. D.; Reeves, M. J.; Smith, E. E.; Schwamm, L. H.; Saver, J. L., Association of Get With The Guidelines-Stroke Program Participation and Clinical Outcomes for Medicare Beneficiaries With Ischemic Stroke. *Stroke* **2016**, *47* (5), 1294-302. <https://doi.org/10.1161/STROKEAHA.115.011874>.
13. Ormseth, C. H.; Sheth, K. N.; Saver, J. L.; Fonarow, G. C.; Schwamm, L. H., The American Heart Association's Get With the Guidelines (GWTG)-Stroke development and impact on stroke care. *Stroke Vasc Neurol* **2017**, *2* (2), 94-105. <https://doi.org/10.1136/svn-2017-000092>.
14. Lee, K. B.; Lim, S. H.; Kim, K. H.; Kim, K. J.; Kim, Y. R.; Chang, W. N.; Yeom, J. W.; Kim, Y. D.; Hwang, B. Y., Six-month functional recovery of stroke patients: a multi-time-point study. *Int J Rehabil Res* **2015**, *38* (2), 173-80. <https://doi.org/10.1097/MRR.0000000000000108>.
15. Reeves, M.; Lisabeth, L.; Williams, L.; Katzan, I.; Kapral, M.; Deutsch, A.; Prvu-Bettger, J., Patient-Reported Outcome Measures (PROMs) for Acute Stroke: Rationale, Methods and Future Directions. *Stroke* **2018**, *49* (6), 1549-1556. <https://doi.org/10.1161/STROKEAHA.117.018912>.
16. Yu, A. Y.; Holodinsky, J. K.; Zerna, C.; Svenson, L. W.; Jette, N.; Quan, H.; Hill, M. D., Use and Utility of Administrative Health Data for Stroke Research and Surveillance. *Stroke* **2016**, *47* (7), 1946-52. <https://doi.org/10.1161/STROKEAHA.116.012390>.
17. Reker, D. M.; Reid, K.; Duncan, P. W.; Marshall, C.; Cowper, D.; Stansbury, J.; Warr-Wing, K. L., Development of an integrated stroke outcomes database within Veterans Health Administration. *J Rehabil Res Dev* **2005**, *42* (1), 77-91. <https://doi.org/10.1682/jrrd.2003.11.0164>.
18. Deutsch, A.; Granger, C. V.; Heinemann, A. W.; Fiedler, R. C.; DeJong, G.; Kane, R. L.; Ottenbacher, K. J.; Naughton, J. P.; Trevisan, M., Poststroke rehabilitation: outcomes and reimbursement of inpatient rehabilitation facilities and subacute rehabilitation programs. *Stroke* **2006**, *37* (6), 1477-82. <https://doi.org/10.1161/01.STR.0000221172.99375.5a>.
19. Paterno, E.; Schneeweiss, S.; George, M. G.; Tong, X.; Franklin, J. M.; Pawar, A.; Mogun, H.; Moura, L.; Schwamm, L. H., Linking the Paul Coverdell National Acute Stroke Program to commercial claims to establish a framework for real-world longitudinal stroke research. *Stroke Vasc Neurol* **2022**, *7* (2), 114-123. <https://doi.org/10.1136/svn-2021-001134>.
20. Xian, Y.; Wu, J.; O'Brien, E. C.; Fonarow, G. C.; Olson, D. M.; Schwamm, L. H.; Bhatt, D. L.; Smith, E. E.; Suter, R. E.; Hannah, D.; Lindholm, B.; Maisch, L.; Greiner, M. A.; Lytle,

B. L.; Pencina, M. J.; Peterson, E. D.; Hernandez, A. F., Real world effectiveness of warfarin among ischemic stroke patients with atrial fibrillation: observational analysis from Patient-Centered Research into Outcomes Stroke Patients Prefer and Effectiveness Research (PROSPER) study. *BMJ* **2015**, *351*, h3786. <https://doi.org/10.1136/bmj.h3786>.

21. Reeves, M. J.; Fonarow, G. C.; Smith, E. E.; Pan, W.; Olson, D.; Hernandez, A. F.; Peterson, E. D.; Schwamm, L. H., Representativeness of the Get With The Guidelines-Stroke Registry: comparison of patient and hospital characteristics among Medicare beneficiaries hospitalized with ischemic stroke. *Stroke* **2012**, *43* (1), 44-9. <https://doi.org/10.1161/STROKEAHA.111.626978>.

22. O'Brien, E. C.; Greiner, M. A.; Xian, Y.; Fonarow, G. C.; Olson, D. M.; Schwamm, L. H.; Bhatt, D. L.; Smith, E. E.; Maisch, L.; Hannah, D.; Lindholm, B.; Peterson, E. D.; Pencina, M. J.; Hernandez, A. F., Clinical Effectiveness of Statin Therapy After Ischemic Stroke: Primary Results From the Statin Therapeutic Area of the Patient-Centered Research Into Outcomes Stroke Patients Prefer and Effectiveness Research (PROSPER) Study. *Circulation* **2015**, *132* (15), 1404-13. <https://doi.org/10.1161/CIRCULATIONAHA.115.016183>.

23. Fonarow, G. C.; Liang, L.; Thomas, L.; Xian, Y.; Saver, J. L.; Smith, E. E.; Schwamm, L. H.; Peterson, E. D.; Hernandez, A. F.; Duncan, P. W.; O'Brien, E. C.; Bushnell, C.; Prvu Bettger, J., Assessment of Home-Time After Acute Ischemic Stroke in Medicare Beneficiaries. *Stroke* **2016**, *47* (3), 836-42. <https://doi.org/10.1161/STROKEAHA.115.011599>.

24. Fonarow, G. C.; Smith, E. E.; Reeves, M. J.; Pan, W.; Olson, D.; Hernandez, A. F.; Peterson, E. D.; Schwamm, L. H.; Get With The Guidelines Steering, C.; Hospitals, Hospital-level variation in mortality and rehospitalization for medicare beneficiaries with acute ischemic stroke. *Stroke* **2011**, *42* (1), 159-66. <https://doi.org/10.1161/STROKEAHA.110.601831>.

25. Kaufman, B. G.; O'Brien, E. C.; Stearns, S. C.; Matsouaka, R.; Holmes, G. M.; Weinberger, M.; Song, P. H.; Schwamm, L. H.; Smith, E. E.; Fonarow, G. C.; Xian, Y., The Medicare Shared Savings Program and Outcomes for Ischemic Stroke Patients: a Retrospective Cohort Study. *J Gen Intern Med* **2019**, *34* (12), 2740-2748. <https://doi.org/10.1007/s11606-019-05283-1>.

26. Kaufman, B. G.; Shah, S.; Hellkamp, A. S.; Lytle, B. L.; Fonarow, G. C.; Schwamm, L. H.; Lesen, E.; Hedberg, J.; Tank, A.; Fita, E.; Bhalla, N.; Atreja, N.; Bettger, J. P., Disease Burden Following Non-Cardioembolic Minor Ischemic Stroke or High-Risk TIA: A GWTG-Stroke Study. *J Stroke Cerebrovasc Dis* **2020**, *29* (12), 105399. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2020.105399>.

27. Reeves, M. J.; Fonarow, G. C.; Xu, H.; Matsouaka, R. A.; Xian, Y.; Saver, J.; Schwamm, L.; Smith, E. E., Is Risk-Standardized In-Hospital Stroke Mortality an Adequate Proxy for Risk-Standardized 30-Day Stroke Mortality Data? Findings From Get With The Guidelines-Stroke. *Circ*

28. Lichtman, J. H.; Leifheit-Limson, E. C.; Goldstein, L. B., Centers for medicare and medicaid services medicare data and stroke research: goldmine or landmine? *Stroke* **2015**, *46* (2), 598-604. <https://doi.org/10.1161/STROKEAHA.114.003255>.

29. Ekker, M. S.; Boot, E. M.; Singhal, A. B.; Tan, K. S.; Debette, S.; Tuladhar, A. M.; de Leeuw, F. E., Epidemiology, aetiology, and management of ischaemic stroke in young adults. *Lancet Neurol* **2018**, *17* (9), 790-801. [https://doi.org/10.1016/S1474-4422\(18\)30233-3](https://doi.org/10.1016/S1474-4422(18)30233-3).

30. Hall, M. J.; Levant, S.; DeFrances, C. J., Hospitalization for stroke in U.S. hospitals, 1989-2009. *NCHS Data Brief* **2012**, (95), 1-8.

31. Bejot, 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* (5). <https://doi.org/10.1161/JAHA.116.003661>.

32. Maaijwee, N. A.; Rutten-Jacobs, L. C.; Arntz, R. M.; Schaapsmeeders, P.; Schoonderwaldt, H. C.; van Dijk, E. J.; de Leeuw, F. E., Long-term increased risk of unemployment after young stroke: a long-term follow-up study. *Neurology* **2014**, *83* (13), 1132-8. <https://doi.org/10.1212/WNL.0000000000000817>.

33. Centers for Medicare and Medicaid Services, Medicare and Medicaid enrollment reports. <https://data.cms.gov/summary-statistics-on-beneficiary-enrollment/medicare-and-medicaid-reports/medicare-monthly-enrollment>.

34. Michigan Department of Health and Human Services (MDHHS), Michigan Stroke Program (MiSP). <https://www.michigan.gov/mdhhs/keep-mi-healthy/communicablediseases/epidemiology/chronicepi/stroke> (accessed 2023).

35. American Heart Association, Get With The Guidelines® - Stroke Case Record Form. <https://www.heart.org/-/media/Files/Professional/Quality-Improvement/Get-With-the-Guidelines/Get-With-The-Guidelines-Stroke/Stroke--Diabetes-CRFJuly21.pdf>.

36. Michigan Value Collaborative, MVC Data Resources. <https://michiganvalue.org/resources-2/> (accessed 2023).

37. American Hospital Association, AHA Annual Survey Database. <https://www.ahadata.com/>.

38. National Institute of Health HIPPA privacy rule. https://privacyruleandresearch.nih.gov/pr_08.asp (accessed 2023).

39. Centers for Medicare and Medicaid Services and Agency for Healthcare Research and Quality, 2016 Measure Information About The 30-Day All-Cause Hospital Readmission Measure, Calculated For The 2018 Value-Based Payment Modifier Program. <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2016-ACR-MIF.pdf>.
40. Hammill, B. G.; Hernandez, A. F.; Peterson, E. D.; Fonarow, G. C.; Schulman, K. A.; Curtis, L. H., Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. *Am Heart J* **2009**, *157* (6), 995-1000. <https://doi.org/10.1016/j.ahj.2009.04.002>.
41. Mao, J.; Moore, K. O.; Columbo, J. A.; Mehta, K. S.; Goodney, P. P.; Sedrakyan, A., Validation of an indirect linkage algorithm to combine registry data with Medicare claims. *J Vasc Surg* **2022**, *76* (1), 266-271 e2. <https://doi.org/10.1016/j.jvs.2022.01.132>.
42. Blakely, T.; Salmond, C., Probabilistic record linkage and a method to calculate the positive predictive value. *Int J Epidemiol* **2002**, *31* (6), 1246-52. <https://doi.org/10.1093/ije/31.6.1246>.
43. Sayers, A.; Ben-Shlomo, Y.; Blom, A. W.; Steele, F., Probabilistic record linkage. *Int J Epidemiol* **2016**, *45* (3), 954-64. <https://doi.org/10.1093/ije/dyv322>.
44. Doidge, J. C.; Harron, K. L., Reflections on modern methods: linkage error bias. *Int J Epidemiol* **2019**, *48* (6), 2050-2060. <https://doi.org/10.1093/ije/dyz203>.
45. Harron, K. L.; Doidge, J. C.; Knight, H. E.; Gilbert, R. E.; Goldstein, H.; Cromwell, D. A.; van der Meulen, J. H., A guide to evaluating linkage quality for the analysis of linked data. *Int J Epidemiol* **2017**, *46* (5), 1699-1710. <https://doi.org/10.1093/ije/dyx177>.
46. Goldstein, H.; Harron, K.; Wade, A., The analysis of record-linked data using multiple imputation with data value priors. *Stat Med* **2012**, *31* (28), 3481-93. <https://doi.org/10.1002/sim.5508>.
47. Harron, K.; Doidge, J. C.; Goldstein, H., Assessing data linkage quality in cohort studies. *Ann Hum Biol* **2020**, *47* (2), 218-226. <https://doi.org/10.1080/03014460.2020.1742379>.
48. Chipperfield, J., A weighting approach to making inference with probabilistically linked data. *Statistica Neerlandica* **2019**, *73* (3), 333-350. <https://doi.org/10.1111/stan.12172>.
49. Brown, A. P.; Randall, S. M.; Ferrante, A. M.; Semmens, J. B.; Boyd, J. H., Estimating parameters for probabilistic linkage of privacy-preserved datasets. *BMC Med Res Methodol* **2017**, *17* (1), 95. <https://doi.org/10.1186/s12874-017-0370-0>.

50. Grannis, S. J.; Overhage, J. M.; Hui, S.; McDonald, C. J., Analysis of a probabilistic record linkage technique without human review. *AMIA Annu Symp Proc* **2003**, 2003, 259-63.
51. Austin, P. C., An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* **2011**, 46 (3), 399-424. <https://doi.org/10.1080/00273171.2011.568786>.
52. Yang, D., & Dalton, A unified approach to measuring the effect size between two groups using SAS. *SAS Global Forum 2012* **2012**.
53. Chun, M.; Qin, H.; Turnbull, I.; Sansome, S.; Gilbert, S.; Hacker, A.; Wright, N.; Zhu, T.; Clifton, D.; Bennett, D.; Guo, Y.; Pei, P.; Lv, J.; Yu, C.; Yang, L.; Li, L.; Lu, Y.; Chen, Z.; Cairns, B. J.; Chen, Y.; Clarke, R., Heterogeneity in the diagnosis and prognosis of ischemic stroke subtypes: 9-year follow-up of 22,000 cases in Chinese adults. *Int J Stroke* **2023**, 18 (7), 847-855. <https://doi.org/10.1177/17474930231162265>.
54. Kaiser Family Foundation, Health Insurance Coverage of the Total Population by State. <https://www.kff.org/other/state-indicator/total-population/?currentTimeframe=2&selectedRows=%7B%22states%22:%7B%22michigan%22:%7B%7D%7D%7D&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D>.
55. Prusynski, R. A.; D'Alonzo, A.; Johnson, M. P.; Mroz, T. M.; Leland, N. E., Differences in Home Health Services and Outcomes Between Traditional Medicare and Medicare Advantage. *JAMA Health Forum* **2024**, 5 (3), e235454. <https://doi.org/10.1001/jamahealthforum.2023.5454>.
56. Prvu Bettger, J.; McCoy, L.; Smith, E. E.; Fonarow, G. C.; Schwamm, L. H.; Peterson, E. D., Contemporary trends and predictors of postacute service use and routine discharge home after stroke. *J Am Heart Assoc* **2015**, 4 (2). <https://doi.org/10.1161/JAHA.114.001038>.
57. Coberly, S., In *Medicare's Post-Acute Care Payment: An Updated Review of the Issues and Policy Proposals*, Washington (DC), 2015.
58. Leppert, M. H.; Sillau, S.; Lindrooth, R. C.; Poisson, S. N.; Campbell, J. D.; Simpson, J. R., Relationship between early follow-up and readmission within 30 and 90 days after ischemic stroke. *Neurology* **2020**, 94 (12), e1249-e1258. <https://doi.org/10.1212/WNL.00000000000009135>.
59. Zhong, W.; Geng, N.; Wang, P.; Li, Z.; Cao, L., Prevalence, causes and risk factors of hospital readmissions after acute stroke and transient ischemic attack: a systematic review and meta-analysis. *Neurol Sci* **2016**, 37 (8), 1195-202. <https://doi.org/10.1007/s10072-016-2570-5>.

60. Leifheit, E. C.; Wang, Y.; Goldstein, L. B.; Lichtman, J. H., Trends in 1-Year Recurrent Ischemic Stroke in the US Medicare Fee-for-Service Population. *Stroke* **2022**, *53* (11), 3338-3347. <https://doi.org/10.1161/STROKEAHA.122.039438>.
61. O'Brien, E. C.; Xian, Y.; Xu, H.; Wu, J.; Saver, J. L.; Smith, E. E.; Schwamm, L. H.; Peterson, E. D.; Reeves, M. J.; Bhatt, D. L.; Maisch, L.; Hannah, D.; Lindholm, B.; Olson, D.; Prvu Bettger, J.; Pencina, M.; Hernandez, A. F.; Fonarow, G. C., Hospital Variation in Home-Time After Acute Ischemic Stroke: Insights From the PROSPER Study (Patient-Centered Research Into Outcomes Stroke Patients Prefer and Effectiveness Research). *Stroke* **2016**, *47* (10), 2627-33. <https://doi.org/10.1161/STROKEAHA.116.013563>.

APPENDIX

Table 3A.1: Listing of all ICD-10 stroke codes and whether the codes is included in the Michigan Value Collaborative definition of acute stroke.

Stroke type	ICD-10 code	Full Description	Included in MVC definition
Hemorrhagic (Subarachnoid)	I6000	Nontraumatic subarachnoid hemorrhage from unspecified carotid siphon and bifurcation	Yes
	I6001	Nontraumatic subarachnoid hemorrhage from right carotid siphon and bifurcation	Yes
	I6002	Nontraumatic subarachnoid hemorrhage from left carotid siphon and bifurcation	Yes
	I6010	Nontraumatic subarachnoid hemorrhage from unspecified middle cerebral artery	Yes
	I6011	Nontraumatic subarachnoid hemorrhage from right middle cerebral artery	Yes
	I6012	Nontraumatic subarachnoid hemorrhage from left middle cerebral artery	Yes
	I602	Nontraumatic subarachnoid hemorrhage from anterior communicating artery	No
	I6020	Nontraumatic subarachnoid hemorrhage from unspecified anterior communicating artery	Yes
	I6021	Nontraumatic subarachnoid hemorrhage from right anterior communicating artery	Yes
	I6022	Nontraumatic subarachnoid hemorrhage from left anterior communicating artery	Yes
	I6030	Nontraumatic subarachnoid hemorrhage from unspecified posterior communicating artery	Yes
	I6031	Nontraumatic subarachnoid hemorrhage from right posterior communicating artery	Yes
	I6032	Nontraumatic subarachnoid hemorrhage from left posterior communicating artery	Yes
	I604	Nontraumatic subarachnoid hemorrhage from basilar artery	Yes
	I6050	Nontraumatic subarachnoid hemorrhage from unspecified vertebral artery	Yes
	I6051	Nontraumatic subarachnoid hemorrhage from right vertebral artery	Yes
	I6052	Nontraumatic subarachnoid hemorrhage from left vertebral artery	Yes
	I606	Nontraumatic subarachnoid hemorrhage from other intracranial arteries	Yes

Table 3A.1 (cont'd)

	I607	Nontraumatic subarachnoid hemorrhage from unspecified intracranial artery	Yes
	I608	Other nontraumatic subarachnoid hemorrhage	Yes
	I609	Nontraumatic subarachnoid hemorrhage, unspecified	Yes
Hemorrhagic (Intracerebral)	I610	Nontraumatic intracerebral hemorrhage in hemisphere, subcortical	Yes
	I611	Nontraumatic intracerebral hemorrhage in hemisphere, cortical	Yes
	I612	Nontraumatic intracerebral hemorrhage in hemisphere, unspecified	Yes
	I613	Nontraumatic intracerebral hemorrhage in brain stem	Yes
	I614	Nontraumatic intracerebral hemorrhage in cerebellum	Yes
	I615	Nontraumatic intracerebral hemorrhage, intraventricular	Yes
	I616	Nontraumatic intracerebral hemorrhage, multiple localized	Yes
	I618	Other nontraumatic intracerebral hemorrhage	Yes
	I619	Nontraumatic intracerebral hemorrhage, unspecified	Yes
	Ischemic	I6300	Cerebral infarction due to thrombosis of unspecified precerebral artery
I63011		Cerebral infarction due to thrombosis of right vertebral artery	Yes
I63012		Cerebral infarction due to thrombosis of left vertebral artery	Yes
I63013		Cerebral infarction due to thrombosis of bilateral vertebral arteries	No
I63019		Cerebral infarction due to thrombosis of unspecified vertebral artery	Yes
I6302		Cerebral infarction due to thrombosis of basilar artery	Yes
I63031		Cerebral infarction due to thrombosis of right carotid artery	Yes
I63032		Cerebral infarction due to thrombosis of left carotid artery	Yes
I63033		Cerebral infarction due to thrombosis of bilateral carotid arteries	No
I63039		Cerebral infarction due to thrombosis of unspecified carotid artery	Yes
I6309		Cerebral infarction due to thrombosis of other precerebral artery	Yes
I6310		Cerebral infarction due to embolism of unspecified precerebral artery	Yes
I63111		Cerebral infarction due to embolism of right vertebral artery	Yes
I63112	Cerebral infarction due to embolism of left vertebral artery	Yes	
I63113	Cerebral infarction due to embolism of bilateral vertebral arteries	No	

Table 3A.1 (cont'd)

I63119	Cerebral infarction due to embolism of unspecified vertebral artery	Yes
I6312	Cerebral infarction due to embolism of basilar artery	Yes
I63131	Cerebral infarction due to embolism of right carotid artery	Yes
I63132	Cerebral infarction due to embolism of left carotid artery	Yes
I63133	Cerebral infarction due to embolism of bilateral carotid arteries	No
I63139	Cerebral infarction due to embolism of unspecified carotid artery	Yes
I6319	Cerebral infarction due to embolism of other precerebral artery	Yes
I6320	Cerebral infarction due to unspecified occlusion or stenosis of unspecified precerebral arteries	Yes
I63211	Cerebral infarction due to unspecified occlusion or stenosis of right vertebral arteries	Yes
I63212	Cerebral infarction due to unspecified occlusion or stenosis of left vertebral arteries	Yes
I63213	Cerebral infarction due to unspecified occlusion or stenosis of bilateral vertebral arteries	No
I63219	Cerebral infarction due to unspecified occlusion or stenosis of unspecified vertebral arteries	Yes
I6322	Cerebral infarction due to unspecified occlusion or stenosis of basilar arteries	Yes
I6323	Cerebral infarction due to unspecified occlusion or stenosis of carotid arteries	No
I63231	Cerebral infarction due to unspecified occlusion or stenosis of right carotid arteries	Yes
I63232	Cerebral infarction due to unspecified occlusion or stenosis of left carotid arteries	Yes
I63233	Cerebral infarction due to unspecified occlusion or stenosis of bilateral carotid arteries	No
I63239	Cerebral infarction due to unspecified occlusion or stenosis of unspecified carotid arteries	Yes
I6329	Cerebral infarction due to unspecified occlusion or stenosis of other precerebral arteries	Yes
I6330	Cerebral infarction due to thrombosis of unspecified cerebral artery	Yes
I63311	Cerebral infarction due to thrombosis of right middle cerebral artery	Yes
I63312	Cerebral infarction due to thrombosis of left middle cerebral artery	Yes

Table 3A.1 (cont'd)

I63313	Cerebral infarction due to thrombosis of bilateral middle cerebral arteries	No
I63319	Cerebral infarction due to thrombosis of unspecified middle cerebral artery	Yes
I63321	Cerebral infarction due to thrombosis of right anterior cerebral artery	Yes
I63322	Cerebral infarction due to thrombosis of left anterior cerebral artery	Yes
I63323	Cerebral infarction due to thrombosis of bilateral anterior cerebral arteries	No
I63329	Cerebral infarction due to thrombosis of unspecified anterior cerebral artery	Yes
I63331	Cerebral infarction due to thrombosis of right posterior cerebral artery	Yes
I63332	Cerebral infarction due to thrombosis of left posterior cerebral artery	Yes
I63333	Cerebral infarction due to thrombosis of bilateral posterior cerebral arteries	No
I63339	Cerebral infarction due to thrombosis of unspecified posterior cerebral artery	Yes
I63341	Cerebral infarction due to thrombosis of right cerebellar artery	Yes
I63342	Cerebral infarction due to thrombosis of left cerebellar artery	Yes
I63343	Cerebral infarction due to thrombosis of bilateral cerebellar arteries	No
I63349	Cerebral infarction due to thrombosis of unspecified cerebellar artery	Yes
I6339	Cerebral infarction due to thrombosis of other cerebral artery	Yes
I6340	Cerebral infarction due to embolism of unspecified cerebral artery	Yes
I6341	Cerebral infarction due to embolism of middle cerebral artery	No
I63411	Cerebral infarction due to embolism of right middle cerebral artery	Yes
I63412	Cerebral infarction due to embolism of left middle cerebral artery	Yes
I63413	Cerebral infarction due to embolism of bilateral middle cerebral arteries	No
I63419	Cerebral infarction due to embolism of unspecified middle cerebral artery	Yes
I63421	Cerebral infarction due to embolism of right anterior cerebral artery	Yes
I63422	Cerebral infarction due to embolism of left anterior cerebral artery	Yes

Table 3A.1 (cont'd)

I63423	Cerebral infarction due to embolism of bilateral anterior cerebral arteries	No
I63429	Cerebral infarction due to embolism of unspecified anterior cerebral artery	Yes
I63431	Cerebral infarction due to embolism of right posterior cerebral artery	Yes
I63432	Cerebral infarction due to embolism of left posterior cerebral artery	Yes
I63433	Cerebral infarction due to embolism of bilateral posterior cerebral arteries	No
I63439	Cerebral infarction due to embolism of unspecified posterior cerebral artery	Yes
I63441	Cerebral infarction due to embolism of right cerebellar artery	Yes
I63442	Cerebral infarction due to embolism of left cerebellar artery	Yes
I63443	Cerebral infarction due to embolism of bilateral cerebellar arteries	No
I63449	Cerebral infarction due to embolism of unspecified cerebellar artery	Yes
I6349	Cerebral infarction due to embolism of other cerebral artery	Yes
I635	Cerebral infarction due to unspecified occlusion or stenosis of cerebral arteries	No
I6350	Cerebral infarction due to unspecified occlusion or stenosis of unspecified cerebral artery	Yes
I63511	Cerebral infarction due to unspecified occlusion or stenosis of right middle cerebral artery	Yes
I63512	Cerebral infarction due to unspecified occlusion or stenosis of left middle cerebral artery	Yes
I63513	Cerebral infarction due to unspecified occlusion or stenosis of bilateral middle cerebral arteries	No
I63519	Cerebral infarction due to unspecified occlusion or stenosis of unspecified middle cerebral artery	Yes
I63521	Cerebral infarction due to unspecified occlusion or stenosis of right anterior cerebral artery	Yes
I63522	Cerebral infarction due to unspecified occlusion or stenosis of left anterior cerebral artery	Yes
I63523	Cerebral infarction due to unspecified occlusion or stenosis of bilateral anterior cerebral arteries	No
I63529	Cerebral infarction due to unspecified occlusion or stenosis of unspecified anterior cerebral artery	Yes

Table 3A.1 (cont'd)

I63531	Cerebral infarction due to unspecified occlusion or stenosis of right posterior cerebral artery	Yes
I63532	Cerebral infarction due to unspecified occlusion or stenosis of left posterior cerebral artery	Yes
I63533	Cerebral infarction due to unspecified occlusion or stenosis of bilateral posterior cerebral arteries	No
I63539	Cerebral infarction due to unspecified occlusion or stenosis of unspecified posterior cerebral arter	Yes
I63541	Cerebral infarction due to unspecified occlusion or stenosis of right cerebellar artery	Yes
I63542	Cerebral infarction due to unspecified occlusion or stenosis of left cerebellar artery	Yes
I63543	Cerebral infarction due to unspecified occlusion or stenosis of bilateral cerebellar arteries	No
I63549	Cerebral infarction due to unspecified occlusion or stenosis of unspecified cerebellar artery	Yes
I6359	Cerebral infarction due to unspecified occlusion or stenosis of other cerebral artery	Yes
I636	Cerebral infarction due to cerebral venous thrombosis, nonpyogenic	Yes
I638	Other cerebral infarction	Yes
I6381	Other cerebral infarction due to occlusion or stenosis of small artery	No
I6389	Other cerebral infarction	No
I639	Cerebral infarction, unspecified	Yes

Table 3A.2: Probabilistic linkage manual review to determine minimum plausible linkage weight.

Linkage method	DOB	Sex	Admission date	Discharge date	Hospital ID	Admission year	Initial number of linkage pairs	Minimum plausible linkage weight (threshold)*	Number of linkage pairs after threshold	Manual review excluded linked pairs**	Final minimum cut off weight	Final linkage pairs (N)
probabilistic	X	X	X	X	X	Blocking	24,625	22.6	24,249	194	32.8	23,918
Probabilistic with EM algorithm (98% sensitivity)	X	X	X	X	X	Blocking	73,920	22.4	24,346	428	27.8	23,918

*Threshold was determined by manually reviewing all the weights of the generated possible linkages to determine the lowest linkage weight of a plausible match.

**Manual review excluded linked pairs that did not exactly match on DOB, gender, or hospital ID or had more than 1 day difference in admission or discharge dates.

Table 3A.3: MiSP and MVC recorded stroke events stratified by stroke event year of discharge.

Database/ year	2016	2017	2018	2019	2020	Row total
MVC (% total)	6,534 (21.3%)	6,690 (21.8%)	6,526 (21.3%)	5,932 (19.3%)	5,003 (16.3%)	30,685
MiSP (% total)	8,966 (19.4%)	9,946 (21.5%)	10,154 (21.9%)	9,205 (19.9%)	8,059 (17.4%)	46,330
Difference (% MiSP)	2,432 (27.1%)	3,256 (32.7%)	3,628 (35.7%)	3,273 (35.6%)	3,056 (37.9%)	15,645 (33.8%)*

*MiSP maximum linkage rate according to the difference in recorded stroke events between MiSP and MVC per year is 66.2% (100% -33.8%)

Table 3A.4: Hospital specific recorded stroke events in MiSP and MVC and their corresponding differences and linkage rates.

Hospital (N=31)	MiSP	MVC	Difference (MiSP – MVC)	Difference (% MiSP)	Linked pairs (n)	Linkage rate (% MiSP)
1	592	397	195	32.9	312	52.7
2	1,305	1,344	-39	-3.0*	762	58.4
3	5,099	2,735	2,364	46.4	2,156	42.3
4	1,976	1,200	776	39.3	977	49.4
5	1,224	792	432	35.3	681	55.6
6	1,369	1,799	-430	-31.4*	893	65.2
7	2,752	1,751	1,001	36.4	1,549	56.3
8	1,072	759	313	29.2	640	59.7
9	1,484	852	632	42.6	750	50.5
10	279	221	58	20.8	191	68.5
11	1,274	611	663	52.0	455	35.7
12	4,107	1,812	2,295	55.9	1,588	38.7
13	969	582	387	39.9	449	46.3
14	2,300	1,477	823	35.8	1,288	56.0
15	1,631	821	810	49.7	700	42.9
16	1,045	443	602	57.6	403	38.6
17	119	237	-118	-99.2*	48	40.3
18	2,652	1,358	1,294	48.8	1,247	47.0
19	3,092	2,196	896	29.0	1,973	63.8
20	397	374	23	5.8	244	61.5
21	1,056	1080	-24	-2.3*	682	64.6
22	443	318	125	28.2	270	60.9
23	1,729	1,211	518	30.0	994	57.5
24	1,774	1,170	604	34.0	1,006	56.7
25	1,607	1,073	534	33.2	770	47.9
26	956	650	306	32.0	574	60.0
27	1,707	1,876	-169	-9.9*	1,075	63.0
28	253	150	103	40.7	124	49.0
29	756	554	202	26.7	459	60.7
30	346	282	64	18.5	215	62.1
31	965	560	405	42.0	443	45.9
Total	46,330	30,685	15,645	33.8%**	23,918	51.6%

* These 5 sampling hospitals were identified because MiSP reported stroke events are less than MVC events even though MVC does not cover all of MiSP population.

**MiSP maximum linkage rate according to hospital recorded events is 66.2% (100%-33.8%)

Table 3A.5: Stratification of MVC and MiSP stroke events according to age groups (<65 and <=65 years old).

	MVC (n=30,685)	MiSP (n= 46,330)	<65 y/o		>= 65 y/o	
			MVC (n= 6,051)	MiSP (n= 16,192)	MVC (n=24,634)	MiSP (n=30,138)
Difference (% MiSP)	15,645 (33.8%)		10,141 (62.6%)*		5,413 (18.3%)**	

* ~63% discrepancy in total numbers observed in <65 age group is because MVC only includes BCBSM private insurance (HMO, PPO) beneficiaries and Medicaid and other insurance providers are missing.

** ~18% discrepancy in total numbers observed in >=65 age group is because Medicare Advantage enrollees outside of BCBSM are not included in MVC.

Table 3A.6: Comparison of demographics, stroke type, and payer characteristics recorded in MVC between linked and unlinked data stratified by age groups.

Variable		All MVC (N= 30,685)		<65 years				>=65 years			
		No.	%	All MVC (N= 6,051)	Linked (N= 4,729, 78.2%)	Unlinked (N= 1,322, 21.8%)	Absolute standardized difference^	All MVC (N= 24,634)	Linked (N= 19,189, 77.9%)	Unlinked (N= 5,445, 22.1%)	Absolute standardized difference^
Age	Mean (SD)	30,685	74.4 (12.8)	55.1 (8.3)	55.1 (8.2)	55.3 (8.4)	0.02	79.1 (8.6)	79.0 (8.5)	79.5 (8.6)	0.06
Sex	Female	16,306	53.1	43.5	43.5	43.3	<0.01	55.5	55.4	56.0	0.01
	Male	14,379	46.9	56.5	56.5	56.7		44.5	44.6	44.0	
Stroke type	Ischemic (%)	26,202	85.4	81.5	80.5	85.0	0.12	86.4	85.0	91.0	0.18
	Hemorrhagic (%)	4,483	14.6	18.5	19.5	15.0		13.6	15.0	8.0	
Payer	BCBSM (HMO and PPO)	4,488	14.6	57.9	59.1	53.4	0.14	4.0	4.1	3.6	0.12
	BSBSM MA	7,069	23.0	5.3	4.7	7.4		27.4	26.2	31.6	
	Medicare FFS	19,128	62.3	36.8	36.1	39.2		68.6	69.7	64.8	

^ A value higher than 0.1 represents a meaningful difference.

Table 3A.7: Sensitivity analysis of outcome event rates between the uncleaned and cleaned linked 1-year episode of care datasets.[^]

Outcome	linked 1-year episode of care before cleaning (N=22,889)			linked 1-year episode of care after cleaning (N= 19,382)		
	30-day event rate % (n)	90-day event rate % (n)	1-year event rate % (n)	30-day event rate % (n)	90-day event rate % (n)	1-year event rate % (n)
Inpatient rehabilitation facility utilization*	21.4 (4,891)	21.9 (5,016)	22.9 (5,250)	24.9 (4,822)	25.5 (4,946)	26.7 (5,171)
Skilled nursing facility utilization*	24.7 (5,643)	27.1 (6,203)	30.4 (6,946)	28.1 (5,449)	31.2 (6,049)	34.9 (6,765)
Home health utilization*	23.9 (5,465)	33.3 (7,611)	38.8 (8,872)	27.5 (5,336)	38.4 (7,436)	44.7 (8,659)
Outpatient visit*	45.3 (10,365)	66.4 (15,202)	79.1 (18,098)	46.4 (8,999)	70.8 (13,720)	85.3 (16,539)
All cause readmission*	12.3 (2,811)	21.7 (4,977)	36.7 (8,405)	14.1 (2,724)	24.9 (4,833)	42.2 (8,169)
Stroke recurrence*	2.9 (670)	4.5 (1,034)	7.4 (1,682)	3.3 (641)	5.1 (991)	8.3 (1,614)
Mortality**	15.8 (2,298)	19.8 (2,876)	27.4 (3,992)	4.0 (486)	9.1 (1,109)	19.8 (2,416)
Home time-Median (IQR)**	18.0 (29.0)	75.0 (69.0)	341.0 (212.0)	22.0 (26.0)	79.0 (40.0)	347 (94.0)

[^]Data cleaning involved excluding patients discharged to hospice care or who left against medical advice and for patients with multiple stroke episodes of care (i.e., another acute stroke admission that occurred at least 1 year apart); only the first episode was included.

* Numerator includes only the first occurrence for a given patient during follow up period.

** Calculated only for Medicare FFS beneficiaries (n= 14,557 before cleaning and n= 12,185 after cleaning).

Table 3A.8: Thirty-day, 90-day, and 1-year post stroke discharge outcome event rates stratified by age groups among linked stroke patients.[^]

Outcome	30-day event rate % (n)			90-day event rate % (n)			1-year event rate % (n)			
	Age category ^{^^}	<65	≥ 65	X ² test	<65	≥ 65	X ² test	<65	≥ 65	X ² test
Inpatient rehabilitation facility utilization*		24.0 (1,001)	25.1 (3,821)	0.15	25.0 (1,041)	25.7 (3,905)	0.34	26.3 (1,097)	26.8 (4,074)	0.56
Skilled nursing facility utilization*		11.9 (496)	32.9 (5,003)	<0.01	14.4 (600)	35.8 (5,449)	<0.01	16.6 (691)	39.9 (6,074)	<0.01
Home health utilization*		17.3 (720)	30.3 (4,616)	<0.01	24.0 (1,000)	42.3 (6,436)	<0.01	29.1 (1,211)	49.0 (7,448)	<0.01
Outpatient visit*		43.6 (1,816)	47.2 (7,183)	<0.01	69.4 (2,890)	71.2 (10,830)	0.02	85.2 (3,548)	85.4 (12,991)	0.70
All cause readmission*		13.7 (570)	14.2 (2,154)	0.43	22.9 (954)	25.5 (3,879)	<0.01	36.6 (1,526)	43.7 (6,643)	<0.01
Stroke recurrence*		4.3 (177)	3.1 (464)	<0.01	6.0 (250)	4.9 (741)	<0.01	9.4 (391)	8.0 (1,223)	<0.01

[^] Mortality and home time outcomes are not available due to limitation in data availability for the whole linked population.

^{^^} Number of patients among <65 = 4,167, and ≥65 = 15,215.

* Numerator includes only the first occurrence for a given patient during follow up period.

Table 3A.9: Thirty-day, 90-day, and 1-year post stroke discharge outcome event rates stratified by race among linked stroke patients.

Outcome	30-day event rate				90-day event rate				1-year event rate			
	% (n)			X ² or F-test	% (n)			X ² or F-test	% (n)			X ² or F-test
Race [^]	White	Black	Other		White	Black	Other		White	Black	Other	
Inpatient rehabilitation facility utilization*	24.7 (3,813)	26.0 (737)	20.0 (51)	0.06	25.3 (3,908)	26.9 (761)	20.0 (51)	0.02	26.3 (4,068)	28.7 (814)	20.8 (53)	<0.01
Skilled nursing facility utilization*	28.3 (4,366)	29.1 (825)	26.7 (68)	0.53	30.8 (4,767)	33.2 (941)	28.6 (73)	0.03	34.3 (5,308)	38.2 (1,083)	31.4 (80)	<0.01
Home health utilization*	27.4 (4,239)	28.7 (813)	31.0 (79)	0.19	37.9 (5,865)	41.6 (1,179)	38.0 (97)	<0.01	43.9 (6,778)	50.6 (1,433)	41.2 (105)	<0.01
Outpatient visit*	47.4 (7,328)	40.0 (1,133)	51.8 (132)	<0.01	71.8 (11,103)	63.5 (1,800)	76.1 (194)	<0.01	86.2 (13,330)	79.5 (2,252)	85.5 (218)	<0.01
All cause readmission*	13.7 (2,110)	15.9 (450)	17.7 (45)	<0.01	24.1 (3,717)	29.6 (838)	26.7 (68)	<0.01	41.0 (6,334)	48.9 (1,384)	38.8 (99)	<0.01
Stroke recurrence*	3.3 (502)	3.4 (96)	5.6 (15)	0.06	5.0 (766)	5.9 (167)	6.7 (17)	0.06	7.9 (1,228)	10.4 (295)	7.8 (20)	<0.01
Mortality**	4.2 (401)	3.1 (59)	5.2 (9)	0.08	9.4 (899)	8.0 (151)	20 (11.6)	0.10	19.8 (1,902)	19.9 (373)	22.1 (38)	0.75
Home time- Median (IQR)**	22.0 (25.0)	19.0 (30.0)	27.0 (24.5)	<0.01	80.0 (38.0)	76.0 (51.0)	82.0 (37.0)	<0.01	349.0 (88.0)	341.0 (118.0)	351.0 (85.5)	0.01

[^] Number of patients who are White = 15,457, Black =2,833, and Other=255. We did not include 837 patients who have missing race data.

* Numerator includes only the first occurrence for a given patient during follow up period.

** Calculated only for Medicare FFS beneficiaries (n=11,671; 9,620 were White, 1,879 were Black, and 172 were Other). Among FFS beneficiaries 514 were not included in this table because of having missing race data.

Table 3A.10: Thirty-day, 90-day, and 1-year post stroke discharge outcome event rates stratified by stroke type among linked stroke patients.

Outcome	30-day event rate			90-day event rate			1-year event rate		
	% (n)		X^2 or t-test	% (n)		X^2 or t-test	% (n)		X^2 or t-test
Stroke type [^]	Ischemic	Hemorrhagic		Ischemic	Hemorrhagic		Ischemic	Hemorrhagic	
Inpatient rehabilitation facility utilization*	24.2 (4,091)	29.7 (731)	<0.01	24.8 (4,190)	30.7 (756)	<0.01	26.0 (4,396)	31.5 (775)	<0.01
Skilled nursing facility utilization*	27.6 (4,668)	33.8 (831)	<0.01	30.2 (5,105)	38.4 (944)	<0.01	34.0 (5,751)	41.2 (1,014)	<0.01
Home health utilization*	28.0 (4,732)	24.5 (604)	<0.01	38.5 (6,517)	37.3 (919)	0.26	44.7 (7,560)	44.7 (1,099)	0.98
Outpatient visit*	46.6 (7,883)	45.4 (1,116)	0.25	70.5 (11,924)	73.0 (1,796)	0.01	85.2 (14,411)	86.5 (2,128)	0.09
All cause readmission*	13.5 (2,289)	17.7 (435)	<0.01	24.2 (4,090)	30.2 (743)	<0.01	41.7 (7,051)	45.4 (1,118)	<0.01
Stroke recurrence*	3.2 (534)	4.4 (107)	<0.01	5.0 (842)	6.1 (149)	0.02	8.2 (1,383)	9.4 (231)	0.04
Mortality**	3.4 (417)	5.0 (69)	0.04	8.9 (957)	11.0 (152)	0.01	19.6 (2,112)	22.0 (304)	0.03
Home time- Median (IQR)**	23.0 (25.0)	13.0 (30.0)	<0.01	80.0 (37.0)	70.0 (60.0)	<0.01	349.0 (87.5)	333.0 (126.0)	<0.01

[^] Number of patients who had ischemic stroke = 16,921 and hemorrhagic stroke = 2,461.

* Numerator includes only the first occurrence for a given patient during follow up period.

** Calculated only for Medicare FFS beneficiaries (n=12,185; 10,804 had ischemic stroke and 1,381 had hemorrhagic stroke).

Table 3A.11: Thirty-day, 90-day, and 1-year post stroke discharge outcome event rates stratified by stroke severity among linked stroke patients.

Outcome	30-day event rate				90-day event rate				1-year event rate			
	Mild (0-4)	Moderate (5-15)	Severe (16-42)	X ² or F-test	Mild (0-4)	Moderate (5-15)	Severe (16-42)	X ² or F- test	Mild (0-4)	Moderate (5-15)	Severe (16-42)	X ² or F-test
Inpatient rehabilitation facility utilization*	20.0 (2,179)	35.6 (1,766)	33.4 (522)	<0.01	20.5 (2,242)	36.2 (1,795)	34.4 (537)	<0.01	21.9 (2,386)	37.2 (1,841)	35.3 (552)	<0.01
Skilled nursing facility utilization*	18.7 (2,044)	39.6 (1,964)	51.6 (806)	<0.01	20.8 (2,272)	42.6 (2,109)	57.3 (896)	<0.01	24.6 (2,686)	46.5 (2,304)	59.1 (924)	<0.01
Home health utilization*	28.0 (3,058)	29.3 (1,451)	21.2 (331)	<0.01	36.1 (22.6)	44.6 (2,211)	36.1 (564)	<0.01	41.6 (4,542)	51.9 (2,573)	43.8 (685)	<0.01
Outpatient visit*	50.5 (5,512)	41.9 (2,074)	36.6 (572)	<0.01	73.4 (8,012)	68.6 (3,400)	63.0 (985)	<0.01	87.5 (9,557)	83.7 (4,149)	78.4 (1,225)	<0.01
All cause readmission*	11.6 (1,267)	15.4 (764)	22.0 (344)	<0.01	20.7 (2,265)	27.8 (1,378)	35.4 (553)	<0.01	37.7 (4,120)	46.0 (2,277)	52.0 (813)	<0.01
Stroke recurrence*	3.1 (338)	3.5 (173)	3.5 (55)	0.35	4.7 (517)	5.6 (277)	5.3 (82)	0.07	7.9 (867)	9.0 (447)	8.1 (127)	0.07
Mortality**	2.0 (135)	5.4 (176)	11.6 (118)	<0.01	5.2 (350)	12.3 (405)	21.4 (219)	<0.01	13.4 (895)	25.3 (829)	39.3 (402)	<0.01
Home time-Median (IQR)**	30.0 (15.0)	13.0 (15.0)	1.0 (16.0)	<0.01	89.0 (19.0)	69.0 (57.0)	36.0 (71.0)	<0.01	358.0 (36.0)	328.0 (152.0)	258.0 (296.0)	<0.01

^ Number of patients who had mild stroke = 10,922, moderate stroke = 4,955, and severe stroke = 1,563. We did not include 1,942 patients who have missing NIHSS data.

* Numerator includes only the first occurrence for a given patient during follow up period.

** Calculated only for Medicare FFS beneficiaries (n=10,994; 6,689 had mild stroke, 3,283 had moderate stroke and 1,022 had severe stroke). Among FFS beneficiaries 1,191 were not included in this table because of having missing NIHSS data.

CHAPTER 4: MANUSCRIPT 2 – PREDICTION OF HOSPITAL READMISSION AFTER STROKE USING MACHINE LEARNING IN A 5-YEAR LINKED COHORT FROM THE MICHIGAN STROKE REGISTRY

4.1 Abstract

Introduction: Hospital readmissions following stroke are common. However, identifying stroke patients at risk of readmission is challenging as predictive models demonstrate modest discrimination (AUC range 0.64 - 0.74), in part because they often rely on limited medical record data collected from single hospitals or healthcare systems. We aimed to develop 30-day and 1-year readmission machine learning (ML) based prediction models using linked registry data and report the most important predictors.

Methods: We probabilistically linked clinical data from acute stroke patients (ICD-10 I61-I63) discharged from any of 31 participating hospitals between 2016 and 2020 from Michigan's Acute Stroke registry to claims data from the Michigan Value Collaborative, a multipayer claims database of Medicare and Blue Cross Blue Shield Michigan beneficiaries. The claims data was used to identify all-cause readmission events within 30-day and 1-year of discharge. We compared the performance of a simple ML method (i.e., LASSO logistic regression) and 2 advanced ML (i.e., extreme gradient boosting (XGBoost), and Artificial neural network (ANN)) methods to predict readmission at 30-days and 1-year post discharge. To evaluate prediction accuracy, we applied a hospital-split internal cross validation method and reported the pooled hospital-specific AUC. Important predictors were identified according to the rank order they were selected by the 31-hospital-specific models using the best performing ML method.

Results: Of 19,382 linked stroke discharges, 2,724 (14.1%) and 8,169 (42.2%) were readmitted within 30-days and 1-year, respectively. The linked population had a mean age of 73.3 (SD= 12.7), 79.7% were white, 52.2% female, 87.3% had an ischemic stroke, 56.4% had a minor

stroke (NIHSS <5), and 50.1% were discharged directly home. LASSO logistic regression model produced similar AUC to XGBoost and ANN ($P>0.05$) with a pooled 30-day and 1-year readmission AUC of 0.68 (95% CI: 0.65-0.70) and 0.67 (95% CI: 0.65-0.69), respectively.

Variables with the highest predictive importance were discharge disposition, length of stay and preexisting comorbidities including chronic renal failure, heart failure, and atrial fibrillation. In contrast, clinical features of stroke (e.g., NIHSS, stroke etiology, and ambulatory status) were far less important and were almost absent from the list of the 15 highest ranked important predictors in the 1-year readmission model.

Conclusions: LASSO regression was able to predict readmission after stroke with similar accuracy as more advanced ML methods. Clinical features of stroke were much less important than the burden of existing comorbidities in predicting post-stroke readmission, especially over longer periods of time.

4.2 Introduction

4.2.1 Stroke Readmission in The US and Its Significance to Payment Reform Policies

In the US, nearly 800,000 patients are diagnosed with new or recurrent stroke every year.¹ Compared to other medical conditions, stroke is associated with high rates of hospital readmission.²⁻⁵ Published post-stroke readmission rates in the US vary (30-day readmission: 8.9%-15.4%, 1-year readmission: 27.2%-48.7%) in large part due to different population inclusion criteria (i.e., age, payer, single vs multi hospital, planned vs unplanned readmission, stroke type).⁶⁻⁹ However, a meta-analysis published in 2016 of 10 reports published between 2006 and 2015, estimated the pooled 30-day and 1-year all-cause readmission rates following stroke to be 17.4% (95% CI, 12.7–23.5%) and 42.5% (95% CI, 34.1–51.3%), respectively.⁶

In 2018 Medicare insured patients (≥ 65 years) had a high all-cause readmission rate of 16.9% with an estimated total cost of about \$26 billion annually.^{2, 5, 10} The Center for Medicare and Medicaid Services (CMS) regards reducing readmissions as one of the central goals of national healthcare reforms.^{5, 9} In 2012, CMS identified readmissions as a measure of hospital quality and integrated them into the Hospital Readmissions Reduction Program (HRRP) payment reform.^{3, 9-11} Although readmission following stroke is not included as one of the qualifying medical conditions in the HRRP program (i.e., acute myocardial infarction, chronic obstructive pulmonary disease, heart failure, pneumonia, coronary artery bypass graft surgery, and elective primary total hip arthroplasty and/or total knee arthroplasty), there is evidence that a spillover effect of the HRRP program may exist that reduces readmissions among patients with conditions not targeted by the program.³ With respect to stroke, a nationwide study reported that 30-day post-stroke readmission rate was reduced on a relative basis by 12% after implementation of the HRRP program compared to the pre implementation period.³

4.2.2 Stroke Readmission Prevention

Identifying specific patient-, hospital-, and systems- level factors associated with readmissions is important as they can guide the development of potential interventions.¹²⁻¹⁵ Factors associated with post-stroke readmissions in US populations have been identified using data from hospital based stroke registries,¹⁶⁻¹⁹ insurance claims,^{7, 9, 16, 17, 20} and electronic medical records (eMR).^{17, 19} Studies have reported that patients who presented with severer strokes^{18, 19}, were Medicare/Medicaid beneficiaries⁷, had a prolonged hospital stay^{7, 17-19}, and were discharged to an intermediate care (e.g., nursing home), hospice, skilled nursing facility or who left against medical advice^{9, 18, 20} had higher risk of readmission. Furthermore, patients with a previous medical history of stroke¹⁹, diabetes^{7, 18}, heart failure^{17, 18}, coronary artery disease¹⁷⁻¹⁹, hypertension^{7, 18}, or renal disease^{16, 17} were also at higher risk of post-stroke readmission.

Research has shown that a proportion – estimated to be between 12% and 31% of unplanned post-stroke readmissions are preventable.^{8, 9, 21} However, it is unclear how much of the preventable readmissions are related to inpatient care, discharge plans, or post discharge care.⁹ Interventions to reduce 30-day readmission after stroke are centered around improvements to transitional care and early follow up; several interventions examining the implementation of improved transitional care programs reported a 48%-54% relative decrease in the risk of post-stroke readmissions²²⁻²⁴. In addition, two population-based studies reported a 2%-16% relative reduced risk of post-stroke readmission among patients who had a primary care visit compared to patients who did not within the first 30-days post discharge.^{25, 26}

4.2.3 Predictive Modelling for Stroke Readmission in the US

Developing prediction models that accurately identify patients at high risk of readmission before discharge could help to target patient to received specific enhanced transitional care

management.²⁷ An example of using valid prediction models to target interventions was evaluated in 19 Kaiser Permanente Northern California (KPNC) hospitals to reduce 30-day post-discharge mortality by identifying and triggering an alert for patients at high risk for clinical deterioration during hospitalization.²⁸ The Kaiser study found that mortality within 30 days after an alert was 16% lower in the intervention cohort compared to the comparison cohort (adjusted relative risk: 0.84, 95% CI: (0.78-0.90)). This example is specific to mortality because as far as we know there are no examples of a validated population based prediction model that is used during hospitalization to identify patient at high risk of readmission.

A review by Lichtman et al. in 2010 did not identify any peer reviewed studies that reported on patient-level prediction model for readmission after stroke either in the US or elsewhere.²⁹ A literature search in PubMed for papers published after 2010 using different combinations of the terms stroke, prediction, prevention, readmission, and machine learning revealed only 5 US-based papers that developed predictive models for post-stroke readmissions, three of which utilized traditional multivariable statistical methods (e.g., logistic or Cox regression)^{9, 30, 31} and two used ML methods.^{5, 32} However, post-stroke readmission prediction models that utilized the traditional statistical methods have several notable limitations including the fact that they utilize a limited number of candidate predictor variables, often relied on a single data source (Medicare FFS claims data), and produced models with modest prediction accuracy as reported using the area under the receiver operating characteristic curve (AUC) (AUC range 0.53 – 0.67).^{9, 30, 31} Logistic regression methods have known limitations including a limited ability to model a large number of interactions, to efficiently deal with high dimensional datasets, and can produce results that tend to be overfitted.³³⁻³⁵ Machine learning (ML) methods that only became widely accessible in the last few years can overcome some of these limitations

and has been used successfully to predict readmission in other conditions (e.g., carotid stenosis,³⁶ heart failure,^{37, 38} myocardial infarction³⁹) as well as in stroke.^{5, 32, 40-42} US based studies that developed ML readmission prediction models for stroke utilized a rich source of data like eMR which permitted including more predictors and produced improved predictive accuracy rates over traditional multivariate regression methods (AUC range 0.64 - 0.74).^{5, 32} Results from 5 studies (including two US based) that compared different ML methods indicate that the most accurate ML methods to predict post-stroke readmission were extreme gradient boosting (XGBoost) and artificial neural networks (ANN) when compared to both traditional modelling methods (e.g., logistic and cox regression) or other ML based methods (e.g., random forest, support vector machine, k-nearest neighbor, and naïve bayes classifier).^{5, 32, 40-42} However, these 5 ML based post-stroke readmission models relied on a single data source (i.e., electronic medical records) and mostly reported only 30-day readmission outcomes.^{5, 32, 40-42}

Different data sources (i.e., electronic medical records, administrative or claims, registry, and hospital survey data) are designed to serve different purposes hence each will have different strengths and limitations.⁴³⁻⁴⁶ Combinations of these data sources through linkage can bridge gaps in the limitations of any single data source, providing a richer source of patient-, hospital-, and system- level data. These data sources once linked together can identify associations that would be impossible to determine otherwise, and can provide outcomes data e.g., readmissions that are often missing or undercounted for in registry and eMR data.⁴⁶⁻⁴⁸

4.2.4 Hypothesis and Objectives

Our central hypothesis is that using patient level information from registry data and ML methods i.e., XGBoost, and ANN will perform better in terms of prediction accuracy of readmission following stroke, compared to simple ML least absolute shrinkage and selection

operator (LASSO) logistic regression. Our primary objectives were to develop 30-day and 1-year all-cause readmission prediction models using LASSO logistic regression, and two advanced ML based methods (i.e., XGBoost and ANN), and to compare the predictive performance of these methods when applied to registry data, and to report the most important predictors from the best performing prediction method. A secondary objective was to examine the impact of using different combinations of data sources (i.e., registry, hospital survey, and administrative data) on the predictive performance of the methods, with the hypothesis that the combination of all data sources will produce the highest predictive performance model.

4.3 Methods

4.3.1 Study Databases

The study was based on the analysis of prospectively collected data of acute ischemic and hemorrhagic stroke discharges (ICD-10 I61-I63) between January 2016 and December 2020 collected by 31 Michigan hospitals participating in the Michigan Stroke Program (MiSP). This data was probabilistically linked to claims data provided by The Michigan Value Collaborative (MVC) registry using indirect identifiers i.e., date of birth, sex, admission date, discharge date, and hospital ID. Both, MiSP and MVC datasets are deidentified and so do not contain any unique patient identifiers. In addition, data from the American Hospital Association's annual survey database were obtained and linked to the admitting hospital unique identification number and admission year.

The MiSP is a representative statewide, hospital-based acute-stroke registry which is part of the CDC Paul Coverdell National Acute Stroke Program (PCNASP) that continuously collected data between 2016-2020 from 31 participating certified stroke hospitals in Michigan. Of the 31 accredited hospitals, 20 were primary stroke centers, 3 were thrombectomy capable

stroke centers, and 8 were comprehensive stroke centers. These 31 hospitals include the majority of the 49 certified stroke centers in Michigan that represents an estimated ~64% of all stroke admissions in the state.^{49, 50} MiSP aims to track and improve stroke care and patient outcomes through the implementation of quality improvement programs.^{49, 50} MiSP identifies stroke discharges using a clinical case definition.⁴⁹ For each discharge detailed clinical data are entered into the GWTG-S comprehensive case record form (CRF).⁵¹ Stroke discharges are reported in MiSP as a standalone anonymized event and so there is no ability to link events related to the same patient, so it is not possible to distinguish stroke discharges as either index stroke events or post-stroke readmissions or recurrences.

MVC is a comprehensive, statewide, claims-based database that includes data from 101 participating hospitals and 40 physician organizations in the state.⁵² The MVC database covers 71% of Michigan's 143 hospitals.⁵² MVC contains claims data for Michigan residents insured by Medicare FFS, Medicaid, and all insurance plans covered by Blue Cross Blue Shield of Michigan (BCBSM). All told, MVC data covers approximately 84% of Michigan's insured population.⁵² Due to restrictions in MVC's DUA with CMS, Medicaid data was not available to be used for this study. Detailed information on MVC database can be found in Chapter 3 of this dissertation.

The American Hospital Association's annual survey is a voluntary survey that represents the most reliable and comprehensive data about hospital facilities in the US.⁵³ The survey is completed annually by nearly 6,300 hospitals and more than 400 health care systems. The survey collects extensive data on a wide variety of topics including hospital organizational structure, facilities and services, utilization data, physician arrangements, staffing, and community orientation.⁵³

This research was approved by Michigan State University (MSU), University of Michigan (UM), and Michigan Department of Health and Human Services (MDHHS) Institutional Review Boards (IRB).

4.3.2 Data Cleaning

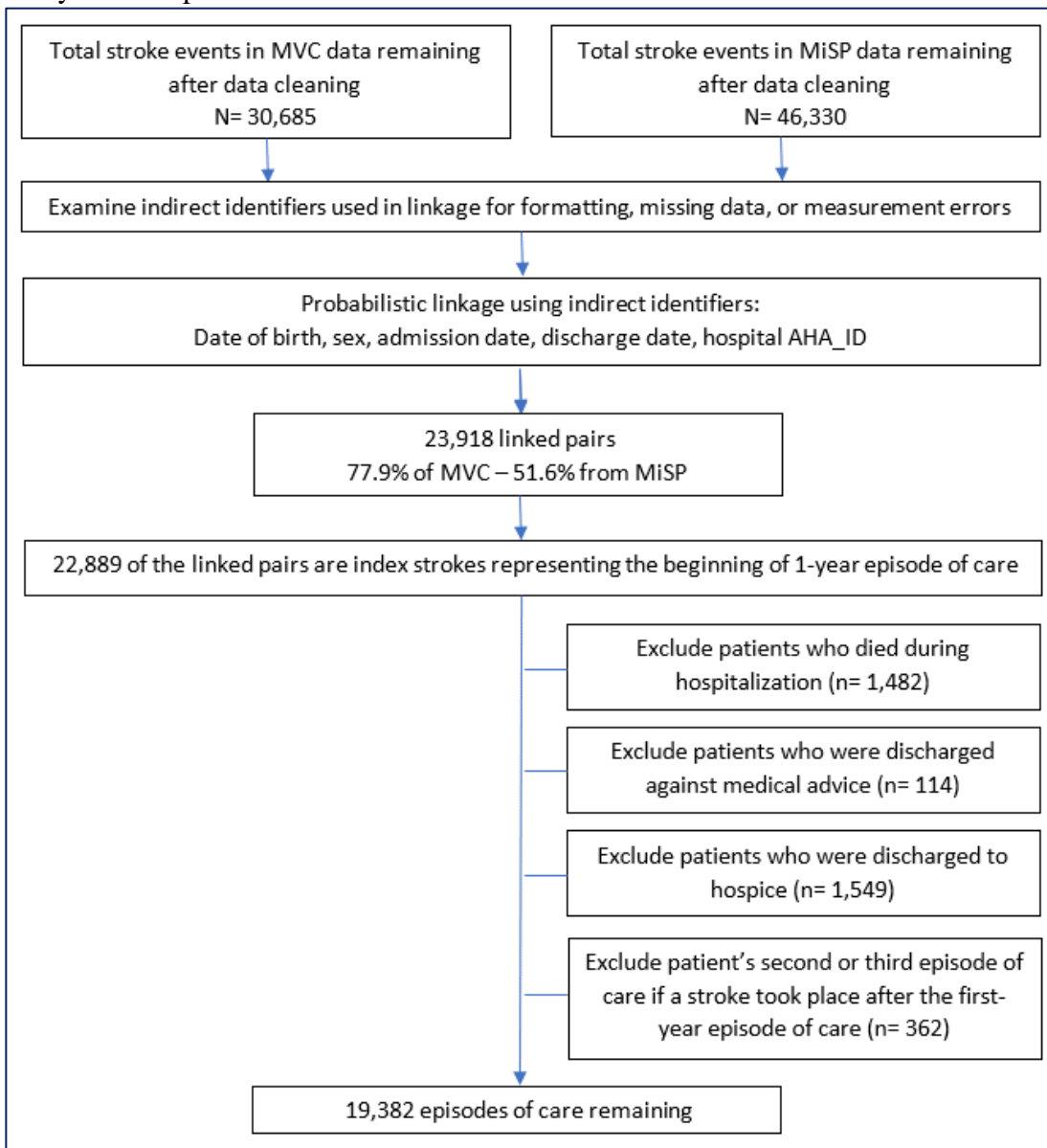
In this research, an index stroke event was defined as patient's first-stroke discharge during the 5-year study period, and a readmission event as any subsequent discharges occurring within one-year of the discharge date of an index stroke event. A stroke related discharge was identified using primary ICD-10 I61-I63 discharge codes. For each index event, all subsequent medical claims reported within the 1-year period following discharge were identified and a comprehensive cleaning process took place to remove duplicate claims submitted for the same health service. In addition, a comprehensive data cleaning process of the MiSP data took place so that it matched MVC's inclusion and exclusion criteria. After cleaning, the number of acute stroke discharges including index and recurrent events in the MiSP and MVC data were 46,330 and 30,685, respectively. All data cleaning, merging, and linkage preparations were done using SAS software v9.4 (Cary, NC). Details on the cleaning process can be found in chapter 3 of this dissertation.

4.3.3 Data Linkage and Study Population

Because the MiSP dataset is unable to distinguish between index events and recurrent stroke events, linkage with MVC must take place at the individual stroke event level. Of the 30,685 identified stroke events in MVC dataset, 28,131 events were index stroke events, and the rest are recurrent stroke events. Using date of birth, sex, admission date, discharge date, and hospital ID linkage variables probabilistic linkage was conducted between the 46,330 MiSP and 30,685 MVC acute stroke discharges. The linkage resulted in 23,918 matched pairs, 22,889 of

which are index strokes that represent the beginning of 1-year stroke episode of care (Figure 4.1). For patients with multiple stroke episodes of care (i.e., another acute stroke admission that occurred at least 1 year apart); only the first episode was included in the analysis; and so all subsequent stroke admissions outside of the first 1-year episode (n=362) were ignored (Figure 4.1). Linkage was done using Match*Pro v2.4.1. Detailed information about the linkage methodology can be found in Chapter 3 of this dissertation.

Figure 4.1: Probabilistic linkage between MiSP and MVC and selection of final analytical sample.



To supplement the analysis data with hospital and system-level variables, data collected by the annual hospital survey of the American Hospital Association were matched to the linked MVC index strokes (N=22,889) dataset according to hospital ID and admission year. Patients were excluded if they died during hospitalization or were discharged to hospice care or against medical advice (Figure 4.1). The final dataset included 19,382 1-year stroke episodes of care.

4.3.4 Data Elements (Potential Predictors)

More than 500 patient-level variables are collected in MiSP using the American Heart Association's GWTG-Stroke case report form (CRF). This includes data on demographics (age, sex, race/ethnicity), clinical stroke presentation (e.g., mode of transportation, last time known well, pre-stroke disability, stroke severity), treatments including tPA and EVT, brain imaging (MRI, CT), ED utilization, more than 20 stroke related comorbidities (patient medical history), in-hospital complications (i.e., pneumonia, DVT, PE, UTI), length of stay, discharge medications and discharge destination. Based on clinical relevance, data availability (missingness), and prior GWTG-Stroke publications, 64 variables were selected for further analysis. A complete list of these variables can be found in Table 4A.1 (Table 1 in Appendix). The process of variable exclusion was conducted independently by two authors, (RH, MR), disagreements were resolved by consensus. Only 4 variables were recorded as continuous variables in the CRF (age, length of stay, onset to door time, and admission NIHSS) but these were recoded to categorical variables using thresholds published in the literature (Table 4A.1).

Nearly all MiSP variables suffered from missing observations. Missing occurs because many variables are listed as optional data fields. In addition, some of the missingness occurs because information was not documented in the patient's medical records, or because the staff responsible for data abstraction for the registry did not record it, however, it's not possible to distinguish between the two scenarios. Observations are also missing due to skip patterns found

in the CRF where some variables or sections will be skipped if a certain criterion is not met according to the information filled previously. This missingness pattern is labeled as not applicable (for example, hemorrhagic stroke variables are labelled not applicable for cases of ischemic stroke). For the 64 selected variables of interest, we decided to include missing observations as its own category in the analysis because missing observations can be medically meaningful. We also reassigned the missing values of some variables to no or absent category through medical reasoning or by using the value of other reported variables in a process called documentation by exception. For example, if the patient had a not documented (ND) record for hospital acquired pneumonia then we recoded the variable to not having pneumonia (No). The complete list of variables from MiSP, American Hospital Association database, and MVC along with their level of missingness is provided in Table 4A.1. Data recording was done using SAS software v9.4 (Cary, NC).

The MVC administrative dataset included all post-acute health services claims submitted to the patient's insurance provider during the 12 months period post hospital discharge post-discharge health services included in-patient rehabilitation (IPR), skilled nursing facility (SNF), long term acute care hospital (LTACH), emergency department (ED), outpatient rehabilitation (OPR), home health care (HHC), outpatient medical visits (OP), and hospital readmissions. Mortality data were only available for Medicare FFS insured patients. In addition to the post-acute claims, MVC data included 79 Hierarchical Condition Category (HCC) binary comorbidity codes related to the index stroke discharge which were used in the model. HCC codes document relevant health conditions for each beneficiary by ranking ICD-10 codes into categories with similar cost patterns.⁵⁴ HCC codes are generated using a model that scans the beneficiaries ICD-10 codes associated with each claim that is submitted by the provider. The HCC codes can be

utilized with other information e.g. demographics to calculate risk adjusted payment rates for Medicare advantage beneficiaries.⁵⁴ Beneficiaries with higher number of HCC codes indicate a higher predicted healthcare cost.⁵⁴ A complete list of the HCC codes can be found in Table 4A.1.

Variable exclusion was conducted by the same clinical researchers (RH and MR) on the hospital and system level variables collected by the American Hospital Association database where 20 variables used previously in similar research projects were selected as potential predictors. A complete list of these variables can be found in Table 4A.1.

4.3.5 Outcome Variables

The primary outcome of this research was all-cause-readmission (recorded as a binary event) to an acute care hospital within 30-day and 1-year of the index acute-stroke discharge as determined by the MVC claims data. All-cause readmission was defined as a post-index discharge admission to an acute care hospital for any reason. We did not distinguish between planned and unplanned admissions because we did not have access in the MVC data to secondary ICD-10 diagnostic, ICD-10 procedural, or clinical classification software (CCS) codes for the readmission events. These variables are required to implement the Agency for Healthcare Research and Quality (AHRQ) algorithm to identify unplanned events.⁵⁵

4.3.6 Descriptive Statistics

Chi-square test was applied to identify the significant difference between the two groups of patients (i.e., readmitted and not readmitted for 30-day and 1-year) for each predictor. In addition, univariate logistic regression analysis was performed and odds ratios with their corresponding confidence intervals were generated. Finally, we also reported the cumulative incidence (%) of readmission for each predictor at each time point (data reported Table 4A.1 in the Appendix). Descriptive statistics of the 31 participating hospitals characteristics are presented

in Table 4A.2. To assess the population variability between hospitals, we reported the summary descriptive statistics of i.e., mean age, proportions by sex, race, insurance coverage, and stroke type for each hospital and reported the chi-square test p-value in Table 4A.3. To assess the potential heterogeneity in 30-day and 1-year readmission outcome rates across hospitals we plotted the hospital-specific and overall average outcome rates and their corresponding 95% confidence interval and implemented a mixed effects model where hospital was designated as a random effect and sex, age, race, and stroke type were designated as fixed effects. Likelihood chi-square test p-value of the mixed effects model was reported. Descriptive statistics was performed using R v4.2.3 in RStudio.

4.3.7 Multivariable Model Development

Our primary objectives were to compare the predictive performance of LASSO logistic regression (an example of a simple ML method), and two advanced ML based methods (i.e., XGBoost and ANN), when applied to 30-day and 1-year all-cause readmissions. Having identified the best performing modelling approach, we then identified the most important patient and hospital level predictors. Prior to building the ML based models we decided to explore the independent effect of potential predictors by adding each independent variable to a logit model that included four priori potential confounders i.e., age, sex, race, and stroke type (ischemic or hemorrhagic) – which we referred to as the base logistic regression model. We examined the independent effect of each of the predictors when added to the base model and reported its statistical significance (based on the likelihood ratio chi square test) and its effect on the discriminant performance as illustrated by the change in the area under the receiver operating characteristic curve (Δ AUC) compared to the base model (Table 4A.4). This process was not used to determine which variables will enter any of the ML models, but rather conducted as an

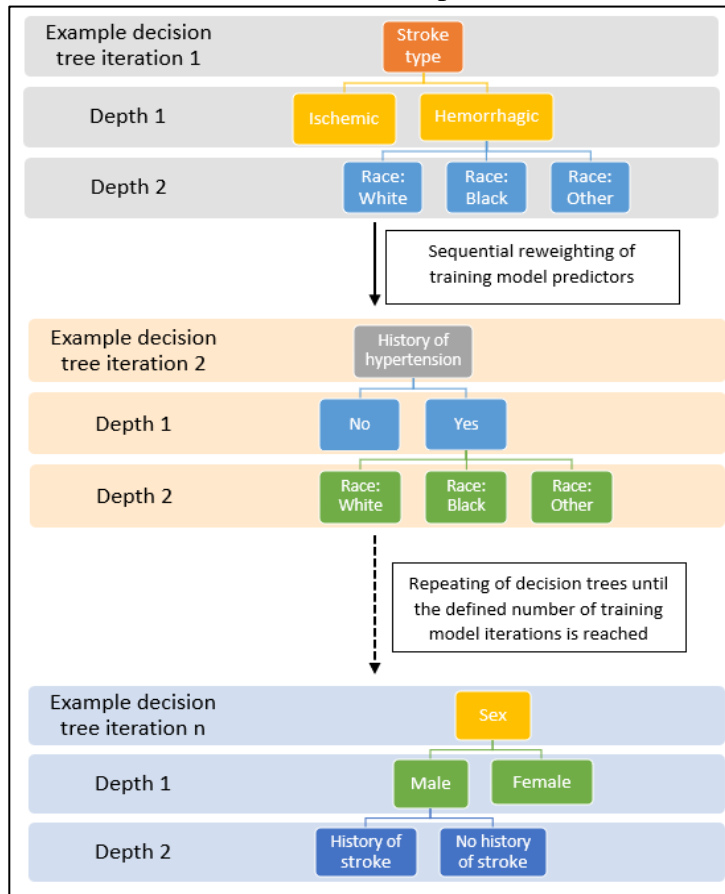
exploratory step to know which variables might end up being the most important in predicting readmission.

To address our primary aims we compared three alternative ML methods to build predictive models i.e., LASSO logistic regression, XGBoost and ANN. Least absolute shrinkage and selection operator (LASSO) logistic regression, is a penalized regression approach that estimates the regression coefficients by minimizing a loss function consisting of the negative - log-likelihood plus a penalty on model complexity proportional to the sum of the absolute values of the regression coefficients, that is $\{\hat{\mu}, \hat{\boldsymbol{\beta}}\}_{\lambda} = \underset{\mu, \boldsymbol{\beta}}{\operatorname{argmin}}\{-\sum_{i=1}^n \log[p(y_i|\mathbf{x}_i, \mu, \boldsymbol{\beta})] + \lambda \sum_{j=1}^p |\beta_j|\}$. Above, $p(y_i|\mathbf{x}_i, \mu, \boldsymbol{\beta})$ is the probability of the i^{th} data point (in our case a Bernoulli likelihood) given the predictors (\mathbf{x}_i), viewed as a function of the intercept (μ) and the regression coefficients ($\boldsymbol{\beta}$), $\lambda \geq 0$ is a penalty hyperparameter, and $\sum_{j=1}^p |\beta_j|$ is the L-1 norm of the regression coefficients (note that the intercept is not penalized).^{56, 57} Setting $\lambda = 0$ leads to a standard logistic regression fitted via maximum likelihood. Large values of λ can lead to sparse solutions (i.e., some estimated effects being equal to zero). Relative to logistic regression LASSO can produce variable selection and renders estimates that shrunk towards zero hence producing more sparse (more interpretable with less number of predictors) models. Usually, LASSO models are fitted over a grid of values of λ (aka the regularization path), starting from the smallest λ that renders all effects equal to zero (λ_{max}) and lowering to values two or three orders of magnitude smaller (e.g., $\lambda_{min} = \lambda_{max}/1000$). The model is fitted in a cross-validation setting, and an optimal value of λ within the grid is obtained by maximizing the prediction accuracy (AUC) in testing data. Then the model is fitted to the entire data set using the chosen value of λ .^{56, 57} The LASSO selection method overcomes limitations present in previous shrinkage methods (i.e., ordinary least squares, ridge regression, and subset selection) by reducing the variance in the

predicted values and providing a more stable and interpretable model which in turn increases the prediction accuracy.⁵⁶ In addition, LASSO is known to have desirable properties for regression models with a large number of covariates, and various efficient optimization algorithms are available to find the estimates.⁵⁷

Extreme gradient boosting (XGBoost) is a type of ensemble machine learning method that combines multiple regression trees- a weak learner's methods in a series where errors of the preceding algorithms are considered to reduce bias in the estimates and obtain a better prediction accuracy.^{58, 59} More specifically, a decision tree method will be repeatedly applied to modified versions of the training model using bootstrapping for a predefined number of iterations (Figure 4.2). These steps will sequentially reweight the predictors in the training model where classified predictors will get lower weights and misclassified predictors will get higher weights according to the previous iterations (Figure 4.2).⁵⁹ After the boosting algorithm is complete the final model is a weighted average of the predictions of each tree.⁵⁹ XGBoost is described as a robust method in dealing with data with complex and nonlinear relationships between variables and outcomes.⁶⁰ After conducting multiple trials of model training using different hyperparameters- model settings (i.e., number of iterations, tree depth, and learning rate) and implementing an internal cross validation technique during model training, the following hyperparameters produced the highest predictive accuracy; 50 iterations, with a maximum tree depth between 1 and 4, and learning rate (weight shrinkage) of 50%.

Figure 4.2: Example of XGBoost architecture consisting of 50 decision trees and a tree depth of 2.

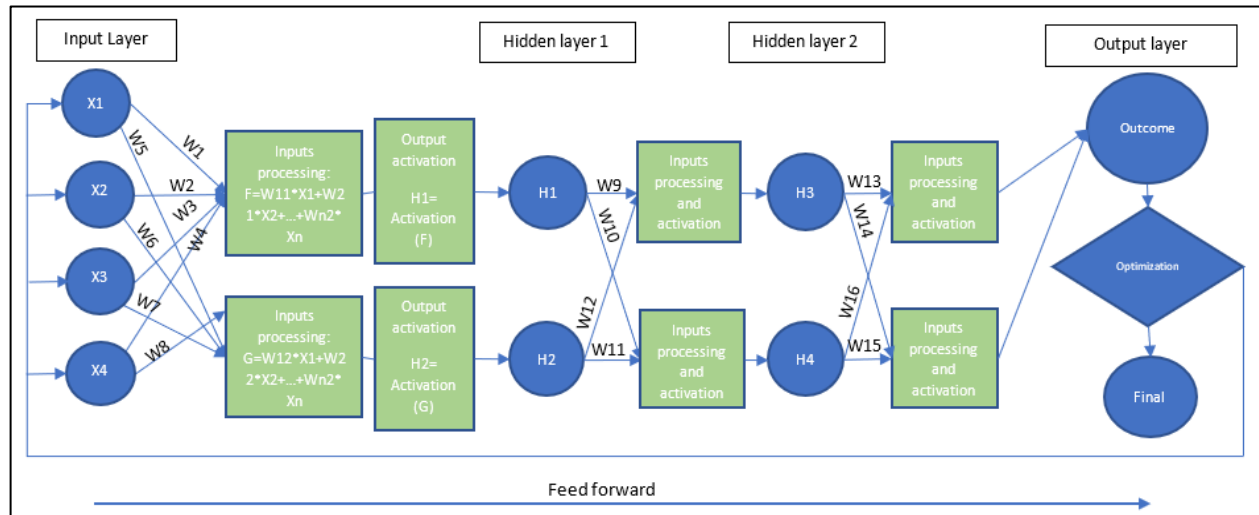


Artificial neural networks (ANN) is an advanced machine learning method used for analysis of high dimensional and big datasets (e.g., images, speech, and unstructured data like natural language processing).⁶¹ ANN is based on a building block of neurons and layers. A neuron consist of a linear combination of inputs and a non-linear activation function that produces an output neuron.⁶² The input layer has all of the potential predictors. The hidden layer houses the neurons which are linear combinations of predefined weighted inputs that the sum of which are nonlinearly transformed using an activation function (e.g., sigmoid and *ReLU* (rectified linear unit)).⁶¹ *ReLU* is the activation function of choice because it is easier to compute and store.⁶¹ The activation function captures nonlinear relationships between the inputs when the neurons are combined to another hidden layer or toward the output layer (Figure 4.3).⁶³ These

neurons are not directly observed and are held in the hidden layer. The final output layer is a linear model that uses the transformed sum of the weighted neurons in the hidden layer as inputs.⁶¹ During each step weights are calculated and used to calculate the output model, but the initial model often has large prediction error. To minimize the prediction error an optimization algorithm has to be implemented to learn the optimal weights. The most common optimization algorithm is scholastic gradient decent (Figure 4.3).⁶³ This algorithm adjusts the weights in small amounts and assess the impact of such changes on the outputs and errors repeatedly using examples from the training dataset.⁶³ The algorithm stops when the prediction errors can't be reduced further. To serve the purpose of predicting in our case a binary outcome, the final step of transformation used a sigmoid function because it transforms the input into a probability that lies between 0 and 1. Our network architecture is a feed-forward network architecture where input signals are flowing only in one direction from the input layer to the output layer. After conducting multiple trials of model training using different hyperparameters (i.e., number of hidden layers, number of hidden neurons in each layer, and dropout rate) and implementing an internal cross validation technique, the following hyperparameters produced the highest predictive accuracy; number of hidden layers between 1 to 5, number of hidden neurons= 30 in each layer, dropout rate in each layer =10%.

It is worth mentioning that our predictive methodologies in terms of how the machine algorithm learns from the data fall under the term of supervised learning. Supervised learning takes place when machine learning algorithms learn from the data with information available on the outcome to develop a prediction model tasked to either classify an outcome (e.g., 1-year readmission into yes or no) or use the regression to calculate a value based on available predictors (e.g., predicting duration of stay).⁶²

Figure 4.3: Artificial neural network detailed architecture consisting of 4 predictors, 2 neurons in each hidden layer (n=2), and a single output.



4.3.8 Comparison Between the Three Machine Learning Methods

Understanding the mechanism by which each of the three ML methods are built on is important but these models have different capabilities in handling high-dimensional data (data with many variables or features), modelling non-linear patterns, ability to select predictors, and modelling interactions. All of the three ML methods are capable of handling high dimensional data and modelling nonlinear patterns (Table 4.1).^{56, 57, 59, 63} However, LASSO logistic regression is the only method from the three that is capable of predictor selection,⁵⁷ thus it is not advised to use it when all of the features of the high dimensional data are included in the model because it can lead to loss of information especially in the event that the number of data points is small (small dataset).⁶⁴ In terms of modelling interactions, all of the three methods can model interactions with LASSO and XGBoost having the option to model all or partially select certain interactions, whereas ANN would model all the interactions by default without the capability to control this feature.^{56, 57, 59, 63} For LASSO only interactions between predictors that were selected would be included in the model.^{56, 57} We did not explore interactions in our LASSO and

XGBoost models because of the big number of interactions that would need to be specified and we relied on ANN model to examine the effect of all the interactions automatically.

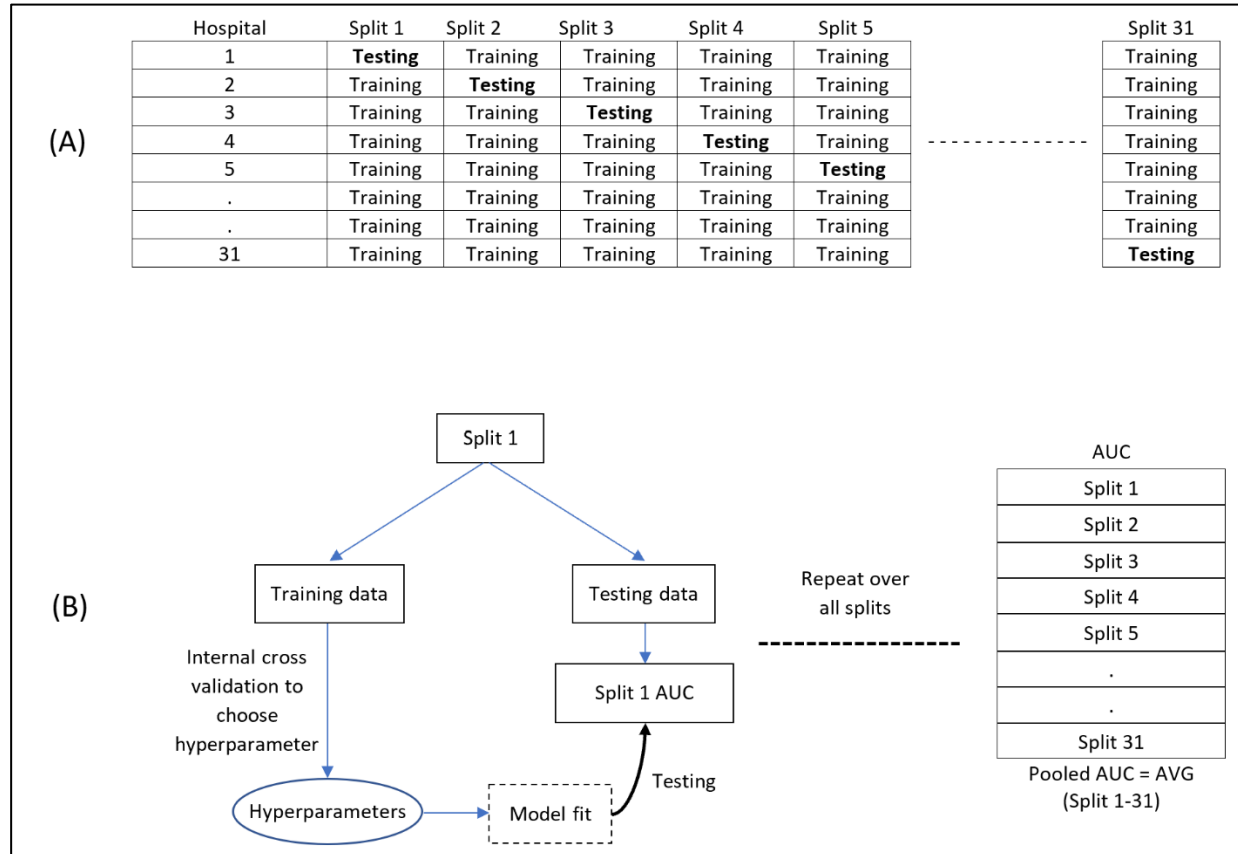
Table 4.1: Comparison of modelling capabilities that can be handled by LASSO logistic regression, XGBoost and ANN machine learning methods.

Model capability	LASSO logistic	XGBoost	ANN
Handling high dimensional data	Yes (not advised in small data sets)	Yes	Yes
Model non-linear patterns	Yes	Yes	Yes
Predictor selection	Yes	No	No
Model Interactions	Yes (user defined)	Yes (user defined)	Yes (by default)

4.3.9 Evaluation of Prediction Performance

We conducted a leave-hospital-out cross validation strategy to evaluate the prediction accuracy of the developed ML models (i.e., LASSO logistic regression, XGBoost, and ANN) in a setting that assumes that the model is used to predict outcomes of hospitals that did not contribute data to the model training. Specifically, hospital site was used as the unit for splitting the data into training and testing datasets, each hospital (n=31) is left out once as a testing dataset for validation of a model based on the remaining hospital sites which are used in the training dataset (Figure 4.4A). This cross-validation approach is what Steyerberg, et al. refers to as internal-external cross validation strategy.⁶⁵ This strategy promises to enhance the external validity of the developed models and accommodate the addition of newly participating hospitals in future model trainings.⁶⁵ The final model performance is based on the pooled (average) AUC of all hospital-specific AUCs (Figure 4.4B).

Figure 4.4: (A) Schematic representation of the internal external cross validation data split approach (B) Schematic representation of training and testing approach and pooled model performance.*

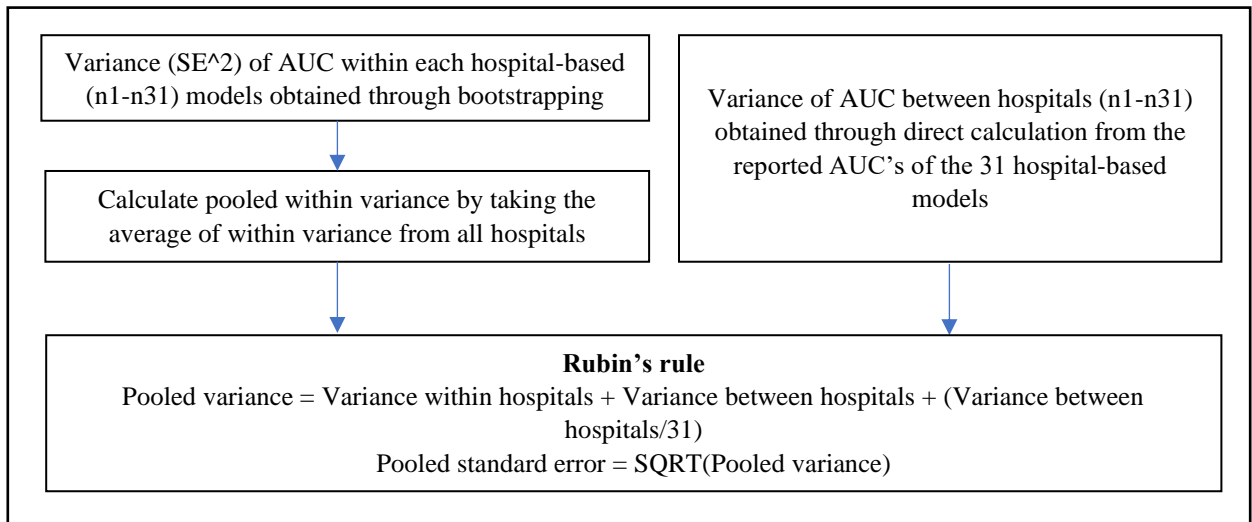


*Hospital site was used as the unit for splitting the data into training and testing datasets. Each hospital (n=31) is left out once as a testing dataset for validation of the model trained on the remaining sites.

To obtain standard errors for the estimated AUCs we bootstrapped (n=1,000) the vectors of outcomes and predictions, for each bootstrap sample we computed the AUC and obtained a within-hospital standard error as the standard deviation of the bootstrap AUCs. Then, to derive our final pooled standard errors we used Rubin's rule which considers both the within hospital standard error as well as between hospital variability (Figure 4.5).⁶⁶ Between hospital variance resembles the variability of hospital-based data split AUC from the pooled AUC. It is calculated by taking the difference between each hospital-based data split AUC and the pooled AUC of all splits, then squaring the differences, and finally dividing the sum of squares by the number of

splits ($n=31$). Within hospital variance represents the variability of the bootstrapped samples AUC from the pooled AUC for each hospital-based split. Model performance (AUC) of our included predictive methods i.e., base logistic regression model, LASSO logistic regression, XGBoost, and ANN using different combination of data sources i.e., MiSP, American Hospital Association, and MVC with their corresponding confidence intervals according to the utilized data source are presented in Table 4A.5.

Figure 4.5: Rubin's rule to calculate pooled standard error from within and in between hospital based models AUC estimates.



We implemented a sign test and reported the P -values in Table 4A.5 to fulfill our primary objectives to compare the predictive performance of pooled model performance of ML methods using only registry data. To visualize effect of adding predictors on predictive performance (AUC) of 30-day and 1-year readmission outcome of each hospital-specific models using only the registry data, we plotted the AUC of the best performing ML method as predictors were added to the model and included the average AUC of all hospital-specific models and the corresponding 95% confidence interval. We ranked the predictors with the highest impact on 30-day and 1-year readmission predictive accuracy according to the order they got selected by the best performing ML model across all the hospital-specific models for the first 15 predictors.

To fulfill our secondary objective to compare the predictive performance of pooled model performance of ML methods using all the combinations of data sources (22 combinations; 3 ML methods*7 data combinations in addition to the base logistic model), we implemented a sign test and presented the results grouped according to data sources using a bar plot and according to methods using a matrix of the 22 combinations of methods and data sources where the cross section of each of these combinations was examined. In addition, to indicate the predictive performance of hospital-specific models when using different combinations of methods and data sources, AUCs produced by the 31 hospital specific models were compared between every two data and method combinations where the proportion of the 31 hospital specific models that had a higher or equal AUC for each cross section of the examined combination of methods and data sources (AUC row \geq AUC column) was presented using a heatmap. Data analysis was performed using R v4.2.3 in RStudio.

4.4 Results

4.4.1 Readmission Event Rates and Patient-Level Characteristics

Using the final linked dataset of 19,382 discharged stroke patients from 31 hospital sites during the 5-year study period, 2,724 (14.1%) and 8,169 (42.1%) were readmitted 30-days and 1-year post discharge, respectively (Table 4.2). Our population was mostly older than 65 years (78.5%), predominantly white (79.7%), 52.2% were female, and 62.9% were insured by Medicare fee for service, 20.9% by Medicare advantage and 16.2% by private plans (Table 4.2). The majority of index hospitalizations were for ischemic stroke (87.3%), with a predominance of minor stroke (NIHSS <5 , 56.4%). Half (50.1%) of index stroke hospitalizations had a final discharge disposition to home (Table 4.2). There was a noticeable decrease in the number of stroke discharges during 2020 compared to years of 2016-2019 (3,062 vs 3,811-4,327). There

were significant univariate associations between 30-day and 1-year post discharge readmission and age, race, insurance provider, stroke type, admission NIHSS, and discharge destination (Table 4.2). A complete set of predictors from the three data sources i.e., MiSP, MVC, and American Hospital Association with their corresponding Chi-square test, odds ratios with their confidence intervals, and incidence of the outcomes were reported in Table 4A.1.

4.4.2 Hospital-Based Characteristics

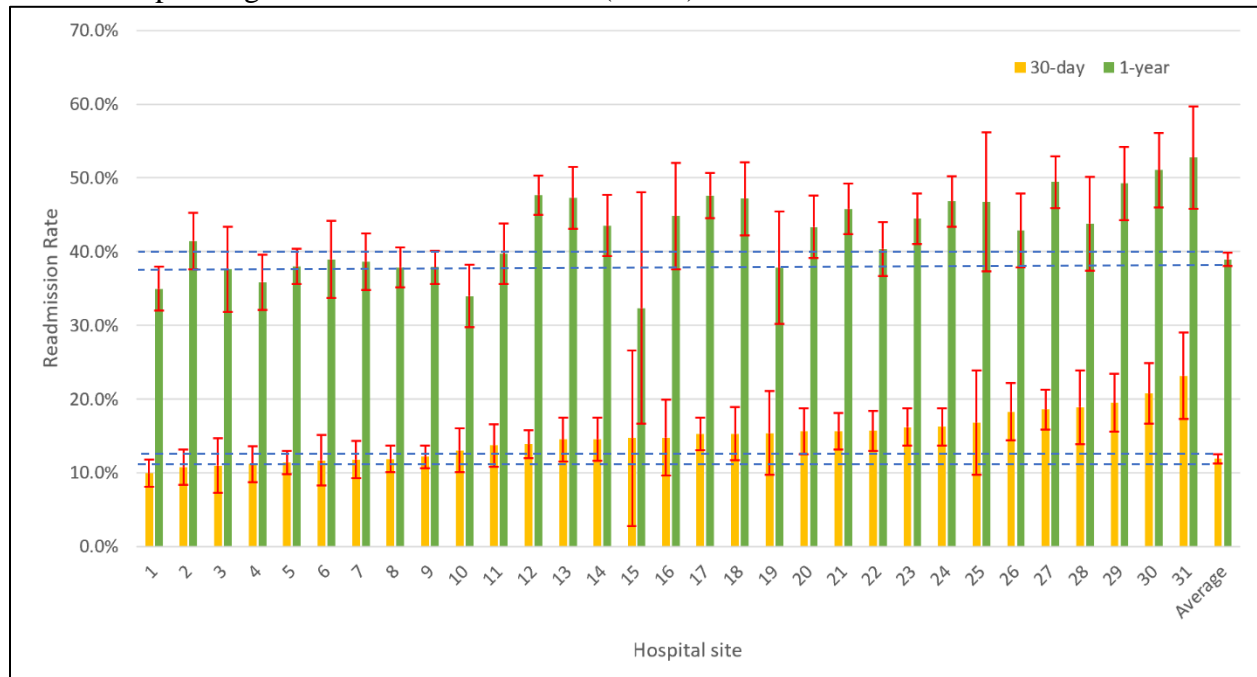
Thirty-day and 1-year readmission rates among 31 participating hospitals ranged between 9.9%-23.1% and 34%-49.4%, respectively (Figure 4.6). The mixed effects model reported a significant difference in 30-day and 1-year readmission rates between hospital sites ($P < 0.001$) indicating heterogeneity.

Among the 31 hospitals, 18 (58%) had over 300 bed capacity, two (6.5%) were operated by for profit organizations, 29 (93.6%) were located in metro areas (defined by the US Census Bureau), 14 (45.2%) were rural referral centers (high-volume acute care rural hospitals that treat a large number of complicated cases), 26 (83.9%) were part of a health system, 20 (64.5%) were accredited as primary stroke centers, two (6.4%) were accredited rehabilitation centers, 24 (77.4%) were teaching hospitals, and 27 (87.1%) had magnetic resonance imaging capabilities (Table 4A.2).

Table 4.2: Study population description using a selected list of potential predictors of readmission from the Michigan Stroke Program registry (n= 19,382).

Predictor	Value	Distribution % (Total=19,382)	30-days all-cause readmission (N= 2,724)					1-year all-cause readmission (N= 8,169)				
			Rate %	OR	95% CI	χ^2 test LRT	χ^2 test p-value	Rate %	OR	95% CI	χ^2 test LRT	χ^2 test p-value
Age category	<65	21.5	13.7		Ref	6.5	0.088	36.6		Ref	98.4	<0.001
	65-74	29.2	13.5	1.0	0.9-1.1			41.0	1.2	1.1-1.3		
	75-84	28.8	15.1	1.1	1-1.3			44.4	1.4	1.3-1.5		
	>=85	20.4	13.8	1.0	0.9-1.1			46.5	1.5	1.4-1.6		
Race	White	79.7	13.7		Ref	12.3	0.006	41.0		Ref	61.5	<0.001
	Black	14.6	15.9	1.2	1.1-1.3			48.9	1.4	1.3-1.5		
	Other	1.3	17.6	1.4	1-1.9			38.8	0.9	0.7-1.2		
	Not Documented	4.3	14.2	1.0	0.9-1.3			42.1	1.0	0.9-1.2		
Latino ethnicity	No/Unable to Determine	96.3	14.1		Ref	1.01	0.313	42.4		Ref	9.7	0.001
	Yes	3.7	12.8	0.9	0.7-1.1			36.5	0.8	0.7-0.9		
sex	Male	47.8	14.3		Ref	0.7	0.417	41.9		Ref	0.5	0.5
	Female	52.2	13.9	1.0	0.9-1			42.4	1.0	1-1.1		
Insurance	BCBSM Preferred Provider Organization (PPO)	12.0	9.3		Ref	84.5	<0.001	27.4		Ref	406.9	<0.001
	BCBSM PPO Medicare Advantage	15.7	14.0	1.6	1.3-1.9			41.3	1.9	1.7-2.1		
	BCN Health Maintenance Organization (HMO)	4.2	9.6	1.0	0.8-1.4			26.1	0.9	0.8-1.1		
	BCN HMO Medicare Advantage	5.2	12.2	1.4	1.1-1.7			39.1	1.7	1.5-2		
	Medicare Fee For Service	62.9	15.4	1.8	1.5-2.1			46.5	2.3	2.1-2.5		
Stroke type	Ischemic	87.3	13.5		Ref	28.9	<0.001	41.7		Ref	12.4	<0.001
	Hemorrhagic	12.7	17.7	1.4	1.2-1.5			45.4	1.2	1.1-1.3		
Admission year	2016	19.7	12.9		Ref	7.2	0.126	43.4		Ref	14.5	0.005
	2017	22.3	14.7	1.2	1-1.3			43.0	1.0	0.9-1.1		
	2018	22.3	13.7	1.1	0.9-1.2			43.0	1.0	0.9-1.1		
	2019	19.9	14.5	1.1	1-1.3			40.7	0.9	0.8-1		
	2020	15.8	14.6	1.1	1-1.3			40.0	0.9	0.8-1		
Admission NIHSS	0	15.4	11.6		Ref	159.9	<0.001	36.1		Ref	226.6	<0.001
	1-4	41.0	11.6	1.0	0.9-1.1			38.3	1.1	1.0-1.2		
	5-15	25.6	15.4	1.4	1.2-1.6			46.0	1.5	1.4-1.6		
	16-20	4.3	22.5	2.2	1.8-2.7			53.5	2.0	1.7-2.4		
	>20	3.8	21.4	2.1	1.7-2.6			50.3	1.8	1.5-2.1		
Not Documented	10.0	18.0	1.7	1.4-2.0	49.4	1.7	1.5-1.9					
Discharge disposition	Home	50.1	10.2		Ref	931.8	<0.001	34.5		Ref	771.0	<0.001
	Skilled Nursing Facility (SNF)	24.9	12.8	1.3	1.2-1.4			44.6	1.5	1.4-1.6		
	Inpatient Rehabilitation Facility (IRF)	1.6	17.8	1.9	1.4-2.6			62.1	3.1	2.5-3.9		
	Long Term Care Hospital (LTCH)	0.6	26.2	3.1	2.0-4.8			54.2	2.2	1.5-3.3		
	Other	21.4	19.0	2.1	1.9-2.3			52.0	2.1	1.9-2.2		
ND	1.5	83.7	45.1	32.8-62.1	90.7	18.4	12.4-27.4					

Figure 4.6: 30-day and 1-year hospital specific and hospital wide average readmission rates and their corresponding 95% confidence intervals (n= 31).



Hospital-specific population characteristics varied where <65 years olds constituted between 13.5%-34.6% of total stroke discharges, 17.6%-98.6% were white, 44.9%-57.4% were females, 54.6%-72.1% were covered by Medicare FFS plan, 12.6%-69.1% covered by Medicare advantage plans, 8.8%-26.6% were covered by private plans, and 78.2%-98.9% had an ischemic stroke (Table 4A.3).

4.4.3 Base Logistic Regression Model Results and Registry Variables with the Highest Prediction Performance

Our simple base logistic regression predictive model that included only sex, age, race, and stroke type produced a 30-day and 1-year pooled predictive accuracy (AUC) of 0.538 and 0.558 respectively (Table 4A.4). When examining the effect of adding a single predictor to the base model on the association with 30-day and 1-year outcomes, almost all of the 64 MiSP registry and 80 MVC data predictors were statistically significantly ($P<.05$) associated with risk of readmission, whereas only a third of the 20 hospital characteristics were statistically

significantly associated with readmission risk (Table 4A.4). The 10 registry derived variables that produced the greater increase in predictive accuracy from the base 30-day readmission model (as determined by the change in AUC) were: discharge disposition, assessment for rehabilitation prior to discharge, ambulatory status on discharge, admission NIHSS, admission duration (length of stay), ambulatory status on admission, persistent or paroxysmal atrial fibrillation/flutter upon discharge, history of heart failure, Cholesterol reducing treatment upon discharge, and history of coronary artery disease/ prior myocardial infarction (Table 4A.4). The 10 variables that produced the greatest increase in predictive accuracy for 1-year readmission were discharge disposition, admission duration, ambulatory status on discharge, history of chronic renal insufficiency, history of diabetes mellitus, admission NIHSS, history of coronary artery disease/ prior myocardial infarction, history of atrial fibrillation/flutter, history of heart failure, and persistent or paroxysmal atrial fibrillation/flutter upon discharge (Table 4A.4).

4.4.4 Relative Performance of Machine Learning Based Predictive Models Using Registry Data and Important Predictors

Utilizing data from only the MiSP registry, the three ML methods i.e., LASSO logistic regression, XGBoost, and ANN produced a 30-day readmission pooled predictive accuracy (AUC) of 0.677 (95% CI: 0.654-0.700), 0.676 (95% CI: 0.653-0.700), and 0.659 (95% CI: 0.640-0.677), respectively (Table 4.3). Prediction of 1-year readmission produced similar values of AUC (Table 4.3). Neither XGBoost nor ANN advanced ML methods produced statistically significantly higher 30-day or 1-year readmission AUC compared to LASSO logistic regression when utilizing data only from MiSP registry (Table 4A.5). LASSO logistic regression prediction method was regarded as the best performing method using registry data because it is the simplest ML method, and it produced similar AUC to XGBoost and ANN.

Table 4.3: Pooled predictive accuracy of 30-day and 1-year readmission using combinations of different methods and data sources.

Method	Data source	30-day readmission		1-year readmission	
		AUC	95% CI [^]	AUC	95% CI [^]
Base Logistic model	Sex, age, race, stroke type	0.526	0.506 - 0.546	0.545	0.529 - 0.560
	MiSP*	0.677	0.654 - 0.700	0.668	0.650 - 0.686
LASSO logistic regression	AHA**	0.535	0.511 - 0.558	0.541	0.527 - 0.555
	MVC***	0.639	0.619 - 0.659	0.668	0.652 - 0.684
	MiSP+AHA	0.678	0.655 - 0.701	0.667	0.648 - 0.685
	MiSP+MVC	0.690	0.664 - 0.715	0.697	0.681 - 0.712
	AHA+MVC	0.638	0.618 - 0.659	0.667	0.651 - 0.683
	MiSP+AHA+MVC	0.690	0.666 - 0.714	0.696	0.681 - 0.712
	MiSP*	0.676	0.653 - 0.700	0.670	0.652 - 0.688
XGBoost	AHA**	0.529	0.510 - 0.548	0.547	0.533 - 0.561
	MVC***	0.639	0.620 - 0.659	0.666	0.651 - 0.681
	MiSP+AHA	0.676	0.653 - 0.699	0.671	0.653 - 0.688
	MiSP+MVC	0.684	0.662 - 0.707	0.691	0.676 - 0.707
	AHA+MVC	0.637	0.618 - 0.657	0.665	0.648 - 0.681
	MiSP+AHA+MVC	0.683	0.659 - 0.707	0.690	0.674 - 0.706
	MiSP*	0.659	0.640 - 0.677	0.665	0.653 - 0.677
ANN	AHA**	0.549	0.533 - 0.564	0.553	0.542 - 0.563
	MVC***	0.641	0.624 - 0.657	0.669	0.659 - 0.680
	MiSP+AHA	0.662	0.643 - 0.681	0.662	0.650 - 0.673
	MiSP+MVC	0.678	0.659 - 0.698	0.690	0.679 - 0.701
	AHA+MVC	0.647	0.632 - 0.662	0.670	0.660 - 0.680
	MiSP+AHA+MVC	0.679	0.661 - 0.697	0.689	0.678 - 0.699

*MiSP: Michigan Stroke Program Registry.

**AHA: American Hospital Association's database.

***MVC: Michigan Value Collaborative.

[^]95% CI includes within and between hospital variances obtained through our hospital split model internal validation methodology.

Figure 4.7 Panel A and Figure 4.8 Panel A present the effect of adding predictors to LASSO logistic regression model that utilizes only MiSP data on the predictive accuracy (AUC) of 30-day and 1-year readmission for the 31 hospital-specific models and their pooled predictive accuracy and 95% confidence interval. In general, adding more predictors increased the AUC, with both

outcomes needing about 30 predictors to reach a plateau. The majority of 30-day and 1-year readmission hospital-specific prediction models produced an AUC between 0.65-0.70. The highest performing hospital-specific models demonstrated a 30-day AUC of 0.8 and 1-year AUC of 0.78. The 95% CI of the pooled predictive accuracy for 30-days and 1-year readmission was narrower at the early stages of predictor selection which may indicate that there is an agreement between the 31 hospital-specific models on the selected predictors. This was observed when we ranked the first 15 predictors that were chosen by the 31 hospital-specific models (Figure 4.7 Panel B and Figure 4.8 Panel B). For 30-day readmission prediction, the 31 hospital-specific models ranked discharge disposition, admission duration, ambulatory status at discharge, history of chronic renal insufficiency, and history of heart failure as the top 5 predictors. For 1-year readmission prediction, the 31 hospital-specific models selected discharge disposition, admission duration, history of chronic renal insufficiency, history of heart failure, and persistent or paroxysmal atrial fibrillation upon discharge as their top 5 predictors. Almost complete agreement in choosing the first 5 predictors in the 30-day and 1-year readmission models was therefore observed. Discharge disposition was selected as the top predictor of 30-day and 1-year readmission by all the 31 hospital-specific models. The majority of 30-day and 1-year readmission hospital-specific models selected predictors that were related to long-term comorbidities over predictors related to the clinical features of stroke, this was especially noted among 1-year readmission prediction models (Figure 4.7 Panel B and Figure 4.8 Panel B).

Figure 4.7: (Panel A) The effect of adding predictors to LASSO logistic regression model that utilizes only MiSP data on the predictive accuracy (AUC) of 30-day readmission for the 31 hospital specific models and their pooled predictive accuracy and 95% confidence interval. (Panel B) Ranking of the first 15 predictors selected by the 31 hospital specific models in (Panel A).

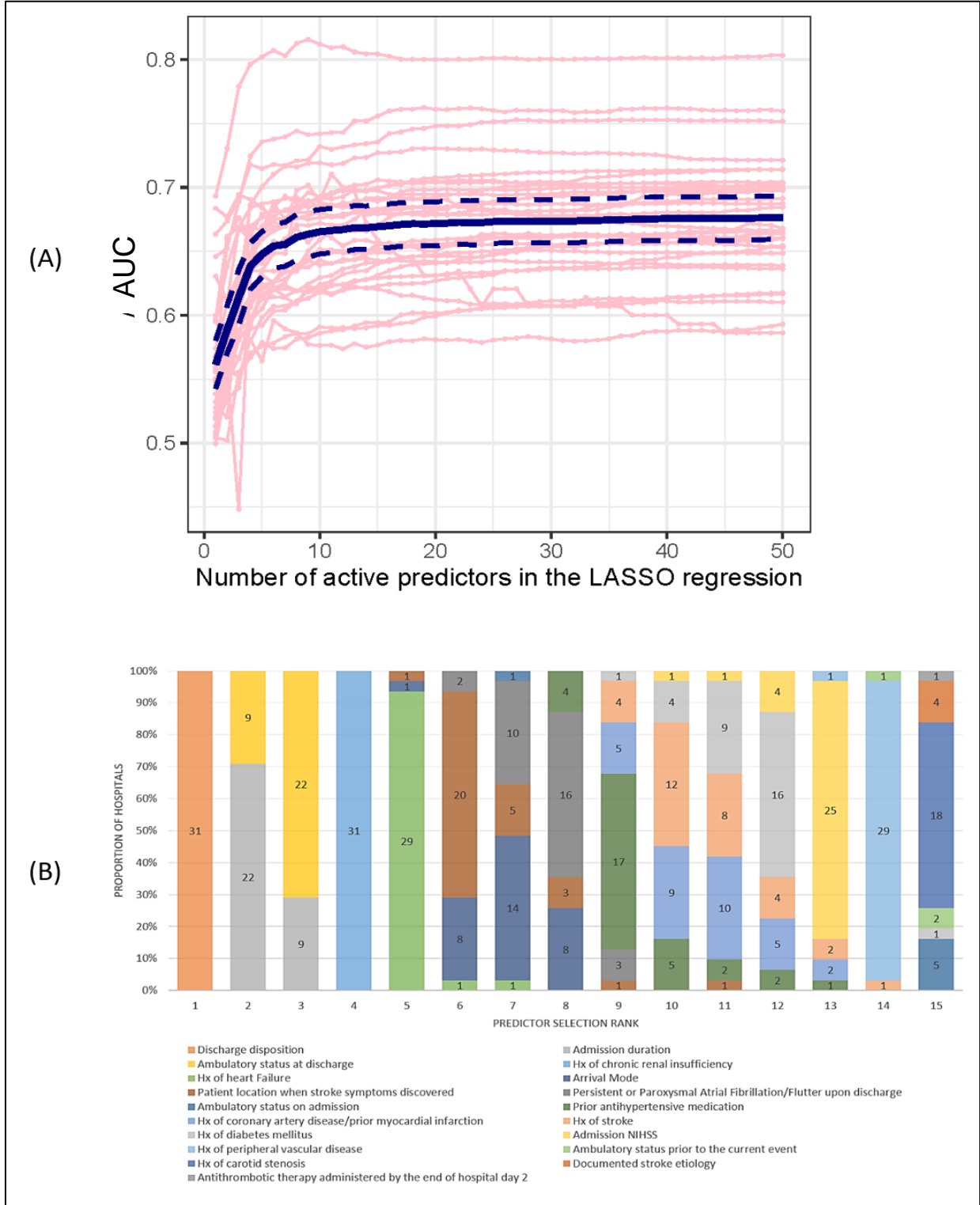
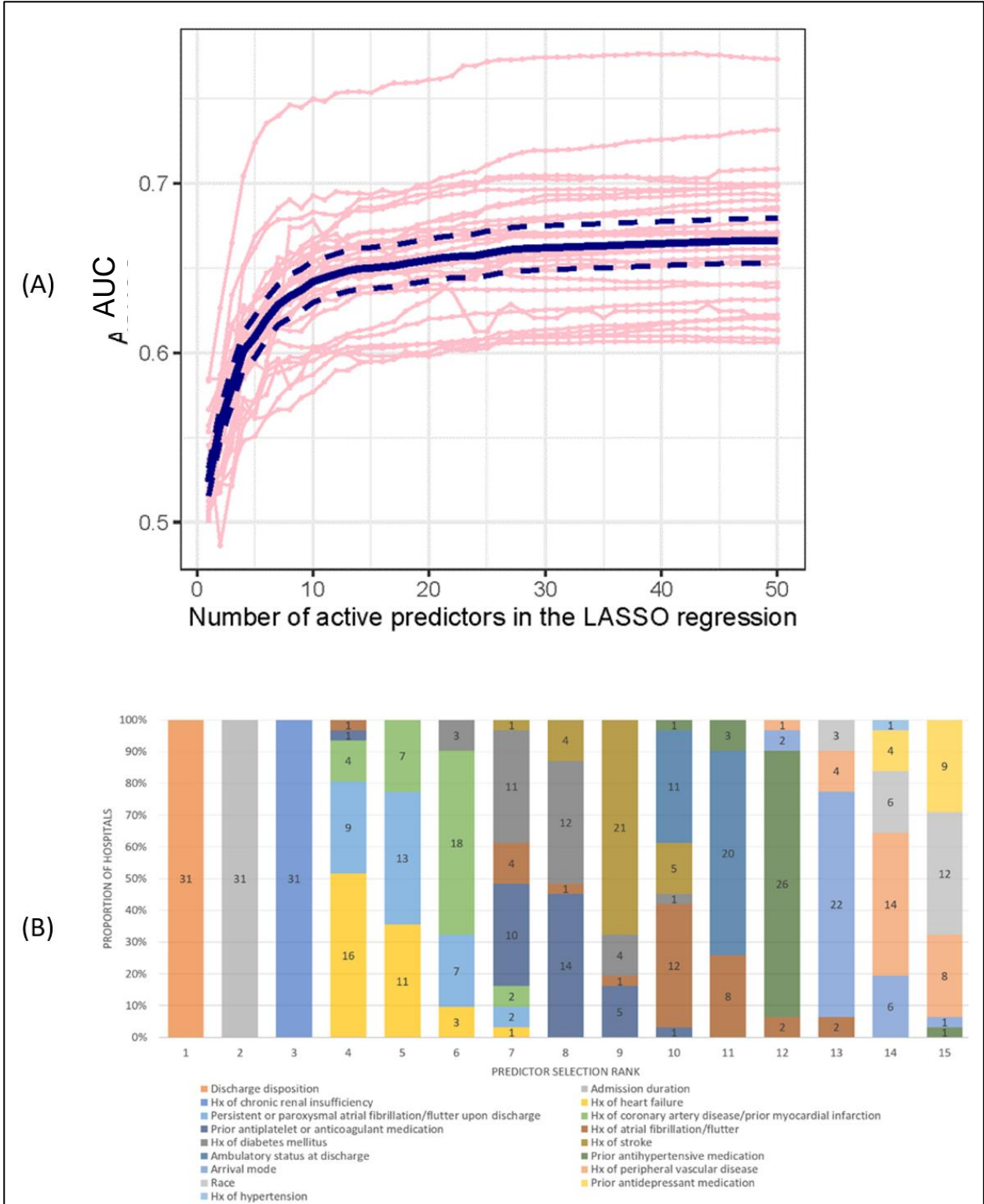


Figure 4.8: (Panel A) The effect of adding predictors to LASSO logistic regression model that utilizes only MiSP data on the predictive accuracy (AUC) of 1-year readmission for the 31 hospital specific models and their pooled predictive accuracy and 95% confidence interval. (Panel B) Ranking of the first 15 predictors selected by the 31 hospital specific models in (Panel A).

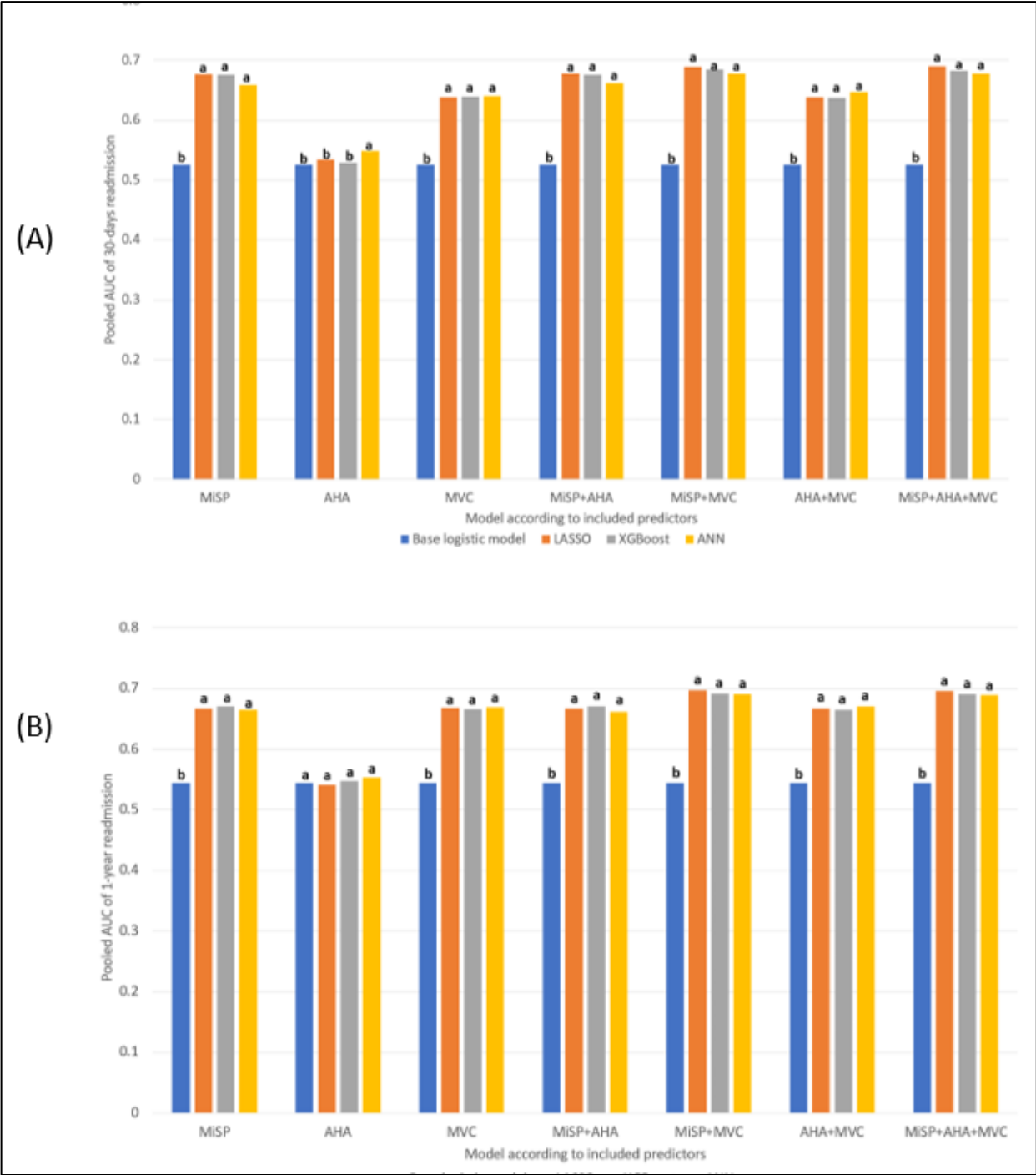


4.4.5 Relative Performance of Machine Learning Based Predictive Models Using Combinations of Different Data Sources

Pooled predictive performance of our models using different combination of data sources i.e., MiSP, MVC, and American Hospital Association, along with their corresponding 95% confidence intervals are presented in Table 4.3. Our ML methods produced statistically significantly higher 30-day and 1-year readmission AUC compared to the base logistic model (30-days AUC = 0.527 and 1-year AUC = 0.545) except when utilizing only American Hospital Association data (Figure 4.9). However, comparing the performance of ML methods when utilizing the same data source did not produce significant differences in 30-day and 1-year AUC except when using American Hospital Association's data to predict 30-day readmission (Figure 4.9).

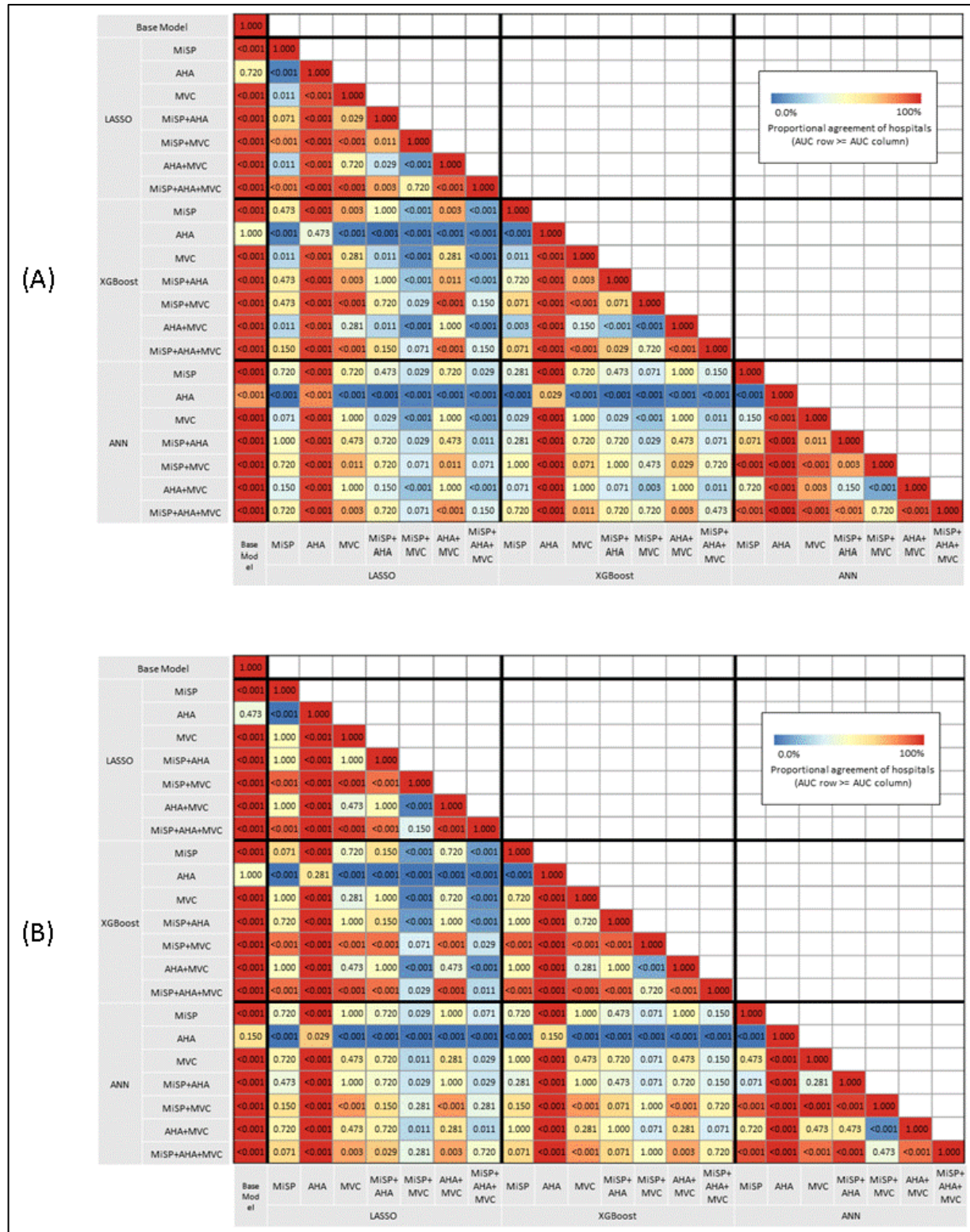
All of the ML methods reported the highest 30-day and 1-year pooled predictive accuracy (AUC) when MiSP registry and MVC claims data were utilized together (Table 4A.5). However, when we examine the 30-day and 1-year predictive accuracy of models that utilize either MiSP or MVC data or the combination of the two, we find that all the models produce similar AUC that are not statistically significant from each other (Figure 4.10). The statistically significant additive effect of adding MVC data to MiSP on 30-day predictive accuracy is not always present, it is apparent when using LASSO logistic regression and ANN methods but not XGBoost ($P < .001$) (Figure 4.10A), where it produced an increase in AUC of 0.013 and 0.019, respectively (Table 4A.5). In the instance of 1-year readmission, the statistically significant additive effect of adding MVC data to MiSP had a higher impact when compared to 30-day, which it is apparent when using LASSO logistic regression, XGBoost, and ANN ($P < .001$) (Figure 4.10B), where it produced an increase in AUC of 0.029, 0.021, and 0.025, respectively (Table 4A.5). When compared to MiSP and MVC, data from the American Hospital Association produced the worst

Figure 4.9: Pooled AUC of (A) 30-day and (B) 1-year readmission using base logistic regression, LASSO logistic regression, XGBoost, and ANN with different combinations of data sources.*



*Predictive models pooled AUC that do not share superscripts indicate significant ($P < 0.05$) sign test difference between methods utilizing the same data source.

Figure 4.10: Heat map of (A) 30-day and (B) 1-year readmission showing the proportion of hospital-specific models with a higher or equal AUC in the row (y-axis) compared to the column (x-axis) different methods and data sources and the corresponding pooled AUC sign test P -value.



30-day and 1-year predictive accuracy (Table 4A.5) (Figure 4.10; denoted by low percentage of hospitals - $P < .001$). It did not produce any statistically significant additive effect when combined with other data sources ($P > .05$) (Figure 4.10). When compared to MiSP data, MVC data alone was able to produce similar 1-year pooled predictive accuracy figures using all ML methods ($P > .05$), this was not the case when predicting 30-day readmission where MiSP outperformed MVC data in all ML methods except for ANN (Figure 4.10). Very high and very low proportion of the 31 hospital specific models that had a higher or equal AUC for each cross section of the examined combination of methods and data sources (AUC row \geq AUC column) was indicative of a higher and lower statically significant difference in pooled AUC across different models, respectively (Figure 4.10).

4.5 Discussion

4.5.1 Predictive Performance and Feature Importance in the Literature

We have taken a novel approach in developing 30-day and 1-year all-cause readmission prediction models using simple ML LASSO logistic regression method, and two advanced ML methods (i.e., XGBoost and ANN). Our 30-day and 1-year readmission rates of 14.1% and 42.2%, respectively were similar to the estimated meta-analysis pooled 30-day and 1-year all-cause post-stroke readmission rates.⁶ Our main objective was to compare the predictive performance of these methods when applied to MiSP registry data, and then identify the most important predictors from the best model. The predictive performance (AUC) of 30-day and 1-year readmission produced by LASSO logistic regression, XGBoost and ANN using MiSP registry data was modest and consistent with two US-based studies that developed post-stroke readmission prediction models using similar ML methods.^{5, 32} The two US-based studies in addition to 3 other international studies relied on a single data source (i.e., electronic medical records), mostly reported only 30-day readmission outcomes, and utilized the traditional internal

validation approach of randomly splitting the data into training and testing datasets which is prone to overfitting.^{5, 32, 40-42} Our study approach attempted to address these limitations. Findings from these previous studies reported that advanced ML methods i.e., XGBoost and ANN were more successful in predicting 30-day post-stroke readmission over other traditional regression or simple ML methods including Logistic regression, COX regression, random forest, support vector machine, k-nearest neighbor, and naïve bayes classifier.^{5, 32, 40-42} None of the previous studies utilized LASSO technique in logistic regression- a simple ML method.^{5, 32, 40-42} Our central hypothesis was not fulfilled because when we compared XGBoost and ANN developed models with the simpler LASSO logistic regression model, both failed to improve predictive accuracy compared to LASSO logistic regression. We therefore chose to provide detailed results of the LASSO logistic regression model performance including important selected predictors not only because it had equivalent predictive accuracy but also due to its relatively simple learning architecture. These findings suggest that either prediction of all cause readmission is limited and very hard to improve or better predictors of post stroke readmission rather than better statistical techniques are needed to improve model performance. These predictors include patient-level socioeconomic factors that are independent of care received at hospital and are associated with readmission including residency zip code, median income, family support, access to home care and transportation, compliance with prescription fillings, and health literacy, that are not usually reported on by either registry or claims databases.^{7, 67}

We used our internal cross validation hospital-based split technique in the LASSO logistic regression model to identify the most important predictors by ranking the order of selected predictors across the 31 hospital-specific models. For 30-day readmission prediction, the 31 hospital-specific models ranked discharge disposition, admission duration, ambulatory status

at discharge, history of chronic renal insufficiency, and history of heart failure as the top 5 predictors. For 1-year readmission prediction, the 31 hospital-specific models selected discharge disposition, admission duration, history of chronic renal insufficiency, history of heart failure, and persistent or paroxysmal atrial fibrillation upon discharge as their top 5 predictors. Results from the 5 prior ML post-stroke readmission prediction models reported that the most important predictive variables were lab results upon admission (e.g., glucose levels, and homocysteine), inpatient procedures (e.g., nasogastric tube insertion, craniectomy, and urinary catheter insertion), stroke clinical features (NIHSS, and stroke etiology) and to a lesser extent patient's clinical history (e.g., hemodialysis, and malnutrition).^{5, 32, 40-42} None of these published papers used a similar approach to ours to find the most important predictors, they mainly used internally coded algorithms embedded into the ML model packages. In addition, the most important predictors they reported are mainly related to inpatient clinical features in contrast to our findings that mainly reported on patient's past medical history as the most important predictors.

Our secondary objective was to examine the impact of using different combinations of linked data sources (i.e., registry, hospital survey, and administrative data) on the predictive performance of the ML methods. All of the ML methods reported the highest 30-day and 1-year pooled predictive accuracy (AUC) when MiSP registry and MVC claims data were added together ($P < .05$). Although, the effect of adding MVC data to MiSP was not statistically significant when using the XGBoost method to predict 30-day readmission. In the instance of 1-year readmission, the statistically significant additive effect of MVC data to MiSP had a higher impact (higher change in AUC) when compared to 30-day and it was observed over all the ML methods. There was no added effect on the pooled model accuracy (AUC) when hospital characteristics from the American Hospital Association were added to either MiSP or MVC or

the combination of both. Across all ML methods MVC data alone was able to produce similar pooled predictive accuracy for 1 year readmission when compared to MiSP data, however this was not the case in prediction of 30-day readmission where MiSP data almost always outperformed MVC data but with modest improvements to AUC that ranged from 1.8% to 3.7%. We hypothesized that the combination of all data sources would produce the highest predictive performance model, but this was not the case. Our finding is similar to a prior ML post-stroke readmission prediction paper that explored the additive effect of predictors on predictive performance of XGBoost ML method using varying number of predictors extracted from eMR data (i.e., 35, 200, and 400 predictors), where a model based on 35 predictors (AUC, 0.62, 95% CI, 0.61–0.63) outperformed models that used 200 (AUC, 0.61; 95% CI, 0.60–0.62) and even 400 predictors (AUC, 0.60; 95% CI, 0.59–0.61).³²

4.5.2 Practicality of the Developed Prediction Model

We followed a prognostic research framework that was published through a series of five papers entitled PROGNosis RESEARCH Strategy (PROGRESS). This series presents standards that should be followed when conducting prognostic research including prediction research that aims to be implemented in medical settings.⁶⁸ Following these standards, we chose to build our prediction model using a set of predictors that are highly associated with readmission and that are readily available through either registry or electronic medical records data. In addition, compared to XGBoost and ANN, our LASSO logistic regression ML technique (our model of choice) is well understood by the research community, and it produces models that have relatively smaller number of predictors. Furthermore, we followed a cross-validation strategy that promises to enhance the external validity and generalizability of the developed model and to accommodate the addition of newly participating hospitals in future model trainings. All of the mentioned steps try to guarantee the interpretability and applicability of the developed model by

healthcare systems or providers to identify patients at high risk of readmission before discharge. Our model ended up including a large number of predictors (more than 50), but the model could potentially be coded and automated to calculate the risk of readmission through the eMR system without human intervention. However, before our model could be adopted in medical practice, external validation must be performed either using MiSP data collected after our study period or similar registry data collected by other hospitals or states.⁶⁸ The final model and its associated estimates are available upon request.

External validation will pave the way to use our model as a tool to identify patients who are eligible for interventions geared toward reducing readmission including case management, post-discharge support, rehabilitation, active follow-up, telemonitoring, discharge planning, coordinated transitional care, home-based care, medication reconciliation, and patient education.^{12, 13, 69}

4.5.3 Strengths and Limitations

One of the major strengths of our study was the utilization of 5-years population-based stroke data from 31 hospitals in Michigan. This was possible by linkage to MVC claims data that provided the outcomes data for our stroke registry population. Our 30-day and 1-year readmission rates were similar to rates published in the literature. Our linked population included patients that are insured by Medicare and BCBSM- the largest health insurer in the state. The linkage enabled the analysis on a set of predictors from the MVC claims data and the American Hospital Association database that were not covered by the MiSP registry. In addition, our ML model development technique of hospital-based splitting into training and testing datasets and the rigorous internal cross validation technique to choose the hyperparameters increased the generalizability and accuracy of the developed models. Furthermore, our techniques and data

sources allowed to easily reproduce such models in the case when the registry expands to include more Michigan hospitals (in fact it now covers 52 hospitals) or external validation is attempted. To the best of our knowledge, this was the first attempt to utilize data from a registry to predict 30-days and 1-year post-stroke readmission through applying ML or hospital-specific model development techniques. The 5 published ML post-stroke readmission prediction models (2 are US-based) relied only on eMR data, utilized traditional internal cross validation methods, and mostly reported only 30-day readmission outcomes.

Our study had several limitations. Prediction models are limited by the number and quality of the predictors they use; the registry data suffered from high levels of non-random missingness in a number of important clinical predictors (e.g., in-patient procedures like intubation and foleys catheter insertion) which limited the number of predictors that were included in the analysis to 64 predictors. In addition, since registries are purpose built, many aspects of patient's past medical history were limited to stroke related comorbidities, specifics to post-acute care utilization were not captured, and important patient socioeconomic characteristics were absent. Due to limitations in data availability from MiSP registry we only included data from 31 hospitals, but we feel that our sample adequately represent the 49 stroke accredited hospitals (i.e., PSC, TSR, and CSC) in Michigan. Limitations in MVC claims data insurance coverage, resulted in excluding many stroke discharges recorded by MiSP registry (e.g., Medicaid, private insurance plans other than BCBS, MA plan other than BCBS, and uninsured population) caused a relatively low linkage rates (51.6% of MiSP population), this limits the generalizability of our results to patients insured by Medicare and BCBSM. XGBoost and ANN model development techniques have a large number of hyperparameters that needs to be preset which may introduce overfitting, we tried to control this by utilizing an internal cross validation

technique to choose one major parameter in both techniques. Our developed models were not externally validated due to lack of external sources to similar data. Nevertheless, given the robustness of our model development technique and the extensive internal validation methods, the results can be considered valid.

4.5.4 Future Directions

In this study, we included data from 31 hospitals with a patient population limited by insurance coverage to Medicare and BCBSM. Therefore, future research should expand the insurance providers including Medicaid and other private insurers to generate a more generalizable model. Additionally, to improve prediction accuracy, future studies should integrate data sources or features that have wider coverage in areas of past medical history, inpatient clinical and post-acute care like lab results, rehabilitation, social support, compliance with outpatient follow up visits, prescription fillings, and income. These characteristics are hard to come by but could be added through data linkage to other data sources (e.g., census bureau or electronic medical records) in the event direct identifiers of patients become available. To improve the representativeness of the hospitals, the developed ML model could get retrained and calibrated through utilizing data from the newly participating hospitals in MiSP now totaling 52 hospitals. Lastly, external validation of the model should take place within Michigan through collecting complete records of stroke patients' data directly from a representative sample of hospitals and follow them up for 1-year or test the models on data collected from the same hospitals but after the study period (2020). Furthermore, external validation should be explored using similar linked data from states participating in Paul Coverdale National Acute Stroke Program to pave the way to adopt them into clinical studies including trials. These points should overcome some of the limitations, reduce bias, and produce an externally valid generalizable model.

4.6 Conclusions

Based on the comparison results of this study, we conclude that simple predictive modelling methods like LASSO logistic regression produces similar prediction accuracy values when compared to more advanced ML methods including XGBoost and ANN. Using a simpler ML method to predict short- and long-term hospital readmissions can help in identifying patients at risk of readmission before they are discharged to improve management of their post-acute care. In addition, it can help health policy makers in developing valid predicted estimates of readmission rates on the hospital and population levels. Our analysis demonstrated that claims data can also be used to predict readmission rates with similar predictive accuracy as registry-based models. The patient's clinical history prior to stroke – including chronic renal insufficiency, heart failure, atrial fibrillation, coronary artery disease, diabetes mellitus, and stroke appears to be of higher importance when predicting long term readmission compared to stroke clinical features such as admission NIHSS, ambulatory status upon discharge, and stroke etiology. These findings indicate that adequate post-acute care can likely contribute to lowering the probability of readmission. Although our modelling methods and registry data demonstrated that readmission after stroke can be predicted, the routine clinical applicability of these models by hospitals prior to discharge to deliver patient-specific post-acute care interventions needs further study. Future studies should utilize data that expands coverage of insurance providers, participating hospitals, and range of clinical and post-acute care variables that are not collected by registries or claims data. Such data could be a source for assessment, development, and implementation of healthcare policies that will improve stroke outcomes.

BIBLIOGRAPHY

1. Tsao, C. W.; Aday, A. W.; Almarzooq, Z. I.; Anderson, C. A. M.; Arora, P.; Avery, C. L.; Baker-Smith, C. M.; Beaton, A. Z.; Boehme, A. K.; Buxton, A. E.; Commodore-Mensah, Y.; Elkind, M. S. V.; Evenson, K. R.; Eze-Nliam, C.; Fugar, S.; Generoso, G.; Heard, D. G.; Hiremath, S.; Ho, J. E.; Kalani, R.; Kazi, D. S.; Ko, D.; Levine, D. A.; Liu, J.; Ma, J.; Magnani, J. W.; Michos, E. D.; Mussolino, M. E.; Navaneethan, S. D.; Parikh, N. I.; Poudel, R.; Rezk-Hanna, M.; Roth, G. A.; Shah, N. S.; St-Onge, M. P.; Thacker, E. L.; Virani, S. S.; Voeks, J. H.; Wang, N. Y.; Wong, N. D.; Wong, S. S.; Yaffe, K.; Martin, S. S.; American Heart Association Council on, E.; Prevention Statistics, C.; Stroke Statistics, S., Heart Disease and Stroke Statistics-2023 Update: A Report From the American Heart Association. *Circulation* **2023**, *147* (8), e93-e621. <https://doi.org/10.1161/CIR.0000000000001123>.
2. Weiss, A. J.; Jiang, H. J., Overview of Clinical Conditions With Frequent and Costly Hospital Readmissions by Payer, 2018. In *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*, Rockville (MD), 2021.
3. Bambhroliya, A. B.; Donnelly, J. P.; Thomas, E. J.; Tyson, J. E.; Miller, C. C.; McCullough, L. D.; Savitz, S. I.; Vahidy, F. S., Estimates and Temporal Trend for US Nationwide 30-Day Hospital Readmission Among Patients With Ischemic and Hemorrhagic Stroke. *JAMA Netw Open* **2018**, *1* (4), e181190. <https://doi.org/10.1001/jamanetworkopen.2018.1190>.
4. Bayliss, W. S.; Bushnell, C. D.; Halladay, J. R.; Duncan, P. W.; Freburger, J. K.; Kucharska-Newton, A. M.; Trogdon, J. G., The Cost of Implementing and Sustaining the Comprehensive Post-Acute Stroke Services Model. *Med Care* **2021**, *59* (2), 163-168. <https://doi.org/10.1097/MLR.0000000000001462>.
5. Darabi, N.; Hosseinichimeh, N.; Noto, A.; Zand, R.; Abedi, V., Machine Learning-Enabled 30-Day Readmission Model for Stroke Patients. *Front Neurol* **2021**, *12*, 638267. <https://doi.org/10.3389/fneur.2021.638267>.
6. Zhong, W.; Geng, N.; Wang, P.; Li, Z.; Cao, L., Prevalence, causes and risk factors of hospital readmissions after acute stroke and transient ischemic attack: a systematic review and meta-analysis. *Neurol Sci* **2016**, *37* (8), 1195-202. <https://doi.org/10.1007/s10072-016-2570-5>.
7. Zhou, L. W.; Lansberg, M. G.; de Havenon, A., Rates and reasons for hospital readmission after acute ischemic stroke in a US population-based cohort. *PLoS One* **2023**, *18* (8), e0289640. <https://doi.org/10.1371/journal.pone.0289640>.
8. Vahidy, F. S.; Donnelly, J. P.; McCullough, L. D.; Tyson, J. E.; Miller, C. C.; Boehme, A. K.; Savitz, S. I.; Albright, K. C., Nationwide Estimates of 30-Day Readmission in Patients With Ischemic Stroke. *Stroke* **2017**, *48* (5), 1386-1388. <https://doi.org/10.1161/STROKEAHA.116.016085>.

9. Lichtman, J. H.; Leifheit-Limson, E. C.; Jones, S. B.; Wang, Y.; Goldstein, L. B., Preventable readmissions within 30 days of ischemic stroke among Medicare beneficiaries. *Stroke* **2013**, *44* (12), 3429-35. <https://doi.org/10.1161/STROKEAHA.113.003165>.
10. Beauvais, B.; Whitaker, Z.; Kim, F.; Anderson, B., Is the Hospital Value-Based Purchasing Program Associated with Reduced Hospital Readmissions? *J Multidiscip Healthc* **2022**, *15*, 1089-1099. <https://doi.org/10.2147/JMDH.S358733>.
11. Fischer, C.; Lingsma, H. F.; Marang-van de Mheen, P. J.; Kringos, D. S.; Klazinga, N. S.; Steyerberg, E. W., Is the readmission rate a valid quality indicator? A review of the evidence. *PLoS One* **2014**, *9* (11), e112282. <https://doi.org/10.1371/journal.pone.0112282>.
12. Leppin, A. L.; Gionfriddo, M. R.; Kessler, M.; Brito, J. P.; Mair, F. S.; Gallacher, K.; Wang, Z.; Erwin, P. J.; Sylvester, T.; Boehmer, K.; Ting, H. H.; Murad, M. H.; Shippee, N. D.; Montori, V. M., Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. *JAMA Intern Med* **2014**, *174* (7), 1095-107. <https://doi.org/10.1001/jamainternmed.2014.1608>.
13. Hansen, L. O.; Young, R. S.; Hinami, K.; Leung, A.; Williams, M. V., Interventions to reduce 30-day rehospitalization: a systematic review. *Ann Intern Med* **2011**, *155* (8), 520-8. <https://doi.org/10.7326/0003-4819-155-8-201110180-00008>.
14. Finkelstein, A.; Taubman, S.; Doyle, J., Health Care Hotspotting - A Randomized, Controlled Trial. Reply. *N Engl J Med* **2020**, *382* (22), 2173-2174. <https://doi.org/10.1056/NEJMc2001920>.
15. Kansagara, D.; Chiovaro, J. C.; Kagen, D.; Jencks, S.; Rhyne, K.; O'Neil, M.; Kondo, K.; Relevo, R.; Motu'apuaka, M.; Freeman, M.; Englander, H., So many options, where do we start? An overview of the care transitions literature. *J Hosp Med* **2016**, *11* (3), 221-30. <https://doi.org/10.1002/jhm.2502>.
16. El Husseini, N.; Fonarow, G. C.; Smith, E. E.; Ju, C.; Sheng, S.; Schwamm, L. H.; Hernandez, A. F.; Schulte, P. J.; Xian, Y.; Goldstein, L. B., Association of Kidney Function With 30-Day and 1-Year Poststroke Mortality and Hospital Readmission. *Stroke* **2018**, *49* (12), 2896-2903. <https://doi.org/10.1161/STROKEAHA.118.022011>.
17. Rao, A.; Barrow, E.; Vuik, S.; Darzi, A.; Aylin, P., Systematic Review of Hospital Readmissions in Stroke Patients. *Stroke Res Treat* **2016**, *2016*, 9325368. <https://doi.org/10.1155/2016/9325368>.
18. Jun-O'Connell, A. H.; Grigoriuc, E.; Silver, B.; Kobayashi, K. J.; Osgood, M.; Moonis, M.; Henninger, N., Association between the LACE+ index and unplanned 30-day hospital readmissions in hospitalized patients with stroke. *Front Neurol* **2022**, *13*, 963733. <https://doi.org/10.3389/fneur.2022.963733>.

19. Loebel, E. M.; Rojas, M.; Wheelwright, D.; Mensching, C.; Stein, L. K., High Risk Features Contributing to 30-Day Readmission After Acute Ischemic Stroke: A Single Center Retrospective Case-Control Study. *Neurohospitalist* **2022**, *12* (1), 24-30. <https://doi.org/10.1177/19418744211027746>.
20. Burke, J. F.; Skolarus, L. E.; Adelman, E. E.; Reeves, M. J.; Brown, D. L., Influence of hospital-level practices on readmission after ischemic stroke. *Neurology* **2014**, *82* (24), 2196-204. <https://doi.org/10.1212/WNL.0000000000000514>.
21. Nahab, F.; Takesaka, J.; Mailyan, E.; Judd, L.; Culler, S.; Webb, A.; Frankel, M.; Choi, D.; Helmers, S., Avoidable 30-day readmissions among patients with stroke and other cerebrovascular disease. *Neurohospitalist* **2012**, *2* (1), 7-11. <https://doi.org/10.1177/1941874411427733>.
22. Leonhardt-Caprio, A. M.; Sellers, C. R.; Palermo, E.; Caprio, T. V.; Holloway, R. G., A Multi-Component Transition of Care Improvement Project to Reduce Hospital Readmissions Following Ischemic Stroke. *Neurohospitalist* **2022**, *12* (2), 205-212. <https://doi.org/10.1177/19418744211036632>.
23. Jun-O'Connell, A. H.; Grigoriuc, E.; Gulati, A.; Silver, B.; Kobayashi, K. J.; Moonis, M.; Henninger, N., Stroke nurse navigator utilization reduces unplanned 30-day readmission in stroke patients treated with thrombolysis. *Front Neurol* **2023**, *14*, 1205487. <https://doi.org/10.3389/fneur.2023.1205487>.
24. Condon, C.; Lycan, S.; Duncan, P.; Bushnell, C., Reducing Readmissions After Stroke With a Structured Nurse Practitioner/Registered Nurse Transitional Stroke Program. *Stroke* **2016**, *47* (6), 1599-604. <https://doi.org/10.1161/STROKEAHA.115.012524>.
25. Terman, S. W.; Reeves, M. J.; Skolarus, L. E.; Burke, J. F., Association Between Early Outpatient Visits and Readmissions After Ischemic Stroke. *Circ Cardiovasc Qual Outcomes* **2018**, *11* (4), e004024. <https://doi.org/10.1161/CIRCOUTCOMES.117.004024>.
26. Leppert, M. H.; Sillau, S.; Lindrooth, R. C.; Poisson, S. N.; Campbell, J. D.; Simpson, J. R., Relationship between early follow-up and readmission within 30 and 90 days after ischemic stroke. *Neurology* **2020**, *94* (12), e1249-e1258. <https://doi.org/10.1212/WNL.00000000000009135>.
27. Marafino, B. J.; Escobar, G. J.; Baiocchi, M. T.; Liu, V. X.; Plimier, C. C.; Schuler, A., Evaluation of an intervention targeted with predictive analytics to prevent readmissions in an integrated health system: observational study. *BMJ* **2021**, *374*, n1747. <https://doi.org/10.1136/bmj.n1747>.
28. Escobar, G. J.; Liu, V. X.; Schuler, A.; Lawson, B.; Greene, J. D.; Kipnis, P., Automated Identification of Adults at Risk for In-Hospital Clinical Deterioration. *N Engl J Med* **2020**, *383* (20), 1951-1960. <https://doi.org/10.1056/NEJMsa2001090>.

29. Lichtman, J. H.; Leifheit-Limson, E. C.; Jones, S. B.; Watanabe, E.; Bernheim, S. M.; Phipps, M. S.; Bhat, K. R.; Savage, S. V.; Goldstein, L. B., Predictors of hospital readmission after stroke: a systematic review. *Stroke* **2010**, *41* (11), 2525-33. <https://doi.org/10.1161/STROKEAHA.110.599159>.
30. Fonarow, G. C.; Smith, E. E.; Reeves, M. J.; Pan, W.; Olson, D.; Hernandez, A. F.; Peterson, E. D.; Schwamm, L. H.; Get With The Guidelines Steering, C.; Hospitals, Hospital-level variation in mortality and rehospitalization for medicare beneficiaries with acute ischemic stroke. *Stroke* **2011**, *42* (1), 159-66. <https://doi.org/10.1161/STROKEAHA.110.601831>.
31. Fehnel, C. R.; Lee, Y.; Wendell, L. C.; Thompson, B. B.; Potter, N. S.; Mor, V., Post-Acute Care Data for Predicting Readmission After Ischemic Stroke: A Nationwide Cohort Analysis Using the Minimum Data Set. *J Am Heart Assoc* **2015**, *4* (9), e002145. <https://doi.org/10.1161/JAHA.115.002145>.
32. Lineback, C. M.; Garg, R.; Oh, E.; Naidech, A. M.; Holl, J. L.; Prabhakaran, S., Prediction of 30-Day Readmission After Stroke Using Machine Learning and Natural Language Processing. *Front Neurol* **2021**, *12*, 649521. <https://doi.org/10.3389/fneur.2021.649521>.
33. Kansagara, D.; Englander, H.; Salanitro, A.; Kagen, D.; Theobald, C.; Freeman, M.; Kripalani, S., Risk prediction models for hospital readmission: a systematic review. *JAMA* **2011**, *306* (15), 1688-98. <https://doi.org/10.1001/jama.2011.1515>.
34. Artetxe, A.; Beristain, A.; Grana, M., Predictive models for hospital readmission risk: A systematic review of methods. *Comput Methods Programs Biomed* **2018**, *164*, 49-64. <https://doi.org/10.1016/j.cmpb.2018.06.006>.
35. Ouwerkerk, W.; Voors, A. A.; Zwinderman, A. H., Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure. *JACC Heart Fail* **2014**, *2* (5), 429-36. <https://doi.org/10.1016/j.jchf.2014.04.006>.
36. Amritphale, A.; Chatterjee, R.; Chatterjee, S.; Amritphale, N.; Rahnavard, A.; Awan, G. M.; Omar, B.; Fonarow, G. C., Predictors of 30-Day Unplanned Readmission After Carotid Artery Stenting Using Artificial Intelligence. *Adv Ther* **2021**, *38* (6), 2954-2972. <https://doi.org/10.1007/s12325-021-01709-7>.
37. Wang, Z.; Chen, X.; Tan, X.; Yang, L.; Kannapur, K.; Vincent, J. L.; Kessler, G. N.; Ru, B.; Yang, M., Using Deep Learning to Identify High-Risk Patients with Heart Failure with Reduced Ejection Fraction. *J Health Econ Outcomes Res* **2021**, *8* (2), 6-13. <https://doi.org/10.36469/jheor.2021.25753>.
38. Sharma, V.; Kulkarni, V.; McAlister, F.; Eurich, D.; Keshwani, S.; Simpson, S. H.; Voaklander, D.; Samanani, S., Predicting 30-Day Readmissions in Patients With Heart Failure

Using Administrative Data: A Machine Learning Approach. *J Card Fail* **2022**, 28 (5), 710-722. <https://doi.org/10.1016/j.cardfail.2021.12.004>.

39. Sarajlic, P.; Simonsson, M.; Jernberg, T.; Back, M.; Hofmann, R., Incidence, associated outcomes, and predictors of upper gastrointestinal bleeding following acute myocardial infarction: a SWEDEHEART-based nationwide cohort study. *Eur Heart J Cardiovasc Pharmacother* **2022**, 8 (5), 483-491. <https://doi.org/10.1093/ehjcvp/pvab059>.

40. Xu, Y.; Yang, X.; Huang, H.; Peng, C.; Ge, Y.; Wu, H.; Wang, J.; Xiong, G.; Yi, Y., Extreme Gradient Boosting Model Has a Better Performance in Predicting the Risk of 90-Day Readmissions in Patients with Ischaemic Stroke. *J Stroke Cerebrovasc Dis* **2019**, 28 (12), 104441. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.104441>.

41. Lv, J.; Zhang, M.; Fu, Y.; Chen, M.; Chen, B.; Xu, Z.; Yan, X.; Hu, S.; Zhao, N., An interpretable machine learning approach for predicting 30-day readmission after stroke. *Int J Med Inform* **2023**, 174, 105050. <https://doi.org/10.1016/j.ijmedinf.2023.105050>.

42. Chen, Y. C.; Chung, J. H.; Yeh, Y. J.; Lou, S. J.; Lin, H. F.; Lin, C. H.; Hsien, H. H.; Hung, K. W.; Yeh, S. J.; Shi, H. Y., Predicting 30-Day Readmission for Stroke Using Machine Learning Algorithms: A Prospective Cohort Study. *Front Neurol* **2022**, 13, 875491. <https://doi.org/10.3389/fneur.2022.875491>.

43. In *Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition, Addendum 2*, Gliklich, R. E.; Leavy, M. B.; Dreyer, N. A., Eds. Rockville (MD), 2019.

44. In *Registries for Evaluating Patient Outcomes: A User's Guide*, 3rd ed.; Gliklich, R. E.; Dreyer, N. A.; Leavy, M. B., Eds. Rockville (MD), 2014.

45. Association, A. H. AHA Annual Survey Database. <https://www.ahadata.com/>.

46. Cadarette, S. M.; Wong, L., An Introduction to Health Care Administrative Data. *Can J Hosp Pharm* **2015**, 68 (3), 232-7. <https://doi.org/10.4212/cjhp.v68i3.1457>.

47. Bradley, C. J.; Penberthy, L.; Devers, K. J.; Holden, D. J., Health services research and data linkages: issues, methods, and directions for the future. *Health Serv Res* **2010**, 45 (5 Pt 2), 1468-88. <https://doi.org/10.1111/j.1475-6773.2010.01142.x>.

48. Dusetzina, S. B.; Tyree, S.; Meyer, A. M.; Meyer, A.; Green, L.; Carpenter, W. R., In *Linking Data for Health Services Research: A Framework and Instructional Guide*, Rockville (MD), 2014.

49. Michigan Department of Health and Human Services (MDHHS), Michigan Stroke Program (MiSP). <https://www.michigan.gov/mdhhs/keep-mi-healthy/communicablediseases/epidemiology/chronicepi/stroke> (accessed 2023).
50. Center of Disease Control and Prevention, Paul Coverdell National Acute Stroke Program. https://www.cdc.gov/dhdsp/programs/stroke_registry.htm (accessed 2023).
51. American Heart Association, Get With The Guidelines® - Stroke Case Record Form. <https://www.heart.org/-/media/Files/Professional/Quality-Improvement/Get-With-the-Guidelines/Get-With-The-Guidelines-Stroke--Diabetes-CRFJuly21.pdf>.
52. Michigan Value Collaborative, MVC Data Resources. <https://michiganvalue.org/resources-2/> (accessed 2023).
53. American Hospital Association, Annual Survey Database. <https://www.ahadata.com/> (accessed 2023).
54. Chan, A. K.; Shahrestani, S.; Ballatori, A. M.; Orrico, K. O.; Manley, G. T.; Tarapore, P. E.; Huang, M.; Dhall, S. S.; Chou, D.; Mummaneni, P. V.; DiGiorgio, A. M., Is the Centers for Medicare and Medicaid Services Hierarchical Condition Category Risk Adjustment Model Satisfactory for Quantifying Risk After Spine Surgery? *Neurosurgery* **2022**, *91* (1), 123-131. <https://doi.org/10.1227/neu.0000000000001980>.
55. Centers for Medicare and Medicaid Services and Agency for Healthcare Research and Quality, 2016 Measure Information About The 30-Day All-Cause Hospital Readmission Measure, Calculated For The 2018 Value-Based Payment Modifier Program. <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2016-ACR-MIF.pdf>.
56. Tibshirani, R., The lasso method for variable selection in the Cox model. *Stat Med* **1997**, *16* (4), 385-95. [https://doi.org/10.1002/\(sici\)1097-0258\(19970228\)16:4<385::aid-sim380>3.0.co;2-3](https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3).
57. Kim, S. M.; Kim, Y.; Jeong, K.; Jeong, H.; Kim, J., Logistic LASSO regression for the diagnosis of breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography. *Ultrasonography* **2018**, *37* (1), 36-42. <https://doi.org/10.14366/usg.16045>.
58. Kagiya, N.; Shrestha, S.; Farjo, P. D.; Sengupta, P. P., Artificial Intelligence: Practical Primer for Clinical Research in Cardiovascular Disease. *J Am Heart Assoc* **2019**, *8* (17), e012788. <https://doi.org/10.1161/JAHA.119.012788>.

59. Wang, G.; Hao, J. X.; Ma, J. A.; Jiang, H. B., A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications* **2011**, *38* (1), 223-230. <https://doi.org/10.1016/j.eswa.2010.06.048>.
60. Wang, Y.; Miao, X.; Xiao, G.; Huang, C.; Sun, J.; Wang, Y.; Li, P.; You, X., Clinical Prediction of Heart Failure in Hemodialysis Patients: Based on the Extreme Gradient Boosting Method. *Front Genet* **2022**, *13*, 889378. <https://doi.org/10.3389/fgene.2022.889378>.
61. Emmert-Streib, F.; Yang, Z.; Feng, H.; Tripathi, S.; Dehmer, M., An Introductory Review of Deep Learning for Prediction Models With Big Data. *Front Artif Intell* **2020**, *3*. <https://doi.org/ARTN 410.3389/frai.2020.00004>.
62. Kagiya, N.; Shrestha, S.; Farjo, P. D.; Sengupta, P. P., Artificial Intelligence: Practical Primer for Clinical Research in Cardiovascular Disease. *Journal of the American Heart Association* **2019**, *8* (17). <https://doi.org/ARTN e01278810.1161/JAHA.119.012788>.
63. LeCun, Y.; Bengio, Y.; Hinton, G., Deep learning. *Nature* **2015**, *521* (7553), 436-444. <https://doi.org/10.1038/nature14539>.
64. Cui, C. W., Dianhui, High dimensional data regression using Lasso model and neural networks with random weights. *Information Sciences* **2016**, *372*. <https://doi.org/10.1016/j.ins.2016.08.060>.
65. Steyerberg, E. W.; Harrell, F. E., Jr., Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* **2016**, *69*, 245-7. <https://doi.org/10.1016/j.jclinepi.2015.04.005>.
66. Marshall, A.; Altman, D. G.; Holder, R. L.; Royston, P., Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol* **2009**, *9*, 57. <https://doi.org/10.1186/1471-2288-9-57>.
67. Murray, F.; Allen, M.; Clark, C. M.; Daly, C. J.; Jacobs, D. M., Socio-demographic and -economic factors associated with 30-day readmission for conditions targeted by the hospital readmissions reduction program: a population-based study. *BMC Public Health* **2021**, *21* (1), 1922. <https://doi.org/10.1186/s12889-021-11987-z>.
68. Hemingway, H.; Croft, P.; Perel, P.; Hayden, J. A.; Abrams, K.; Timmis, A.; Briggs, A.; Udumyan, R.; Moons, K. G.; Steyerberg, E. W.; Roberts, I.; Schroter, S.; Altman, D. G.; Riley, R. D.; Group, P., Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* **2013**, *346*, e5595. <https://doi.org/https://doi.org/10.1136/bmj.e5595>.

69. Kripalani, S.; Theobald, C. N.; Anctil, B.; Vasilevskis, E. E., Reducing hospital readmission rates: current strategies and future directions. *Annu Rev Med* **2014**, *65*, 471-85. <https://doi.org/10.1146/annurev-med-022613-090415>.

APPENDIX

Table 4A.1: Univariate descriptive statistics of potential predictors of readmission from the Michigan Stroke Program (MiSP), American Hospital Association (AHA), and Michigan Value Collaborative (MVC) databases (n= 19,382 linked stroke discharges).

Predictors group	Predictor	Value	% total=19,382	30-days all-cause readmission					1-year all-cause readmission						
				%	OR	95% CI	χ^2 test LRT	χ^2 test p-value	%	OR	95% CI	χ^2 test LRT	χ^2 test p-value		
Demographics (from MiSP data)	Age category	1: <65	21.5	13.7	Ref	6.5	0.088	36.6	Ref	98.4	<0.001				
		2: 65-74	29.2	13.5	1.0			0.9-1.1	41.0			1.2	1.1-1.3		
		3: 75-84	28.8	15.1	1.1			1.0-1.3	44.4			1.4	1.3-1.5		
		4: >=85	20.4	13.8	1.0			0.9-1.1	46.5			1.5	1.4-1.6		
	Race	1: White	79.7	13.7	Ref	12.3	0.006	41.0	Ref	61.5	<0.001				
		2: Black	14.6	15.9	1.2			1.1-1.3	48.9			1.4	1.3-1.5		
		3: Other	1.3	17.6	1.4			1.0-1.9	38.8			0.9	0.7-1.2		
		ND	4.3	14.2	1.0			0.9-1.3	42.1			1.0	0.9-1.2		
	Latino Ethnicity	2: No/UTD	96.3	14.1	Ref	1.0	0.313	42.4	Ref	9.7	0.001				
		1: Yes	3.7	12.8	0.9			0.7-1.1	36.5			0.8	0.7-0.9		
	sex	1: Male	47.8	14.3	Ref	0.7	0.417	41.9	Ref	0.5	0.5				
		2: Female	52.2	13.9	1.0			0.9-1.0	42.4			1.0	1.0-1.1		
	Insurance (From MVC claims data)	BCBSM PPO Comm	12.0	9.3	Ref	84.5	<0.001	27.4	Ref	406.9	<0.001				
		BCBSM PPO MA	15.7	14.0	1.6			1.3-1.9	41.3			1.9	1.7-2.1		
		BCN Comm	4.2	9.6	1.0			0.8-1.4	26.1			0.9	0.8-1.1		
BCN MA Other		5.2	12.2	1.4	1.1-1.7			39.1	1.7			1.5-2.0			
Medicare FFS		62.9	15.4	1.8	1.5-2.1			46.5	2.3			2.1-2.5			
Admission year	2016	19.7	12.9	Ref	7.2	0.126	43.4	Ref	14.5	0.005					
	2017	22.3	14.7	1.2			1-1.3	43.0			1.0	0.9-1.1			
	2018	22.3	13.7	1.1			0.9-1.2	43.0			1.0	0.9-1.1			
	2019	19.9	14.5	1.1			1.0-1.3	40.7			0.9	0.8-1			
	2020	15.8	14.6	1.1			1.0-1.3	40.0			0.9	0.8-1			
Administrative related (from MiSP data)	Documented stroke etiology	1: Large-artery atherosclerosis (e.g., carotid, or basilar artery stenosis)	9.2	15.5	Ref	62.5	<0.001	45.0	Ref	109.0	<0.001				
		2: Cardioembolism (e.g., atrial fibrillation/flutter, prosthetic heart valve, recent MI)	14.5	15.1	1.0			0.8-1.1	46.3			1.1	0.9-1.2		
		3: Small-vessel disease (e.g., Subcortical or brain stem lacunar infarction <1.5 cm)	13.6	10.6	0.6			0.5-0.8	35.4			0.7	0.6-0.8		
		4: Stroke of other determined etiology	1.5	14.0	0.9			0.6-1.3	42.3			0.9	0.7-1.1		
		5: Cryptogenic Stroke	21.3	13.8	0.9			0.7-1.0	40.2			0.8	0.7-0.9		
		6: Hemorrhagic Intracerebral	9.8	18.1	1.2			1.0-1.4	47.5			1.1	1.0-1.3		
		7: Hemorrhagic subarachnoid	2.9	15.9	1.0			0.8-1.3	37.7			0.7	0.6-0.9		
		ND	27.2	13.3	0.8			0.7-1.0	42.4			0.9	0.8-1.0		
		Only comfort measures	1 - Day 0 or 1	0.5	9.1			Ref	2.5			0.288	Ref	8.2	0.016

Table 4A.1 (cont'd)

	2 - Day 2 or after	0.5	15.7	1.9	0.8-4.4			1.0	0.6-1.8				
	ND/UTD	99.0	14.1	1.6	0.8-3.3			1.5	1.0-2.3				
Admission duration	1: 0-2	31.6	11.0		Ref			Ref					
	2: 3-6	48.6	13.5	1.3	1.1-1.4	166.0	<0.001	1.6	1.4-1.7	413.4	<0.001		
	3: >6	19.9	20.3	2.1	1.8-2.3			2.3	2.1-2.5				
	1: Home	50.1	10.2		Ref			34.5	Ref				
Discharge disposition	9: Skilled Nursing Facility (SNF)	24.9	12.8	1.3	1.2-1.4			44.6	1.5	1.4-1.6			
	10: Inpatient Rehabilitation Facility (IRF)	1.6	17.8	1.9	1.4-2.6			62.1	3.1	2.5-3.9			
	11: Long Term Care Hospital (LTCH)	0.6	26.2	3.1	2.0-4.8	931.8	<0.001	54.2	2.2	1.5-3.3	771.0	<0.001	
	13: Other	21.4	19.0	2.1	1.9-2.3			52.0	2.1	1.9-2.2			
	ND	1.5	83.7	45.1	32.8-62.1			90.7	18.4	12.4-27.4			
Onset to door time	1: >=0 and <4.5	38.1	14.2		Ref			41.6	Ref				
	2: >4.5 and <=12	17.1	13.5	0.9	0.8-1.1	9.2	0.27	41.5	1.0	0.9-1.1	11.8	0.008	
	3: >12	31.4	13.4	0.9	0.8-1.0			41.8	1.0	0.9-1.1			
	ND	13.4	15.7	1.1	1-1.3			45.2	1.2	1.1-1.3			
Stroke type	1: Ischemic	87.3	13.5		Ref			41.7	Ref				
	2: Hemorrhagic	12.7	17.7	1.4	1.2-1.5	828.9	<0.001	45.4	1.2	1.1-1.3	12.4	<0.001	
Ambulatory status on admission	1: Able to ambulate independently (no help from another person) w/ or w/o device	22.1	10.8		Ref			35.6	Ref				
	2: With assistance (from person)	21.7	14.7	1.4	1.2-1.6	110.6	<0.001	45.1	1.5	1.4-1.6	177.1	<0.001	
	3: Unable to ambulate	11.4	20.5	2.1	1.8-2.5			51.9	1.9	1.8-2.2			
	ND	44.9	13.7	1.3	1.2-1.5			41.5	1.3	1.2-1.4			
Ambulatory status prior to the current event	1: Able to ambulate independently (no help from another person) w/ or w/o device	77.5	13.1		Ref			40.8	Ref				
	2: With assistance (from person)	3.0	21.6	1.8	1.5-2.2	72.8	<0.001	55.9	1.8	1.6-2.2	127.2	<0.001	
	3: Unable to ambulate	2.2	23.9	2.1	1.7-2.6			62.3	2.4	2.0-2.9			
	ND	17.4	15.8	1.2	1.1-1.4			43.5	1.1	1.0-1.2			
Admission related (from MiSP data)	Prior Antihypertensive medication	2: No/ND	37.6	11.8		Ref		36.5	Ref				
		1: Yes	62.4	15.4	1.4	1.2-1.5	48.3	<0.001	45.5	1.5	1.4-1.5	151.7	<0.001
	Prior cholesterol reducer medication	2: No/ND	48.4	13.2		Ref		38.5	Ref				
		1: Yes	51.6	14.8	1.1	1.1-1.2	10.2	0.001	45.6	1.3	1.3-1.4	101.3	<0.001
	Prior anti-hyperglycemic medication	2: No/ND	79.1	13.4		Ref		40.4	Ref				
		1: Yes	20.9	16.5	1.3	1.2-1.4	23.8	<0.001	48.9	1.4	1.3-1.5	94.0	<0.001
	Prior antiplatelet or anticoagulant medication	2: No/ND	42.8	12.9		Ref		36.8	Ref				
		1: Yes	57.2	14.9	1.2	1.1-1.3	16.4	<0.001	46.1	1.5	1.4-1.6	171.0	<0.001
	Prior antidepressant medication	2: No/ND	85.2	13.7		Ref		40.9	Ref				
		1: Yes	14.8	16.2	1.2	1.1-1.4	11.9	<0.001	49.3	1.4	1.3-1.5	69.3	<0.001
	Admission NIHSS	0: 0	15.4	11.6		Ref			36.1	Ref			
		1: 1-4	41.0	11.6	1.0	0.9-1.1			38.3	1.1	1.0-1.2		
2: 5-15		25.6	15.4	1.4	1.2-1.6	159.9	<0.001	46.0	1.5	1.4-1.6	226.6	<0.001	
3: 16-20		4.3	22.5	2.2	1.8-2.7			53.5	2.0	1.7-2.4			

Table 4A.1 (cont'd)

	4: >20	3.8	21.4	2.1	1.7-2.6			50.3	1.8	1.5-2.1		
	ND	10.0	18.0	1.7	1.4-2.0			49.4	1.7	1.5-1.9		
Arrival Mode	1: EMS from home/scene	44.0	16.9		Ref	112.0	<0.001	47.0		Ref	159.5	<0.001
	2: Private transport/taxi/other from home/scene"	34.8	11.0	0.6	0.6-0.7			37.1	0.7	0.6-0.7		
	3: Transfer from other hospital	19.9	13.3	0.8	0.7-0.8			40.4	0.8	0.7-0.8		
	ND	1.2	11.3	0.6	0.4-0.9			38.7	0.7	0.5-0.9		
Where did the patient first receive care?	1: Emergency Department/Urgent Care	64.1	14.7		Ref	13.3	0.004	43.0		Ref	16.1	0.001
	2: Direct Admit, not through ED	7.4	12.2	0.8	0.7-0.9			40.5	0.9	0.8-1.0		
	3: Imaging suite	0.9	15.4	1.1	0.7-1.6			49.1	1.3	0.9-1.7		
	ND	27.6	13.1	0.9	0.8-1.0			40.3	0.9	0.8-1.0		
Patient location when stroke symptoms discovered	1: Not in a healthcare setting	93.5	13.5		Ref	68.1	<0.001	41.4		Ref	67.1	<0.001
	2: Another acute care facility	0.7	18.7	1.5	1.0-2.3			50.4	1.4	1.0-2.0		
	3: Chronic health care facility	4.3	24.3	2.1	1.7-2.4			55.2	1.7	1.5-2.0		
	5: Outpatient healthcare setting	1.0	15.4	1.2	0.8-1.7			47.9	1.3	1.0-1.7		
	ND	0.6	13.3	1.0	0.6-1.7			39.8	0.9	0.6-1.4		
ED patient	2: No	11.2	12.3		Ref	8.8	0.013	40.5		Ref	6.9	0.031
	1: Yes	80.3	14.1	1.2	1-1.3			42.1	1.1	1-1.2		
	ND	8.6	15.5	1.3	1.1-1.6			44.8	1.2	1-1.4		
Atrial fibrillation/flutter	2: No	74.1	13.3		Ref	43.1	<0.001	39.9		Ref	195.2	<0.001
	1: Yes	21.7	17.1	1.3	1.2-1.5			51.3	1.6	1.5-1.7		
	ND	4.2	11.1	0.8	0.6-1.0			34.1	0.8	0.7-0.9		
Dyslipidemia	2: No	41.5	13.8		Ref	8.4	0.015	40.7		Ref	42.4	<0.001
	1: Yes	54.3	14.5	1.1	1.0-1.2			43.9	1.1	1.1-1.2		
	ND	4.2	11.1	0.8	0.6-1.0			34.1	0.8	0.6-0.9		
Heart failure	2: No	84.2	13.4		Ref	72.3	<0.001	42.5		Ref	208.1	<0.001
	1: Yes	11.6	20.0	1.6	1.4-1.8			30.0	1.8	1.7-2.0		
	ND	4.2	11.1	0.8	0.6-1.0			34.1	0.8	0.7-0.9		
Sickle cell	2: No	95.8	14.2		Ref	8.3	0.016	40.1		Ref	23.9	<0.001
	1: Yes	0.1	30.0	2.6	0.7-1.0			50.8	0.6	0.1-2.2		
	ND	4.2	11.1	0.8	0.6-0.9			34.1	0.7	0.6-0.8		
Previous stroke	2: No	74.3	13.3		Ref	48.3	<0.001	42.0		Ref	174.3	<0.001
	1: Yes	21.5	17.3	1.4	1.2-1.5			47.0	1.5	1.4-1.7		
	ND	4.2	11.1	0.8	0.6-1.0			34.1	0.8	0.7-0.9		
Previous Transient ischemic attack	2: No	85.7	14.2		Ref	6.7	0.036	42.4		Ref	41.3	<0.001
	1: Yes	10.1	14.2	1.0	0.9-1.1			44.2	1.2	1.1-1.3		
	ND	4.2	11.1	0.8	0.6-0.9			34.1	0.7	0.6-0.8		
Drug/alcohol abuse	2: No	89.6	14.0		Ref	11.5	0.003	43.0		Ref	24.7	<0.001
	1: Yes	6.2	16.4	1.2	1.0-1.4			38.7	1.1	1.0-1.2		
	ND	4.2	11.1	0.8	0.6-1.0			34.1	0.7	0.6-0.8		
Family history of stroke	2: No	84.7	14.4		Ref	11.6	0.003	42.5		Ref	37.8	<0.001
	1: Yes	11.1	12.6	0.9	0.8-1.0			37.7	0.8	0.8-0.9		
	ND	4.2	11.1	0.7	0.6-0.9			34.1	0.7	0.6-0.8		

Table 4A.1 (cont'd)

Hormonal replacement therapy	2: No	95.2	14.2	Ref		7.4	0.024	42.4		Ref		24.4	<0.001
	1: Yes	0.6	11.5	0.8	0.4-1.4			51.9	0.8	0.6-1.2			
	ND	4.2	11.1	0.8	0.6-0.9			34.1	0.7	0.6-0.8			
Prosthetic heart valve	2: No	94.3	14.1	Ref		12.1	0.002	42.4		Ref		33.8	<0.001
	1: Yes	1.5	19.1	1.4	1.1-1.9			51.9	1.5	1.2-1.8			
	ND	4.2	11.1	0.8	0.6-0.9			34.1	0.7	0.6-0.8			
Migraine	2: No	92.1	14.3	Ref		10.2	0.006	42.7		Ref		32.9	<0.001
	1: Yes	3.7	11.8	0.8	0.6-1.0			36.9	0.8	0.7-0.9			
	ND	4.2	11.1	0.7	0.6-0.9			34.1	0.7	0.6-0.8			
Obesity overweight	2: No	51.6	14.2	Ref		6.7	0.036	42.3		Ref		23.5	<0.001
	1: Yes	44.2	14.2	1.0	0.9-1.1			42.7	1.0	1.0-1.1			
	ND	4.2	11.1	0.8	0.6-0.9			34.1	0.7	0.6-0.8			
Chronic renal insufficiency	2: No	83.2	13.3	Ref		86.3	<0.001	40.5		Ref		226.6	<0.001
	1: Yes	12.6	20.3	1.7	1.5-1.9			55.9	1.9	1.7-2.0			
	ND	4.2	11.1	0.8	0.7-1.0			34.1	0.8	0.7-0.9			
Sleep apnea	2: No	87.0	14.2	Ref		6.7	0.036	42.2		Ref		28.6	<0.001
	1: Yes	8.9	14.2	1.0	0.9-1.2			45.2	1.1	1.0-1.2			
	ND	4.2	11.1	0.8	0.6-0.9			34.1	0.7	0.6-0.8			
Depression	2: No	77.6	13.8	Ref		16.6	<0.001	41.3		Ref		69.7	<0.001
	1: Yes	18.2	15.9	1.2	1.1-1.3			47.6	1.3	1.2-1.4			
	ND	4.2	11.1	0.8	0.6-1.0			34.1	0.7	0.6-0.9			
Deep vein thrombosis/ pulmonary embolism	2: No	93.7	14.1	Ref		12.4	0.002	42.3		Ref		33.8	<0.001
	1: Yes	2.1	18.4	1.4	1.1-1.8			50.4	1.4	1.1-1.7			
	ND	4.2	11.1	0.8	0.6-0.9			34.1	0.7	0.6-0.8			
Familial hypercholesterolemia	2: No	95.6	14.2	Ref		8.7	0.013	42.5		Ref		25.6	<0.001
	1: Yes	0.2	7.1	0.5	0.1-1.5			31.0	0.6	0.3-1.2			
	ND	4.2	11.1	0.8	0.6-0.9			34.1	0.7	0.6-0.8			
Vaping	2: No	95.8	14.2	Ref		8.5	0.014	42.5		Ref		29.8	<0.001
	1: Yes	0.0	0.0	0.0	0.0-1.5E+10 8			31.0	0.0	0.0-3.6E+63			
	ND	4.2	11.1	0.8	0.6-0.9			34.1	0.7	0.6-0.8			
Emerging infectious diseases	2: No	95.7	14.2	Ref		6.9	0.031	42.5		Ref		23.2	<0.001
	1: Yes	0.1	18.2	1.3	0.5-4.0			0.0	0.9	0.4-2.2			
	ND	4.2	11.1	0.8	0.6-0.9			34.1	0.7	0.6-0.8			
Dementia	2: No	95.3	14.1	Ref		15.1	<0.001	42.5		Ref		30.0	<0.001
	1: Yes	0.5	25.3	2.1	1.3-3.2			40.9	1.7	1.1-2.5			
	ND	4.2	11.1	0.8	0.6-0.9			34.1	0.7	0.6-0.8			
Coronary artery disease/ prior myocardial infarction	2: No	70.2	13.2	Ref		46.6	<0.001	42.4		Ref		196.7	<0.001
	1: Yes	25.6	16.9	1.3	1.2-1.5			55.6	1.6	1.5-1.7			
	ND	4.2	11.1	0.8	0.7-1.0			34.1	0.8	0.7-0.9			
Carotid stenosis	2: No	90.9	13.9	Ref		24.1	<0.001	42.0		Ref		56.5	<0.001
	1: Yes	4.9	19.0	1.4	1.2-1.7			51.6	1.5	1.3-1.7			
	ND	4.2	11.1	0.8	0.6-1.0			34.1	0.7	0.6-0.8			

Table 4A.1 (cont'd)

Diabetes mellitus	2: No	63.9	13.0	Ref		46.8	<0.001	39.9			Ref		180.0	<0.001
	1: Yes	31.9	16.5	1.3	1.2-1.4			48.9	1.5	1.4-1.6				
	ND	4.2	11.1	0.8	0.7-1.0			34.1	0.8	0.7-0.9				
Peripheral vascular disease	2: No	89.7	13.8	Ref		30.9	<0.001	41.6			Ref		105.4	<0.001
	1: Yes	6.1	19.2	1.5	1.3-1.7			55.2	1.7	1.5-1.9				
	ND	4.2	11.1	0.8	0.6-1.0			34.1	0.7	0.6-0.8				
Hypertension	2: No	21.8	11.8	Ref		34.0	<0.001	35.9			Ref		123.3	<0.001
	1: Yes	74.0	14.9	1.3	1.2-1.5			44.5	1.4	1.3-1.5				
	ND	4.2	11.1	0.9	0.7-1.2			34.1	0.9	0.8-1.1				
Smoking	2: No	79.0	14.2	Ref		6.8	0.033	42.8			Ref		26.3	<0.001
	1: Yes	16.8	13.9	1.0	0.9-1.1			41.1	0.9	0.9-1.0				
	ND	4.2	11.1	0.7	0.6-0.9			34.1	0.7	0.6-0.8				
Antithrombotic therapy administered by the end of hospital day 2	0: No/ND	6.5	17.5	Ref		67.8	<0.001	43.9			Ref		47.4	<0.001
	1: Yes	75.4	13.2	0.7	0.6-0.8			41.2	0.9	0.8-1.0				
	2: NC	15.8	17.7	1.0	0.9-1.2			46.9	1.1	1.0-1.3				
	Missing	2.2	8.3	0.4	0.3-0.6			33.8	0.7	0.5-0.8				
Completed brain imaging	0: No/ND	1.3	11.3	Ref		19.4	<0.001	38.1			Ref		18.8	<0.001
	1: Yes	80.6	14.4	1.3	0.9-2			42.7	1.2	0.9-1.6				
	2: NC	17.5	13.1	1.2	0.8-1.8			40.5	1.1	0.9-1.4				
	Missing	0.6	4.0	0.3	0.1-0.9			27.4	0.6	0.4-1.0				
Documented DVT or PE	0: No	99.0	14.0	Ref		4.9	0.026	42.0			Ref		11.8	<0.001
	1: Yes	1.0	19.9	1.5	1.1-2.2			54.4	1.6	1.2-2.2				
Catheter-based stroke treatment	1: No	5.0	16.8	Ref		6.0	0.014	46.1			Ref		6.6	0.010
	2: Yes	95.0	13.9	0.8	0.7-1.0			41.9	0.8	0.7-1.0				
IV thrombolytic initiated	0: No	73.2	13.8	Ref		30.7	<0.001	42.1			Ref		17.5	<0.001
	1: Yes	14.7	12.5	0.9	0.8-1.0			39.5	0.9	0.8-1.0				
	ND	12.1	17.7	1.3	1.2-1.5			45.3	1.1	1-1.2				
Patient NPO throughout the entire hospital stay	0: No	97.3	13.6	Ref		82.2	<0.001	41.7			Ref		64.3	<0.001
	1: Yes	2.7	29.1	2.6	2.1-3.1			59.2	2.0	1.7-2.4				
Treatment for Hospital-Acquired Pneumonia:	0: No/ND	48.3	13.8	Ref		20.3	<0.001	41.5			Ref		38.3	<0.001
	1: Yes	1.2	24.8	2.1	1.5-2.8			61.6	2.3	1.7-2.9				
	2: NC	50.4	14.1	1.0	0.9-1.1			42.2	1.0	1.0-1.1				
Treatment for urinary tract infection (UTI)	0: No	97.0	13.9	Ref		9.6	0.002	Ref			Ref		22.4	<0.001
	1: Yes	3.0	18.7	1.4	1.1-1.8			1.5	1.3-1.8					
Antidepressant treatment	0: No	84.5	13.8	Ref		4.6	0.032	41.1			Ref		44.8	<0.001
	1: Yes	15.5	15.3	1.1	1-1.3			47.7	1.3	1.2-1.4				
Antihypertensive treatment	0: No	28.1	14.0	Ref		0.0	0.845	39.3			Ref		24.6	<0.001
	1: Yes	71.9	14.1	1.0	0.9-1.1			43.2	1.2	1.1-1.3				
Cholesterol reducing treatment	0: No	12.7	20.3	Ref		83.0	<0.001	46.4			Ref		20.6	<0.001
	1: Yes	87.3	13.1	0.6	0.5-0.7			41.5	0.8	0.8-0.9				
Ambulatory status at discharge	2: Able to ambulate independently (no help from another person) w/ or w/o device	42.8	10.9	Ref		245.0	<0.001	36.1			Ref		360.1	<0.001
	3: With assistance (from person)	27.0	15.3	1.5	1.3-1.6			48.0	1.6	1.5-1.8				

Table 4A.1 (cont'd)

	4: Unable to ambulate	8.0	26.6	3.0	2.6-3.4			58.1	2.5	2.2-2.7		
	ND	22.3	14.2	1.3	1.2-1.5			41.1	1.2	1.1-1.3		
Persistent or Paroxysmal Atrial Fibrillation/Flutter	0: No	75.5	13.0	Ref		56.2	<0.001	39.4	Ref		193.0	<0.001
	1: Yes	24.0	17.5	1.4	1.3-1.6			50.9	1.6	1.5-1.7		
	ND	0.6	12.7	1.0	0.6-1.7			33.6	0.8	0.5-1.2		
Assessed for Rehabilitation Services	0: No	0.9	11.4	Ref		403.2	<0.001		Ref		94.3	<0.001
	1: Yes	96.3	13.1	1.2	0.7-1.8				1.2	0.9-1.7		
	ND	2.8	49.8	7.7	4.7-12.5				2.9	2.1-4.1		
Inpatient avg length of stay in days	>5.4 (high)	23.3	14.2	Ref		0.5	0.771	42.6	Ref		0.7	0.696
	<3.9 (low)	4.2	14.8	1.1	0.9-1.3			42.7	1.0	0.9-1.2		
	3.9-5.4 (normal)	72.5	14.0	1.0	0.9-1.1			42.0	1.0	0.9-1.0		
Hospital participates in any bundled payment arrangements	0: No	24.4	14.2	Ref		4.5	0.216	41.2	Ref		14.0	0.003
	1: Yes	45.3	14.5	1.0	0.9-1.1			42.7	1.1	1.0-1.1		
	2: Did previously but no longer doing so	5.5	13.9	1.0	0.8-1.2			37.6	0.9	0.7-1.0		
	ND	24.8	13.2	0.9	0.8-1.0			43.1	1.1	1.0-1.2		
Bed size	3: 50-99	0.8	15.4	Ref		32.9	<0.001	37.8	Ref		19.0	0.002
	4: 100-199	8.4	17.1	1.1	0.7-1.8			42.3	1.2	0.9-1.7		
	5: 200-299	13.5	16.3	1.1	0.7-1.7			45.0	1.3	1.0-1.9		
	6: 300-399	22.7	14.0	0.9	0.6-1.4			43.3	1.3	0.9-1.7		
	7: 400-499	17.8	13.7	0.9	0.6-1.4			41.8	1.2	0.8-1.6		
	8: >=500	36.8	12.8	0.8	0.5-1.3	40.7	1.1	0.8-1.6				
Core-based statistical area type	Metro	94.3	14.2	Ref		3.4	0.18	42.5	Ref		17.6	<0.001
	Micro	3.2	11.8	0.8	0.6-1.0			38.6	0.9	0.7-1.0		
	Rural	2.5	13.0	0.9	0.7-1.2			34.0	0.7	0.6-0.8		
Contracts with commercial payers where payment is tied to performance on quality/safety metrics	0: No	10.5	15.5	Ref		5.6	0.062	44.4	Ref		8.5	0.014
	1: Yes	65.9	14.1	0.9	0.8-1.0			41.5	0.9	0.8-1.0		
	ND	23.6	13.3	0.8	0.7-1.0			43.1	0.9	0.9-1.0		
Type of authority responsible for establishing policy concerning overall operations	14: Government, non-federal (city)	2.0	15.3	Ref		17.1	<0.001	47.2	Ref		15.5	0.002
	21: NGO, not for profit (Church)	12.4	15.6	1.0	0.8-1.4			42.1	0.8	0.7-1.0		
	23: NGO, not for profit (other)	82.8	13.6	0.9	0.7-1.2			41.8	0.8	0.7-1.0		
	33: investor owned, for profit (corporation)	2.9	18.8	1.3	0.9-1.8			49.0	1.1	0.8-1.4		
Rural Referral Center	0: No	75.1	13.9	Ref		1.9	0.169	42.2	Ref		0.1	0.797
	1: Yes	24.9	14.7	1.1	1-1.2			42.0	1.0	0.9-1.1		
Stroke accreditation certification program	1: Joint Commission International (JCI)	91.4	13.9	Ref		4.5	0.215	41.9	Ref		7.4	0.059
	2: Det Norske Veritas (DNV)	0.2	14.7	1.1	0.4-2.8			32.4	0.7	0.3-1.4		
	3: Healthcare Facilities Accreditation Program (HFAP)	6.3	14.4	1.0	0.9-1.2			43.7	1.1	1-1.2		
	ND	2.1	17.7	1.3	1-1.7			47.4	1.2	1-1.5		
Stroke accreditation	CSC: Comprehensive Stroke Center.	45.3	12.4	Ref		35.9	<0.001	40.3	Ref		23.1	<0.001
	PSC: Primary Stroke Center	44.2	15.4	1.3	1.2-1.4			43.9	1.2	1.1-1.2		

Table 4A.1 (cont'd)

	TSR: Thrombectomy Capable Stroke Center	10.6	15.6	1.3	1.1-1.5			42.8	1.1	1-1.2		
Stroke rehab accreditation	0: No	90.0	14.1		Ref	1.2	0.273	42.5		Ref	10.3	0.001
	1: Yes	10.0	13.2	0.9	0.8-1.1			38.7	0.9	0.8-0.9		
Accreditation Council for Graduate Medical Education accredited programs	0: No	6.3	17.0		Ref	8.7	0.003	41.1		Ref	0.6	0.426
	1: Yes	93.7	13.9	0.8	0.7-0.9			42.2	1.0	0.9-1.2		
Medical school affiliation reported to American Medical Association	0: No	16.8	15.5		Ref	6.4	0.011	42.8		Ref	0.7	0.409
	1: Yes	83.2	13.8	0.9	0.8-1.0			42.0	1.0	0.9-1.0		
Accreditation by Commission on Accreditation of Rehabilitation Facilities (CARF)	0: No	88.1	14.0		Ref	1.1	0.284	41.8		Ref	5.2	0.022
	1: Yes	11.9	14.8	1.1	0.9-1.2			44.4	1.1	1-1.2		
Member of Council of Teaching Hospital of the Association of American Medical Colleges (COH)	0: No	66.9	14.5		Ref	6.9	0.009	42.1		Ref	0.0	0.969
	1: Yes	33.1	13.1	0.9	0.8-1.0			42.2	1.0	0.9-1.1		
System member	0: No	14.5	13.4		Ref	1.0	0.307	41.5		Ref	0.6	0.444
	1: Yes	85.5	14.2	1.1	0.9-1.2			42.3	1.0	1.0-1.1		
Magnetic resonance imaging (MRI) capable hospital	0: No	1.9	20.7		Ref	13.4	0.001	51.1		Ref	12.5	0.002
	1: Yes	93.8	13.9	0.6	0.5-0.8			42.0	0.7	0.6-0.9		
	ND	4.3	14.9	0.7	0.5-0.9			41.2	0.7	0.5-0.9		
Neurological services hospital	0: No	2.8	10.3		Ref	7.5	0.02	38.0		Ref	4.4	0.111
	1: Yes	92.9	14.1	1.4	1.1-1.9			42.3	1.2	1-1.4		
	ND	4.3	14.9	1.5	1.1-2.1			41.2	1.1	0.9-1.4		
Occupancy rate	>0.8 (high)	9.0	13.8		Ref	0.2	0.917	40.8		Ref	3.1	0.212
	<.65 (low)	34.7	14.2	1.0	0.9-1.2			41.7	1.0	0.9-1.2		
	0.65-0.80 (normal)	56.3	14.0	1.0	0.9-1.2			42.7	1.1	1-1.2		
Physical rehabilitation care hospital	0: No	28.4	14.7		Ref	3.1	0.210	43.7		Ref	7.9	0.020
	1: Yes	67.3	13.7	0.9	0.8-1.0			41.5	0.9	0.9-1.0		
	ND	4.3	14.9	1.0	0.8-1.2			41.2	0.9	0.8-1.0		
Physical rehabilitation outpatient services hospital	0: No	6.2	12.4		Ref	3.3	0.190	39.0		Ref	5.8	0.055
	1: Yes	89.5	14.1	1.2	1-1.4			42.4	1.2	1-1.3		
	ND	4.3	14.9	1.2	1-1.6			41.2	1.1	0.9-1.3		
Skilled nursing care hospital	0: No	81.6	14.1		Ref	1.1	0.588	42.5		Ref	3.9	0.143
	1: Yes	14.1	13.5	1.0	0.8-1.1			40.5	0.9	0.9-1.0		
	ND	4.3	14.9	1.1	0.9-1.3			41.2	1.0	0.8-1.1		
Telehealth stroke care hospital	0: No	36.0	13.6		Ref	2.2	0.332	41.3		Ref	4.0	0.135
	1: Yes	59.7	14.3	1.1	1.0-1.2			42.7	1.1	1-1.1		
	ND	4.3	14.9	1.1	0.9-1.4			41.2	1.0	0.9-1.2		

Table 4A.1 (cont'd)

Comorbidities (HCC codes from claims data)	Telehealth stroke care - health system	0: No	62.3	14.2	Ref	1.2	0.542	42.2	Ref	0.3	0.853			
		1: Yes	33.4	13.7	1.0			0.9-1	42.1			1.0	0.9-1.1	
		ND	4.3	14.9	1.1			0.9-1.3	41.2			1.0	0.8-1.1	
		Hospital maintains a separate nursing home type of long-term care unit	0: No	94.5	14.0	Ref	1.4	0.233	42.0	Ref	3.2	0.074		
			1: Yes	5.5	15.3	1.1			0.9-1.3	44.8			1.1	1-1.3
		Amyotrophic Lateral Sclerosis and Other Motor Neuron Disease (HCC 73)	0: No	99.6	14.0	Ref	9.0	0.011	42.1	Ref	6.5	0.038		
			1: Yes	0.1	18.8	1.4			0.4-5.0	37.5			0.8	0.3-2.3
			ND	0.3	29.6	2.6			1.4-4.6	59.3			2.0	1.2-3.4
		Acute Myocardial Infarction (HCC 86)	0: No	94.4	13.4	Ref	90.1	<0.001	41.0	Ref	162.1	<0.001		
			1: Yes	5.3	24.2	2.1			1.8-2.4	60.9			2.2	2.0-2.5
			ND	0.3	29.6	2.7			1.5-4.9	59.3			2.1	1.2-3.6
		Angina Pectoris (HCC 88)	0: No	94.9	13.8	Ref	26.4	<0.001	41.4	Ref	73.3	<0.001		
			1: Yes	4.8	18.8	1.5			1.2-1.7	55.1			1.7	1.5-2.0
			ND	0.3	29.6	2.6			1.5-4.7	59.3			2.1	1.2-3.5
		Specified Heart Arrhythmias (HCC 96)	0: No	67.1	12.6	Ref	72.5	<0.001	38.1	Ref	263.1	<0.001		
			1: Yes	32.7	16.9	1.4			1.3-1.5	50.3			1.6	1.5-1.7
			ND	0.3	29.6	2.9			1.6-5.2	59.3			2.4	1.4-4.1
		Artificial Openings for Feeding or Elimination (HCC 188)	0: No	98.5	13.9	Ref	26.4	<0.001	41.9	Ref	43.3	<0.001		
			1: Yes	1.2	24.3	2.0			1.5-2.7	61.7			2.2	1.7-2.9
			ND	0.3	29.6	2.6			1.5-4.7	59.3			2.0	1.2-3.5
		Atherosclerosis of the Extremities with Ulceration or Gangrene (HCC 106)	0: No	99.0	13.9	Ref	17.6	<0.001	41.9	Ref	32.9	<0.001		
			1: Yes	0.7	23.5	1.9			1.3-2.8	64.0			2.5	1.7-3.5
			ND	0.3	29.6	2.6			1.4-4.7	59.3			2.0	1.2-3.5
		Aspiration and Specified Bacterial Pneumonias (HCC 114)	0: No	97.4	13.7	Ref	76.1	<0.001	41.6	Ref	87.6	<0.001		
			1: Yes	2.3	28.7	2.5			2.1-3.1	62.9			2.4	2.0-2.9
			ND	0.3	29.6	2.7			1.5-4.8	59.3			2.0	1.2-3.5
		Bone/Joint/Muscle Infections/Necrosis (HCC 39)	0: No	98.6	13.9	Ref	17.0	<0.001	41.9	Ref	31.6	<0.001		
		1: Yes	1.2	21.1	1.7	1.2-2.3			58.7	2.0			1.5-2.6	
		ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0			1.2-3.5	
	Breast, Prostate, and Other Cancers and Tumors (HCC 12)	0: No	92.0	13.7	Ref	22.6	<0.001	41.7	Ref	23.6	<0.001			
		1: Yes	7.7	17.3	1.3			1.1-1.5	47.2			1.3	1.1-1.4	
		ND	0.3	29.6	2.6			1.5-4.8	59.3			2.0	1.2-3.5	
	Cardio-Respiratory Failure and Shock (HCC 84)	0: No	91.2	13.0	Ref	145.4	<0.001	40.3	Ref	265.2	<0.001			
		1: Yes	8.5	24.3	2.1			1.9-2.4	60.9			2.3	2.1-2.6	
		ND	0.3	29.6	2.8			1.6-5.0	59.3			2.2	1.2-3.7	
	Cerebral Hemorrhage (HCC 99)	0: No	84.4	13.3	Ref	52.0	<0.001	41.5	Ref	23.8	<0.001			
		1: Yes	15.3	18.0	1.4			1.3-1.6	45.6			1.2	1.1-1.3	
		ND	0.3	29.6	2.7			1.5-4.9	59.3			2.1	1.2-3.5	
	Cerebral Palsy (HCC 74)	0: No	99.6	14.0	Ref	8.7	0.013	42.1	Ref	8.4	0.015			
		1: Yes	0.1	14.8	1.1			0.4-3.1	55.6			1.7	0.8-3.7	
		ND	0.3	29.6	2.6			1.4-4.6	59.3			2.0	1.2-3.4	
	Congestive Heart Failure (HCC 85)	0: No	71.6	12.3	Ref	129.7	<0.001	37.7	Ref	398.9	<0.001			
		1: Yes	28.1	18.5	1.6			1.5-1.8	53.4			1.9	1.8-2.0	
		ND	0.3	29.6	3.0			1.7-5.4	59.3			2.4	1.4-4.1	

Table 4A.1 (cont'd)

Chronic Hepatitis (HCC 29)	0: No	98.8	13.9	Ref		18.0	<0.001	41.9			Ref		33.0	<0.001
	1: Yes	0.9	22.5	1.8	1.3-2.6			61.2	2.2	1.6-3.0				
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5				
Chronic Kidney Disease, Severe (Stage 4) (HCC 137)	0: No	96.0	13.7	Ref		42.9	<0.001	41.3			Ref		129.6	<0.001
	1: Yes	3.7	21.9	1.8	1.5-2.1			62.3	2.3	2.0-2.7				
	ND	0.3	29.6	2.7	1.5-4.8			59.3	2.1	1.2-3.6				
Chronic Kidney Disease, Stage 5 (HCC 136)	0: No	97.1	13.6	Ref		100.3	<0.001	41.2			Ref		225.9	<0.001
	1: Yes	2.6	30.4	2.8	2.3-3.4			74.2	4.1	3.4-5.0				
	ND	0.3	29.6	2.7	1.5-4.8			59.3	2.1	1.2-3.6				
Chronic Pancreatitis (HCC 34)	0: No	99.3	14.0	Ref		15.8	<0.001	42.0			Ref		18.2	<0.001
	1: Yes	0.4	25.0	2.1	1.3-3.4			60.7	2.1	1.4-3.3				
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5				
Chronic Ulcer of Skin, Except Pressure (HCC 161)	0: No	96.3	13.7	Ref		42.6	<0.001	41.5			Ref		75.1	<0.001
	1: Yes	3.4	22.2	1.8	1.5-2.2			57.8	1.9	1.6-2.3				
	ND	0.3	29.6	2.6	1.5-4.8			59.3	2.0	1.2-3.5				
Cirrhosis of Liver (HCC 28)	0: No	98.7	13.9	Ref		35.5	<0.001	41.9			Ref		27.0	<0.001
	1: Yes	1.0	28.0	2.4	1.8-3.3			58.0	1.9	1.4-2.5				
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5				
Coagulation Defects and Other Specified Hematological Disorders (HCC 48)	0: No	90.1	13.3	Ref		72.2	<0.001	40.7			Ref		160.2	<0.001
	1: Yes	9.6	20.4	1.7	1.5-1.9			55.7	1.8	1.7-2.0				
	ND	0.3	29.6	2.7	1.5-4.9			59.3	2.1	1.2-3.7				
Coma, Brain Compression/Anoxic Damage (HCC 80)	0: No	90.3	13.3	Ref		85.2	<0.001	41.2			Ref		67.6	<0.001
	1: Yes	9.4	21.2	1.8	1.6-2.0			50.8	1.5	1.3-1.6				
	ND	0.3	29.6	2.8	1.5-4.9			59.3	2.1	1.2-3.6				
Chronic Obstructive Pulmonary Disease (HCC 111)	0: No	78.6	13.0	Ref		66.0	<0.001	39.2			Ref		259.8	<0.001
	1: Yes	21.1	17.7	1.4	1.3-1.6			53.1	1.8	1.6-1.9				
	ND	0.3	29.6	2.8	1.6-5.1			59.3	2.3	1.3-3.9				
Colorectal, Bladder, and Other Cancers (HCC 11)	0: No	97.0	13.8	Ref		27.0	<0.001	41.8			Ref		25.4	<0.001
	1: Yes	2.7	20.8	1.6	1.3-2.0			51.4	1.5	1.2-1.8				
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5				
Cystic Fibrosis (HCC 110)	0: No	99.7	14.0	Ref		10.2	0.006	42.1			Ref		6.4	0.040
	1: Yes	0.0	50.0	6.1	0.4-98.2			50.0	1.4	0.1-22.0				
	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.4				
Diabetes with Acute Complications (HCC 17)	0: No	98.8	13.9	Ref		32.7	<0.001	41.9			Ref		34.3	<0.001
	1: Yes	1.0	27.7	2.4	1.7-3.3			61.2	2.2	1.6-2.9				
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5				
Diabetes with Chronic Complications (HCC 18)	0: No	69.8	12.6	Ref		77.6	<0.001	38.3			Ref		265.0	<0.001
	1: Yes	30.0	17.2	1.4	1.3-1.6			50.8	1.7	1.6-1.8				
	ND	0.3	29.6	2.9	1.6-5.2			59.3	2.3	1.4-4.0				
Diabetes (HCC 17-18-19)	0: No	58.9	12.4	Ref		69.5	<0.001	37.6			Ref		235.9	<0.001
	1: Yes	40.8	16.4	1.4	1.3-1.5			48.6	1.6	1.5-1.7				
	ND	0.3	29.6	3.0	1.7-5.4			59.3	2.4	1.4-4.2				
Diabetes without Complication (HCC 19)	0: No	88.9	14.0	Ref		8.7	0.013	42.1			Ref		6.4	0.041
	1: Yes	10.8	13.9	1.0	0.9-1.1			42.0	1.0	0.9-1.1				

Table 4A.1 (cont'd)

	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.4		
	0: No	97.9	13.7	Ref				41.4	Ref			
Dialysis Status (HCC 134)	1: Yes	1.8	32.3	3.0	2.4-3.8	84.4	<0.001	79.7	5.5	4.3-7.2	213.0	<0.001
	ND	0.3	29.6	2.7	1.5-4.8			59.3	2.1	1.2-3.5		
	0: No	95.9	13.9	Ref				41.8	Ref			
Drug/Alcohol Dependence (HCC 55)	1: Yes	3.8	16.0	1.2	1.0-1.4	11.2	0.004	48.5	1.3	1.1-1.5	19.3	<0.001
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5		
	0: No	99.5	14.0	Ref				42.1	Ref			
Drug/Alcohol Psychosis (HCC 54)	1: Yes	0.2	22.0	1.7	0.8-3.6	10.6	0.005	51.2	1.4	0.8-2.7	7.8	0.020
	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.4		
	0: No	99.2	13.9	Ref				42.0	Ref			
End-Stage Liver Disease (HCC 27)	1: Yes	0.5	30.4	2.7	1.8-4.1	26.8	<0.001	65.7	2.6	1.8-4.0	29.5	<0.001
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5		
	0: No	92.6	13.4	Ref				40.9	Ref			
Other Significant Endocrine and Metabolic Disorders (HCC 23)	1: Yes	7.1	21.5	1.8	1.5-2.0	70.2	<0.001	57.6	2.0	1.8-2.2	150.0	<0.001
	ND	0.3	29.6	2.7	1.5-4.9			59.3	2.1	1.2-3.6		
	0: No	57.9	13.1	Ref				40.6	Ref			
Hemiplegia/Hemiparesis (HCC 103)	1: Yes	41.8	15.3	1.2	1.1-1.3	27.0	<0.001	44.2	1.2	1.1-1.2	31.5	<0.001
	ND	0.3	29.6	2.8	1.6-5.0			59.3	2.1	1.2-3.7		
	0: No	98.7	14.0	Ref				42.0	Ref			
Hip Fracture/Dislocation (HCC 170)	1: Yes	1.0	17.5	1.3	0.9-1.9	10.6	0.005	48.5	1.3	1.0-1.7	9.8	0.008
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5		
	0: No	99.5	14.0	Ref				42.1	Ref			
HIV/AIDS (HCC 1)	1: Yes	0.2	20.5	1.6	0.8-3.3	10.1	0.007	54.5	1.7	0.9-3.0	9.2	0.010
	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.4		
	0: No	98.5	14.0	Ref				42.0	Ref			
Inflammatory Bowel Disease (HCC 35)	1: Yes	1.2	17.9	1.3	1.0-1.9	11.4	0.003	48.5	1.3	1.0-1.7	10.2	0.006
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5		
	0: No	97.8	13.8	Ref				41.9	Ref			
Disorders of Immunity (HCC 47)	1: Yes	2.0	22.3	1.8	1.4-2.3	28.1	<0.001	53.5	1.6	1.3-2.0	27.0	<0.001
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5		
	0: No	97.2	13.7	Ref				41.5	Ref			
Complications of Specified Implanted Device or Graft (HCC 176)	1: Yes	2.5	26.9	2.3	1.9-2.9	66.4	<0.001	65.7	2.7	2.2-3.3	120.0	<0.001
	ND	0.3	29.6	2.7	1.5-4.8			59.3	2.1	1.2-3.5		
	0: No	98.4	13.9	Ref				41.8	Ref			
Intestinal Obstruction/Perforation (HCC 33)	1: Yes	1.3	21.5	1.7	1.3-2.3	19.7	<0.001	61.5	2.2	1.7-2.9	46.6	<0.001
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5		
	0: No	98.6	13.9	Ref				41.8	Ref			
Amputation Status, Lower Limb/Amputation Complications (HCC 189)	1: Yes	1.1	25.8	2.2	1.6-2.9	29.5	<0.001	66.7	2.8	2.1-3.7	59.1	<0.001
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5		
	0: No	97.3	13.8	Ref				41.8	Ref			
Lung and Other Severe Cancers (HCC 9)	1: Yes	2.4	21.5	1.7	1.4-2.1	28.4	<0.001	55.9	1.8	1.5-2.1	43.1	<0.001
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5		

Table 4A.1 (cont'd)

Fibrosis of Lung and Other Chronic Lung Disorders (HCC 112)	0: No	97.2	13.8	Ref		29.8	<0.001	41.7			Ref		60.1	<0.001
	1: Yes	2.5	21.6	1.7	1.4-2.1			58.4	2.0	1.6-2.4				
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5				
Lymphoma and Other Cancers (HCC 10)	0: No	97.7	13.9	Ref		26.1	<0.001	41.9			Ref		18.9	<0.001
	1: Yes	2.0	21.7	1.7	1.4-2.2			50.9	1.4	1.2-1.8				
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5				
Exudative Macular Degeneration (HCC 124)	0: No	97.5	14.0	Ref		8.9	0.012	42.1			Ref		7.2	0.028
	1: Yes	2.2	14.7	1.1	0.8-1.4			44.2	1.1	0.9-1.3				
	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.5				
Major Depressive, Bipolar, and Paranoid Disorders (HCC 58)	0: No	90.1	13.6	Ref		36.5	<0.001	41.1			Ref		77.4	<0.001
	1: Yes	9.6	18.2	1.4	1.2-1.6			51.3	1.5	1.4-1.7				
	ND	0.3	29.6	2.7	1.5-4.8			59.3	2.1	1.2-3.6				
Major Head Injury (HCC 167)	0: No	98.1	13.9	Ref		17.0	<0.001	41.9			Ref		33.3	<0.001
	1: Yes	1.7	19.8	1.5	1.2-2.0			56.3	1.8	1.4-2.2				
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5				
Metastatic Cancer and Acute Leukemia (HCC 8)	0: No	97.9	13.8	Ref		31.2	<0.001	41.9			Ref		32.1	<0.001
	1: Yes	1.8	23.3	1.9	1.5-2.4			55.3	1.7	1.4-2.1				
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5				
Monoplegia, Other Paralytic Syndromes (HCC 104)	0: No	95.1	14.0	Ref		9.2	0.010	42.1			Ref		6.4	0.041
	1: Yes	4.6	13.2	0.9	0.8-1.1			42.2	1.0	0.9-1.2				
	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.4				
Morbid Obesity (HCC 22)	0: No	89.9	13.8	Ref		16.6	<0.001	41.6			Ref		26.9	<0.001
	1: Yes	9.8	16.2	1.2	1.1-1.4			47.0	1.2	1.1-1.4				
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5				
Multiple Sclerosis (HCC 77)	0: No	98.9	14.0	Ref		8.7	0.013	42.1			Ref		6.5	0.039
	1: Yes	0.8	13.7	1.0	0.6-1.5			41.0	1.0	0.7-1.3				
	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.4				
Muscular Dystrophy (HCC 76)	0: No	99.7	14.0	Ref		9.5	0.009	42.1			Ref		6.5	0.039
	1: Yes	0.1	23.1	1.8	0.5-6.7			46.2	1.2	0.4-3.5				
	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.4				
Myasthenia Gravis/Myoneural Disorders, Inflammatory and Toxic Neuropathy (HCC 75)	0: No	98.8	13.9	Ref		14.3	0.001	41.9			Ref		27.7	<0.001
	1: Yes	1.0	20.3	1.6	1.1-2.3			58.8	2.0	1.5-2.7				
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5				
Opportunistic Infections (HCC 6)	0: No	99.5	14.0	Ref		13.4	0.001	42.0			Ref		17.0	<0.001
	1: Yes	0.3	25.5	2.1	1.1-4.0			64.7	2.5	1.4-4.5				
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5				
Major Organ Transplant or Replacement Status (HCC 186)	0: No	99.2	14.0	Ref		14.4	0.001	42.0			Ref		27.1	<0.001
	1: Yes	0.5	23.2	1.9	1.2-3.0			65.3	2.6	1.7-4.0				
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5				
Paraplegia (HCC 71)	0: No	99.4	14.0	Ref		9.5	0.009	42.1			Ref		10.0	0.007
	1: Yes	0.3	18.0	1.4	0.7-2.6			54.1	1.6	1-2.7				
	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.5				

Table 4A.1 (cont'd)

Parkinson s and Huntington s Diseases (HCC 78)	0: No	97.4	14.0	Ref		9.1	0.011	41.9		Ref		15.5	<0.001
	1: Yes	2.4	15.0	1.1	0.8-1.4			49.0	1.3	1.1-1.6			
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5			
Proliferative Diabetic Retinopathy and Vitreous Hemorrhage (HCC 122)	0: No	99.3	14.0	Ref		8.7	0.013	42.1		Ref		10.8	0.004
	1: Yes	0.4	13.4	1.0	0.5-1.8			53.7	1.6	1.0-2.5			
	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.5			
Pneumococcal Pneumonia, Empyema, Lung Abscess (HCC 115)	0: No	98.6	13.9	Ref		21.1	<0.001	41.9		Ref		43.9	<0.001
	1: Yes	1.1	22.8	1.8	1.3-2.5			62.6	2.3	1.8-3.1			
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5			
Pressure Ulcer of Skin with Full Thickness Skin Loss (HCC 158)	0: No	98.8	13.9	Ref		16.2	<0.001	41.9		Ref		36.5	<0.001
	1: Yes	0.9	21.8	1.7	1.2-2.5			62.9	2.4	1.7-3.2			
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5			
Pressure Ulcer of Skin with Necrosis Through to Muscle, Tendon, or Bone (HCC 157)	0: No	99.6	14.0	Ref		10.0	0.007	42.1		Ref		13.0	0.001
	1: Yes	0.1	22.2	1.8	0.7-4.4			66.7	2.8	1.2-6.1			
	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.5			
Protein-Calorie Malnutrition (HCC 21)	0: No	94.7	13.5	Ref		79.3	<0.001	41.4		Ref		85.2	<0.001
	1: Yes	5.0	23.8	2.0	1.7-2.3			55.9	1.8	1.6-2.0			
	ND	0.3	29.6	2.7	1.5-4.8			59.3	2.1	1.2-3.6			
Quadriplegia (HCC 70)	0: No	99.4	14.0	Ref		11.8	0.003	42.0		Ref		13.4	0.001
	1: Yes	0.3	22.2	1.8	1.0-3.2			58.7	2.0	1.2-3.2			
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5			
Acute Renal Failure (HCC 135)	0: No	84.3	12.5	Ref		178.1	<0.001	39.4		Ref		329.5	<0.001
	1: Yes	15.4	22.0	2.0	1.8-2.2			57.1	2.1	1.9-2.2			
	ND	0.3	29.6	2.9	1.6-5.3			59.3	2.2	1.3-3.9			
Respiratory Arrest (HCC 83)	0: No	99.7	14.0	Ref		10.8	0.004	42.1		Ref		7.1	0.029
	1: Yes	0.0	33.3	3.1	0.8-12.3			55.6	1.7	0.5-6.4			
	ND	0.3	29.6	2.6	1.4-4.6			59.3	2.0	1.2-3.4			
Respirator Dependence/Tracheostomy Status (HCC 82)	0: No	99.1	13.9	Ref		33.2	<0.001	42.0		Ref		31.7	<0.001
	1: Yes	0.6	32.4	3.0	2.0-4.4			65.8	2.7	1.8-3.9			
	ND	0.3	29.6	2.6	1.5-4.7			59.3	2.0	1.2-3.5			
Rheumatoid Arthritis and Inflammatory Connective Tissue Disease (HCC 40)	0: No	92.1	14.0	Ref		9.1	0.011	41.6		Ref		28.2	<0.001
	1: Yes	7.7	14.6	1.1	0.9-1.2			47.9	1.3	1.2-1.4			
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5			
Schizophrenia (HCC 57)	0: No	98.5	13.9	Ref		17.2	<0.001	41.9		Ref		27.8	<0.001
	1: Yes	1.3	20.8	1.6	1.2-2.2			56.7	1.8	1.4-2.3			
	ND	0.3	29.6	2.6	1.4-4.7			59.3	2.0	1.2-3.5			
Seizure Disorders and Convulsions (HCC 79)	0: No	92.2	13.6	Ref		46.5	<0.001	41.2		Ref		87.9	<0.001
	1: Yes	7.5	19.6	1.6	1.4-1.8			53.4	1.6	1.5-1.8			
	ND	0.3	29.6	2.7	1.5-4.8			59.3	2.1	1.2-3.6			
Septicemia, Sepsis, Systemic Inflammatory Response Syndrome/Shock (HCC 2)	0: No	95.7	13.5	Ref		81.5	<0.001	41.1		Ref		187.1	<0.001
	1: Yes	4.0	25.3	2.2	1.8-2.6			65.6	2.7	2.3-3.2			
	ND	0.3	29.6	2.7	1.5-4.8			59.3	2.1	1.2-3.6			

Table 4A.1 (cont'd)

Severe Skin Burn or Condition (HCC 162)^	0: No	-	14.0	Ref	9.9	0.007	42.1	Ref		13.3	0.001	
	1: Yes	-	0.0	0.0			3.94E+79	100.0	145,0			0.0-9.84E+88
	ND	-	29.6	2.6			1.4-4.6	59.3	2.0			1.2-3.4
Severe Head Injury (HCC 166)^	0: No	-	14.0	Ref	10.2	0.006	42.1	Ref		6.4	0.040	
	1: Yes	-	50.0	6.1			0.4-98.2	50.0	1.4			0.1-22.0
	ND	-	29.6	2.6			1.4-4.6	59.3	2.0			1.2-3.4
Severe Hematological Disorders (HCC 46)	0: No	99.1	14.0	Ref	9.1	0.010	42.0	Ref		17.3	<0.001	
	1: Yes	0.6	16.2	1.2			0.7-2.0	57.7	1.9			1.3-2.7
	ND	0.3	29.6	2.6			1.4-4.6	59.3	2.0			1.2-3.5
Spinal Cord Disorders/Injuries (HCC 72)	0: No	98.6	14.0	Ref	13.2	0.001	42.0	Ref		16.0	<0.001	
	1: Yes	1.1	19.2	1.5			1.0-2.1	52.6	1.5			1.2-2.0
	ND	0.3	29.6	2.6			1.4-4.7	59.3	2.0			1.2-3.5
Ischemic or Unspecified Stroke (HCC 100)	0: No	14.8	18.5	Ref	61.3	<0.001	47.8	Ref		50.4	<0.001	
	1: Yes	84.9	13.2	0.7			0.6-0.7	41.1	0.8			0.7-0.8
	ND	0.3	29.6	1.9			1.0-3.4	59.3	1.6			0.9-2.8
Traumatic Amputations and Complications (HCC 173)	0: No	99.2	14.0	Ref	15.3	<0.001	42.0	Ref		15.4	<0.001	
	1: Yes	0.5	23.5	1.9			1.2-3.0	56.9	1.8			1.2-2.7
	ND	0.3	29.6	2.6			1.4-4.7	59.3	2.0			1.2-3.5
Unstable Angina and Other Acute Ischemic Heart Disease (HCC 87)	0: No	96.1	13.6	Ref	65.0	<0.001	41.4	Ref		105.8	<0.001	
	1: Yes	3.6	24.5	2.1			1.7-2.5	60.6	2.2			1.9-2.5
	ND	0.3	29.6	2.7			1.5-4.8	59.3	2.1			1.2-3.5
Vascular Disease (HCC 108)	0: No	73.9	12.8	Ref	76.9	<0.001	38.9	Ref		237.5	<0.001	
	1: Yes	25.8	17.6	1.5			1.3-1.6	51.3	1.7			1.5-1.8
	ND	0.3	29.6	2.9			1.6-5.2	59.3	2.3			1.3-3.9
Vascular Disease with Complications (HCC 107)	0: No	95.3	13.8	Ref	28.1	<0.001	41.7	Ref		39.6	<0.001	
	1: Yes	4.5	19.3	1.5			1.3-1.8	51.6	1.5			1.3-1.7
	ND	0.3	29.6	2.6			1.5-4.7	59.3	2.0			1.2-3.5
Vertebral Fractures without Spinal Cord Injury (HCC 169)	0: No	98.0	13.9	Ref	22.0	<0.001	41.8	Ref		39.5	<0.001	
	1: Yes	1.7	21.2	1.7			1.3-2.2	57.5	1.9			1.5-2.3
	ND	0.3	29.6	2.6			1.5-4.7	59.3	2.0			1.2-3.5

^Due to Medicare DUA data restrictions fields that represent less than 11 must not be presented.

Table 4A.2: Descriptive statistics of hospital sites characteristics from the American Hospital Association's database (n=31).

Hospital Characteristic	Category	Number of hospitals (%)
Bed size	50-99	1 (3.2)
	100-199	7 (22.6)
	200-299	5 (16.1)
	300-399	8 (25.8)
	400-499	5 (16.1)
	>=500	5 (16.1)
Type of authority responsible for establishing policy concerning overall operations	Government, non-federal (city)	1 (3.2)
	NGO, not for profit (Church)	4 (12.9)
	NGO, not for profit (other)	24 (77.4)
	Investor owned, for profit (corporation)	2 (6.5)
Core-based statistical area	Metro	29 (93.6)
	Micro	1 (3.2)
	Rural	1 (3.2)
Rural referral center	No	17 (54.8)
	Yes	14 (45.2)
System member	No	5 (16.1)
	Yes	26 (83.9)
Stroke accreditation	CSC: Comprehensive Stroke Center	8 (25.8)
	PSC: Primary Stroke Center	20 (64.5)
	TSR: Thrombectomy Capable Stroke Center	3 (9.7)
Stroke rehabilitation accreditation	No	29 (93.6)
	Yes	2 (6.4)
Accreditation Council for Graduate Medical Education accredited programs	No	4 (12.9)
	Yes	27 (87.1)
Medical school affiliation reported to American Medical Association	No	7 (22.6)
	Yes	24 (77.4)
Magnetic resonance imaging (MRI) capable hospital	No	1 (3.2)
	Yes	27 (87.1)
	Not determined (ND)	3 (9.7)

Table 4A.3: Descriptive statistics of readmission risk and patient characteristics by hospital site (n=31).

Hospital	Total number of stroke admissions (N=19,382)	30-day readmission proportion (N=2,724)	1-year readmission proportion (N=8,169)	Age^				Race*			Sex^			Insurance^				Stroke Type^	
				<65	65-74	75-84	>=85	White	Black	Other	M	F	PPO	PPO MA	HMO	HMO MA	Medicare FFS	Ischemic	Hemorrhagic
1	998	9.9	35	20.1	30.7	30.5	18.7	87.7	9.1	1.1	47.4	52.6	11.8	14.8	2.4	3.1	67.8	86.6	13.4
2	632	10.8	41.5	20.1	29.8	27.2	22.9	85.3	11.4	1.0	45.9	54.1	6.2	10.3	4.0	10.8	68.8	89.1	10.9
3	274	10.9	37.6	20.4	29.9	33.9	15.7	81.8	14.2	0.4	55.1	44.9	11.3	13.1	4.7	8.0	62.8	98.9	1.1
4	636	11.2	35.8	13.2	27.8	32.9	26.1	83.2	14.2	0.5	46.5	53.5	7.1	12.6	1.7	0.0	78.6	92.3	7.2
5	1,576	11.4	38.0	23.3	30.6	28.9	17.3	85.1	8.0	1.8	50.7	49.3	13.5	15.6	6.2	8.6	56.1	85.8	14.2
6	334	11.7	38.9	24.0	26.4	28.1	21.6	89.5	6.6	2.1	49.7	50.3	13.5	9.3	4.5	6.6	66.2	85.6	14.4
7	621	11.8	38.6	13.4	29.0	35.4	22.2	98.6	0.3	0.8	55.1	44.9	8.9	15.6	3.1	3.4	69.1	83.7	16.3
8	1,247	11.9	37.9	30.3	30.8	23.6	15.3	82.9	10.2	3.1	50.0	50.0	16.3	14.2	10.3	4.6	54.6	78.2	21.8
9	1,740	12.2	37.9	21.8	26.8	29.5	21.8	85.4	5.7	1.0	49.0	51.0	12.5	12.2	4.7	6.7	63.9	85.0	15.0
10	483	13.0	34.0	15.1	26.1	35.6	23.2	94.8	0.2	3.1	50.7	49.3	10.1	17.4	2.3	1.7	68.5	90.7	9.3
11	547	13.7	39.7	25.6	26.9	29.3	18.3	79.5	8.6	2.4	48.8	51.2	13.9	11.0	5.3	13.5	56.3	87.9	12.1
12	1,320	13.9	47.7	26.6	31.3	26.2	13.9	36.4	54.3	0.8	48.4	51.6	14.3	16.3	3.9	1.6	63.9	85.4	14.6
13	537	14.5	47.3	18.4	27.2	34.8	19.6	93.1	1.5	0.6	47.5	52.5	10.1	18.4	1.9	5.8	63.9	91.1	8.9
14	556	14.6	43.5	16.4	27.5	30.8	25.4	52.9	18.7	2.0	47.3	52.7	12.4	18.0	4.1	1.6	63.9	88.5	11.5
15	34	14.7	32.4	17.7	35.3	23.5	23.5	91.2	5.9	2.9	47.1	52.9	11.8	14.7	2.9	8.8	61.8	88.2	11.8
16	183	14.8	44.8	13.7	30.6	30.1	25.7	95.1	1.1	0.6	51.9	48.1	7.7	18	2.7	6.0	65.6	96.2	3.8
17	1,009	15.3	47.6	19.5	32.9	27.6	20	77.6	18.2	0.7	48.5	51.5	11.4	19.7	3.4	5.9	59.7	87.8	12.2
18	386	15.3	47.2	29.3	31.6	26.4	12.7	48.5	48.2	0.5	46.1	53.9	16.3	16.3	3.6	4.7	59.1	90.7	9.3
19	156	15.4	37.8	13.5	24.4	31.4	30.8	96.8	3.2	0.0	46.2	53.8	7.1	17.3	4.5	6.4	64.7	96.2	3.8
20	519	15.6	43.4	13.5	24.3	28.9	33.3	89.4	8.9	1.7	44.3	55.7	9.4	19.1	3.7	50	62.8	87.7	12.3
21	819	15.6	45.8	21.1	26.3	27.2	25.4	81.1	15.9	0.2	45.0	55.0	12.1	15.5	2.2	2.9	67.3	90.3	9.7
22	694	15.7	40.3	21.2	29.3	28.5	21.0	82.6	14.1	1.7	48.0	52.0	12.5	16.7	5.8	8.9	56.1	79.8	20.2
23	803	16.2	44.5	22.2	30.0	29.1	18.7	79.5	18	1.5	46.6	53.4	15.7	18.6	3.6	5.9	56.3	85.9	14.1
24	814	16.2	46.8	16.1	26.4	32.8	24.7	93	4.9	1.0	45.8	54.2	11.2	22.1	3.7	4.2	58.9	88.1	11.9
25	107	16.8	46.7	34.6	23.4	28.0	14.0	70.1	25.2	0.0	34.6	65.4	15.9	12.2	5.6	4.7	61.7	91.6	8.4
26	378	18.3	42.9	19.8	31.8	26.2	22.2	94.7	2.1	1.1	51.6	48.4	10.1	15.9	1.6	3.4	69.1	98.7	1.3
27	781	18.6	49.4	21.3	31.8	27.5	19.5	87.2	8.8	0.6	47.0	53.0	11.7	19.1	3.2	4.6	61.5	89.8	10.2
28	233	18.9	43.8	21.9	26.2	25.8	26.2	93.6	4.7	1.3	43.8	56.2	11.6	15.9	3.0	30	66.5	97.0	3.0
29	390	19.5	49.2	19.2	28.5	28.7	23.6	86.2	11.3	1.3	41.0	59.0	8.2	13.9	3.9	2.1	72.1	96.4	3.6
30	376	20.7	51.1	34.6	37.5	16.2	11.7	17.6	77.4	3.7	42.6	57.4	12.0	9.6	4.8	4.5	69.2	76.9	23.1
31	199	23.1	52.8	18.1	31.7	28.1	22.1	98.0	0.5	0.0	51.3	48.7	10.6	19.6	3.0	6.5	60.3	91.5	8.5
X^2 Test p-value				<0.001				<0.001			<0.001			<0.001					

^Proportions might add slightly more than 100 due to rounding.

*Proportions might not add to 100% because of missingness.

Table 4A.4: Multivariate logistic regression analysis of the effect of adding individual predictors to the base logit model (i.e., sex, age, race, and stroke type) on the discriminant performance of the model illustrated by the change in the area under the receiver operating characteristic curve (AUC).

Predictors group	Predictor	30-day readmission			1-year readmission		
		χ^2 test LRT	χ^2 test p-value	Δ AUC (Base model AUC = 0.538)	χ^2 test LRT	χ^2 test p-value	Δ AUC (Base model AUC = 0.558)
Demographics	Latino Ethnicity	1.1	0.301	0.001	5.5	0.019	0.001
	Insurance (From claims data)	113.3	<0.001	0.034	335.3	<0.001	0.034
Administrative related	Admission year	6.9	0.139	0.004	13.8	0.008	0.002
	Documented stroke etiology	36.9	<0.001	0.015	83.5	<0.001	0.011
	Only comfort measures	2.8	0.252	0.001	13.6	0.001	0.002
	Admission duration	135.5	<0.001	0.040	356.3	<0.001	0.039
	Discharge disposition	913.3	<0.001	0.085	677.9	<0.001	0.055
	Onset to door time	8.2	0.042	0.006	8.4	0.039	0.001
Admission related	Ambulatory status on admission	100.2	<0.001	0.032	139.8	<0.001	0.017
	Ambulatory status prior to the current event	72.9	<0.001	0.018	103.8	<0.001	0.012
	Prior Antihypertensive medication	48.3	<0.001	0.018	113.9	<0.001	0.015
	Prior cholesterol reducer medication	9.9	0.002	0.007	89.4	<0.001	0.011
	Prior anti-hyperglycemic medication	24.6	<0.001	0.012	96.8	<0.001	0.013
	Prior antiplatelet or anticoagulant medication	17.5	<0.001	0.007	141.1	<0.001	0.018
	Prior antidepressant medication	15.4	<0.001	0.007	87.7	<0.001	0.012
	Admission NIHSS	137.3	<0.001	0.039	183.4	<0.001	0.023
	Arrival Mode	108.3	<0.001	0.035	121.3	<0.001	0.016
	Where did the patient first receive care?	15.7	0.001	0.007	11.6	0.009	0.003
	Patient location when stroke symptoms discovered	69.6	<0.001	0.014	52.5	<0.001	0.008
	ED patient	11.8	0.003	0.005	6.3	0.042	0.001
	Previous medical history	Atrial fibrillation/flutter	44.1	<0.001	0.017	157.7	<0.001
Dyslipidemia		8.6	0.013	0.004	36.8	<0.001	0.005
Heart failure		72.3	<0.001	0.021	178.6	<0.001	0.022
Sickle cell		7.5	0.023	0.001	22.3	<0.001	0.003
Previous stroke		46.9	<0.001	0.016	152.4	<0.001	0.020
Previous Transient ischemic attack		6.3	0.043	0.002	36.4	<0.001	0.004
Drug/alcohol abuse		11.7	0.003	0.005	28.6	<0.001	0.004
Family history of stroke		10.3	0.006	0.004	32.1	<0.001	0.004
Hormonal replacement therapy		6.7	0.034	0.002	21.6	<0.001	0.003
Prosthetic heart valve		11.9	0.003	0.003	31.7	<0.001	0.004
Migraine		8.8	0.012	0.004	23.8	<0.001	0.003
Obesity overweight		6.5	0.038	0.002	25.2	<0.001	0.003
Chronic renal insufficiency		82.5	<0.001	0.019	195.6	<0.001	0.024
Sleep apnea	6.2	0.044	0.001	32.7	<0.001	0.004	

Table 4A.4 (cont'd)

	Depression	20.3	<0.001	0.009	92.7	<0.001	0.012
	Deep vein thrombosis/ pulmonary embolism	11.8	0.003	0.003	30.9	<0.001	0.004
	Familial hypercholesterolemia	8.2	0.017	0.002	23.1	<0.001	0.003
	Vaping	8.1	0.017	0.002	26.7	<0.001	0.003
	Emerging infectious diseases	6.5	0.038	0.002	21.1	<0.001	0.003
	Dementia	13.8	0.001	0.002	25.3	<0.001	0.003
	Coronary artery disease/ prior myocardial infarction	47.2	<0.001	0.019	178.4	<0.001	0.023
	Carotid stenosis	26.5	<0.001	0.009	55.9	<0.001	0.008
	Diabetes mellitus	46.2	<0.001	0.018	178.7	<0.001	0.024
	Peripheral vascular disease	31.0	<0.001	0.010	100.2	<0.001	0.011
	Hypertension	32.5	<0.001	0.013	85.1	<0.001	0.010
	Smoking	6.3	0.044	0.002	23.2	<0.001	0.003
	Antithrombotic therapy administered by the end of hospital day 2	39.0	<0.001	0.015	33.2	<0.001	0.005
	Completed brain imaging	20.6	<0.001	0.003	13.5	0.004	0.002
	Documented DVT or PE	4.5	0.035	0.003	10.6	0.001	0.001
Inpatient related	Catheter-based stroke treatment	9.3	0.002	0.004	11.3	0.001	0.002
	IV thrombolytic initiated	3.0	0.225	0.002	3.4	0.184	0.001
	Patient NPO throughout the entire hospital stay	70.3	<0.001	0.007	52.3	<0.001	0.006
	Treatment for Hospital-Acquired Pneumonia:	16.6	<0.001	0.005	32.7	<0.001	0.003
	Treatment for urinary tract infection (UTI)	8.4	0.004	0.003	14.3	<0.001	0.002
	Antidepressant treatment	6.8	0.009	0.003	61.3	<0.001	0.008
	Antihypertensive treatment	0.0	0.854	<0.001	11.2	0.001	0.002
Discharge related	Cholesterol reducing treatment	61.3	<0.001	0.019	13.5	<0.001	0.002
	Ambulatory status at discharge	225.8	<0.001	0.050	284.4	<0.001	0.033
	Persistent or Paroxysmal Atrial Fibrillation/Flutter	61.5	<0.001	0.023	159.4	<0.001	0.021
	Assessed for Rehabilitation Services	397.5	<0.001	0.034	91.5	<0.001	0.013
	Inpatient avg length of stay in days	1.0	0.611	0.001	0.3	0.856	<0.001
	Hospital participates in any bundled payment arrangements	3.2	0.363	0.002	9.8	0.020	0.002
	Bed size	38.8	<0.001	0.017	19.7	0.001	0.004
	Core-based statistical area type	3.1	0.210	0.002	14.6	0.001	0.002
	Contracts with commercial payers where payment is tied to performance on quality/safety metrics	5.9	0.054	0.003	10.1	0.006	0.001
	Type of authority responsible for establishing policy concerning overall operations	12.7	0.005	0.004	5.9	0.118	0.001
Hospital and system characteristics (from American Hospital Association data)	Rural Referral Center	1.2	0.267	0.002	0.7	0.392	<0.001
	Stroke accreditation certification program	4.6	0.203	0.001	8.0	0.046	0.001
	Stroke accreditation	42.0	<0.001	0.014	23.8	<0.001	0.004
	Stroke rehab accreditation	2.0	0.156	0.002	7.5	0.006	0.002
	Accreditation Council for Graduate Medical Education accredited programs	11.7	0.001	0.005	0.0	0.920	<0.001
	Medical school affiliation reported to American Medical Association	9.0	0.003	0.004	1.6	0.211	<0.001
	Accreditation by Commission on Accreditation of Rehabilitation Facilities (CARF)	2.1	0.143	0.002	8.7	0.003	0.002
	Member of Council of Teaching Hospital of the Association of American Medical Colleges (COTH)	10.9	0.001	0.003	0.8	0.365	<0.001
	System member	1.4	0.236	0.002	0.3	0.602	<0.001

Table 4A.4 (cont'd)

	Magnetic resonance imaging (MRI) capable hospital	8.2	0.016	0.003	3.5	0.173	<0.001
	Neurological services hospital	7.4	0.024	0.003	4.6	0.100	0.001
	Occupancy rate	1.6	0.451	0.001	2.2	0.327	0.001
	Physical rehabilitation care hospital	2.6	0.276	0.003	3.4	0.187	0.001
	Physical rehabilitation outpatient services hospital	3.1	0.214	0.002	3.7	0.156	0.001
	Skilled nursing care hospital	1.2	0.548	0.001	2.8	0.251	<0.001
	Telehealth stroke care hospital	2.6	0.272	0.003	2.0	0.364	0.001
	Telehealth stroke care - health system	2.5	0.284	0.002	1.6	0.440	<0.001
	Hospital maintains a separate nursing home type of long-term care unit	1.9	0.170	0.002	3.2	0.074	0.001
Comorbidities (HCC codes from claims data)	Amyotrophic Lateral Sclerosis and Other Motor Neuron Disease (HCC 73)	9.3	0.009	0.001	6.8	0.033	0.001
	Acute Myocardial Infarction (HCC 86)	91.5	<0.001	0.021	157.5	<0.001	0.016
	Angina Pectoris (HCC 88)	27.4	<0.001	0.010	73.6	<0.001	0.011
	Specified Heart Arrhythmias (HCC 96)	78.1	<0.001	0.027	219.0	<0.001	0.028
	Artificial Openings for Feeding or Elimination (HCC 188)	26.2	<0.001	0.004	44.8	<0.001	0.005
	Atherosclerosis of the Extremities with Ulceration or Gangrene (HCC 106)	18.1	<0.001	0.004	33.0	<0.001	0.004
	Aspiration and Specified Bacterial Pneumonias (HCC 114)	70.2	<0.001	0.011	84.5	<0.001	0.009
	Bone/Joint/Muscle Infections/Necrosis (HCC 39)	17.0	<0.001	0.004	33.3	<0.001	0.004
	Breast, Prostate, and Other Cancers and Tumors (HCC 12)	20.9	<0.001	0.004	18.2	<0.001	0.002
	Cardio-Respiratory Failure and Shock (HCC 84)	134.6	<0.001	0.029	266.1	<0.001	0.028
	Cerebral Hemorrhage (HCC 99)	23.5	<0.001	0.008	12.5	0.002	0.001
	Cerebral Palsy (HCC 74)	9.0	0.011	0.001	9.7	0.008	0.001
	Congestive Heart Failure (HCC 85)	131.2	<0.001	0.039	344.5	<0.001	0.039
	Chronic Hepatitis (HCC 29)	16.5	<0.001	0.003	31.5	<0.001	0.003
	Chronic Kidney Disease, Severe (Stage 4) (HCC 137)	40.7	<0.001	0.007	111.0	<0.001	0.013
	Chronic Kidney Disease, Stage 5 (HCC 136)	95.5	<0.001	0.017	222.4	<0.001	0.016
	Chronic Pancreatitis (HCC 34)	15.5	<0.001	0.002	19.0	<0.001	0.003
	Chronic Ulcer of Skin, Except Pressure (HCC 161)	42.9	<0.001	0.010	74.2	<0.001	0.009
	Cirrhosis of Liver (HCC 28)	34.1	<0.001	0.006	29.4	<0.001	0.004
	Coagulation Defects and Other Specified Hematological Disorders (HCC 48)	66.7	<0.001	0.022	156.4	<0.001	0.020
	Coma, Brain Compression/Anoxic Damage (HCC 80)	58.9	<0.001	0.014	56.7	<0.001	0.008
	Chronic Obstructive Pulmonary Disease (HCC 111)	69.1	<0.001	0.026	263.9	<0.001	0.032
	Colorectal, Bladder, and Other Cancers (HCC 11)	26.5	<0.001	0.007	22.7	<0.001	0.003
	Cystic Fibrosis (HCC 110)	10.4	0.006	0.001	6.7	0.035	0.001
	Diabetes with Acute Complications (HCC 17)	33.4	<0.001	0.005	37.0	<0.001	0.004
	Diabetes with Chronic Complications (HCC 18)	79.3	<0.001	0.026	263.7	<0.001	0.033
	Diabetes (HCC 17-18-19)	70.4	<0.001	0.025	232.1	<0.001	0.029
	Diabetes without Complication (HCC 19)	9.1	0.011	0.001	6.8	0.034	0.001
	Dialysis Status (HCC 134)	80.2	<0.001	0.013	213.0	<0.001	0.013
	Drug/Alcohol Dependence (HCC 55)	11.7	0.003	0.003	28.4	<0.001	0.004
	Drug/Alcohol Psychosis (HCC 54)	10.6	0.005	0.002	8.3	0.015	0.001
	End-Stage Liver Disease (HCC 27)	25.7	<0.001	0.005	31.5	<0.001	0.003
	Other Significant Endocrine and Metabolic Disorders (HCC 23)	61.7	<0.001	0.016	137.2	<0.001	0.018

Table 4A.4 (cont'd)

Hemiplegia/Hemiparesis (HCC 103)	29.3	<0.001	0.009	28.2	<0.001	0.003
Hip Fracture/Dislocation (HCC 170)	11.0	0.004	0.002	8.5	0.015	0.001
HIV/AIDS (HCC 1)	10.4	0.006	0.002	10.0	0.007	0.001
Inflammatory Bowel Disease (HCC 35)	12.0	0.002	0.003	11.8	0.003	0.002
Disorders of Immunity (HCC 47)	28.4	<0.001	0.007	29.9	<0.001	0.004
Complications of Specified Implanted Device or Graft (HCC 176)	64.4	<0.001	0.015	119.0	<0.001	0.011
Intestinal Obstruction/Perforation (HCC 33)	18.8	<0.001	0.004	43.4	<0.001	0.005
Amputation Status, Lower Limb/Amputation Complications (HCC 189)	30.1	<0.001	0.006	65.9	<0.001	0.006
Lung and Other Severe Cancers (HCC 9)	29.4	<0.001	0.007	45.3	<0.001	0.007
Fibrosis of Lung and Other Chronic Lung Disorders (HCC 112)	30.4	<0.001	0.007	57.1	<0.001	0.006
Lymphoma and Other Cancers (HCC 10)	26.3	<0.001	0.006	18.5	<0.001	0.003
Exudative Macular Degeneration (HCC 124)	9.3	0.009	0.002	6.7	0.035	0.001
Major Depressive, Bipolar, and Paranoid Disorders (HCC 58)	40.2	<0.001	0.015	98.2	<0.001	0.014
Major Head Injury (HCC 167)	14.0	0.001	0.004	28.6	<0.001	0.003
Metastatic Cancer and Acute Leukemia (HCC 8)	32.5	<0.001	0.008	34.9	<0.001	0.005
Monoplegia, Other Paralytic Syndromes (HCC 104)	9.2	0.010	0.002	6.8	0.034	0.001
Morbid Obesity (HCC 22)	19.0	<0.001	0.007	44.2	<0.001	0.005
Multiple Sclerosis (HCC 77)	9.0	0.011	0.001	6.9	0.031	0.001
Muscular Dystrophy (HCC 76)	9.9	0.007	0.001	7.0	0.030	0.001
Myasthenia Gravis/Myoneural Disorders, Inflammatory and Toxic Neuropathy (HCC 75)	14.7	0.001	0.004	28.5	<0.001	0.003
Opportunistic Infections (HCC 6)	13.8	0.001	0.002	17.3	<0.001	0.002
Major Organ Transplant or Replacement Status (HCC 186)	15.2	0.001	0.004	33.6	<0.001	0.004
Paraplegia (HCC 71)	9.8	0.007	0.002	10.3	0.006	0.001
Parkinson s and Huntington s Diseases (HCC 78)	9.4	0.009	0.001	13.0	0.002	0.002
Proliferative Diabetic Retinopathy and Vitreous Hemorrhage (HCC 122)	9.0	0.011	0.002	12.1	0.002	0.002
Pneumococcal Pneumonia, Empyema, Lung Abscess (HCC 115)	21.4	<0.001	0.004	44.1	<0.001	0.005
Pressure Ulcer of Skin with Full Thickness Skin Loss (HCC 158)	16.1	<0.001	0.003	33.7	<0.001	0.004
Pressure Ulcer of Skin with Necrosis Through to Muscle, Tendon, or Bone (HCC 157)	10.1	0.006	0.002	12.8	0.002	0.001
Protein-Calorie Malnutrition (HCC 21)	73.7	<0.001	0.014	69.5	<0.001	0.010
Quadriplegia (HCC 70)	11.7	0.003	0.001	12.0	0.003	0.002
Acute Renal Failure (HCC 135)	172.5	<0.001	0.038	289.9	<0.001	0.034
Respiratory Arrest (HCC 83)	10.8	0.004	0.001	7.3	0.026	0.001
Respirator Dependence/Tracheostomy Status (HCC 82)	31.0	<0.001	0.004	32.4	<0.001	0.003
Rheumatoid Arthritis and Inflammatory Connective Tissue Disease (HCC 40)	9.4	0.009	0.002	28.5	<0.001	0.003
Schizophrenia (HCC 57)	17.3	<0.001	0.004	31.0	<0.001	0.004
Seizure Disorders and Convulsions (HCC 79)	42.0	<0.001	0.014	85.8	<0.001	0.012
Septicemia, Sepsis, Systemic Inflammatory Response Syndrome/Shock (HCC 2)	80.6	<0.001	0.016	183.3	<0.001	0.016
Severe Skin Burn or Condition (HCC 162)	10.3	0.006	0.001	13.7	0.001	0.001
Severe Head Injury (HCC 166)	10.0	0.007	0.001	6.7	0.035	0.001
Severe Hematological Disorders (HCC 46)	9.4	0.009	0.002	16.6	<0.001	0.002

Table 4A.4 (cont'd)

Spinal Cord Disorders/Injuries (HCC 72)	13.2	0.001	0.004	16.1	<0.001	0.002
Ischemic or Unspecified Stroke (HCC 100)	33.1	<0.001	0.009	40.9	<0.001	0.005
Traumatic Amputations and Complications (HCC 173)	15.4	<0.001	0.003	14.8	0.001	0.002
Unstable Angina and Other Acute Ischemic Heart Disease (HCC 87)	65.6	<0.001	0.015	102.7	<0.001	0.012
Vascular Disease (HCC 108)	76.4	<0.001	0.025	207.3	<0.001	0.027
Vascular Disease with Complications (HCC 107)	29.7	<0.001	0.009	44.1	<0.001	0.007
Vertebral Fractures without Spinal Cord Injury (HCC 169)	22.1	<0.001	0.003	34.8	<0.001	0.004

Table 4A.5: Sign test *P*-value of the pooled predictive accuracy of 30-day and 1-year readmission of LASSO, XGBoost and ANN ML methods using MiSP registry data.

Outcome/method	30-day readmission			1-year readmission		
	LASSO	XGBoost	ANN	LASSO	XGBoost	ANN
30-day readmission	LASSO	-	0.473			
	XGBoost	0.473	-			
	ANN	0.720	0.281			
1-year readmission	LASSO			-	0.071	0.720
	XGBoost			0.071	-	0.720
	ANN			0.720	0.720	-

CHAPTER 5: MANUSCRIPT 3 – THE COMPARATIVE EFFECTIVENESS OF INPATIENT REHABILITATION FACILITY VERSUS SKILLED NURSING FACILITY IN PATIENTS DISCHARGED FOLLOWING ACUTE STROKE IN A MICHIGAN COHORT

5.1 Abstract

Background and objectives: Early post stroke rehabilitation therapy is used to improve functional outcomes, but to maximize its effect patients need to be discharged to an appropriate rehabilitation setting that optimizes the likelihood of achieving their highest level of functional recovery. However, the clinical decision to discharge a patient to a particular rehabilitation setting is complex, especially when it comes to choosing between care at an inpatient rehabilitation facility (IRF) versus a skilled nursing facility (SNF). The aim of this paper is to examine the comparative effectiveness of IRFs versus SNFs on functional recovery among acute stroke patients using home time as the primary outcome measure among fee-for-service (FFS) Medicare beneficiaries.

Methods: We probabilistically linked data from acute stroke patients discharged from 31 hospitals participating in the Michigan Acute Stroke registry between 2016-2020 to the Michigan Value Collaborative - a multipayer claims database. We restricted data to Medicare FFS beneficiaries. Claims data was used to identify admission to IRF and SNF following hospital discharge for stroke and to track utilization and mortality up to 1-year post discharge. Our primary outcome was home time (number of days alive and outside of inpatient care) within 90-days and 1-year following discharge from the acute stroke hospital setting. We quantified the comparative effectiveness of IRF vs SNF by reporting the crude and inverse probability of treatment weighted (IPTW) mean differences in home time (with 95% confidence intervals). IPTW were estimated using a multivariable logit model that included 35 patient- and hospital-level factors. As secondary outcomes we reported 90-day and 1-year all-cause mortality rate, and

restricted mean survival time following discharge from the acute setting. In addition, we conducted sensitivity analysis to assess the effect of the length of stay in the initial rehabilitation setting and mortality on home time.

Results: We identified 5,943 Medicare FFS beneficiaries discharged alive from the acute setting to either an IRF (n= 2,995, 50.4%) or SNF (n= 2,948, 49.6%). Compared to SNF patients, IRF patients were younger, had shorter acute hospital length of stay, were less likely to be female, had severe stroke (NIHSS >20), and were more likely to be ambulatory at discharge. Patients discharged to IRF also had lower prevalence of atrial fibrillation, heart failure, and previous stroke but were more likely to be smokers. The mean unadjusted 90-day home time for IRF and SNF patients was 57.6 and 42.0 days, respectively, while for 1 year it was 287.7 and 220.1 days, respectively. After IPTW adjustment, mean home time for IRF patients was 11.1 days (95% CI: 9.5 – 12.57) longer at 90-days, and 46.3 days (95% CI: 39.8 – 52.9) longer at 1-year, compared to SNF patients. However, after accounting for differences in rehabilitation length of stay during the first 30-days post discharge, the mean difference in adjusted 90-day mean home time disappeared (mean 0.5 days; 95% CI: -1.1 – 2.1), however there remained a significant but smaller difference at 1-year (35.7 days; 95% CI: 29.1 – 42.2). IRF patients had a 48% and 45% lower adjusted odds of death at 90-days and 1-year post discharge, respectively. Compared to SNF patients, IRF patients also had a higher adjusted restricted mean survival time of 4.3 days over 90-days and 32.4 days over 1-year of follow up. After excluding patients who died within 90-days and 1-year, the mean difference in adjusted 90-day and 1-year home time decreased from 11.1 days to 9.7 days (95% CI: 8.1 – 11.3) and from 46.3 days to 23.2 days (19.0 – 27.4), respectively.

Conclusions: We quantified the comparative effectiveness of IRF and SNF rehabilitation care settings using home time. However, our findings suggest that home time might not be a valid proxy measure of functional improvement because it is heavily impacted by rehabilitation length of stay. Nevertheless, this approach has the potential to deliver stronger evidence needed to conduct future studies that utilize more stable functional outcome measures.

5.2 Introduction

5.2.1 Post Acute Rehabilitation Settings

In the US, nearly 800,000 patients are diagnosed with new or recurrent stroke every year.¹ More than 90% of stroke survivors live with daily functional limitations, 50% of which are related to mobility impairments.^{2,3} To promote recovery following stroke approximately two thirds of stroke survivors receive post-acute care after hospital discharge that typically involves rehabilitation care.^{4,5} The post discharge rehabilitation services are delivered either at home using home health (HH) services, at an outpatient rehabilitation setting, or at designated rehabilitation facilities including inpatient rehabilitation facilities (IRF) and skilled nursing facilities (SNF).⁶ Across the US in 2018 19% of Medicare FFS stroke patients were discharged to IRF, 25% to SNF, and another 12% receive HH care.⁷ Nationally representative Get With The Guideline – Stroke (GWTG-S) registry data reported similar frequencies with 25.4%, 19.5%, and 11.5% discharged to IRF, SNF, and HH post-acute care services, respectively.⁵

Despite the structural and clinical differences between IRF and SNF, they are often directly compared due to the extensive overlap in the clinical populations served by the two types of facilities.^{8,9} IRFs provide intensive, interdisciplinary rehabilitation care under the direct supervision of a physician,¹⁰ whereas SNFs provide less intensive rehabilitation care (also known as subacute rehabilitation) to stroke survivors.^{4,10} The 2016 American Heart Association stroke rehabilitation guidelines specified that post-acute care discharge decisions should be based on the patient's expected degree of functional recovery and medical needs.² According to Medicare regulations, indications for discharge to an IRF include the ability to tolerate three hours of therapy a day for at least 5 days per week, and the expectation of significant improvement over reasonable period of time (e.g., 2 weeks) with eventual return to the community.^{4,11} SNFs are

indicated for patients for whom only partial improvements are expected over a 3-5 weeks, and where more intense therapy is unlikely to be tolerated.^{4, 11} Unlike SNF admissions, IRF admissions do not require a prior 3-day hospital stay and can take place directly from the community.¹² In the event there is a break in skilled nursing care (e.g. readmission to hospital), a patient is eligible to return to a SNF and continue their rehabilitation care if their break period is less than 30 days.¹³

5.2.2 The Decision to Be Discharged to IRF or SNF

Ideally, patients should be discharged to the post-acute care setting that maximizes their likelihood of functional recovery, but the lack of clear clinical guidance and validated clinical tools makes the clinical decision to discharge a given patient to one of these facilities (i.e., IRF or SNF) complex and somewhat subjective.^{9, 14} This is evident in the myriad of clinical factors that are associated with discharge destination decisions. Studies show that patient level factors including high number of comorbidities, premorbid dementia, the availability of family support, and patient or family preference play a major role in allocation to either an IRF or SNF.^{14, 15} These factors result in a constellation of differences in patient characteristics discharged to rehabilitation care following stroke; when compared to patients who are discharged to SNF, IRF patients are more likely to be white, male, younger, insured, have a shorter acute hospital stay, have milder stroke etiology, fewer comorbidities, and higher ambulatory scores on admission to and discharge from an IRF.¹⁶⁻²⁰

Various hospital- and system-level factors also influence the decision to be discharged to an IRF or SNF including geographic (regional) location, presence of a facility closely affiliated with the acute hospital, bed or facility availability, insurance coverage (e.g., in network vs out-of-network facility, number of treatment days covered), hospital for-profit status, and teaching

status.^{14, 15, 21-24} These factors are associated with large differences in access to rehabilitation care as evident in the wide state-to-state variation in utilization of IRF (e.g. in Arizona 19% of stroke patients were discharged to an IRF compared to only 4% in Florida).⁸

What is also contributing to the complexity of clinical decision making for IRF and SNF is the considerable uncertainty regarding the comparative effectiveness of the two settings on the long-term functional recovery for individual stroke patients.^{20, 21} Although there is a general consensus that discharge to IRF is associated with better functional outcomes compared to SNF,^{4, 9, 25} all of the comparative studies conducted to date in the US (total of 11) have relied on observational designs,^{10, 16, 26-32} utilized data from large administrative (claims) databases (i.e., Medicare FFS^{10, 16, 27, 29, 30, 32, 33} or Veteran Affairs²⁶) or medical records from large healthcare systems (e.g., in California^{25, 28}), and hospitals from several US cities³¹. Statistically adjusted findings from these studies concluded that compared to SNF, poststroke rehabilitation in IRF settings resulted in greater improvements in functional outcomes whether measured by the functional independent measures (FIM),^{10, 32} activity of daily living (ADL),³¹ mobility and self-care,¹⁶ Activity Measure for Post Acute Care (AM-PAC)²⁵, home time,²⁹ successful community discharge (home for >30 consecutive days),³³ or survival.^{28, 30, 32}

Of the comparative studies that utilized claims data from Medicare FFS population (7 out of 11 total), only one linked their data to a stroke registry database (i.e., GWTG-S) to obtain comprehensive clinical data on the acute index hospitalization.²⁹ A major limitation of all of these studies is the lack of data on long term functional recovery following discharge from rehabilitation care.⁴ Obtaining functional recovery metrics such as modified Rankin Scale (mRS), activities of daily living (ADL), and the Barthel Index (BI) relies on individual patient follow-up that is costly and hard to achieve for large population based studies³⁴

5.2.3 Home Time: A Valid Functional Recovery Metric

A practical alternative approach to quantifying functional recovery when using administrative (claims) data is the calculation of home time. Home time is defined as the amount of time post discharge spent alive and out of an inpatient care setting which includes acute hospital admissions, IRF, SNF, and long-term care hospital.²⁹ Home time has been validated as a metric of stroke functional recovery by 2 clinical trials (one conducted in 24 countries and the other in the UK^{35, 36}), 1 single center cohort study from Canada,³⁷ and 4 nationally representative cohort studies (1 in the US,³⁸ 2 in Canada,³⁹ 3 in the UK,⁴⁰ and 1 in Australia⁴¹). Results from this diverse set of studies have found that greater home time is significantly and positively associated with changes in mRS,³⁵⁻³⁹ BI,³⁶ Functional Independence Measure (FIM),⁴¹ Six Simple Variable (SSV)⁴⁰ scores over 90-days post discharge, and with changes in mRS scores over 1-year post discharge.³⁸ In the US, the variation in mean home time measured at the hospital level was examined by O'Brien and colleagues in linked GWTG-S and Medicare FFS data.⁴² O'Brien et. al. reported that higher annual ischemic stroke admission volume and rural hospital location (versus urban) were associated with higher adjusted home time post stroke. Home time was also used by Bettger et. al., in 2019 to examine the comparative effectiveness of IRF versus SNF on stroke outcomes measured at 90-days and 1 year post discharge using linked GWTG-S and Medicare FFS data.²⁹ This study showed that IRF patients had higher adjusted 90-day and 1-year home time compared to SNF patients.²⁹

5.2.4 Aims

Our aims are to utilize the MiSP-MVC linked dataset to examine the comparative effectiveness of IRFs versus SNFs on functional recovery among Medicare FFS stroke patients using 90-day and 1-year home time as the primary outcome measure. We will use inverse

probability of treatment selection weights - a propensity score method to account for confounding. We hypothesize that patients discharged to IRFs would have higher mean home time than those discharged to SNFs. As a secondary outcome we report on 90-day and 1-year all-cause mortality.

5.3 Methods

5.3.1 Study Data Bases

The study was based on the analysis of prospectively collected data of acute ischemic and hemorrhagic stroke discharges (ICD-10 I61-I63) between January 2016 and December 2020 collected by 31 Michigan hospitals participating in the Michigan Stroke Program (MiSP). This data was probabilistically linked to claims data provided by The Michigan Value Collaborative (MVC) database using indirect identifiers i.e., date of birth, sex, admission date, discharge date, and hospital ID. Both, MiSP and MVC datasets are deidentified and so do not contain any unique patient identifiers. In addition, data on hospital characteristics were obtained from the American Hospital Association's annual survey database which was linked to the admitting hospital unique identification number and admission year.

The MiSP is a representative statewide, hospital-based acute-stroke registry which is part of the CDC Paul Coverdell National Acute Stroke Program (PCNASP) that continuously collected data between 2016-2020 from 31 participating certified stroke hospitals in Michigan. Of the 31 accredited hospitals, 20 were primary stroke centers, 3 were thrombectomy capable stroke centers, and 8 were comprehensive stroke centers. These 31 hospitals include the majority of the 49 certified stroke centers in Michigan that represents an estimated ~64% of all stroke admissions in the state.^{43, 44} MiSP aims to track and improve stroke care and patient outcomes through the implementation of quality improvement programs.^{43, 44} MiSP identifies stroke

discharges using a clinical case definition.⁴³ For each discharge detailed clinical data are entered into the GWTG-S comprehensive case record form (CRF).⁴⁵ Stroke discharges are reported in MiSP as a standalone anonymized event and so there is no ability to link events related to the same patient, so it is not possible to distinguish stroke discharges as either index stroke events or stroke recurrences.

MVC is a comprehensive, statewide, claims-based database that includes data from 101 participating hospitals and 40 physician organizations in the state.⁴⁶ The MVC database covers 71% of Michigan's 143 hospitals.⁴⁶ MVC contains claims data for Michigan residents insured by Medicare FFS, Medicaid, and all insurance plans covered by Blue Cross Blue Shield of Michigan (BCBSM). All told, MVC data covers approximately 84% of Michigan's insured population.⁴⁶ Due to restrictions in MVC's DUA with CMS, Medicaid data was not available to be used for this study. Detailed information on MVC database can be found in Chapter 3 of this dissertation.

The American Hospital Association's annual survey is a voluntary survey that represents the most reliable, and comprehensive data about hospital facilities in the US.⁴⁷ The survey is completed annually by nearly 6,300 hospitals and more than 400 health care systems. The survey collects extensive data on a wide variety of topics including hospital organizational structure, facilities and services, utilization data, physician arrangements, staffing, and community orientation.⁴⁷

This research was approved by Michigan State University (MSU), University of Michigan (UM), and Michigan Department of Health and Human Services (MDHHS) Institutional Review Boards (IRB).

5.3.2 Data cleaning

In this research, an index stroke event was defined as patient's first-stroke discharge during the 5-year study period, and a readmission event as any subsequent hospital discharges occurring within one-year of the discharge date of the index stroke event. A stroke related discharge was identified using primary ICD-10 I61-I63 discharge codes. For each index event, all subsequent medical claims reported within the 1-year period following discharge were identified and a comprehensive cleaning process took place to remove duplicate claims submitted for the same health service. In addition, a comprehensive data cleaning process of the MiSP data took place so that it matched MVC's inclusion and exclusion criteria. After cleaning, the number of acute stroke discharges including index and recurrent events in the MiSP and MVC data were 46,330 and 30,685, respectively. All data cleaning, merging, and linkage preparations were done using SAS software v9.4 (Cary, NC). Details on the cleaning process can be found in chapter 3 of this dissertation.

5.3.3 Data Linkage and Study Population

Because the MiSP dataset is unable to distinguish between index events and recurrent stroke events, linkage with MVC must take place at the individual stroke event level. Of the 30,685 identified stroke events in MVC dataset, 28,131 events were index stroke events, and the rest are recurrent stroke events. Using date of birth, sex, admission date, discharge date, and hospital ID linkage variables probabilistic linkage was conducted between the 46,330 MiSP and 30,685 MVC acute stroke discharges. The linkage resulted in 23,918 matched pairs, 22,889 of which were identified as index strokes that represent the beginning of 1-year stroke episode of care (Figure 5.1). For patients with multiple stroke episodes of care (i.e., another acute stroke admission that occurred at least 1 year apart), only the first episode was included in the analysis.

Linkage was done using Match*Pro v2.4.1. Detailed information about the linkage methodology can be found in Chapter 3 of this dissertation.

5.3.4 Identifying IRF and SNF Discharges

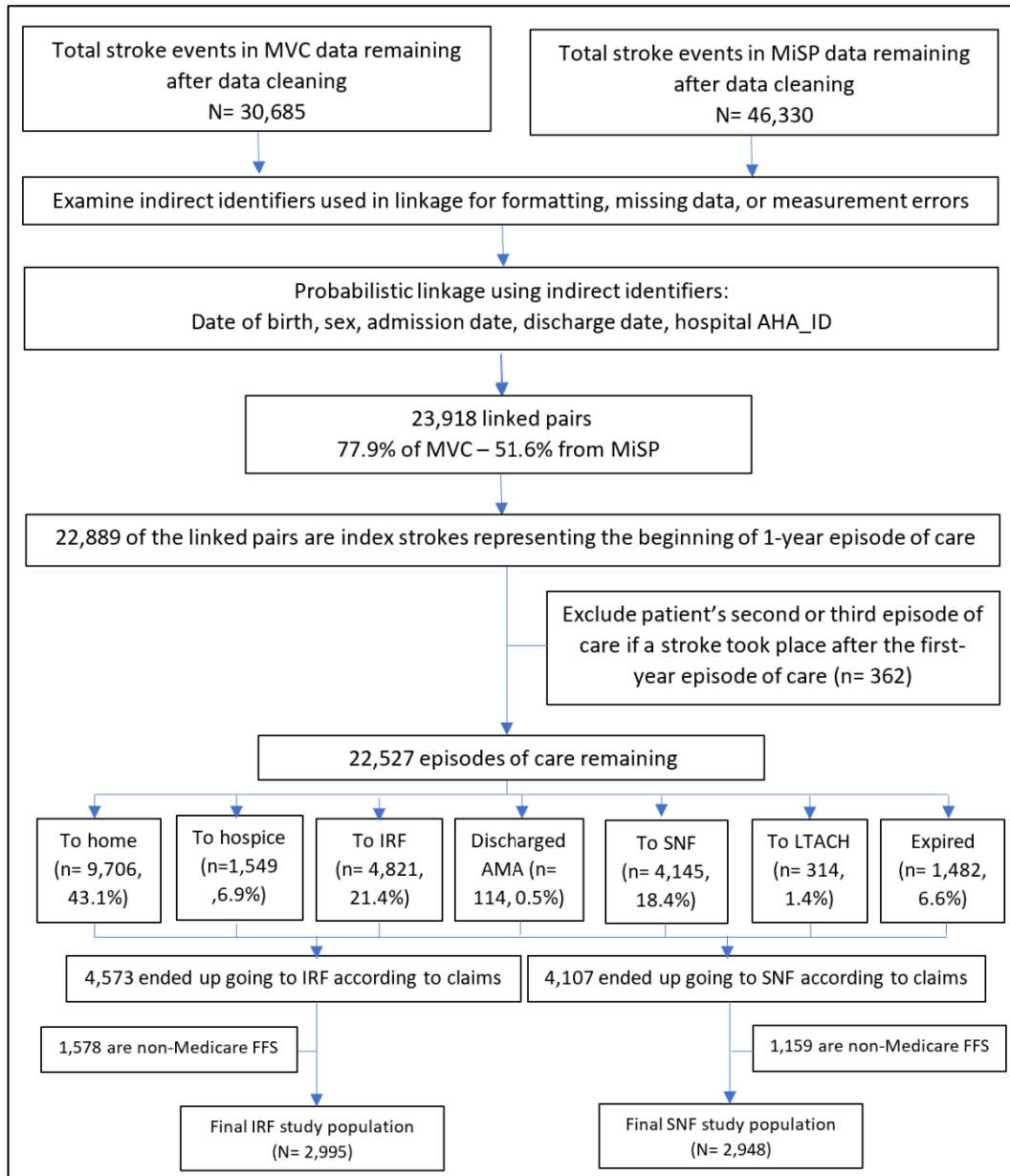
Of the 22,527 1-year stroke episodes of care, 4,821 and 4,145 had a discharge destination to IRF and SNF, respectively according to the MiSP discharge status codes (Figure 5.1). However, using the claims data, we scanned all the patients regardless of their discharge destination and found that 4,573 and 4,107 were confirmed as going directly to IRF and SNF, respectively. Of the patients who had a discharge destination to IRF (n=4,821) and SNF (n=4,145) recorded in MiSP, 91% (n= 4,385) and 89.8% (n=3,722) ended up going to IRF and SNF, respectively. Because not all patients ended up going to the designated medical discharge destination, we chose to use claims data to ascertain whether a patient actually received care in an IRF or SNF. After excluding patients who were not Medicare FFS beneficiaries, the final analysis included 2,995 and 2,948 stroke episodes of care who received IRF and SNF care, respectively (Figure 5.1).

5.3.5 Available Patient Characteristics and Techniques Used to Deal with Missingness

Available MiSP variables included data on demographics (age, sex, race/ethnicity), clinical stroke presentation (e.g., mode of transportation, last time known well, pre-stroke disability (e.g., pre stroke and current ambulatory status), stroke severity (e.g., NIHSS), clinical procedures including tPA and EVT, brain imaging (MRI, CT), more than 20 medical comorbidities (patient medical history), in-hospital complications (i.e., pneumonia, DVT, PE, UTI), length of stay, discharge medications and discharge destination. Nearly all of these variables suffered from missingness. For the variables involved in the analysis process, we

decided to include missing observations as its own category because missing observations can be medically meaningful. We also reassigned the missing values of some variables to no or absent

Figure 5.1: Probabilistic linkage between MiSP and MVC and selection of final analytical sample.



category through medical reasoning or by using value of other reported variables in a process called documentation by exception. Data recoding was done using SAS software v9.4 (Cary, NC).

5.3.6 Selection of Patient and Hospital Characteristics for the Analysis

To serve the purpose of this study, we included all available variables that were prognostic of the outcome (i.e., home time) regardless of whether they were also associated with the rehabilitation assignment to IRF or SNF.^{48, 49} Based on clinical prognostic relevance to the main outcome (i.e. home time),³⁵⁻⁴¹ data availability (missingness), and prior GWTG-Stroke home time publications,^{29, 38, 42} 29 variables were selected to be included in the analysis including age, sex, race, ethnicity, 21 stroke related comorbidities, stroke type, NIHSS upon admission, ambulatory status upon discharge, and duration of hospital stay. Only 5 of the 29 selected variables did not suffer from any data missingness: these included age, sex, ethnicity, stroke type, and admission duration. Frequency of missingness of the included variables is reported in the results section. Only 3 variables were recorded as continuous variables (age, length of stay, and admission NIHSS), of which admission NIHSS was recoded to a categorical variable using thresholds published in the literature (this variable also including a category for missing).⁵⁰ We opted to leave age and length of stay variables as continuous variables because they did not suffer from any missingness. In addition, the following six variables from the American Hospital Association database: hospital bed size, urban and rural classification, whether the hospital is part of a healthcare system, hospital stroke accreditation, hospital rehabilitation certification, and teaching status were added to the 29 patient level variables. None of these variables suffered from any data missingness.

5.3.7 Outcome Variables

The primary outcome was the mean difference in home time between IRF and SNF patients at 90-days and 1-year following discharge from the acute hospital setting. Home time is defined as the amount of time following acute stroke discharge that was spent alive and out of an

inpatient care setting. It was calculated by subtracting the cumulative number of days spent in an acute hospital, inpatient rehabilitation, skilled nursing facility, and long-term care hospital from the total number of days alive in 90-days or 365 days post discharge. As secondary outcomes we reported on 90-day and 1-year post discharge all-cause mortality (%) as well as the restricted mean survival time (RMST)- a time-dependent measure that is used to estimate the average survival time for a group during a defined time period.^{51, 52} RMST resemble the area under the Kaplan Meijer survival curve.⁵³ In our study none of the included patients were not lost to follow up (no censoring), hence RMST is simply calculated as the average of patients survival time.⁵³

We examined RMST in addition to mortality rate because the calculated effect measures from mortality rates (e.g., hazard ratio) have a built in selection bias in representing a single time point and/or in ignoring the distribution of events between the start of follow up and that time point.^{51, 52} In addition, in the case of hazard ratio, it is hard to guarantee that the proportional hazard assumption (the hazard ratio is constant over time) of the treatment effect will hold over the follow up period.⁵⁴ Therefore, a better approach that overcome previous methods limitations would be to analyze death as a time-dependent continuous variable.^{51, 52}

5.3.8 Descriptive Statistics and Study Population

To understand the selection process by which patients are discharged to IRF or SNF versus home, we first undertook a comparison between the combined IRF-SNF discharged group (N= 8,966) and those discharged home (with or without home health services) (N= 9,706) (Table 5A.1 – Appendix). Further, to understand the differences within Medicare eligible patients we made a comparison between the combined IRF-SNF population that had Medicare FFS (N= 5,943) to the combined IRF-SNF population that had MA insurance (i.e., BCBS Medicare Advantage) (N= 1,898) (Table 5A.2).

However, our primary descriptive comparison was to compare the characteristics of the IRF (n= 2,995) and SNF (n= 2,948) groups. We quantified between group differences by reporting the absolute standardized differences (ASD), where a value higher than 0.1 represents a meaningful difference (Figure 5.2).^{55, 56} The standardized difference is the method of choice to use in large datasets because unlike *p*-values ASDs are not influenced by the sample size.^{55, 56} Figure 5.2: Absolute Standardized Difference equations for continuous and categorical variables.

$$\text{Standardized difference for continuous variables} = \frac{(\bar{x}_{IRF} - \bar{x}_{SNF})}{\sqrt{\frac{s_{IRF}^2 + s_{SNF}^2}{2}}}$$

where \bar{x}_{IRF} and \bar{x}_{SNF} denotes mean of the variable in the IRF and SNF discharges, where s_{IRF}^2 and s_{SNF}^2 denote the variance of the variable in IRF and SNF discharges, respectively.

$$\text{Standardized difference for dichotomous variables}^* = \frac{(\hat{P}_{IRF} - \hat{P}_{SNF})}{\sqrt{\frac{\hat{P}_{IRF}(1 - \hat{P}_{IRF}) + \hat{P}_{SNF}(1 - \hat{P}_{SNF})}{2}}}$$

where \hat{P}_{IRF} and \hat{P}_{SNF} denotes the prevalence or mean of the dichotomous variable in the IRF and SNF discharges. Multilevel categorical variables can be calculated using multivariate Mahalanobis distance method.

We examined home time, all-cause mortality rates, and RMST by reporting the crude (unadjusted) 90-days and 1-year rates. We also generated Kaplan-Meijer survival (mortality) curves stratified by rehabilitation setting (IRF or SNF). We quantified the mean difference in home time and RMST between IRF and SNF by reporting the crude mean difference and 95% confidence interval. We intended to utilize Cox proportional hazards regression analysis to evaluate the relationship of being discharged to IRF versus SNF with mortality, but the proportional hazards assumption was violated for the entire follow up period.³² This was evident using “estat phtest” command after fitting the model with “stcox” in STATA where the proportional-hazard assumption test generated a *p*-value of <0.05. Even though we used RMST as a better alternative to hazard ratios, we also reported the odds ratios and 95% confidence

intervals generated from a logistic regression model in order to present an effect measure that is commonly used in the literature.³²

5.3.9 Propensity Score Balancing Methods

Since our study is an observational comparative effectiveness design, the non-randomized nature of rehabilitation allocation will de facto produce treatment-selection bias if IRF patients are systematically different than SNF patients with respect to known and unknown confounders.^{29, 48, 49, 57} Statistical methods based on propensity score- defined in our case as patient's probability of being allocated to IRF conditional on observed baseline covariates, are used to reduce the effect of confounding.^{49, 57} The propensity score is considered a balancing score where patient's discharged to IRF and SNF with the same propensity score should have similar distributions of observed baseline covariates.^{48, 49, 57} Thus, when we balance by adjusting, stratifying, matching, or inverse probability of treatment weighting (IPTW) using the propensity score, allocation theoretically will be independent from the potential outcomes given the baseline covariates thus allowing observational studies to mimic randomized experiments.^{49, 49, 57} IPTW uses the inverse of propensity scores to produce patient specific weights that are used in model adjustment.^{49, 57}

5.3.10 Choosing the Correct Estimand and the Associated Propensity Score Balancing Method

Before choosing our propensity score weighting methods and estimating the effect of rehabilitation setting allocation on patient recovery using home time or by calculating RMST, it is important to define the population for which the effect is being estimated.⁵⁸ This step is essential in order to correctly interpret the results, and choose the appropriate propensity score balancing method.⁵⁸ The effect of interest considering a particular population is referred to as the "estimand". For propensity score methods relevant estimands include: 1) The average treatment

effect in the population (ATE) defined as the difference in average outcomes for IRF and SNF populations if they were all allocated to IRF vs SNF, or 2) the average treatment effect in the treated (ATT) defined as the difference between the average outcomes observed for the IRF patients and the average outcomes they would have experienced had they instead been allocated to SNF. In our study we are interested in answering the question of how would home time differ on average were IRF allocation be given to all patients versus were SNF allocation was given to all patients, hence the ATE is our estimand of choice.⁵⁸ The ATE estimate is useful to generate clinical recommendations when current rehabilitation allocation practices are not well informed; the ATE can act as a proxy outcome measure to a clinical trial setting.⁵⁸ Since our estimand of choice is the ATE, the best propensity score balancing methods to use are either matching (without loss of study subjects) on the propensity score or IPTW.^{48, 58} In our study we decided to use IPTW because previous studies indicated that IPTW possess similar balancing capabilities to matching on propensity scores without the potential risk of losing a substantial number of study subjects.^{49, 57}

5.3.11 Calculating and Evaluating Propensity Scores and IPTWs

Since we determined that our estimand is ATE, IPTW weight is calculated as $w = \frac{Z}{e} + \frac{1-Z}{1-e}$, where $Z=1$ denotes allocation to IRF, $Z=0$ denotes allocation to SNF, and e is the propensity score. Each subject's weight is equal to the inverse of the probability of receiving the treatment that the subject received ($Z=1$ then $w = 1/e$, $Z=0$ then $w = 1/(1-e)$). The propensity score (e) for each subject was estimated using a logistic regression model where our binary rehabilitation allocation variable (IRF=1 vs SNF=0) was regressed on the 35 baseline variables ($0 < e < 1$) without implementation of variable selection techniques or including interaction terms. The developed model had a high discrimination power (AUC) of 0.747 (95% CI: 0.735 –

0.760) and demonstrated a good fit using the Hosmer and Lemeshow goodness-of-fit test (p -value: 0.295). The model covariate estimates from the full logit propensity score model are reported in Table 5A.3. In addition, we reported on descriptive statistics of propensity scores and IPTW weights using the mean, median, mode, standard deviation, minimum value, maximum value, lower quartile, upper quartile, interquartile range, 1st, 5th, 10th, 90th, 95th, and 99th percentiles (Table 5A.4). In addition, IPTW box plots and overlapping histograms were constructed (Figure 5A.1 and 5A.2).

Examining the distribution of the weights for the presence of extreme values (Table 5A.4, Figure 5A.1 and 5A.2) is important because the effect of patients that have a very high or very low weight due to IRF patients having very low or SNF patients having very high propensity scores, respectively on the estimated treatment effect can result in unstable effect estimates (home time and RMST).⁴⁹ To alleviate this undue influence on the variability of the estimated effect, we truncated the weights by omitting patients with extreme IPTW values defined as being above the 99th percentile.

5.3.12 Evaluating Propensity Score Balance Assumptions

To make sure that weighting has removed any observed systematic differences between the IRF and SNF groups, we performed balance diagnostic by comparing the IRF and SNF weighted characteristics and reported the weighted ASD for each model covariate.⁴⁹ The weighted ASD was calculated using the equations presented earlier (Figure 5.2) using the weighted data. We again used a threshold of < 0.1 as an acceptable difference. We reported the changes in the calculated weighted and unweighted standardized differences using a Love plot.⁵⁹ However, even if the IRF and SNF groups were balanced this does not guarantee the absence of unmeasured confounders. Also stroke severity score (NIHSS), ambulatory status at discharge,

medical history, and race variables had missing data and even though we handled this by creating a missing category, it is possible that residual confounding could result if the data are not missing at random.²⁹ Thus, a sensitivity analysis was conducted using a complete case analysis approach to assess the potential bias in the estimates related to data missingness (See section 5.3.14).

Additional assumptions of the propensity score method include positivity (each patient has a nonzero probability or no absolute contradiction of receiving IRF or SNF), consistency (each patient's potential outcome is equal to the observed outcome when the potential allocation is the same as the observed allocation), and that the propensity score model is correctly specified (e.g., interactions of baseline covariates in the logistic regression are omitted incorrectly or important confounding factors are not included).⁴⁹ However these assumptions are very hard to verify.⁴⁹

5.3.13 Calculating the Weighted Outcomes

After we ascertained that our IPTW method produced balanced IRF and SNF populations, we quantified differences of home time and RMST between IRF and SNF by reporting the weighted mean difference and 95% confidence intervals. The IPTW weighted mean difference was calculated by estimating the ATE through subtracting the mean weighted home time or RMST in the IRF group (\widehat{y}_1) from the mean weighted home time or RMST in the SNF group (\widehat{y}_0) ($\widehat{ATE} = \frac{1}{N} (\sum_i^N (IPTW * y_{1i}) - \sum_i^N (IPTW * y_{0i}))$). For mortality we also reported the weighted odds ratios and 95% confidence intervals from a logit model that included only a single term for IRF or SNF treatment. The IPTW weighted mean difference and odds ratio were carried out using the “teffects” and “strmst2” command in STATA software. Descriptive

analysis was done using SAS software v9.4 (Cary, NC), and adjusted analysis was done using STATA v18.0. Statistical significance was defined as $\alpha = 0.05$.

5.3.14 Sensitivity Analysis

We conducted four different sensitivity analyses. The first was conducted to examine the effect of the amount of cumulative time spent in the IRF or SNF during the first 30-days following discharge from the acute hospital on the calculated home time. The decision to examine the total rehabilitation admission duration over 30 day period rather than the duration of the first rehabilitation admission stems from the fact that SNF patients in particular can be discharged and then readmitted to SNF within the 30-day period post-acute discharge to continue their rehabilitation. This most often occurs when SNF patient are readmitted to the acute care setting and then return to SNF. In our data the mean length of stay of the initial IRF and SNF care settings were 14.6 (SD= 8.0) and 11.5 (SD= 8.2) days, respectively (Table 5A.5). However, the mean cumulative length of stay in the same rehabilitation setting over the first 30-day period was 15.3 (SD= 8.1) days for IRF patients and 26.4 (SD= 21.3) days for SNF patients (Table 5A.5). This change in length of stay was driven by the fact that during the first 30-days post-acute stroke discharge SNF patients get admitted to a SNF setting on average 2.1 (SD= 0.8) times compared to 1.3 (SD= 0.5) times in IRF patients (Table 5A.5). Nationally, it is well documented that the average number of rehabilitation days in a SNF is almost double that of IRF among Medicare and Medicaid beneficiaries (28 vs 16 days).¹³ Thus, this difference in length of stay has a direct effect on the home time calculation and likely impacts the calculated mean differences in home time especially over the short time horizon of 90-days post discharge. To execute this sensitivity analysis, we added the cumulative length of stay within 30-days post

discharge for both IRF or SNF admissions to the original calculated home time of each patient and recalculated the 90-day and 1-year crude and adjusted home time mean differences.

The second sensitivity analysis was conducted to assess the difference between the calculated adjusted outcomes (both home time and mortality) using the original study population (n= 2,995 (for IRF), n=2,948 (for SNF)) versus a complete case analysis approach after excluding patients with any missing data on NIHSS, ambulatory status on discharge, medical history, and race which resulted in n = 2,148 (71.7%) IRF patients and n=1,893 (64.2%) SNF patients) (Table 5A.6). This is important because we want to increase the degree of confidence that our approach of including all patients did not introduce systematic errors into the propensity score model which results in increased confidence in our balance diagnostics and adjusted outcomes estimates. We reported on the comparison between the smaller IRF (n=2,148) and SNF (n= 1,893) populations by comparing patient- and hospital-level characteristics using standardized differences (Table 5A.6). Also, we compared the generated ASDs using the sensitivity analysis dataset with ASDs generated by the study dataset. Further, we reported on the crude and IPTW adjusted outcomes using the same methods described earlier and compared them to the estimates produced by the study population (Table 5A.7).

We conducted a third sensitivity analysis to assess the degree of bias exerted by patients who expired during the first rehabilitation admission (5 in IRF and 41 in SNF) on home time. Some comparative effectiveness studies comparing IRF and SNF patients exclude patients who die during rehabilitation as these patients did not complete their full term of rehabilitation.^{10, 25, 30} Thus, this analysis was done on 2,990 IRF vs 2,907 SNF patients who were discharged alive from rehabilitation setting and we reported the crude and IPTW adjusted home time (Table 5A.8).

Finally, the fourth sensitivity analysis was conducted to determine how much of the difference in home time estimates are due to mortality after acute stroke discharge. It is well documented that IRF patients are healthier compared to SNF patients, thus SNF patients are more likely to experience death post discharge.¹⁶⁻²⁰ To execute this sensitivity analysis, we developed survivor only cohorts by dropping patients who experienced death within 90-days (208 in IRF and 504 in SNF) and 1-year (504 in IRF and 1,040 in SNF) post discharge and recalculated the 90-day and 1-year crude and adjusted home time mean differences only among the survivors who lived to 90 days (2,787 in IRF and 2,444 in SNF) and 1 year (2,491 in IRF and 1,908 in SNF).

5.4 Results

5.4.1 Descriptive Statistics comparing Patients Discharged Home to Those Discharge to IRF or SNF

When comparing patients discharged home with and without home health services (N= 9,706) to those discharged to an IRF or SNF (N= 8,966), patients discharged home were younger, less likely to be female, more likely to have ischemic stroke, minor stroke (NIHSS <5), and to be able to ambulate independently at discharge (Table 5A.1). In addition, patients discharged home had shorter hospital stays at index admission, were less likely to have atrial fibrillation, hypertension, heart failure, and previous stroke but had a higher prevalence of migraines and were more likely to be smokers. No differences were observed in hospital characteristics between the two groups. Out of the 35 examined characteristics, 12 had meaningful differences (ASD > 0.1).

5.4.2 Descriptive Statistics Comparing IRF and SNF Patients Insured by Medicare FFS and MA

Among the combined IRF-SNF discharge group, there were only minor differences between Medicare FFS (N= 5,943) and MA (i.e., BCBS Medicare Advantage) (N= 1,898) insured beneficiaries. Only 4 out of the 35 examined characteristics had meaningful differences (ASD > 0.1) (Table 5A.2); the FFS group were younger, were more likely to be female, have a shorter stay at index admission, and were more likely to be smokers. This comparison is important because we were unable to include MA patients due to absence of mortality data and thus the minor differences between FFS and MA population indicate that our findings could possibly be extrapolated to the MA population.

5.4.3 Descriptive Statistics Comparing IRF and SNF Medicare FFS Patients (The Study Sample)

Our study sample included 5,943 Medicare FFS beneficiaries discharged to either an IRF (50%, n= 2,995) or SNF (50%, n= 2,948) (Figure 5.1). Patients discharged to IRF were younger, less likely to be female, less likely to have major strokes (NIHSS >20) and more likely to have minor strokes (NIHSS 1-4), and be able to ambulate with assistance at discharge (ASD > 0.1) (Table 5.1). In addition, patients discharged from IRF had shorter hospital stays at index admission, were less likely to have atrial fibrillation, heart failure, and previous stroke but were more likely to be smokers. Furthermore, patients discharged to IRF were less likely to be discharged from thrombectomy capable stroke centers, but more likely to be discharged from hospitals with 300 or more beds and from hospitals located in rural areas. Out of the 35 examined characteristics, 12 had meaningful differences (ASD > 0.1).

Table 5.1: Descriptive statistics of inpatient rehabilitation facility (IRF) and skilled nursing facility (SNF) populations characteristics and their corresponding absolute standardized difference.

	Variable	IRF (N= 2,995)	SNF (N= 2,948)	Absolute standardized difference
Demographics				
Age*	Mean (SD)	75.5 (10.6)	80.0 (10.5)	0.46
Sex*	Female	54.4	62.9	0.18
	Male	45.6	37.1	
Race	White	79.1	78.6	0.01
	Black	15.7	15.5	0.01
	Other	1.1	1.2	<0.01
	Missing	4.0	4.7	0.03
	Latino ethnicity*	No	97.1	96.7
	Yes	2.9	3.2	
Characteristics of stroke hospitalization				
Stroke Type*	Hemorrhagic	13.2	13.5	0.01
	Ischemic	86.8	86.5	
Admission NIHSS category	0	7.9	6.9	0.04
	1-4	38.1	29.7	0.18
	5-15	37.5	35.2	0.05
	16-20	5.4	7.7	0.09
	>20	4.4	6.8	0.11
	Missing	6.8	13.7	0.23
Admission duration*	Mean (SD)	5.1 (4.0)	6.5 (4.9)	0.34
Ambulatory status on discharge	Able to ambulate independently (no help from another person) w/ or w/o device	21.1	18.6	0.06
	With assistance (from person)	51.6	39.3	0.25
	Unable to ambulate	6.7	18.5	0.36
	Missing	20.5	23.5	0.07
Past medical history				
Missing medical history*	Yes	4.6	4.4	0.01
	No	95.4	95.6	
Atrial fibrillation/flutter	Yes	23.5	32.2	0.19
	No	71.9	63.4	
Prosthetic heart valve	Yes	1.2	1.8	0.05
	No	94.1	93.7	
Coronary artery disease/ prior myocardial infarction	Yes	27.4	28.8	0.03
	No	68.0	66.7	
Carotid stenosis	Yes	4.9	4.8	0.01
	No	90.5	90.8	
Diabetes mellitus	Yes	33.6	35.7	0.04
	No	61.8	59.9	
Peripheral vascular disease	Yes	6.4	7.8	0.06
	No	89.0	87.7	
Hypertension	Yes	75.5	77.3	0.02
	No	18.9	18.3	

Table 5.1 (cont'd)

Smoking	Yes	17.3	12.4	0.14
	No	78.1	83.2	
Dyslipidemia	Yes	53.8	54.3	0.01
	No	41.6	41.3	
Heart failure	Yes	11.2	17.1	0.17
	No	84.2	78.4	
Previous stroke	Yes	22.7	29.7	0.16
	No	72.7	65.9	
Previous Transient ischemic attack	Yes	9.3	11.5	0.07
	No	86.1	84.1	
Drug/alcohol abuse	Yes	5.9	4.5	0.06
	No	89.5	91.0	
Family history of stroke	Yes	10.4	8.3	0.07
	No	84.9	87.3	
Migraine	Yes	0.6	0.4	0.05
	No	84.7	95.1	
Obesity overweight	Yes	2.6	2.0	0.09
	No	92.7	93.6	
Chronic renal insufficiency	Yes	42.8	38.3	0.10
	No	52.5	57.3	
Sleep apnea	Yes	13.8	17.2	0.02
	No	81.6	78.3	
Depression	Yes	8.0	7.5	0.07
	No	84.4	88.1	
Deep vein thrombosis/ pulmonary embolism	Yes	18.8	21.6	0.01
	No	76.6	73.9	
Dementia	Yes	0.5	1.0	0.06
	No	94.9	94.6	
Hospital Characteristics				
Bed Size*	50-99	0.7	0.6	0.02
	100-199	6.0	9.4	0.13
	200-299	10.5	16.1	0.17
	300-399	27.0	21.6	0.13
	400-499	19.3	17.1	0.06
	>=500	36.6	35.3	0.03
Core-based statistical area*	Metro	93.5	94.5	0.04
	Micro	3.2	4.2	0.05
	Rural	3.2	1.3	0.13
System member*^	Yes	84.8	86.6	0.05
	No	15.2	13.4	
Stroke accreditation*	CSC: Comprehensive Stroke Center	45.7	45.7	<0.01
	PSC: Primary Stroke Center	46.3	42.5	0.08
	TSR: Thrombectomy Capable Stroke Center	8.0	11.8	0.13
Stroke rehabilitation accreditation*^	Yes	8.8	9.9	0.07
	No	91.2	90.1	

Table 5.1 (cont'd)

Teaching hospital (Medical school affiliation reported to American Medical Association)*	Yes	84.2	82.2	0.05
	No	15.8	17.8	

*Covariate did not have any missing values.

^System member indicates that the hospital is part of a healthcare system. Stroke rehabilitation accreditation is given by The Joint Commission.

5.4.4 Evaluation of Balance Assumption

After excluding 29 IRF and 29 SNF discharges (total of 58) because they had an extreme IPTW above the 99th percentile (Table 5A.4), the generated weighted standardized differences of the 35 characteristics were all below the 0.1 threshold indicating that conditioning on the IPTW truncated weights (n= 5,885) removed the observed systematic differences between the IRF and SNF groups (Figure 5.3).

Figure 5.3: Love plot of the absolute standardized differences in unweighted (n= 5,943) and weighted (n= 5,885) data of patient- and hospital-level characteristics.*



*58 discharges were deleted because they had an extreme IPTW above 99th percentile.

5.4.5 Unadjusted and Adjusted Weighted Outcomes

The observed unadjusted 90-day and 1-year mean home time was 15.6 (95% CI: 14.2 - 17.1) and 67.6 (95% CI: 61.5 – 73.7) days higher among IRF patients compared to SNF patients, respectively. After adjusting using IPTW, IRF patients were found to have 11.1 (95% CI: 9.5 – 12.7) and 46.3 (95% CI: 39.8 – 52.9) days higher home time at 90-days and 1-year, respectively (Table 5.2).

Table 5.2: Unadjusted and IPTW adjusted differences in home time, mortality, and restricted mean survival time (RMST) between patients discharged to inpatient rehabilitation facility (IRF) or skilled nursing facility (SNF).

Outcome, time point	IRF (N= 2,995)	SNF (N= 2,948)	Unadjusted effect measure (N= 5,943)		IPTW adjusted effect measure (N= 5,885)*	
	Mean (SD) or N (%)		Mean difference or odds ratio (95% CI)	<i>p</i> -value**	Mean difference or odds ratio (95% CI)	<i>p</i> -value**
Home Time (days)						
90-day	57.6 (27.8)	42.0 (29.5)	15.6 (14.2 -17.1)	<0.001	11.1 (9.5 – 12.7)	<0.001
1-year	287.7 (104.2)	220.1 (133.1)	67.6 (61.5 – 73.7)	<0.001	46.3 (39.8 – 52.9)	<0.001
Mortality rate						
90-day	208 (6.9%)	504 (17.1%)	0.36 (0.30 – 0.42)^	<0.001	0.52 (0.42 – 0.62)^	<0.001
1-year	504 (16.8%)	1,040 (35.3%)	0.37 (0.33 – 0.42)^	<0.001	0.55 (0.48 – 0.62)^	<0.001
Restricted mean survival time (RMST) (days)						
90-day	86.8 (13.3)	80.9 (22.2)	5.9 (4.9 – 6.8)	<0.001	4.3 (3.6 – 4.9)	<0.001
1-year	327.9 (93.0)	279.7 (130.8)	48.2 (42.5 – 54.0)	<0.001	32.4 (28.3 – 36.6)	<0.001

*58 discharges were deleted because they had an extreme IPTW above 99th percentile.

**Independent t-test or X² test.

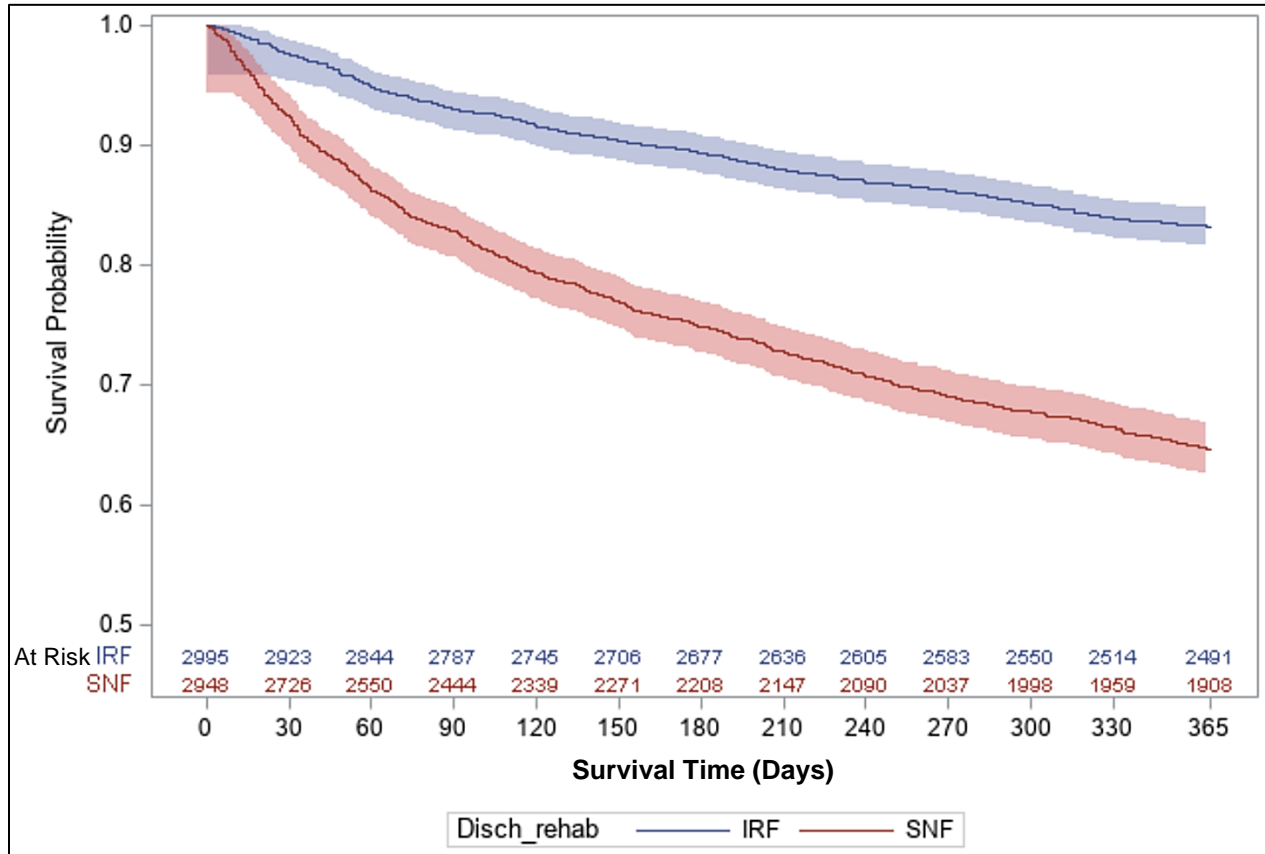
^Odds ratio and 95% confidence interval.

Compared to SNF patients, a much lower proportion of IRF patients died within 90-days (6.9% vs 17.1%) and 1-year (16.8% vs 35.3%) following acute stroke discharge (Table 5.2, Figure 5.4). IRF patients had a 64% and 63% lower unadjusted 90-day and 1-year odds of death, respectively. After adjusting using IPTW, IRF patients were found to have 48% and 45% lower adjusted 90-day and 1-year odds of death, respectively (Table 5.2).

The observed unadjusted 90-day and 1-year RMST differences were 5.9 (95% CI: 4.9 – 6.8) and 48.2 (95% CI: 42.5 – 54.0) days higher among the IRF patients compared to SNF patients, respectively. After adjusting using IPTW, IRF patients were found to have 4.3 (95% CI:

3.6 – 4.9) and 32.4 (95% CI: 28.3 – 36.6) higher RMST difference in 90-days and 1-year, respectively (Table 5.2).

Figure 5.4: Kaplan-Meier survival curve over 1-year follow up with 95% confidence intervals stratified by initial rehabilitation setting.*



* At risk defines the number of IRF and SNF population that did not experience death at a certain follow up time.

5.4.6 Sensitivity Analysis Outcomes

5.4.6.1 Sensitivity Analysis #1 - Accounting for Rehabilitation Facility Length of Stay

The first sensitivity analysis that amended the home time calculation by accounting for the cumulative number of days spent in IRF or SNF during the 30-day post discharge period, resulted in very different results especially for the 90-day follow up period. Compared to the originally calculated 90-days adjusted home time mean difference of 11.1 days (95% CI: 9.5 – 12.7), the adjusted 90-days amended home time mean difference was almost zero and not significantly different (0.5 days, 95% CI: -1.1 – +2.1) (Table 5.3). This indicates that 30-days

IRF and SNF rehabilitation admission duration was responsible for the vast majority of the original 11 days mean difference in 90-day home time. Similar effects were observed in the 1-year home time calculations; the amended mean difference was approximately 11 days lower than the originally calculated home time mean difference (i.e., 35.7 vs 46.3 days) (Table 5.3). In our data IRF and SNF patients had a mean 30-day post discharge rehabilitation length of stay of 15.3 (SD: 8.1) and 26.4 (SD:21.3) days, respectively and 12.6% and 19.6% were readmitted to the acute care hospital at least once during the 30-day period post discharge, respectively (Table 5A.5).

Table 5.3: Unadjusted and IPTW adjusted differences in amended home time between inpatient rehabilitation facility (IRF) and skilled nursing facility (SNF).*

Outcome, time point	IRF (N= 2,995)	SNF (N= 2,948)	Unadjusted effect measure (N= 5,943)		IPTW adjusted effect measure (N= 5,885)**	
	Mean (SD)		Mean difference (95% CI)	p-value [^]	Mean difference (95% CI)	p-value [^]
Home Time amended for 30-days post discharge rehabilitation admission duration**						
90-day	72.9 (26.0)	68.3 (30.2)	4.6 (3.1 – 6.0)	<0.001	0.5 (-1.1 – 2.1)	0.55
1-year	303.0 (103.6)	246.5 (133.5)	56.5 (50.4 – 62.6)	<0.001	35.7 (29.1 – 42.2)	<0.001

*Amended home time calculation was done by adding the number of days a patient spent in IRF or SNF rehabilitation 30-days post discharge. Mean 30-days IRF and SNF admission duration were 15.3 (8.1) and 26.6 (21.3) days, respectively.

**58 discharges were deleted because they had an extreme IPTW above 99th percentile.

[^]Independent t-test.

5.4.6.2 Sensitivity Analysis #2 - Complete Case Analysis

Comparing the IRF (n=2,148) and SNF (n= 1,893) groups in the complete case analysis cohort (n =4,041) where discharges with missing values in NIHSS, ambulatory status, medical history, and race were excluded, resulted in meaningful differences in 14 out of 35 characteristics (ASD > 0.1). This finding is similar to the original study dataset (n =5,943) where 12 out of 35 characteristics were different. The two additional variables that were different included prevalence of overweight or obesity, and likelihood of being discharged from a system member hospital (Table 5A.6).

Outcomes calculated using the complete case analysis dataset, were similar to the outcomes calculated using the study dataset (Table 5A.7). Compared to the study dataset the complete case analysis dataset yielded slightly higher adjusted 90-days home time mean difference (11.6 vs 11.1), and lower adjusted 1-years home time mean difference (44.4 vs 46.3) (Table 5.2 and 5A.7). In addition, compared to the original dataset, the complete case analysis dataset reported lower adjusted 90-day odds of death (42% vs 48%), and lower adjusted 1-year odd of death (41% vs 45%). Further, the complete case analysis dataset had a lower adjusted 90-days (3.9 vs 4.3) and 1-year (30.5 vs 32.4) RMST differences compared to the full dataset (Table 5.2 and 5A.7).

5.4.6.3 Sensitivity Analysis #3 – Accounting for Deaths That Occurred During IRF or SNF Admission

There were 5 (0.2%) and 41 (1.4%) patients who died during the initial IRF or SNF admission. After dropping these subjects from the home time calculation, the observed unadjusted 90-day and 1-year mean home time were slightly higher compared to the study dataset (15.6 vs 15.2) and (67.6 vs 65.1), respectively. This was also the case for the adjusted 90-days and 1-year home time estimates (11.1 vs 10.9) and (46.3 vs 45.3), respectively (Table 5.2 and 5A.8). These results are within expectations since excluded patients who died in rehabilitation would otherwise have a home time value of zero.

5.4.6.4 Sensitivity Analysis #4 – Accounting for Mortality Throughout the Follow Up Period

Lastly, we observed that the mortality rate for IRF and SNF patients were 6.9% and 17.1% at 90-days and 16.8% and 35.3% at 1-year post-acute stroke discharge, respectively. Comparing the IRF and SNF weighted home time mean differences in the original study population and the population that survived to either 90-days or at 1-year of follow up, we found that mortality was responsible for a modest decrease of 12.6% in 90-days adjusted home time

mean difference (11.1 vs 9.7), but a large decrease of 50% in 1-year adjusted home time mean difference (46.3 vs 23.2) (Table 5.2 and 5.4). These findings indicate that even though SNF patients experienced a substantially higher mortality rate throughout the 1 year follow up period compared to IRF patients, during the first 90-days most of the difference is due to the higher time spent in SNF rehabilitation compared to IRF rehabilitation and not mortality. But over the 1 year follow up, half of the home time difference was solely due to mortality- a measure that is not directly influenced by rehabilitation setting, which indicate that much of the home time advantage of IRF is due to the bias from the unmeasured selection of patients who are sicker to SNF.

Table 5.4: Unadjusted and IPTW adjusted mean differences of home time in inpatient rehabilitation facility (IRF) compared with skilled nursing facility (SNF) among patients who remain alive at 90-days and 1-year post discharge.

Outcome	90-days home time		1-year home time	
	IRF (n= 2,787)	SNF (n=2,444)	IRF (n=2,491)	SNF (n=1,908)
Mean (SD)	60.7 (26.0)	48.1 (28.0)	326.6 (47.5)	299.3 (74.4)
Unadjusted mean difference (95% CI)	N= 5,231 12.6 (11.2 – 14.1)		N= 4,399 27.3 (23.5 – 31.2)	
Adjusted mean difference (95% CI)	N= 5,178^ 9.7 (8.1 – 11.3)		N= 4,354^ 23.2 (19.0 – 27.4)	

^Discharges with extreme IPTW above 99th percentile were deleted.

5.5 Discussion

5.5.1 Comparative Effectiveness of IRF Versus SNF on Functional Recovery Using Home Time

In this observational retrospective of Medicare FFS beneficiaries that included linked data from MiSP registry and MVC claims database, we compared the effectiveness of IRF versus SNF rehabilitation setting on patient’s functional recovery up to 1-year post stroke as defined by time spent at home. Our research was driven by the considerable overlap of the patients populations served by these two settings, and the considerable uncertainty regarding the comparative effectiveness of the two settings on the long term functional recovery for stroke patients.^{9, 14, 20, 21} In addition, there is a debate around the cost effectiveness between IRF and

SNF rehabilitation setting where stroke rehabilitation at IRFs costs approximately double that at SNFs despite the longer length of stay for SNF patients.^{8, 13}

Our study confirms the findings of previously published literature showing that patients discharged to IRF are younger, healthier, have milder stroke presentation, and have better survival compared to SNF patients.^{10, 16, 26-29, 32} Our study findings also concurs with the findings of previous literature that IRF patients have better functional outcomes compared to SNF patients over the longer 1-year term of follow up but not over a shorter (i.e., 90 days) duration of follow up.^{29, 33}

Comparative effectiveness research that compares long term (1-year) functional outcome of stroke patients following IRF versus SNF rehabilitation is limited to only two studies published in the last 10 years.^{8, 29, 33} However, neither study used an objective functional outcome measures (e.g., mRS or ADL) and instead relied on home time²⁹ or a variant of home time (% successful community discharge- home for >30 consecutive days)³³ that were derived from claims data.

Using linked GWTG-S Medicare FFS data from 2006 to 2008, Bettger et. al., reported that unadjusted mean home time among Medicare FFS beneficiaries was higher for IRF patients than SNF patients at both 90-days (51.8 (SD: 31.2) vs 32.5 (SD: 30.7)) and 1-year (271.2 (SD: 112.5) vs 195.5 (SD: 138.5)).²⁹ Compared to Bettger et. al. study, our unadjusted home time findings were a little higher. Although Bettger et. al. did not use the difference in home time as their effect measure, we used their data to estimate the mean unadjusted difference in 90-day home time (51.8 minus 32.5 = 19.3 days) and 1-year home time (271.2 minus 195.5 = 75.7 days) which are similar to our 90-day (15.6 days) and 1-year (67.6 days) estimates. The effect measure (hazard ratio) used to quantify the crude and adjusted differences in home time in Bettger et. al.

study is a questionable approach as home time should be analyzed as a continuous and not a time to event variable.⁵⁸

A second study by Simmonds et. al., that used successful community discharge- a home time derived outcome to compare IRF and SNF outcomes was conducted among a national sample of Medicare FFS acute stroke discharges between 2012 and 2013. It found that compared to SNF patients, IRF patients had a higher 90-day (68.2% vs 44.7%) and 1-year (81.4% vs 60.1%) adjusted risk of successful community discharge.³³ Simmond et. al. recorded the instance when a patient has spent more than 30 consecutive days at home and outside of inpatient setting during certain post discharge follow up period into a binary outcome (yes/no) called successful community discharge, the motivation to develop this outcome measure was because home time differences are very hard to be clinically interpreted and because this measure is not likely to be recorded until the end of rehabilitation period.³³ Although not directly comparable, the adjusted risk differences at 90-day (23.5%) and 1-year (21.3%) reported by the Simmonds study were statistically significant and are concordant with our adjusted home time findings that IRF patients have longer home time compared to SNF patients but only over longer 1-year of follow up.

5.5.2 Comparative Effectiveness of IRF Versus SNF on Mortality

Bettger et. al. also reported lower mortality among patients who were discharged from the hospital to an IRF rather than to a SNF at 90 days (7.2% vs 21.1%), and 1 year (17.9% vs 38.6%).²⁹ Simmonds et. al. study reported very similar mortality rates at 90 days (7.1% for IRF vs 21.0% for SNF), and 1 year (18.2% for IRF vs 38.8% for SNF).³³ Our mortality rates are a little lower than both the Bettger et. al. and Simmonds et. al. studies; we found that 6.9% of IRF patients and 17.1% of SNF patients died within 90 days, and 16.8% of IRF patients and 35.3% of

SNF patient died within 1 year. These small differences in unadjusted statistics could be attributable to older data used by Bettger et. al. (2006-2008) and Simmond et al. (2012-2013) and the much larger number of hospitals from across the nation (1,192 hospitals for Bettger et. al.; 891 hospitals for Simmonds et. al.) which would include geographic areas of the US with known higher stroke mortality.^{29, 33}

In addition, Bettger et. al. reported that IRF care is associated with 48% (0.52 - 95% CI: 0.49 – 0.55) and 35% lower adjusted hazard ratio (0.65 95% CI: 0.62 – 0.68) of death within 90-days and 1-year post discharge, respectively.²⁹ Simmond et. al. also reported that IRF care was associated with 46% lower adjusted odds (odds ratio 0.54 – 95% CI: 0.51-0.57) of death in 1-year post discharge (90-day odds ratios were not reported).³³ These estimates are very similar to our adjusted mortality odds ratio estimates of 0.52 (90 days) and 0.55 (1 year) that illustrated that IRF patients have substantially better survival than SNF patients up to 1-year following acute stroke discharge. As expected, these data were also replicated in the longer adjusted RMST in IRF patients compared to SNF patients both at 90 days and 1 year.

There are two other studies that compared mortality between IRF and SNF settings. A study by Wang et. al. among Kaiser Permanente health care system acute stroke discharges between 1996 and 2004 reported a lower 1-year adjusted mortality risk among patients who received IRF rehabilitation within 14-days post discharge compared to SNF patients (relative risk of 0.33 – 95% CI: 0.24 – 0.45).²⁸ Second, a study by Hong et. al. using national Medicare FFS acute stroke discharges between 2013 and 2014 reported a statistically significant adjusted lower odds of mortality 365 days post-acute stroke discharge among IRF patients compared to SNF patients (odds ratio of 0.75 – 95% CI: 0.72 – 0.77).¹⁶

Despite the fact that these prior studies had differences in hospital sites, patient inclusion criteria and used different effect measures, they all agreed as did our study on the fact that IRF patients had better survivability than SNF patients up to 1-year post-acute stroke discharge.

5.5.3 Sensitivity Analysis to Evaluate the Effect of Rehabilitation Duration of Admission

Although the validity of home time as a proxy of functional recovery has been examined by several studies ranging from secondary analyses of trials and observational cohorts based analyses either at 90 days and 1 year post discharge,³⁵⁻⁴¹ home time might not be an appropriate measure for our comparative effectiveness study. This stems from the fact that none of the validation studies of home time compared different rehabilitation settings,³⁵⁻⁴¹ and that home time is calculated from the point of acute stroke discharge and the typical rehabilitation length of stay at a SNF is about double that of IRF, so home time will de facto be lower for SNF patients. Our sensitivity analysis findings highlighted these concerns; after amending the originally calculated home time by adding the number of days a patients spent in IRF or SNF in the 30 days post-acute discharge, we found no adjusted mean difference in home time between IRF and SNF admissions within 90-days post discharge. However, 1-year mean differences remained significantly different although were attenuated from a mean of 29 to 42 days. These findings provide evidence that questions the validity of using home time as a functional recovery measure for rehabilitation studies especially over short term follow up (e.g., 90-days). The approach used by Simonds et. al. to generate a different outcome measure termed successful community discharge (achieving >30 consecutive days at home) could serve as an alternative to home time because it is not directly affected by duration of rehabilitation.

5.5.4 Sensitivity Analysis Approach to Deal with Missingness

We originally hypothesized that missingness in our data was likely informative. Our strategy to include all case observations with missing data by coding missing values as their own category produced similar weighted outcomes compared to the complete case analysis dataset. The latter excluded 1,902 cases that has missing data on key prognostic variables (i.e., NIHSS, ambulatory status, medical history, and race). In addition, univariate comparison of differences between IRF and SNF populations in the original and complete case analysis datasets were similar. This indicates that missingness in our data did not appear to occur in a systematic pattern, which increases our level of confidence in the validity of our propensity score model and adjusted outcomes estimates.

5.5.5 Sensitivity Analysis to Assess the Effect of Mortality on Home Time

We conducted two sensitivity analyses to assess the effect of mortality of home time. Because the inclusion criteria of similar comparative effectiveness studies varies such as only including patients who survive the rehabilitation period,¹⁰ patient who survive up to 30 days post discharge,³⁰ or patients who survive six months post discharge,²⁵ we conduct a sensitivity analysis that examines the effect of including or excluding patients who die during rehabilitation. Our finding suggests that patients who die during rehabilitation do not contribute a high difference in home time, but we should note that there were only a small number of cases who died at the IRF or SNF setting.

Further, because SNF patients have lower survival compared to IRF patients,^{16, 28, 29, 33} we quantified how much mortality contributed to changes in home time up to 1-year post discharge. Our findings that mortality was responsible for a modest 10% relative decrease (absolute decrease of 1.4 days) in 90-days home time mean difference, but a large 50% relative decrease of

23.1 days in 1-year home time mean difference suggests that mortality has a much bigger relative impact on home time when comparing IRF to SNF patients over the long term follow up period not on the short term. This finding highlights that much of the longer-term difference in home time between IRF and SNF patients is driven by the higher mortality in SNF patients, which indicate that much of the home time advantage of IRF is due to the bias from the unmeasured selection of patients who are sicker to SNF.

5.5.6 IPTW as a Propensity Score Weighting Method in Observational Studies

Since our study is a retrospective observational comparative effectiveness design, the non-randomized nature of rehabilitation allocation will de facto produce treatment-selection bias in which IRF patients are systematically different than SNF patients due to known and unknown confounders.^{29, 48, 49, 57} Thus, to reduce the effect of confounding when using observational data, we chose to adjust the data using IPTW propensity score method. Our study reported in detail on the process of choosing the correct weight balancing method, proper selection of study variables, determining the estimand that serves the purpose of this research, and reporting on the propensity scores, and IPTW model estimates and descriptive statistics.

In addition, our study evaluated the weighting balance assumption through reporting the weighted standardized differences in a Love plot where all of our study variables had a standardized differences well below the 0.1 threshold indicating that conditioning on IPTW has removed any observed systematic differences between the IRF and SNF groups in the available observed data.

5.5.7 Comparative Effectiveness Applicability to This Research

In this study we examined two comparable treatment settings that provide rehabilitation services for substantially similar clinical stroke patients populations.³³ We generated evidence

that support discharge to IRF care over SNF care after stroke. This study was done using retrospective observational study design where selection of treatment (IRF versus SNF) is biased due to a myriad of factors mentioned earlier in the introduction section. Propensity score adjustment methods were used to eliminate the observed selection bias using factors that are available in the stroke registry but residual confounding effects due to unmeasured factors are likely still present. Our propensity score model had a modest discrimination of 0.74 and was successful in balancing the two populations. The calculated estimand (average treatment effect in the population) is well documented to act as a proxy outcome measure to a clinical trial setting. We also conducted a sensitivity analysis to address issues related to missing data. All of the points mentioned earlier fulfill the definition of comparative effectiveness research and all of the steps required to conduct such a research using observational data.⁶⁰

5.5.8 Strengths and Limitations

This study has important strengths. To our knowledge, this is one of a few studies to compare the effectiveness of IRF versus SNF on functional recovery up to 1-year post discharge using home time. Our sensitivity analysis approach of examining the effect of rehabilitation length of stay and mortality on home time is novel and provided evidence that questions the validity of home time as a functional outcome for studies of institutional rehabilitation. We used claims data to identify patients who used IRF or SNF rather than discharge destination information. Furthermore, unlike most of the previous population-based studies that utilized administrative data to compare IRF and SNF,^{10, 16, 27, 29, 30, 32, 33} our linked registry-claims data structure allowed for including a list of prognostic variables that cover demographics, stroke presentation, discharge ambulatory status, and a comprehensive list of comorbidities, which provided more clinical information (compared to studies that rely on only claim data) that could

help reduce confounding. Finally, we reported on the process of choosing the correct weight balancing methods, selection of our study variables, determining the estimand that serves the purpose of this research, and the propensity scores and IPTW model estimates and their descriptive statistics.

This study should be interpreted in the context of the following limitations. Our study population was limited to stroke patients discharged from 31 stroke certified hospitals in Michigan, which might limit the generalizability of our study patients discharged from certified stroke centers. Due to limited availability of mortality data in the MVC claims database, our study focused on Medicare FFS beneficiaries only. Nevertheless, our study results can probably be extrapolated to BCBSM Medicare Advantage population because of the high degree of similarity between the Medicare FFS and MA populations that were discharged to IRF or SNF. Because of the observational design our data are likely prone to selection bias and residual unmeasured confounders which could reduce the validity of our comparisons. We tried to overcome the selection bias by including all the relevant prognostic variables in our analysis and by conducting a sensitivity analysis that excluded all the observations that suffered from missing data where we reported similar outcomes suggesting that missingness in our data was at random and thus were non informative. Even though we tried to reduce the risk of unmeasured residual confounding by including all of the available potential prognostic factors in our propensity score model and by conducting an evaluation of the weighted data balance assumption, the list of variables that we used did not cover patient-level important confounders related to physician decision making, hospital policies, bed availability, family and patient preferences, and social support.⁸ The large adjusted differences in mortality likely reflect these residual unmeasured

factors. Finally, due to the scarcity of studies that utilized home time as a functional outcome measure, we were only able to compare our study results with one study.

5.6 Future Directions and Conclusions

This study included data of Medicare FFS beneficiaries from 31 stroke certified hospitals in Michigan. Therefore, future research should attempt to expand data collection from additional hospitals and insurance providers. One solution to overcome the limited availability of mortality data is through conducting a data linkage with the national death index. The possibility of examining the comparative effectiveness of IRF versus SNF among the currently available BCBSM private insurance data and future available data from other insurance providers including Medicaid and other private insurers (i.e., Health Alliance Plan (Henry Ford Health System), Spectrum Health, and United Health) would be important in order to generate more generalizable outcomes, examine the disparities introduced by insurance providers on functional outcomes, and provide further evidence that support the consensus that IRF patients have better functional outcomes compared to SNF patients over long periods of follow up. Due to the uncertainty in using home time for the comparative effectiveness between IRF and SNF, future studies should avoid using home time as a proxy of functional recovery over 90 days short period of follow up and should rely on more stable measures like mRS or successful community discharge (home for >30 consecutive days).³³ Finally, to overcome all the comparative effectiveness analytical limitations introduced by the non-randomized observational study nature of population-based data, there is a need to conduct randomized clinical trials to produce unbiased comparative effectiveness estimates of the functional recovery following IRF and SNF rehabilitation or other modes of post-acute care.^{8,33} To conclude, our finding suggest that home time might not be a valid proxy measure of functional improvement because it is heavily

impacted by rehabilitation length of stay. Nevertheless, this approach has the potential to deliver stronger evidence needed to conduct future studies that utilize more stable functional outcome measures.

BIBLIOGRAPHY

1. Tsao, C. W.; Aday, A. W.; Almarzooq, Z. I.; Anderson, C. A. M.; Arora, P.; Avery, C. L.; Baker-Smith, C. M.; Beaton, A. Z.; Boehme, A. K.; Buxton, A. E.; Commodore-Mensah, Y.; Elkind, M. S. V.; Evenson, K. R.; Eze-Nliam, C.; Fugar, S.; Generoso, G.; Heard, D. G.; Hiremath, S.; Ho, J. E.; Kalani, R.; Kazi, D. S.; Ko, D.; Levine, D. A.; Liu, J.; Ma, J.; Magnani, J. W.; Michos, E. D.; Mussolino, M. E.; Navaneethan, S. D.; Parikh, N. I.; Poudel, R.; Rezk-Hanna, M.; Roth, G. A.; Shah, N. S.; St-Onge, M. P.; Thacker, E. L.; Virani, S. S.; Voeks, J. H.; Wang, N. Y.; Wong, N. D.; Wong, S. S.; Yaffe, K.; Martin, S. S.; American Heart Association Council on, E.; Prevention Statistics, C.; Stroke Statistics, S., Heart Disease and Stroke Statistics-2023 Update: A Report From the American Heart Association. *Circulation* **2023**, *147* (8), e93-e621. <https://doi.org/10.1161/CIR.0000000000001123>.
2. Writing Group, M.; Mozaffarian, D.; Benjamin, E. J.; Go, A. S.; Arnett, D. K.; Blaha, M. J.; Cushman, M.; Das, S. R.; de Ferranti, S.; Despres, J. P.; Fullerton, H. J.; Howard, V. J.; Huffman, M. D.; Isasi, C. R.; Jimenez, M. C.; Judd, S. E.; Kissela, B. M.; Lichtman, J. H.; Lisabeth, L. D.; Liu, S.; Mackey, R. H.; Magid, D. J.; McGuire, D. K.; Mohler, E. R., 3rd; Moy, C. S.; Muntner, P.; Mussolino, M. E.; Nasir, K.; Neumar, R. W.; Nichol, G.; Palaniappan, L.; Pandey, D. K.; Reeves, M. J.; Rodriguez, C. J.; Rosamond, W.; Sorlie, P. D.; Stein, J.; Towfighi, A.; Turan, T. N.; Virani, S. S.; Woo, D.; Yeh, R. W.; Turner, M. B.; American Heart Association Statistics, C.; Stroke Statistics, S., Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. *Circulation* **2016**, *133* (4), e38-360. <https://doi.org/10.1161/CIR.0000000000000350>.
3. Tong, X.; Kuklina, E. V.; Gillespie, C.; George, M. G., Medical complications among hospitalizations for ischemic stroke in the United States from 1998 to 2007. *Stroke* **2010**, *41* (5), 980-6. <https://doi.org/10.1161/STROKEAHA.110.578674>.
4. Winstein, C. J.; Stein, J.; Arena, R.; Bates, B.; Cherney, L. R.; Cramer, S. C.; Deruyter, F.; Eng, J. J.; Fisher, B.; Harvey, R. L.; Lang, C. E.; MacKay-Lyons, M.; Ottenbacher, K. J.; Pugh, S.; Reeves, M. J.; Richards, L. G.; Stiers, W.; Zorowitz, R. D.; American Heart Association Stroke Council, C. o. C.; Stroke Nursing, C. o. C. C.; Council on Quality of, C.; Outcomes, R., Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke* **2016**, *47* (6), e98-e169. <https://doi.org/10.1161/STR.0000000000000098>.
5. Prvu Bettger, J.; McCoy, L.; Smith, E. E.; Fonarow, G. C.; Schwamm, L. H.; Peterson, E. D., Contemporary trends and predictors of postacute service use and routine discharge home after stroke. *J Am Heart Assoc* **2015**, *4* (2). <https://doi.org/10.1161/JAHA.114.001038>.
6. Skolarus, L. E.; Feng, C.; Burke, J. F., No Racial Difference in Rehabilitation Therapy Across All Post-Acute Care Settings in the Year Following a Stroke. *Stroke* **2017**, *48* (12), 3329-3335. <https://doi.org/10.1161/STROKEAHA.117.017290>.
7. (MedPAC), M. P. A. C. *Report to the Congress: Medicare Payment Policy*; 2013.

8. Simmonds, K. P.; Burke, J.; Kozlowski, A. J.; Andary, M.; Luo, Z.; Reeves, M. J., Rationale for a Clinical Trial That Compares Acute Stroke Rehabilitation at Inpatient Rehabilitation Facilities to Skilled Nursing Facilities: Challenges and Opportunities. *Arch Phys Med Rehabil* **2022**, *103* (6), 1213-1221. <https://doi.org/10.1016/j.apmr.2021.08.004>.
9. Alcusky, M.; Ulbricht, C. M.; Lapane, K. L., Postacute Care Setting, Facility Characteristics, and Poststroke Outcomes: A Systematic Review. *Arch Phys Med Rehabil* **2018**, *99* (6), 1124-1140 e9. <https://doi.org/10.1016/j.apmr.2017.09.005>.
10. Deutsch, A.; Granger, C. V.; Heinemann, A. W.; Fiedler, R. C.; DeJong, G.; Kane, R. L.; Ottenbacher, K. J.; Naughton, J. P.; Trevisan, M., Poststroke rehabilitation: outcomes and reimbursement of inpatient rehabilitation facilities and subacute rehabilitation programs. *Stroke* **2006**, *37* (6), 1477-82. <https://doi.org/10.1161/01.STR.0000221172.99375.5a>.
11. Miller, E. L.; Murray, L.; Richards, L.; Zorowitz, R. D.; Bakas, T.; Clark, P.; Billinger, S. A.; American Heart Association Council on Cardiovascular, N.; the Stroke, C., Comprehensive overview of nursing and interdisciplinary rehabilitation care of the stroke patient: a scientific statement from the American Heart Association. *Stroke* **2010**, *41* (10), 2402-48. <https://doi.org/10.1161/STR.0b013e3181e7512b>.
12. Chan, T. C.; Brennan, J. J.; Castillo, E. M., Impact of skilled nursing facility (SNF) 3-day hospitalization requirement waiver during the COVID-19 pandemic on emergency department and inpatient SNF discharges in California. *J Am Coll Emerg Physicians Open* **2024**, *5* (1), e13094. <https://doi.org/10.1002/emp2.13094>.
13. *Skilled nursing facility (SNF) quality reporting program (QRP) public reporting*; Center for Medicare and Medicaid Services: 2023.
14. Cormier, D. J.; Frantz, M. A.; Rand, E.; Stein, J., Psychiatrist referral preferences for postacute stroke rehabilitation. *Medicine (Baltimore)* **2016**, *95* (33), e4356. <https://doi.org/10.1097/MD.0000000000004356>.
15. Stein, J.; Rodstein, B. M.; Levine, S. R.; Cheung, K.; Sicklick, A.; Silver, B.; Hedeman, R.; Egan, A.; Borg-Jensen, P.; Magdon-Ismail, Z.; Northeast Cerebrovascular Consortium Stroke, R.; Recovery Delphi Study, G., Which Road to Recovery?: Factors Influencing Postacute Stroke Discharge Destinations: A Delphi Study. *Stroke* **2022**, *53* (3), 947-955. <https://doi.org/10.1161/STROKEAHA.121.034815>.
16. Hong, I.; Goodwin, J. S.; Reistetter, T. A.; Kuo, Y. F.; Mallinson, T.; Karmarkar, A.; Lin, Y. L.; Ottenbacher, K. J., Comparison of Functional Status Improvements Among Patients With Stroke Receiving Postacute Care in Inpatient Rehabilitation vs Skilled Nursing Facilities. *JAMA Netw Open* **2019**, *2* (12), e1916646. <https://doi.org/10.1001/jamanetworkopen.2019.16646>.

17. Freburger, J. K.; Holmes, G. M.; Ku, L. J.; Cutchin, M. P.; Heatwole-Shank, K.; Edwards, L. J., Disparities in postacute rehabilitation care for stroke: an analysis of the state inpatient databases. *Arch Phys Med Rehabil* **2011**, *92* (8), 1220-9. <https://doi.org/10.1016/j.apmr.2011.03.019>.
18. Hong, I.; Karmarkar, A.; Chan, W.; Kuo, Y. F.; Mallinson, T.; Ottenbacher, K. J.; Goodwin, J. S.; Andersen, C. R.; Reistetter, T. A., Discharge Patterns for Ischemic and Hemorrhagic Stroke Patients Going From Acute Care Hospitals to Inpatient and Skilled Nursing Rehabilitation. *Am J Phys Med Rehabil* **2018**, *97* (9), 636-645. <https://doi.org/10.1097/PHM.0000000000000932>.
19. Xian, Y.; Thomas, L.; Liang, L.; Federspiel, J. J.; Webb, L. E.; Bushnell, C. D.; Duncan, P. W.; Schwamm, L. H.; Stein, J.; Fonarow, G. C.; Hoenig, H.; Montalvo, C.; George, M. G.; Lutz, B. J.; Peterson, E. D.; Bettger, J. P., Unexplained Variation for Hospitals' Use of Inpatient Rehabilitation and Skilled Nursing Facilities After an Acute Ischemic Stroke. *Stroke* **2017**, *48* (10), 2836-2842. <https://doi.org/10.1161/STROKEAHA.117.016904>.
20. Hayes, H. A.; Mor, V.; Wei, G.; Presson, A.; McDonough, C., Medicare Advantage Patterns of Poststroke Discharge to an Inpatient Rehabilitation or Skilled Nursing Facility: A Consideration of Demographic, Functional, and Payer Factors. *Phys Ther* **2023**, *103* (4). <https://doi.org/10.1093/ptj/pzad009>.
21. Ottenbacher, K. J.; Graham, J. E., The state-of-the-science: access to postacute care rehabilitation services. A review. *Arch Phys Med Rehabil* **2007**, *88* (11), 1513-21. <https://doi.org/10.1016/j.apmr.2007.06.761>.
22. Magdon-Ismail, Z.; Sicklick, A.; Hedeman, R.; Bettger, J. P.; Stein, J., Selection of Postacute Stroke Rehabilitation Facilities: A Survey of Discharge Planners From the Northeast Cerebrovascular Consortium (NECC) Region. *Medicine (Baltimore)* **2016**, *95* (16), e3206. <https://doi.org/10.1097/MD.00000000000003206>.
23. Stein, J.; Borg-Jensen, P.; Sicklick, A.; Rodstein, B. M.; Hedeman, R.; Bettger, J. P.; Hemmitt, R.; Silver, B. M.; Thode, H. C.; Magdon-Ismail, Z.; Northeast Cerebrovascular, C., Are Stroke Survivors Discharged to the Recommended Postacute Setting? *Arch Phys Med Rehabil* **2020**, *101* (7), 1190-1198. <https://doi.org/10.1016/j.apmr.2020.03.006>.
24. Buntin, M. B.; Garten, A. D.; Paddock, S.; Saliba, D.; Totten, M.; Escarce, J. J., How much is postacute care use affected by its availability? *Health Serv Res* **2005**, *40* (2), 413-34. <https://doi.org/10.1111/j.1475-6773.2005.00365.x>.
25. Chan, L.; Sandel, M. E.; Jette, A. M.; Appelman, J.; Brandt, D. E.; Cheng, P.; Teselle, M.; Delmonico, R.; Terdiman, J. F.; Rasch, E. K., Does postacute care site matter? A longitudinal study assessing functional recovery after a stroke. *Arch Phys Med Rehabil* **2013**, *94* (4), 622-9. <https://doi.org/10.1016/j.apmr.2012.09.033>.

26. Hoenig, H.; Sloane, R.; Horner, R. D.; Zolkewitz, M.; Reker, D., Differences in rehabilitation services and outcomes among stroke patients cared for in veterans hospitals. *Health Serv Res* **2001**, *35* (6), 1293-318.
27. Kind, A. J.; Smith, M. A.; Liou, J. I.; Pandhi, N.; Frytak, J. R.; Finch, M. D., Discharge destination's effect on bounce-back risk in Black, White, and Hispanic acute ischemic stroke patients. *Arch Phys Med Rehabil* **2010**, *91* (2), 189-95. <https://doi.org/10.1016/j.apmr.2009.10.015>.
28. Wang, H.; Sandel, M. E.; Terdiman, J.; Armstrong, M. A.; Klatsky, A.; Camicia, M.; Sidney, S., Postacute care and ischemic stroke mortality: findings from an integrated health care system in northern California. *PM R* **2011**, *3* (8), 686-94. <https://doi.org/10.1016/j.pmrj.2011.04.028>.
29. Prvu Bettger, J.; Thomas, L.; Liang, L. *Comparing Recovery Options for Stroke Patients*; Patient-Centered Outcomes Research Institute (PCORI): Washington (DC), 2019.
30. Buntin, M. B.; Colla, C. H.; Deb, P.; Sood, N.; Escarce, J. J., Medicare spending and outcomes after postacute care for stroke and hip fracture. *Med Care* **2010**, *48* (9), 776-84. <https://doi.org/10.1097/MLR.0b013e3181e359df>.
31. Kane, R. L.; Chen, Q.; Finch, M.; Blewett, L.; Burns, R.; Moskowitz, M., The optimal outcomes of post-hospital care under medicare. *Health Serv Res* **2000**, *35* (3), 615-61.
32. Springer, M. V.; Skolarus, L. E.; Feng, C.; Burke, J. F., Functional Impairment and Postacute Care Discharge Setting May Be Useful for Stroke Survival Prognostication. *J Am Heart Assoc* **2022**, *11* (6), e024327. <https://doi.org/10.1161/JAHA.121.024327>.
33. Simmonds, K. P.; Burke, J.; Kozlowski, A. J.; Andary, M.; Luo, Z.; Reeves, M. J., Emulating 3 Clinical Trials That Compare Stroke Rehabilitation at Inpatient Rehabilitation Facilities With Skilled Nursing Facilities. *Arch Phys Med Rehabil* **2022**, *103* (7), 1311-1319. <https://doi.org/10.1016/j.apmr.2021.12.029>.
34. ElHabr, A. K.; Katz, J. M.; Wang, J.; Bastani, M.; Martinez, G.; Gribko, M.; Hughes, D. R.; Sanelli, P., Predicting 90-day modified Rankin Scale score with discharge information in acute ischaemic stroke patients following treatment. *BMJ Neurol Open* **2021**, *3* (1), e000177. <https://doi.org/10.1136/bmjno-2021-000177>.
35. Mishra, N. K.; Shuaib, A.; Lyden, P.; Diener, H. C.; Grotta, J.; Davis, S.; Davalos, A.; Ashwood, T.; Wasiewski, W.; Lees, K. R.; Stroke Acute Ischemic, N. X. Y. T. I. T., Home time is extended in patients with ischemic stroke who receive thrombolytic therapy: a validation study of home time as an outcome measure. *Stroke* **2011**, *42* (4), 1046-50. <https://doi.org/10.1161/STROKEAHA.110.601302>.

36. Quinn, T. J.; Dawson, J.; Lees, J. S.; Chang, T. P.; Walters, M. R.; Lees, K. R.; Gain; Investigators, V., Time spent at home poststroke: "home-time" a meaningful and robust outcome measure for stroke trials. *Stroke* **2008**, *39* (1), 231-3. <https://doi.org/10.1161/STROKEAHA.107.493320>.
37. Yu, A. Y. X.; Fang, J.; Porter, J.; Austin, P. C.; Smith, E. E.; Kapral, M. K., Hospital-based cohort study to determine the association between home-time and disability after stroke by age, sex, stroke type and study year in Canada. *BMJ Open* **2019**, *9* (11), e031379. <https://doi.org/10.1136/bmjopen-2019-031379>.
38. Fonarow, G. C.; Liang, L.; Thomas, L.; Xian, Y.; Saver, J. L.; Smith, E. E.; Schwamm, L. H.; Peterson, E. D.; Hernandez, A. F.; Duncan, P. W.; O'Brien, E. C.; Bushnell, C.; Prvu Bettger, J., Assessment of Home-Time After Acute Ischemic Stroke in Medicare Beneficiaries. *Stroke* **2016**, *47* (3), 836-42. <https://doi.org/10.1161/STROKEAHA.115.011599>.
39. Yu, A. Y. X.; Rogers, E.; Wang, M.; Sajobi, T. T.; Coutts, S. B.; Menon, B. K.; Hill, M. D.; Smith, E. E., Population-based study of home-time by stroke type and correlation with modified Rankin score. *Neurology* **2017**, *89* (19), 1970-1976. <https://doi.org/10.1212/WNL.0000000000004631>.
40. McDermid, I.; Barber, M.; Dennis, M.; Langhorne, P.; Macleod, M. J.; McAlpine, C. H.; Quinn, T. J., Home-Time Is a Feasible and Valid Stroke Outcome Measure in National Datasets. *Stroke* **2019**, *50* (5), 1282-1285. <https://doi.org/10.1161/STROKEAHA.118.023916>.
41. Gattellari, M.; Goumas, C.; Jalaludin, B.; Worthington, J., Measuring stroke outcomes for 74 501 patients using linked administrative data: System-wide estimates and validation of 'home-time' as a surrogate measure of functional status. *Int J Clin Pract* **2020**, *74* (6), e13484. <https://doi.org/10.1111/ijcp.13484>.
42. O'Brien, E. C.; Xian, Y.; Xu, H.; Wu, J.; Saver, J. L.; Smith, E. E.; Schwamm, L. H.; Peterson, E. D.; Reeves, M. J.; Bhatt, D. L.; Maisch, L.; Hannah, D.; Lindholm, B.; Olson, D.; Prvu Bettger, J.; Pencina, M.; Hernandez, A. F.; Fonarow, G. C., Hospital Variation in Home-Time After Acute Ischemic Stroke: Insights From the PROSPER Study (Patient-Centered Research Into Outcomes Stroke Patients Prefer and Effectiveness Research). *Stroke* **2016**, *47* (10), 2627-33. <https://doi.org/10.1161/STROKEAHA.116.013563>.
43. Michigan Department of Health and Human Services (MDHHS), Michigan Stroke Program (MiSP). <https://www.michigan.gov/mdhhs/keep-mi-healthy/communicablediseases/epidemiology/chronicepi/stroke> (accessed 2023).
44. Center of Disease Control and Prevention, Paul Coverdell National Acute Stroke Program. https://www.cdc.gov/dhdsp/programs/stroke_registry.htm (accessed 2023).

45. American Heart Association, Get With The Guidelines® - Stroke Case Record Form. <https://www.heart.org/-/media/Files/Professional/Quality-Improvement/Get-With-the-Guidelines/Get-With-The-Guidelines-Stroke/Stroke--Diabetes-CRFJuly21.pdf>.
46. Michigan Value Collaborative, MVC Data Resources. <https://michiganvalue.org/resources-2/> (accessed 2023).
47. American Hospital Association, Annual Survey Database. <https://www.ahadata.com/> (accessed 2023).
48. Austin, P. C.; Yu, A. Y. X.; Vyas, M. V.; Kapral, M. K., Applying Propensity Score Methods in Clinical Research in Neurology. *Neurology* **2021**, *97* (18), 856-863. <https://doi.org/10.1212/WNL.0000000000012777>.
49. Austin, P. C.; Stuart, E. A., Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* **2015**, *34* (28), 3661-79. <https://doi.org/10.1002/sim.6607>.
50. Kogan, E.; Twyman, K.; Heap, J.; Milentijevic, D.; Lin, J. H.; Alberts, M., Assessing stroke severity using electronic health record data: a machine learning approach. *BMC Med Inform Decis Mak* **2020**, *20* (1), 8. <https://doi.org/10.1186/s12911-019-1010-x>.
51. Hernan, M. A., The hazards of hazard ratios. *Epidemiology* **2010**, *21* (1), 13-5. <https://doi.org/10.1097/EDE.0b013e3181c1ea43>.
52. Ni, A.; Lin, Z.; Lu, B., Stratified Restricted Mean Survival Time Model for Marginal Causal Effect in Observational Survival Data. *Ann Epidemiol* **2021**, *64*, 149-154. <https://doi.org/10.1016/j.annepidem.2021.09.016>.
53. Uno, H.; Claggett, B.; Tian, L.; Inoue, E.; Gallo, P.; Miyata, T.; Schrag, D.; Takeuchi, M.; Uyama, Y.; Zhao, L.; Skali, H.; Solomon, S.; Jacobus, S.; Hughes, M.; Packer, M.; Wei, L. J., Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol* **2014**, *32* (22), 2380-5. <https://doi.org/10.1200/JCO.2014.55.2208>.
54. Royston, P.; Parmar, M. K., Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* **2013**, *13*, 152. <https://doi.org/10.1186/1471-2288-13-152>.
55. Patorno, E.; Schneeweiss, S.; George, M. G.; Tong, X.; Franklin, J. M.; Pawar, A.; Mogun, H.; Moura, L.; Schwamm, L. H., Linking the Paul Coverdell National Acute Stroke Program to commercial claims to establish a framework for real-world longitudinal stroke research. *Stroke Vasc Neurol* **2022**, *7* (2), 114-123. <https://doi.org/10.1136/svn-2021-001134>.

56. Austin, P. C., Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* **2009**, 28 (25), 3083-107. <https://doi.org/10.1002/sim.3697>.
57. Austin, P. C., An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* **2011**, 46 (3), 399-424. <https://doi.org/10.1080/00273171.2011.568786>.
58. Greifer, N. S. A. E., Choosing the Estimand When Matching or Weighting in Observational Studies. *arXiv* **2021**, 2106 (10577).
59. Wu, A. H.; Pitt, B.; Anker, S. D.; Vincent, J.; Mujib, M.; Ahmed, A., Association of obesity and survival in systolic heart failure after acute myocardial infarction: potential confounding by age. *Eur J Heart Fail* **2010**, 12 (6), 566-73. <https://doi.org/10.1093/eurjhf/hfq043>.
60. Merkow, R. P.; Schwartz, T. A.; Nathens, A. B., Practical Guide to Comparative Effectiveness Research Using Observational Data. *JAMA Surg* **2020**, 155 (4), 349-350. <https://doi.org/10.1001/jamasurg.2019.4395>.

APPENDIX

Table 5A.1: Descriptive statistics of population characteristics stratified by discharge destination to home versus skilled nursing and inpatient rehabilitation facilities and their corresponding absolute standardized difference.

Variable	Home ± home health (N= 9,706)	IRF or SNF (N= 8,966)	Absolute standardized difference	
Demographics				
Age*	Mean (SD)	70.7 (12.8)	76.4 (17.7)	0.47
Sex*	Female	48.9	55.8	0.14
	Male	51.1	44.2	
Race	White	81.0	79.1	0.04
	Black	13.5	15.2	0.05
	Other	1.5	1.1	0.04
	Missing	4.0	4.6	0.03
Latino ethnicity*	Yes	4.0	3.4	0.03
	No	96.0	96.6	
Characteristics of stroke hospitalization				
Stroke Type*	Ischemic	90.6	85.5	0.16
	Hemorrhagic	9.4	14.5	
Admission NIHSS category	0	23.5	7.1	0.47
	1-4	49.4	33.7	0.32
	5-15	15.7	36.3	0.49
	16-20	1.4	6.9	0.28
	>20	1.1	5.9	0.26
	Missing	9.0	10.0	0.03
Admission duration*	Mean (SD)	3.2 (2.7)	6.2 (5.1)	0.75
Ambulatory status on discharge	Able to ambulate independently (no help from another person) w/ or w/o device	65.8	20.2	1.03
	With assistance (from person)	11.4	44.3	0.79
	Unable to ambulate	1.5	13.0	0.45
	Missing	21.2	22.5	0.03
Past medical history				
Missing medical history*	Yes	4.2	4.0	<0.01
	No	95.8	96.0	
Atrial fibrillation/flutter	Yes	17.1	26.7	0.23
	No	78.7	69.3	
Prosthetic heart valve	Yes	1.4	1.6	0.01
	No	94.4	94.4	
Coronary artery disease/ prior myocardial infarction	Yes	25.0	26.6	0.04
	No	70.9	69.4	
Carotid stenosis	Yes	5.0	4.7	0.02

Table 5A.1 (cont'd)

	No	90.8	91.2	
Diabetes mellitus	Yes	30.5	33.3	0.06
	No	65.4	62.7	
Peripheral vascular disease	Yes	5.8	6.4	0.02
	No	90.0	89.6	
Hypertension	Yes	71.8	76.4	0.11
	No	24.0	19.6	
Smoking	Yes	18.7	14.7	0.11
	No	77.1	81.3	
Dyslipidemia	Yes	55.0	54.0	0.02
	No	40.8	41.9	
Heart failure	Yes	9.9	13.4	0.11
	No	85.9	82.6	
Previous stroke	Yes	19.1	24.2	0.13
	No	76.7	71.7	
Previous Transient ischemic attack	Yes	10.4	10.0	0.02
	No	85.4	86.0	
Drug/alcohol abuse	Yes	6.4	5.8	0.03
	No	89.4	90.2	
Family history of stroke	Yes	12.5	9.8	0.09
	No	83.3	86.2	
Migraine	Yes	4.9	2.5	0.12
	No	90.9	93.4	
Obesity overweight	Yes	46.5	41.8	0.10
	No	49.3	54.1	
Chronic renal insufficiency	Yes	11.1	14.2	0.09
	No	84.7	81.8	
Sleep apnea	Yes	9.7	7.9	0.06
	No	86.1	88.0	
Depression	Yes	17.1	19.6	0.06
	No	78.7	76.4	
Deep vein thrombosis/ pulmonary embolism	Yes	1.9	2.3	0.03
	No	93.9	93.7	
Dementia	Yes	0.3	0.8	0.07
	No	95.5	95.2	
Hospital Characteristics				
Bed Size*	50-99	0.9	0.6	0.04
	100-199	9.0	7.6	0.05
	200-299	13.8	13.1	0.02
	300-399	22.2	23.0	0.02
	400-499	17.2	18.6	0.04
	>=500	36.9	37.1	<0.01
Core-based statistical area*	Metro	94.2	94.2	<0.01
	Micro	3.0	3.5	0.03

Table 5A.1 (cont'd)

	Rural	2.9	2.2	0.04
System member*	Yes	85.4	85.4	<0.01
	No	14.6	14.6	
Stroke accreditation*	CSC: Comprehensive Stroke Center	44.5	47.1	0.05
	PSC: Primary Stroke Center	44.9	42.6	0.05
	TSR: Thrombectomy Capable Stroke Center	10.6	10.3	0.01
Stroke rehabilitation accreditation*	Yes	10.3	9.8	0.02
	No	89.7	90.2	
Teaching hospital (Medical school affiliation reported to American Medical Association)*	Yes	82.7	83.9	0.03
	No	17.3	16.1	

*Covariate did not have any missing values.

Table 5A.2: Descriptive statistics of IRF and SNF populations characteristics stratified by Medicare FFS and Medicare Advantage beneficiaries and their corresponding absolute standardized difference.

Variable		Medicare FFS (N= 5,943)	Medicare Advantage (N= 1,898)	Absolute standardized difference
Demographics				
Age*	Mean (SD)	77.7 (10.8)	79.0 (8.78)	0.13
Sex*	Female	58.6	53.4	0.11
	Male	41.4	46.6	
Race	White	78.9	82.8	0.10
	Black	15.6	12.4	0.09
	Other	1.2	0.7	0.05
	Missing	4.4	4.1	0.01
Latino ethnicity*	Yes	3.1	3.2	0.01
	No	96.9	96.8	
Characteristics of stroke hospitalization				
Stroke Type*	Ischemic	86.7	84.4	0.06
	Hemorrhagic	13.3	15.6	
Admission NIHSS category	0	7.4	6.8	0.02
	1-4	33.9	34.8	0.02
	5-15	36.4	35.4	0.02
	16-20	6.5	7.6	0.04
	>20	5.6	5.5	<0.01
	Missing	10.2	9.9	0.01
Admission duration*	Mean (SD)	5.8 (4.5)	6.7 (5.0)	0.19
Ambulatory status on discharge	Able to ambulate independently (no help from another person) w/ or w/o device	19.9	19.7	<0.01
	With assistance (from person)	45.5	43.8	0.03

Table 5A.2 (cont'd)

	Unable to ambulate	12.5	13.1	0.02
	Missing	22.0	23.4	0.03
Past medical history				
Missing medical history*	Yes	4.5	3.4	0.06
	No	95.5	96.7	
Atrial fibrillation/flutter	Yes	27.8	29.4	0.03
	No	67.7	67.2	
Prosthetic heart valve	Yes	1.5	2.1	0.05
	No	93.9	94.5	
Coronary artery disease/ prior myocardial infarction	Yes	28.1	27.4	0.02
	No	67.4	69.2	
Carotid stenosis	Yes	4.8	5.8	0.04
	No	90.6	90.8	
Diabetes mellitus	Yes	34.6	31.6	0.06
	No	60.8	65.0	
Peripheral vascular disease	Yes	7.1	5.2	0.08
	No	88.4	91.4	
Hypertension	Yes	76.9	79.0	0.05
	No	18.6	17.6	
Smoking	Yes	14.9	10.3	0.14
	No	80.6	86.3	
Dyslipidemia	Yes	54.0	59.1	0.10
	No	41.4	37.5	
Heart failure	Yes	14.1	13.3	0.03
	No	81.3	83.3	
Previous stroke	Yes	26.1	22.4	0.08
	No	69.4	74.2	
Previous Transient ischemic attack	Yes	10.4	11.3	0.03
	No	85.1	85.3	
Drug/alcohol abuse	Yes	5.2	4.0	0.06
	No	90.2	92.6	
Family history of stroke	Yes	9.4	10.3	0.03
	No	86.1	86.3	
Migraine	Yes	93.2	94.0	0.02
	No	2.3	2.6	
Obesity overweight	Yes	40.6	42.7	0.04
	No	54.9	53.9	
Chronic renal insufficiency	Yes	15.5	13.2	0.07
	No	80.0	83.4	
Sleep apnea	Yes	7.7	8.3	0.02
	No	87.7	88.3	
Depression	Yes	20.2	18.3	0.05
	No	75.3	78.3	

Table 5A.2 (cont'd)

Deep vein thrombosis/ pulmonary embolism	Yes	2.1	3.5	0.09
	No	93.4	93.1	
Dementia	Yes	0.7	1.1	0.04
	No	94.8	95.5	
Hospital Characteristics				
Bed Size*	50-99	0.6	0.4	0.03
	100-199	7.6	7.5	0.01
	200-299	13.3	12.4	0.03
	300-399	24.3	22.9	0.03
	400-499	18.2	19.8	0.04
	>=500	36.0	37.0	0.02
Core-based statistical area*	Metro	94.0	95.2	0.05
	Micro	3.7	2.7	0.06
	Rural	2.3	2.1	0.01
System member*	Yes	85.7	85.7	<0.01
	No	14.3	14.3	
Stroke accreditation*	CSC: Comprehensive Stroke Center	45.7	46.4	0.01
	PSC: Primary Stroke Center	44.4	41.7	0.05
	TSR: Thrombectomy Capable Stroke Center	9.9	11.9	0.06
Stroke rehabilitation accreditation*	Yes	8.9	10.7	0.06
	No	91.1	89.3	
Teaching hospital (Medical school affiliation reported to American Medical Association)*	Yes	83.2	85.4	0.06
	No	16.8	14.6	

*Covariate did not have any missing values.

Table 5A.3: Propensity Score logistic regression model covariate estimates odds ratios and 95% confidence intervals (event = IRF).

Variable	Odds ratio	95% CI	Chi square <i>p</i> -value	
Demographics				
Age*	0.94	0.94 - 0.95	<0.01	
Sex*	Female	ref	<0.01	
	Male	1.35		1.19 - 1.52
Race	Black	ref	0.80	
	White	1.04		0.88 - 1.23
	Other	0.84		0.49 - 1.43
	Missing	0.97		0.71 - 1.33
Latino ethnicity*	No	ref	0.94	
	Yes	0.99		0.69 - 1.41
Characteristics of stroke hospitalization				
Stroke Type*	Hemorrhagic	ref	<0.01	
	Ischemic	0.68		0.57 - 0.82
Admission NIHSS category	0	ref	<0.01	
	1-4	1.06		0.85 - 1.33

Table 5A.3 (cont'd)

	5-15	1.02	0.82 – 1.29	
	16-20	0.88	0.64 – 1.20	
	>20	0.84	0.60 – 1.16	
	Missing	0.44	0.33 – 0.59	
Admission duration*		0.92	0.90 – 0.93	<0.01
Ambulatory status on discharge	Able to ambulate independently (no help from another person) w/ or w/o device		ref	<0.01
	With assistance (from person)	1.28	1.09 – 1.50	
	Unable to ambulate	0.41	0.33 – 0.52	
	Missing	0.95	0.78 – 1.16	
Past medical history				
Missing medical history**	No		ref	<0.01
	Yes	2.36	1.28 – 4.34	
Atrial fibrillation/flutter	No		ref	0.02
	Yes	0.85	0.74 – 0.97	
Prosthetic heart valve	No		Ref	0.28
	Yes	0.79	0.48 – 1.23	
Coronary artery disease/ prior myocardial infarction	No		ref	0.69
	Yes	1.03	0.89 – 1.18	
Carotid stenosis	No		ref	0.74
	Yes	1.05	0.80 – 1.37	
Diabetes mellitus	No		ref	<0.01
	Yes	0.71	0.62 – 0.81	
Peripheral vascular disease	No		ref	0.34
	Yes	0.89	0.71- 1.13	
Hypertension	No		ref	<0.01
	Yes	1.26	1.07 – 1.48	
Smoking	No		ref	0.19
	Yes	0.89	0.74 – 1.06	
Dyslipidemia	No		ref	0.14
	Yes	1.10	0.97 – 1.26	
Heart failure	No		ref	<0.01
	Yes	0.74	0.62 – 0.89	
Previous stroke	No		ref	<0.01
	Yes	0.64	0.56 – 0.73	
Previous Transient ischemic attack	No		ref	0.64
	Yes	0.96	0.79 – 1.16	
Drug/alcohol abuse	No		ref	0.13
	Yes	0.81	0.62 – 1.06	
Family history of stroke	No		ref	0.01
	Yes	1.30	1.06 – 1.57	
Migraine	No		ref	0.53
	Yes	1.13	0.77 – 1.67	
Obesity overweight	No		ref	0.12
	Yes	1.11	0.97 – 1.26	
Chronic renal insufficiency	No		ref	0.06
	Yes	0.85	0.73 – 1.01	
Sleep apnea	No		ref	0.37
	Yes	0.90	0.72 – 1.13	

Table 5A.3 (cont'd)

Depression	No		ref	<0.01
	Yes	0.73	0.63 – 0.85	
Deep vein thrombosis/ pulmonary embolism	No		ref	0.76
	Yes	1.07	0.71 – 1.60	
Dementia	No		ref	0.62
	Yes	0.84	0.42 – 1.68	
Hospital Characteristics				
Bed Size*	50-99		ref	<0.01
	100-199	0.28	0.13 – 0.60	
	200-299	0.70	0.34 – 1.46	
	300-399	1.46	0.71 – 3.00	
	400-499	1.53	0.74 – 3.17	
	>=500	1.25	0.60 – 2.60	
Core-based statistical area*	Metro		ref	<0.01
	Micro	0.37	0.26 – 0.52	
	Rural	7.84	4.80 – 12.81	
System member*	No		ref	<0.01
	Yes	0.75	0.60 – 0.92	
Stroke accreditation*	CSC: Comprehensive Stroke Center		ref	<0.01
	PSC: Primary Stroke Center	1.35	1.15 – 1.58	
	TSR: Thrombectomy Capable Stroke Center	0.85	0.67 – 1.07	
Stroke rehabilitation accreditation*	No		ref	0.03
	Yes	0.75	0.57 – 0.98	
Teaching hospital (Medical school affiliation reported to American Medical Association)*	No		ref	0.78
	Yes	1.03	0.85 – 1.24	

*Covariate did not have any missing values.

**Because of identical missing proportions among medical history variables, missingness entered the model as a standalone variable.

Table 5A.4: Descriptive statistics of propensity scores and IPTW stratified by rehabilitation allocation to IRF and SNF.

Descriptive statistic	IRF (n= 2,995)		SNF (n= 2,948)	
	Propensity score	IPTW	Propensity score	IPTW
Mean	0.60	2.00	0.41	2.03
Median	0.61	1.63	0.40	1.66
Mode	0.37	1.32	0.29	1.41
Standard deviation	0.18	1.58	0.20	1.21
Minimum	0.03	1.03	0.01	1.01
Maximum	0.97	38.20	0.95	18.15
Lower quartile	0.47	1.36	0.26	1.34
Upper quartile	0.74	2.14	0.56	2.27
Interquartile range	0.27	0.78	0.30	0.93
1 st percentile	0.14	1.08	0.05	1.05
5 th percentile	0.26	1.17	0.10	1.12
10 th percentile	0.34	1.22	0.15	1.17
90 th percentile	0.82	2.96	0.68	3.12
95 th percentile	0.86	3.83	0.76	4.22
99 th percentile	0.93	7.38	0.86	7.14

Figure 5A.1: IPTW box plots stratified by rehabilitation allocation to IRF and SNF.

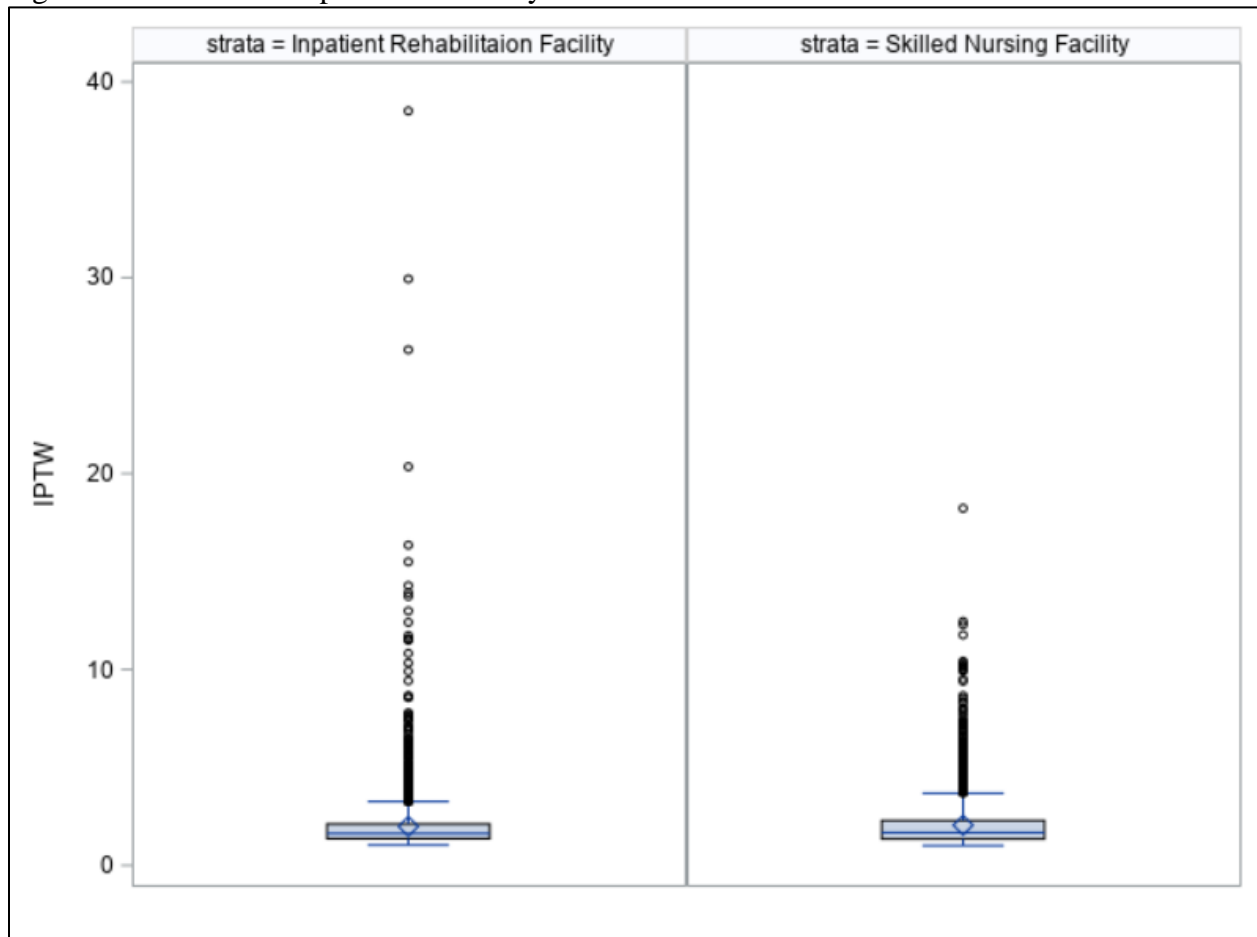
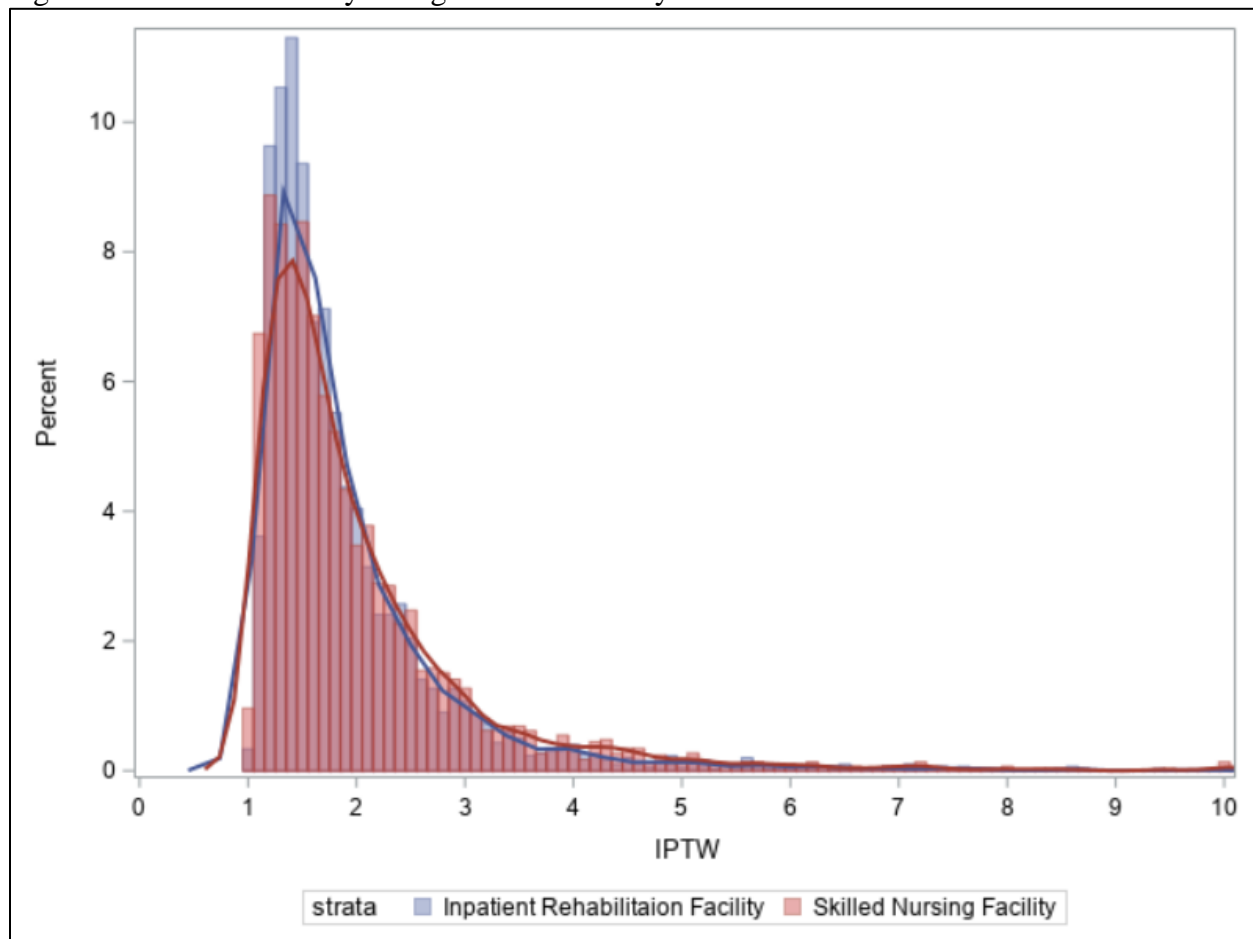


Figure 5A.2: IPTW overlay histograms stratified by rehabilitation allocation to IRF and SNF.*



*Histogram width is 0.5 and x-axis is limited to IPTW=10.

Table 5A.5: Descriptive statistics of the initial and cumulative 30-day rehabilitation and readmission events stratified by IRF and SNF populations.

Measure/population	IRF	SNF	P-value*
	Average and SD OR N (%)	Average and SD OR N (%)	
All population	n=2,995	n=2,948	-
Initial rehabilitation length of stay (days)	14.6 (8.0)	11.5 (8.2)	<0.01
Total 30-day rehabilitation length of stay in the same setting (days)	15.3 (8.1)	26.4 (21.3)	<0.01
30-days all-cause readmission	376 (12.6%)	577 (19.6%)	<0.01
Readmitted population	n=376	n=577	-
30-day rehabilitation admission rate in the same setting	1.3 (0.5)	2.1 (0.8)	<0.01
30-days all-cause readmission rate	1.1 (0.4)	1.2 (0.4)	0.67
Time to first all-cause readmission within 30-days of acute stroke discharge (days)	14.0 (9.1)	12.4 (8.5)	<0.01
Total 30-day readmission length of stay (days)	7.0 (6.7)	6.9 (6.3)	0.76

*X²- and t-test were used according to the type of data.

Table 5A.6: Descriptive statistics of IRF and SNF populations characteristics among Medicare FFS population using the complete case analysis dataset and their corresponding absolute standardized difference.*

Variable	IRF (N=2,148)	SNF (N= 1,893)	Absolute standardized difference	
Demographics				
Age	Mean (SD)	75.9 (10.5)	80.4 (10.3)	0.43
Sex	Female	55.2	63.2	0.16
	Male	44.8	36.8	
Race	White	83.5	84.0	0.01
	Black	15.6	15.1	0.01
	Other	0.9	1.0	<0.01
Latino ethnicity	Yes	1.3	1.1	0.01
	No	98.7	98.9	
Characteristics of stroke hospitalization				
Stroke Type	Ischemic	89.3	90.5	0.04
	Hemorrhagic	10.7	9.5	
Admission NIHSS category	0	9.0	8.6	0.01
	1-4	41.5	35.5	0.13
	5-15	39.7	40.0	0.02
	16-20	5.1	8.1	0.12
	>20	4.7	7.8	0.12
Admission duration	Mean (SD)	4.9 (3.5)	6.2 (4.5)	0.33
Ambulatory status on discharge	Able to ambulate independently (no help from another person) w/ or w/o device	27.1	25.0	0.05
	With assistance (from person)	64.5	50.5	0.29
	Unable to ambulate	8.4	24.5	0.44
Past medical history				
Atrial fibrillation/flutter	Yes	23.0	33.3	0.23
	No	77.0	66.7	
Prosthetic heart valve	Yes	1.4	2.3	0.07
	No	98.6	97.7	
Coronary artery disease/ prior myocardial infarction	Yes	30.8	32.3	0.03
	No	69.2	67.7	
Carotid stenosis	Yes	5.8	5.8	<0.01
	No	94.2	94.2	
Diabetes mellitus	Yes	37.5	39.2	0.04
	No	62.5	60.8	
Peripheral vascular disease	Yes	7.1	9.3	0.08
	No	92.9	90.7	
Hypertension	Yes	82.8	84.4	0.04
	No	17.2	15.6	

Table 5A.6 (cont'd)

Smoking	Yes	17.7	13.0	0.13
	No	82.3	87.0	
Dyslipidemia	Yes	59.8	61.0	0.03
	No	40.2	39.0	
Heart failure	Yes	12.3	18.4	0.17
	No	87.7	81.6	
Previous stroke	Yes	24.9	32.5	0.17
	No	75.1	67.5	
Previous Transient ischemic attack	Yes	10.7	13.4	0.08
	No	89.3	86.6	
Drug/alcohol abuse	Yes	6.4	4.4	0.09
	No	93.6	95.6	
Family history of stroke	Yes	11.7	8.8	0.10
	No	88.3	91.2	
Migraine	Yes	3.1	2.2	0.06
	No	96.9	97.8	
Obesity overweight	Yes	51.3	45.9	0.11
	No	48.7	54.1	
Chronic renal insufficiency	Yes	14.7	18.6	0.10
	No	85.3	81.4	
Sleep apnea	Yes	8.4	8.3	<0.01
	No	91.6	91.7	
Depression	Yes	21.4	24.6	0.08
	No	78.6	75.4	
Deep vein thrombosis/ pulmonary embolism	Yes	2.1	2.5	0.03
	No	97.9	97.5	
Dementia	Yes	0.5	0.8	0.04
	No	99.5	99.2	
Hospital Characteristics				
Bed Size	50-99	0.9	0.7	0.02
	100-199	6.2	9.1	0.11
	200-299	10.5	17.1	0.19
	300-399	30.9	25.3	0.13
	400-499	22.5	21.6	0.02
	>=500	28.9	26.3	0.06
Core-based statistical area	Metro	91.3	91.7	0.01
	Micro	4.3	6.3	0.09
	Rural	4.3	1.9	0.14
System member	Yes	88.9	92.1	0.11
	No	11.1	7.9	
Stroke accreditation	CSC: Comprehensive Stroke Center	39.5	38.5	0.02
	PSC: Primary Stroke Center	50.6	47.0	0.07
	TSR: Thrombectomy Capable Stroke Center	9.9	14.5	0.14

Table 5A.6 (cont'd)

Stroke rehabilitation accreditation	Yes	4.8	4.6	0.01
	No	95.2	95.4	
Teaching hospital (Medical school affiliation reported to American Medical Association)	Yes	82.0	79.5	0.06
	No	18.0	20.5	

*The complete case analysis dataset excluded all patients with missing values for admission NIHSS, discharge ambulatory status, medical history, and race.

Table 5A.7: Unadjusted and IPTW adjusted differences in home time, mortality, and restricted mean survival time (RMST) between patients discharged to inpatient rehabilitation facility (IRF) or skilled nursing facility (SNF) in the complete case analysis dataset.*

Outcome, time point	IRF (N= 2,148)	SNF (N= 1,893)	Unadjusted effect measure (N= 4,041)		IPTW adjusted effect measure (N= 4,000)**	
	Mean (SD) or N (%)		Mean difference or odds ratio (95% CI)	<i>p</i> -value***	Mean difference or odds ratio (95% CI)	<i>p</i> -value***
Home Time						
90-day	57.9 (27.8)	41.6 (29.4)	16.3 (14.3 – 17.8)	<0.001	11.6 (9.7 – 13.5)	<0.001
1-year	287.8 (103.8)	218.9 (133.4)	68.9 (61.4 – 76.3)	<0.001	44.4 (36.5 – 52.2)	<0.001
Mortality rate						
90-day	124 (5.8%)	276 (14.6%)	0.36 (0.28 – 0.44)^	<0.001	0.58 (0.44 – 0.72)^	<0.001
1-year	368 (17.1%)	672 (35.5%)	0.38 (0.32 – 0.43)^	<0.001	0.59 (0.50 – 0.68)^	<0.001
Restricted mean survival time (RMST)						
90-day	86.9 (13.1)	81.0 (22.1)	5.9 (4.8 – 7.0)	<0.001	3.9 (3.1 – 4.7)	<0.001
1-year	327.7 (92.7)	278.8 (130.9)	48.9 (41.8 – 56.0)	<0.001	30.5 (25.5 – 35.5)	<0.001

*The complete case analysis dataset excluded all patients with missing values for admission NIHSS, discharge ambulatory status, medical history, and race.

**41 discharges (22 IRF and 19 SNF) were deleted because they had an extreme IPTW above 99th percentile.

***Independent t-test or X^2 test.

^ Odds ratio and 95% confidence interval.

Table 5A.8: Unadjusted and IPTW adjusted differences in home time between patients discharged to inpatient rehabilitation facility (IRF) or skilled nursing facility (SNF) among patients discharged alive from rehabilitation.

Outcome, time point	IRF (N= 2,990)	SNF (N= 2,907)	Unadjusted effect measure (N= 5,897)		IPTW adjusted effect measure (N= 5,839)*	
	Mean (SD)		Mean difference (95% CI)	<i>p</i> -value^	Mean difference (95% CI)	<i>p</i> -value^
Home Time						
90-day	57.7 (27.8)	42.5 (29.3)	15.2 (13.7 -16.6)	<0.001	10.9 (9.3 – 12.5)	<0.001
1-year	288.2 (103.6)	223.1 (131.5)	65.1 (59.0 – 71.1)	<0.001	45.3 (38.7 – 51.8)	<0.001

*58 discharges were deleted because they had an extreme IPTW above 99th percentile.

^Independent t-test.

CHAPTER 6: SUMMARY AND DISCUSSION

6.1 Summary of Findings and Limitations

6.1.1 Summary of Findings

In this dissertation, we set out to provide an assessment of the long-term recovery from stroke in Michigan using data from Michigan Stroke Program (MiSP) registry. Specifically, we addressed the following primary objectives:

- 1- 1a) Generate a linked database by linking a 5-year retrospective cohort of all acute stroke discharges entered into MiSP registry between 2016-2020 with Michigan Value Collaborative (MVC) - a claims database using both deterministic and probabilistic matching techniques.
- 1b) Use the linked data to generate descriptive data on 30-day, 90-day and 1-year outcome event rates including mortality, all-cause hospital readmissions, stroke recurrence, use of post-acute care services (i.e., inpatient rehabilitation facility (IRF), skilled nursing facility (SNF), and home health), out-patient visits, and home time.
- 2- Develop 30-day and 1-year all-cause readmission prediction models using LASSO logistic regression, and two non-linear machine learning based methods (i.e., XGBoost and ANN), compare the predictive performance of these methods, and report the most important predictors from the best performing prediction models.
- 3- Estimate the comparative effectiveness of inpatient rehabilitation facility (IRF) versus skilled nursing facility (SNF) institutional rehabilitation care on functional recovery in Medicare fee-for-service (FFS) acute stroke hospitalizations over 90 days and 1 year post discharge using home time and report on all-cause mortality.

The data used in this dissertation was from acute stroke patients discharged from 31 stroke certified hospitals in Michigan which were linked to administrative claims data from Blue Cross Blue Shield of Michigan (BCBSM) (private and Medicare Advantage plans) or Medicare fee-for-service (FFS). The major findings are summarized as follows:

- 1- 1a) Probabilistic linkage of MiSP and MVC produced a higher number of unique linked pairs (n= 23,918) compared to deterministic linkage (n= 22,660). Of the 46,330 MiSP stroke events, 23,918 (51.6%) were linked to the MVC claims database; these links represent 77.9% of the 30,685 MVC acute-stroke claims. As anticipated based on the coverage of the MVC claims data we found lower linkage rates in MiSP data among the <65 age group compared to >=65 age group (29.2% vs 63.7%).

1b) Stroke outcome event rates were similar to previously published rates in the literature. Among the 19,382 linked 1-year stroke episodes of care, 24.9%, 28.1%, 27.5%, and 46.4% utilized IRF, SNF, home health, and outpatient care at least once within 30-days of hospital discharge, respectively. A total of 14.1%, 24.9%, and 42.2% of the linked population were readmitted at least once within 30-days, 90-days, and 1-year post discharge, respectively. Only 3.3% of our linked population had a stroke recurrence within 30-days; this increased to 5.1% at 90-days and to 8.3% at 1-year post discharge. Among the 12,185 Medicare FFS linked stroke cases; mortality rates were 4.0%, 9.1%, and 19.8% within 30-days, 90-days, and 1-year post discharge, respectively. In addition, among the FFS population, median home time was found to be 22.0, 79.0, 347.0 days within 30-days, 90-days, and 1-year post discharge, respectively.
- 2- The linked population had a mean age of 73.3 (SD= 12.7), 79.7% were white, 52.2% were female, 87.3% had an ischemic stroke, 56.4% had a minor stroke (NIHSS <5), and

50.1% were discharged directly home. Of 19,382 linked stroke discharges, 2,724 (14.1%) and 8,169 (42.2%) were readmitted within 30-days and 1-year, respectively. Using registry data, LASSO logistic regression model produced similar AUC to XGBoost and ANN (p -value >0.05) with a 30-day and 1-year readmission AUC of 0.68 (95% CI: 0.65-0.70) and 0.67 (95% CI: 0.65-0.69), respectively. Variables with the highest predictive importance were discharge disposition, acute hospital length of stay, and preexisting comorbidities including chronic renal failure, heart failure, and atrial fibrillation. In contrast, clinical features of stroke (e.g., NIHSS, stroke etiology, and ambulatory status) were less important and were almost absent from the 1-year readmission model. Models that utilized either MiSP or MVC data or the combination of the two produced similar 30-day and 1-year AUC that were not statistically significantly different from each other.

- 3- Of the included 5,943 Medicare FFS beneficiaries, 2,995 and 2,948 were discharged alive to either an IRF or SNF, respectively. Compared to SNF patients, IRF patients were younger, had shorter acute hospital length of stay, were less likely to be females, were less likely to have very severe stroke (NIHSS >20), and were more likely to be able to ambulate at discharge. In terms of comorbidities and past medical history IRF patients had lower prevalence of atrial fibrillation, heart failure, and previous stroke but were more likely to be smokers. After Inverse probability of treatment weighting (IPTW) adjustment, compared to SNF, IRF patients had increased mean home time of 11.1 days (9.5 – 12.57 and 46.3 days (39.8 – 52.9) at 90-days and 1-year, respectively. However, in sensitivity analyses that accounted for differences in rehabilitation length of stay during the first 30-days post discharge, the mean difference in adjusted 90-day home time disappeared (mean 0.5 days; 95% CI: -1.1 – 2.1), although there remained a significant

difference at 1-year (35.7 days; 95% CI: 29.1 – 42.2). Mortality was noticeably lower in patients discharged to IRF; IRF patients were associated with 48% and 45% lower adjusted odds of death over 90-days and 1-year post discharge, respectively. In the sensitivity analysis that excluded patients who died within 90-days and 1-year, the mean difference in adjusted 90-day and 1-year home time decreased to 9.7 days (95% CI: 8.1 – 11.3) and 23.2 days (19.0 – 27.4), respectively.

These findings illustrate that probabilistic linkage between MiSP acute stroke registry and MVC claims data using indirect identifiers is feasible and that these data can be used to generate several stroke outcomes including stroke recurrence, all cause readmission, mortality, home time, and outpatient and rehabilitation care utilization up to 1-year post discharge – which were not readily available previously. Further, the linked data demonstrated that prediction of all cause readmission can be achieved with relatively high accuracy, that LASSO regression was able to predict readmission after stroke with similar accuracy to more advanced ML methods, and that clinical features of stroke were much less important than the burden of existing comorbidities in predicting post-stroke readmission, especially over longer periods of time (1-year). The linked data also provided an important opportunity to examine differences in home time between IRF and SNF patients.

Our findings provided further evidence that in Medicare FFS stroke patients in need of post-acute rehabilitation, discharge to IRF versus SNF was associated with longer home time (a previously validated measure of functional recovery) and lower mortality over one year post discharge. However, our sensitivity analysis illustrated that home time especially in the short term is heavily impacted by rehabilitation length of stay hence future studies should avoid using home time as a valid proxy of functional recovery over 90 days short period of

follow up and should rely on more stable measures like mRS or successful community discharge (home for >30 consecutive days). Our sensitivity analysis also illustrated that mortality is a major contributor to home time differences over the longer 1-year follow up period.

6.1.2 Limitations

This work has several important limitations, each of which has been discussed in some length in Chapters 3-5. First, our linkage work (described in Chapter 3) was affected by limitations in MVC claims data insurance coverage that resulted in excluding many stroke discharges recorded by MiSP registry because they were covered by Medicaid, private insurance plans other than BCBSM, Medicare Advantage plans other than BCBSM, or were uninsured. This resulted in a low linkage rate (51.6% of MiSP population). In addition, when compared to the unlinked population, the linked population was older, more white, more likely to be females, and carried a higher burden of comorbidities. These facts may limit the generalizability of our results to patients insured by Medicare FFS and BCBSM.

Despite the implementation of rigorous linkage evaluation techniques, limitations in the availability of valid negative and positive controls, the lack of personal identifiers, and the lack of a linkage gold standard (linkage that produces a reference dataset where true match status is known with certainty) made it difficult to generate measures of linkage accuracy (i.e., sensitivity, specificity, and positive predictive value). These limitations were carried over to Chapters 4 and 5.

In the readmission analysis (described in Chapter 4), the registry data suffered from high levels of non-random missingness in several important clinical predictors (e.g., in-patient procedures like intubation and foleys catheter insertion) which limited the number of predictors

that were included in the analysis. This may have reduced the accuracy of our prediction model. Further, we did not externally validate our models because we had no access to external sources similar to our data. Nevertheless, given the robustness of our model development technique and the extensive leave-hospital-out cross validation method, the results can be considered as representative of the expected prediction accuracy for the models developed in this study to predict readmission for hospitals not included in our data set.

The major limitation of the comparative effectiveness analysis of IRF vs SNF (described in Chapter 5) is the fact that the study population was limited to Medicare FFS beneficiaries which limits the generalizability of our findings. Also, all of our analyses are based on stroke cases discharged from stroke certified hospitals which further limits generalizability. The observational analysis design of this study which limits the availability of the factors that could be studied means that our data are prone to selection bias and residual measured and unmeasured confounders which could reduce the validity of our comparisons. We tried to overcome the selection bias by including all the relevant prognostic variables in our analysis and by conducting a sensitivity analysis that excluded all the observations that suffered from missing data. The findings from the sensitivity analysis suggest that data missingness was at random and thus was non informative. Also, balance diagnostics supported that our propensity score approach controlled for measured confounding. Finally, due to the scarcity of studies that utilized home time as a functional outcome measure, we were only able to compare our study results with one study.

6.2 Direction of Future Research

There are several potential routes through which future work can build on this dissertation. First, future research should attempt to expand access to claims data from other

insurance providers including Medicaid and other private insurers (i.e., Health Alliance Plan (Henry Ford Health System), Priority Health, and United Health) which would allow the generation of more generalizable data and the ability to investigate differences between insurance providers. Additionally, if personal identifiers become available, internal cross validation of the linkage should take place within Michigan using gold standard MiSP and MVC linked datasets, this would provide stronger evidence that allow for the external validation of our linkage method using similar data linkages in other states or regions.

Second, to improve the prediction accuracy of readmission, future studies should explore integrating electronic medical records and claims data features that cover a wider range of comorbidities and areas of inpatient clinical and post-acute care like lab results, rehabilitation, outpatient follow up visits, and prescription fillings. Additionally, external validation of the prediction model should be explored using similar linked data from states participating in Paul Coverdale National Acute Stroke Program or national GWTG-S data. Most importantly, future research could also explore developing prediction models specific for ischemic or hemorrhagic strokes as the linked data allows for such investigations. We did not use the data to investigate stroke recurrence, but this would be another important line of investigation as many clinical interventions post stroke have the prevention of stroke as their primary goal.

Thirdly, due to limitations in the availability of data from the registry and mortality data from MVC this dissertation studied the comparative effectiveness of IRF vs SNF only among Medicare FFS beneficiaries. Therefore, future research should attempt to expand data collection from additional hospitals in Michigan and obtain more robust mortality data from other insurance providers. For example, mortality data could be obtained by conducting a linkage between the national vital records and the MiSP-MVC dataset which is very important to

generate a more generalizable findings and vitally research functional outcomes among the younger stroke patients (< 65 years old). Further, future studies should assess the validity of home time as a valid proxy of functional recovery following rehab. Alternative definitions of zero time i.e., calculating home time from the point of discharge from the end of rehabilitation care and not from acute care should be explored. In addition, given the limitations of retrospective observational population-based data, there is an indispensable need to conduct randomized clinical trials that examine the functional recovery of IPR and SNF rehabilitation allocation.

Finally, in addition to stroke recurrence, the linked dataset is rich in other outcomes that could be further analyzed in future research projects including post-acute home health, outpatient rehabilitation, and primary or specialized outpatient follow up care utilization.

6.3 Implications for Public Health, Clinical Practice, and Public Policy

The findings of this dissertation have important implications for both public health, clinical practice, and public policy. First, our linked dataset provided important and often hard to come by descriptive statistics of post stroke outcomes up to 1-year post discharge that indicated that Michigan residents have similar stroke outcomes to nationally published reports. Robust longitudinal stroke outcome information may provide the necessary data to serve the purpose of evaluating health systems, health insurance providers, current health policies, and hospitals in Michigan through improvements in stroke care and outcomes. In addition, we provided evidence that using a simple machine learning method to predict post stroke readmission, (1) could help to identify patients at risk of readmission before they are discharged to improve management of their post-acute care, and (2) can help health policy makers address high readmission rates through utilizing the models as a tool to evaluate hospital specific performance. Finally, we

provided additional evidence that being discharged to inpatient rehabilitation facility versus skilled nursing facility is associated with better functional (i.e., home time) and survival over 1 year of follow up which will (1) help clinicians in their complex and subjective decision to determine which rehabilitation setting will maximize patients odds of optimum recovery, and (2) help policy makers in incentivizing discharge to certain rehabilitation destinations, amending current health coverage laws to improve access, or introduce changes to the current facilities to improve patient outcomes.

6.4 Conclusions

In this dissertation, we generated a linked dataset that permitted the assessment of long-term (up to 1-year) outcomes following hospitalization for acute stroke. We probabilistically linked data between MiSP acute stroke registry and MVC claims database using indirect identifiers and produced a valid linked dataset that has acceptable representation of Medicare FFS and BCBSM insured population in Michigan. The stroke outcomes data generated up to 1-year post discharge using the linked dataset were similar to previously published literature in the US. Further, we found that very small number of previous linkage studies conducted a thorough evaluation of their linkages, hence the detailed linkage evaluation steps and techniques presented in this dissertation can serve as an example to guide future linkages studies using GWTG-S data (or other stroke registries) with claims data.

We also concluded that simple predictive modelling methods like LASSO logistic regression produced similar prediction accuracy values when compared to more advanced ML methods including XGBoost and ANN. My analysis also demonstrated that claims data can also be used alone to predict readmission rates with similar predictive accuracy as registry-based models. Moreover, the patient's clinical history prior to stroke, particularly chronic renal failure,

atrial fibrillation, heart failure and hospitalization including admission duration and discharge destination were found to be of higher importance when predicting long term readmission compared to clinical features of stroke such as NIHSS and stroke etiology. Those findings indicate that adequate post-acute care including adequate post-acute primary and neurology care follow up and rehabilitation can likely contribute to lowering the probability of readmission.

Finally, we provided further evidence that discharge to the IRF versus SNF is associated with better adjusted functional outcomes and survival during follow up among Medicare FFS stroke patients. My detailed approach of using and evaluating weighted outcomes in this retrospective observation study design and the nature of our rich linked dataset compared to the previous studies that did not report on their weighting approach and mainly utilized claims data, delivered stronger evidence needed to better guide the current complex nature of the clinical decision to discharge a stroke patient to an IRF or SNF rehabilitation setting to maximize their odds of functional recovery.