BMJ Open

External validation of risk prediction models for post-stroke mortality in Berlin

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-089320
Article Type:	Original research
Date Submitted by the Author:	27-May-2024
Complete List of Authors:	Reitzle, Lukas; Robert Koch Institute, Department of Epidemiology and Health Monitoring; Charité Universitätsmedizin Berlin, Center for Stroke Research Berlin Rohmann, Jessica; Charité Universitätsmedizin Berlin, Center for Stroke Research Berlin; Charité Universitätsmedizin Berlin, Institute of Public Health Kurth, Tobias; Charité Universitätsmedizin Berlin, Institute of Public Health Audebert, Heinrich; Charite Universitatsmedizin Berlin, Center for Stroke Research Berlin; Charité Universitätsmedizin Berlin, Klinik und Hochschulambulanz für Neurologie, Campus Benjamin Franklin Piccininni, Marco; Hasso Plattner Institute for Digital Engineering, Digital Health—Machine Learning Research Group; Charité Universitätsmedizin Berlin, Center for Stroke Research Berlin
Keywords:	Stroke < NEUROLOGY, Prognosis, Mortality, Clinical Decision-Making

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

External validation of risk prediction models for post-stroke mortality in Berlin

Authors: Lukas Reitzle, MD [1,2], Jessica L. Rohmann, PhD [1, 3], Tobias Kurth, MD, ScD [3], Heinrich J. Audebert, MD [1,4], Marco Piccininni, PhD [1,3,5,6]

[1] Center for Stroke Research Berlin, Charité – Universitätsmedizin Berlin, Berlin, Germany
[2] Department of Epidemiology and Health Monitoring, Robert Koch Institute, Berlin, Germany
[3] Institute of Public Health, Charité – Universitätsmedizin Berlin, Berlin, Germany
[4] Klinik und Hochschulambulanz für Neurologie, Campus Benjamin Franklin, Charité – Universitätsmedizin Berlin, Berlin, Germany
[5] Digital Health—Machine Learning Research Group, Hasso Plattner Institute for Digital Engineering, Potsdam, Germany
[6] Digital Engineering Faculty, University of Potsdam, Potsdam, Germany

Corresponding author:
Lukas Reitzle, MD
Robert Koch Institute
Department of Epidemiology and Health Monitoring
Nordufer 20
13353 Berlin
+49 30 18754 3738
reitzlel@rki.de

Address of all authors:
Jessica L. Rohmann: jessica.rohmann@charite.de
Tobias Kurth: tobias.kurth@charite.de
Heinrich J. Audebert: heinrich.audebert@charite.de
Marco Piccininni: marco.piccininni@hpi.de

Keywords: Stroke, prognosis, mortality, clinical decision-making

Word count: 3,088

Abstract

Objectives: Prediction models for post-stroke mortality can support medical decision-making. Although numerous models have been developed, external validation studies determining the models' transportability beyond the original settings are lacking. We aimed to assess the performance of two prediction models for post-stroke mortality in Berlin, Germany.

Design: We used data from the Berlin-SPecific Acute Treatment in Ischemic or hAemorrhagic stroke with Long term follow–up (B–SPATIAL) registry.

Setting: Multicenter stroke registry in Berlin, Germany

Participants: Adult patients, admitted within 6 hours after symptom onset and with an ICD-10 discharge diagnosis of ischemic stroke, hemorrhagic stroke or transient ischemic attack at one of 15 hospitals with stroke units between January 1st, 2016 and January 31st, 2021.

Primary Outcome Measures: We evaluated calibration (calibration-in-the-large and calibration plot) and discrimination performance (c-statistic) of *Bray et al*.'s 30-day mortality and *Smith et al*.'s in-hospital mortality prediction models. Information on mortality was supplemented by Berlin city registration office records.

Results: For the validation of *Bray et al*.'s model, we included 7,879 patients (mean age 75; 55.0% men). We observed 763 (9.7%) deaths within 30 days of stroke compared to 680 (8.6%) predicted. The model's c-statistic was 0.865 (95%CI: 0.851-0.879). For *Smith et al*.'s model, we performed the validation among 1,931 patients (mean age 75; 56.2% men), observing 105 (5.4%) in-hospital deaths compared to the 92 (4.8%) predicted. The c-statistic was 0.891 (95%CI: 0.864-0.918). The calibration plots of both models revealed an underestimation of the mortality risk for high-risk patients.

Conclusions: Among Berlin stroke patients, both models showed high discrimination and good overall calibration performance, despite underestimation of risk among high-risk patients. The good performance of *Bray et al*.'s model in Berlin illustrates how a small number of routinely collected variables can be sufficient for valid prediction of post-stroke mortality.

Strengths and limitations of this study

- The prospective, multicenter B-SPATIAL registry included all stroke/TIA patients presenting to 15 Berlin hospitals with stroke units during a 5-year period.
- Loss-to-follow-up was low, data completeness was high, and outcome information was reliable.
- Since Berlin is a densely populated city with several stroke units, our findings may be transportable to urban Central European areas but not necessarily to other settings.
- The prediction model for in-hospital mortality after stroke by Smith et al. could only be validated with data from 3 of the 15 hospitals, routinely collecting data on all relevant predictors.

Introduction

In 2019, stroke was the second leading cause of death and the third leading cause of combined death and disability worldwide.[1] In the context of stroke aftercare, prediction models have been developed to predict functional outcomes and mortality risk after acute stroke. These tools support (shared) clinical decision-making by providing information about likely prognosis to health professionals, patients, and their families.[2, 3] Yet, before implementing prediction models in routine clinical practice, the transportability from the original development population to the population of interest should be assessed; models with low performance in the setting of interest may generate non-accurate predictions and lead to sub-optimal decisions.[4]

According to the systematic review of *Fahey et al.,* prior to September 2015, 38 prediction models for post-stroke mortality had been developed.[2] Despite the abundance of existing prediction models for post-stroke outcomes, only a small fraction have been externally validated.[2]

Two models including routinely collected variables were developed by *Bray et al*. (2014) using data from the Sentinel Stroke National Audit Program (SSNAP) in the United Kingdom[3] and by *Smith et al.* (2010) using data from the Get With the Guidelines (GWTG) Stroke Program in the United States.[5] Though these models have already been subjected to validation studies in their respective originating countries[6], to date, both models have only undergone external validation in the China National Stroke Registry.[7-9] Our study thus aimed to assess calibration and discrimination performance of *Bray et al.* (2014)[3] and *Smith et al.* (2010)[5] prediction models for post-stroke mortality among Berlin stroke patients.

Methods

Data source

We used data from the Berlin – SPecific Acute Treatment in Ischemic or hAemorrhagic stroke with Long term follow–up (B–SPATIAL) registry (Clinicaltrials.gov identifier: NCT03027453), a multicenter

registry for adult stroke patients in Berlin. Data were collected from patients aged 18 years or older, admitted within 6 hours after symptom onset and with ICD-10 discharge diagnoses of ischemic stroke (I63/I64), hemorrhagic stroke (I61), non-traumatic subdural hemorrhage (I62) or transient ischemic attack (TIA; G45.0–G45.3 and G45.5–G45.9) at one of 15 hospitals with stroke units in Berlin, Germany, between January 1st, 2016 and January 31st, 2021. Patients with no symptoms upon arrival of emergency medical services and without neurological symptoms at hospital arrival were not included in the registry. In this external validation study, we did not include patients for whom a mobile stroke unit was dispatched, as part of the B_PROUD interventional study, which was linked to the registry.[10] We further excluded patients who opted out of data collection.[11]

Prediction models for post-stroke mortality

We evaluated the performance of *Bray et al.'s model A* (2014) including the full NIHSS (all items) for 30-day all-cause mortality[3] (hereafter: *Bray et al.*'s model) and *Smith et al*.'s model (2010) including the NIHSS for in-hospital mortality.[5]

Predictors

Bray et al.'s model included the following predictors: age group (<60, 60-69, 70-79, 80-89, ≥90 years), stroke type (ischemic or hemorrhagic), atrial fibrillation and NIHSS at admission.[3] Using the B-SPATIAL registry data, stroke type was determined using available ICD-10 codes (I63 or I64 for ischemic, I61 for hemorrhagic). Atrial fibrillation was considered present if the patient had a known history of atrial fibrillation or if atrial fibrillation was diagnosed by the emergency medical service or at admission. *Smith et al*.'s model included the following predictors: age as a continuous variable, sex (male vs. non-male), NIHSS at admission, atrial fibrillation, history of stroke or TIA, coronary artery disease, diabetes mellitus and dyslipidemia.[5] Additionally, the model included a variable indicating the mode of hospital arrival, categorized as arrival by private transport, by ambulance, or other arrival not via the emergency department (ED) (e.g. direct admission from the hospital ward) as a predictor. For the

purposes of this external validation, we assumed a prior history of stroke or TIA if indicated by imaging performed while in hospital or if documentation of an ischemic stroke or TIA was available. We defined the presence of coronary artery disease as documented previous myocardial infarction, coronary stent placement, or corresponding diagnostic coronary angiography result. In the B-SPATIAL registry, diabetes was defined as documented history of diabetes, the use of anti-diabetic medication, or a measured A1C level above 6.5% or blood glucose above 200 mg/dl (non-fasting) or 126 mg/dl (fasting). We defined dyslipidemia to include either a reported history of the condition, having measured Low-Density-Lipoprotein (LDL)-cholesterol levels above 130 mg/dl, or having total cholesterol levels above 220 mg/dl. In line with the original publication, in cases of missing documentation or unknown mode of arrival, we assumed arrival by private transport. For the few cases with documented secondary transfer but no documentation of transfer from an external hospital, we assumed the patient was internally transferred within the same hospital and thus did not arrive via the ED.

Outcomes

While *Bray et al.*'s model used 30-day all-cause mortality as outcome, *Smith et al.*'s model predicted in-hospital mortality. In the B-SPATIAL registry, we defined in-hospital mortality as death documented as the discharge reason or an mRS score of 6 at discharge. In cases where both documentation of discharge reason and mRS at discharge were missing, we assumed patients were alive at discharge. To create the 30-day all-cause mortality variable, we considered patients with in-hospital death and hospital stays ≤30 days, and patients for whom the date of death was within 30 days of stroke. We obtained information about the date of death from the Berlin city registration office at two and four months after stroke.[11]

Study population

For *Bray et al.*'s model, we included patients with either acute ischemic or hemorrhagic stroke. For *Smith et al.*'s model, we included only ischemic stroke patients, similar to inclusion in the original publication (which retrospectively identified patients using the ICD-9 codes). Since the predictors history of stroke or TIA, coronary artery disease, and dyslipidemia were only routinely recorded in three of the B-SPATIAL registry hospitals, we excluded patients from the remaining hospitals for the validation of *Smith et al.*'s model. For both models, we excluded patients who were transferred from a hospital not participating in the B-SPATIAL registry or with missing values for one of the predictors or the outcome. When information about transfer status was missing, we assumed the patient was not transferred.

Analyses of the data from the B-SPATIAL registry, including the external validation of clinical risk scores, were approved by the ethics committee of the Charité - Universitätsmedizin Berlin (EA1/208/21). The B-SPATIAL registry used an opt-out mechanism for patient inclusion. Two months after their index event, patients were informed in writing about the inclusion of their record in the B-SPATIAL registry and had multiple opportunities to opt out.[11]

Statistical analysis

We used the prediction models' published formulas to calculate risk of 30-day all-cause mortality[3] or in-hospital mortality[5] for each included individual (see Supplemental R Code).

To assess model calibration, we evaluated the calibration-in-the-large by comparing the actual ("observed") number of deaths to the one predicted by the model ("expected") using the observed-to-expected (O/E) event ratio. We then used a calibration plot to graphically compare the observed mortality risk with the mean predicted risk within decile groups of predicted risk. We estimated 95% confidence intervals for the observed risk using the exact method. Furthermore, we calculated the calibration intercept and slope using the logistic calibration framework.[12] We assessed the discriminatory ability of the two prediction models by calculating the concordance statistic (c-

statistic) and visualizing the receiver operating characteristics (ROC)-curve. We computed 95% confidence intervals for the c-statistic based on 2,000 stratified bootstrap replicates.

For both models, we conducted sensitivity analyses. For *Smith et al*.'s model, to assess the robustness of our assumption of arrival by private transport when the mode of arrival was missing or unknown, we reran our analysis excluding patients with unknown or missing mode of arrival. In the original prediction model development studies, *Smith et al*. explicitly excluded patients with TIA, and *Bray et al*. did not specify how these patients were handled in the data management stage.[3, 5] However, at the time of admission, TIA patients presenting with neurological symptoms compatible with stroke are not distinguishable from ischemic stroke patients. Therefore, in an additional sensitivity analysis, we investigated the performance of both models when classifying all patients with final diagnosis of TIA as ischemic stroke patients.

All statistical analyses were performed using R v4.2.1 and RStudio 2022.07.1. The rms package was used for calculation of calibration intercept and slope and the pROC package for the c-statistic and ROC-curve.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this study.

Results

***Bray et al*.'s model for 30-day all-cause mortality**

We included 7,879 stroke patients in the external validation of *Bray et al*.'s model (**Figure 1**). The median age of the B-SPATIAL patients included in this validation was 75 years and 55.0% were male (**Table 1**). A final diagnosis of ischemic stroke was considerably more common (92.4%) than hemorrhagic stroke. Median NIHSS at admission was 5 (IQR:2-11). In total, 763 (9.7%) of included patients died within 30 days of admission. We found that *Bray et al.*'s model underestimated the

mortality risk, predicting an average mortality of 8.6%, corresponding to 680 deaths. The resulting observed-to-expected ratio was 1.12, indicating that 12% more deaths were observed in our sample than were predicted by the model. The calibration plot showed good alignment of predicted and observed mortality for eight decile groups (**Table S1**, **Figure 2**). The model underestimated the observed mortality risk for the two decile groups with the highest predicted mortality. The calibration intercept was 0.34 (95%CI: 0.19-0.49) and the slope was 1.10 (95%CI: 1.03-1.17). **Figure 2** shows the ROC-curve illustrating the discriminatory ability of the model in our sample. The c-statistic was 0.865 (95%CI: 0.851-0.879). In the sensitivity analysis for *Bray et al*.'s model, we included an additional 1,932 patients ultimately diagnosed with TIA into the validation sample. Among the 9,811 ischemic stroke/TIA patients, the observed 30-day mortality was 7.9%, which was very similar to the 7.4% predicted by *Bray et al*.'s model. The calibration plot, calibration intercept (0.37 [95%CI: 0.22-0.52]) and slope (1.15 [95%CI: 1.08-1.21]) from this sensitivity analysis were similar to the ones of the main analysis, and the c-statistic was 0.880 (95%CI: 0.867-0.893) (**Figure S1**).

***Smith et al*.'s model for in-hospital mortality**

For the external validation of *Smith et al*.'s prediction model, we included 1,931 ischemic stroke patients (**Figure 1**). The median age in this sample was 75 and 56.2% were male (**Table 2**). The median NIHSS was 4 (IQR: 2-10) and most patients arrived by ambulance. In total, 105 (5.4%) ischemic stroke patients died in hospital.

Smith et al.'s model predicted an average risk of in-hospital mortality of 4.8%, equal to 92 deaths. The corresponding observed-to-expected ratio of 1.14 indicated an underestimation of in-hospital mortality by 14%. The calibration plot revealed that the model underestimated the mortality risk in the decile groups with highest predicted mortality. For decile groups with lower risk, the observed and predicted risk were well-aligned, albeit with high uncertainty, as only few events were observed (**Table S2**, **Figure 3**). The corresponding calibration intercept was 1.20 (95%CI: 0.66-1.76), and the slope was 1.43 (95%CI: 1.22-1.66). We depicted the discriminatory ability of the model predicting in-

hospital mortality as a ROC-curve (**Figure 3**). The corresponding c-statistic was 0.891 (95%CI: 0.864-0.918). In the sensitivity analysis for *Smith et al*.'s model, in which we excluded all patients with unknown or missing mode of arrival, the observed in-hospital mortality was 5.6% compared to 5.0% predicted by the model. The calibration plot, calibration intercept (1.11 [95%CI: 0.56-1.67]) and slope (1.41 [95%CI: 1.19-1.64]), as well as the c-statistic of 0.883 (95%CI: 0.854-0.912), were comparable to those estimated in the main analysis (**Figure S2**). In the second sensitivity analysis, we additionally included 597 TIA patients. Among the 2,528 included ischemic stroke/TIA patients, 4.3% died in the hospital compared to a mortality of 4.0% predicted by the model. The calibration plot, calibration intercept (1.24 [95%CI: 0.71-1.77]) and slope (1.47 [95%CI: 1.27-1.68]), as well as the c-statistic (0.902 [95%CI: 0.875-0.929]) obtained in this second sensitivity analysis only slightly deviated from the main analysis results (**Figure S3**).

Discussion

In this study, we externally validated two prognostic prediction models for mortality after stroke; *Bray et al*.'s model for 30-day all-cause mortality and *Smith et al*.'s model for in-hospital mortality, using data from a multicenter registry of adult stroke patients in Berlin, Germany.

Bray et al.'s prediction model was developed in the United Kingdom in 2014 using data from the SSNAP, which is the national registry for acute stroke in England and Wales.[3] The original publication included an external validation study using data from the South London Stroke Register, which showed good calibration performance and high discrimination of the model, with a c-statistic of 0.87.[3] The model was later externally validated in two other studies. The first was a temporal validation study using SSNAP data from a different time period, which found a slightly worse discrimination (c-statistic of 0.774).[6] The second was conducted in the China National Stroke Registry; despite substantial differences in the study population's composition, the model showed a good discrimination ability (c-statistic of 0.80) and good calibration in this setting.[8] Our findings add to this

evidence body; in the Berlin setting, we found good alignment between the predicted and observed mortality risk with a slight underestimation among high risk patients. We also observed a higher discriminatory performance of the model in our setting (c-statistic of 0.865).

Smith et al.'s prediction model was developed using data from the GWTG Stroke Program in the United States.[5] Thereafter, two external validation studies using different cohorts from the China National Stroke Registry found good calibration and high discriminatory ability of this model with c-statistic of 0.867[9] and 0.86[7], similar to the results of the internal validation from the development study (c-statistic of 0.85[5]). In the Berlin setting, we observed high discriminatory performance (c-statistic of 0.891); however, the *Smith et al.* model underestimated the risk in high-risk individuals and the calibration intercept and slope indicated a less optimal calibration in our setting compared to the original publication and both external validation studies. More uncertainty was present for this validation, due to the low sample size.

The calibration of both models shows an underestimation of the mortality risk especially for high risk-patients in Berlin. The underestimation could have been due to different factors. In our analysis, we excluded patients who have opted out of study participation. Opting-out was only possible if patients survived the stroke, which might have introduced a selection of stroke cases with higher severity in our sample.

For the validation of both models, we excluded patients with final diagnosis of TIA in our main analysis, since in the original publications and validation studies, TIA patients were either explicitly excluded[5, 6] or their inclusion was not specified.[3, 7-9] Within the Berlin setting, both models performed similarly or even better after the inclusion of TIA patients. Different definitions of TIA exist, and the diagnostic discrepancy may explain why researchers might hesitate to include patients with TIA in studies for the development of prediction models. However, since patients with a TIA present with symptoms comparable to an ischemic stroke upon admission, from a clinical and methodological perspective, we believe future work should consider including TIA patients also in the development and validation of post-stroke prediction models.

A systematic review of prediction models for post-stroke outcomes found that models with a high number of predictors do not necessarily show a better performance.[2] Models with few variables such as *Bray et al*.'s model, including only four predictors, already showed high discrimination.[3, 8] Models that perform sufficiently well with fewer, routinely-measured variables should be preferred over models with several predictors, since they are more likely to be used in practice. For this reason, as has also been argued for other clinical applications,[13, 14] we believe that future prediction model studies in the context of post-stroke outcomes should compare newly developed models' performance with that of well-established models, or preferably focus on the external validation and updating of existing models rather than the development of new ones.

Strengths and limitations

The strengths of this study include the prospective, multicenter design of the B-SPATIAL registry with coverage of all 15 Berlin stroke units during a 5-year period. Therefore, the B-SPATIAL registry for adult stroke patients can be considered representative for the population of stroke patients in Berlin and comprises detailed information on demographics and clinical characteristics, with low loss to follow-up, especially for mortality endpoints.[11] The recording of vital status during follow-up is considered particularly reliable because the information was supplemented by city registration office records.

Some limitations should be considered when interpreting our results. As Berlin is a densely populated city with several stroke units, the availability of stroke care might be different compared to other regional settings in Germany and Central Europe. Therefore, our results may not generalize to different settings, such as rural areas. Furthermore, the B-SPATIAL registry only contains information on patients with hospital arrival within 6 hours of symptom onset, since this was the time window for reperfusion treatment eligibility when the registry commenced. However, a substantial proportion of stroke patients present to hospitals later than 6 hours after onset,[15, 16] and the performance of these prediction models might differ for these patients.

Only three of 15 registry hospitals routinely documented history of hyperlipidemia, coronary artery disease and history of stroke or TIA. Therefore, we could only validate *Smith et al*.'s model in this subsample, which composed 30% of the full validation sample. Furthermore, as overall in-hospital mortality risk was low in our setting, only 105 in-hospital deaths were observed in this subsample, which decreased the power of the analysis and somewhat limits the interpretation of the calibration plot due to higher uncertainty, especially in the lower decile groups, in which only few events were observed. For both models, we excluded patients with missing information on at least one of the predictors (except mode of arrival), as this information was missing only for a small number of stroke patients.

Conclusion

Despite being developed outside of Germany, the external validation of *Smith et al*.'s model and *Bray et al*.'s model for post-stroke mortality both demonstrated good overall calibration in a large stroke registry in Berlin. Both models showed high discrimination ability, but underestimated risk in high-risk patients. The performance of *Bray et al*.'s model indicated a good transportability to the Berlin setting and illustrates how a small number of variables that can be routinely obtained at hospital admission can be sufficient for valid prediction of post-stroke mortality.

Acknowledgment section

Contributions

MP and JR conceived the study. LR, TK, JR and MP designed the study. LR selected the models for external validation. JLR and HJA provided B-SPATIAL registry support, for which HJA obtained funding. The statistical analyses were planned and conducted by LR in consultation with MP and JR. LR created the R code, tables, and figures, which were revised by MP. LR wrote the first draft of the manuscript with input from JR and MP. MP provided project supervision. All authors reviewed and edited the manuscript for intellectual content and approved the final version of the manuscript.

Data Access, Responsibility, and Analysis

Dr. Reitzle had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Acknowledgements

We are grateful to all collaborating hospitals and the study nurses for their engagement and thank Jakob Beilstein for assistance with data management.

Declaration of Conflicting Interests

Dr Reitzle reports having received research grants from the Federal Ministry of Health in Germany.

Dr Rohmann reports having received a grant from Novartis Pharma for conducting a self-initiated research project about migraine.

Dr Kurth reports having received research grants from the Federal Ministry of Health in Germany. He also received personal compensation from Eli Lilly and Company, The BMJ, and Frontiers.

Dr Audebert reports personal fees from Bayer Vital, Boehringer Ingelheim, Bristol- Myers Squibb, Novo Nordisk, Pfizer, and from Sanofi.

Dr Piccininni reports having received partial funding from Novartis Pharma and being awarded a research grant from the Center for Stroke Research Berlin (private donations).

Funding/Support

The B-SPATIAL registry was funded by the German Federal Ministry of Education and Research (BMBF) and the German Research Foundation (DFG), granted to the Center for Stroke Research Berlin and by the DFG to the Excellence Cluster NEUROCURE.

Role of the Funder/Sponsor

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; nor decision to submit the manuscript for publication.

Data sharing statement

B-SPATIAL registry data can be made available in a de-identified manner to researchers who provide a methodologically sound proposal (to the extent allowed by the registry's data protection agreement). Data access requests should be directed to jessica.rohmann (at) charite.de.

References

1. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol* 2021; 20: 795-820. DOI: 10.1016/s1474-4422(21)00252-0.

2. Fahey M, Crayton E, Wolfe C and Douiri A. Clinical prediction models for mortality and functional outcome following ischemic stroke: A systematic review and meta-analysis. *PLoS One* 2018; 13: e0185402. DOI: 10.1371/journal.pone.0185402.

3. Bray BD, Campbell J, Cloud GC, et al. Derivation and external validation of a case mix model for the standardized reporting of 30-day stroke mortality rates. *Stroke* 2014; 45: 3374-3380. DOI: 10.1161/strokeaha.114.006451.

4. Debray TP, Vergouwe Y, Koffijberg H, et al. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015; 68: 279-289. DOI: 10.1016/j.jclinepi.2014.06.018.

5. Smith EE, Shobha N, Dai D, et al. 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.

6. Dutta D, Cannon A and Bowen E. Validation and comparison of two stroke prognostic models for in hospital, 30-day and 90-day mortality. *European Stroke Journal* 2017; 2: 327-334. DOI: 10.1177/2396987317703581.

7. Sun S, Pan Y, Bai L, et al. GWTG Risk Model for All Stroke Types Predicts In-Hospital and 3-Month Mortality in Chinese Patients with Acute Stroke. *J Stroke Cerebrovasc Dis* 2019; 28: 800-806. DOI: 10.1016/j.jstrokecerebrovasdis.2018.11.024.

8. Yu P, Pan Y, Wang Y, et al. External Validation of a Case-Mix Adjustment Model for the Standardized Reporting of 30-Day Stroke Mortality Rates in China. *PLoS One* 2016; 11: e0166069. DOI: 10.1371/journal.pone.0166069.

9. Zhang N, Liu G, Zhang G, et al. A risk score based on Get With the Guidelines-Stroke program data works in patients with acute ischemic stroke in China. *Stroke* 2012; 43: 3108-3109. DOI: 10.1161/strokeaha.112.669085.

10. Ebinger M, Siegerink B, Kunz A, et al. Association Between Dispatch of Mobile Stroke Units and Functional Outcomes Among Patients With Acute Ischemic Stroke in Berlin. *Jama* 2021; 325: 454-466. DOI: 10.1001/jama.2020.26345.

11. Napierkowski I, Lorenz-Meyer I, Hille A, et al. Follow-up of Patients With Stroke, Based on Opt-Out Choice Potential Approach for Acute Care Quality Registries or Observational Studies. *Neurology* 2022; 99: e1335-e1344. DOI: 10.1212/wnl.0000000000200916.

12. Van Calster B, Nieboer D, Vergouwe Y, et al. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016; 74: 167-176. DOI: 10.1016/j.jclinepi.2015.12.005.

13. Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016; 353: i2416. DOI: 10.1136/bmj.i2416.

14. Steyerberg EW and Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014; 35: 1925-1931. 20140604. DOI: 10.1093/eurheartj/ehu207.

15. Hillmann S, Wiedmann S, Rücker V, et al. Stroke unit care in germany: the german stroke registers study group (ADSR). *BMC Neurol* 2017; 17: 49. DOI: 10.1186/s12883-017-0819-0.

16. Tong D, Reeves MJ, Hernandez AF, et al. Times from symptom onset to hospital arrival in the Get with the Guidelines--Stroke Program 2002 to 2009: temporal trends and implications. *Stroke* 2012; 43: 1912-1917. DOI: 10.1161/strokeaha.111.644963.

Figure legends

Figure 1. Flow chart showing eligibility criteria applied to the B-SPATIAL study population for the external validation of Bray et al.'s model (2014) for post-stroke 30-day all-cause mortality [3] (left) and Smith et al.'s model (2010) for post-stroke in-hospital mortality [5] (right). *Patients, who opted out or for whom a mobile stroke unit was dispatched as part of the B-PROUD study were not included.

Figure 2. External validation of Bray et al.'s model for post-stroke 30-day all-cause mortality in the B-SPATIAL registry (Main analysis). Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.

Figure 3. External validation of Smith et al.'s model for post-stroke in-hospital mortality in the B-SPATIAL registry (main analysis). Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.

Tables

Table 1. Characteristics of study population of stroke patients from the B-SPATIAL registry included in the external validation of Bray et al.'s model for post-stroke 30-day all-cause mortality, stratified by outcome status.

Characteristic		Alive at 30d N = 7,116	Deceased at 30d N = 763	All N = 7,879
Age	median (IQR)	75 (63, 81)	84 (78, 90)	75 (64, 82)
Male sex	n (%)	3,994 (56.1%)	342 (44.8%)	4,336 (55.0%)
Stroke type				
Hemorrhagic	n (%)	445 (6.3%)	156 (20.4%)	601 (7.6%)
Ischemic	n (%)	6,671 (93.7%)	607 (79.6%)	7,278 (92.4%)
NIHSS at admission	median (IQR)	4 (2, 9)	17 (10, 22)	5 (2, 11)
Atrial fibrillation	n (%)	1,937 (27.2%)	353 (46.3%)	2,290 (29.1%)

NIHSS: National Institutes of Health Stroke Scale; IQR: Interquartile range

Table 2. Characteristics of study population of ischemic stroke patients from the B-SPATIAL registry included in the external validation of Smith et al.'s model for post-stroke in-hospital mortality, stratified by outcome status

Characteristic		Alive at discharge N = 1,826	Dead In-hospital N = 105	All N = 1,931
Age	median (IQR)	74 (62, 81)	83 (78, 88)	75 (63, 82)
Male sex	n (%)	1,038 (56.8%)	47 (44.8%)	1,085 (56.2%)
NIHSS at admission	median (IQR)	4 (2, 9)	17 (13, 22)	4 (2, 10)
Mode of arrival				
private	n (%)	290 (15.9%)	4 (3.8%)	294 (15.2%)
not via ED	n (%)	68 (3.7%)	2 (1.9%)	70 (3.6%)
ambulance	n (%)	1,468 (80.4%)	99 (94.3%)	1,567 (81.1%)
Atrial fibrillation	n (%)	521 (28.5%)	50 (47.6%)	571 (29.6%)
Previous stroke or TIA	n (%)	489 (26.8%)	22 (21.0%)	511 (26.5%)
Coronary artery disease	n (%)	309 (16.9%)	25 (23.8%)	334 (17.3%)
Diabetes mellitus	n (%)	438 (24.0%)	31 (29.5%)	469 (24.3%)
Dyslipidemia	n (%)	1,119 (61.3%)	50 (47.6%)	1,169 (60.5%)

NIHSS: National Institutes of Health Stroke Scale; IQR: Interquartile range;
ED: Emergency department; TIA: Transient ischemic attack

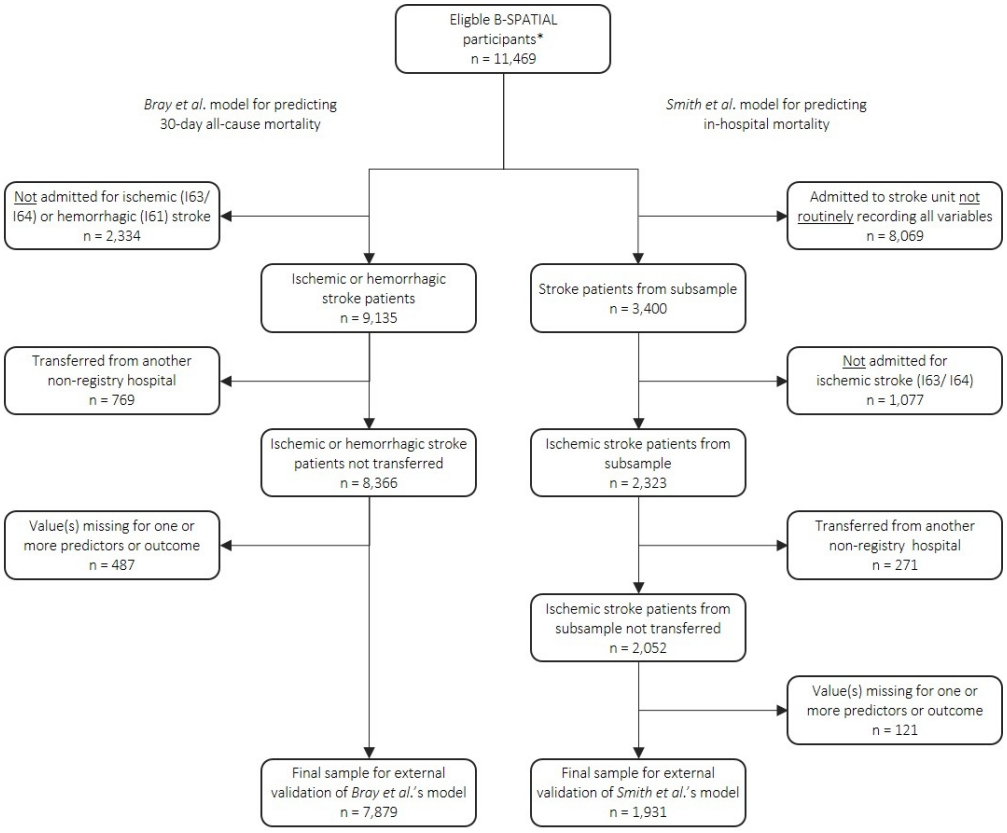


Figure 1. Flow chart showing eligibility criteria applied to the B-SPATIAL study population for the external validation of Bray et al.'s model (2014) for post-stroke 30-day all-cause mortality 3 (left) and Smith et al.'s model (2010) for post-stroke in-hospital mortality 9 (right). *Patients, who opted out or for whom a mobile stroke unit was dispatched as part of the B-PROUD study were not included.

181x150mm (150 x 150 DPI)

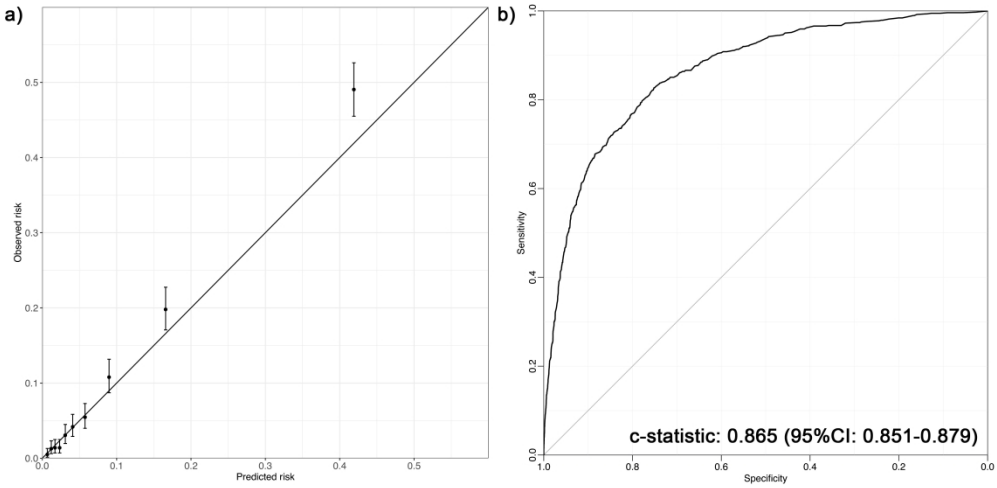


Figure 2. External validation of Bray et al.'s model for post-stroke 30-day all-cause mortality in the B-SPATIAL registry (Main analysis). Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.

1375x661mm (96 x 96 DPI)

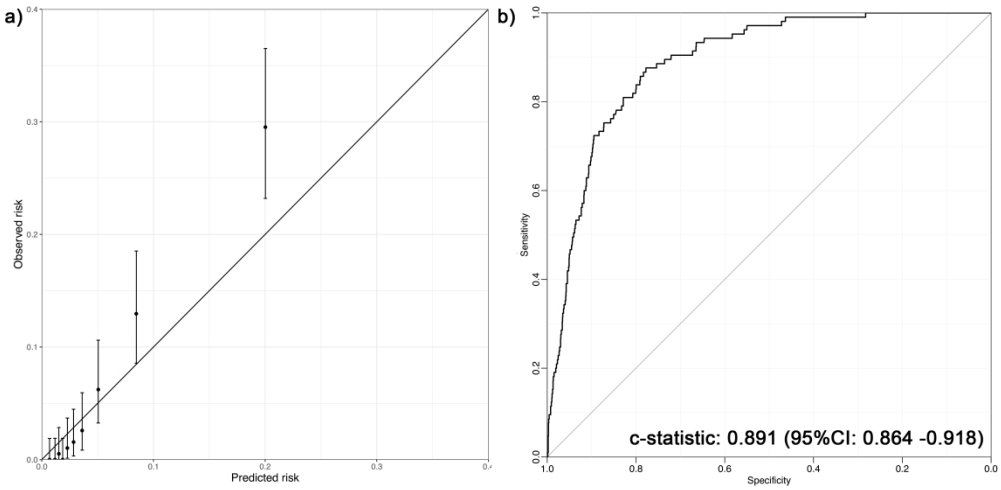


Figure 3. External validation of Smith et al.'s model for post-stroke in-hospital mortality in the B-SPATIAL registry (main analysis). Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.

1834x881mm (72 x 72 DPI)

Supplement - External validation of risk prediction models for post-stroke mortality in Berlin

Supplemental Tables

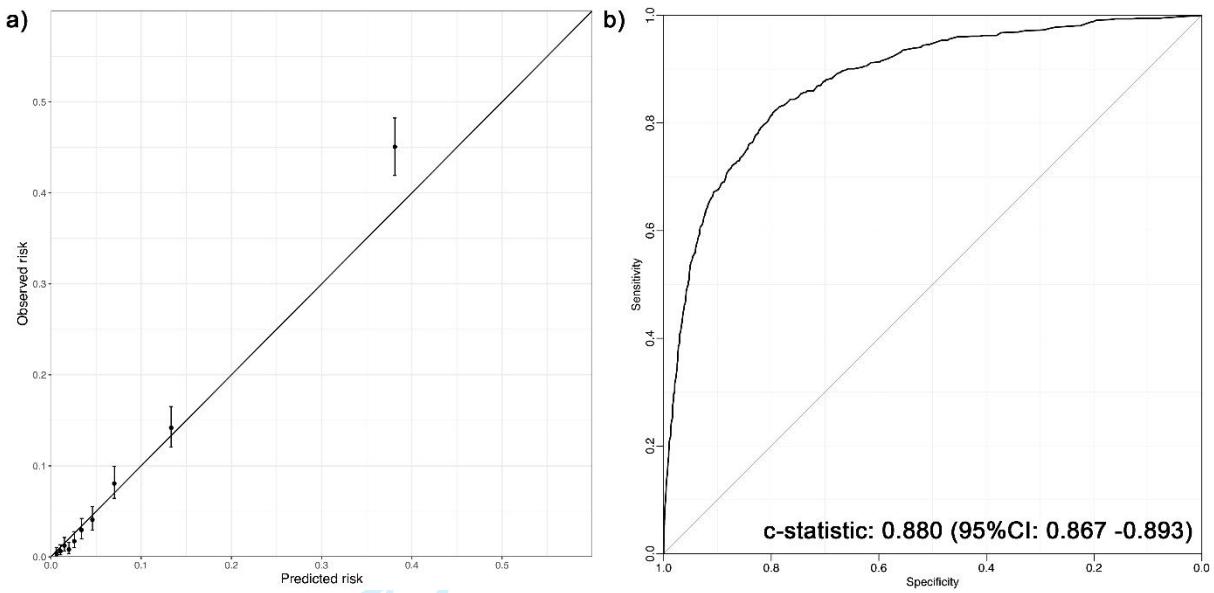
Supplementary Table 1. *External validation of Bray et al.'s model for post-stroke 30-day all-cause mortality in the B-SPATIAL registry (Main analysis). Observed and predicted 30-day mortality risk after stroke in decile groups.*

Decile	n	Predicted mortality	Observed mortality	
		%	%	95%CI
1	788	0.68	0.51	0.14 - 1.29
2	788	1.20	1.27	0.61 - 2.32
3	788	1.70	1.40	0.70 - 2.48
4	788	2.33	1.40	0.70 - 2.48
5	788	3.09	3.05	1.96 - 4.50
6	788	4.08	4.19	2.90 - 5.83
7	788	5.74	5.46	3.98 - 7.28
8	788	8.98	10.8	8.71 - 13.2
9	788	16.6	19.8	17.1 - 22.8
10	787	41.9	49.0	45.5 - 52.6

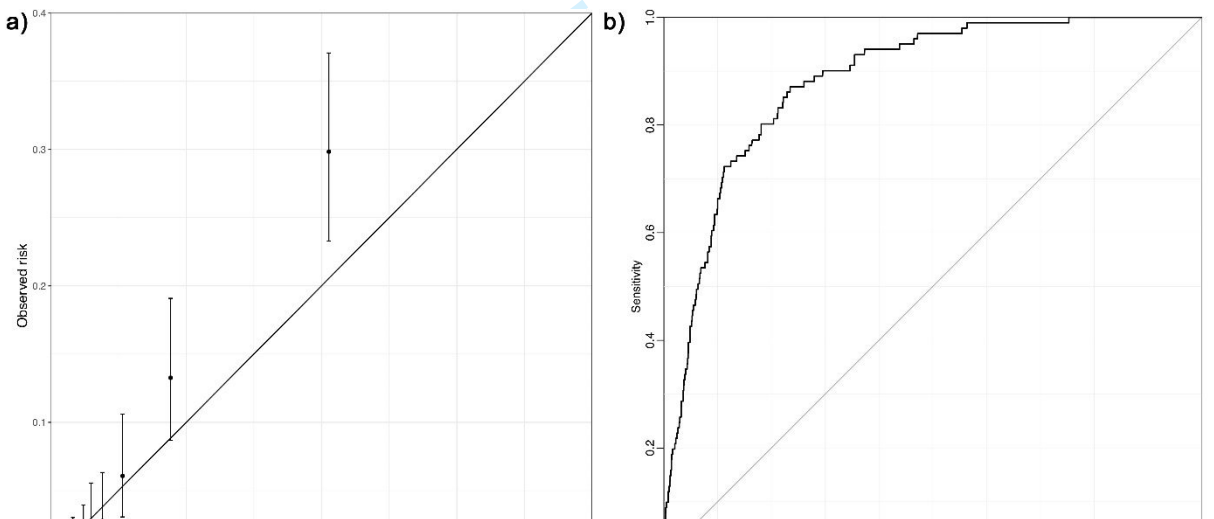
Supplementary Table 2. *External validation of Smith et al.'s model for post-stroke in-hospital mortality in the B-SPATIAL registry (Main analysis). Observed and predicted in-hospital mortality risk after stroke in decile groups.*

Decile	n	Predicted mortality	Observed mortality	
		%	%	95%CI
1	194	0.69	0.00	0.00 - 1.88
2	193	1.18	0.00	0.00 - 1.89
3	193	1.52	0.52	0.01 - 2.85
4	193	1.87	0.00	0.00 - 1.89
5	193	2.29	1.04	0.13 - 3.69
6	193	2.84	1.55	0.32 - 4.48
7	193	3.61	2.59	0.85 - 5.94
8	193	5.05	6.22	3.25 - 10.6
9	193	8.46	13.0	8.56 - 18.5
10	193	20.0	29.5	23.2 - 36.5

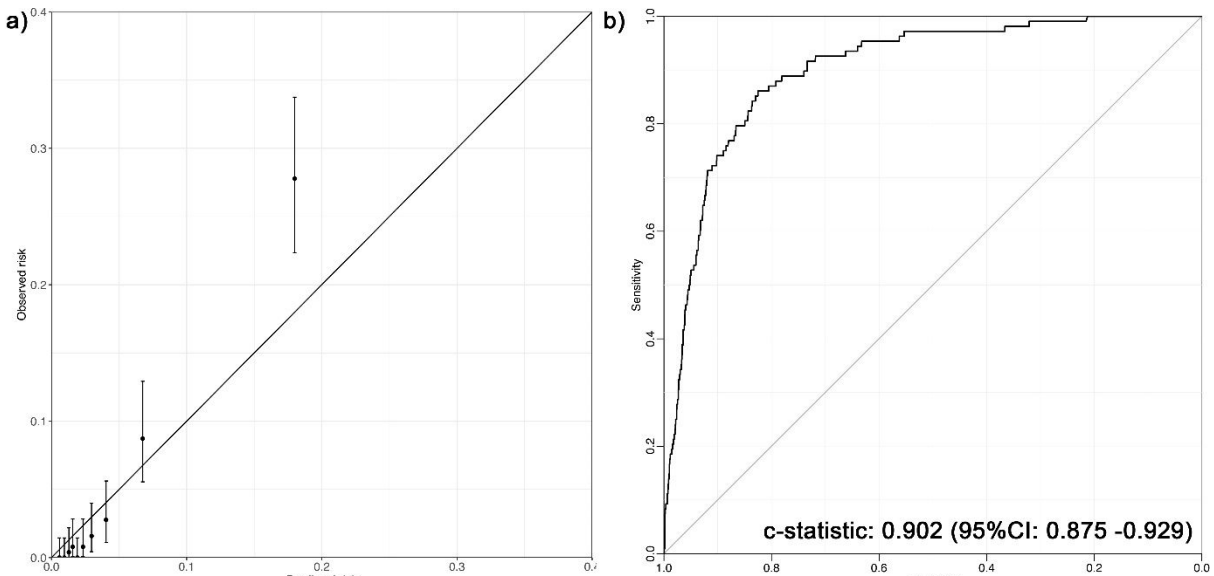
Supplement Figures



Supplementary Figure 1. *Sensitivity analysis for the external validation of Bray et al.'s model for post-stroke 30-day all-cause mortality in the B-SPATIAL registry. In this sensitivity analysis patients diagnosed with TIA were included as ischemic stroke patients. Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC curve).*



Supplementary Figure 2. *Sensitivity analysis 1 for the external validation of Smith et al.'s model for post-stroke in-hospital mortality in the B-SPATIAL registry. In the first sensitivity analysis all patients with unknown or missing mode of arrival were excluded. Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.*



Supplementary Figure 3. *Sensitivity analysis 2 for the external validation of Smith et al.'s model for post-stroke in-hospital mortality in the B-SPATIAL registry. In the second* ***s****ensitivity analysis patients diagnosed with TIA were included as ischemic stroke patients. Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.*

R code

```
# Title: External validation of prediction models for post-stroke mortality within
the "Berlin - SPecific Acute Treatment in Ischemic or hemorrhAgic stroke with Long
term follow-up" (B-SPATIAL) registry

# Date: 2024-02-02

# Note: the following packages must be installed (if not already installed):
# install.packages("tidyverse")
# install.packages("lubridate")
# install.packages("rms")
# install.packages("gtsummary")
# install.packages("fastDummies")
# install.packages("pROC")




# External validation of Bray et al.'s model A for 30-day mortality after stroke

# Definition of study population (Main analysis): Bray et al.'s model A
#
# Exclusion of patients…
# - with no ischemic or hemorrhagic stroke (Variable: stroke_type)
# - transferred to or form another hospital (Variable: transfer)
# - with missing age or age <18 or >120 (Variable: age)
# - with missing information on atrial fibrillation (Variable: atr_fibr)
# - with missing information on NIHSS at admission (Variable: nihss_adm)
# - with implausible NIHSS (<0 or >42)
# - with no information on death within 30 days (Variable: dead_30d)

data_pms_01 <- data %>%
  filter(stroke_type %in% c("ischemic", " hemorrhagic"),
         !transf == "Yes" | is.na(transf),
         !is.na(age),
         age >= 18 & age <= 120,
         !is.na(atr_fibr),
         !is.na(nihss_adm),
         nihss_adm >= 0 & nihss_adm <= 42,
         !is.na(dead_30d))




# Definition of study population (Sensitivity analysis 1): Bray et al.'s model A
#
# Exclusion of patients as in the main analysis except
# - Patients with TIA considered as ischemic stroke

data_sens2 <- data
data_sens2$stroke_type[data_sens2$stroke_type == "TIA"] <- "ischemic"

data_pms_01 <- data_sens2 %>%
  filter(stroke_type %in% c("ischemic", " hemorrhagic"),
         !transf == "Yes" | is.na(transf),
         !is.na(age),
         age >= 18 & age <= 120,
         !is.na(atr_fibr),
         !is.na(nihss_adm),
         nihss_adm >= 0 & nihss_adm <= 42,
         !is.na(dead_30d))
```

```
# Calculate predicted probabilities using model equation: Bray et al.'s model A
#
# Predictors:
# - Age in years (categorical): <60, 60-69, 70-79, 80-89, >= 90
# - NIHSS at admission (continuous)
# - Atrial fibrillation (dichotomous)
# - Stroke type (categorical): Ischemic stroke, intracerebral hemorrhage

data_pms_01  <- data_pms_01 %>%
  mutate(age_cat = factor(case_when(age < 60 ~ "<60",
                                    age >= 60 & age < 70 ~ "60-69",
                                    age >= 70 & age < 80 ~ "70-79",
                                    age >= 80 & age < 90 ~ "80-89",
                                    age >= 90 ~ ">= 90"),
                          levels = c("<60", "60-69", "70-79", "80-89", ">= 90"))) %>%
  dummy_cols(select_columns = c("age_cat", "stroke_type", "atr_fibr")) %>%
  mutate(pms_01_pred = -5.250 + `age_cat_60-69` * 0.624 + `age_cat_70-79` * 1.033 +
         `age_cat_80-89` * 1.488 + `age_cat_>= 90` * 1.781 + nihss_adm * 0.137 +
         atr_fibr_1 * 0.425 + stroke_type_hemorrhagic * 0.870) %>%
  mutate(pms_01_prob = 1/(1+exp(-pms_01_pred)))



# Description of study population: Bray et al.'s model A

# Print Table 1 - Characteristics of included stroke patients
tbl_summary(data_pms_01 %>%
              select(age, sex, stroke_type, atr_fibr, nihss_adm, dead_30d) %>%
              mutate(dead_30d = as.factor(dead_30d)) %>%
              mutate(dead_30d = recode_factor(dead_30d, `0` = "Survived",
                                              `1` = "Died")),
            digits = list(all_categorical() ~ c(0, 1)),
            label = list(age ~ "Age, median (IQR)",
                         sex ~ "Sex",
                         stroke_type ~ "Stroke type",
                         atr_fibr ~ "Atrial fibrillation",
                         nihss_adm ~ "NIHSS at admission, median (IQR)"),
            by = dead_30d) %>%
  add_overall()



# Calibration: Bray et al.'s model A
# Calibration-in-thelarge

# Predicted 30-day mortality risk
mean(data_pms_01$pms_01_prob)

# Observed 30-day mortality risk
mean(data_pms_01$dead_30d)

# Predicted absolute number of death within 30-days
round(mean(data_pms_01$pms_01_prob) * nrow(data_pms_01), digits = 0)

# Predicted absolute number of death within 30-days
mean(data_pms_01$dead_30d) * nrow(data_pms_01)

# Observed to expected ratio
mean(data_pms_01$dead_30d) / mean(data_pms_01$pms_01_prob)


# Calibration plot

# Grouping of stroke patients into deciles of observed mortality
data_pms_01$pms_01_prob_group = ntile(data_pms_01$pms_01_prob, 10)

# Table of predicted and observed 30-day mortality by decile
# Including calculation 95%-confidence intervals for observed mortality
```

```
pp <- data_pms_01 %>%
  group_by(pms_01_prob_group) %>%
  summarise(rel_pred = mean(pms_01_prob), rel_obs = mean(dead_30d),
            abs_obs = sum(dead_30d), n = n()) %>%
  mutate(ci = binconf(abs_obs, n, alpha=0.05, method = c("exact"))) %>%
  mutate(ci_l = ci[,"Lower"], ci_u = ci[,"Upper"]) %>%
  select(-ci)


# Visualization of calibration plot
ggplot(data = pp, aes(x = rel_pred, y = rel_obs)) +
  geom_point() +
  theme_bw() +
  geom_abline(intercept = 0, slope = 1, col = "black") +
  geom_errorbar(aes(min = ci_l, max = ci_u)) +
  xlab("Predicted risk") +
  ylab("Observed risk") +
  scale_x_continuous(expand = c(0, 0), limits = c(0,0.4)) +
  scale_y_continuous(expand = c(0, 0), limits = c(0,0.4)) +
  theme(axis.text = element_text(size = 10),
        axis.title = element_text(size = 14),
        plot.margin = margin(4, 4, 4, 4))


# Calibration intercept and slope
# Fit logistic regression model
fit <- glm(dead_30d ~ pms_01_pred, data = data_pms_01, family = binomial)

# Print calibration intercept and slope
fit$coefficients
confint(fit)


# Discrimination: Bray et al.'s model A

# AUC and ROC-Curve
roc <- roc(data_pms_01$dead_30d, data_pms_01$pms_01_prob, ci = TRUE,
           ci.alpha = 0.95)

# Print c-statistic and 95%-confidence interval
auc(roc)
ci.auc(roc)

# Visualize ROC-curve
plot(roc, grid=TRUE, print.auc=FALSE, xlim=c(1,0), ylim=c(0,1),
     xaxs = "i", yaxs = "i", asp=NA, mar=c(2.5, 2.5, .5, .5)+.1,
     mgp=c(1.5, 0.5, 0), grid.lty=1, grid.lwd=.5, grid.col="#EEEEEE")



# External validation of Smith et al.'s model including NIHSS for in-hospital
mortality after stroke

# Definition of study population (Main analysis): Smith et al.'s model
#
# Exclusion of patients…
# - from hospitals not recoding the relevant predictors routinely
# - with no ischemic stroke (Variable: stroke_type)
# - transferred to or form another hospital (Variable: transfer)
# - with missing information on sex
# - with missing information on age or age <18 or >120 (Variable: age)
# - with missing information on atrial fibrillation (Variable: atr_fibr)
# - with missing information on previous stroke or TIA (Variable: prior_event)
# - with missing information on coronary artery disease (Variable: chd)
# - with missing information on diabetes mellitus (Variable: dm)
# - with missing information on hyperlipidemia (Variable: hlp)
# - with missing information on NIHSS at admission (Variable: nihss_adm)
# - with implausible NIHSS (<0 or >42)
# - with no information on in-hospital death (Variable: dead_inhospital)
```

```
data_pms_02 <- data %>%
  filter(stroke_type %in% c("ischemic"),
         !transf == "Yes" | is.na(transf),
         !is.na(sex),
         !is.na(age),
         age >= 18 & age <= 120,
         !is.na(atr_fibr),
         !is.na(prior_event),
         !is.na(chd),
         !is.na(dm),
         !is.na(hlp),
         !is.na(nihss_adm),
         nihss_adm >= 0 & nihss_adm <= 42,
         !is.na(dead_inhospital))


# Definition of study population (Sensitivity analysis 1): Smith et al.'s model
#
# Exclusion of patients as in the main analysis except
# - information on mode of arrival is missing or unknown (variable transp)

data_pms_02 <- data %>%
  filter(stroke_type %in% c("ischemic"),
         !transf == "Yes" | is.na(transf),
         !is.na(sex),
         !is.na(age),
         age >= 18 & age <= 120,
         !is.na(atr_fibr),
         !is.na(prior_event),
         !is.na(chd),
         !is.na(dm),
         !is.na(hlp),
         !is.na(nihss_adm),
         nihss_adm >= 0 & nihss_adm <= 42,
         !is.na(dead_inhospital),
         !is.na(transp) | transp == "unknown")


# Definition of study population (Sensitivity analysis 2): Smith et al.'s model
#
# Exclusion of patients as in the main analysis except
# - Patients with TIA considered as ischemic stroke

data_sens2 <- data
data_sens2$stroke_type[data_sens2$stroke_type == "TIA"] <- "ischemic"

data_pms_02 <- data_sens2 %>%
  filter(stroke_type %in% c("ischemic"),
         !transf == "Yes" | is.na(transf),
         !is.na(sex),
         !is.na(age),
         age >= 18 & age <= 120,
         !is.na(atr_fibr),
         !is.na(prior_event),
         !is.na(chd),
         !is.na(dm),
         !is.na(hlp),
         !is.na(nihss_adm),
         nihss_adm >= 0 & nihss_adm <= 42,
         !is.na(dead_inhospital))
```

```
# Calculate Probabilities for PMS using model equation: Smith et al.'s model
# Model 2 (including NIHSS)
#
# Predictors:
# - Age (continuous): per 1 year over 60
# - Sex (categorical): male, non-male
# - NIHSS at admission (continuous)
# - Mode of arrival (categorical): By private transport, by Ambulance, Not via ED
# - Atrial fibrillation (dichotomous)
# - Previous stroke or TIA (dichotomous)
# - Coronary artery disease (dichotomous)
# - Diabetes mellitus (dichotomous)
# - History of dyslipidemia (dichotomous)
#
# Imputation of mode of arrival missing or unknown to "private" (Variable: transp)

data_pms_02 <- data_pms_02 %>%
  mutate(age_60 = case_when(age <= 60 ~ 0, age >= 60 ~ age-60),
         transp_pms = factor(case_when(transp %in% c("private", "unknown") |
                          is.na(transp) ~ "private",
                          transp %in% c("clinical acute event",
                          "secondary relocation") ~ "not via ED",
                          transp == "ambulance service" ~ "ambulance"),
                             levels = c("private", "not via ED", "ambulance"))) %>%
  dummy_cols(select_columns = c("sex", "transp_pms", "atr_fibr",
                               "prior_event", "chd", "dm", "hlp")) %>%
  mutate(pms_02_pred = -5.3169 + age_60 * 0.0176 + sex_male * 0.167 +
         nihss_adm * 0.116 + `transp_pms_not via ED` * 0.9611 +
         transp_pms_ambulance * 0.7654 + atr_fibr_1 * 0.300 +
         prior_event_1 * -0.112 + chd_1 * 0.268 + dm_1 * 0.124 + hlp_1 * -0.132)%>%
  mutate(pms_02_prob = 1/(1+exp(-pms_02_pred)))



# Description of study population: Smith et al.'s model

# Print Table 1 - Characteristics of included stroke patients
tbl_summary(data_pms_02 %>%
                        select(age, sex, nihss_adm, transp_pms, atr_fibr,
                               prior_event, chd, dm, hlp, dead_inhospital) %>%
                        mutate(dead_inhospital = as.factor(dead_inhospital)) %>%
                        mutate(dead_inhospital = recode_factor(dead_inhospital,
                                            `0` = "Survived", `1` = "Died")),
                      by = dead_inhospital,
                      digits = list(all_categorical() ~ c(0, 1)),
                      label = list(age ~ "Age, median (IQR)",
                                   sex ~ "Sex",
                                   nihss_adm ~ "NIHSS at admission, median (IQR)",
                                   transp_pms ~ "Mode of arrival",
                                   atr_fibr ~ "Atrial fibrillation",
                                   prior_event ~ "Previous stroke or TIA",
                                   chd ~ "Coronary artery disease",
                                   dm ~ "Diabetes mellitus",
                                   hlp ~ "Dyslipidemia")) %>%
  add_overall()



# Calibration: Smith et al.'s model

# Calibration-in-the-large

# Predicted in-hospital mortality risk
mean(data_pms_02$pms_02_prob)

# Observed in-hospital mortality risk
mean(data_pms_02$dead_inhospital)
```

```
# Predicted absolute number of in-hospital death
round(mean(data_pms_02$pms_02_prob) * nrow(data_pms_02), digits = 0)

# Observed absolute number of in-hospital death
mean(data_pms_02$dead_inhospital) * nrow(data_pms_02)

# Observed to expected ratio
mean(data_pms_02$dead_inhospital) / mean(data_pms_02$pms_02_prob)


# Calibration plot

# Grouping of stroke patients into deciles of observed mortality
data_pms_02$pms_02_prob_group = ntile(data_pms_02$pms_02_prob, 10)

# Table of predicted and observed in-hospital mortality by decile
# Including calculation 95%-confidence intervals for observed mortality
pp <- data_pms_02 %>%
  group_by(pms_02_prob_group) %>%
  summarise(rel_pred = mean(pms_02_prob), rel_obs = mean(dead_inhospital),
            abs_obs = sum(dead_inhospital), n = n()) %>%
  mutate(ci = binconf(abs_obs, n, alpha=0.05,method=c("exact"))) %>%
  mutate(ci_l = ci[,"Lower"], ci_u = ci[,"Upper"]) %>%
  select(-ci)


# Visualization of calibration plot
ggplot(data = pp, aes(x = rel_pred, y = rel_obs)) +
  geom_point() +
  theme_bw() +
  geom_abline(intercept = 0, slope = 1, col = "black") +
  geom_errorbar(aes(min = ci_l, max = ci_u)) +
  xlab("Predicted risk") +
  ylab("Observed risk") +
  scale_x_continuous(expand = c(0, 0), limits = c(0,0.4)) +
  scale_y_continuous(expand = c(0, 0), limits = c(0,0.4)) +
  theme(axis.text = element_text(size = 10),
        axis.title = element_text(size = 14),
        plot.margin = margin(4, 4, 4, 4))


# Calibration intercept and slope
# Fit logistic regression model
fit <- glm(dead_inhospital ~ pms_02_pred, data = data_pms_02, family = binomial)

# Print calibration intercept and slope
fit$coefficients
confint(fit)


# Discrimination: Smith et al.'s model

# AUC and ROC-Curve
roc <- roc(data_pms_02$dead_inhospital, data_pms_02$pms_02_prob, ci = TRUE,
           ci.alpha = 0.95)

# Print c-statistic and 95%-confidence interval
auc(roc)
ci.auc(roc)

# Visualize ROC-curve
plot(roc, grid=TRUE, print.auc=FALSE, xlim=c(1,0), ylim=c(0,1),
     xaxs = "i", yaxs = "i", asp=NA, mar=c(2.5, 2.5, .5, .5)+.1,
     mgp=c(1.5, 0.5, 0), grid.lty=1, grid.lwd=.5, grid.col="#EEEEEE")
```

BMJ Open

External validation of risk prediction models for post-stroke mortality in Berlin

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-089320.R1
Article Type:	Original research
Date Submitted by the Author:	28-Mar-2025
Complete List of Authors:	Reitzle, Lukas; Robert Koch Institute, Department of Epidemiology and Health Monitoring; Charité - Universitätsmedizin Berlin, Center for Stroke Research Berlin Rohmann, Jessica; Charité - Universitätsmedizin Berlin, Center for Stroke Research Berlin; Charité - Universitätsmedizin Berlin, Institute of Public Health Kurth, Tobias; Charité - Universitätsmedizin Berlin, Institute of Public Health Audebert, Heinrich; Charité - Universitätsmedizin Berlin, Center for Stroke Research Berlin; Charité - Universitätsmedizin Berlin, Klinik und Hochschulambulanz für Neurologie, Campus Benjamin Franklin Piccininni, Marco; Hasso Plattner Institute for Digital Engineering, Digital Health - Machine Learning Research Group; Charité - Universitätsmedizin Berlin, Institute of Public Health; Charité - Universitätsmedizin Berlin, Center for Stroke Research Berlin; University of Potsdam, Digital Engineering Faculty
Primary Subject Heading:	Neurology
Secondary Subject Heading:	Cardiovascular medicine
Keywords:	Stroke < NEUROLOGY, Prognosis, Mortality, Clinical Decision-Making

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

External validation of risk prediction models for post-stroke mortality in Berlin

Authors: Lukas Reitzle, MD [1,2], Jessica L. Rohmann, PhD [1, 3], Tobias Kurth, MD, ScD [3], Heinrich J. Audebert, MD [1,4], Marco Piccininni, PhD [1,3,5,6]

[1] Center for Stroke Research Berlin, Charité – Universitätsmedizin Berlin, Berlin, Germany
[2] Department of Epidemiology and Health Monitoring, Robert Koch Institute, Berlin, Germany
[3] Institute of Public Health, Charité – Universitätsmedizin Berlin, Berlin, Germany
[4] Klinik und Hochschulambulanz für Neurologie, Campus Benjamin Franklin, Charité – Universitätsmedizin Berlin, Berlin, Germany
[5] Digital Health - Machine Learning Research Group, Hasso Plattner Institute for Digital Engineering, Potsdam, Germany
[6] Digital Engineering Faculty, University of Potsdam, Potsdam, Germany

Corresponding author:
Lukas Reitzle, MD
Robert Koch Institute
Department of Epidemiology and Health Monitoring
Nordufer 20
13353 Berlin
+49 30 18754 3738
reitzlel@rki.de

Email Addresses of all authors:
Jessica L. Rohmann: jessica.rohmann@charite.de
Tobias Kurth: tobias.kurth@charite.de
Heinrich J. Audebert: heinrich.audebert@charite.de
Marco Piccininni: marco.piccininni@hpi.uni-potsdam.de

Keywords: Stroke, prognosis, mortality, clinical decision-making

Word count: 3,765

Abstract

Objectives: Prediction models for post-stroke mortality can support medical decision-making. Although numerous models have been developed, external validation studies determining the models' transportability beyond the original settings are lacking. We aimed to assess the performance of two prediction models for post-stroke mortality in Berlin, Germany.

Design: We used data from the Berlin-SPecific Acute Treatment in Ischemic or hAemorrhagic stroke with Long term follow–up (B–SPATIAL) registry.

Setting: Multicenter stroke registry in Berlin, Germany

Participants: Adult patients, admitted within 6 hours after symptom onset and with an ICD-10 discharge diagnosis of ischemic stroke, hemorrhagic stroke or transient ischemic attack at one of 15 hospitals with stroke units between January 1st, 2016 and January 31st, 2021.

Primary Outcome Measures: We evaluated calibration (calibration-in-the-large, intercept, slope and plot) and discrimination performance (c-statistic) of *Bray et al*.'s 30-day mortality and *Smith et al*.'s in-hospital mortality prediction models. Information on mortality was supplemented by Berlin city registration office records.

Results: For the validation of *Bray et al*.'s model, we included 7,879 patients (mean age 75; 55.0% men). We observed 763 (9.7%) deaths within 30 days of stroke compared to 680 (8.6%) predicted. The model's c-statistic was 0.865 (95%CI: 0.851-0.879). For *Smith et al*.'s model, we performed the validation among 1,931 patients (mean age 75; 56.2% men), observing 105 (5.4%) in-hospital deaths compared to the 92 (4.8%) predicted. The c-statistic was 0.891 (95%CI: 0.864-0.918). The calibration plots of both models revealed an underestimation of the mortality risk for high-risk patients.

Conclusions: Among Berlin stroke patients, both models showed good calibration performance for low and medium risk patients and high discrimination, while underestimating risk among high-risk patients. The acceptable performance of *Bray et al*.'s model in Berlin illustrates how a small number of routinely collected variables can be sufficient for valid prediction of post-stroke mortality.

Strengths and limitations of this study

- The prospective, multicenter B-SPATIAL registry included all stroke/TIA patients presenting to 15 Berlin hospitals with stroke units during a 5-year period.
- Loss-to-follow-up was low, data completeness was high, and outcome information was reliable.
- Since Berlin is a densely populated city with several stroke units, our findings may be transportable to urban Central European areas but not necessarily to other settings.
- The prediction model for in-hospital mortality after stroke by Smith et al. could only be validated with data from 3 of the 15 hospitals, routinely collecting data on all relevant predictors.

Introduction

In 2019, stroke was the second leading cause of death and the third leading cause of combined death and disability worldwide.[1] In the context of stroke aftercare, prediction models have been developed to predict functional outcomes and mortality risk after acute stroke. These tools support (shared) clinical decision-making by providing information about likely prognosis to health professionals, patients, and their families.[2, 3] Yet, before implementing prediction models in routine clinical practice, the transportability from the original development population to the population of interest should be assessed; models with low performance in the setting of interest may generate non-accurate predictions and lead to sub-optimal decisions.[4]

According to the systematic review by *Fahey et al.,* prior to September 2015, 38 prediction models for post-stroke mortality had been developed.[2] Despite the abundance of existing prediction models for post-stroke outcomes, only a small fraction have been externally validated.[2] Among the most frequently used predictors were demographic characteristics (e.g., age and sex), stroke severity as measured by the National Institutes of Health Stroke Scale (NIHSS), stroke type, and comorbidities.[2] The variable NIHSS, alone, has shown high predictive performance for early mortality after acute stroke[5] and is often used in prediction models for post-stroke mortality.[6, 7]

Two prediction models for post-stroke mortality including the NIHSS and other routinely collected variables were developed by *Bray et al*. (2014) using data from the Sentinel Stroke National Audit Program (SSNAP) in the United Kingdom[3] and by *Smith et al.* (2010) using data from the Get With the Guidelines (GWTG) Stroke Program in the United States.[8] Though these models have already been subjected to validation studies in their respective originating countries[9], to date, both models have only undergone external validation in the China National Stroke Registry.[10-12] Our aim was to conduct an external validation[4] study to assess calibration and discrimination performances of *Bray et al.* (2014)[3] and *Smith et al.* (2010)[8] prediction models for post-stroke mortality among Berlin stroke patients.

Methods

Data source

We used data from the Berlin – SPecific Acute Treatment in Ischemic or hAemorrhagic stroke with Long term follow–up (B–SPATIAL) registry (Clinicaltrials.gov identifier: NCT03027453), a multicenter registry for adult stroke patients in Berlin. Data were collected from patients aged 18 years or older, admitted within 6 hours after symptom onset and with discharge diagnoses according to the 10th revision of the International Classification of Diseases (ICD-10) of ischemic stroke (I63/I64), hemorrhagic stroke (I61), non-traumatic subdural hemorrhage (I62) or transient ischemic attack (TIA; G45.0–G45.3 and G45.5–G45.9) at one of the 15 hospitals with stroke units in Berlin, Germany, between January 1st, 2016 and January 31st, 2021. Patients with no symptoms upon the arrival of emergency medical services and without neurological symptoms at hospital arrival were not included in the registry. In this external validation study, we did not include patients for whom a mobile stroke unit was dispatched, as part of the B_PROUD interventional study,[13] which was linked to the registry. We further excluded patients who opted out of data collection.[14]

Prediction models for post-stroke mortality

We evaluated the performance of *Bray et al.'s model A* (2014) including the full NIHSS (all items) for 30-day all-cause mortality[3] (hereafter: *Bray et al.*'s model) and *Smith et al.*'s model (2010) including the NIHSS for in-hospital mortality.[8]

Predictors

Bray et al.'s model included the following predictors: age group (<60, 60-69, 70-79, 80-89, ≥90 years), stroke type (ischemic or hemorrhagic), atrial fibrillation and NIHSS at admission.[3] In the original development study, all variables were directly entered in a secure web portal by clinical teams in accordance with the SSNAP registry.[3] In our external validation using the B-SPATIAL registry data,

stroke type was determined using available ICD-10 codes (I63 or I64 for ischemic, I61 for hemorrhagic). Atrial fibrillation was considered present if the patient had a known history of atrial fibrillation or if atrial fibrillation was diagnosed by the emergency medical service or at admission. *Smith et al*.'s model included the following predictors: age as a continuous variable, sex (male vs. non-male), NIHSS at admission, atrial fibrillation, history of stroke or TIA, coronary artery disease, diabetes mellitus and dyslipidemia.[8] Additionally, the model included a variable indicating the mode of hospital arrival, categorized as arrival by private transport, by ambulance, or other arrival not via the emergency department (ED) (e.g. direct admission from the hospital ward) as a predictor. In the GWTG registry, used in the development study, clinicians used an internet-based tool for data entry.[8] In our external validation, we assumed a prior history of stroke or TIA if indicated by imaging performed while in hospital or if documentation of an ischemic stroke or TIA was available. We defined the presence of coronary artery disease as documented previous myocardial infarction, coronary stent placement, or corresponding diagnostic coronary angiography result. In the B-SPATIAL registry, diabetes was defined as documented history of diabetes, the use of anti-diabetic medication, or a measured A1C level above 6.5% or blood glucose above 200 mg/dl (non-fasting) or 126 mg/dl (fasting). We defined dyslipidemia to include either a reported history of the condition, having measured Low-Density-Lipoprotein (LDL)-cholesterol levels above 130 mg/dl, or having total cholesterol levels above 220 mg/dl. In line with the original development study, in cases of missing documentation or an unknown mode of arrival, we assumed arrival by private transport. For the few cases with documented secondary transfer but no documentation of transfer from an external hospital, we assumed the patient was internally transferred within the same hospital and thus did not arrive via the ED.

Outcomes

Smith et al.'s model predicted in-hospital mortality. We defined in-hospital mortality as death documented as the discharge reason or a modified Rankin Scale (mRS) score of 6 at discharge. In

cases where both documentation of discharge reason and mRS at discharge were missing, we assumed patients were alive at discharge. *Bray et al.*'s model used 30-day all-cause mortality as the outcome. To create the 30-day all-cause mortality variable, we counted those patients with in-hospital death and hospital stays ≤30 days, and patients for whom the date of death was within 30 days of hospital admission. We obtained information about the date of death from the Berlin city registration office at two and four months after stroke.[14]

Study population

For *Bray et al.*'s model, we included patients with either an acute ischemic or hemorrhagic stroke diagnosis. For *Smith et al.*'s model, we included only ischemic stroke patients, similar to the inclusion in the original publication (which retrospectively identified patients using the ICD-9 codes). Since the predictors history of stroke or TIA, coronary artery disease, and dyslipidemia were only routinely recorded in three of the B-SPATIAL registry hospitals, we excluded patients from the remaining hospitals in the validation of *Smith et al.*'s model. For both models, we excluded patients who were transferred from a hospital not participating in the B-SPATIAL registry, and, for the main analysis, we also excluded patients with missing values for one of the predictors. When information about transfer status was missing, we assumed the patient was not transferred.

Analyses of the data from the B-SPATIAL registry, including the external validation of clinical risk scores, were approved by the ethics committee of the Charité - Universitätsmedizin Berlin (EA1/208/21). The B-SPATIAL registry used an opt-out mechanism for patient inclusion. Two months after their index event, patients were informed in writing about the inclusion of their record in the B-SPATIAL registry and had multiple opportunities to opt out.[14]

Statistical analysis

We used the prediction models' published formulas to calculate risk of 30-day all-cause mortality[3] or in-hospital mortality[8] for each included individual (see Supplemental **R Code**).

To assess model calibration, we evaluated the calibration-in-the-large by comparing the actual ("observed") number of deaths to the one predicted by the model ("expected") using the observed-to-expected (O/E) event ratio. We then used a calibration plot to graphically compare the observed mortality risk with the mean predicted risk within decile groups of predicted risk. We estimated 95% confidence intervals for the observed risk using the binomial exact method. Furthermore, we calculated the calibration intercept and slope using the logistic recalibration framework.[15] We assessed the discriminatory ability of the two prediction models by calculating the concordance statistic (c-statistic) and corresponding 95% confidence intervals and visualizing the receiver operating characteristics (ROC) curve.

In addition, we assessed the discriminatory ability of the NIHSS alone for both outcomes and computed the c-statistic for Bray et al.'s model predicting in-hospital mortality and Smith et al.'s model predicting 30-day mortality. In a subgroup analysis, we evaluated the models' performances in terms of calibration and discrimination separately by sex.

For both models, we conducted multiple sensitivity analyses. For *Smith et al*.'s model, to assess the robustness of our assumption of arrival by private transport when the mode of arrival was missing or unknown, we reran our analysis excluding patients with unknown or missing mode of arrival. In the original prediction model development studies, *Smith et al*. explicitly excluded patients with TIA, and *Bray et al*. did not specify how these patients were handled in the data management stage.[3, 8] However, at the time of admission, TIA patients presenting with neurological symptoms compatible with stroke are not distinguishable from ischemic stroke patients. Therefore, in an additional sensitivity analysis, we investigated the performance of both models when classifying all patients with final diagnosis of TIA as ischemic stroke patients. Finally, we assessed calibration and discrimination of both models after imputing the predictors' missing values by Multiple Imputation by Chained Equations. Specifically, for each model's validation, we imputed 5 datasets using only the model-specific predictors and outcome in the imputation.

All statistical analyses were performed using R v4.2.1 and RStudio 2022.07.1. The pROC package was used for the calculation of the c-statistic and ROC curve and the c-statistic's confidence intervals were derived using the package's ci.auc function with default settings. The mice package and the miceafter package were used for the imputation and pooling of results.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this study.

Results

***Bray et al.*'s model for 30-day all-cause mortality**

We included 7,879 stroke patients in the external validation of *Bray et al.*'s model (**Figure 1**). The median age of the B-SPATIAL patients included in this validation was 75 years and 55.0% were male (**Table 1**). A final diagnosis of ischemic stroke was considerably more common (92.4%) than hemorrhagic stroke. Median NIHSS at admission was 5 (interquartile range [IQR]:2-11). In total, 763 (9.7%) of included patients died within 30 days of admission. We found that *Bray et al.*'s model underestimated the mortality risk, predicting an average mortality of 8.6%, corresponding to 680 deaths. The resulting observed-to-expected ratio was 1.12, indicating that 12% more deaths were observed in our sample than were predicted by the model. The calibration plot showed good alignment of predicted and observed mortality for eight decile groups (**Table S1**, **Figure 2**). The model underestimated the observed mortality risk for the two decile groups with the highest predicted mortality. The calibration intercept was 0.34 (95%CI: 0.19-0.49) and the slope was 1.10 (95%CI: 1.03-1.17). **Figure 2** shows the ROC curve illustrating the discriminatory ability of the model in our sample. The c-statistic for 30-day mortality, the model's intended outcome, was 0.865 (95%CI: 0.851-0.879). For comparison, the NIHSS alone showed a c-statistic of 0.838 (95%CI: 0.823-0.853). When instead using *Bray et al.*'s model to predict in-hospital mortality in this validation dataset, we obtained a c-

statistic of 0.873 (95%CI: 0.858-0.888). In the subgroup analysis by sex, *Bray et al.*'s model showed similar performance for male and non-male patients with regard to calibration (**Figure S1**) and discrimination (c-statistic of 0.858 [95%CI: 0.836-0.880] for males and 0.865 [95%CI: 0.847-0.883] for non-males; **Figure S2**).

In the sensitivity analysis for *Bray et al*.'s model, we included an additional 1,932 patients who were ultimately diagnosed with TIA into the validation sample. Among the 9,811 ischemic stroke/TIA patients, the observed 30-day mortality was 7.9%, which was very similar to the 7.4% predicted by *Bray et al*.'s model. The calibration plot, calibration intercept (0.37 [95%CI: 0.22-0.52]) and slope (1.15 [95%CI: 1.08-1.21]) from this sensitivity analysis were similar to the ones of the main analysis, and the c-statistic was 0.880 (95%CI: 0.867-0.893; **Figure S3**).

In a second sensitivity analysis, in which we used multiple imputation, a total of 8,366 stroke patients were included, of whom 951 (11.4%) died within 30 days. The observed mortality was higher compared to the main analysis, but the conclusions did not fundamentally change. The model underestimated 30-day mortality in the highest risk individuals (to a slightly greater extent than as was assessed in the main analysis; **Figure S4**). The model's calibration intercept was 0.52 [95%CI: 0.37-0.67]), and the calibration slope was 1.12 [95%CI: 1.04-1.19]). The c-statistic obtained after multiple imputation was 0.870 [95%CI: 0.855-0.884], similar to the main analysis.

Smith et al.'s model for in-hospital mortality

For the external validation of *Smith et al*.'s prediction model, we included 1,931 ischemic stroke patients (**Figure 1**). The median age in this sample was 75, and 56.2% patients were male (**Table 2**). The median NIHSS was 4 (IQR: 2-10) and most patients arrived by ambulance (81.1%). In total, 105 (5.4%) ischemic stroke patients died during the hospital stay.

Smith et al.'s model predicted an average risk of in-hospital mortality of 4.8%, corresponding to 92 deaths. The observed-to-expected ratio of 1.14 indicated an underestimation of in-hospital mortality by 14%. The calibration plot revealed that the model underestimated the mortality risk in the decile groups with the highest predicted mortality. For decile groups with low and medium risk, the

observed and predicted risk were well-aligned, albeit with high uncertainty, as only few patients were observed (**Table S2**, **Figure 3**). The corresponding calibration intercept was 1.20 (95%CI: 0.66-1.76), and the slope was 1.43 (95%CI: 1.22-1.66). We depicted the discriminatory ability of the model predicting in-hospital mortality as a ROC curve (**Figure 3**). The corresponding c-statistic was 0.891 (95%CI: 0.864-0.918). For comparison, the c-statistic for in-hospital mortality of NIHSS alone was 0.868 (95%CI: 0.833-0.903). When instead using *Smith et al.*'s model to predict 30-day mortality in this validation dataset, the c-statistic was 0.873 (95%CI: 0.847-0.899). Compared to non-male patients, the calibration of *Smith et al.*'s model seemed slightly better among male patients, as the underestimation of the predicted risk was lower in the highest risk decile groups (**Figure S5**). Discrimination ability seemed higher for male patients with a c-statistic of 0.914 (95%CI: 0.881-0.946) compared to non-male patients (c-statistic: 0.867 [95%CI: 0.825-0.908]) (**Figure S6**).

In the sensitivity analysis excluding patients with unknown or missing mode of arrival for *Smith et al.*'s model, the observed in-hospital mortality was 5.6% compared to 5.0% predicted by the model. The calibration plot (**Figure S7**), calibration intercept (1.11 [95%CI: 0.56-1.67]) and slope (1.41 [95%CI: 1.19-1.64]), as well as the c-statistic of 0.883 (95%CI: 0.854-0.912), were comparable to those estimated in the main analysis.

In the second sensitivity analysis for *Smith et al.*'s model, we additionally included 597 TIA patients. Among the 2,528 included ischemic stroke/TIA patients, 4.3% died in the hospital compared to a mortality of 4.0% predicted by the model. The calibration plot (**Figure S8**), calibration intercept (1.24 [95%CI: 0.71-1.77]) and slope (1.47 [95%CI: 1.27-1.68]), as well as the c-statistic (0.902 [95%CI: 0.875-0.929]) obtained in this second sensitivity analysis only slightly deviated from the main analysis results.

In a further sensitivity analysis, in which we used multiple imputation, a total of 2,052 ischemic stroke patients were included, of whom 117 (5.7%) died in-hospital. After imputation, the model's

calibration intercept (1.26 [95%CI: 0.70-1. 81]), slope (1.44 [95%CI: 1.22-1.66]), and calibration plot (**Figure S9**) as well as c-statistic (0.893 [95%CI: 0.864-0.917]) were similar to the main analysis.

Discussion

In this study, we externally validated two prognostic prediction models for mortality after stroke; *Bray et al*.'s model for 30-day all-cause mortality and *Smith et al*.'s model for in-hospital mortality, using data from a multicenter registry of adult stroke patients presenting to 15 stroke units in Berlin, Germany.

Bray et al.'s prediction model was originally developed in the United Kingdom in 2014 using data from the SSNAP, which is the national registry for acute stroke in England and Wales.[3] The original publication included an external validation study using data from the South London Stroke Register, which showed good calibration performance and high discrimination of the model, with a c-statistic of 0.87.[3] The model was later externally validated in two other studies. The first was a temporal validation study using SSNAP data from a different time period, which found a slightly worse discrimination ability (c-statistic of 0.774).[9] The second was conducted in the China National Stroke Registry; despite substantial differences in the study population's composition, the model showed a good discrimination ability (c-statistic of 0.80) and good calibration in this setting.[11] Our findings add to this evidence body, providing an external validation study from Germany. In the Berlin setting, we found good alignment between the predicted and observed 30-day mortality in low and medium risk individuals; however, the *Bray et al.* model underestimated risk among high-risk patients. Compared to other external validations, we observed a higher discriminatory performance of the model in our setting (c-statistic of 0.865). Overall, our conclusions did not differ after multiple imputation or when stratifying by sex.

Smith et al.'s prediction model was originally developed using data from the GWTG Stroke Program in the United States.[8] Thereafter, two external validation studies using different cohorts from the China National Stroke Registry found good calibration and high discrimination ability of this model

with c-statistics of 0.867[12] and 0.86[10], similar to the results of the internal validation in the development study (c-statistic of 0.85[8]). In the Berlin setting, we observed higher discrimination performance (c-statistic of 0.891); however, the *Smith et al.* model underestimated the risk in high-risk individuals, showing non-optimal calibration in our setting. Higher uncertainty was present for this validation, due to the low sample size. The findings of the external validation did not differ substantially after multiple imputation. We observed slightly better model performance for male stroke patients.

The calibration of both models showed an underestimation of the mortality risk for high-risk patients in Berlin, which may have been due to different factors. In our analysis, we excluded patients who opted out of study participation. Opting-out was only possible for patients who survived the stroke, which might have introduced a selection of higher severity stroke in our sample.

For the validation of both models, we excluded patients with final diagnosis of TIA in our main analysis, since in the original publications and validation studies, TIA patients were either explicitly excluded[8, 9] or their inclusion was not specified.[3, 10-12] Within the Berlin setting, we found that both models performed similarly or even better after the inclusion of TIA patients. Different definitions of TIA exist, and the diagnostic discrepancy may explain why researchers might hesitate to include patients with TIA in studies for the development of prediction models. However, since patients with a TIA present with symptoms comparable to an ischemic stroke upon admission, from a clinical and methodological perspective, we believe future work should consider including TIA patients also in the development and validation of post-stroke prediction models.

A systematic review of prediction models for post-stroke outcomes found that models with a high number of predictors do not necessarily show a better performance.[2] Our results underscore the high predictive ability of NIHSS, which as a single predictor attained a c-statistic of more than 0.83 for both mortality outcomes, comparable to previous studies.[7, 16] Even models with few variables such as *Bray et al.*'s model, including only four predictors, showed high discrimination.[3, 11] Models that perform sufficiently well with fewer, routinely-measured variables should be preferred over models

with several predictors, since they are more likely to be used in practice. For this reason, as has also been argued for other clinical applications,[17, 18] we believe that future prediction model studies in the context of post-stroke outcomes should compare newly developed models' performances with that of well-established models, or preferably focus on the external validation or updating of existing models rather than developing new ones.

Strengths and limitations

The strengths of this study include the prospective, multicenter design of the B-SPATIAL registry with coverage of all 15 Berlin stroke units over a 5-year period. Therefore, the B-SPATIAL registry for adult stroke patients can be considered representative for the population of stroke patients in Berlin and comprises detailed information on demographics and clinical characteristics, with low loss to follow-up, especially for mortality endpoints.[14] The recording of vital status during follow-up is considered particularly reliable because the information was supplemented by city registration office records. Some limitations should be considered when interpreting our results. As Berlin is a densely populated city with several stroke units, the availability of stroke care might be different compared to other regional settings in Germany and Central Europe. Therefore, our results may not generalize to different settings, such as rural areas. Furthermore, the B-SPATIAL registry only contains information on patients with hospital arrival within 6 hours of symptom onset, since this was the eligibility window for reperfusion treatments when the registry commenced. However, we acknowledge that a substantial proportion of stroke patients present to hospitals later than 6 hours after onset,[19, 20] and the performance of these prediction models might differ for these patients.

Only three of 15 registry hospitals routinely documented history of hyperlipidemia, coronary artery disease, and history of stroke or TIA. Therefore, we could only validate *Smith et al*.'s model in this subsample, which composed 30% of the full validation sample. Furthermore, as overall in-hospital mortality risk was low in our setting, only 105 in-hospital deaths were observed in this subsample, which decreased the power of the analysis and somewhat limits the interpretation of the calibration

plot due to higher uncertainty. For both models, in the main analysis, we excluded patients with missing information on at least one of the predictors (except mode of arrival). However, the sensitivity analysis in which predictors' missing values were imputed showed similar behavior in terms of calibration and discrimination compared to the main analysis for both models.

Conclusion

Despite being developed outside of Germany, the external validation of *Smith et al*.'s model and *Bray et al*.'s model for post-stroke mortality both demonstrated good calibration for low and medium risk stroke patients in a large stroke registry in Berlin. Both models showed high discrimination ability, but underestimated risk in high-risk patients. The performance of *Bray et al*.'s model indicated an overall acceptable transportability to the Berlin setting and illustrates how a small number of variables that can be routinely obtained at hospital admission can suffice for valid prediction of post-stroke mortality.

Acknowledgment section

Contributions

MP and JR conceived the study. LR, TK, JR and MP designed the study. LR selected the models for external validation. JLR and HJA provided B-SPATIAL registry support, for which HJA obtained funding. The statistical analyses were planned and conducted by LR in consultation with MP and JR. LR created the R code, tables, and figures, which were revised by MP. LR wrote the first draft of the manuscript with input from JR and MP. MP provided project supervision. All authors reviewed and edited the manuscript for intellectual content and approved the final version of the manuscript. LR had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. LR is the guarantor.

Acknowledgements

This work builds directly on the Master's thesis project of Dr. Reitzle (supervised by Dr. Piccininni and Dr. Rohmann) and was conducted in accordance with the registered research proposal, which was approved by the Master of Science in Epidemiology Program at the Berlin School of Public Health, Charité – Universitätsmedizin Berlin. We are grateful to all collaborating hospitals and the study nurses for their engagement and thank Jakob Beilstein for assistance with data management.

Declaration of Conflicting Interests

Dr. Reitzle reports having received research grants from the Federal Ministry of Health in Germany. Dr. Rohmann reports having received a grant from Novartis Pharma for conducting a self-initiated research project about migraine.

Outside the content of this paper, Dr. Kurth declares having received research grants from the German Federal Joint Committee (G-BA) and personal compensation from the North-East German Society for Gynecological Oncology (NOGGO), AbbVie, Eli Lilly & Company, Novartis, the BMJ, and Frontiers.

Dr. Audebert reports personal fees from Bayer Vital, Boehringer Ingelheim, Bristol- Myers Squibb, Novo Nordisk, Pfizer, and from Sanofi.

Dr. Piccininni reports having received partial funding from Novartis Pharma and being awarded a research grant from the Center for Stroke Research Berlin (private donations).

Funding/Support

The B-SPATIAL registry was funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung; BMBF) and the German Research Foundation (Deutsche Forschungsgemeinschaft; DFG), granted to the Center for Stroke Research Berlin and by the DFG to the Excellence Cluster NEUROCURE.

Role of the Funder/Sponsor

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; nor decision to submit the manuscript for publication.

Data sharing statement

B-SPATIAL registry data can be made available in a de-identified manner to researchers who provide a methodologically sound proposal (to the extent allowed by the registry's data protection agreement). Data access requests should be directed to jessica.rohmann (at) charite.de.

References

1. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol* 2021; 20: 795-820. DOI: 10.1016/s1474-4422(21)00252-0.

2. Fahey M, Crayton E, Wolfe C, et al. Clinical prediction models for mortality and functional outcome following ischemic stroke: A systematic review and meta-analysis. *PLoS One* 2018; 13: e0185402. DOI: 10.1371/journal.pone.0185402.

3. Bray BD, Campbell J, Cloud GC, et al. Derivation and external validation of a case mix model for the standardized reporting of 30-day stroke mortality rates. *Stroke* 2014; 45: 3374-3380. DOI: 10.1161/strokeaha.114.006451.

4. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012; 98: 691-698. 20120307. DOI: 10.1136/heartjnl-2011-301247.

5. Rost NS, Bottle A, Lee JM, et al. Stroke Severity Is a Crucial Predictor of Outcome: An International Prospective Validation Study. *J Am Heart Assoc* 2016; 5. DOI: 10.1161/jaha.115.002433.

6. Weimar C, König IR, Kraywinkel K, et al. Age and National Institutes of Health Stroke Scale Score within 6 hours after onset are accurate predictors of outcome after cerebral ischemia: development and external validation of prognostic models. *Stroke* 2004; 35: 158-162. DOI: 10.1161/01.Str.0000106761.94985.8b.

7. Kumar A, Roy I, Bosch PR, et al. Medicare Claim-Based National Institutes of Health Stroke Scale to Predict 30-Day Mortality and Hospital Readmission. *J Gen Intern Med* 2022; 37: 2719-2726. 20211026. DOI: 10.1007/s11606-021-07162-0.

8. Smith EE, Shobha N, Dai D, et al. 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.

9. Dutta D, Cannon A and Bowen E. Validation and comparison of two stroke prognostic models for in hospital, 30-day and 90-day mortality. *European Stroke Journal* 2017; 2: 327-334. DOI: 10.1177/2396987317703581.

10. Sun S, Pan Y, Bai L, et al. GWTG Risk Model for All Stroke Types Predicts In-Hospital and 3-Month Mortality in Chinese Patients with Acute Stroke. *J Stroke Cerebrovasc Dis* 2019; 28: 800-806. DOI: 10.1016/j.jstrokecerebrovasdis.2018.11.024.

11. Yu P, Pan Y, Wang Y, et al. External Validation of a Case-Mix Adjustment Model for the Standardized Reporting of 30-Day Stroke Mortality Rates in China. *PLoS One* 2016; 11: e0166069. DOI: 10.1371/journal.pone.0166069.

12. Zhang N, Liu G, Zhang G, et al. A risk score based on Get With the Guidelines-Stroke program data works in patients with acute ischemic stroke in China. *Stroke* 2012; 43: 3108-3109. DOI: 10.1161/strokeaha.112.669085.

13. Ebinger M, Siegerink B, Kunz A, et al. Association Between Dispatch of Mobile Stroke Units and Functional Outcomes Among Patients With Acute Ischemic Stroke in Berlin. *JAMA* 2021; 325: 454-466. DOI: 10.1001/jama.2020.26345.

14. Napierkowski I, Lorenz-Meyer I, Hille A, et al. Follow-up of Patients With Stroke, Based on Opt-Out Choice Potential Approach for Acute Care Quality Registries or Observational Studies. *Neurology* 2022; 99: e1335-e1344. DOI: 10.1212/wnl.0000000000200916.

15. Van Calster B, Nieboer D, Vergouwe Y, et al. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016; 74: 167-176. DOI: 10.1016/j.jclinepi.2015.12.005.

16. Fonarow GC, Saver JL, Smith EE, et al. Relationship of national institutes of health stroke scale to 30-day mortality in medicare beneficiaries with acute ischemic stroke. *J Am Heart Assoc* 2012; 1: 42-50. DOI: 10.1161/jaha.111.000034.

17. Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016; 353: i2416. DOI: 10.1136/bmj.i2416.

18. Steyerberg EW and Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014; 35: 1925-1931. 20140604. DOI: 10.1093/eurheartj/ehu207.

19. Hillmann S, Wiedmann S, Rücker V, et al. Stroke unit care in germany: the german stroke registers study group (ADSR). *BMC Neurol* 2017; 17: 49. DOI: 10.1186/s12883-017-0819-0.

20. Tong D, Reeves MJ, Hernandez AF, et al. Times from symptom onset to hospital arrival in the Get with the Guidelines--Stroke Program 2002 to 2009: temporal trends and implications. *Stroke* 2012; 43: 1912-1917. DOI: 10.1161/strokeaha.111.644963.

Figure legends

Figure 1. Flow chart showing eligibility criteria applied to the B-SPATIAL study population for the external validation of Bray et al.'s model (2014) for post-stroke 30-day all-cause mortality [3] (left) and Smith et al.'s model (2010) for post-stroke in-hospital mortality[8] (right). *Patients, who opted out or for whom a mobile stroke unit was dispatched as part of the B-PROUD study were not included.

Figure 2. External validation of Bray et al.'s model for post-stroke 30-day all-cause mortality in the B-SPATIAL registry (Main analysis). Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.

Figure 3. External validation of Smith et al.'s model for post-stroke in-hospital mortality in the B-SPATIAL registry (main analysis). Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.

Tables

Table 1. Characteristics of study population of stroke patients from the B-SPATIAL registry included in the external validation of Bray et al.'s model for post-stroke 30-day all-cause mortality, stratified by outcome status.

Characteristic		Alive at 30d N = 7,116	Deceased at 30d N = 763	All N = 7,879
Age	median (IQR)	75 (63, 81)	84 (78, 90)	75 (64, 82)
Male sex	n (%)	3,994 (56.1%)	342 (44.8%)	4,336 (55.0%)
Stroke type				
Hemorrhagic	n (%)	445 (6.3%)	156 (20.4%)	601 (7.6%)
Ischemic	n (%)	6,671 (93.7%)	607 (79.6%)	7,278 (92.4%)
NIHSS at admission	median (IQR)	4 (2, 9)	17 (10, 22)	5 (2, 11)
Atrial fibrillation	n (%)	1,937 (27.2%)	353 (46.3%)	2,290 (29.1%)
Diabetes mellitus[1]	n (%)	1,885 (26.6%)	218 (28.6%)	2,103 (26.8%)
Hypertension[1]	n (%)	5,728 (80.6%)	658 (86.7%)	6,386 (81.2%)
Length of hospitalization in days	median (IQR)	6 (4, 9)	5 (3, 9)	6 (4, 9)
Systemic thrombolysis[1]	n (%)	3,102 (46.6%)	289 (47.6%)	3,391 (46.7%)

NIHSS: National Institutes of Health Stroke Scale; IQR: Interquartile range;
[1] Information was missing for diabetes (n = 35), hypertension (n = 12) and thrombolytic therapy (n = 612)

Table 2. Characteristics of study population of ischemic stroke patients from the B-SPATIAL registry included in the external validation of Smith et al.'s model for post-stroke in-hospital mortality, stratified by outcome status

Characteristic		Alive at discharge N = 1,826	Dead In-hospital N = 105	All N = 1,931
Age	median (IQR)	74 (62, 81)	83 (78, 88)	75 (63, 82)
Male sex	n (%)	1,038 (56.8%)	47 (44.8%)	1,085 (56.2%)
NIHSS at admission	median (IQR)	4 (2, 9)	17 (13, 22)	4 (2, 10)
Mode of arrival				
private	n (%)	290 (15.9%)	4 (3.8%)	294 (15.2%)
not via ED	n (%)	68 (3.7%)	2 (1.9%)	70 (3.6%)
ambulance	n (%)	1,468 (80.4%)	99 (94.3%)	1,567 (81.1%)
Length of hospitalization in days	median (IQR)	5 (4, 8)	5 (2, 11)	5 (3, 8)

Systemic thrombolysis[1]		953 (52.3%)	60 (57.1%)	1,013 (52.6%)
Atrial fibrillation	n (%)	521 (28.5%)	50 (47.6%)	571 (29.6%)
Previous stroke or TIA	n (%)	489 (26.8%)	22 (21.0%)	511 (26.5%)
Coronary artery disease	n (%)	309 (16.9%)	25 (23.8%)	334 (17.3%)
Diabetes mellitus	n (%)	438 (24.0%)	31 (29.5%)	469 (24.3%)
Dyslipidemia	n (%)	1,119 (61.3%)	50 (47.6%)	1,169 (60.5%)
Hypertension[1]	n (%)	1,428 (78.3%)	86 (84.3%)	1,514 (78.6%)

NIHSS: National Institutes of Health Stroke Scale; IQR: Interquartile range;
ED: Emergency department; TIA: Transient ischemic attack
[1] Information was missing for hypertension (n = 5) and thrombolytic therapy (n = 4)

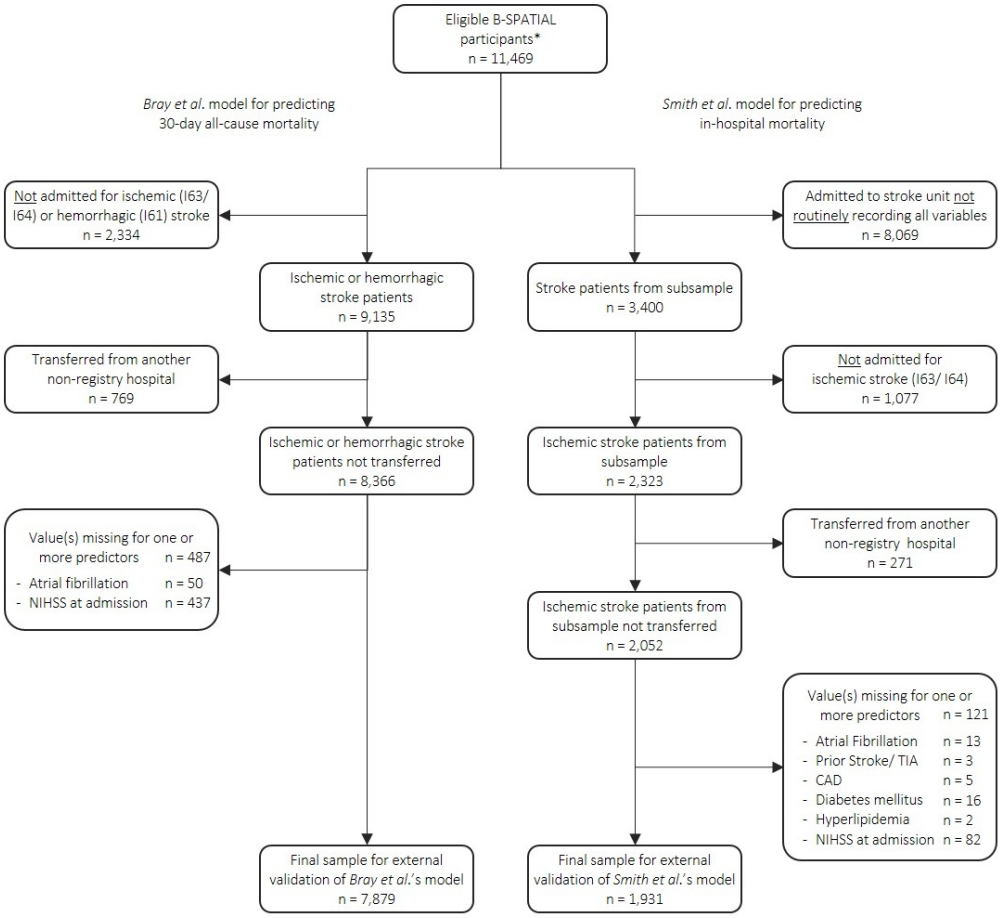


Figure 1. Flow chart showing eligibility criteria applied to the B-SPATIAL study population for the external validation of Bray et al.'s model (2014) for post-stroke 30-day all-cause mortality 3 (left) and Smith et al.'s model (2010) for post-stroke in-hospital mortality8 (right). *Patients, who opted out or for whom a mobile stroke unit was dispatched as part of the B-PROUD study were not included.

182x166mm (150 x 150 DPI)

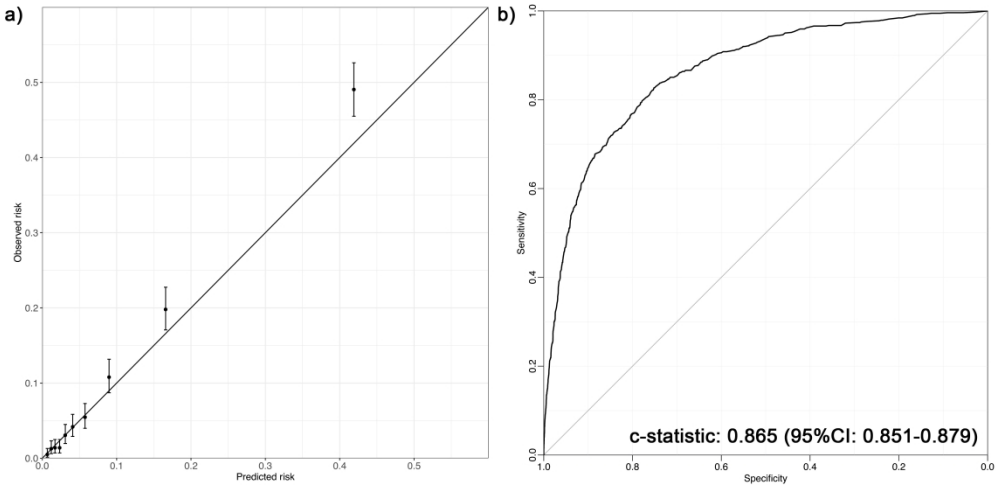


Figure 2. External validation of Bray et al.'s model for post-stroke 30-day all-cause mortality in the B-SPATIAL registry (main analysis). Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.

1375x661mm (96 x 96 DPI)

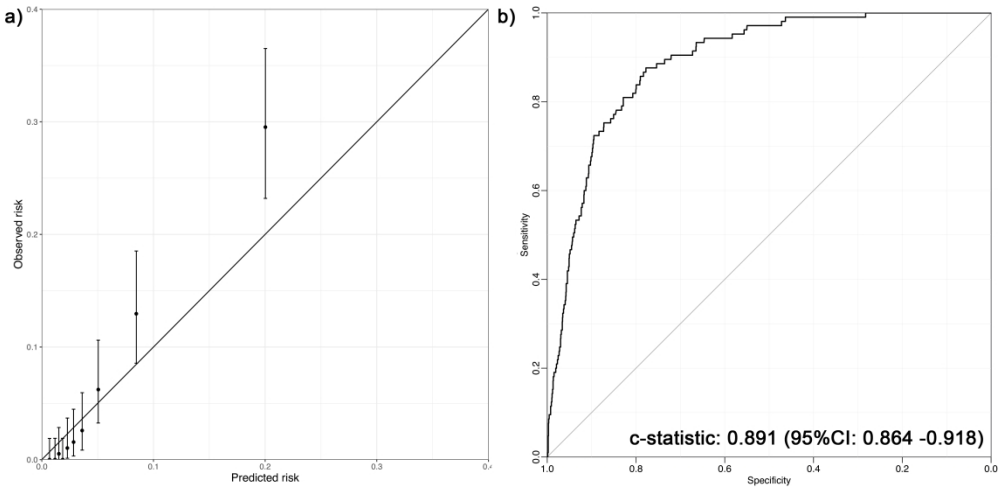


Figure 3. External validation of Smith et al.'s model for post-stroke in-hospital mortality in the B-SPATIAL registry (main analysis). Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.

1834x881mm (72 x 72 DPI)

Supplement - External validation of risk prediction models for post-stroke mortality in Berlin

Supplemental Tables

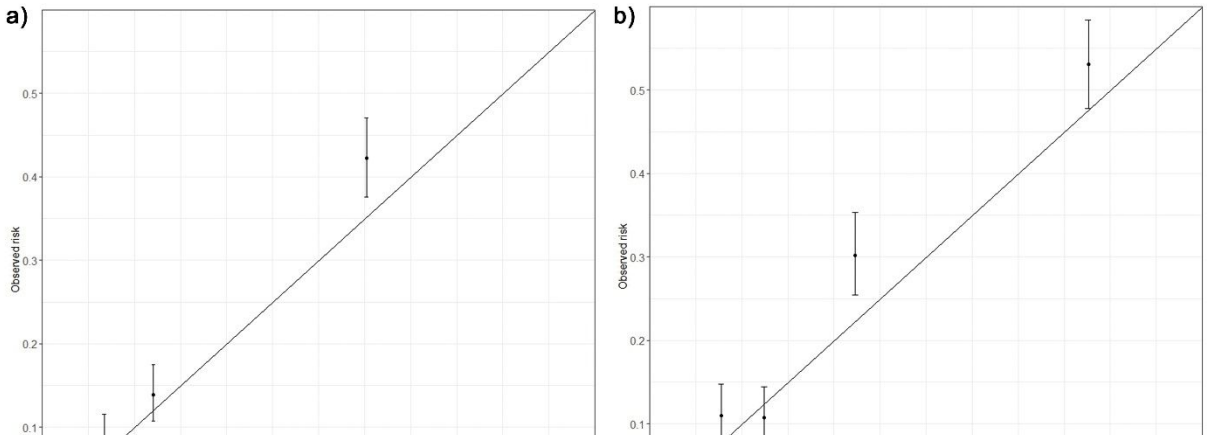
Supplementary Table 1. *External validation of Bray et al.'s model for post-stroke 30-day all-cause mortality in the B-SPATIAL registry (Main analysis). Observed and predicted 30-day mortality risk after stroke in decile groups.*

Decile	n	Predicted mortality	Observed mortality	
		%	%	95%CI
1	788	0.68	0.51	0.14 - 1.29
2	788	1.20	1.27	0.61 - 2.32
3	788	1.70	1.40	0.70 - 2.48
4	788	2.33	1.40	0.70 - 2.48
5	788	3.09	3.05	1.96 - 4.50
6	788	4.08	4.19	2.90 - 5.83
7	788	5.74	5.46	3.98 - 7.28
8	788	8.98	10.8	8.71 - 13.2
9	788	16.6	19.8	17.1 - 22.8
10	787	41.9	49.0	45.5 - 52.6

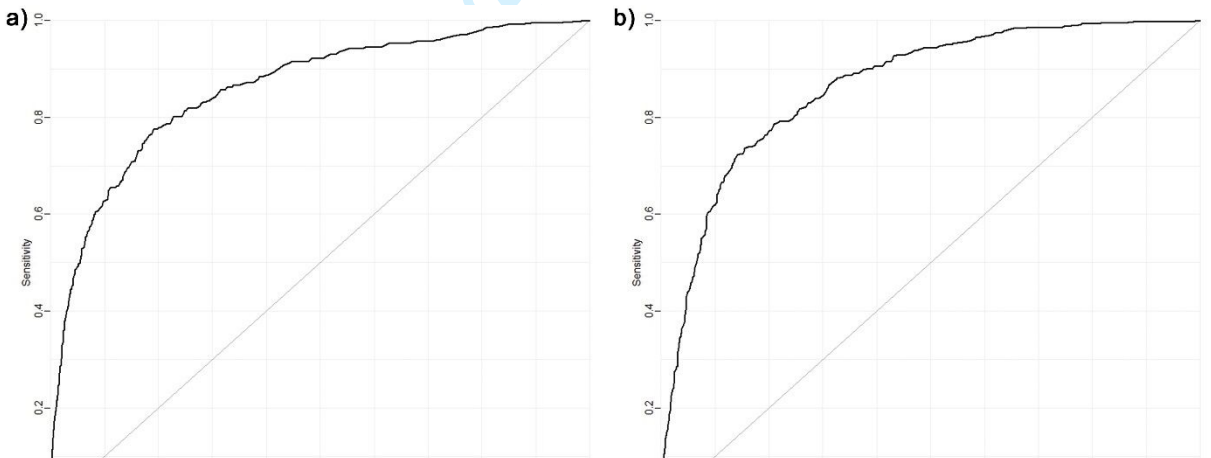
Supplementary Table 2. *External validation of Smith et al.'s model for post-stroke in-hospital mortality in the B-SPATIAL registry (Main analysis). Observed and predicted in-hospital mortality risk after stroke in decile groups.*

Decile	n	Predicted mortality	Observed mortality	
		%	%	95%CI
1	194	0.69	0.00	0.00 - 1.88
2	193	1.18	0.00	0.00 - 1.89
3	193	1.52	0.52	0.01 - 2.85
4	193	1.87	0.00	0.00 - 1.89
5	193	2.29	1.04	0.13 - 3.69
6	193	2.84	1.55	0.32 - 4.48
7	193	3.61	2.59	0.85 - 5.94
8	193	5.05	6.22	3.25 - 10.6
9	193	8.46	13.0	8.56 - 18.5
10	193	20.0	29.5	23.2 - 36.5

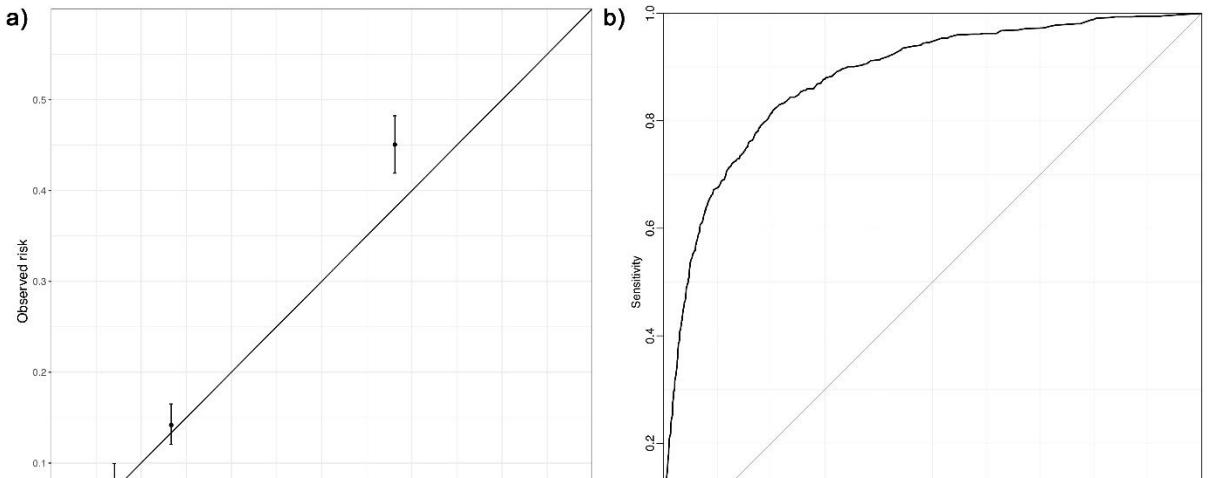
Supplement Figures



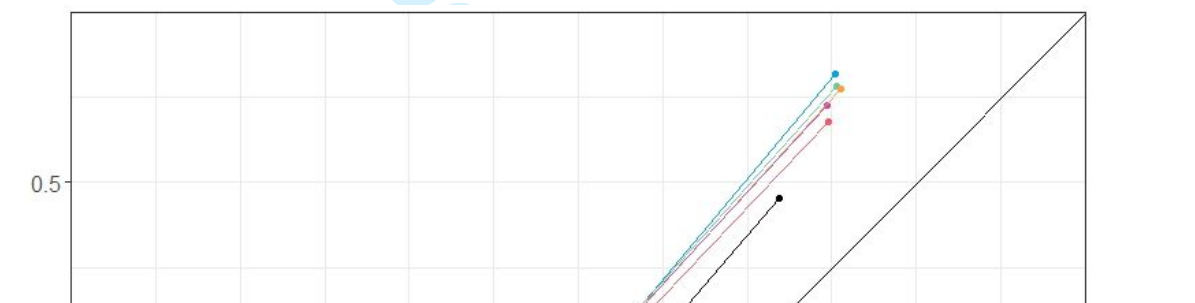
Supplementary Figure S1. *Calibration plots for the external validation of Bray et al.'s model for post-stroke 30-day all-cause mortality in a) male and b) non-male patients of the B-SPATIAL registry.*



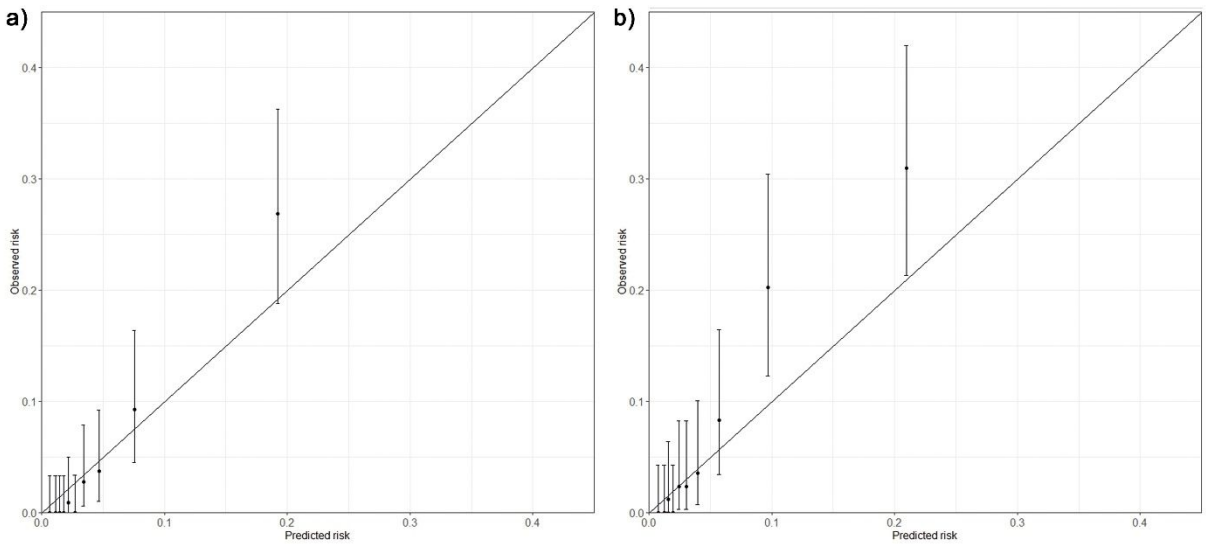
Supplementary Figure S2. *Receiver operating characteristic (ROC) curve for the external validation of Bray et al.'s model for post-stroke 30-day all-cause mortality in a) male and b) non-male patients of the B-SPATIAL registry.*



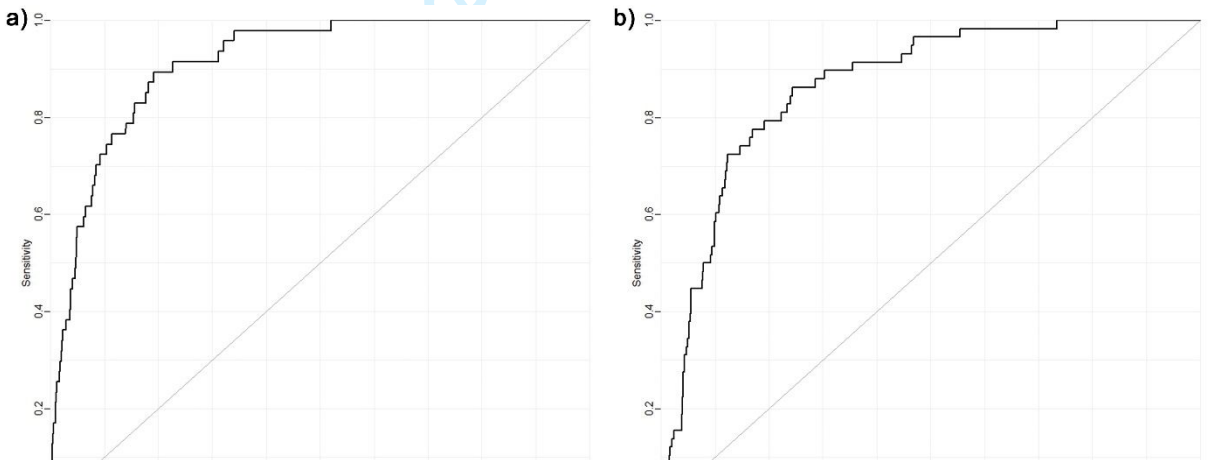
Supplementary Figure S3. *Sensitivity analysis for the external validation of Bray et al.'s model for post-stroke 30-day all-cause mortality in the B-SPATIAL registry. In this sensitivity analysis patients diagnosed with TIA were included as ischemic stroke patients. Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.*



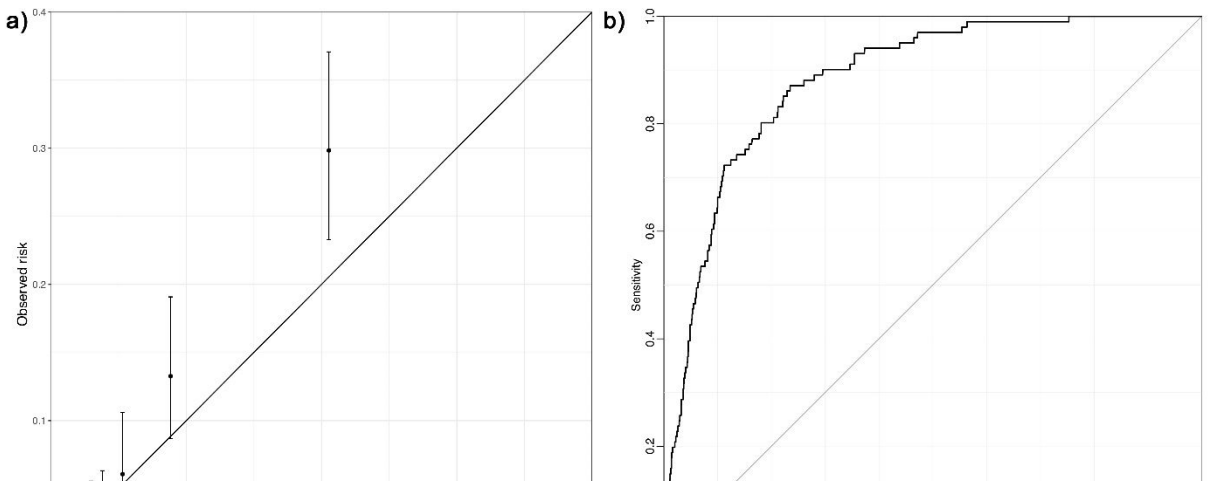
Supplementary Figure S4. *Sensitivity analysis for the external validation of Bray et al.'s model for post-stroke 30-day all-cause mortality in the B-SPATIAL registry. In this sensitivity analysis missing values for predictors were imputed by multiple imputation. Calibration plot by imputed data set (0: original without imputation; 1-5: imputed data sets).*



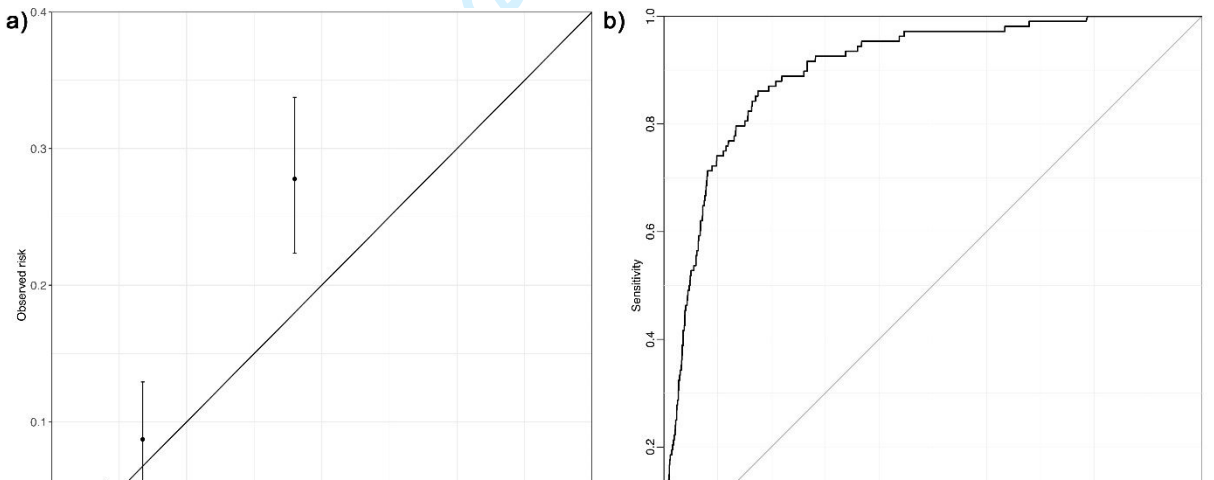
***Supplementary Figure S5**. Calibration plots for the external validation of Smith et al.'s model for post-stroke in-hospital mortality in a) male and b) non-male patients of the B-SPATIAL registry.*



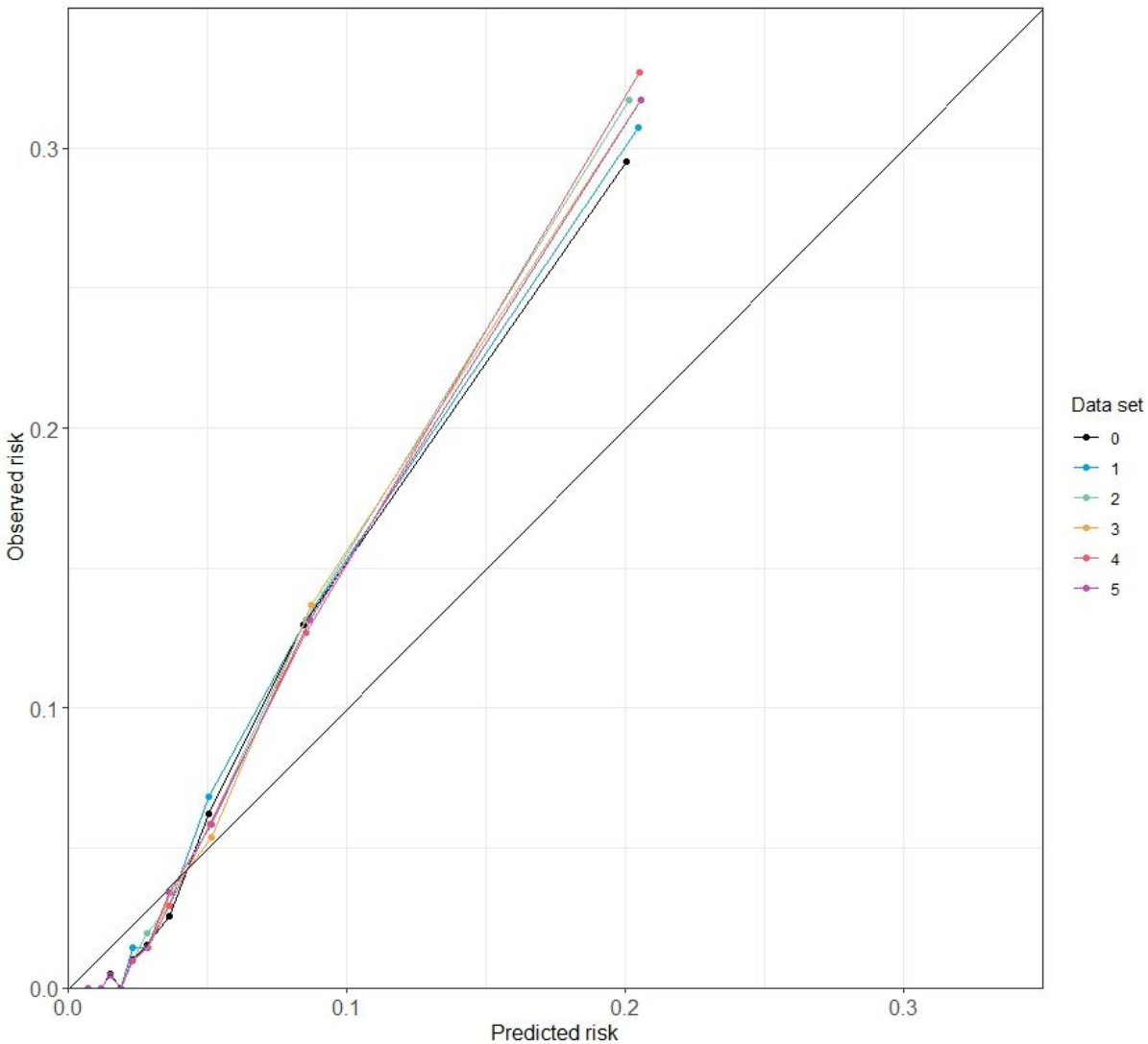
***Supplementary Figure S6**. Receiver operating characteristic (ROC) curve for the external validation of Smith et al.'s model for in-hospital mortality in a) male and b) non-male patients of the B-SPATIAL registry.*



Supplementary Figure S7. *Sensitivity analysis for the external validation of Smith et al.'s model for post-stroke in-hospital mortality in the B-SPATIAL registry. In this sensitivity analysis all patients with unknown or missing mode of arrival were excluded. Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.*



Supplementary Figure S8. *Sensitivity analysis for the external validation of Smith et al.'s model for post-stroke in-hospital mortality in the B-SPATIAL registry. In this sensitivity analysis patients diagnosed with TIA were included as ischemic stroke patients. Panel a) shows the calibration plot and Panel b) the receiver operating characteristic (ROC) curve.*



Supplementary Figure S9. *Sensitivity analysis for the external validation of Smith et al.'s model for in-hospital mortality in the B-SPATIAL registry. In this sensitivity analysis missing values for predictors were imputed by multiple imputation. Calibration plot by imputed data set (0: original without imputation; 1-5: imputed data sets).*

R code

```
# Title: External validation of prediction models for post-stroke mortality within
the "Berlin – SPecific Acute Treatment in Ischemic or hemorrhAgic stroke with Long
term follow-up" (B-SPATIAL) registry

# Date: 2024-02-02

# Note: the following packages must be installed (if not already installed):
# install.packages("tidyverse")
# install.packages("lubridate")
# install.packages("gtsummary")
# install.packages("fastDummies")
# install.packages("pROC")
#
# The file includes only the R-Code for the main analysis



# External validation of Bray et al.'s model A for 30-day mortality after stroke

# Definition of study population: Bray et al.'s model A
#
# Exclusion of patients…
# - with no ischemic or hemorrhagic stroke (Variable: stroke_type)
# - transferred to or form another hospital (Variable: transfer)
# - with missing age or age <18 or >120 (Variable: age)
# - with missing information on atrial fibrillation (Variable: atr_fibr)
# - with missing information on NIHSS at admission (Variable: nihss_adm)
# - with implausible NIHSS (<0 or >42)
# - with no information on death within 30 days (Variable: dead_30d)

data_pms_01 <- data %>%
  filter(stroke_type %in% c("ischemic", "hemorrhagic"),
         !transf == "Yes" | is.na(transf),
         !is.na(age),
         age >= 18 & age <= 120,
         !is.na(atr_fibr),
         !is.na(nihss_adm),
         nihss_adm >= 0 & nihss_adm <= 42,
         !is.na(dead_30d))


# Calculate predicted probabilities using model equation: Bray et al.'s model A
#
# Predictors:
# - Age in years (categorical): <60, 60-69, 70-79, 80-89, >= 90
# - NIHSS at admission (continuous)
# - Atrial fibrillation (dichotomous)
# - Stroke type (categorical): Ischemic stroke, intracerebral hemorrhage

data_pms_01  <- data_pms_01 %>%
  mutate(age_cat = factor(case_when(age < 60 ~ "<60",
                                    age >= 60 & age < 70 ~ "60-69",
                                    age >= 70 & age < 80 ~ "70-79",
                                    age >= 80 & age < 90 ~ "80-89",
                                    age >= 90 ~ ">= 90"),
                  levels = c("<60", "60-69", "70-79", "80-89", ">= 90"))) %>%
  dummy_cols(select_columns = c("age_cat", "stroke_type", "atr_fibr")) %>%
  mutate(pms_01_pred = -5.250 + `age_cat_60-69` * 0.624 + `age_cat_70-79` * 1.033 +
         `age_cat_80-89` * 1.488 + `age_cat_>= 90` * 1.781 + nihss_adm * 0.137 +
         atr_fibr_1 * 0.425 + stroke_type_hemorrhagic * 0.870) %>%
  mutate(pms_01_prob = 1/(1+exp(-pms_01_pred)))
```

```
# Description of study population: Bray et al.'s model A

# Print Table 1 - Characteristics of included stroke patients
tbl_summary(data_pms_01 %>%
              select(age, sex, stroke_type, atr_fibr, nihss_adm, dead_30d) %>%
              mutate(dead_30d = as.factor(dead_30d)) %>%
              mutate(dead_30d = recode_factor(dead_30d, `0` = "Survived",
                                              `1` = "Died")),
            digits = list(all_categorical() ~ c(0, 1)),
            label = list(age ~ "Age, median (IQR)",
                         sex ~ "Sex",
                         stroke_type ~ "Stroke type",
                         atr_fibr ~ "Atrial fibrillation",
                         nihss_adm ~ "NIHSS at admission, median (IQR)"),
            by = dead_30d) %>%
  add_overall()



# Calibration: Bray et al.'s model A

# Calibration-in-the-large

# Predicted 30-day mortality risk
mean(data_pms_01$pms_01_prob)

# Observed 30-day mortality risk
mean(data_pms_01$dead_30d)

# Predicted absolute number of death within 30-days
round(mean(data_pms_01$pms_01_prob) * nrow(data_pms_01), digits = 0)

# Predicted absolute number of death within 30-days
mean(data_pms_01$dead_30d) * nrow(data_pms_01)

# Observed to expected ratio
mean(data_pms_01$dead_30d) / mean(data_pms_01$pms_01_prob)


# Calibration plot
# Grouping of stroke patients into deciles of observed mortality
data_pms_01$pms_01_prob_group = ntile(data_pms_01$pms_01_prob, 10)

# Table of predicted and observed 30-day mortality by decile
# Including calculation 95%-confidence intervals for observed mortality

pp <- data_pms_01 %>%
  group_by(pms_01_prob_group) %>%
  summarise(rel_pred = mean(pms_01_prob), rel_obs = mean(dead_30d),
            abs_obs = sum(dead_30d), n = n()) %>%
  mutate(ci = binconf(abs_obs, n, alpha=0.05, method = c("exact"))) %>%
  mutate(ci_l = ci[,"Lower"], ci_u = ci[,"Upper"]) %>%
  select(-ci)


# Visualization of calibration plot
ggplot(data = pp, aes(x = rel_pred, y = rel_obs)) +
  geom_point() +
  theme_bw() +
  geom_abline(intercept = 0, slope = 1, col = "black") +
  geom_errorbar(aes(min = ci_l, max = ci_u)) +
  xlab("Predicted risk") +
  ylab("Observed risk") +
  scale_x_continuous(expand = c(0, 0), limits = c(0,0.4)) +
  scale_y_continuous(expand = c(0, 0), limits = c(0,0.4)) +
  theme(axis.text = element_text(size = 10),
        axis.title = element_text(size = 14),
        plot.margin = margin(4, 4, 4, 4))
```

```
# Calibration intercept and slope
# Fit logistic regression model
fit <- glm(dead_30d ~ pms_01_pred, data = data_pms_01, family = binomial)

# Print calibration intercept and slope
fit$coefficients
confint(fit)


# Discrimination: Bray et al.'s model A

# AUC and ROC-Curve
roc <- roc(data_pms_01$dead_30d, data_pms_01$pms_01_prob, ci = TRUE,
           ci.alpha = 0.95)

# Print c-statistic and 95%-confidence interval
auc(roc)
ci.auc(roc)

# Visualize ROC-curve
plot(roc, grid=TRUE, print.auc=FALSE, xlim=c(1,0), ylim=c(0,1),
     xaxs = "i", yaxs = "i", asp=NA, mar=c(2.5, 2.5, .5, .5)+.1,
     mgp=c(1.5, 0.5, 0), grid.lty=1, grid.lwd=.5, grid.col="#EEEEEE")
```

```
# External validation of Smith et al.'s model including NIHSS for in-hospital
mortality after stroke

# Definition of study population: Smith et al.'s model
#
# Exclusion of patients…
# - from hospitals not recoding the relevant predictors routinely
# - with no ischemic stroke (Variable: stroke_type)
# - transferred to or form another hospital (Variable: transfer)
# - with missing information on sex
# - with missing information on age or age <18 or >120 (Variable: age)
# - with missing information on atrial fibrillation (Variable: atr_fibr)
# - with missing information on previous stroke or TIA (Variable: prior_event)
# - with missing information on coronary artery disease (Variable: chd)
# - with missing information on diabetes mellitus (Variable: dm)
# - with missing information on hyperlipidemia (Variable: hlp)
# - with missing information on NIHSS at admission (Variable: nihss_adm)
# - with implausible NIHSS (<0 or >42)
# - with no information on in-hospital death (Variable: dead_inhospital)

data_pms_02 <- data %>%
  filter(stroke_type %in% c("ischemic"),
         !transf == "Yes" | is.na(transf),
         !is.na(sex),
         !is.na(age),
         age >= 18 & age <= 120,
         !is.na(atr_fibr),
         !is.na(prior_event),
         !is.na(chd),
         !is.na(dm),
         !is.na(hlp),
         !is.na(nihss_adm),
         nihss_adm >= 0 & nihss_adm <= 42,
         !is.na(dead_inhospital))


# Calculate Probabilities for PMS using model equation: Smith et al.'s model
# Model 2 (including NIHSS)
#
# Predictors:
# - Age (continuous): per 1 year over 60
# - Sex (categorical): male, non-male
# - NIHSS at admission (continuous)
# - Mode of arrival (categorical): By private transport, by Ambulance, Not via ED
# - Atrial fibrillation (dichotomous)
# - Previous stroke or TIA (dichotomous)
# - Coronary artery disease (dichotomous)
# - Diabetes mellitus (dichotomous)
# - History of dyslipidemia (dichotomous)
#
# Imputation of mode of arrival missing or unknown to "private" (Variable: transp)

data_pms_02 <- data_pms_02 %>%
  mutate(age_60 = case_when(age <= 60 ~ 0, age >= 60 ~ age-60),
         transp_pms = factor(case_when(transp %in% c("private", "unknown") |
                           is.na(transp) ~ "private",
                           transp %in% c("clinical acute event",
                            "secondary relocation") ~ "not via ED",
                           transp == "ambulance service" ~ "ambulance"),
                              levels = c("private", "not via ED", "ambulance"))) %>%
  dummy_cols(select_columns = c("sex", "transp_pms", "atr_fibr",
                                "prior_event", "chd", "dm", "hlp")) %>%
  mutate(pms_02_pred = -5.3169 + age_60 * 0.0176 + sex_male * 0.167 +
         nihss_adm * 0.116 + `transp_pms_not via ED` * 0.9611 +
         transp_pms_ambulance * 0.7654 + atr_fibr_1 * 0.300 +
         prior_event_1 * -0.112 + chd_1 * 0.268 + dm_1 * 0.124 + hlp_1 * -0.132)%>%
  mutate(pms_02_prob = 1/(1+exp(-pms_02_pred)))
```

```
# Description of study population: Smith et al.'s model

# Print Table 1 - Characteristics of included stroke patients
tbl_summary(data_pms_02 %>%
                        select(age, sex, nihss_adm, transp_pms, atr_fibr,
                               prior_event, chd, dm, hlp, dead_inhospital) %>%
                        mutate(dead_inhospital = as.factor(dead_inhospital)) %>%
                        mutate(dead_inhospital = recode_factor(dead_inhospital,
                                                               `0` = "Survived", `1` = "Died")),
                      by = dead_inhospital,
                      digits = list(all_categorical() ~ c(0, 1)),
                      label = list(age ~ "Age, median (IQR)",
                                   sex ~ "Sex",
                                   nihss_adm ~ "NIHSS at admission, median (IQR)",
                                   transp_pms ~ "Mode of arrival",
                                   atr_fibr ~ "Atrial fibrillation",
                                   prior_event ~ "Previous stroke or TIA",
                                   chd ~ "Coronary artery disease",
                                   dm ~ "Diabetes mellitus",
                                   hlp ~ "Dyslipidemia")) %>%
  add_overall()


# Calibration: Smith et al.'s model
# Calibration-in-the-large

# Predicted in-hospital mortality risk
mean(data_pms_02$pms_02_prob)

# Observed in-hospital mortality risk
mean(data_pms_02$dead_inhospital)

# Predicted absolute number of in-hospital death
round(mean(data_pms_02$pms_02_prob) * nrow(data_pms_02), digits = 0)

# Observed absolute number of in-hospital death
mean(data_pms_02$dead_inhospital) * nrow(data_pms_02)

# Observed to expected ratio
mean(data_pms_02$dead_inhospital) / mean(data_pms_02$pms_02_prob)


# Calibration plot
# Grouping of stroke patients into deciles of observed mortality
data_pms_02$pms_02_prob_group = ntile(data_pms_02$pms_02_prob, 10)

# Table of predicted and observed in-hospital mortality by decile
# Including calculation 95%-confidence intervals for observed mortality
pp <- data_pms_02 %>%
  group_by(pms_02_prob_group) %>%
  summarise(rel_pred = mean(pms_02_prob), rel_obs = mean(dead_inhospital),
            abs_obs = sum(dead_inhospital), n = n()) %>%
  mutate(ci = binconf(abs_obs, n, alpha=0.05,method=c("exact"))) %>%
  mutate(ci_l = ci[,"Lower"], ci_u = ci[,"Upper"]) %>%
  select(-ci)

# Visualization of calibration plot
ggplot(data = pp, aes(x = rel_pred, y = rel_obs)) +
  geom_point() +
  theme_bw() +
  geom_abline(intercept = 0, slope = 1, col = "black") +
  geom_errorbar(aes(min = ci_l, max = ci_u)) +
  xlab("Predicted risk") +
  ylab("Observed risk") +
  scale_x_continuous(expand = c(0, 0), limits = c(0,0.4)) +
  scale_y_continuous(expand = c(0, 0), limits = c(0,0.4)) +
  theme(axis.text = element_text(size = 10),
        axis.title = element_text(size = 14),
        plot.margin = margin(4, 4, 4, 4))
```

```
# Calibration intercept and slope
# Fit logistic regression model
fit <- glm(dead_inhospital ~ pms_02_pred, data = data_pms_02, family = binomial)

# Print calibration intercept and slope
fit$coefficients
confint(fit)


# Discrimination: Smith et al.'s model

# AUC and ROC-Curve
roc <- roc(data_pms_02$dead_inhospital, data_pms_02$pms_02_prob, ci = TRUE,
           ci.alpha = 0.95)

# Print c-statistic and 95%-confidence interval
auc(roc)
ci.auc(roc)

# Visualize ROC-curve
plot(roc, grid=TRUE, print.auc=FALSE, xlim=c(1,0), ylim=c(0,1),
     xaxs = "i", yaxs = "i", asp=NA, mar=c(2.5, 2.5, .5, .5)+.1,
     mgp=c(1.5, 0.5, 0), grid.lty=1, grid.lwd=.5, grid.col="#EEEEEE")
```