Journal of the American Heart Association

ORIGINAL RESEARCH

# Claims-Based Machine Learning Classifier of Modified Rankin Scale in Acute Ischemic Stroke

Mamoon Habib, MSc; Rafaella Cazé de Medeiros, MD; Syed Muhammad Ahsan, MBBS, MD; Aidan McDonald Wojciechowski, BS; Maria A. Donahue, MD; Deborah Blacker, MD, ScD; Joseph P. Newhouse, PhD; Lee H. Schwamm, MD; M. Brandon Westover, MD, PhD*; Lidia M. V. R. Moura, MD, PhD, MPH*

**BACKGROUND:** We developed a classifier to infer acute ischemic stroke severity from Medicare claims using the modified Rankin Scale at discharge. The classifier can be used to improve stroke outcomes research and support the development of national surveillance tools.

**METHODS:** This multistate study included all participating centers in the Paul Coverdell National Acute Stroke Program database from 9 US states. This database was linked to Medicare data sets for patients hospitalized with acute ischemic stroke, employing demographics, admission details, and diagnosis codes to create unique patient matches. We included Medicare beneficiaries aged 65 and older who were hospitalized for an initial acute ischemic stroke from January 2018 to December 2020. Using Lasso-penalized logistic regression, we developed and validated a binary classifier for modified Rankin Scale outcomes and as a secondary analysis we used ordinal regression to model the full modified Rankin Scale. Performance was evaluated on held-out test data using area under the receiver operator characteristic curve, receiver operator characteristic precision-recall, sensitivity, and specificity.

**RESULTS:** We analyzed data from 68 636 eligible patients. The mean age was 79.5 years old. Seventy-seven and a half percent of beneficiaries were White, 14% were Black, 2.6% were Asian, and 2% were Hispanic. The classifier achieved an area under the receiver operator characteristic curve score of 0.86 (95% CI, 0.85–0.86), sensitivity of 0.81 (95% CI, 0.80–0.81), specificity of 0.73 (95% CI, 0.72–0.74), and precision-recall area under the curve of 0.90 (95% CI, 0.90–0.91) on the test set.

**CONCLUSIONS:** Among Medicare beneficiaries hospitalized for acute ischemic stroke, the claims-based classifier demonstrated excellent performance in area under the receiver operator characteristic curve, precision-recall area under the curve, sensitivity, and acceptable specificity for modified Rankin Scale classification.

**Key Words:** acute ischemic stroke ■ classifier ■ Medicare ■ modified Rankin Scale ■ Paul Coverdell National Acute Stroke Program

Every 40 seconds, someone in the United States has a stroke.[1] Stroke is one of the leading causes of long-term disability, affecting about 795 000 people in the United States annually.[2] Acute ischemic stroke (AIS) severity can be variable, with a significant portion of discharged patients presenting with declining functionality, leading to increased needs for rehabilitation and admission to nursing facilities.[3] Both

Correspondence to: Lidia M. V. R. Moura, MD, PhD, MPH, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Wang 739D, Boston, MA 02114. Email: lidia.moura@mgh.harvard.edu

*M. Brandon Westover and L. M. V. R. Moura are co-senior authors.

This article was sent to Jose Rafael Romero, MD, Associate Editor, for review by expert referees, editorial decision, and final disposition.

Preprint posted on MedRxiv February 11, 2025. doi: https://doi.org/10.1101/2025.02.06.25321827

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.125.041635

For Sources of Funding and Disclosures, see page 10.



*JAHA* is available at: www.ahajournals.org/journal/jaha

## CLINICAL PERSPECTIVE

### What Is New?

- A novel claims-based classifier to infer acute ischemic stroke severity using the modified Rankin Scale at discharge was developed.
- Medicare claims were integrated with clinical data from the stroke registry using penalized logistic regression for both binary and ordinal classification.

### What Are the Clinical Implications?

- The classifier provides a robust tool for assessing stroke severity, which can enhance stroke outcomes research and quality improvement initiatives, and supports the development of national surveillance tools, guiding resource allocation in stroke care.

### Nonstandard Abbreviations and Acronyms

| | |
|---|---|
| **AIS** | acute ischemic stroke |
| **MEDPAR** | Medicare Provider Analysis and Review |
| **PCNASP** | Paul Coverdell National Acute Stroke Program |

modifiable (ie, obesity, diabetes, cardiovascular disease, certain medications, physical inactivity, etc.) and nonmodifiable stroke risk factors (ie, age, sex, race, ethnicity, genetics) can help determine disease severity and patient prognosis.[4]

Functional outcome prediction in AIS affects the quality of patient care decisions.[5,6] Recent advances in computational and software technologies have greatly influenced the rise of machine learning (ML) studies, offering more precise outcome measures.[7–9] ML models have identified several crucial factors to predict and classify functional outcomes, such as an initial National Institutes of Health Stroke Scale (NIHSS) score, age, fasting blood glucose, creatinine levels, and the modified Rankin Scale (mRS).[10,11] mRS has been widely used to assess AIS severity and clinical prognosis in electronic health records and registries.[12] The creation of models and classifiers can be personalized to assess outcomes in patients with AIS, including the classification of mRS.[8,9,13] However, limited valid measures of stroke severity have hindered national, large-scale, claims-based studies.[14]

Despite this limitation, claims data may offer indirect clues about a patient's level of disability based on the types of claims filed.[15,16] Leveraging a data set that links claims to mRS scores, we explored whether supervised ML could develop a classifier to infer mRS from claims information. Such a model could enable the personalization of outcome assessments for patients with AIS and the classification of mRS in large claims-based studies, thereby configuring a tool for national surveillance of stroke severity.

We linked the Paul Coverdell National Acute Stroke Program (PCNASP) and Medicare claims-based inpatient data of older adults presenting with AIS to develop and validate the mRS classifier of stroke severity at discharge.

## METHODS

The Medicare data supporting this study's findings are retrospective, were routinely collected by The Centers for Medicare & Medicaid Services for billing purposes and were made available with no direct identifiers. The requirement for informed consent was waived, as this research posed minimal risk to patients, would not affect clinical decisions about the individuals care, and could not be practicably carried out without the waiver of consent. All results were aggregated following Centers for Medicare & Medicaid Services Cell Suppression Policies. Restrictions apply to the availability of these data, which were used under license for this study. Medicare data are available through the Centers for Medicare & Medicaid Services with their permission. The materials for replication of the classifiers are included in Data S1. PCNASP data are available through the Centers for Disease Control and Prevention with their permission.

This study was approved by the Mass General Brigham Institutional Review Board's ethical guidelines and followed the Strengthening the Reporting of Observational Studies in Epidemiology (Table S1) guidelines for observational studies[17] and the transparent reporting of multivariable prediction models developed or validated using clustered data[18] (Table S2) and the updated guidance for reporting clinical prediction models that use regression or ML methods (transparent reporting of multivariable prediction models developed or validated using clustered data-artificial intelligence) (Table S3).[19]

### Study Design

We conducted a retrospective analysis of claims data from patients with AIS using a sample from 9 large US states. We aimed to develop and validate a classifier based on claims data that infers mRS at discharge.

### Data Source

We accessed data from the PCNASP registry and Medicare Claims data. PCNASP collects data on stroke cases and captures discharge mRS scores reported by

clinicians or hospital staff.[20] The PCNASP registry includes information from 2008 to 2020 from the following states: California, Georgia, Massachusetts, Michigan, Minnesota, New York, Ohio, Washington, and Wisconsin.

We then matched the PCNASP data on individuals aged 65 or older with data from fee-for-service Medicare, a national health insurance program administered by Centers for Medicare & Medicaid Services.[21] The Medicare Provider Analysis and Review (MEDPAR) files contain extensive information about these beneficiaries, including patient demographics, admission and discharge dates, diagnosis, procedure codes, provider identifiers, and comorbidities.[22]

## Study Population

We analyzed Medicare claims data for beneficiaries aged 65 and older hospitalized for AIS from January 2018 to December 2020. We included beneficiaries who were enrolled in traditional Medicare Part A (inpatient hospital insurance; care in a skilled nursing facility, hospice care, and some home health care) and Part B (physician and other medical provider services; outpatient care, medical supplies, and preventive services) who had mRS values documented in the PCNASP clinical database.

We used a multistep exclusion and inclusion process to refine our patient population. First, we excluded patients with missing mRS scores in the PCNASP patient-level data and then linked the remaining data with Medicare claims data. We found patients with a diagnosis of AIS in the Medicare claims data during 2018 to 2020 and used only their first stroke encounter. We identified a beneficiaries' first episode of stroke based on the first event date in the study period. Next, we created 2 groups based on the availability of an mRS score for any stroke on hospital-level data (Figure 1). The first group included patients admitted to the hospital with a≥90% or more completion rate of mRS, and the second group included patients admitted to hospitals with <90% of mRS completion. We used 20% of the first group and all of the second group as a training sample; the remaining 80% of the first group was set aside as an independent test sample.

## Linking Databases

Because there were no unique patient identifiers common to both databases, we applied a matching strategy to link

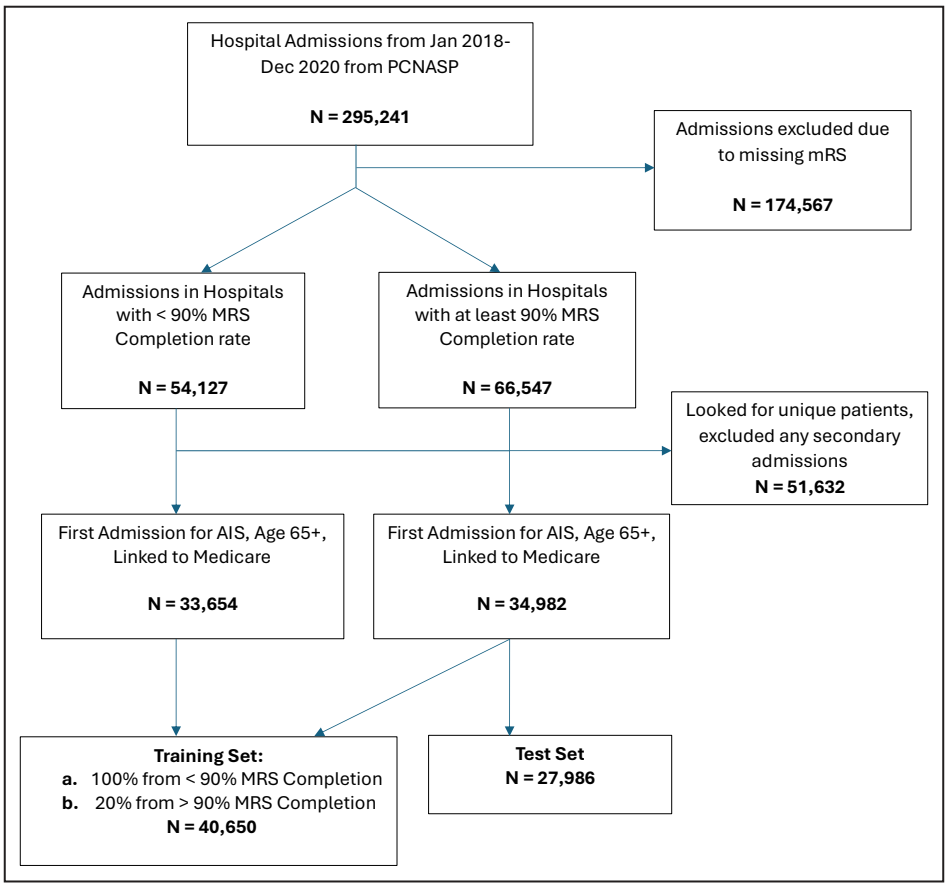


**Figure 1. Inclusion and exclusion criteria chart.**
A consort diagram of our inclusion/exclusion criteria and model development. AIS indicates acute ischemic stroke; MRS, modified Rankin Scale; and PCNASP, Paul-Coverdell National Acute Stroke Program.

individuals in the PCNASP and Medicare data sets.[23] For this linkage we used variables such as age, sex, admission and discharge dates, diagnosis code, hospitals, and state. After linkage, we retained patients with unique matches, excluding cases where PCNASP IDs corresponded to multiple Medicare Beneficiary IDs and vice versa. Due to limited access to baseline institutionalized (nonoutpatient) data, we excluded patients transferred from another hospital, skilled nursing facility, or other health care facilities.

## Variables

We included demographic variables, medical history, treatments, and discharge outcomes. Most variables were extracted from the MEDPAR files. Those not included in MEDPAR were extracted from hospital-level data by linking MEDPAR data with provider-level data and included variables such as bed size and hospital location, category, and level. We included 2 stroke-related variables for inpatient conditions and procedures such as tissue plasminogen activator and endovascular treatment. We used the value "1" if the condition or procedure was present and the value "0" if not. For continuous variables such as age and length of stay, we standardized their values using a normalization process.[24] Categorical variables, such as race and admission type, were converted into dummy variables for use in the model. We used the variables included in the Chronic Conditions Warehouse from Medicare, which contains algorithms for 27 chronic conditions to determine comorbidities and relevant patient medical history in our patient population (Table S4).[22] We selected the first-ever criteria a beneficiary met for the chronic condition.

## Construct of Interest (End Point)

Our primary end point was the accurate classification of mRS score at discharge. We dichotomized the mRS into "favorable" if valued as equal or less than 2 (from no symptoms to slight disabilities) and "unfavorable" if the mRS score was >2 (interval from moderate disability to death).[12,25]

As a secondary analysis, we developed ordinal classifiers using the previous sampling approach to obtain more granularity among mRS categories. The 2 approaches of ordinal classification consist of a full mRS, one represented by 0: no symptoms; 1: no significant disabilities, despite symptoms; 2: slight disabilities; 3: moderate disability; 4: moderate to severe disability; 5: severe disability; 6: death.[23,24] The second ordinal model consists of the same full scale but excludes the death category.

## Statistical Analysis

### *Primary Analysis: Binary Classifier*

The binary classifier outputs probabilities for each class. A threshold of 0.5 was used to convert the probabilities into binary values. Predictions with a probability ≥0.5 were assigned to the unfavorable mRS category and those <0.5 to the favorable class.

For development of our binary classifier, binary logistic regression with a lasso penalty was trained to predict the binary mRS category (favorable versus unfavorable). The best hyperparameters were determined through a grid-search hyperparameter tuning process. The hyperparameters included a range of the inverse regularization strength C ($10^{-4}$ to 100), tolerance values (1e-4 to 1e-1), maximum iterations (5000 to 50 000), solver methods ('liblinear' and 'saga'), and class weight settings (none and balanced). The hyperparameters that generated the largest area under the receiver operator characteristic curve (ROC AUC) were chosen. Stratified 5-fold cross-validation was used to evaluate the classifier's performance within the training set. The model was separately evaluated on the test set, which was not used in model development. Additionally, we reported the performance metrics for various patient subsets, stratifying them by race, both independently and in groups, as well as ethnicity, sex, and hospital location (Table S5). We report the performance of the model including *International Classification of Diseases, Tenth Revision* (*ICD-10*) NIHSS score as a feature (Table S6).

### *Secondary Analysis: Ordinal Classifier*

We also trained a classifier on the full-scale mRS values using ordinal regression. The ordinal regression model outputs probabilities for each class. To assign class labels, we selected the class with the maximum predicted probability.

We fitted the model as a parallel classifier with a logit link and Lasso L1 penalty using the ordinalNet R package. Grid-search hyperparameter tuning was performed on the training data set to select the best model based on lambda and family values. We defined a sequence of lambda values (ranging from 0.001 to 0.01) and multiple family values (cumulative, acat, sratio, cratio).

For each family type in the classifier, models were fitted across a range of lambda values and log-likelihood was used to evaluate model performance. The optimal lambda for each family type was selected as the value that achieved the highest log-likelihood, once we selected the optimal family type and lambda value, we refitted the final classifier on the training data with the chosen parameters. We tested the refitted model on the test data set to check for its generalizability.

## Performance Metrics

For both primary and secondary analyses, we evaluated classifier's performance using ROC AUC and area

under the precision-recall curve to assess the model's ability to distinguish between classes. Sensitivity and specificity were included to evaluate the model's ability in identifying true positives and true negatives.

To calculate CIs for our performance metrics, we performed 10 000 iterations of bootstrap random sampling with replacement in each iteration. We created a distribution for each metric and calculated 95% CI to show the classifier's performance variability.

# RESULTS

## Characteristics of the Samples

We assessed 295 241 hospital admissions for AIS between January 2018 and December 2020 for eligibility. After applying the inclusion and exclusion criteria, the sample included 68 636 unique Medicare beneficiaries who were 65 years old or older with a first admission for AIS and available discharge mRS scores. We obtained distinctive patient hospital encounters with < or ≥90% completion of the mRS (N=33 654 and N=34 982, respectively) (Figure 1).

The mean age for the full sample was 79.53 (SD 8.7), and 77.5% of beneficiaries were White, 14% were Black, 2.7% were Asian, and 2% were Hispanic (Table 1). The mean age for the test data was 79.76 (SD 8.7). Approximately 91% of our patient sample was admitted through emergency care. Regarding discharge disposition, the test set data were more evenly distributed between home, skilled nursing facilities, and inpatient rehabilitation facilities with 28%, 23%, and 19%, respectively. The remaining percentage was distributed between approximately 100 other discharge disposition variables, according to the MEDPAR files code.[26] The test set was also evenly distributed for patient interventions, such as receipt of tissue plasminogen activator and endovascular intervention. Concerning comorbidities, 71% of beneficiaries had hypertension, 39% diabetes, and 29% congestive heart failure. A further breakdown of the full sample, training, and test set demographics can be found in Table 1. We used 63 covariates to predict a scale score, such as demographics, medical history, treatments, and discharge outcomes (a list can be found in Figure 2 and Table S7).

## Binary Classifier

On the held-out test data, our binary classifier achieved an ROC AUC score of 0.86 (95% CI, 0.85–0.86, Figure 3), sensitivity of 0.81 (95% CI, 0.80–0.81), specificity of 0.73 (95% CI, 0.72–0.74), and area under the precision-recall curve of 0.91 (95% CI, 0.90–0.91, Figure 4). Figure 1 shows the model's feature coefficients sorted/ranked by their contribution to its predictions. The best selected hyperparameters are in Table S8. Palliative care was the strongest predictor (2.02) of unfavorable mRS outcomes. Similarly, coded hemiplegia (0.71), and the use of ventilator during the AIS hospitalization (0.61) were strong predictors of unfavorable outcomes. Several features were also associated with a lower likelihood of unfavorable outcomes. For instance, binary discharge disposition (home versus others) had the strongest negative coefficient (–1.95), suggesting that favorable discharge outcomes strongly predict better recovery. Transesophageal echocardiogram (–0.31) and tissue plasminogen activator administration (–0.25) were associated with favorable outcomes.

## Ordinal Classifier

For our secondary analysis, the ordinal model's overall performance on the test data is presented in Table 2. The best hyperparameters are shown in Table S9. The model demonstrates a stronger ability to distinguish between mRS scores 0 (no symptoms) and 5/6 (severe disability/death) compared with its performance in differentiating intermediate outcomes (1–4) (see Figures S1 and S2).

Classes 2 (slight disability) and 3 (moderate disability) showed the lowest ROC AUC and area under the precision-recall curve scores. Figure S3 presents a box plot of grouped probabilities, highlighting how the model conflates mRS scores 2 and 3 with mRS score 4. The model's ability to distinguish between mRS scores 0 (no symptoms) and 5/6 (severe disability/death) is higher compared with its performance in differentiating intermediate outcomes (1–4) (see Figure S1).

Additionally, we excluded death to evaluate whether the model's performance improves in predicting intermediate outcomes 2 and 3; however, no significant changes in performance were observed. The coefficients from both ordinal models (Tables S10 and S11) were consistent with those observed in the binary model. For instance, in the full-scale mRS ordinal model, discharge disposition (ie, discharged home; coefficient=1.99) increased the odds of falling into a lower (better) mRS category, whereas palliative care (coefficient=–2.72) increased the odds of a higher (worse) category.

# DISCUSSION

Considering the clinical burden of AIS and its influence on patient mortality, rate of disability, medical complications, and health care expenditures, it is fundamental to monitor the impact, severity, and prognosis of this condition.[1,27,28] Our interpretation of the identified factors driving the classification highlights their strong face validity and consistency with existing literature as they align with clinical

**Table 1. Demographic Characteristics**

| Characteristics | Full sample (N=68 636) | Training/validation (n=40 650) | Test (n=27 986) |
|---|---|---|---|
| Age, y, mean±SD | 79.53±8.67 | 79.38±8.63 | 79.76±8.71 |
| Sex (%) | | | |
| Female | 37 439 (54.54) | 22 045 (54.23) | 15 394 (55.00) |
| Male | 31 197 (45.45) | 18 605 (45.76) | 12 592 (45.00) |
| Race (%) | | | |
| White | 53 192 (77.49) | 31 794 (78.21) | 21 398 (76.45) |
| Black | 9629 (14.02) | 5394 (13.26) | 4235 (15.13) |
| Asian | 1821 (2.65) | 1146 (2.81) | 675 (2.41) |
| Hispanic | 1361 (1.98) | 753 (1.85) | 608 (2.17) |
| Other | 1483 (2.16) | 876 (2.15) | 607 (2.16) |
| Unknown | 997 (1.45) | 593 (1.45) | 404 (1.44) |
| North American Native | 153 (0.22) | 94 (0.23) | 59 (0.21) |
| Admission type (%) | | | |
| Emergency | 62 639 (91.26) | 36 657 (90.18) | 25 982 (92.83) |
| Urgent | 4911 (7.15) | 3375 (8.30) | 1536 (5.48) |
| Trauma center | 559 (0.81) | 326 (0.80) | 233 (0.83) |
| ICU type (%) | | | |
| Intermediate ICU | 13 325 (19.41) | 8379 (20.61) | 4946 (17.6) |
| General | 11 569 (16.85) | 6786 (16.69) | 4783 (17.09) |
| Medical | 3033 (4.41) | 1599 (3.93) | 1434 (5.12) |
| Surgical | 1501 (2.18) | 1073 (2.63) | 428 (1.52) |
| Trauma | 153 (0.22) | 117 (0.28) | 36 (0.12) |
| Other | 144 (0.20) | 63 (0.15) | 81 (0.28) |
| Discharge disposition (%)* | | | |
| Home/self-care | 18 931 (27.58) | 11 233 (27.63) | 7698 (27.50) |
| Skilled nursing facility | 15 426 (22.47) | 9056 (22.27) | 6370 (22.76) |
| Inpatient rehabilitation facility | 12 856 (18.73) | 7213 (18.67) | 5266 (18.81) |
| Interventions (%) | | | |
| Tissue plasminogen activator | 9001 (13.11) | 5579 (13.72) | 3422 (12.22) |
| Endovascular intervention | 3089 (4.50) | 1780 (4.37) | 1309 (4.67) |
| Comorbidities (%) | | | |
| Acute myocardial infarction | 4290 (6.25) | 2544 (6.26) | 1746 (6.24) |
| Atrial fibrillation | 13 304 (19.38) | 7700 (18.94) | 5604 (20.02) |
| Diabetes | 26 708 (38.91) | 15 581 (38.33) | 11 127 (39.76) |
| Congestive heart failure | 19 766 (28.80) | 11 555 (28.43) | 8211 (29.34) |
| Hypertension | 48 451 (70.59) | 28 418 (69.91) | 20 033 (71.58) |

Baseline demographics, admission type, ICU type, and comorbidities stratified by sample, training, and test groups. ICU indicates intensive care unit.
*We did not include all discharge disposition variable in the table, as there are >100 existing items. We reported the most relevant ones in this table.

expectations and prior studies. Indirect measures of stroke severity such as palliative care, hemiplegia, endotracheal intubation, and feeding device usage were strong predictors of unfavorable mRS outcomes, which is consistent with established knowledge on poor prognostic factors in AIS. Similarly, favorable discharge disposition (eg, discharged home), tissue plasminogen activator administration and brain imaging (computed tomography or magnetic resonance imaging) were associated with better outcomes, reinforcing the importance of early and effective stroke management.

We developed and validated a claims-based classifier to accurately identify stroke severity measured by mRS at discharge in patients aged 65 or older who experienced AIS. By leveraging administrative claims data, our classifier demonstrated strong predictive performance for categorizing stroke severity. This tool holds significant potential for facilitating large-scale research on stroke outcomes and

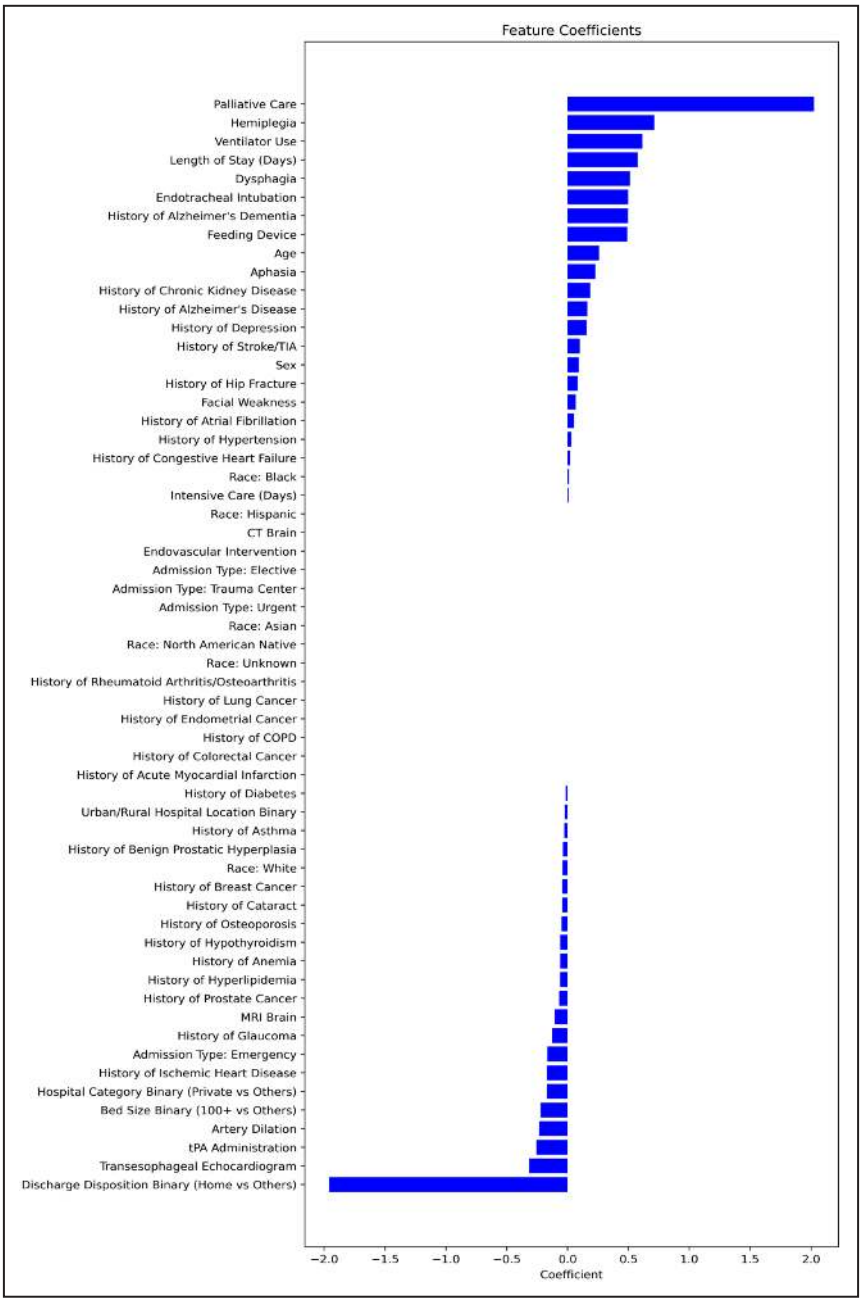


**Figure 2. Model features.**
The full list of the classifier's features and their coefficient values. COPD indicates chronic obstructive pulmonary disease; CT, computed tomography; MRI, magnetic resonance imaging; TIA, transient ischemic attack; and tPA, tissue plasminogen activator.

improving national surveillance efforts, enabling more effective monitoring of stroke care quality and recovery outcomes. Validated claims-based classifiers for AIS surveillance are also important for observing geographic trends and are essential for population health research, which in turn can inform public health policy and national guidelines to improve clinical practice.[3,29]

Previous studies have used ML methods for stroke functional outcome assessment.[5,13,30] Joon Nyung Heo et al. measured mRS scores 90 days after hospital discharge using 3 learning algorithm models: deep neural network, random forest, and logistic regression. The study had similar results with the logistic regression model (AUC 0.85), and the best performance was by the deep neural network model (AUC 0.88)[30]

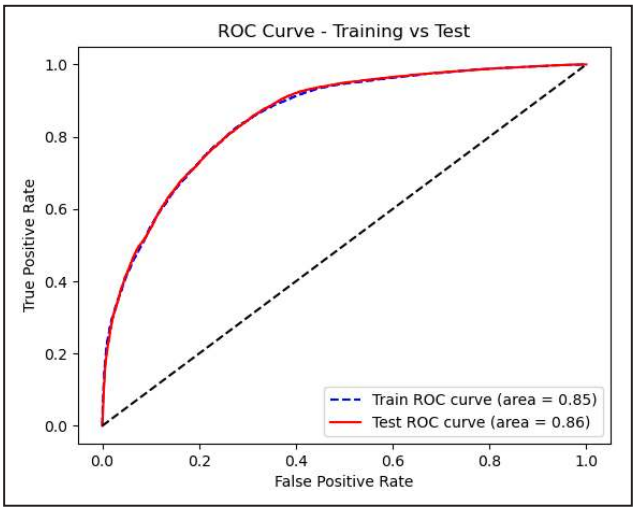


**Figure 3. ROC curves training versus test.**
Comparison of the ROC in both the training and test sets of the classifier. ROC indicates receiver operating characteristic.

In our study, logistic regression for mRS classification at discharge yielded positive results with the ROC AUC score of 0.86, reiterating the results seen in other models.[5,13,27,30]

Most important, the previous studies were limited by selection bias due to their sampling from single regions of the United States.[5,13,27,30] Our study overcomes this challenge by including a national, large-scale sample with representation of patients and practices from 9 US states spanning all regions of the United States. Therefore, our cohort provides a more robust, inclusive, and representative claims-based classifier for beneficiaries with AIS than has been heretofore available.

Prior studies creating mRS stroke severity classifiers used a random assignment approach within hospitals to create training and test sets.[9,13] This approach is potentially biased because random sampling does not account for hospital-level patterns in patient intake and reporting. We addressed this by categorizing the training and test data sets depending on whether hospitals reported < or ≥90% mRS completion. We used data only from those with ≥90% mRS completion as the test set, with a random 20% allocated to the training set for representativeness, allowing the classifier to be trained and tested with higher-quality data and partially accounting for potential bias in random sampling.

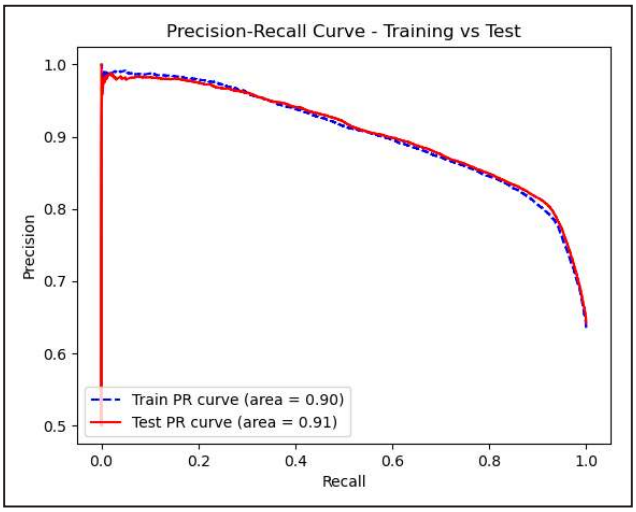


**Figure 4. Precision-recall area under the curve for binary classifier.**
Comparison of precision-recall curve of the classifier in the training and test sets. PR indicates precision-recall.

**Table 2. Full-Scale Ordinal Model Performance**

| Metric | Score [CI] |
|---|---|
| ROC AUC | 0.81 [0.80–0.81] |
| Precision-recall AUC | 0.39 [0.37–0.39] |
| Sensitivity | 0.42 [0.41–0.42] |
| Specificity | 0.89 [0.88–0.89] |

Performance metrics from full-scale ordinal. We report micro-average ROC AUC and precision-recall AUC. ROC AUC indicates receiver operating characteristic curve area under the curve.

Furthermore, our study used binary and ordinal regression methods to classify the mRS score in patients with AIS. To assess potential bias, we conducted additional analyses for multiple subgroups where the model performance was consistent, indicating the model's robustness and generalizability. Although binary analyses yield results that are more easily interpreted by examining the absolute risk reduction between the 2 groups, they do not account for within-group variation.[25] We therefore implemented an ordinal approach to achieve better use of the data.[25,31] The use of the ordinal method increased statistical power and decreased loss of information when compared with previous studies.[5,32]

Other research groups have focused on validating admission stroke severity, such as electronic health record-based classifiers of NIHSS score at admission.[33] This is important work, as classifiers of stroke severity at admission can inform resource allocation while patients are admitted and guide other care measures. However, our study focused on leveraging claims data to classify stroke functional outcomes at discharge using the mRS. The mRS is important because it provides information on patient functional outcomes, which can inform the prioritization of postdischarge stroke care allocation and predictions of long-term outcomes, among other applications.[34,35] The score's ability to predict the level of functionality makes it an essential tool for national-level surveillance using administrative databases.[5]

## Limitations

Although we used a nationally representative stroke registry covering 9 US states and its major stroke centers linked to administrative claims data, results may not be generalized to states not included in our data or smaller community health care centers. In addition, our selection of older adults ≥65 covered by fee-for-service Medicare may not represent other patient populations. Slightly over half of eligible Medicare beneficiaries are now enrolled in Medicare Advantage Part C instead of traditional Medicare. Beneficiaries must also be enrolled in Parts A and B, as well as Part B premium. Recent studies have shown that enrollment in lower-cost Medicare Advantage plans has increased among low-income and racial and ethnic minority groups.[36] Future studies assessing these groups would benefit these populations.

In our study, the choice of mRS as the scale to determine stroke severity can have limitations. For example, a patient's mobility level may directly influence a higher score. This factor, however, may not be related to an episode of stroke but to previous comorbidities such as arthritis, congestive heart failure, and others. This might yield an incorrect impression of increased stroke severity. We acknowledge that the NIHSS is a well-established predictor of functional outcomes, and for that reason, we included the mean and median of *ICD-10* NIHSS scores in Table S12 and not as a covariate in the main analysis, due to data missingness.

We excluded 12 894 patients transferred from another hospital, skilled nursing facility, or other health care facilities from the analytical sample, which may have omitted a subset of the population with AIS with a higher burden of baseline comorbidities. We selected this approach due to limited access to predictor data from these groups. Including these patients could have enhanced classifier representativeness and performance by increasing the sample size and introducing greater variability. Nevertheless, our classifier demonstrated high performance while capturing a broad and still nationally representative segment of the population with AIS.

We were limited by data availability for the Medicare and PCNASP data sets. Although use of administrative claims linked to data registries represents a vast source of information for research purposes,[37] some inherent limitations (eg, human-type errors of scores and clinical scales, and missing data) (eg, missing mRS scores and other stroke-related variables) are surely present. Medicare data do not contain a separate variable for ethnicity. Instead, the variable Hispanic is included as a variable within the race category. Therefore, in our analysis of Hispanic versus Non-Hispanic, patients who identified as the races White, Black, Asian, and others were assumed to be non-Hispanic. This may lead to misinterpretation of patient diversity. Despite these limitations, national administrative claims data remain valuable in representing large-sized populations and their reflections.[38,39] Particularly, the choice of PCNASP and Medicare allowed the use of both patient and hospital-level data, which contributes valuable information from current patients with stroke to future patients with stroke.[23] This approach is generalizable to other stroke databases that can be linked with Medicare data.[40]

Furthermore, the chosen period for this study (January 2018 to December 2020) was based on the access to enough baseline patient data, as well as the inclusion of *ICD-10* codes during the linkage process between PCNASP and Medicare. This period still allows for generalizability of our model and future studies could benefit from addressing future data from 2021 and forward.

Lastly, the replicability of our classifier can present some challenges, for example, requiring at least 2 databases to perform linkage of common unique identifiers and extract multiple variables. Users looking to replicate should have experience in Python and R Programming and can refer to Data S1 for replication.

## CONCLUSIONS

We developed a claims-based classifier to identify stroke severity in patients with AIS using discharge mRS score. Importantly, we partially addressed potential bias by accounting for hospital-level patterns in sampling using mRS completion rates. Our classifier has expanded on previous research by using PCNASP and Medicare-linked data from several states to assess stroke severity.

## ARTICLE INFORMATION

Received February 6, 2025; accepted August 7, 2025.

### Affiliations

Department of Neurology (M.H., R.C.M., S.M.A., A.M.W., M.A.D., L.M.V.R.M.) and Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA (D.B.), Department of Health Care Policy, Harvard Medical School, Boston, MA (J.P.N.); Department of Epidemiology (D.B.) and Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Boston, MA (J.P.N.); Harvard Kennedy School, Cambridge, MA (J.P.N.); Yale School of Medicine, New Haven, CT (L.H.S.); Department of Neurology, Beth Israel Lahey Health Medical System, Boston, MA (M.B.W.); and National Bureau of Economic Research, Cambridge, MA (J.P.N.).

### Sources of Funding

This study was supported by the National Institutes of Health (1R01AG073410).

### Disclosures

Mamoon Habib, Rafaella Cazé de Medeiros, Syed Muhammad Ahsan, Aidan McDonald Wojciechowski, Maria A. Donahue, and Joseph P. Newhouse have no conflict of interest to disclose. Deborah Blacker receives support from the National Institute of Neurological Disorders and Stroke and the National Institute on Aging and reports no conflict of interest. M. Brandon Westover is a cofounder, scientific advisor, and consultant to Beacon Biosignals and has a personal equity interest in the company. Lee H. Schwamm receives research support from the National Institute of Neurological Disorders and Stroke and the National Institute on Aging and reports no conflict of interest. Lidia M. V. R. Moura receives research support from the Epilepsy Foundation of America, the National Institute of Neurological Disorders and Stroke, and the National Institute on Aging and reports no conflict of interest.

### Supplemental Material

Data S1
Tables S1–S12
Figures S1–S3

## REFERENCES

1. Tsao CW, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, Baker-Smith CM, Beaton AZ, Boehme AK, Buxton AE, et al. Heart disease and stroke statistics—2023 update: a report from the American Heart Association. *Circulation*. 2023;147:e93–e621. doi: 10.1161/CIR.0000000000001123
2. CDC. Stroke Facts. Stroke. October 24, 2024. Accessed November 22, 2024. https://www.cdc.gov/stroke/data-research/facts-stats/index.html
3. Ziaeian B, Xu H, Matsouaka RA, Xian Y, Khan Y, Schwamm LS, Smith EE, Fonarow GC. US surveillance of acute ischemic stroke patient characteristics, care quality, and outcomes for 2019. *Stroke*. 2022;53:3386–3393. doi: 10.1161/STROKEAHA.122.039098
4. Boehme AK, Esenwa C, Elkind MSV. Stroke risk factors, genetics, and prevention. *Circ Res*. 2017;120:472–495. doi: 10.1161/CIRCRESAHA.116.308398
5. Zhang MY, Mlynash M, Sainani KL, Albers GW, Lansberg MG. Ordinal prediction model of 90-day modified Rankin Scale in ischemic stroke. *Front Neurol*. 2021;12:12. doi: 10.3389/fneur.2021.727171
6. Fonarow GC, Kapral MK, Schwamm LH. Future of quality and outcomes research in stroke. *Circ Cardiovasc Qual Outcomes*. 2015;8:S66–S68. doi: 10.1161/CIRCOUTCOMES.115.002309
7. Fralick M, Colak E, Mamdani M. Machine learning in medicine. *N Engl J Med*. 2019;380:2588–2590. doi: 10.1056/NEJMc1906060
8. Zihni E, McGarry B, Kelleher J. *Moving toward explainable decisions of artificial intelligence models for the prediction of functional outcomes of ischemic stroke patients*. Exon Publ. 2022:73–90. doi: 10.36255/exon-publications-digital-health-explainable-decisions
9. Su PY, Wei YC, Luo H, Liu CH, Huang WY, Chen KF, Lin CP, Wei HY, Lee TH. Machine learning models for predicting influential factors of early outcomes in acute ischemic stroke: registry-based study. *JMIR Med Inform*. 2022;10:e32508. doi: 10.2196/32508
10. Li X, Pan X, Jiang C, Wu MR, Liu YK, Wang FS, Zheng XH, Yang J, Sun C, Zhu YB, et al. Predicting 6-month unfavorable outcome of acute ischemic stroke using machine learning. *Front Neurol*. 2020;11:11. doi: 10.3389/fneur.2020.539509
11. Liu Y, Yu Y, Ouyang J, Jiang B, Yang G, Ostmeier S, Wintermark M, Michel P, Liebeskind DS, Lansberg MG, et al. Functional outcome prediction in acute ischemic stroke using a fused imaging and clinical deep learning model. *Stroke*. 2023;54:2316–2327. doi: 10.1161/STROKEAHA.123.044072
12. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials. *Stroke*. 2007;38:1091–1096. doi: 10.1161/01.STR.0000258355.23810.c6
13. Lee J, Park KM, Park S. Interpretable machine learning for prediction of clinical outcomes in acute ischemic stroke. *Front Neurol*. 2023;14:1234046. doi: 10.3389/fneur.2023.1234046
14. ElHabr AK, Katz JM, Wang J, Bastani M, Martinez G, Gribko M, Hughes DR, Sanelli P. Predicting 90-day modified Rankin Scale score with discharge information in acute ischaemic stroke patients following treatment. *BMJ Neurol Open*. 2021;3:e000177. doi: 10.1136/bmjno-2021-000177
15. Festa N, Price M, Moura LMVR, Blacker D, Normand SL, Newhouse JP, Hsu J. Evaluation of claims-based ascertainment of Alzheimer disease and related dementias across health care settings. *JAMA Health Forum*. 2022;3:e220653. doi: 10.1001/jamahealthforum.2022.0653
16. Festa N, Price M, Weiss M, Moura LMVR, Benson NM, Zafar S, Blacker D, Normand ST, Newhouse JP, Hsu J. Evaluating the accuracy of Medicare risk adjustment for Alzheimer's disease and related dementias. *Health Aff*. 2022;41:1324–1332. doi: 10.1377/hlthaff.2022.00185
17. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370:1453–1457. doi: 10.1016/S0140-6736(07)61602-X
18. Debray TPA, Collins GS, Riley RD, Snell KIE, van Calster B, Reitsma JB, Moons KGM. Transparent reporting of multivariable prediction models developed or validated using clustered data (TRIPOD-cluster): explanation and elaboration. *BMJ*. 2023;380:e071058. doi: 10.1136/bmj-2022-071058
19. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, van Calster B, Ghassemi M, Liu X, Reitsma JB, van Smeden M, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that

use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi: 10.1136/bmj-2023-078378
20. Taha M, Habib M, Lomachinsky V, Hadar P, Newhouse JP, Schwamm LH, Blacker D, Moura LMVR. Evaluating the concordance between International Classification of Diseases, Tenth Revision code and stroke severity as measured by the National Institutes of Health Stroke Scale. *BMJ Neurol Open*. 2024;6:e000831. doi: 10.1136/bmjno-2024-000831
21. Centers for Medicare and Medicaid Services. Centers for Medicare and Medicaid Services. Accessed January 7, 2025. https://www.cms.gov
22. Medicare Provider Analysis and Review (MedPAR). ResDAC (Research Data Assistance Center). Accessed January 7, 2025. https://resdac.org/cms-data/files/medpar
23. Patorno E, Schneeweiss S, George MG, Tong X, Franklin JM, Pawar A, Mogun H, Moura LMVR, Schwamm LH. 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;7:114–123. doi: 10.1136/svn-2021-001134
24. Bobbit, Z. Z-Score Normalization: Definition & Examples. August 2021. https://www.statology.org/z-score-normalization/
25. Ganesh A, Luengo-Fernandez R, Wharton RM, Rothwell PM, on behalf of the Oxford Vascular Study. Ordinal vs dichotomous analyses of modified Rankin Scale, 5-year outcome, and cost of stroke. *Neurology*. 2018;91:e1951–e1960. doi: 10.1212/WNL.0000000000006554
26. MEDPar. Destination upon discharge from facility code. ResDAC (Research Data Assistance Center). https://resdac.org/cms-data/variables/destination-upon-discharge-facility-code
27. Daidone M, Ferrantelli S, Tuttolomondo A. Machine learning applications in stroke medicine: advancements, challenges, and future prospectives. *Neural Regen Res*. 2024;19:769–773. doi: 10.4103/1673-5374.382228
28. Ovbiagele B, Goldstein LB, Higashida RT, Howard VJ, Johnston SC, Khavjou OA, Lackland DT, Lichtman JH, Mohl S, Sacco RL, et al. Forecasting the future of stroke in the United States: a policy statement from the American Heart Association and American Stroke Association. *Stroke*. 2013;44:2361–2375. doi: 10.1161/STR.0b013e31829734f2
29. Barnett ML, Linder JA, Clark CR, Sommers BD. Low-value medical Services in the Safety-net Population. *JAMA Intern Med*. 2017;177:829–837. doi: 10.1001/jamainternmed.2017.0401
30. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning–based model for prediction of outcomes in acute stroke. *Stroke*. 2019;50:1263–1265. doi: 10.1161/STROKEAHA.118.024293
31. Risselada R, Lingsma HF, Molyneux AJ, Kerr RSC, Yarnold J, Sneade M, Steyerberg EW, Sturkenboom MCJM. Prediction of two month modified Rankin Scale with an ordinal prediction model in patients with aneurysmal subarachnoid haemorrhage. *BMC Med Res Methodol*. 2010;10:86. doi: 10.1186/1471-2288-10-86
32. Roozenbeek B, Lingsma HF, Perel P, Edwards P, Roberts I, Murray GD, Maas AI, Steyerberg EW; IMPACT (International Mission on Prognosis and Clinical Trial Design in Traumatic Brain Injury) Study Group, CRASH (Corticosteroid Randomisation After Significant Head Injury) Trial Collaborators. The added value of ordinal analysis in clinical trials: an example in traumatic brain injury. *Crit Care*. 2011;15:R127. doi: 10.1186/cc10240
33. Brott T, Adams HP, Olinger CP, Marler JR, Barsan WG, Biller J, Spilker J, Holleran R, Eberle R, Hertzberg V. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke*. 1989;20:864–870. doi: 10.1161/01.STR.20.7.864
34. Rankin J. Cerebral vascular accidents in patients over the age of 60: II. Prognosis. *Scott Med J*. 1957;2:200–215. doi: 10.1177/003693305700200504
35. de Haan R, Limburg M, Bossuyt P, van der Meulen J, Aaronson N. The clinical meaning of Rankin 'handicap' grades after stroke. *Stroke*. 1995;26:2027–2030. doi: 10.1161/01.STR.26.11.2027
36. Meyers DJ, Mor V, Rahman M, Trivedi AN. Growth in Medicare advantage greatest among Black and Hispanic enrollees. *Health Aff (Millwood)*. 2021;40:945–950. doi: 10.1377/hlthaff.2021.00118
37. Fernandes M, Cardall A, Jing J, Ge W, Moura LMVR, Jacobs C, McGraw C, Zafar SF, Westover MB. Identification of patients with epilepsy using automated electronic health records phenotyping. *Epilepsia*. 2023;64:1472–1481. doi: 10.1111/epi.17589
38. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58:323–337. doi: 10.1016/j.jclinepi.2004.10.012
39. MacKay EJ, Stubna MD, Chivers C, Draugelis ME, Hanson WJ, Desai ND, Groeneveld PW. Application of machine learning approaches to administrative claims data to predict clinical outcomes in medical and surgical patient populations. *PLoS One*. 2021;16:e0252585. doi: 10.1371/journal.pone.0252585
40. Lichtman JH, Leifheit-Limson EC, Goldstein LB. Centers for Medicare and Medicaid Services Medicare data and stroke research: goldmine or landmine? *Stroke*. 2015;46:598–604. doi: 10.1161/STROKEAHA.114.003255