




## ORIGINAL RESEARCH

# Development and Validation of Prediction Models for Severe Complications After Acute Ischemic Stroke: A Study Based on the Stroke Registry of Northwestern Germany

Anna K. Bonkhoff , MD; Nicole Rüksamen, PhD; Christian Grefkes , MD; Natalia S. Rost , MD; Klaus Berger, MD; André Karch, MD

**BACKGROUND:** The treatment of stroke has been undergoing rapid changes. As treatment options progress, prediction of those under risk for complications becomes more important. Available models have, however, frequently been built based on data no longer representative of today's care, in particular with respect to acute stroke management. Our aim was to build and validate prediction models for 4 clinically important, severe outcomes after stroke.

**METHODS AND RESULTS:** We used German registry data from 152 710 patients with acute ischemic stroke obtained in 2016 (development) and 2017 (validation). We took into account potential predictors that were available at admission and focused on in-hospital mortality, intracranial mass effect, secondary intracerebral hemorrhage, and deep vein thrombosis as outcomes. Validation cohort prediction and calibration performances were assessed using the following 4 statistical approaches: logistic regression with backward selection,  $l_1$ -regularized logistic regression, k-nearest neighbor, and gradient boosting classifier. In-hospital mortality and intracranial mass effects could be predicted with high accuracy (both areas under the curve, 0.90 [95% CI, 0.90–0.90]), whereas the areas under the curve for intracerebral hemorrhage (0.80 [95% CI, 0.80–0.80]) and deep vein thrombosis (0.73 [95% CI, 0.73–0.73]) were considerably lower. Stroke severity was the overall most important predictor. Models based on gradient boosting achieved better performances than those based on logistic regression for all outcomes. However, area under the curve estimates differed by a maximum of 0.02.

**CONCLUSIONS:** We validated prediction models for 4 severe outcomes after acute ischemic stroke based on routinely collected, recent clinical data. Model performance was superior to previously proposed approaches. These predictions may help to identify patients at risk early after stroke and thus facilitate an individualized level of care.

**Key Words:** ischemic stroke ■ machine learning ■ mortality ■ prediction ■ severe outcomes

Stroke mortality has decreased globally in recent years.<sup>1</sup> Nonetheless, severe complications after stroke continue to reduce the chance of a good functional outcome, as they may delay and impede recovery after stroke.<sup>2</sup> Many research studies have focused on identifying those at highest risk for severe

complications after stroke in an aim to offer tailored therapies for individual patients with stroke and hence optimize clinical workflows.<sup>3</sup> These efforts have focused on the prediction of early mortality, intracranial mass effect, and secondary intracerebral hemorrhage as well as deep vein thrombosis (DVT; eg, in the lower

Correspondence: Anna K. Bonkhoff, MD, J. Philip Kistler Stroke Research Center, Massachusetts General Hospital, 175 Cambridge St, Suite 300, Boston, MA 02114. E-mail: abonkhoff@mgh.harvard.edu

Supplemental Material for this article is available at <https://www.ahajournals.org/doi/suppl/10.1161/JAHA.121.023175>

For Sources of Funding and Disclosures, see page 12.

© 2022 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

JAHA is available at: [www.ahajournals.org/journal/jaha](http://www.ahajournals.org/journal/jaha)

## CLINICAL PERSPECTIVE

### What Is New?

- The present study is the first to rely on a large stroke registry comprising recent clinical data on 152 710 patients with acute ischemic stroke and test the performances of several machine-learning algorithms in their prediction of 4 severe complications (in-hospital mortality, intracranial mass effects, secondary intracerebral hemorrhage, and deep vein thrombosis) after stroke.
- In-hospital mortality and intracranial mass effects in particular could be predicted with high accuracies and areas under the curve of  $\approx 0.90$ , with admission stroke severity being the most important predictor.
- Gradient boosting–based models were shown to consistently outperform logistic regression; however, differences in areas under the curve were small.

### What Are the Clinical Implications?

- Severe complications after stroke can be accurately predicted by employing routinely collected clinical data, as exemplarily recorded in stroke registries.
- Patients at risk for certain complications may be identified early after stroke by the model predictions presented here and receive an individualized level of care, for example, featuring specific additional diagnostic tests, such as venous Doppler of the legs.
- Validated models are openly available for the generation of predictions for new patients: [https://github.com/AnnaBonkhoff/Predict\\_severe\\_complications\\_after\\_stroke](https://github.com/AnnaBonkhoff/Predict_severe_complications_after_stroke)

## Nonstandard Abbreviations and Acronyms

**NIHSS** National Institutes of Health Stroke Scale

extremity) and provided promising results.<sup>4–7</sup> However, many of these currently available prediction models have been developed on stroke data gathered years ago or were trained using data from a larger time span to increase sample sizes. This circumstance may hamper a smooth implementation of the models in current clinical routine given that stroke care has changed considerably during the past 2 decades. These changes comprise the widespread use of intravenous thrombolysis and mechanical thrombectomy, as well as the extension of time windows after symptom onset when these acute therapies can be administered, to

name only a few examples.<sup>8,9</sup> Conceivably, all of these changes have a substantial influence on the occurrence of complications and their prediction.

We used recent data from a large German stroke registry<sup>10,11</sup> to predict intrahospital mortality, intracranial mass effect, secondary intracerebral hemorrhage, and DVT after ischemic stroke. To this end, we developed and validated 4 individual prediction models. In addition, we systematically evaluated if the use of advanced model-building strategies might improve the prediction accuracy.

## METHODS

### Data and Code Availability

Data analysis was conducted using jupyter notebooks in a python3.7 environment, particularly using implementations offered through the package scikit-learn.<sup>12</sup>

The code and trained models to generate outcome predictions for new patient data and a corresponding step-by-step documentation is openly available online: [https://github.com/AnnaBonkhoff/Predict\\_severe\\_complications\\_after\\_stroke](https://github.com/AnnaBonkhoff/Predict_severe_complications_after_stroke). The authors agree to make the data available to any researcher for the express purposes of reproducing the results presented here based on a written data transfer agreement and with the explicit permission for data sharing by the local institutional review board. The original data collection tool can be accessed at <https://www.medizin.uni-muenster.de/qsnwd/downloads.html> (“Spezifikationen”), whereas a version translated into English can be found in Data S1.

### Study Population

Data on patients with stroke originated from the Stroke Registry of Northwestern Germany. This registry has prospectively collected demographic and clinical characteristics of adult patients with stroke since the year 2000, consisted of a network of 155 hospitals at the time of data collection for this study, and has been described in detail previously.<sup>10,11</sup> Hospital participation in the stroke registry project is voluntary. Participation is, however, a mandatory prerequisite for stroke unit certification through the German Stroke Society, which serves as motivating factor. In this study, we included data on any patient hospitalized in 2016 or 2017 with ischemic stroke and the *International Classification of Diseases, Tenth Revision (ICD-10)* diagnosis code I63. Data from 2016 were used for model development, and data from 2017 were used for model validation. Registry data consisted of routine clinical stroke data, such as the National Institutes of Health Stroke Scale (NIHSS)–defined stroke severity, that were obtained by experienced physicians. Data were anonymized, centrally quality controlled, and stored at the coordinating

center at the University of Muenster, Germany. The ethics committee of the Westphalian Board of Physicians and the University of Muenster approved the study design. Because the identity of each documented patient is completely anonymized at the point of data collection, no study-specific informed consent was obtained. This study complies with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis reporting guideline.<sup>13</sup>

## Predictor Variables and Outcomes

The aim of this study was to develop practically applicable prediction models for severe complications after stroke in a broad, unselected sample of patients with acute ischemic stroke. We considered every variable of the standardized registry data set as potential predictor if it was continuously acquired in both 2016 and 2017 and was recorded within the first 24 hours after admission. This latter limitation was included to ensure applicability of the prediction model as early as possible. We focused on the prediction of severe adverse outcomes after stroke: early mortality (within the first 7 days after admission), intracranial mass effect, secondary intracerebral hemorrhage, and DVT, for example, of veins in the lower extremity. In the case of secondary intracerebral hemorrhage, we performed a subgroup analysis after the exclusion of all patients receiving intravenous thrombolysis. All events were recorded as present versus not present (and not in any graded way). Diagnoses themselves were passed on by each participating hospital and relied on their clinical pathways. For example, a patient might have been diagnosed as having an intracranial mass effect non-invasively based on classic neuroimaging findings (eg, compression of ventricles and midline shift) in combination with clinical symptoms (eg, reduced level of consciousness). We conducted analyses on a complete case basis, that is, we excluded patients with missing data for any of our considered variables (excluded patients, 4.3% overall; cf. Data S2). We compared the key covariates age, sex, and Rankin Scale upon admission between included and excluded patients to assess potential selection biases. A full list of all included 47 predictor variables and outcomes is presented in Table 1 and Table S1. Exact numbers of included and excluded patients and data of key comparisons are presented in Tables S2 through S4.

## Model Development

We developed 4 different models to predict each of the adverse outcomes after stroke. The classic model relied on logistic regression in combination with a backward stepwise procedure. In this scenario, we started with the full model considering any available input variable to predict the specific outcome. The variable

**Table 1. Stroke Sample Characteristics**

	2016 and 2017, N=146 062
Age, y	72.7 (13.1)
Female sex	69 234 (47.4)
Situation of living, before stroke	
Independently in own home	117 055 (80.1)
Care at home	15 847 (10.9)
Nursing home	13 160 (9.0)
Comorbidities	
Diabetes	42 944 (29.4)
Hypertension	124 754 (85.4)
Previous myocardial infarct	14 246 (9.8)
Previous stroke	38 089 (26.1)
Hypercholesterinaemia	84 644 (58.0)
Atrial fibrillation	
Yes, known before stroke	28 962 (19.8)
Yes, previously unknown	13 455 (9.2)
Stroke severity and symptoms at admission	
Stroke severity (NIHSS)	5.9 (6.2) 4 (6)
Motor impairments	95 636 (65.5)
Language impairments	44 684 (30.6)
Speech impairments	63 962 (43.8)
Swallowing impairments	32 168 (22.0)
Consciousness	
Awake	134 357 (92.0)
Soporific-stuporous	10 034 (6.9)
Comatose	1671 (0.01)
Rankin scale	
0	7749 (5.3)
1	20 541 (14.1)
2	35 672 (24.4)
3	34 920 (23.9)
4	23 839 (16.3)
5	23 341 (16.0)
Median (interquartile range)	3 (2)
Barthel index: bladder function	
0	29 459 (20.2)
5	18 893 (12.9)
10	97 710 (66.9)
Median (interquartile range)	10 (5)
Barthel index: transfer	
0	27 596 (18.9)
5	24 527 (16.8)
10	35 666 (24.4)
15	58 273 (39.9)
Median (interquartile range)	10 (10)
Barthel index: mobility	
0	34 310 (23.5)
5	27 645 (18.9)

(Continued)

**Table 1. Continued**

	2016 and 2017, N=146 062
10	36 117 (24.7)
15	47 990 (32.9)
Median (interquartile range)	10 (10)
Admission, times, and therapies	
Intravenous thrombolysis	24 989 (17.1)
Intraarterial thrombectomy and thrombolysis	10 706 (7.3)
Time from symptom onset until admission	
<1 h	11 825 (8.1)
1–2 h	23 177 (15.9)
2–3 h	13 889 (9.5)
3–3.5 h	4697 (3.2)
3.5–4 h	4229 (2.9)
4–6 h	13 366 (9.2)
6–24 h	29 721 (20.4)
24–48 h	10 923 (7.5)
>48 h	17 539 (12.0)
Imaging before admission	15 270 (10.5)
Intensive care admission	7752 (5.3)
Stroke characteristics	
TOAST classification	
Atherothrombotic	33 314 (22.8)
Embolic	46 097 (31.6)
Microangiopathic	30 003 (20.5)
Competing	5754 (3.9)
Other	5153 (3.5)
Uncertain	25 741 (17.6)
Large vessel stenosis	
Stenosis	137 411 (94.1)
No stenosis	5335 (3.7)
Unknown, no diagnostic tests	3316 (2.3)
Complications	
In-hospital mortality	7683 (5.3)
Intracranial mass effect	2411 (1.7)
Secondary intracerebral hemorrhage	2580 (1.8)
Deep vein thrombosis	606 (0.4)

Please note that the variable "Admission to intensive care" was only included in the prediction models of early mortality. Although admission to intensive care necessarily occurs before a fatal outcome, the temporal order was not known for any of the other 3 complications (ie, we could not exclude that admission to intensive care was a consequence of a complication). Continuous variables are presented as mean (SD) and categorical variables as absolute count (percentage). NIHSS indicates National Institutes of Health Stroke Scale. TOAST stands for the Trial of Org 10172 in Acute Stroke Treatment.

associated with the highest *P* value was then dropped from the model if it additionally exceeded a threshold of  $P > 0.01$ . The variable selection process stopped once all remaining variables were significantly associated with the outcome (level of significance,  $P < 0.01$ ). In addition, we employed 3 modern model-building

strategies representing different types of learning algorithms typically applied in the field of machine learning. This includes a  $l_1$ -regularized regression model<sup>14</sup>; a *k*-nearest neighbor classifier<sup>15</sup>; and a tree-based model, the gradient boosting classifier.<sup>16</sup>

During model development, we first performed a downsampling step. There were substantially fewer patients presenting with a specific complication than patients without this complication. For example, only 1.8% of all patients experienced an intracerebral hemorrhage, while 98.2% did not. Hence, we randomly selected (without replacement) a subset of the larger, nonaffected patient group to establish a group balance (cf. Data S3 for further motivations). By these means, we developed models in samples of 50% affected and 50% nonaffected patients. This downsampling step was repeated 100 times. The subsequent internal validation scheme depended on the respective prediction model. In case of logistic regression, we initiated a 4:1 train:test set split after each downsampling step. The backward stepwise selection of input variables and model fitting was performed in the train data set, whereas the prediction performance was then obtained for the hold-out test set. The 3 other approaches, the  $l_1$ -regularized logistic regression, *k*-nearest neighbor, and gradient boosting classifier, entered a nested cross-validation to securely run a hyperparameter optimization step intended to maximize prediction performance.<sup>17</sup> After an initial 4:1 train:test set split, we conducted a grid search to find the best hyperparameter settings via 5-fold cross-validation in the train set (cf. Table S5 for details on hyperparameter choices). The best performing model within this inner loop was then tested in the hold-out test set to get an estimate of an unbiased internal validation prediction performance. We measured out-of-sample prediction performance as area under the receiver operating curve (AUC) in the test sets. This measure therefore took into account the true positive and false positive rates at decision thresholds varying from 0 to 1. Lastly, we recorded in how many of the 100 downsampling repetitions a variable was chosen to stay in the model in the case of backward stepwise regression. In the case of the gradient boosting classifier, we noted the average feature importance. This importance can be explicitly computed for each input variable, allows their ranking and comparison (cf. Data S3 for details). We present odds ratios for the most stably selected input variables of logistic regression models as well as group averages (ie, patients with a specific outcome versus those without) for the most relevant input variables to allow for conclusions on the likely directionality of effects.

### Validation in Time

We implemented a temporal validation by testing the developed models on registry data obtained in the

subsequent year from January 1, 2017, to December 31, 2017. Included input and output variables remained the same. Although we had inserted a downsampling step during model development and had thus considered the same number of patients with and without a certain complication, we considered the entire sample without downsampling for model validation. We calibrated the models developed in the 2016 data set by weighing the model’s probabilistic predictions according to the fractions of patients with severe outcomes in the development and validation cohort. The final models were then employed to obtain a nonoptimistic estimate of the AUC. In addition, we visually evaluated the calibration of our prediction models in calibration plots and obtained Brier scores.<sup>18</sup>

## RESULTS

Overall, we considered trajectories of 152 710 patients with ischemic stroke included in the Stroke Registry of Northwestern Germany. Patients admitted to a participating hospital from January 1, 2016, to December 31, 2016, with complete data contributed to the development cohort (N=74 749 of 76 019, 98.3%), whereas patients of the subsequent year were assigned to the validation cohort (N=71 313 of 76 691, 93.0%; cf. Tables S2 through S4 for comparisons of included and excluded patients). The mean age of all patients was 72.7 years (SD, 13.1 years), 47.4% were women, median stroke severity at admission was determined as an NIHSS score of 4 (interquartile range [IQR], 6). Further baseline characteristics are summarized in Table 1.

Calculated during the 2 years of data recording, complication rates were 1.7% for intracranial mass effect (n=2411), 1.8% for intracerebral hemorrhage (n=2580), and 0.4% for DVT (n=606). Mortality within the first week after admission was 5.3% (n=7683).

### Prediction Results in the Validation Data Set (2017)

After developing the 4 competing models in the 2016 data set, we observed best 2017 validation data set prediction performances for the outcomes early

mortality and intracranial mass effect with AUC values of 0.90 (95% CI, 0.90–0.90). Validation data set prediction performances for secondary intracerebral hemorrhage and DVT were 0.80 (95% CI, 0.80–0.80) and 0.73 (95% CI, 0.72–0.73), respectively (Table 2). Validation data set prediction performance for intracerebral hemorrhage remained almost the same when restricting analyses to patients who did not receive any thrombolytic therapy (AUC, 0.79 [95% CI, 0.79–0.79]; Table S6). The gradient boosting classifier achieved best results for all 4 outcomes and outperformed logistic regression in all cases. However, prediction performances across classifiers were generally comparable and within a very narrow range of AUCs (largest difference: DVT, 0.73 [95% CI, 0.72–0.73] versus 0.71 [95% CI, 0.71–0.71] gradient boosting classifier versus logistic regression).

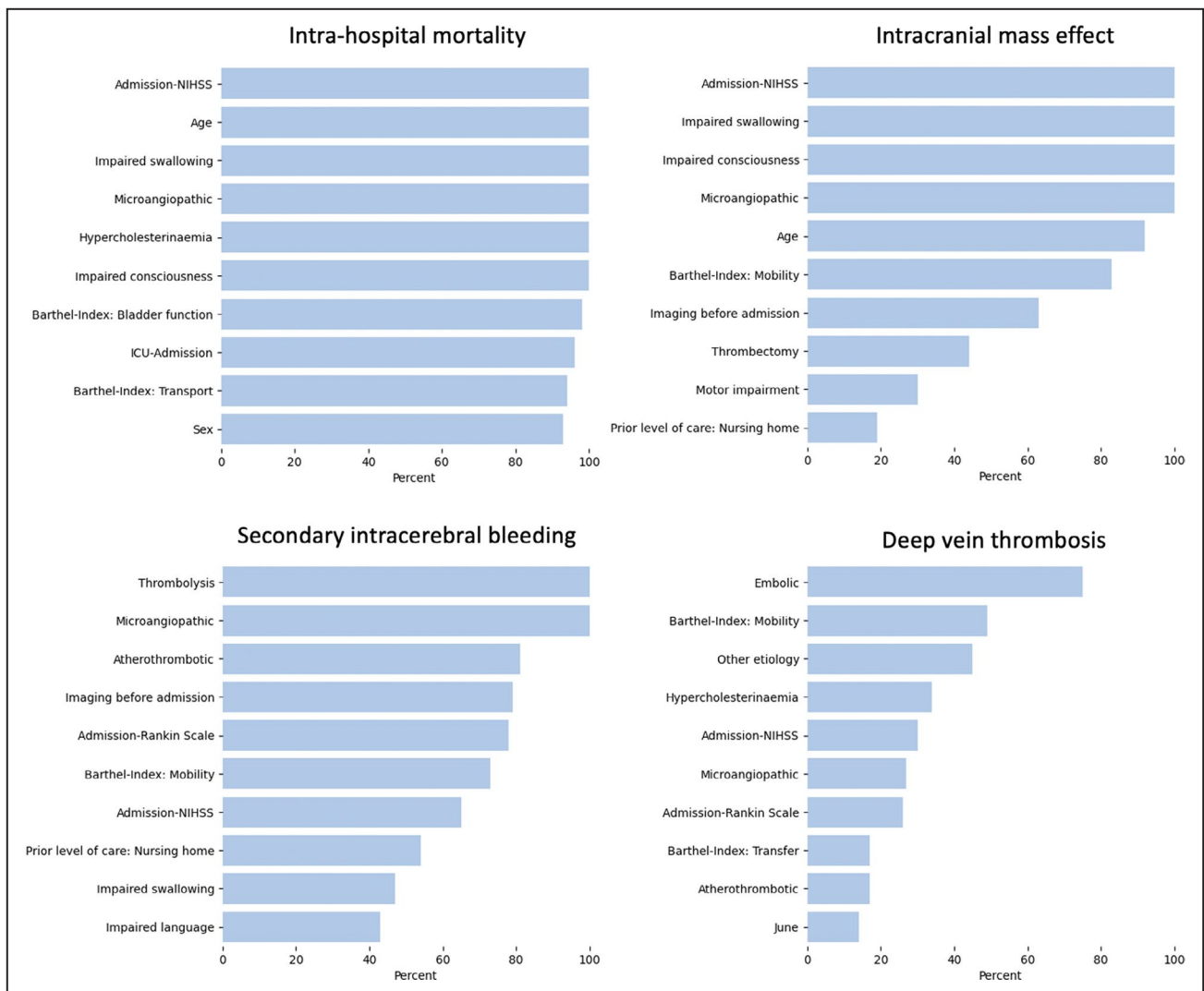
Furthermore, calibration plots indicated good calibration of all the classifiers for all 4 outcome scores in the validation set (Figure S1). Best Brier scores were achieved when predicting early mortality (range, 0.109–0.118) and intracranial mass effect (range, 0.113–0.132). Slightly higher scores were obtained for intracerebral hemorrhage (range, 0.165–0.173) and DVT (range, 0.191–0.213).

The most relevant input variables of each outcome’s backward stepwise regression model are presented in Figure 1. Because we repeated the backward stepwise selection after each of the 100 downsampling steps in total, an input variable could have, at maximum, been selected 100 times, or in 100% of the cases. The more often a variable was selected, the more stably important it may be for outcome prediction. The input variables with the highest feature importance in the gradient boosting models are shown in Figure 2. Altogether, NIHSS upon admission was the input variable most frequently ranked first for both the backward stepwise regression models and the gradient boosting models. The median of the admission stroke severity was generally several points higher in the groups of patients who experienced a severe complication than in the control group (eg, in-hospital mortality NIHSS: patients who died during the hospital stay, 17 [IQR, 11]; patients who survived, 4 [IQR, 5]; cf. Tables S7 through S10), which indicates a likely positive association

**Table 2. Prediction Results for All 4 Outcomes and Prediction Models in the Temporal Validation Cohort**

Classifier	In-hospital mortality	Intracranial mass effect	Secondary intracerebral hemorrhage	Deep vein thrombosis
Logistic	0.90 (0.90–0.90)	0.89 (0.89–0.89)	0.79 (0.79–0.79)	0.71 (0.71–0.71)
l1-regularized logistic regression	0.90 (0.90–0.90)	0.90 (0.89–0.90)	0.80 (0.79–0.80)	0.73 (0.72–0.73)
k-nearest neighbor classifier	0.89 (0.89–0.89)	0.88 (0.88–0.88)	0.78 (0.78–0.78)	0.71 (0.71–0.72)
Gradient boosting classifier	0.90 (0.90–0.90)	0.90 (0.90–0.90)	0.80 (0.80–0.80)	0.73 (0.72–0.73)

Data are shown as area under the curve (95% CI).

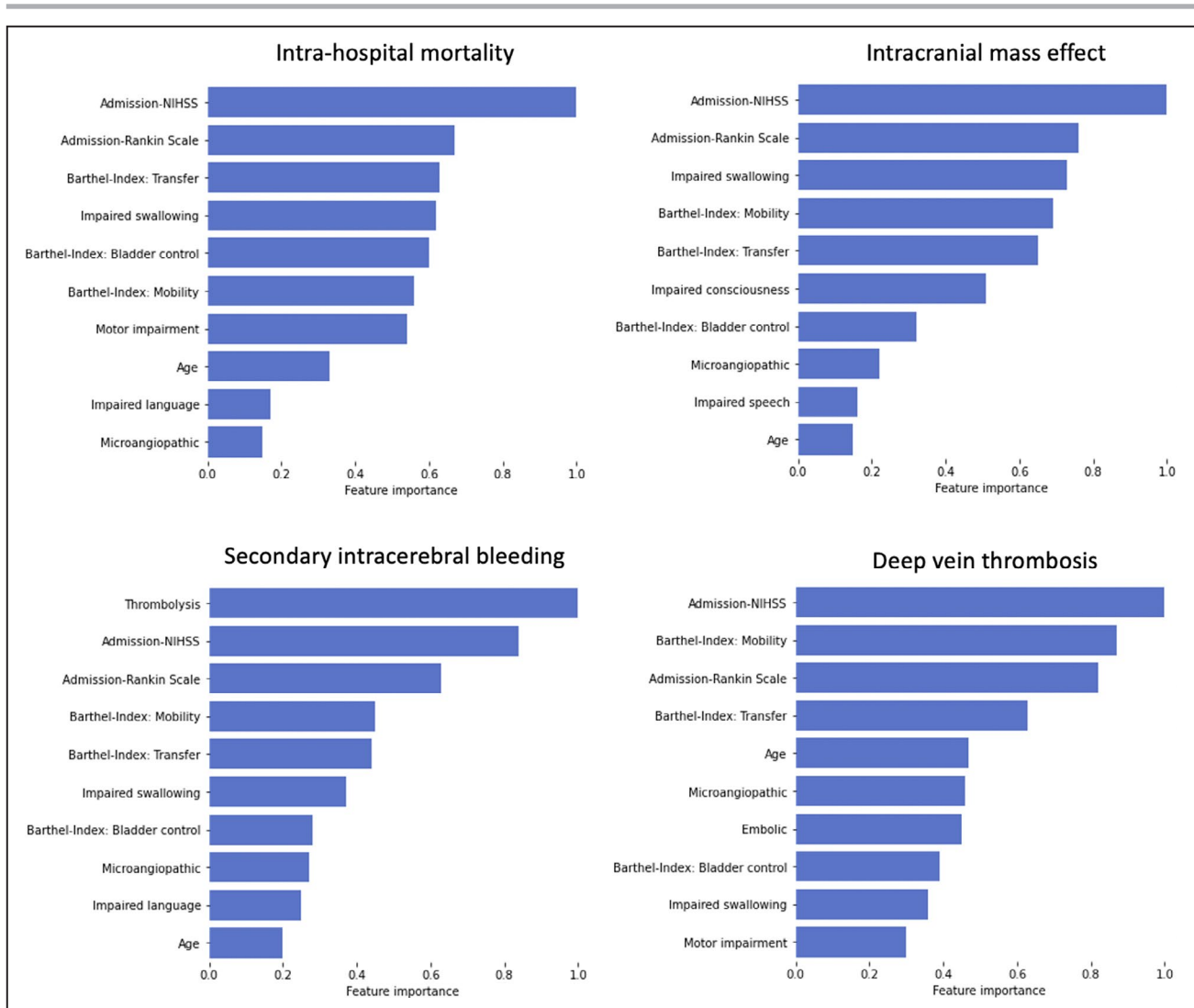


**Figure 1. The 10 most frequently selected variables in the backward stepwise logistic regression models.**

After each initial downsampling step, we performed backward stepwise variable selection, that is, we only kept those input variables in the model that were significantly associated with the outcome. Because we repeated the downsampling step 100 times, an input variable could have, at maximum, been selected 100 times, or in 100% of the cases (x axis). Altogether, a variable may be considered more important in the prediction of a specific outcome, the more often it is selected. In case of secondary intracerebral hemorrhage, thrombolysis and microangiopathic stroke etiology were, for example, selected in all 100 downsampling scenarios and may thus possess the highest predictive capacity. Atherothrombotic stroke etiology and imaging before admission were selected in ~80% of the downsampling scenarios and hence did not contribute to prediction models in ~20% of the cases, indicating a less consistent predictive capacity. Of note, we here only measured the overall relevance, yet not the direction of the association. Each variable could thus have had a positive or negative effect on the outcome. In a second step, we retrained logistic models with the 10 most stable input variables in 100 further downsampling scenarios to compute odds ratios informing about the directionality of effects (Table S11). Tables S7 through S10 furthermore present the group averages for patients with and without a specific outcome. Because the outcome deep vein thrombosis could not be predicted as well as the other outcomes, the relevance of input variables was not as certain either, which may explain the lower overall percentages for selected variables. ICU indicates intensive care unit; NIHSS, National Institutes of Health Stroke Scale.

between stroke severity and the adverse outcomes. This notion is reinforced by an odds ratio >1 in all 4 logistic regression models (cf. Table S11). For intrahospital mortality and intracranial mass effect, impaired swallowing and consciousness at admission as well as age had high overall rankings. The groups of patients who died during the hospital stay or were diagnosed

with an intracranial mass effect had substantially higher percentages of impaired swallowing and impaired consciousness (eg, in-hospital mortality and impaired swallowing: patients who died, 75.4%; patients who survived, 19.6%). In the case of in-hospital mortality, patients who died were on average 8.5 years older than patients who survived; in the case of intracranial



**Figure 2. Feature importance for the 10 most important input variables for each of the 4 outcomes based on the gradient boosting classifier models.**

Feature importance, a measure inherent to tree-based algorithms, is higher the more a variable contributes to the prediction of a specific outcome. Accordingly, the NIHSS score on admission was the most important variable in prediction of in-hospital mortality, increased intracranial pressure, and deep vein thrombosis, whereas administration of thrombolytic therapy was the most telling variable in the prediction of an intracerebral hemorrhage. Individual feature importance has been normalized by the top-ranked input variable and therefore range from 0 to 1. NIHSS indicates National Institutes of Health Stroke Scale.

mass effect, this age difference was less pronounced. Odds ratios for these input variables were consistently >1, with a maximum of 2.75 (95% CI, 2.69–2.81) for impaired swallowing function in the prediction of intracranial mass effect. The uptake of intravenous thrombolysis was the most relevant input feature for the prediction of secondary intracerebral hemorrhages. Thrombolysis was ≈3 times more frequent in patients with secondary intracerebral hemorrhages than in patients without (48.1% versus 16.9%) and had an odds ratio of 3.49 (95% CI, 3.44–3.55). Of note, some high-ranked input variables, such as microangiopathic etiology, were generally less frequent in the groups of patients with an adverse outcome, indicating negative

associations to the adverse outcome. The exhaustive list of odds ratios and group averages for all the input variables stated in Figures 1 and 2 that inform about the likely directionality of effect can be found in Tables S7 through S11. Prediction results based on the left-out test set for the 2016 development data set were very similar to the 2017 validation set results and can be found in Table S12.

## DISCUSSION

We developed and validated prediction models for mortality and 3 further key severe complications after

acute ischemic stroke. In-hospital mortality as well as intracranial mass effects could be predicted with high accuracy and AUCs of 0.9. Prediction performances of intracerebral hemorrhage as well as DVT were capped at slightly lower levels, achieving AUCs of 0.80 and 0.73, respectively. Herewith we consistently achieved prediction performances beyond those reported in the literature.

### Early Mortality

Various studies aimed at constructing robust predictions models for mortality at varying time points after stroke. Frequently, these studies were also based on stroke registry data. Considered time points ranged from the early acute phase (<7 days)<sup>4</sup> and in-hospital mortality<sup>19,20</sup> to the subacute (eg, 30 days) and chronic phase 1 year after stroke.<sup>21,22</sup> In a recent study, Gatttringer et al<sup>4</sup> developed the predicting early mortality of acute ischemic stroke score for mortality within the first week after stroke, relying on Austrian stroke registry data of 77 653 patients with stroke. These data were pooled from the years 2006 to 2016. They achieved an AUC of 0.88 (95% CI, 0.86–0.91) in the temporal validation data set that considered data from patients with stroke in 2017. In view of our validation cohort AUC of 0.90 (95% CI, 0.90–0.90), we thus present a slighter better prediction performance that is, however, still included in the upper end of the 95% CI in Gatttringer et al.<sup>4</sup>

Moreover, our results compare favorably with those obtained in the largest study in the field that considered data from 274 988 patients with ischemic stroke. These patients originated from 1036 hospitals participating in the American Get With the Guidelines Stroke Program between 2001 and 2007.<sup>20</sup> The authors focused on in-hospital mortality—thus an outcome comparable with ours—and established a validation data set prediction performance AUC of 0.72. Although this performance was obtained without any information on stroke severity, they augmented this performance to an AUC of 0.85 in a subsample of 109 187 patients who had admission NIHSS scores readily available. Interestingly, they reported an AUC of 0.83 for a model considering the NIHSS score only, which demonstrates the high relevance of initial stroke severity in the prediction of mortality.

All of these studies used slightly different clinical variables compared with those in our study. In most cases, these included age, some measure of stroke severity, and information on prestroke health status (eg, information on prior stroke; comorbidities, such as diabetes, preexisting heart disease). Extracted features of our best performing model, the gradient boosting classifier, also comprised measures of NIHSS-derived stroke severity and Rankin Scale–based degree of

disability. In addition, our model singled out impaired swallowing and several Barthel Index items, such as mobility and bladder control, as relevant predictive features.

### Intracranial Mass Effects

Numerous previous studies have focused on predicting mortality attributed to cerebral edema or cerebral edema with mass effects,<sup>5,23</sup> whereas our aim here was to predict the presence or absence of intracranial mass effects. A recent study more comparable with ours recruited 572 patients with ischemic stroke and described a prediction algorithm for cerebral edema with mass effect as detected on follow-up scans 3 days after the acute event.<sup>24</sup> Cerebral edema were observed in a fourth of the patients and could be predicted with an AUC of 0.78 (in-sample estimate). Key predictors were total anterior circulation syndrome, hyperdense appearance of middle cerebral artery, closed eyes, vomiting (all positively associated), lacunar cerebral syndrome, and white matter lesions (negatively associated).

We observed a substantially higher prediction performance in our study as we achieved an AUC of 0.90 despite a simpler nature of our input variables. The most important predictors of our best performing model resembled those for the prediction of mortality and included the NIHSS, Rankin Scale, and Barthel Index items at admission as well as information on impairments of swallowing, consciousness, and stroke etiology (microangiopathic).

Our estimate was obtained in an independent validation data set, whereas Muscari et al<sup>24</sup> reported in-sample estimates that might be affected by overfitting. The study by Muscari et al and our study differed in their inclusion and exclusion criteria: Muscari et al excluded patients with visible edema at admission, which was not possible in the current study. Information on the exact time point of edema diagnosis was not available. Furthermore, the incidence of intracerebral edemas differed notably: 27.6% versus 1.7%. This difference might be partly, but likely not fully, explained by a more stringent and complete imaging follow-up in Muscari et al and varying definitions of cerebral edema. It seems likely that their substantially smaller sample included patients who were more severely affected than would be observed in a more general sample as our stroke registry.

### Secondary Intracerebral Hemorrhage

Secondary intracerebral hemorrhage after initial ischemic stroke was previously mainly predicted in samples of patients who had received intravenous thrombolysis before study inclusion. The 2 most recent external validation studies report AUCs between



0.56 and 0.76 for various published intracerebral hemorrhage prediction algorithms.<sup>25,26</sup> Although Strbian et al<sup>25</sup> declared a score abbreviated to SEDAN<sup>6</sup> (SEDAN: baseline blood Sugar, Early infarct signs, [hyper] Dense cerebral artery sign on admission computed tomography scan, Age, NIH Stroke Scale on admission) as the best model (AUC=0.70), Asuzu et al<sup>26</sup> instead reported best prediction performances for the DRAGON<sup>27</sup> ([hyper] Dense cerebral artery sign/early infarct signs on admission CT scan, prestroke modified Rankin Scale score, Age, Glucose level at baseline, Onset-to-treatment time, NIH Stroke Scale on admission) score with an AUC of 0.73. Varying validation set sizes—3012 patients in the case of Strbian et al and 210 in Asuzu et al—may partially explain these differences. Furthermore, the DRAGON score was not tested in Strbian et al.

In contrast to these studies, we did not reduce our sample to only those patients who received intravenous thrombolysis. Instead, we predicted the occurrence of symptomatic secondary intracerebral hemorrhages in an unselected sample of patients with ischemic stroke. We considered this complementary approach worthwhile given that in our sample less than half of the patients with stroke experiencing a symptomatic cerebral hemorrhage had actually received a thrombolytic treatment (46.8% of 2576 patients with intracerebral hemorrhage). In this way, we also tried to prevent potential sample biases as, conceivably, multiple factors lead to the decision of intravenous thrombolysis in the first place.

Altogether, our results are promising because we could predict secondary intracerebral hemorrhages with an AUC of 0.80—higher than those reported in the literature. Furthermore, we only used routinely obtained clinical variables, whereas both the SEDAN as well as the DRAGON score build on more elaborated information. In addition to age and stroke severity, they consider blood glucose levels and imaging-derived information on early infarct signs, such as a hyperdense cerebral artery sign, in the admission scans. Unsurprisingly, the administration of intravenous thrombolytic therapy was the input variable with the highest importance in predicting secondary intracerebral hemorrhages. Measures of stroke severity and stroke etiology were further important input features. Their importance in predicting secondary intracerebral hemorrhages may explain why prediction performance remained high, almost unchanged at an AUC of 0.79, when restricting the analyses to only those patients who did not receive any intravenous thrombolysis.

## Deep Vein Thrombosis

Especially in the presence of limb paralysis and immobility, DVT is a relevant outcome after stroke that

can have severe consequences as it can cause a potentially fatal pulmonary embolism.<sup>2</sup> Early mobilization, sufficient hydration, and also prophylactic anticoagulation can be seen as examples of effective preventive strategies.<sup>28,29</sup> Previous prediction modeling studies for DVT have resulted in rather mediocre prediction performances, importantly, despite comparatively comprehensive sets of clinical input features. Dennis et al<sup>7</sup> used data from 2664 patients with stroke who were immobile originating from the CLOTS (Clots in Legs or Stockings After Stroke) trial to build a classifier of DVT occurrence. Forward and backward input variable selection processes led to the selection of the following variables as the most discriminative: dependence before stroke, unable to lift arms off bed, history of DVT/pulmonary embolism, and diabetes. The AUC, however, was only 0.57 in the validation data set. Smaller studies, recruiting 671 and 862 patients, reported slightly higher prediction performances with AUCs of 0.65 and 0.70, respectively.<sup>30,31</sup> In part, these scores also relied on more elaborate input variables, such as information on obesity, active cancer, or the level of low-density lipoprotein. Because our highest validation data set prediction performance showed an AUC of 0.73, it was found to be higher than previous estimates—despite the simple nature of our input variables. However, clinical utility may still be limited in view of the absolute AUC value. An important difference between studies can once again be seen in markedly differing outcome proportions: only 0.4% of patients in our data set had a documented event of DVT, whereas signs of DVT were detected in 10.9% of the patients in Dennis et al,<sup>7</sup> 12.4% of the patients in Liu et al,<sup>31</sup> and 22.1% of the patients in Li et al.<sup>30</sup> These differences may be partly explained by varying inclusion criteria—several of the previous studies focused on patients with an increased baseline risk, for example, attributed to severe hemiparesis. In addition, all of the aforementioned studies applied ultrasonography to screen for DVT. Currently, there is no such general ultrasonography screening established for unselected patient collectives, such as the collective represented in our stroke registry. As a result, the number of undetected cases may be high.

## General Considerations and Potential Outreach

In summary, we present evidence of augmented performance in predicting intrahospital mortality and 3 further important severe adverse outcomes after acute ischemic stroke. In contrast to the majority of previous studies, which primarily relied on logistic regression, we provide a comprehensive overview of the performance of multiple statistical learning approaches.<sup>32</sup> We opted for approaches representing various different learning

architectures. Some of them, the k-nearest neighbor and gradient boosting classifiers, can automatically extract nonlinear effects. In the case of logistic regression, this is only possible if these effects are inserted manually and thus intentionally, likely relying on preexisting notions, often in the form of expert knowledge. However, we could, at best, detect marginal improvements through the application of advanced statistical learning approaches: gradient boosting outperformed logistic regression in all 4 cases. Nonetheless, AUC estimates differed by an amount of a maximum of 0.02 only. This finding of no substantial superiority of more advanced machine learning–like, model-building strategies is well in line with several previous reports.<sup>33,34</sup> For instance, Evangelia et al<sup>33</sup> reviewed 71 clinical prediction studies that compared logistic regression to machine-learning approaches and could not detect any generally added benefit.

Although we did not observe a substantial improvement by advanced learning algorithms in comparison with logistic regression, our models still compared favorably with previously reported prediction models. Importantly, we achieved this performance despite generally simple input variables that were acquired in clinical routine. The observed increase in prediction performance may be explained by a larger data set size in comparison with earlier studies: our models had access to more incident cases to learn classification rules from. We also took into account a great variety of input variables and did not reduce our model to very few select variables. Moreover, increased data homogeneity may have been instrumental, as data were exclusively obtained in 2016 and 2017. Acute stroke treatments have experienced monumental changes, ranging from the introduction of stroke unit care in the 1990s to thrombolysis in the 2000s, more advanced imaging, extension of pertinent time windows, and nationwide thrombectomy in the 2010s.<sup>8</sup> In addition, the population of patients with stroke as a whole has experienced substantial alterations, primarily attributed to aging effects. All of these changes in the characteristics and care of patients with stroke have potential effects on functional outcomes and complications after stroke as well as on how these end points can be predicted.

Thus, any prediction model and its implementation should match these dynamics. We concentrated on a short, homogeneous time window from 2016 to 2017 that already reflects the most fundamental changes in acute stroke management, such as thrombolysis and thrombectomy. Our models would, however, have to be updated regularly to adopt to gradually changing environments in parallel. Such an update would be a feasible plan in the case of German stroke registry data, as these kinds of data are forwarded to a central storage in short time intervals.

A strength of our study can be seen in the large number of individual patients who originated from 155 hospitals, which underlines the chances of successful generalization. As our stroke registry data are anonymized and already stored centrally, its use does not interfere with data privacy concerns or often experienced barriers to shared data.

Altogether, our findings may argue for the use of routinely acquired clinical data to build outcome prediction models with the goal to stratify patients according to their risk profiles—the prediction models used here were all able to generate probability estimates of an outcome. Although it may not be feasible to apply preventive measures for unselected patient collectives, such a risk stratification could then allow for more tailored approaches. Preventive actions could then only be considered for those patients with the highest risks and potentially help to alleviate the detrimental effects of all of these complications—in the short and long term. Concretely, we could, for example, identify a subgroup of patients with stroke with a high DVT risk, say >80%, and administer venous Doppler of the legs before hospital discharge specifically to this subgroup. At best, this personalized diagnostic scheme would help uncover relevant cases of DVT early while being economical and not overstraining clinical resources. By these means, this risk stratification approach is conceptually different from prognostication: we assume that a specific outcome is still malleable in the first case, for example, by treatments, whereas it is fixed in the second case.

A limitation of our study may be the decreased level of model transparency. Because we strived for the highest possible prediction accuracy, we accepted a decrease in interpretability.<sup>35–37</sup> We could, however, generate rankings of most stably selected and important input variables (Figures 1 and 2). We then gained some insights on the likely directionality of effects for these relevant input variables via comparing group averages for patients with and without a specific outcome. Further options may be seen in general post hoc explanation methods, such as LIME<sup>38</sup> (locally interpretable model-agnostic explanations) and SHAP<sup>39</sup> (Shapley Additive Explanations), which can be employed to any model to generate explanations of the model's output.

A further limitation of our work may be that we did not test any deep-learning approaches. Typical studies in areas where deep learning excels, for example, in image or language processing, consider training examples in the millions.<sup>40</sup> In addition, deep learning has been shown to be particularly beneficial if complex interactions are present and exploitable.<sup>41,42</sup> Consequently, a significant increase in performance appears rather unlikely given the limited data set size and lack of signs of nonlinear or interaction effects.

Furthermore, although the usage of routinely acquired, basic clinical data may increase the feasibility of application in clinical settings, it may also be considered a limitation of our study: conceivably, information on more elaborated laboratory values<sup>43</sup> or advanced neuroimaging findings, such as vessel occlusions or infarct size or locations,<sup>44,45</sup> which we did not have access to in our stroke registry, can enhance prediction performances beyond those observed here. In addition, future studies may consider predicting continuous instead of our dichotomous outcome measures to further increase clinical utility. Most stroke registry data are acquired by experienced physicians, the data itself are continuously curated and carefully quality controlled at the Institute of Epidemiology, University of Muenster. Nonetheless, our stroke registry data represent observational data and feature some missing data (up to 7%, as demonstrated here), which may warrant future (prospective) studies to ensure a reliable generalization of our models to completely new patient data. Lastly, differences in sample characteristics and rates of outcomes may hamper direct comparisons between studies. Given that these differences were, for example, very pronounced in the case of DVT, future studies could examine whether our improved prediction performance was primarily attributed to predicting more severe and clinically particularly relevant cases of DVT that were documented by the clinical team.

In future work, we plan to extend prediction scenarios to intracerebral hemorrhage after intravenous thrombolysis, mortality after intracranial mass effects, and infectious events, such as pneumonia, after acute ischemic stroke. Our code and current versions of our trained models, which can be used to generate outcome predictions for new patients, is openly available on GitHub. We will furthermore aim to update models regularly to match training sets to current stroke populations as closely as possible. Lastly, it will be important to validate models in temporally as well as spatially independent data sets.

## CONCLUSIONS

Using data from 152 710 patients included in the stroke registry, we presented validated prediction models for 4 key adverse outcomes after acute ischemic stroke. We achieved high prediction accuracies for in-hospital mortality as well as intracranial mass effects with AUCs of 0.90. Intracerebral hemorrhage as well as DVT were predicted with AUCs of 0.80 and 0.73, respectively. Herewith, our prediction models consistently performed favorably when compared with previously established models. The NIHSS-defined stroke severity upon admission was the most predictive input variable for all outcomes. In view of their excellent AUCs,

prediction models for in-hospital mortality as well as intracranial mass effects could be of clinical importance and augment the clinical decision-making process.

## ARTICLE INFORMATION

Received July 12, 2021; accepted January 18, 2022.

### Affiliations

J. Philip Kistler Stroke Research Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA (A.K.B., N.S.R.); Institute of Epidemiology and Social Medicine, University of Muenster, Albert-Schweitzer-Campus 1, Muenster, Germany (A.K.B., N.R., K.B., A.K.); Cognitive Neuroscience, Institute of Neuroscience and Medicine, Research Centre Juelich, Juelich, Germany (C.G.); and Department of Neurology, Department of Neurology, University Hospital Cologne and Medical Faculty, University of Cologne, Germany (C.G.).

### Acknowledgments

We are grateful to our colleagues at the J. Philip Kistler Stroke Research Center, Massachusetts General Hospital, the Institute of Epidemiology and Social Medicine, University of Muenster, and Department of Neurology, University Hospital Cologne for valuable support and discussions. Additionally, we would like to specifically acknowledge valuable discussions on model calibration with Amy Nelson, MD. The continuous engagement and effort in the documentation of patients with stroke in all centers participating in the Stroke Registry of Northwestern Germany is highly appreciated and acknowledged. The following centers contributed to the registry: Johannes Wesling Klinikum Minden; Krankenhaus Lübbecke; Klinikum Herford; Evangelisches Krankenhaus Unna; Marienhospital Iserlohn Letmathe; St. Elisabeth Krankenhaus Dorsten; St. Josef-Hospital Bochum; Knappschaftskrankenhaus Bochum-Langendreer; BG Bergmannsheil Bochum; Paracelsus Klinik Marl; Evangelisches Krankenhaus Hattingen; Gemeinschaftskrankenhaus Herdecke; Evangelisches Krankenhaus Gelsenkirchen; Evangelisches Krankenhaus Herne; Universitätsklinikum Münster; Herz-Jesu-Krankenhaus Münster; Clemenshospital Münster; Klinikum Ibbenbüren; Helios Klinik Lengerich; LWL-Klinik Lengerich; St. Elisabeth-Hospital Iserlohn; St. Marien-Hospital Hamm; St. Barbara Klinik Heesen Hamm; Evangelisches Krankenhaus Hamm; St. Franziskus-Hospital Ahlen; St. Johannes Hospital Hagen; Klinikum Arnsberg-St. Johannes Hospital; Städt. Krankenhaus Marienhospital Arnsberg; Klinikum Lüdenscheld; Kreisklinikum Siegen; Klinikum Lippe-Lemgo; St. Vincenz Krankenhaus Menden; Ev. Krankenhaus Bielefeld Gilead I; LWL-Klinik Gütersloh; St. Vincenz Krankenhaus Paderborn; Ev. Krankenhaus Bielefeld Johannesstift; St. Elisabeth-Hospital; Gütersloh; Städtische Kliniken Dortmund; Klinikum Westfalen GmbH, Knappschaftskrankenhaus Dortmund; Hüttenhospital Dortmund; Kath. Krankenhaus Dortmund-West; Knappschaftskrankenhaus Recklinghausen; Christophorus-Kliniken GmbH Dülmen; St. Marien-Hospital Borken; Evangelisches Krankenhaus Castrop-Rauxel; Elisabeth-Krankenhaus Recklinghausen; Knappschaftskrankenhaus Bottrop; Prosper-Hospital Recklinghausen; St. Barbara Hospital Gladbeck; St. Vincenz-Krankenhaus Datteln; St. Rochus Hospital Castrop-Rauxel; KKRN Katholisches Klinikum Ruhrgebiet Nord GmbH Marl; St. Sixtus-Hospital Haltern am See; Gertrudis-Hospital Herten-Westerholt; St. Elisabeth-Hospital Herten; Klinikum Stadt Soest; St. Marien-Hospital GmbH Lünen; Ev. Krankenhaus Lippstadt; St. Ansgar-Krankenhaus Höxter; Asklepios Klinik Weißenfels; Carl von Basedow Klinikum Merseburg; Klinikum Niederlausitz GmbH Senftenberg; Altmark-Klinikum Gardelegen; Havelland Kliniken GmbH–Rathenow; Helios Klinik Blankenhain; DRK Manniske Krankenhaus Bad Frankenhausen; KMG Klinikum Kyritz; DRK Krankenhaus Luckenwalde; Havelland Kliniken GmbH–Klinik Nauen; Evangelisches Krankenhaus Ludwigsfelde-Teltow; Krankenhaus Märkisch-Oderland Strausberg; Sana-Krankenhaus Bergen auf Rügen; DRK Krankenhaus Teterow; Städtisches Klinikum Dresden–Standort Neustadt; IIm-Kreis-Kliniken Arnstadt; IIm-Kreis-Kliniken Ilmenau; SRH Krankenhaus Waltershausen-Friedrichroda; Waldkrankenhaus "Rudolf Elle" Eisenberg; Klinikum Barnim GmbH, Werner Forßmann KH Eberswalde; St. Georg Klinikum Eisenach; Ameos Klinikum Haldensleben; Gesundheitszentrum Bitterfeld/Wolfen; Helios-Klinik Sangerhausen; Helios Klinikum Bad Saarow; Universitätsklinikum Magdeburg; Paracelsus-Klinik Zwickau; Sana Krankenhaus Templin; Kreiskrankenhaus Rudolf Virchow Glauchau; Städtisches Klinikum Görlitz; Asklepios Klinik Pasewalk; Kreiskrankenhaus

Prenzlau; Klinikum Ernst von Bergmann Potsdam; Immanuel Klinik Rüdersdorf; Kreiskrankenhaus Prignitz Perleberg; Universitätsklinikum Carl Gustav Carus Dresden; Städtisches Klinikum Dresden—Standort Friedrichstadt; Sana Klinikum Borna; HELIOS Vogtland-Klinikum Plauen; Klinikum St. Georg gGmbH Leipzig; Martin Gropius Krankenhaus Eberswalde; BG-Kliniken Bergmannstrost Halle (Saale); Ruppiner Kliniken GmbH Neuruppin; AMEOS Klinikum Bernburg; Sächsisches Krankenhaus Altscherbitz Schkeuditz; Klinikum Magdeburg; Klinikum Dorothea Christiane Erleben Wernigerode; Universität Halle-Wittenberg; Universität Leipzig; Martha-Maria Krankenhaus Halle-Dölau; Klinikum Chemnitz; Kreiskrankenhaus Freiberg; Sächsisches Krankenhaus Arnsdorf; Heinrich-Braun-Krankenhaus Zwickau; Hanse-Klinikum Stralsund; KMG Klinikum Güstrow GmbH; Universität Rostock; AMEOS Diakonie-Klinikum Uckermark Uckermark; Hanse-Klinikum Wismar; Universitätsmedizin Greifswald; Helios Kliniken Schwerin; Krankenhaus Plau am See; Asklepios Fachklinikum Lübben; Asklepios Fachklinikum Teupitz; Asklepios Fachklinikum Brandenburg; Dietrich-Bonhoeffer-Klinikum Neubrandenburg; Carl-Thiem-Klinikum Cottbus; Asklepios Klinikum Uckermark Schwedt; Klinikum Frankfurt (Oder); Klinikum Meiningen; Sofien- und Hufeland-Klinikum Weimar; Helios-Klinikum Gotha; Oberhavel Kliniken GmbH—Klinik Hennigsdorf; Südharz-Krankenhaus Nordhausen; Klinikum Altenburger Land Altenburg; SRH Zentralklinikum Suhl; Universitätsklinikum Jena; SRH Waldklinikum Gera; Ökumenisches Hainich Klinikum Mühlhausen; Zentralklinik Bad Berka; Asklepios Fachklinikum Stadtroda; Helios Kliniken Aue; Helios Klinikum Erfurt; Fachkrankenhaus Hubertusburg Wernsdorf; Sächsisches Krankenhaus Rodewisch; Thüringen-Kliniken "Georgius Agricola" Rudolstadt; Klinikum Burgenlandkreis Naumburg; Diakoniekrankenhaus Chemnitzer Land; Diakomed gGmbH Hartmannsdorf; Muldentalkliniken; Krankenhaus Wurzen; Kreiskrankenhaus Greiz; Kliniken Erlabrunn gGmbH Breitenbrunn; Elblandkliniken Meißen; Uniklinikum Schleswig-Holstein Campus Kiel; Sana Kliniken Lübeck; Diakoniekrankenhaus Flensburg; Elbe-Klinikum Stade; Kreiskrankenhaus Schleiz; Eichsfeld Klinikum Kleinbartloff; Katholisches Krankenhaus St. Johann Nepomuk Erfurt; Hufeland Klinikum GmbH Standort Mühlhausen; Klinikum Bremerhaven-Reinkenheide; Klinikum Bad Salzungen; Robert-Koch-Krankenhaus-Appolda; Klinikum Bremen-Mitte; Gesundheit Nord - Klinikum Bremen Nord; KRH Klinikum Robert Koch Gehrden; Klinikum Großburgwedel; Helios Klinikum Hildesheim; Helios-Albert-Schweitzer-Klinikum Northeim; HELIOS Klinikum Salzgitter GmbH; HELIOS Klinikum Gifhorn; Städtisches Klinikum Wolfenbüttel; St. Bonifatius-Hospital Lingen; Christliches Klinikum Melle; Agaplesion Diakoniekrankenhaus Rotenburg; Mittelweser Kliniken GmbH Nienburg; Paracelsus Klinik Osnabrück; Heidekreis-Klinikum Soltau; Städtisches Klinikum Lüneburg; Klinikum Uelzen; KRH Klinikum Nordstadt Hannover; Nordwest-Krankenhaus Sanderbusch Sande; KRH Klinikum Neustadt am Rübenberge; Universitätsklinikum Göttingen; BDH-Klinik Hessisch Oldendorf; DIAKOVERE Henriettenstift Hannover; Agnes-Karll-Krankenhaus Laatzen; Klinikum Braunschweig; Asklepios-Kliniken Schildaualt Seesen; Klinikum Peine; Marienhospital Osnabrück; St. Bernward Krankenhaus Hildesheim; Medizinische Hochschule Hannover; Krankenhaus St. Elisabeth-Stift Damme; Klinikum Osnabrück; Hans-Susemihl-Krankenhaus Emden; Ev. Bathildis-Krankenhaus Bad Pyrmont; Diakovere Friederikenstift Hannover; Christliches Krankenhaus Quakenbrück; Allgemeines Krankenhaus Celle; Ammerland Klinik GmbH Westerstede; Evangelisches Krankenhaus Oldenburg; Ludmillenstift Meppen; St. Josefs Hospital Cloppenburg; Euregio-Klinik Nordhorn; Agaplesion Evangelisches Krankenhaus Holzminden; Klinikum Wolfsburg; Rhein-Maas Klinikum Marienhöhe Würselen; Universitätsklinikum Aachen; CaritasKlinikum Saarbrücken; Klinikum Saarbrücken; Knappschaftsklinikum Saar Sulzbach; Diakonie Klinikum Neunkirchen/Saar; DRK Krankenhaus Saarlouis; Marienhospital Saarbrücken-Dillingen; Universitätsklinikum Homburg/Saar; SHG Klinikum Merzig; Knappschaftskrankenhaus Püttlingen; Marienkrankenhaus St. Wendel; Sana Kliniken Duisburg; Kliniken Maria Hilf Mönchengladbach; Kreiskrankenhaus Gummersbach; Helios Klinikum Wuppertal-Barmen; Agaplesion Bethesda Krankenhaus Wuppertal; Centre Hospitalier de Luxembourg; Centre Hospitalier Emile Mayrisch Esch-sur-Alzette; Hôpital St-Louis Ettelbruck Luxembourg; ZithaKlinik Luxembourg; and Hôpitaux Robert Schuman Luxembourg-Kirchberg.

### Sources of Funding

Dr Bonkhoff is supported by a Massachusetts General Hospital Executive Committee on Research Fund for Medical Discovery Clinical Research Fellowship Award. Dr Grefkes is in part funded by the Deutsche Forschungsgemeinschaft (German Research Foundation; project

431549029) and SFB 1451 projects B05 and C05. Dr Rost is supported in part by the National Institutes of Health—National Institute of Neurological Disorders and Stroke (R01NS082285, R01NS086905, U19NS115388).

### Disclosures

Dr Rost has received compensation as scientific advisory consultant from Omnio, Sanofi Genzyme, and AbbVie Inc. The remaining authors have no disclosures to report.

### Supplemental Material

Data S1–S3  
Tables S1–S12  
Figure S1  
References 46,47

## REFERENCES

- Feigin VL, Forouzanfar MH, Krishnamurthi R, Mensah GA, Connor M, Bennett DA, Moran AE, Sacco RL, Anderson L, Truelsen T, et al. Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease Study 2010. *Lancet*. 2014;383:245–255. doi: 10.1016/S0140-6736(13)61953-4
- Kumar S, Selim MH, Caplan LR. Medical complications after stroke. *Lancet Neurol*. 2010;9:14. doi: 10.1016/S1474-4422(09)70266-2
- Hinman JD, Rost NS, Leung TW, Montaner J, Muir KW, Brown S, Arenillas JF, Feldmann E, Liebeskind DS. Principles of precision medicine in stroke. *J Neurol Neurosurg Psychiatry*. 2017;88:54–61. doi: 10.1136/jnnp-2016-314587
- Gattringer T, Posekany A, Niederkorn K, Knoflach M, Poltrum B, Mutzenbach S, Haring H-P, Ferrari J, Lang W, Willeit J, et al. Predicting early mortality of acute ischemic stroke: score-based approach. *Stroke*. 2019;50:349–356. doi: 10.1161/STROKEAHA.118.022863
- Thomalla G, Hartmann F, Juettler E, Singer OC, Lehnhardt F-G, Köhnlmann M, Kersten JF, Krüzelmann A, Humpich MC, Sobesky J, et al. Prediction of malignant middle cerebral artery infarction by magnetic resonance imaging within 6 hours of symptom onset: a prospective multicenter observational study. *Ann Neurol*. 2010;68:435–445. doi: 10.1002/ana.22125
- Strbian D, Engelster S, Michel P, Meretoja A, Sekoranja L, Ahlhelm FJ, Mustanoja S, Kuzmanovic I, Sairanen T, Forss N, et al. Symptomatic intracranial hemorrhage after stroke thrombolysis: the SEDAN score. *Ann Neurol*. 2012;71:634–641. doi: 10.1002/ana.23546
- Dennis M, Sandercock P, Reid J, Graham C, Murray G, Venables G, Rudd A, Bowler G; The CLOTS Trials Collaboration. Can clinical features distinguish between immobile patients with stroke at high and low risk of deep vein thrombosis? Statistical modelling based on the CLOTS trials cohorts. *J Neurol Neurosurg Psychiatry*. 2011;82:1067–1073. doi: 10.1136/jnnp.2010.235945
- Lee M, Ovbiagele B. Stroke in 2017: intensive and extensive—advances in stroke management. *Nat Rev Neurol*. 2018;14:68–70. doi: 10.1038/nrneurol.2017.187
- Norrving B. Stroke management—recent advances and residual challenges. *Nat Rev Neurol*. 2019;15:69–71. doi: 10.1038/s41582-019-0129-1
- Minnerup J, Wersching H, Ringelstein EB, Schilling M, Schäbitz W-R, Wellmann J, Berger K. Impact of the extended thrombolysis time window on the proportion of recombinant tissue-type plasminogen activator-treated stroke patients and on door-to-needle time. *Stroke*. 2011;42:2838–2843. doi: 10.1161/STROKEAHA.111.616565
- Bonkhoff AK, Karch A, Weber R, Wellmann J, Berger K. Female stroke: sex differences in acute treatment and early outcomes of acute ischemic stroke. *Stroke*. 2021;52:406–415. doi: 10.1161/STROKEAHA.120.032850
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.
- Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med*. 2015;13:1. doi: 10.1186/s12916-014-0241-z
- Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc: Ser B*. 1996;58:267–288.

15. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat*. 1992;46:175–185.
16. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29:1189–1232. doi: 10.1214/aos/1013203451
17. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *NeuroImage*. 2017;145:166–179. doi: 10.1016/j.neuroimage.2016.10.038
18. Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev*. 1950;78:1–3. doi: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2
19. Heuschmann PU, Kolominsky-Rabas PL, Misselwitz B, Hermanek P, Leffmann C, Janzen RWC, Rother J, Buecker-Nott H-J, Berger K. Predictors of in-hospital mortality and attributable risks of death after ischemic stroke: the German Stroke Registers Study Group. *Arch Intern Med*. 2004;164:1761–1768. doi: 10.1001/archinte.164.16.1761
20. Smith EE, Shobha N, Dai D, Olson DM, Reeves MJ, Saver JL, Hernandez AF, Peterson ED, Fonarow GC, Schwamm LH. Risk score for in-hospital ischemic stroke mortality derived and validated within the get with the guidelines-stroke program. *Circulation*. 2010;122:1496–1504. doi: 10.1161/CIRCULATIONAHA.109.932822
21. Saposnik G, Kapral MK, Liu Y, Hall R, O'Donnell M, Raptis S, Tu JV, Mamdani M, Austin PC. IScore: a risk score to predict death early after hospitalization for an acute ischemic stroke. *Circulation*. 2011;123:739–749. doi: 10.1161/CIRCULATIONAHA.110.983353
22. O'Donnell MJ. The PLAN score: a bedside prediction rule for death and severe disability following acute ischemic stroke. *Arch Intern Med*. 2012;172:1548. doi: 10.1001/2013.jamainternmed.30
23. Kasner SE, Demchuk AM, Berrouschot J, Schmutzhard E, Harms L, Verro P, Chalela JA, Abbur R, McGrade H, Christou I, et al. Predictors of fatal brain edema in massive hemispheric ischemic stroke. *Stroke*. 2001;32:2117–2123. doi: 10.1161/hs0901.095719
24. Muscari A, Faccioli L, Lega MV, Lorusso A, Trossello MP, Puddu GM, Spinardi L, Zoli M. Predicting cerebral edema in ischemic stroke patients. *Neuro Sci*. 2019;40:745–752. doi: 10.1007/s10072-019-3717-y
25. Strbian D, Michel P, Seiffge DJ, Saver JL, Numminen H, Meretoja A, Murao K, Weder B, Forss N, Parkkila A-K, et al. Symptomatic intracranial hemorrhage after stroke thrombolysis: comparison of prediction scores. *Stroke*. 2014;45:752–758. doi: 10.1161/STROKEAHA.113.003806
26. Asuzu D, Nystrom K, Amin H, Schindler J, Wira C, Greer D, Chi NF, Halliday J, Sheth KN. Comparison of 8 scores for predicting symptomatic intracerebral hemorrhage after IV thrombolysis. *Neurocrit Care*. 2015;22:229–233. doi: 10.1007/s12028-014-0060-2
27. Strbian D, Meretoja A, Ahlhelm FJ, Pitkaniemi J, Lyrer P, Kaste M, Engelter S, Tatlisumak T. Predicting outcome of IV thrombolysis-treated ischemic stroke patients: the DRAGON score. *Neurology*. 2012;78:427–432. doi: 10.1212/WNL.0b013e318245d2a9
28. Diener H-C, Ringelstein EB, von Kummer R, Landgraf H, Koppenhagen K, Harenberg J, Rektor I, Csányi A, Schneider D, Klingelhöfer J, et al. Prophylaxis of thrombotic and embolic events in acute ischemic stroke with the low-molecular-weight heparin certoparin: results of the PROTECT Trial. *Stroke*. 2006;37:139–144. doi: 10.1161/01.STR.0000195182.67656.ee
29. Kamphuisen PW, Agnelli G. What is the optimal pharmacological prophylaxis for the prevention of deep-vein thrombosis and pulmonary embolism in patients with acute ischemic stroke? *Thromb Res*. 2007;119:265–274. doi: 10.1016/j.thromres.2006.03.010
30. Li S-Y, Feng L, Xiao M-J, Chen S-Y, He J-C, Wang Z. Derivation and validation of a clinical prediction scale for isolated distal deep venous thrombosis in patients after acute ischemic stroke. *J Stroke Cerebrovasc Dis*. 2017;26:2087–2092. doi: 10.1016/j.jstrokecerebrovasdis.2017.04.027
31. Liu L-P, Zheng H-G, Wang DZ, Wang Y-L, Hussain M, Sun H-X, Wang A-X, Zhao X-Q, Dong K-H, Wang C-X, et al. Risk assessment of deep-vein thrombosis after acute stroke: a prospective study using clinical factors. *CNS Neurosci Ther*. 2014;20:403–410. doi: 10.1111/cns.12227
32. Bonkhoff AK, Grefkes C. Precision medicine in stroke: towards personalized outcome predictions using artificial intelligence. *Brain*. 2021;awab439. doi: 10.1093/brain/awab439
33. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22. doi: 10.1016/j.jclin.2019.02.004
34. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. *JAMA Network Open*. 2020;3:e1918962. doi: 10.1001/jamanetworkopen.2019.18962
35. Shmueli G. To Explain or to predict? *Stat Sci*. 2010;25:289–310. doi: 10.1214/10-STS330
36. Bzdok D, Altman N, Krzywinski M. Points of significance: statistics versus machine learning. *Nat Methods*. 2018;15:233–234. doi: 10.1038/nmeth.4642
37. Holm EA. In defense of the black box. *Science*. 2019;364:26–27. doi: 10.1126/science.aax0162
38. Ribeiro MT, Singh S, Guestrin C. “Why should i trust you?” Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016:1135–1144.
39. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017:4768–4777.
40. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;25:1097–1105.
41. Hinton G. Deep learning—a technology with the potential to transform health care. *JAMA*. 2018;320:1101–1102. doi: 10.1001/jama.2018.11100
42. Schulz M-A, Yeo BT, Vogelstein JT, Mourao-Miranada J, Kather JN, Kording K, Richards B, Bzdok D. Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. *Nat Commun*. 2020;11:1–15.
43. De Marchis GM, Dankowski T, König IR, Fladt J, Fluri F, Gensicke H, Foerch C, Findling O, Kurmann R, Fischer U, et al. A novel biomarker-based prognostic score in acute ischemic stroke: the CoRisk score. *Neurology*. 2019;92:e1517–e1525. doi: 10.1212/WNL.00000000000007177
44. Ospel JM, Mayank A, Hill MD, Menon BK, Demchuk A, McTaggart R, Nogueira R, Poppe A, Haussen D & Qiu W et al. Strength of association between infarct volume and clinical outcome depends on the magnitude of infarct size: results from the ESCAPE-NA1 Trial. 5.
45. Bonkhoff AK, Schirmer MD, Bretzner M, Hong S, Reegenhardt RW, Brudfors M, Donahue KL, Nardin MJ, Dalca AV, Giese A-K, et al. Outcome after acute ischemic stroke is linked to sex-specific lesion patterns. *Nat Commun*. 2021;12:3289. doi: 10.1038/s41467-021-23492-3
46. Kubat M, Matwin S. Addressing the curse of imbalanced training sets: one-sided selection. In *ICML vol. 97* 179–186 (Citeseer, 1997).
47. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. New York, NY: Springer; 2013:112.

# **SUPPLEMENTAL MATERIAL**

## **Data S1. Data collection tool 2016 – English translation**

We here present an exhaustive list of obtained items. We only considered those variables for our prediction tool that were readily available upon admission and continuously documented for the years 2016 and 2017. The original German version can be accessed online: <https://www.medizin.uni-muenster.de/qsnwd/downloads.html> (“Spezifikationen”).

**Center**

**Date of admission**

**Date of discharge**

**Year of birth**

**Patient’s provision available** – yes/no

**Sex** – Male/Female

**Time between symptom onset and admission** – <1h, 1-2h, 2-3h, 3-3.5h, 3.5h-4h, 4-6h, 6-24h, 24-48h, >48h, unknown

**Duration of symptoms** – <1h, 1-24h, >24h

**Transport to Hospital** – private, emergency doctor/helicopter with emergency doctor, ambulance (without emergency doctor), other

**Admission initiated by** – self, emergency doctor, doctor in private practice, external hospital, internally in hospital

**Barthel-Index upon admission and at discharge** - Bladder control, Transition Bed-Chair, Mobility

**Initial admission on** – normal ward, intensive care, stroke unit

**Stay on Stroke Unit** – yes/no

**Days in Stroke Unit**

**Symptoms upon admission:**

**Motor (Arm/Hand and/or Leg/Foot)** – yes/no

**Speech** – yes/no

**Language** – yes/no

**Swallowing** – yes/no

**Consciousness** – awake, somnolent, comatose

**Comorbidity:**

**Diabetes mellitus** – yes/no

**Hypertension** – yes/no

**Previous myocardial infarction** – yes/no

**Previous stroke** – yes/no

**Hypercholesterinaemia** – yes/no

**Atrial fibrillation** – yes, known, yes, but newly diagnosed, no

**ICD-10 classification**

**Etiology** – no infarct, atherothrombotic, cardiogen-embolic, microangiopathic, other cause, unknown cause, competing causes

**Diagnostics after event:**

**Long-term ECG (at least 24h)** – yes/no

**Swallowing test** – yes/no

**Neuroimaging: CCT** – yes/no

**Neuroimaging: MRI** – yes/no

**If Neuroimaging: New visible lesion** – yes/no

**Extracranial vascular diagnostics** – no, yes in <48h, yes in >48h

**Intracranial vascular diagnostics** – no, yes in <48h, yes in >48h

**If vascular diagnostics: Symptomatic ipsilesional stenosis of Ateria carotis interna** – no, <50%, 50-70%, 70-99%, 100%, not investigated

**Recommendation of revascularization** – no, surgery documented, stenting documented, transfer for surgery, transfer for stenting, other

**Therapy/Prophylaxis:**

**Antiplatelet agent <48h after symptom onset** – yes/no

**Antiplatelet agent at discharge** – yes/no

**Thromboprophylaxis** – yes/no

**Oxygen** – yes/no

**Antihypertensives** – yes/no

**Antidiabetics** – yes/no

**Statins** – yes/no

**Anticoagulation (e.g. Warfarin)** – no, Vitamin K-Antagonists, new anticoagulants

**Complications** – no complications, urinary tract infection, pneumonia, re-stroke, increased intracranial pressure, intracerebral bleeding, thrombosis, other complication

**Rehabilitative therapy:**

**Physio-/ Occupational therapy** – no, yes (begin <2days after onset), yes (begin >2days after onset)

**Speech therapy** – no, yes (begin <2days after onset), yes (begin >2days after onset)

**Mobilization** – no, yes (begin <2days after onset), yes (begin >2days after onset)

**Thrombolysis i.v.** – yes/no

**Thrombolysis i.a.** – yes/no

**Thrombectomy** – yes/no

**NIHSS upon admission**

**Time between admission and first neuroimage** – no image, first image before admission, <0.5h, 0,5-1h, 1-3h, 3-6h, >6h



**Time between admission and thrombolysis (i.v., i.a.)** – no thrombolysis, <0.5, 0.5-1h, 1-2h, 2-3h, 3-4h, 4-6h, >6h

**Planned rehabilitation after discharge from acute care hospital** – no rehabilitation planned, neurological rehabilitation Phase B, neurological rehabilitation Phase C, neurological rehabilitation Phase D (inpatient), neurological rehabilitation Phase D (outpatient), geriatric rehabilitation, other rehabilitation

**Discharge** – regular discharge, discharge due to other reasons, transfer to other facility <2h after admission, transfer to other hospital, death, discharge to rehabilitation, discharge to nursing home, internal transfer

## **Data S2. Complete case analyses**

We performed complete case analyses and thus excluded subjects with missing information for any of the considered variables. This led to the exclusion of 1,270 out of 76,019 patients for the year 2016 (1.7%) and the exclusion of 5,378 patients out of 76,691 patients (7.0%) for the year 2017. The exclusion of more patients in 2017 than 2016 (7% vs. 1.7%) was primarily due to an increased frequency of missing information on “Swallowing impairments at admission” (3.0% in 2017 vs. 0.4% in 2016), “Speech impairments at admission” (2.1% vs. 0.2%), “Language impairments at admission” (1.4% vs. 0.1%) and the comorbidities hypercholesterinaemia (1.8% vs. 0.4%) and prior myocardial infarct (1.7% vs. 0.2%). This increase between 2016 and 2017 may be due to the fact that it was possible to record these items as “present/absent/not possible to determine” in 2017, but only as “present/absent” in 2016.

The comparison of included and excluded patients based on the complete case criterion can be found in **Tables S2-S3**. The groups of excluded patients in 2016 and 2017 were slightly older, comprised more female patients and had a slightly higher median Rankin Score in 2017 than the one of included patients. Included and excluded subjects in the 2016 derivation cohort, used for model training, differed in only very small absolute amounts. Differences were slightly more pronounced in the 2017 validation cohort. However, importantly, differences in key characteristics (age, sex, Rankin Scale) and rates of the four adverse outcomes of *included* subjects in 2016 and 2017 were very small in magnitude. This later aspect is reassuring with respect to the suitability of the 2017 cohort as validation cohort.

## **Data S3. Details on our prediction pipeline**

### **Downsampling**

The downsampling step was motivated by pronounced class imbalances in case of all four of our adverse outcomes (rates: 5% mortality, 1.7% ICP, 1.8% ICH and 0.4% DVT). Downsampling the majority class, as implemented in our study, has been shown to improve classifier performance on minority class cases, i.e., those patients that are usually of interest.<sup>48</sup>

### Variable Normalization

As some of the used approaches, i.e., the  $l_1$ -regularized logistic regression and k-nearest neighbor classifier, are sensitive to variable scales, we normalized input variables by subtracting the sample mean and dividing by the sample standard deviation in the training set and applied the same normalization to the test set.

### Feature importance

The feature importance computed for the gradient boosting classifier quantifies how much each variable improves the prediction performance based on the increase in the Gini index-derived purity. The Gini index itself describes the total variance across various classes; a low value indicates that each node contains only observations from a single class.<sup>47</sup>

**Table S1.** Stroke sample characteristics (continuation).

		<b>2016 and 2017</b>
<b>Months</b>	<b>Jan</b>	12,464 (8.5)
	<b>Feb</b>	11,603 (7.9)
	<b>Mar</b>	12,541 (8.6)
	<b>April</b>	11,924 (8.2)
	<b>May</b>	12,596 (8.6)
	<b>June</b>	12,019 (8.4)
	<b>July</b>	12,275 (8.4)
	<b>August</b>	12,345 (8.5)
	<b>September</b>	11,847 (8.1)
	<b>October</b>	12,043 (8.2)
	<b>November</b>	12,110 (8.3)
	<b>December</b>	11,530 (7.7)

**Table S2.** Key characteristics of included and excluded subjects in the year 2016.

	Included subjects (N=74,749)	Excluded subjects (N=1,270)	Statistical comparison (two-sided t-test or Fisher's exact test as appropriate)
Age (in years)	72.7 (13.1)	73.3 (12.7)	$p=0.08$
Sex (female)	47.6%	48.1%	$p=0.77$
Rankin Scale upon admission	3 (2)	3 (2)	$p=0.01$

**Table S3.** Key characteristics of included and excluded subjects in the year 2017.

	Included subjects (N=71,313)	Excluded subjects (N=5,378)	Statistical comparison (two-sided t-test or Fisher's exact test as appropriate)
Age (in years)	72.7 (13.0)	74.9 (12.8)	$p<0.01$
Sex (female)	47.1%	52.7%	$p<0.01$
Rankin Scale upon admission	3 (2)	4 (2)	$p<0.01$

**Table S4.** Key characteristics and frequencies of adverse outcomes of included subjects in 2016 and 2017.

	Derivation cohort (2016) (N=74,749)	Validation cohort (2017) (N=71,313)
Age (in years)	72.7 (13.1)	72.7 (13.0)
Sex (females)	35,616 (47.7)	33,618 (47.1)
Rankin Scale upon admission	3 (3)	3 (2)
In-hospital mortality	4,046 (5.4)	3,637 (5.1)
Increased intracranial pressure	1,339 (1.8)	1,072 (1.5)
Secondary intracerebral hemorrhage	1,361 (1.8)	1,219 (1.7)
Deep vein thrombosis	337 (0.5)	269 (0.4)

**Table S5.** Model building approaches and their respective hyperparameter settings entered in the grid search.

Approach	Hyperparameter	Description of hyperparameter	Subset of values assessed in the grid search
l1-regularized	'C'	'C': inverse of regularization strength. Smaller values imply a stronger regularization.	'C' = 0.01, 0.1, 1, 10, 100, 1000

logistic regression			
k-nearest neighbor classifier	'n_neighbors'	'n_neighbors': number of nearest neighbors to respect for majority vote (i.e., a new data point is classified into the same group than the majority of considered neighbors)	'n_neighbors' = 1, 5, 10, 50
Gradient boosting classifier	'n_estimators' 'max_depth' 'loss'	'n_estimators': the number of boosting stages to conduct 'max_depth': maximum depth of individual classifiers 'loss': loss function that is optimized	'n_estimators' = 100, 300, 500 'max_depth' = 1, 3, 5 'loss' = 'deviance', 'exponential'

**Table S6.** Validation dataset AUCs for the prediction of secondary intracerebral hemorrhage in the sample of stroke patients that did not receive any thrombolytic therapy.

Classifier	Secondary intracerebral hemorrhage
Logistic	0.78 (0.78-0.78)
l1-regularized logistic regression	0.79 (0.79-0.79)
kNN	0.77 (0.77-0.77)
GBC	0.79 (0.79-0.79)

**Table S7.** Group averages for most stable and important input variables for in-hospital mortality in 2016.

	Patients without in-hospital mortality	Patients with in-hospital mortality
<b>Both logistic regression and gradient boosting</b>		
Admission-NIHSS	4 (5)	17 (11)
Age	72.2 (13.1)	80.7 (10.2)
Impaired swallowing	19.6%	75.4%
Microangiopathic etiology	20.9%	3.3%
Barthel index: Bladder function (less than full score)	30.9%	86.1%
Barthel Index: Transfer (less than full score)	58.5%	96.2%
<b>Only logistic regression</b>		
Hypercholesterinaemia	57.1%	43.3%
Impaired consciousness	6.2%	52.6%
ICU-admission	4.4%	22.9%
Sex (female)	47.0%	58.3%
<b>Only Gradient boosting</b>		
Admission-Rankin Scale	3 (5)	5 (1)
Barthel Index: Mobility (less than full score)	65.7%	97.3%

Motor impairment	64.7%	92.9%
Impaired language	29.5%	70.8%

**Table S8. Group averages for most stable and important input variables for increased intracranial pressure in 2016.**

	Patients without ICP	Patients with ICP
<b>Both logistic regression and gradient boosting</b>		
Admission-NIHSS	4 (6)	18 (11)
Impaired swallowing	21.5%	79.3%
Impaired consciousness	7.9%	55.0%
Microangiopathic etiology	20.3%	1.6%
Age	72.7 (13.0)	73.3 (14.5)
Barthel Index: Mobility (less than full score)	66.8%	96.2%
<b>Only logistic regression</b>		
Imaging before admission	9.5%	17.3%
Thrombectomy	7.0%	25.9%
Motor impairment	65.7%	92.0%
Prior level of care: Nursing home	9.1%	12.8%
<b>Only Gradient boosting</b>		
Admission-Rankin Scale	3 (2)	5 (0)
Barthel Index: Transfer (less than full score)	59.9%	94.7%
Barthel Index: Bladder control (less than full score)	33.0%	80.0%
Impaired speech	43.5%	77.6%

**Table S9. Group averages for most stable and important input variables for secondary intracerebral hemorrhage in 2016.**

	Patients without ICB	Patients with ICB
<b>Both logistic regression and gradient boosting</b>		
Thrombolysis	16.9%	48.1%
Microangiopathic etiology	20.2%	5.1%
Admission-Rankin Scale	3 (2)	4 (2)
Barthel Index: Mobility (less than full score)	67.0%	91.3%
Admission-NIHSS	4 (6)	12 (12)
Impaired swallowing	22.0%	55.9%
Impaired language	31.3%	59.1%
<b>Only logistic regression</b>		
Atherothrombotic etiology	23.3%	19.5%
Imaging before admission	9.5%	14.1%
Prior level of care: Nursing home	9.1%	11.9%
<b>Only Gradient boosting</b>		
Barthel Index: Transfer (less than full score)	60.0%	88.8%

<b>Barthel Index: Bladder control (less than full score)</b>	33.3%	65.8%
<b>Age</b>	72.6 (13.1)	75.4 (12.2)

**Table S10. Group averages for most stable and important input variables for deep vein thrombosis in 2016.**

	<b>Patients without DVT</b>	<b>Patients with DVT</b>
<b>Both logistic regression and gradient boosting</b>		
<b>Embolic etiology</b>	31.9%	51.9%
<b>Barthel Index: Mobility (less than full score)</b>	75.6%	82.5%
<b>Admission-NIHSS</b>	4 (6)	9 (12)
<b>Microangiopathic etiology</b>	20.0%	6.5%
<b>Admission-Rankin Scale</b>	3 (2)	4 (2)
<b>Barthel Index: Transfer (less than full score)</b>	76.0%	80.4%
<b>Only logistic regression</b>		
<b>Other etiology</b>	3.4%	7.4%
<b>Hypercholesterinaemia</b>	56.4%	44.8%
<b>Atherothrombotic etiology</b>	23.2%	15.7%
<b>June</b>	8.3%	11.6%
<b>Only Gradient boosting</b>		
<b>Age</b>	72.7 (13.1)	72.6 (13.1)
<b>Barthel Index: Bladder control (less than full score)</b>	33.8%	57.6%
<b>Impaired swallowing</b>	22.5%	45.4%
<b>Motor impairment</b>	66.1%	81.3%

**Table S11. Odds ratios for the ten most stable input variables of all four severe adverse outcomes.** Logistic regression models for all four outcomes were re-run in 100 downsampling scenarios, this time considering the most stably selected input variables (instead of backward stepwise selection).

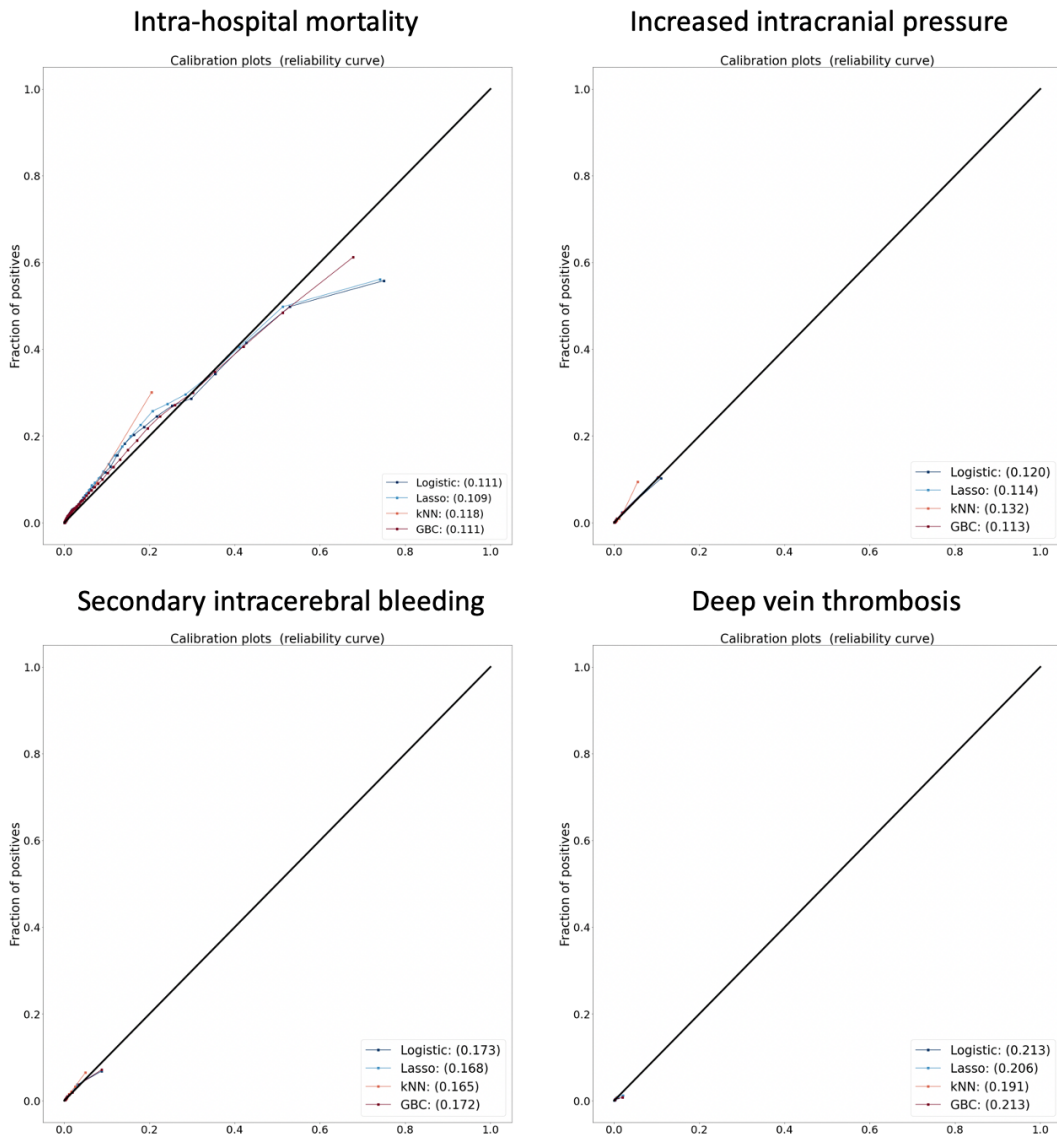
	<b>Odds ratio (95% confidence intervals)</b>
<b>In-Hospital mortality</b>	
<b>Admission-NIHSS per point increase</b>	1.08 (1.08-1.08)
<b>Age per year increase</b>	1.05 (1.05-1.05)
<b>Impaired swallowing</b>	2.15 (2.13-2.17)
<b>Microangiopathic etiology</b>	0.33 (0.32-0.33)
<b>Hypercholesterinaemia</b>	0.67 (0.66-0.67)
<b>Impaired consciousness</b>	2.13 (2.10-2.16)
<b>Barthel index: Bladder function (higher score higher function)</b>	0.95 (0.95-0.95)
<b>ICU-Admission</b>	1.68 (1.64-1.71)
<b>Sex (female)</b>	0.76 (0.75-0.77)

<b>Barthel Index: Transfer (higher score higher function)</b>	0.93 (0.93-0.93)
<b>Increased intracranial pressure</b>	
<b>Admission-NIHSS per point increase</b>	1.08 (1.07-1.08)
<b>Impaired swallowing</b>	2.75 (2.69-2.81)
<b>Impaired consciousness</b>	2.03 (1.99-2.08)
<b>Microangiopathic etiology</b>	0.21 (0.20-0.21)
<b>Age per year increase</b>	0.98 (0.98-0.98)
<b>Barthel Index: Mobility (higher score higher function)</b>	0.90 (0.90-0.90)
<b>Imaging before admission</b>	1.64 (1.60-1.68)
<b>Thrombectomy</b>	1.53 (1.50-1.57)
<b>Motor impairment</b>	0.68 (0.67-0.69)
<b>Prior level of care: Nursing home</b>	0.73 (0.70-0.75)
<b>Secondary intracerebral hemorrhages</b>	
<b>Thrombolysis</b>	3.49 (3.44-3.55)
<b>Microangiopathic etiology</b>	0.36 (0.36-0.37)
<b>Atherothrombotic etiology</b>	0.62 (0.61-0.63)
<b>Imaging before admission</b>	1.69 (1.66-1.73)
<b>Admission-Rankin Scale per point increase</b>	1.22 (1.21-1.23)
<b>Barthel Index: Mobility (higher score higher function)</b>	0.95 (0.95-0.96)
<b>Admission-NIHSS per point increase</b>	1.04 (1.04-1.04)
<b>Prior level of care: Nursing home</b>	0.67 (0.66-0.68)
<b>Impaired swallowing</b>	1.41 (1.38-1.43)
<b>Impaired language</b>	1.30 (1.28-1.32)
<b>Deep vein thrombosis</b>	
<b>Embolic etiology</b>	1.58 (1.53-1.64)
<b>Barthel Index: Mobility (higher score higher function)</b>	0.96 (0.96-0.96)
<b>Other etiology</b>	2.26 (2.09-2.45)
<b>Hypercholesterinaemia</b>	0.64 (0.63-0.66)
<b>Admission-NIHSS per point increase</b>	1.03 (1.03-1.04)
<b>Microangiopathic etiology</b>	0.51 (0.49-0.53)
<b>Admission-Rankin Scale per point increase</b>	1.16 (1.15-1.18)
<b>Barthel Index: Transfer (higher score higher function)</b>	0.99 (0.98-0.99)
<b>Atherothrombotic etiology</b>	0.68 (0.66-0.71)
<b>June</b>	1.87 (1.79-10.96)

**Table S12.** Test AUCs (95%-CI) in the training data from 2016.

<b>Classifier</b>	<b>Early mortality</b>	<b>Increased intracranial pressure</b>	<b>Secondary intracerebral hemorrhage</b>	<b>Deep vein thrombosis</b>
<b>Logistic</b>	0.91 (0.91-0.91)	0.89 (0.89-0.89)	0.80 (0.80-0.81)	0.70 (0.69-0.71)
<b>l1-regularized logistic regression</b>	0.91 (0.91-0.91)	0.89 (0.89-0.89)	0.80 (0.80-0.81)	0.71 (0.70-0.72)
<b>kNN</b>	0.90 (0.90-0.90)	0.88 (0.87-0.88)	0.79 (0.79-0.79)	0.71 (0.69-0.71)

<b>GBC</b>	0.90 (0.90-0.90)	0.89 (0.89-0.89)	0.81 (0.80-0.81)	0.71 (0.71-0.72)
------------	------------------	------------------	------------------	------------------



**Figure S1.** Calibration plots and Brier-Scores (c.f., figure legends) in the validation cohort. Calibration plots contrast the predicted probability of an outcome with the observed proportions. A model is considered well calibrated if those two measures correspond to one another closely. Brier scores measure the accuracy of probabilistic predictions, with lower values indicating a more favorable calibrations of predictions.