The SeLECT score: development and validation of a novel prognostic model to predict late seizures after ischemic stroke.

Marian Galovic MD[1,2], Nico Döhler[2], Barbara Erdélyi-Canavese MD,[2] Ansgar Felbecker MD,[2] Philip Siebel MD,[2] Julian Conrad MD[3,4], Stefan Evers PhD[3,5], Michael Winklehner[6], Tim J von Oertzen FRCP[6], Hans-Peter Haring MD[6], Anna Serafini MD[7], Giorgia Gregoraci MD[8,9], Mariarosaria Valente MD[7,10], Francesco Janes PhD,[7] Gian Luigi Gigli MD[7,10], Mark R Keezer MDCM[1,11,12], John S Duncan FRCP[1], Josemir W Sander FRCP[1,12], Matthias J Koepp PhD[1], Barbara Tettenborn MD[2]

1. National Institute for Health Research University College London Hospitals Biomedical Research Centre, Department of Clinical and Experimental Epilepsy, UCL Institute of Neurology, Queen Square, London, WC1N 3BG, and Chalfont Centre for Epilepsy, Chalfont St. Peter, SL9 0RJ, United Kingdom
2. Department of Neurology, Kantonsspital St. Gallen, St. Gallen, Switzerland.
3. Department of Neurology, University of Münster, Münster, Germany.
4. Department of Neurology and German Center for Vertigo and Balance Disorders-IFB, Ludwig-Maximilians-University, Munich, Germany.
5. Department of Neurology, Krankenhaus Lindenbrunn, Coppenbrügge, Germany.
6. Department of Neurology 1, Kepler Universitätsklinikum - Neuromed Campus, Linz, Austria.
7. Neurology Unit, University-Hospital, Udine, Italy.
8. Department of Medicine, Section of Statistics, University of Udine, Italy.
9. Department of Medicine, University of Udine, Italy.
10. Department of Experimental and Clinical Medical Sciences, University of Udine Medical School, Udine, Italy.
11. Centre hospitalier de l'Université de Montréal (CHUM), Hôpital Notre-Dame, Montréal, QC, Canada.
12. Stichting Epilepsie Instellingen Nederland (SEIN), Heemstede 2103SW, The Netherlands.

Corresponding author:
Barbara Tettenborn, MD
Department of Neurology, Kantonsspital St. Gallen
St. Gallen, Switzerland
Tel: +41 71 494 1652
Email: barbara.tettenborn@kssg.ch

Abstract

Background: Stroke is one of the leading causes of acquired epilepsy in adults. An instrument to predict those at high risk of developing post-stroke seizures is not available. We aimed to develop and validate a prognostic model of late (>7 days) seizures after ischemic stroke.

Methods: The SeLECT score was developed in 1200 people who had an ischemic stroke in Switzerland using backward elimination of a multivariable Cox proportional hazards model. We externally validated this score in 1169 participants from three independent international cohorts (Austria, Germany, Italy) and assessed its performance with the concordance statistic and calibration plots.

Findings: Overall, late seizures occurred in 4% (95% confidence interval [CI] 4% - 5%) of people during the first year and in 8% (95% CI 6% - 9%) within five years after stroke. The final model included five variables and was termed SeLECT based on the first letters of the included parameters (Severity of stroke, Large-artery atherosclerotic aetiology, Early seizures, Cortical involvement, Territory of MCA involvement). The lowest SeLECT value (0 points) was associated with a 0·7% risk of late seizures within one year after stroke (1·2% within five years), whereas the highest value (9 points) predicted a 63% risk of late seizures within one year (83% within five years). The model had an overall concordance statistic of 0·77 in the validation cohorts. Calibration plots indicated high agreement of predicted and observed outcomes.

Interpretation: This easily applied instrument was a good predictor of the risk of late seizures after stroke in triple external validation and is freely available as a smartphone app. The SeLECT score has the potential to identify those at high risk of seizures and is a step towards more personalised medicine. It can inform the selection of an enriched population for antiepileptogenic treatment trials and will guide the recruitment for biomarker studies of epileptogenesis.

Funding: None.

Introduction

Stroke is the third leading cause of death and a major cause of disability in Europe, affecting one in six adults with an estimated 3 to 6 million stroke cases annually.[1] Stroke survivors have an increased risk of spontaneous seizures with stroke being the major cause of acquired epilepsy in adults.[2] Seizures can negatively influence post-stroke outcome[3] and quality of life[4] and may increase in-hospital costs.[5]

Seizures after stroke are defined as early (occurring ≤ 7 days after the insult) or late (> 7 days).[6] According to the current International League Against Epilepsy (ILAE) definition, a single late seizure after stroke qualifies as structural epilepsy due to the high (> 60%) risk of recurrence over the next 10 years.[7] Early seizures alone are not sufficient to make the diagnosis of epilepsy as they are considered to be provoked.[6]

Typically, there is a latent period of several weeks to years between stroke and the first late seizure. During this time, epileptogenic processes lead to the development of tissue capable of generating spontaneous seizures.[8] Recent developments of potentially antiepileptogenic compounds in animal models emphasised the need for an early intervention within the initial weeks after stroke, before an epileptogenic cascade of events sets in.[8] All previous treatment trials to prevent epileptogenesis after stroke in humans were unsuccessful.[9]

A contributory reason for these failures was the difficulty in identifying those at high risk of seizures following stroke and the need for prolonged follow-up. Trials in an unselected population would require large sample sizes and thus be costly, as late seizures only affect up to 9% of people who have suffered an ischemic stroke within the previous five years.[10,11]

Defining a prognostic biomarker of seizures after stroke is a crucial goal to advance the development of antiepileptogenic treatments[9,12] and could be important to guide the clinical management of stroke survivors.

A multitude of studies identified variables associated with late seizures after stroke. A validated instrument to synthesise a patient's clinical characteristics to predict the risk of

seizures individually and objectively is lacking. A previous attempt to develop a prognostic score[13] yielded promising results, but was of limited applicability as it was based on a small sample, only took recurrent seizures into account, and was not validated.

We aimed to develop and externally validate a prognostic risk score for late seizures after ischemic stroke that can be calculated from readily available routine clinical variables.

Methods

Study populations and procedures

We developed the model using a prospective registry of post-stroke seizures at a tertiary referral centre in St. Gallen, a major regional acute neurology centre in Eastern Switzerland. We included consecutive people with acute first-ever neuroimaging-confirmed ischemic stroke admitted between January 2002 and December 2008. Excluded were those with transient ischemic attacks (n = 495), prior history of stroke (n = 250), primary haemorrhagic stroke (n = 94), prior history of seizures (n = 43), re-infarction during follow-up (n = 9), and those who had potentially epileptogenic comorbidities (alcohol or drug abuse, n = 60; intracranial tumours, n = 28; cerebral venous thrombosis, n = 11; history of severe traumatic brain injury, n = 12; history of brain surgery, n = 4; other [including cerebral arteriovenous malformations, large cerebral aneurysms, cerebral vasculitis, hydrocephalus, and cerebral abnormalities of undetermined aetiology], n = 15). Eighty-five were lost to follow-up or died before follow-up was performed, leaving 1200 subjects for the final analysis.

Baseline characteristics in the derivation cohort were analysed by a neurologist at admission and diagnosis of stroke was confirmed at discharge. Brain scan analysis was performed using the best available imaging modality (MRI in 80%, CT in 20%) at discharge. Follow-up was carried out in all subjects after median 28 months (interquartile range [IQR] 21 to 47) with a structured telephone interview based on a validated questionnaire to detect seizures.[14] In subjects lacking capacity to perform the questionnaire, we interviewed close relatives and

additionally nursing staff or their general practitioner. Positive answers triggered a face-to-face neurological consultation and an electroencephalogram (EEG) to determine the epileptic nature of these episodes and to exclude seizure mimics. If the neurologist suspected a cause of epileptic seizures other than the index ischemic stroke, follow-up imaging was requested to rule out a co-pathology or re-infarction.

We validated the model in three external cohorts from Austria (n = 459), Germany (n = 311), and Italy (n = 399). The German (Münster)[15] and Italian (Udine)[10] cohorts have been described previously and are briefly summarised below.

In the Austrian nested case-control validation study, cases (people with late seizures) and controls (people without late seizures) were randomly selected from a larger cohort of consecutive people with a primary stroke diagnosis admitted to a tertiary referral centre in Linz between January 2005 and December 2014. Excluded were subjects with transient ischemic attack (TIA, n = 11), pre-existing brain lesions (i.e. intracranial tumour, trauma or other, n = 7), haemorrhagic stroke (n = 97), history of seizures (n = 8), cerebral venous thrombosis (n = 5), death within days after stroke (n = 9), and those with insufficient follow-up (n = 48). Baseline and follow-up data were retrospectively extracted from medical records. Twenty-eight (6%) cases received endovascular thrombectomy either with a suction device (until 2011) or using a stent-retriever (from 2011 onwards).[16] All subjects included had face-to-face follow-up neurological interviews three to six months after stroke and then yearly. Follow-up was terminated after median 10 (IQR 2 to 41) months and it was noted whether they had suffered late seizures. If seizures were suspected, additional EEG and brain imaging (MRI or CT) were performed.

The German validation cohort[15] included those with a first-ever hemispheric stroke admitted to a tertiary referral centre between January 2003 and March 2010. Excluded were subjects with recurrent stroke (n = 225), only infratentorial stroke (n = 322), haemorrhagic stroke (n = 32), transient ischemic attack (n = 64), or cerebral venous thrombosis (n = 14). Subjects who

died in hospital (n = 195), died before the interview (n = 21) or were lost to follow-up (n = 139) and those declining participation (n = 12) were also excluded. Baseline characteristics were extracted from medical records. Follow-up was completed after median 23 (IQR 12 to 44) months and all subjects received a structured telephone interview.[14]

The Italian validation cohort[10] was part of a population-based study in the Udine district with 153,312 residents. Included were all first-ever strokes occurring between April 2007 and March 2009. Excluded were those with TIAs (n = 178), previous brain lesions (i.e. brain tumour, exact number of subjects not recorded), non-ischemic stroke (n = 156), prior history of seizures or epilepsy (n = 22), those with missing time-to-event data (n = 108), and those deceased or lost to follow-up (n = 94). Participants were evaluated by a neurologist within 48 hours of admission. All subjects were followed up by a face-to-face interview with study neurologists at 1, 6, and 24 months after the stroke.

Approval was granted from all local ethical committees. The informed consent procedures are described in the Appendix.

Definitions

We used WHO definitions for stroke and ILAE definitions for seizures, which were classified as early (≤ 7 days post-stroke) or late (spontaneous unprovoked seizures > 7 days post-stroke).[6] Stroke severity was measured with the National Institutes of Health Stroke Scale (NIHSS) at admission and stratified into mild (≤ 3), moderate (4 to 10), and severe (≥ 11).[17,18] Missing NIHSS values were imputed using a previously validated algorithm.[19] Stroke aetiology was categorized according to the Trial of Org 10172 in Acute Stroke Treatment classification. Arterial territory was classified according to a published atlas[20] and white matter hyperintensities were defined as a score ≥ 1 on the age-related white matter changes score.[21]

Statistics

We searched for predictors of late seizures that (i) were repeatedly reported in studies or systematic reviews, (ii) can be easily ascertained in different settings with various clinical experience, and (iii) are part of the routine work up of people who had a stroke. We identified six potential parameters: stroke severity,[15,22-24] cortical involvement,[10,11,23,25,26] early seizures,[23,24,26] infarction in the middle cerebral artery (MCA) territory,[25,27] large-artery atherosclerotic stroke aetiology,[15,22] and age.[10,11,24] Inconsistent evidence exists for the role of thrombolysis[28] and cardioembolic stroke aetiology.[23,24] Lesion size[24,25] was not included as a parameter as it is not routinely determined in a clinical setting, the measurement approaches were not comparable between sites, and some data were not available. To identify any potential novel predictors not previously reported we also performed a univariable analysis with Cox proportional hazards regression within the derivation cohort. A minimum follow-up of at least 8 days was required in order to capture any late (> 7 days) seizures.

For model development in the Swiss derivation cohort, we included significant ($p<0{\cdot}05$) predictors from univariable analysis and those consistently reported in previous studies. All candidate variables were entered into Cox proportional hazards regression analysis and the assumption of proportional hazards was confirmed. We used a backward stepwise elimination approach to simplify the model on the basis of the Akaike Information Criterion (AIC). The AIC estimates the fit of each statistical model, penalises overfitting, and provides a means to select relevant variables that improve the model even if they do not reach the $p<0{\cdot}05$ threshold for statistical significance.[29] The final integer-based scoring system was developed by dividing the adjusted hazards ratios (aHR) of the remaining items in the derivation cohort (Table 3) by the median of the lowest three values (i.e. 1·7) and rounding to the nearest integer.

We assessed the predictive accuracy of the prognostic instrument in three validation cohorts with discrimination and calibration. Discrimination, i.e. the degree to which a model

differentiates between those with or without late seizures, was calculated with concordance (c) statistic, ranging from 0·5 (no discrimination) to 1·0 (perfect discrimination). Calibration, i.e. the agreement between the predicted and observed risk of seizures, was assessed with calibration plots. Perfect calibration is implied by a 45-degree diagonal line whereas relevant deviation above or below this line reflects underprediction or overprediction.

To increase precision, we calculated prediction estimates of late seizures in the combined derivation and validation cohorts. Confidence intervals (95% CI) for risk predictions were generated with bootstrapping methods to account for residual uncertainty.

Missing information was imputed with the multiple imputations method and the median of these values was chosen for final calculations. Sensitivity analyses of the original datasets were performed to test the robustness of the imputation approach. Calculations were done with R statistical software 3·3·3. The present study is reported in compliance with standard guidelines for prediction models (table in appendix).[30]

Role of the funding source

This study was not funded by an external agency. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Results

The derivation cohort included 1200 participants and three validation cohorts comprised 1169 people. Baseline characteristics for all cohorts are summarised in table 1. The Kaplan Meier plot of late seizures after stroke is displayed in figure 1A. Overall, late seizures occurred in 4% (95% CI 4% - 5%) of people during the first year after stroke and in 8% (95% CI 6% - 9%) within 5 years after stroke.

Model development

Univariable analysis in the derivation cohort (table 2) found a significant association of late seizures with several predictors which were consistently reported previously; stroke severity, stroke aetiology, location in MCA territory, cortical involvement, thrombolysis, and early seizures. Two novel potential predictors were also identified: stroke laterality and white-matter hyperintensities. These variables were entered into a multivariable model. Age was also included in the model despite being non-significant in the univariable analysis, as it was previously reported as a relevant predictor. Data on early treatment with antiepileptic drugs is described in the Appendix.

Five predictors remained in the final multivariable model (table 3) after simplification: cortical involvement, early seizure, NIHSS at admission, territory of MCA involvement, and large-artery atherosclerosis (figures 1B to 1F). Assigning point values to these items, an integer-based estimation system was developed and termed SeLECT based on the first letters of the five included parameters (panel 1).

Model performance

SeLECT was a significant predictor of late seizures in the pooled data of all three validation cohorts (HR 1·8 per point, 95% CI 1·6 – 2·1; p <0·0001). Model performance in the overall validation cohort showed a *c* statistic of 0·77 (95% CI 0·71 – 0·82). Discrimination in individual validation cohorts was: Austria 0·78 (95% CI 0·70 - 0·87), Germany 0·74 (95% CI 0·63 - 0·85), and Italy 0·81 (95% CI 0·69 - 0·93). Calibration plots indicated good fit of predicted and observed data (figures 2A to H).

Data was complete for 99·2% of the predictors (Switzerland 99·2%, Austria 100%, Germany 97%, Italy 99·7%) and 100% of the outcome parameters. A sensitivity analysis of available data generated similar results to the main analysis in all cohorts. The detailed results of this and other sensitivity analyses are outlined in the Appendix.

Model discrimination (*c* statistic 0·73, 95% CI 0·63 – 0·83) remained consistent in a subset of people receiving intravenous thrombolysis (n = 186) in the validation cohorts. Data for both intravenous thrombolysis (n = 102; *c* statistic 0·83, 95% CI 0·75 – 0·92) and endovascular thrombectomy (n = 28; *c* statistic 0·83, 95% CI 0·68 – 0·99) was available in the Austrian cohort and the findings suggest good discrimination in both treatment subgroups.

Model prediction

Prediction estimates of the SeLECT score are displayed in figure 3 and in a prediction chart in figure 4. The lowest SeLECT value (0 points) predicts a 0·7% risk of late seizures within the first year and a 1·2% risk within five years after stroke. In contrast, the highest SeLECT value (9 points) indicates a 63% risk of late seizures within one year and an 83% risk within five years after stroke. In addition to these two time-points, the SeLECT model offers flexible predictions at any chosen time after stroke. An exemplary estimation with further explanation and a cut-off analysis can be found in the appendix. To facilitate SeLECT bedside estimations and prediction we developed a practical smartphone and tablet app ('SeLECT score'), available for iOS and Android.

Discussion

We developed a novel practical prognostic instrument to predict the risk of late seizures after stroke. The SeLECT score was successfully externally validated and showed good discrimination and calibration. The model incorporates five items: severity of stroke, large-artery atherosclerotic aetiology, early seizures, cortical involvement, territory of MCA involvement. The rationale for these factors is discussed in the Appendix.

Development and validation of the SeLECT score followed established recommendations.[30] We carefully selected a list of candidate predictors. Such a process involves making

compromises, such as the exclusion of parameters which are not routinely assessed in a clinical setting or which do not support sufficient validation data. Future research might refine SeLECT predictions by including lesion size, biomarkers, EEG findings, psychiatric co-morbidities, genetic data, and advanced neuroimaging.

This study has several strengths. The assembled population is one of the largest cohorts of post-stroke seizures. Good performance of SeLECT in triple external validation and inclusion of a broad spectrum of people with ischemic strokes support the wide use of this instrument in diverse clinical settings and populations. SeLECT is a practical instrument that can be used at the bedside. The predictors are well-defined, easily measured and routinely available. All individual cohorts were adequately powered to demonstrate a good discrimination of the SeLECT model. This is indicated by the 95% confidence intervals of concordance statistics, that exceeded 0.60 in all cohorts.

The results are only applicable to ischemic strokes. We did not include people with primary haemorrhage as previous research suggests different mechanism of epileptogenesis in these individuals.[26,27] An alternative model should be used for primarily haemorrhagic cases.[31]

This study has several limitations. (1) There were missing data but this was limited in scope and managed with established imputation techniques.[30] Sensitivity analysis did not detect discrepancies.

(2) Selection bias needs to be considered as severe strokes were more likely to have died or been lost to follow-up. The potential for selection bias was mitigated by retaining these cases in the survival analysis and censoring them after death or last follow-up. Due to the retrospective approach in the Austrian cohort, ascertainment bias is a concern. A referral bias is unlikely as the model performed well in the population based Italian cohort.

(3) There was a variability of follow-up duration between and within cohorts. These differences were handled with established censoring methods in survival analysis and an impact on the results was reduced to a minimum.

(4) Different imaging modalities (MRI or CT) were used to determine stroke location. This might have increased data variability and reduced the reliability to detect cortical involvement in those with CT scans, but such an approach mimics a real-life situation, making the model applicable in clinical practice.

(5) At times there was deviation from the perfect slope in calibration plots (figure 2). These deviations were limited in scope and within the estimated 95% confidence interval.

(6) We did not collect data on antiepileptic drugs used for indications other than epilepsy (i.e. neuropathic pain or psychiatric conditions) in the Swiss cohort and the influence of post-stroke treatment was not considered in model development. Most participants were, however, unlikely to have started antiepileptic treatment before they had seizures and the overall impact on the results would be minimal.

(7) There were differences in baseline characteristics and observed frequency of seizures in the cohorts. This variance could be attributed to differences in study design (cohort vs. case control study, prospective vs. retrospective), inclusion and exclusion criteria, type of follow-up (face-to-face vs. telephonic), and setting (tertiary centre vs. population-based). SeLECT performed well in all populations and this demonstrates the robustness and reliability of the model when applied to different populations.

(8) Telephonic follow-up is not the gold-standard to diagnose seizures. We combined telephone screening based on a validated questionnaire[14] and subsequent face-to-face evaluation by a neurologist in the Swiss derivation cohort. We believe that this approach reduces loss to follow-up[32] while retaining diagnostic accuracy. The model was subsequently validated in cohorts using both face-to-face (Austria, Italy) and telephonic (Germany) follow-up demonstrating robustness irrespective of follow-up method.

(9) Recent advances in acute stroke treatment are difficult to consider in longitudinal studies requiring long follow-up. The SeLECT model demonstrated good discrimination in subjects receiving intravenous thrombolysis or endovascular thrombectomy in the contemporary

(recruitment until 2014) validation cohort from Austria. More data from novel stroke cohorts might, however, be required to address fully the impact of new treatments on post-stroke epilepsy.

The SeLECT score has several practical applications. Firstly, it is a step towards a much needed biomarker of post-stroke epilepsy that may optimise recruitment for antiepileptogenic treatment trials.[12] Targeted selection of people at high risk of seizures is an efficient and cost-effective enrolment strategy. For example, a treatment trial to reduce the risk of late seizures by 50% overall would require a sample size of about 1500 participants (cumulative seizure incidence 6% within 2 years, power 80%, $p < 0{\cdot}05$). SeLECT could be used to enrol an enriched population with a seizure risk above 20%, reducing the required sample size to less than 400 people. Minimising financial burden and reducing exposure of participants could make such a study possible.

Secondly, SeLECT may be used to stratify participants in clinical trials according to seizure risk. This might aid group comparisons in non-randomized studies.

Third, SeLECT is a first step towards more personalised medicine as it could identify individuals at high risk of late seizures and post-stroke epilepsy. There is currently, however, no prophylactic treatment available for these individuals.[9] Now that identification of high risk cases with the SeLECT score is feasible, their optimal medical management needs to be the focus of future studies and expert panels. Several questions need addressing: (i) What are appropriate measures to reduce seizure-provoking factors, including sleep deprivation, intercurrent illness or depression, in very high risk subjects (e.g. risk above 60% within five years, i.e. SeLECT $\geq$ 8 points)? Are these measures beneficial? (ii) Should people with high risk of late seizures (e.g. risk above 20% within five years, i.e. SeLECT $\geq$ 6 points) receive regular, at least yearly, follow-ups with a neurologist? Should these evaluations include an EEG; are such follow-ups cost-effective? (iii) Does a very high risk of late seizures have an impact on the risk of motor vehicle accidents and the ability to drive? (iv) Antiepileptic

treatment is usually indicated after a first spontaneous late seizure after stroke.[7] Is there a benefit from initiating prophylactic treatment in high risk cases even before a first late seizure? (v) Some drugs, in particular enzyme-inducers and benzodiazepines, might have detrimental effects on functional recovery after stroke.[33] The choice of an optimal potentially prophylactic antiepileptic needs to be determined in future studies and the SeLECT score could be used to control for case mix variation in such a study.

In the future, if antiepileptogenic treatments become available, it may be appropriate that only those at high risk receive them. SeLECT might be a useful instrument to identify those who are most likely to benefit from antiepileptogenic interventions.

Research in context

Evidence before this study

We searched PubMed for articles published before April 1, 2016. Search terms were ("Stroke" OR "Brain Ischemia" OR "Intracranial Embolism and Thrombosis"[MeSH Terms]) AND ("Epilepsy" OR "Seizures"[MeSH Terms]). Reference lists of relevant articles, reviews and meta-analyses were also searched for additional sources. Previous studies have most consistently identified stroke severity and location, cortical involvement, early seizures, stroke aetiology, young age, and lesion size as possible indicators of late seizures after stroke. No validated instrument exists to synthetize these single variables into an individual prediction of seizure risk. One non-validated study explored the possibility to prognosticate seizure recurrences after stroke. The small sample size (n = 264 overall; n = 10 with recurrent seizures), absence of multivariate modelling, failure to report established performance parameters, and lack of validation hindered the translation of this model into clinical practice.

Added value of this study

This study used a large derivation cohort and three independent external validation cohorts to develop and validate a prognostic instrument of late seizures after stroke. The SeLECT score can generate individualised estimates for the risk of late seizures within the first years after stroke. The main strengths of this model are (i) the use of routinely available predictors that can be calculated at the bedside, (ii) generalisability to a broad spectrum of people with ischemic stroke, and (iii) successful validation in different countries and clinical settings. A free app for smartphones and tablets (“SeLECT score”) has been developed to make this model widely available and easily applicable in clinical practice.

Implications of all the available evidence

We present a method to prognosticate late seizures after stroke that has several practical applications. It can inform people who had a stroke and relatives about the risk of seizures and has the potential to guide clinical responsiveness in future. From the perspective of stroke and epilepsy researchers, such prognostic information will be valuable to target enriched populations for recruitment in prospective studies and to control for between-group differences in nonrandomized trials. Should antiepileptogenic treatments become available in future, such a prognostic model could be used to select individuals at risk who would benefit from these procedures.

Contributors

Data for the study was collected in Switzerland (MG, ND, BE, AF, PS, BT), Germany (JC, SE), Austria (MW, TJO, HPH), and Italy (AS, GG, MV, FJ, GLG). These authors were involved in the design and/or supervision of local sub-studies. Literature search was done by ND and MG. Statistical analyses in the combined cohort were performed at UCL and Kantonsspital St. Gallen (MG, ND) with input from MKo, MKe, JWS, and JSD. All authors were involved in data interpretation and revising the manuscript for intellectual content. The first draft of the report was written by the lead author (MG).

Declaration of interests

Marian Galovic has nothing to disclose.

Nico Döhler has nothing to disclose.

Barbara Erdélyi has nothing to disclose.

Ansgar Felbecker has nothing to disclose.

Philip Siebel has nothing to disclose.

Julian Conrad has nothing to disclose.

Stefan Evers received honoraria for consulting and for lectures in the last 5 years from the companies Allergan, Mundipharma, Pfizer, and Reckitt Benckiser.

Michael Winklehner has nothing to disclose.

Tim J von Oertzen reports personal fees and non-financial support from Eisai Pharma GmbH Vienna, grants, personal fees and non-financial support from UCB Pharma GmbH Vienna, non-financial support from Medtronic Österreich GmbH, grants, personal fees and non-financial support from Novartis Phama, personal fees from Roche Pharma, personal fees from Biogen Idec Austria, personal fees from Liva Nova, personal fees from Sanofi-Aventis GmbH, grants from Grossegger & Drbal GmbH, outside the submitted work; and is vize-

president of the Österreichische Gesellschaft für Epileptologie (Austrian ILAE chapter) and webeditor in chief of the European Academy of Neurology (EAN).

Hans-Peter Haring has nothing to disclose.

Anna Serafini has nothing to disclose.

Giorgia Gregoraci has nothing to disclose.

Mariarosaria Valente has nothing to disclose.

Francesco Janes has nothing to disclose.

Gian Luigi Gigli has nothing to disclose.

Mark R Keezer reports grants and personal fees from UCB, personal fees from Sage Therapeutics as well as from Novartis, outside of the submitted work.

John S Duncan is supported by UK NIHR UCLH/UCL BRC as a Senior Investigator.

Josemir W Sander reports personal fees from Eisai, grants from UCB, grants and personal fees from Eisai, grants and personal fees from UCB, grants from GSK, grants from Eisai, grants from WHO, grants from NEF, personal fees from Lundbeck, personal fees from Teva, grants and personal fees from Eisai, grants and personal fees from UCB, outside the submitted work; and Prof Sander's current position is endowed by the Epilepsy Society, he is a member of the Editorial Board of the Lancet Neurology, and receives research support from the Marvin Weil Epilepsy Research Fund.

Matthias J Koepp has nothing to disclose.

Barbara Tettenborn has nothing to disclose.

Acknowledgements

We thank Friederike Theresia Wagner and Felix Peter for their work in implementing and maintaining the post-stroke epilepsy registry in St. Gallen, Switzerland. We thank Dominik Zieglgänsberger for critical input. We thank Robert Sasák for designing and programming the SeLECT-score app.

This work was partly done at UCLH/UCL Comprehensive Biomedical Research Centre, which receives a proportion of funding from the UK Department of Health's National Institute for Health Research Biomedical Research Centres funding scheme. JWS receives research support from the Dr Marvin Weil Epilepsy Research Fund.

Role of the funding source

The study was not funded by an external agency.

References

1 Seshadri S, Wolf PA. Lifetime risk of stroke and dementia: current concepts, and estimates from the Framingham Study. *The Lancet Neurology* 2007; **6**: 1106–14.

2 Hauser WA, Annegers JF, Kurland LT. Incidence of Epilepsy and Unprovoked Seizures in Rochester, Minnesota: 1935–1984. *Epilepsia* 1993; **34**: 453–8.

3 Huang C-W, Saposnik G, Fang J, Steven DA, Burneo JG. Influence of seizures on stroke outcomes: a large multicenter study. *Neurology* 2014; **82**: 768–76.

4 Kim JS, Choi-Kwon S, Kwon SU, Lee HJ, Park K-A, Seo YS. Factors affecting the quality of life after ischemic stroke: young versus old patients. *J Clin Neurol* 2005; **1**: 59–68.

5 Guekht A, Mizinova M, Ershov A, *et al*. In-hospital costs in patients with seizures and epilepsy after stroke. *Epilepsia* 2015; **56**: 1309–13.

6 Beghi E, Carpio A, Forsgren L, *et al*. Recommendation for a definition of acute symptomatic seizure. *Epilepsia* 2010; **51**: 671–5.

7 Fisher RS, Acevedo C, Arzimanoglou A, *et al*. ILAE official report: a practical clinical definition of epilepsy. *Epilepsia* 2014; **55**: 475–82.

8 Pitkänen A, Lukasiuk K. Mechanisms of epileptogenesis and potential treatment targets. *The Lancet Neurology* 2011; **10**: 173–86.

9 Trinka E, Brigo F. Antiepileptogenesis in humans: disappointing clinical evidence and ways to move forward. *Curr Opin Neurol* 2014; **27**: 227–35.

10 Serafini A, Gigli GL, Gregoraci G, *et al*. Are Early Seizures Predictive of Epilepsy after a Stroke? Results of a Population-Based Study. *Neuroepidemiology* 2015; **45**: 50–8.

11 Graham NSN, Crichton S, Koutroumanidis M, Wolfe CDA, Rudd AG. Incidence and associations of poststroke epilepsy: the prospective South London Stroke Register. *Stroke* 2013; **44**: 605–11.

12 Jozwiak S, Becker A, Cepeda C, *et al*. WONOEP appraisal: Development of epilepsy biomarkers-What we can learn from our patients? *Epilepsia* 2017; **58**: 951–61.

13 Strzelczyk A, Haag A, Raupach H, Herrendorf G, Hamer HM, Rosenow F. Prospective evaluation of a post-stroke epilepsy risk scale. *J Neurol* 2010; **257**: 1322–6.

14 Placencia M, Sander JW, Shorvon SD, Ellison RH, Cascante SM. Validation of a screening questionnaire for the detection of epileptic seizures in epidemiological studies. *Brain* 1992; **115 (Pt 3)**: 783–94.

15 Conrad J, Pawlowski M, Dogan M, Kovac S, Ritter MA, Evers S. Seizures after cerebrovascular events: Risk factors and clinical features. *Seizure* 2013; **22**: 275–82.

16 Stockinger MS, Trenkler J, Oertzen von T, *et al*. Effective cerebrovascular thrombectomy requires well-organized structures : Real world experiences of a regional stroke network between 2012-2015. *Wien Klin Wochenschr* 2017; **129**: 96–101.

17 Fischer U, Arnold M, Nedeltchev K, *et al*. NIHSS score and arteriographic findings in acute ischemic stroke. *Stroke* 2005; **36**: 2121–5.

18 Fischer U, Baumgartner A, Arnold M, *et al*. What is a minor stroke? *Stroke* 2010; **41**: 661–6.

19 Williams LS, Yilmaz EY, Lopez-Yunez AM. Retrospective assessment of initial stroke severity with the NIH Stroke Scale. *Stroke* 2000; **31**: 858–62.

20 Tatu L, Moulin T, Bogousslavsky J, Duvernoy H. Arterial territories of the human brain: cerebral hemispheres. *Neurology* 1998; **50**: 1699–708.

21 Wahlund LO, Barkhof F, Fazekas F, *et al*. A new rating scale for age-related white matter changes applicable to MRI and CT. *Stroke* 2001; **32**: 1318–22.

22 Belcastro V, Vidale S, Gorgone G, *et al*. Non-convulsive status epilepticus after ischemic stroke: a hospital-based stroke cohort study. *J Neurol* 2014; **261**: 2136–42.

23 Bladin CF, Alexandrov AV, Bellavance A, *et al*. Seizures after stroke: a prospective multicenter study. *Arch Neurol* 2000; **57**: 1617–22.

24 Kammersgaard LP, Olsen TS. Poststroke Epilepsy in the Copenhagen Stroke Study: Incidence and Predictors. *J Stroke and Cerebrovasc Dis* 2005; **14**: 210–4.

25 Wang G, Jia H, Chen C, *et al*. Analysis of risk factors for first seizure after stroke in Chinese patients. *Biomed Res Int* 2013; **2013**: 702871.

26 Ferlazzo E, Gasparini S, Beghi E, *et al*. Epilepsy in cerebrovascular diseases: Review of experimental and clinical data with meta-analysis of risk factors. *Epilepsia* 2016; **57**: 1205–14.

27 Benbir G, Ince B, Bozluolcay M. The epidemiology of post-stroke epilepsy according to stroke subtypes. *Acta Neurol Scand* 2006; **114**: 8–12.

28 Bentes C, Martins H, Peralta AR, *et al*. Epileptic manifestations in stroke patients treated with intravenous alteplase. *European Journal of Neurology* 2017; **24**: 755–61.

29 Bozdogan H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 1987; **52**: 345–70.

30 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; **162**: 55–63.

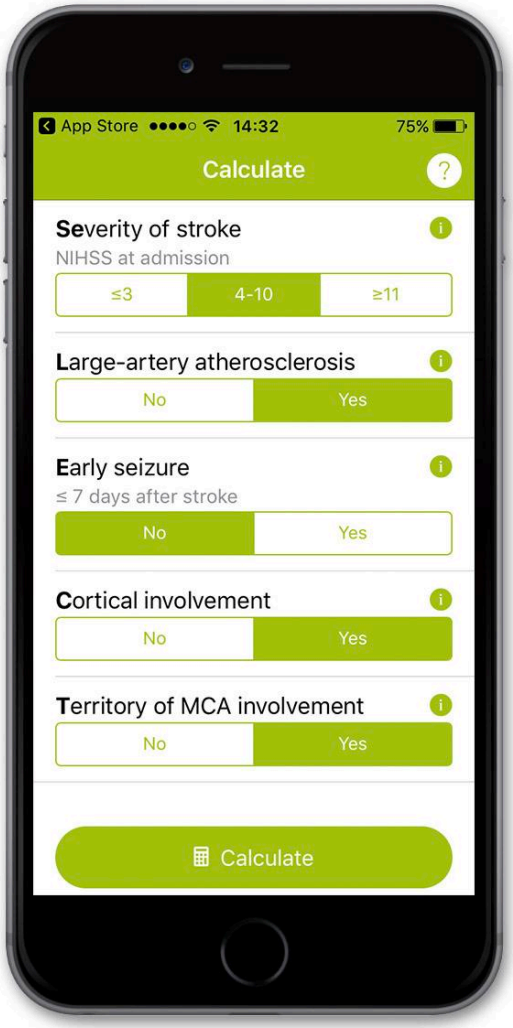
31 Haapaniemi E, Strbian D, Rossi C, *et al*. The CAVE score for predicting late seizures after intracerebral hemorrhage. *Stroke* 2014; **45**: 1971–6.

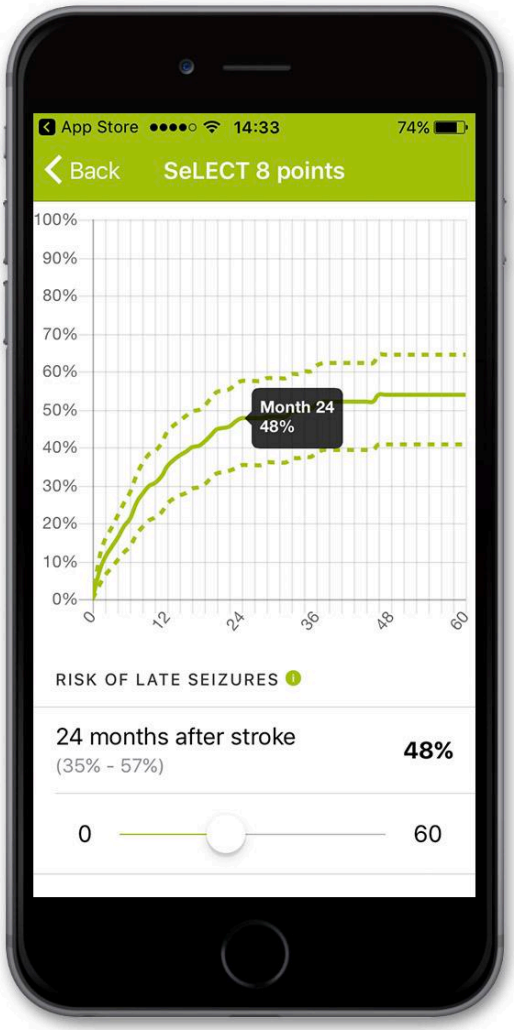
32 Pendlebury ST, Chen P-J, Welch SJV, *et al*. Methodological Factors in Determining Risk of Dementia After Transient Ischemic Attack and Stroke: (II) Effect of Attrition on Follow-Up. *Stroke* 2015; **46**: 1494–500.

33 Goldstein LB. Common drugs may influence motor recovery after stroke. The Sygen In Acute Stroke Study Investigators. *Neurology* 1995; **45**: 865–71.

Free smartphone and tablet app

A free smartphone and tablet app called ‘SeLECT score’ is available to facilitate bedside estimations and prediction. It is available on
Apple iOS: https://itunes.apple.com/us/app/select-score/id1241429202
Android: https://play.google.com/store/apps/details?id=sk.sasak.select





Panels

Panel 1: *Calculation of the SeLECT score*.

SeLECT Score:	
(Se) Severity of stroke	
NIHSS ≤ 3	0 points
NIHSS 4 to 10	1 point
NIHSS ≥ 11	2 points
(L) Large-artery atherosclerosis	
No	0 points
Yes	1 point
(E) Early seizure (≤ 7 days)	
No	0 points
Yes	3 points
(C) Cortical involvement	
No	0 points
Yes	2 points
(T) Territory of MCA	
No	0 points
Yes	1 point

To calculate an individual's SeLECT score, the points associated with each predictor can be added to obtain the total risk score. As an example, a person suffering a stroke with initially 12 points on NIHSS due to large-artery atherosclerosis, no early seizures, and with infarction involving the cortex in the MCA territory will have a risk score of 2+1+0+2+1 = 6 points. According to Figure 3 this corresponds to a late seizure risk of 18% within one year and of 29% within five years after stroke. NIHSS = National Institutes of Health Stroke Scale, MCA = middle cerebral artery.

Tables

Table 1: *Demographic and clinical characteristics*.

	Switzerland (n = 1200)	Austria (n = 459)	Germany (n = 311)	Italy (n = 399)
Age	71 (61 - 78)	77 (68 - 86)	68 (55 - 76)	78 (70 - 85)
Sex				
Male	699 (58%)	207 (45%)	166 (53%)	198 (50%)
Female	501 (42%)	252 (55%)	145 (47%)	201 (50%)
NIHSS at admission				
≤ 3	727 (61%)	170 (37%)	139 (45%)	183 (46%)
4 – 10	312 (26%)	190 (41%)	119 (38%)	124 (31%)
≥ 11	161 (13%)	99 (22%)	53 (17%)	92 (23%)
Stroke aetiology				
Small-vessel occlusion	400 (32%)	134 (29%)	15 (5%)	78 (20%)
Large-artery atherosclerosis	198 (17%)	121 (26%)	64 (21%)	55 (14%)
Cardioembolism	369 (31%)	155 (34%)	92 (30%)	87 (22%)
Other determined aetiology	19 (2%)	49 (11%)	19 (6%)	11 (3%)
Undetermined aetiology	214 (18%)	0 (0%)	121 (39%)	168 (42%)
Thrombolysis				
IV-thrombolysis	139 (12%)	102 (22%)	73 (23%)	11 (3%)
IA-thrombolysis	--	28 (6%)	unknown	--
Post-stroke seizures				
Early seizures	37 (3%)	3 (1%)	12 (4%)	11 (3%)
AED treatment after early seizure	17/33 (52%)	2/3 (67%)	11/12 (92%)	3/11 (27%)
Late seizures during follow-up	71 (6%)	30 (9%)	23 (7%)	13 (3%)
Late seizures after 1 year*	4% (3% - 5%)	6% (3% - 8%)	7% (4% - 10%)	3% (1% - 5%)
Late seizures after 5 years*	7% (5% - 9%)	13% (8% - 17%)	8% (5% - 11%)	--
Type of first late seizure				
Focal aware	10/71 (14%)	5/30 (17%)	5/23 (22%)	4/13 (31%)
Focal impaired awareness	20/71 (28%)	5/30 (17%)	7/23 (30%)	2/13 (15%)
Focal to bilateral tonic-clonic	31/71 (44%)	18/30 (60%)	7/23 (30%)	3/13 (23%)
Status epilepticus	7/71 (10%)	2/30 (7%)	4/23 (17%)	0/13 (0%)
Not classifiable	3/71 (4%)	0/30 (0%)	0/23 (0%)	4/13 (31%)
Recurrent late seizures during follow-up**	43 (4%)	14 (3%)	16 (5%)	10 (3%)
Study information				
Recruitment period	2002-08	2005-14	2003-10	2007-09
Duration of follow-up *(months)*	28 (21 – 47)	10 (2 – 41)	23 (12 – 44)	24 (7 – 24)
SeLECT score	3 (1 – 4)	3 (1 – 4)	2 (1 – 3)	2 (1 – 3)

Data displayed as N (%) or median (interquartile range). * Kaplan Meier estimates of cumulative absolute risk of late seizures after 1 year or 5 years are given with 95% confidence interval in brackets. Due to a maximum follow-up of 24 months in the Italian cohort, a 5-year estimate was not calculated for the Italian cohort. NIHSS = National Institutes of Health Stroke Scale, IV = intravenous, IA = intraarterial, AED = antiepileptic drug. ** The incidence of recurrent seizures is influenced by differences in antiepileptic treatment protocols after a first late seizure and duration of follow-up in individual cohorts, hence, these numbers must be interpreted with caution.

Table 2: *Univariable analysis of predictors associated with time to first late seizure in the derivation cohort (n = 1200).*

Variable	N (%) or median (IQR)	HR (95% CI)	P value
Age *(per 10 years)*	71 (61 – 78)	0·9 (0·7 – 1·1)	0·16
Sex			
Male	699 (58%)	0·8 (0·5 – 1·3)	0·31
Female	501 (42%)	1·3 (0·8 – 2·0)	0·31
Stroke risk factors			
Hypertension	908 (76%)	1·3 (0·7 – 2·3)	0·39
Diabetes mellitus	205 (17%)	1·3 (0·7 – 2·2)	0·41
Smoking	497 (41%)	0·8 (0·5 – 1·3)	0·40
Dyslipidaemia	764 (64%)	1·1 (0·6 – 1·7)	0·83
Stroke severity at admission			
NIHSS ≤ 3	727 (61%)	Reference category	--
NIHSS 4 to 10	312 (26%)	2·1 (1·2 – 3·7)	**0·01**
NIHSS ≥ 11	161 (13%)	4·7 (2·7 – 8·2)	**<0·0001**
Stroke aetiology			
Small-vessel occlusion	400 (32%)	0·3 (0·2 – 0·6)	**0·0005**
Large-artery atherosclerosis	198 (17%)	2·5 (1·5 – 4·1)	**0·0003**
Cardioembolism	369 (31%)	1·0 (0·6 – 1·6)	0·97
Other determined origin	19 (2%)	1·0 (0·1 – 7·0)	0·98
Undetermined aetiology	214 (18%)	1·3 (0·8 – 2·3)	0·31
Stroke laterality			
Left	571 (48%)	1·1 (0·7 – 1·8)	0·58
Right	572 (48%)	1·7 (1·0 – 2·7)	**0·03**
Stroke location			
Cortical involvement	665 (55%)	6·7 (3·2 – 14·0)	**<0·0001**
ACA territory involvement	44 (4%)	1·8 (0·6 – 4·8)	0·28
MCA territory involvement	857 (71%)	3·7 (1·7 – 8·1)	**0·001**
PCA territory involvement	120 (10%)	1·1 (0·5 – 2·4)	0·75
Other imaging findings			
Secondary haemorrhage	60 (5%)	1·8 (0·8 – 4·2)	0·15
White matter hyperintensities	685 (57%)	0·6 (0·4 – 0·9)	**0·02**
Early seizure	37 (3%)	6·8 (3·5 – 12·9)	**<0·0001**
Poststroke treatment			
Thrombolysis	139 (12%)	2·1 (1·2 – 3·8)	**0·009**
Antiplatelet therapy	767 (64%)	0·7 (0·4 – 1·2)	0·17
Anticoagulation therapy	488 (41%)	1·3 (0·8 – 2·0)	0·32

NIHSS = National Institutes of Health Stroke Scale, ACA = anterior cerebral artery, MCA = middle cerebral artery, PCA = posterior cerebral artery, HR = hazards ratio, CI = confidence interval.

Table 3: *Final multivariable Cox proportional hazards model of time to first late seizure in the derivation cohort (n = 1200).*

Variable	aHR (95% CI)	P value	ΔAIC
Cortical involvement	4·2 (1·9 – 9·0)	0·0003	-15·7
Early seizure	4·8 (2·5 – 9·3)	<0·0001	-13·7
Stroke severity at admission			
NIHSS ≤ 3	Reference category	--	--
NIHSS 4 to 10	1·7 (1·0 – 3·1)	0·06	-3·4
NIHSS ≥ 11	2·7 (1·5 – 4·9)	0·0008	-9·3
Territory of MCA involvement	1·8 (0·8 – 4·1)	0·12	-0·3
Large-artery atherosclerosis	1·5 (0·9 – 2·5)	0·15	-0·3
Stroke laterality: right	Eliminated in step 5 (p = 0·24, ΔAIC = 0·7)		
Age	Eliminated in step 4 (p = 0·35, ΔAIC = 1·2)		
Small-vessel occlusion	Eliminated in step 3 (p = 0·65, ΔAIC = 1·8)		
White matter hyperintensities	Eliminated in step 2 (p = 0·86, ΔAIC = 2·0)		
Thrombolysis	Eliminated in step 1 (p = 0·88, ΔAIC = 2·0)		

NIHSS = National Institutes of Health Stroke Scale, MCA = middle cerebral artery, aHR = adjusted hazards ratio, CI = confidence interval, ΔAIC = Change in Akaike information criterion – a negative value implies that the variable improves the model and should not be eliminated.

Figure 1: *Kaplan Meier estimates of time to first late seizure in the overall study population (n = 2369).*

(A) Plot of time to first late seizure, 95% confidence intervals are shaded in grey. (B-F) Kaplan Meier estimates for individual predictors in the SeLECT model. Separate lines are displayed for each cut-off used in SeLECT. NIHSS = National Institutes of Health Stroke Scale, MCA = middle cerebral artery.

Figure 2: *Calibration plots for predicting late seizures within one year (A-D) and within five years (E-H) after stroke in external validation cohorts*.

A logarithmic scale was used to improve the presentation of low probabilities. Diagonal dotted line indicates perfect calibration. Vertical bars show 95% confidence intervals.

*Follow-up in the Italian cohort was terminated after two years, hence a calibration plot five years after stroke could not be calculated. Instead, we present a calibration chart two years after stroke for the Italian cohort.

Figure 3: *Predicted risk of late seizures according to SeLECT score*.

The prediction plot **(A)** displays the predicted risk of late seizures 0 to 72 months after stroke. Each curve represents the estimates for a SeLECT value, ranging from 0 to 10. Lower panels display risk estimate charts of late seizures one year **(B)** and five years **(C)** after stroke according to SeLECT score values. Vertical lines are 95% confidence intervals (CI).

Figure 4: *Prediction chart of late seizures after stroke*.

Numbers in the prediction chart correspond to the risk of late seizures one year after stroke, numbers in brackets are risks five years after stroke.

MCA = middle cerebral artery, NIHSS = National Institutes of Health Stroke Scale.

Figure 1

Click here to download Figure: SeLECT_Figure_1.pdf

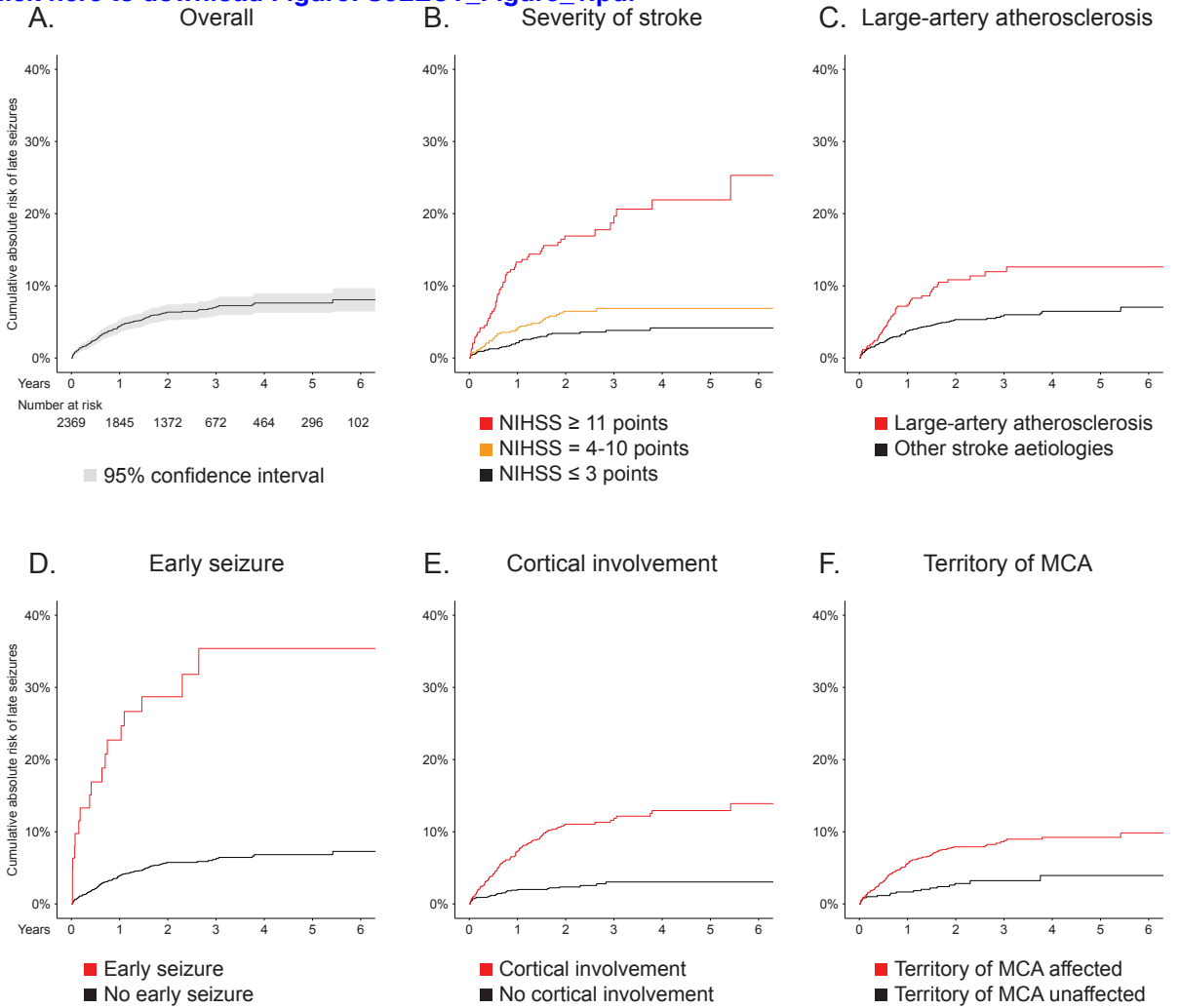


Figure 2

Click here to download Figure: SeLECT_Figure2.pdf

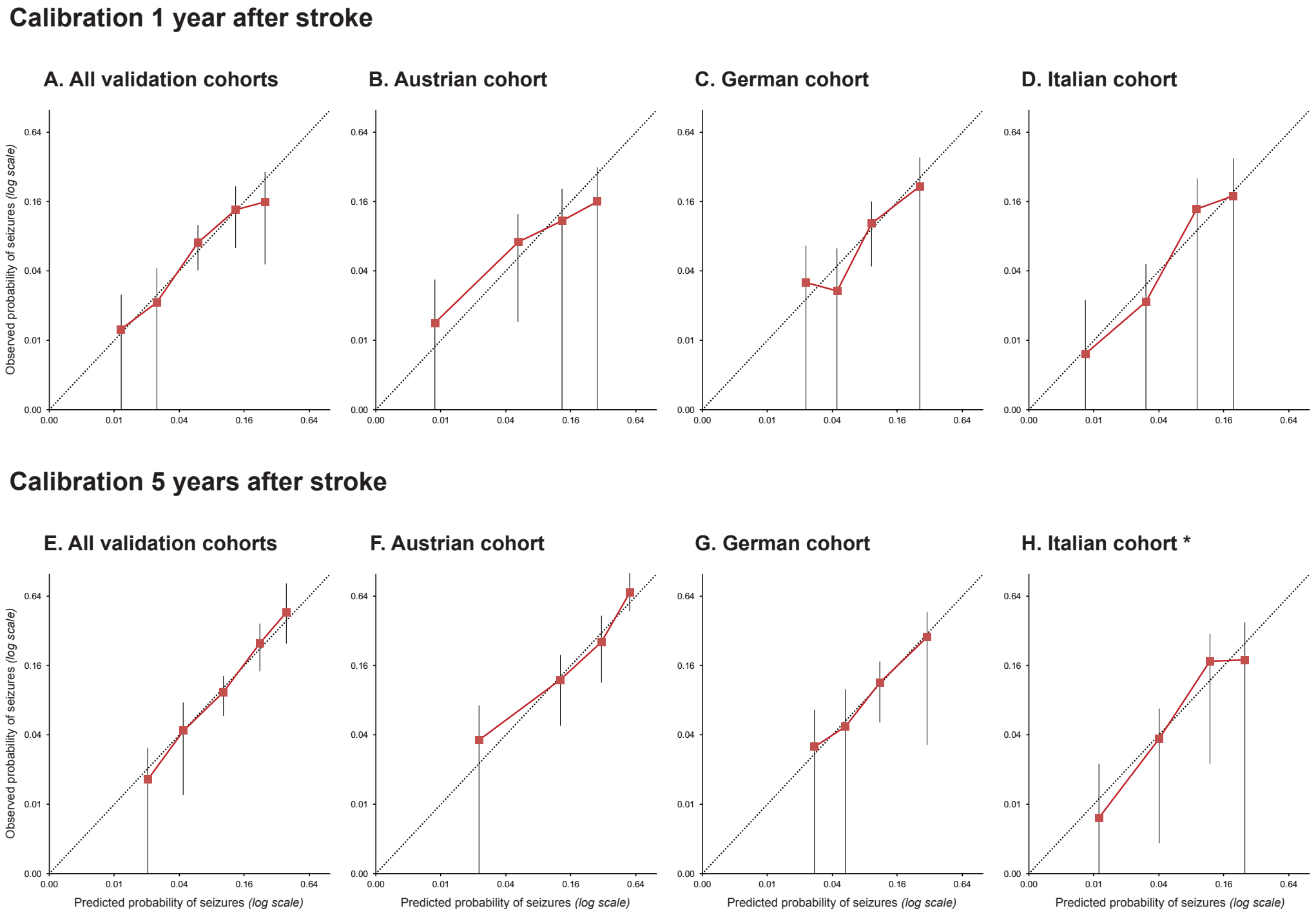
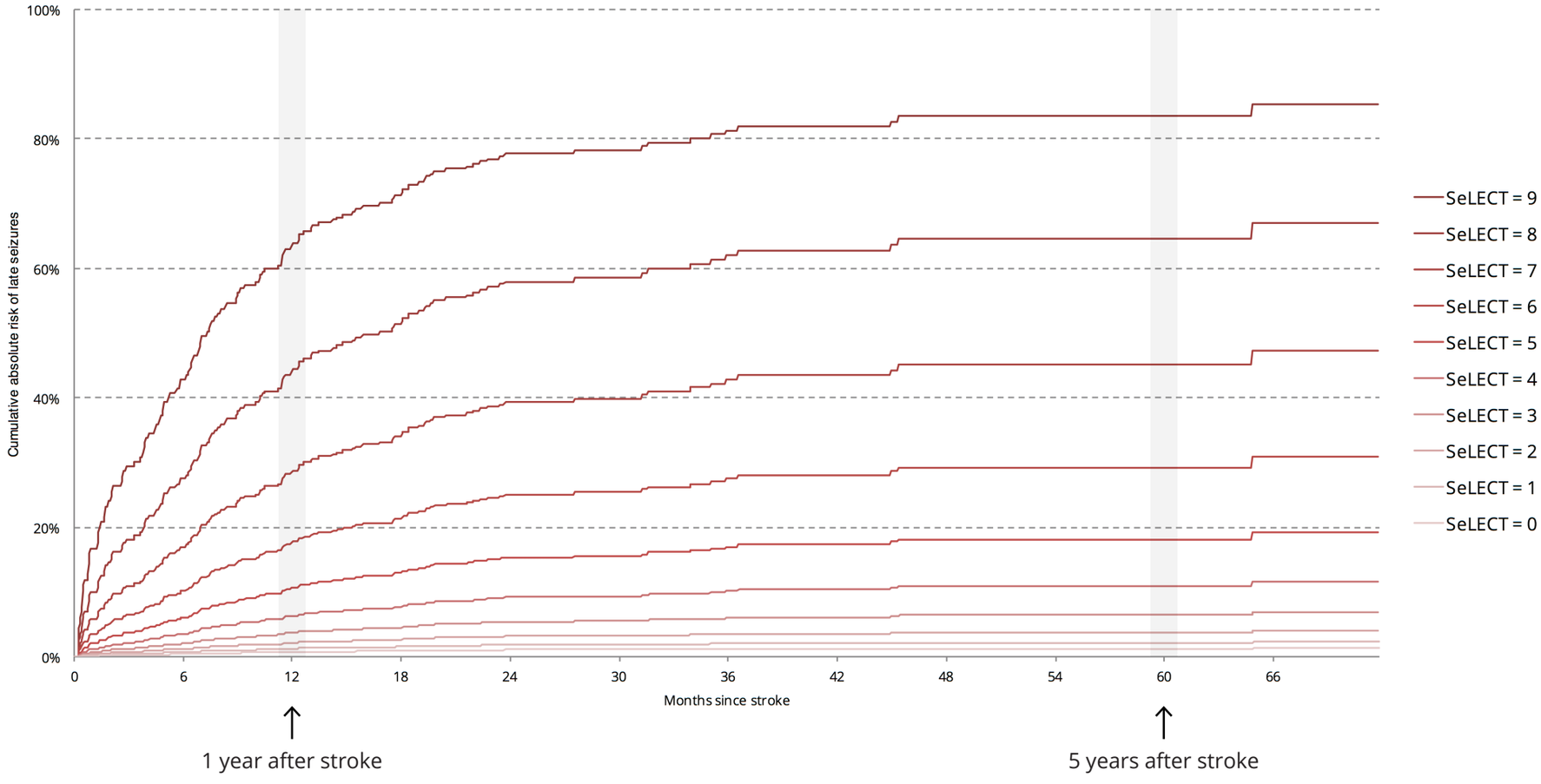


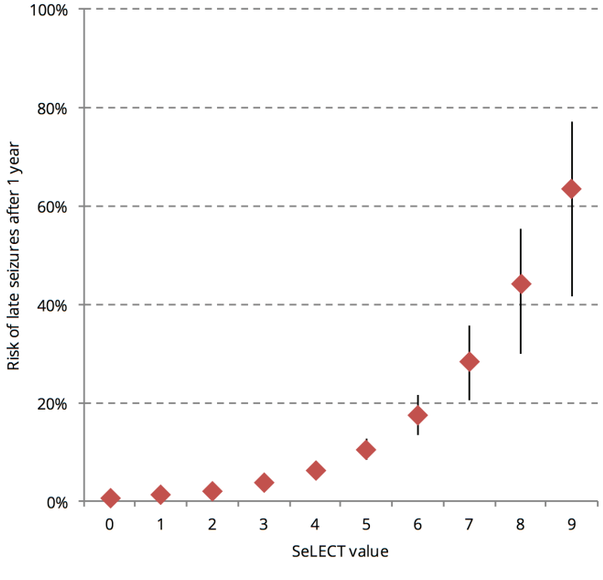
Figure 3

Click here to download Figure: SeLECT_Figure3.pdf

A. Prediction plot

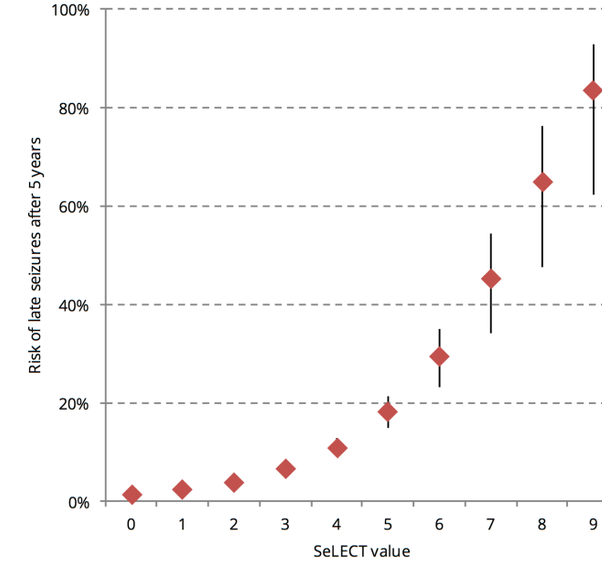


B. Risk estimate chart 1 year after stroke



SeLECT value	Risk estimate (95% CI)
0	0.7% (0.4%-1%)
1	1% (1%-2%)
2	2% (2%-3%)
3	4% (3%-4%)
4	6% (5%-7%)
5	11% (8%-13%)
6	18% (13%-22%)
7	28% (20%-36%)
8	44% (30%-55%)
9	63% (42%-77%)

C. Risk estimate chart 5 years after stroke



SeLECT value	Risk estimate (95%CI)
0	1% (1%-2%)
1	2% (1%-3%)
2	4% (3%-5%)
3	6% (5%-8%)
4	11% (9%-13%)
5	18% (15%-21%)
6	29% (23%-35%)
7	45% (34%-54%)
8	65% (48%-76%)
9	83% (62%-93%)

Figure 4

Click here to download Figure: SeLECT_Figure4.pdf

No early seizures

	No cortical involvement				Cortical involvement			
Territory of MCA	No		Yes		No		Yes	
Large-artery atherosclerosis	No	Yes	No	Yes	No	Yes	No	Yes
NIHSS 0 to 3	0.7% (1%)	1% (2%)	1% (2%)	2% (4%)	2% (4%)	4% (6%)	4% (6%)	6% (11%)
NIHSS 4 to 10	1% (2%)	2% (4%)	2% (4%)	4% (6%)	4% (6%)	6% (11%)	6% (11%)	11% (18%)
NIHSS ≥ 11	2% (4%)	4% (6%)	4% (6%)	6% (11%)	6% (11%)	11% (18%)	11% (18%)	18% (29%)

Very low risk
Low risk
Moderate risk
High risk
Very high risk

Early seizures

	No cortical involvement				Cortical involvement			
Territory of MCA	No		Yes		No		Yes	
Large-artery atherosclerosis	No	Yes	No	Yes	No	Yes	No	Yes
NIHSS 0 to 3	4% (6%)	6% (11%)	6% (11%)	11% (18%)	11% (18%)	18% (29%)	18% (29%)	28% (45%)
NIHSS 4 to 10	6% (11%)	11% (18%)	11% (18%)	18% (29%)	18% (29%)	28% (45%)	28% (45%)	44% (65%)
NIHSS ≥ 11	11% (18%)	18% (29%)	18% (29%)	28% (45%)	28% (45%)	44% (65%)	44% (65%)	63% (83%)

The SeLECT score: development and validation of a novel prognostic model to predict late seizures after ischemic stroke.

APPENDIX

The SeLECT score is a simple prognostic instrument for the prediction of late seizure risk after ischemic stroke. Although the calculation of SeLECT score values and prediction estimates is easy, some readers might not be familiar with the use of prediction plots and prediction charts. The following pages describe the practical calculation of the SeLECT score using an example subject.
We also present several cut-off, missing-value, and sensitivity analyses to complement the data in the main manuscript. Additional notes on clinical rationale for SeLECT factors and informed consent procedures are outlined to complement the methods and discussion sections in the main manuscript.

Table of Contents

1 Example calculations

1.1 Example subject

A 73-year-old male has been admitted to the hospital with acute ischemic stroke. He had an initial ***NIH Stroke Scale (NIHSS) of 12 points*** *and the infarction* ***involved the cortex*** *and the* ***territory of the middle cerebral artery (MCA)****. It was determined that his stroke was due to* ***large-artery atherosclerosis*** *(severe stenosis of the internal carotid artery). He* ***did not suffer from early seizures*** *after stroke.*

1.2 Classification of example subject

Severity of stroke NIHSS ≥ 11: Yes
Large-artery atherosclerosis: Yes
Early seizure: No
Cortical involvement: Yes
Territory of MCA affected: Yes

1.3 Calculation with prediction plot

Panel A1 can be used to compute the total SeLECT score value. For the example subject, the total SeLECT score is 6 points (2 points for both cortical involvement and severity of stroke, 1 point for both large-artery atherosclerosis and territory of MCA, 0 points for early seizures).

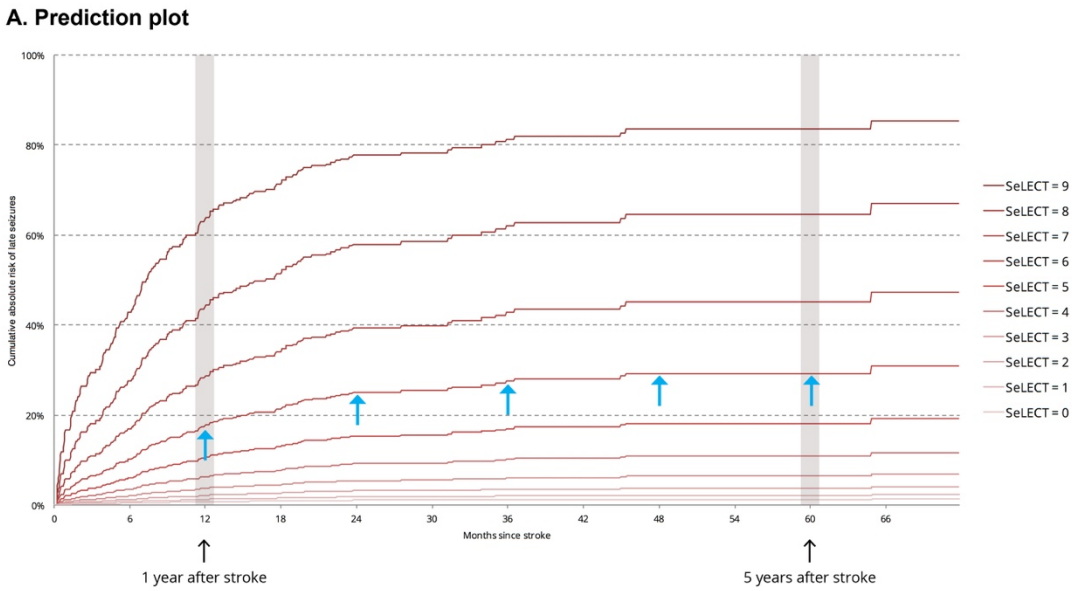
Panel A1: *Numerical calculation of SeLECT score values.*

SeLECT Score:

(Se) Severity of stroke	
NIHSS ≤ 3	0 points
NIHSS 4 to 10	1 point
NIHSS ≥ 11	2 points
(L) Large-artery atherosclerosis	
No	0 points
Yes	1 point
(E) Early seizure (≤ 7 days)	
No	0 points
Yes	3 points
(C) Cortical involvement	
No	0 points
Yes	2 points
(T) Territory of MCA	
No	0 points
Yes	1 point

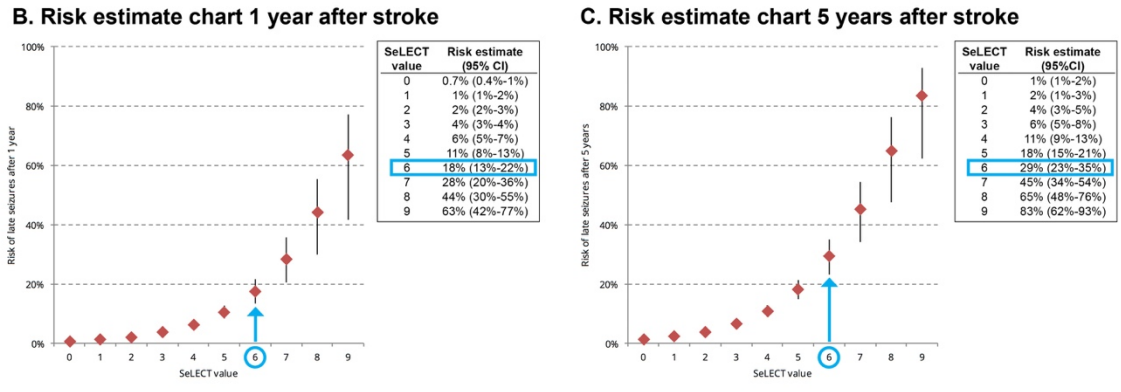
Next, locate the appropriate curve corresponding to the total SeLECT score value in the prediction plot (blue arrows in Figure A1). This curve represents the predicted proportion of cases who will suffer late seizures within a time-frame after stroke. If a clinician is interested in the risk of late seizures 2 years after stroke, the prediction curve indicates a risk of around 25% at 24 months post-stroke.

Figure A1: Example calculation with the prediction plot.



Alternatively, risk estimate charts can be used for two pre-specified time-points (1 year and 5 years after stroke). First, locate the total SeLECT score value on the X axis of the risk charts (blue arrows in Figure A2). The corresponding data points indicates the absolute risk of late seizures after 1 or 5 years. Alternatively, the risk predictions can be found in the tables (blue rectangles in Figure A2). For the example subject, this corresponds to a late seizure risk of 18% (95% CI 13% – 22%) within one year and of 29% (95% CI 23% – 35%) within five years after stroke.

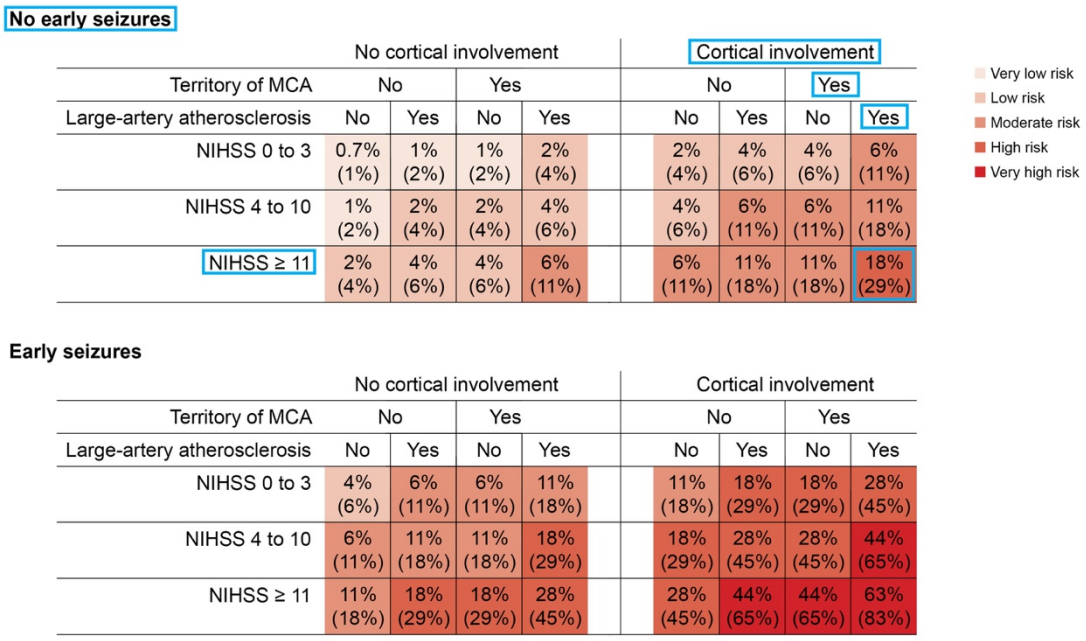
Figure A2: Example calculation with the risk estimate charts.



1.4 Calculation with prediction chart

To use the upper or lower part of the chart, first determine whether the subject did or did not suffer early seizures. Next, locate the subject's characteristics according to the four remaining items (blue rectangles in Figure A3). Locate the cell that matches the characteristics of all items.
The number in this cell corresponds to the late seizure risk one year after stroke, the number in brackets indicates late seizure risk five years after stroke. For the example subject, this corresponds to a risk of 18% within one year and of 29% within five years after stroke.

Figure A3: *Example calculation with prediction charts.*



2 Cut-off analysis

The sensitivity, specificity, positive and negative predictive values of different SeLECT cut-offs were estimated in the overall cohort using the timeROC package in R statistical software 3.3.3.

Blanche, P., Dartigues, J. F., & Jacqmin-Gadda, H. (2013). Estimating and comparing time- dependent areas under receiver operating characteristic curves for censored event times with competing risks. Statistics in medicine, 32(30), 5381-5397.

Table A1: *SeLECT cut-off analysis 1 year after stroke.*

SeLECT cut-off	Sensitivity	Specificity	Positive predictive value	Negative predictive value
≥ 1	96.8%	14.2%	4.9%	99.0%
≥ 2	93.4%	37.4%	6.3%	99.2%
≥ 3	88.0%	52.5%	7.8%	99.0%
≥ 4	74.8%	73.7%	11.4%	98.5%
≥ 5	46.2%	87.3%	14.2%	97.3%
≥ 6	17.6%	96.7%	19.6%	96.3%
≥ 7	5.6%	99.4%	29.7%	95.9%
≥ 8	4.5%	99.7%	38.2%	95.8%
≥ 9	3.4%	99.9%	74.1%	95.8%

Table A2: *SeLECT cut-off analysis 5 years after stroke.*

SeLECT cut-off	Sensitivity	Specificity	Positive predictive value	Negative predictive value
≥ 1	97.8%	13.8%	7.2%	98.9%
≥ 2	93.6%	36.8%	9.1%	98.8%
≥ 3	86.3%	51.9%	10.9%	98.2%
≥ 4	72.0%	73.8%	15.7%	97.5%
≥ 5	44.1%	87.9%	19.9%	95.9%
≥ 6	18.2%	96.7%	27.2%	94.6%
≥ 7	4.7%	99.4%	33.1%	93.9%
≥ 8	3.9%	99.8%	52.5%	93.9%
≥ 9	2.3%	99.9%	66.3%	93.8%

3 Early treatment with antiepileptic drugs

Treatment with antiepileptic drugs after early seizures was started in 17 (52%) out of 33 people with early seizures in the derivation cohort and 16 (62%) out of 26 in the overall validation cohort (table 1). Neither in the derivation cohort (adjusted hazards ration [HR] 2·1, 95% CI 0·6-7·1; p = 0·25) nor in the validation cohort (adjusted HR 3·0, 95% CI 0·4-26·0; p = 0·31) was AED treatment after an early seizure significantly associated with late seizure risk after correction for the presence of early seizures.

Participants receiving antiepileptic treatment at time of stroke and a history of seizures or epilepsy were excluded in all cohorts. There were no stroke survivors receiving antiepileptic treatment used for indications other than epilepsy (i.e. neuropathic pain or psychiatric conditions) in the Austrian, German, and Italian cohorts, whereas this data was not explicitly collected in the Swiss cohort.

4 Missing data sensitivity analysis

The Swiss, German and Italian cohorts had missing data, which was imputed using a multiple imputations approach. Below is a detailed breakdown of missing data per cohort.

4.1 Missing data in individual cohorts

Table A3: *Missing data in the Swiss cohort (n=1200).*

Variable	Valid	Valid percent	Missing	Missing percent
Anticoagulation therapy	1152	96%	48	4%
Antiplatelet therapy	1168	97.3%	32	2.7%
Smoking	1182	98.5%	18	1.5%
Stroke aetiology	1184	98.7%	16	1.3%
Dyslipidaemia	1188	99%	12	1%
Diabetes mellitus	1188	99%	12	1%
Hypertension	1191	99.3%	9	0.7%
White matter hyperintensities	1193	99.4%	7	0.6%
Secondary hemorrhage	1195	99.6%	5	0.4%
ACA territory involvement	1196	99.7%	4	0.3%
MCA territory involvement	1196	99.7%	4	0.3%
PCA territory involvement	1196	99.7%	4	0.3%
Cortical involvement	1196	99.7%	4	0.3%
Stroke laterality	1196	99.7%	4	0.3%
Anticonvulsant treatment	1198	99.8%	2	0.2%
Age	1200	100%	0	0%
Sex	1200	100%	0	0%
Stroke severity	1200	100%	0	0%
Thrombolysis	1200	100%	0	0%
Early seizure	1200	100%	0	0%
Total	**23819**	**99.2%**	**181**	**0.8%**

Table A4: *Missing data in the Austrian cohort (n=459).*

Variable	Valid	Valid percent	Missing	Missing percent
NIHSS	459	100%	0	0%
Stroke aetiology	459	100%	0	0%
Early seizures	459	100%	0	0%
Cortical involvement	459	100%	0	0%
Territory of MCA	459	100%	0	0%
Total	**2295**	**100%**	**0**	**0%**

Table A5: *Missing data in the German cohort (n=311).*

Variable	Valid	Valid percent	Missing	Missing percent
NIHSS	271	87.1%	40	12.9%
Stroke aetiology	311	100%	0	0%
Early seizures	311	100%	0	0%
Cortical involvement	311	100%	0	0%

Territory of MCA	311	100%	0	0%
Total	**1515**	**97.4%**	**40**	**2.6%**

Table A6: *Missing data in the Italian cohort (n=399).*

Variable	Valid	Valid percent	Missing	Missing percent
NIHSS	394	98.7%	5	1.3%
Stroke aetiology	399	100%	0	0%
Early seizures	399	100%	0	0%
Cortical involvement	399	100%	0	0%
Territory of MCA	399	100%	0	0%
Total	**1990**	**99.7%**	**5**	**0.3%**

4.2 Sensitivity analysis using available data

A sensitivity analysis was performed using only available data. Similar results to the main analysis were generated in all cohorts. The difference in adjusted hazards ratios (aHR) between sensitivity and main analyses was ≤ 0.3. The difference in c statistics between sensitivity and main analyses was ≤ 0.02 (Table A1) in all cohorts and the model remained well-calibrated (Figure A4).

Table A7: *Multivariable Cox proportional hazards model of time to first late seizure in the derivation cohort using only available data (n=1090).*

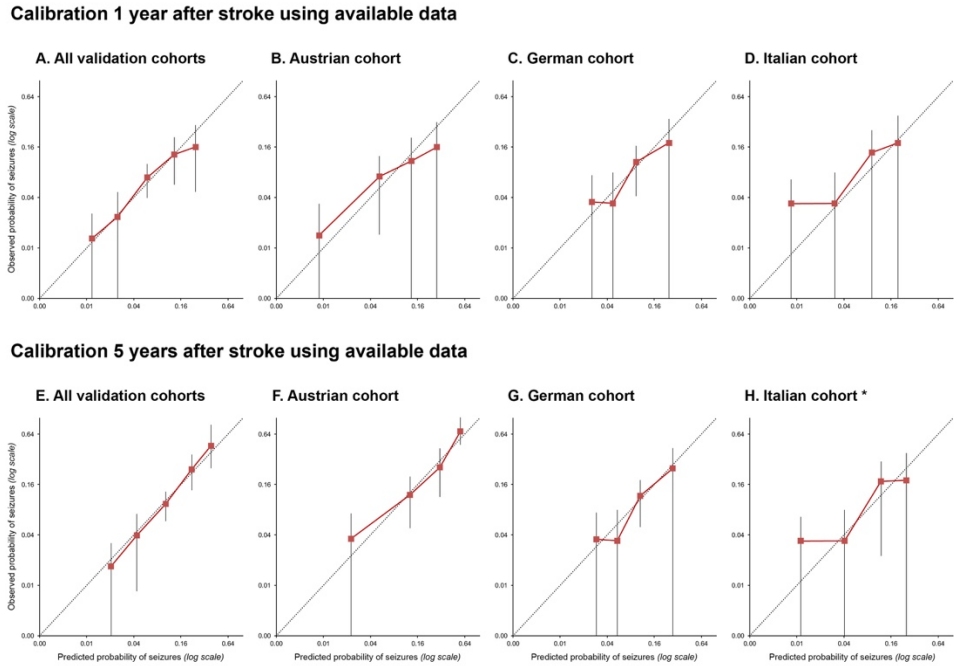
Variable	aHR (95% CI)	P value	ΔAIC
Cortical involvement	3.9 (1.8 – 8.4)	0·0005	-13.5
Early seizure	4.7 (2.4 – 9.2)	<0·0001	-13.3
Stroke severity at admission			
NIHSS ≤ 3	Reference category	--	--
NIHSS 4 to 10	1.5 (0.9 – 2.8)	0.15	1.9
NIHSS ≥ 11	2.6 (1.4 – 4.7)	0.002	-5.4
Territory of MCA involvement	1.7 (0.7 – 3.8)	0·21	0.2
Large-artery atherosclerosis	1.5 (0.9 – 2.6)	0·12	-0.4
Age	Eliminated in step 5 (p = 0.29, ΔAIC = 0.9)		
Stroke laterality: right	Eliminated in step 4 (p = 0.35, ΔAIC = 1.1)		
White matter hyperintensities	Eliminated in step 3 (p = 0.76, ΔAIC = 1.9)		
Thrombolysis	Eliminated in step 2 (p = 0.88, ΔAIC = 1.9)		
Small-vessel occlusion	Eliminated in step 1 (p = 0.93, ΔAIC = 2.0)		

NIHSS = National Institutes of Health Stroke Scale, MCA = middle cerebral artery, aHR = adjusted hazards ratio, CI = confidence interval, ΔAIC = Change in Akaike information criterion.

Table A8: *SeLECT discrimination using only available data.*

Cohort	Number of subjects	C statistic (95% confidence interval)
Overall validation	1124	0.76 (0.70-0.82)
Austria	459	0.78 (0.70-0.87)
Germany	271	0.72 (0.60-0.84)
Italy	394	0.81 (0.68-0.93)

Figure A4: *Calibration plots after 1 and 5 years using available data. *Data after 5 years was not available in the Italian cohort and the calibration plot after two years is displayed instead-*



5 Derivation in those receiving MRI scans

In the derivation cohort, 80% of subjects received MRI scans whereas 20% received CT scans only. People receiving follow-up CT were older ($p<0.001$) and had more severe strokes ($p<0.001$) compared to those receiving MRI. Excluding those who did not receive MRI would have introduced relevant selection bias. Nevertheless, we performed a sensitivity analysis including only those who received MRI scans in the derivation cohort (see table below). The multivariable Cox proportional hazards model generated similar results to the main analysis, although some deviation was noted that can be attributed due to including a mode mildly affected cohort.

Table A9: *Multivariable Cox proportional hazards model of time to first late seizure in the derivation cohort in those subjects who received MRI scans (n = 954).*

Variable	aHR (95% CI)	P value	ΔAIC
Cortical involvement	3.2 (1.5 – 7.0)	0.004	-8.5
Early seizure	5.3 (2.7 – 10.3)	<0.0001	-14.8
Stroke severity at admission			
NIHSS ≤ 3	Reference category	--	--
NIHSS 4 to 10	1.7 (0.9 – 3.2)	0.08	-1.0
NIHSS ≥ 11	2.3 (1.2 – 4.5)	0.01	-3.8
Territory of MCA involvement	1.6 (0.7 – 3.7)	0.26	0.6
Large-artery atherosclerosis	1.4 (0.8 – 2.5)	0.23	0.6
Stroke laterality: right	Eliminated in step 5 (p = 0.37, ΔAIC = 1.2)		
Small-vessel occlusion	Eliminated in step 4 (p = 0.41, ΔAIC = 1.3)		
Age	Eliminated in step 3 (p = 0.70, ΔAIC = 1.9)		
White matter hyperintensities	Eliminated in step 2 (p = 0.87, ΔAIC = 2.0)		
Thrombolysis	Eliminated in step 1 (p = 0.94, ΔAIC = 2.0)		

NIHSS = National Institutes of Health Stroke Scale, MCA = middle cerebral artery, aHR = adjusted hazards ratio, CI = confidence interval, ΔAIC = Change in Akaike information criterion.

6 Additional analyses in individual cohorts

6.1 Electrolytes, glucose, and fever

Sodium, potassium, and glucose blood levels and fever were measured at admission in the Italian (n = 399) cohort. As discussed in the Italian reference manuscript (Serafini et al. 2015, Neuroepidemiology) and displayed in the table below, there was no association of these factors with late seizures after stroke.
Blood glucose at admission was also captured in the German (n = 311) cohort and there was no association between blood glucose levels and late seizures (hazards ratio 1.08 per mmol/l, 95% CI 0.97 – 1.20; p = 0.17).

Table A10: *Association of predisposing factors with late seizures in the Italian cohort (n = 399).*

Variable	N (%) in Italian cohort	Hazards ratio (95% CI)	P value
Sodium			
> 145 mmol/l	4 (1%)	0.05 (0-10^{10})	0.80
< 135 mmol/l	23 (6%)	3.1 (0.7-14.3)	0.14
Potassium			
> 5.1 mmol/l	12 (3%)	2.0 (0.3-15.4)	0.51
< 3.5 mmol/l	43 (11%)	0.6 (0.1-4.9)	0.66
Glycemia ≥ 11.1 mmol/l	222 (56%)	1.1 (0.3-3.8)	0.84
Fever at stroke onset	3 (1%)	0.05 (0-10^{16})	0.89

6.2 Positive family history for epilepsy

Family history data was available from the German (n = 311) cohort. A positive family history for epilepsy was not associated with late seizures in the German cohort (19/311 [6%] with positive family history for epilepsy; hazards ratio 0.7, 95% CI 0.1 – 5.4; p = 0.76).

7 Late seizures with and without preceding early seizures

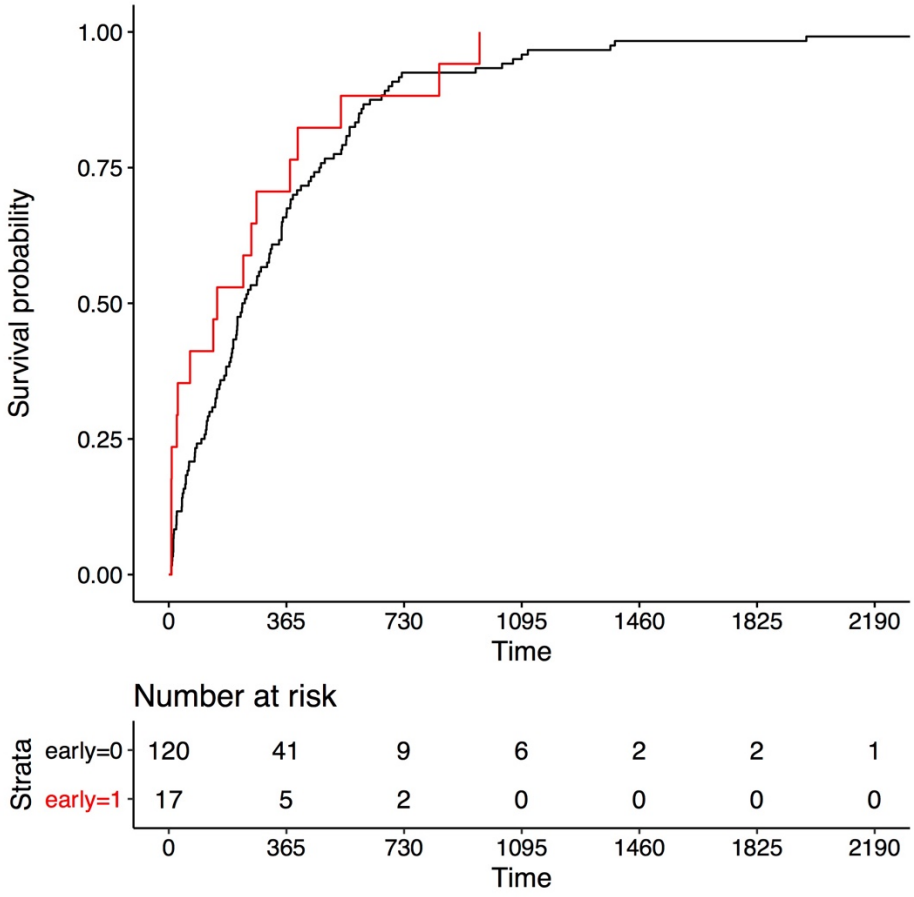
7.1 Recurrence rate

People with early seizures had more than one late seizure during follow-up in 64% of cases which is comparable to those without early seizures (60%).

7.2 Time to first late seizure

There was no difference in the time to first late seizure (HR 0.7, 95% CI 0.4 – 1.2; p = 0.27) in those who had a late seizure following an early seizure as compared to those who had a late seizure without preceding early seizures.

Figure A5: *Kaplan Meier estimates of time to first late seizure in people with (n=17, red line) and without (n = 120, black line) early seizures. There was no statistically significant difference between groups (p = 0.27). Time is given in days after stroke.*



7.3 Provoking factors

Electrolyte disturbance, blood sugar levels and signs of infection (fever, C-reactive protein [CRP]) were captured at admission in the Austrian and Italian cohorts. No subjects with late seizures precipitated by an early seizure had fever or disturbances of serum sodium, potassium or glucose levels at admission. Only one person had mildly elevated CRP at admission (25.8 mg/l) which is not uncommon in severe stroke (NIHSS 26 points at admission in this case) (Christensen and Boysen 2004 Cerebrovasc Dis).

8 Clinical rationale for factors included in the SeLECT model

The factors included in the SeLECT model are clinically reasonable. The cerebral cortex and regions in the MCA territory, i.e. the temporal and frontal lobes, might be more susceptible to epileptogenesis after stroke than other brain areas (Ferlazzo *et al.,* 2016; Pitkänen *et al.,* 2015). Severe strokes might disrupt a larger area of the cortex and they might be associated with more severe inflammation (Danton and Dietrich, 2003) and increased release of excitatory neurotransmitters (Castillo *et al.,* 1996). Large-artery atherosclerotic stroke is associated with a specific activation of inflammatory pathways involving the CD40/CD40L system (Antoniades *et al.,* 2009) that has also been implied in post-stroke epilepsy (Pitkänen *et al.,* 2015; Zhang *et al.,* 2014). Atherosclerosis might also be associated with blood-brain-barrier disruption that could facilitate epileptogenesis (Seiffert *et al.,* 2004). People suffering early seizures after stroke might have an increased predisposition to generate epileptic seizures. In other words, early seizures could be a marker of a lower overall 'threshold' for seizures that also increases the susceptibility to epileptogenesis (Engel, 2013).

We consider it unlikely that those late seizures, that were preceded by an early seizure, were provoked by an ongoing precipitating factor, e.g. electrolyte disturbance or infection after stroke. Provoked late seizures were explicitly excluded according to our definitions and there were no differences in recurrence rates and time to first late seizure in people who had or did not have early seizures (see section 8 in the Appendix).

References:

Antoniades C, Bakogiannis C, Tousoulis D, Antonopoulos AS, Stefanadis C. The CD40/CD40 ligand system: linking inflammation with atherothrombosis. *J Am Coll Cardiol* 2009; **54**: 669–77.

Castillo J, Dávalos A, Naveiro J, Noya M. Neuroexcitatory amino acids and their relation to infarct size and neurological deficit in ischemic stroke. *Stroke* 1996; **27**: 1060–5.

Danton GH, Dietrich WD. Inflammatory mechanisms after ischemia and stroke. *J Neuropathol Exp Neurol* 2003; **62**: 127–36.

Engel J. Causes of Human Epilepsy. In: Engel J Jr., ed. Seizures and Epilepsy, 2nd edn. Oxford University Press, 2013: 157–90.

Ferlazzo E, Gasparini S, Beghi E, *et al.* Epilepsy in cerebrovascular diseases: Review of experimental and clinical data with meta-analysis of risk factors. *Epilepsia* 2016; **57**: 1205–14.

Pitkänen A, Roivainen R, Lukasiuk K. Development of epilepsy after ischaemic stroke. *The Lancet Neurology* 2015; **15**: 185–97.

Seiffert E, Dreier JP, Ivens S, *et al.* Lasting blood-brain barrier disruption induces epileptic focus in the rat somatosensory cortex. *J Neurosci* 2004; **24**: 7829–36.

Zhang B, Chen M, Yang H, Wu T, Song C, Guo R. Evidence for involvement of the CD40/CD40L system in post-stroke epilepsy. *Neurosci Lett* 2014; **567**: 6–10.

9 Informed consent procedures

All subjects in the Italian cohort and those having a face-to-face interview in the Swiss cohort gave written informed consent. All subjects evaluated by telephone in the Swiss and German cohorts gave verbal informed consent. According to Swiss and German law the regional ethical committees exempted these cohorts from requiring written informed consent. The Austrian case-control study was classified as retrospective service evaluation by the regional ethical committee and informed consent was not required.

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	4-5
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	5
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	5-7
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	5-7, Table 1
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	5-7
	5b	D;V	Describe eligibility criteria for participants.	5-7
	5c	D;V	Give details of treatments received, if relevant.	Appendix, p.6
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	7
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	NA
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	7
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	NA
Sample size	8	D;V	Explain how the study size was arrived at.	5-7
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	8
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	9
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	8
	10c	V	For validation, describe how the predictions were calculated.	9
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	8-9
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	NA
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	Fig. 4
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	5-7, 13, Table 1
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	9, Table 1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Table 1
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	Table 1
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	9-11, Fig. 1
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	Table 2
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Table 3
	15b	D	Explain how to the use the prediction model.	14, Panel 1, Appendix
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	10-11
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	NA
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	12-14
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	NA
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	11-15
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	14-15
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	23, SeLECT app
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	9

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.