Risk Prediction of Cognitive Decline after Stroke

Youssef Hbid, Marion Fahey, Charles D A Wolfe, Majed Obaid, Abdel Douiri

Risk prediction of cognitive decline after stroke

Youssef Hbid[123], Marion Fahey[4], Charles D.A. Wolfe[45] , Majed Obaid[4] , Abdel Douiri[45]

[1] LMDP, Cadi Ayyad University, Marrakech, Morocco
[2] UMMISCO, IRD, France
[3] Sorbonne University, Laboratoire Jacques-Louis Lions, Paris, France
[4] King's College London, School of Population Health and Environmental Sciences, London, United Kingdom
[5] National Institute for Health Research Biomedical Research Centre, Guy's and St Thomas' NHS Foundation Trust and King's College London, London, United Kingdom

Corresponding author: Youssef Hbid
Corresponding author's e-mail address: Youssef.hbid@etu.upmc.fr

Abstract

Background and Purpose

Cognitive decline is one of the major outcomes after stroke. We have developed and evaluated a risk predictive tool of post-stroke cognitive decline and assessed its clinical utility.

Methods

In this population-based cohort, 4,783 patients with first-ever stroke from the South London Stroke Register (1995-2010) were included in developing the model. Cognitive impairment was measured using the Mini Mental State Examination (cut off 24/30) and the Abbreviated Mental Test (cut off 8/10) at 3-months and yearly thereafter. A penalised mixed-effects linear model was developed and temporal-validated in a new cohort consisted of 1,718 stroke register participants recruited from (2011-2018). Prediction errors on discrimination and calibration were assessed. The clinical utility of the model was evaluated using prognostic accuracy measurements and decision curve analysis.

Results

The overall predictive model showed good accuracy, with root mean squared error of 0.12 and R2 of 73%. Good prognostic accuracy for predicting severe cognitive decline was observed AUC: (88%, 95% CI [85-90]), (89.6%, 95% CI [86-92]), (87%, 95% CI [85-91]) at 3 months, one and 5 years respectively. Average predicted recovery patterns were analysed by age, stroke subtype, Glasgow-coma scale, and left-stroke and showed variability.

Decision curve analysis showed an increased clinical benefit, particularly at threshold probabilities of above 15% for predictive risk of cognitive impairment.

Conclusions

The derived prognostic model seems to accurately screen the risk of post-stroke cognitive decline. Such prediction could support the development of more tailored management evaluations and identify groups for further study and future trials.

Keywords: Post-stroke, rehabilitation, Cognitive decline, monitoring, recovery, clinical prediction, case mix, mixed-effects model.

1. Introduction

Stroke is a common long-term condition with an increasing incidence as the population ages. Patients who have had a stroke have an increased likelihood of cognitive deficit compared to those who have not had a stroke [1]. It remains persistently high up to fifteen years post-stroke and is associated with higher disability, lower quality of life and depression. An increasingly ageing population coupled with the decline in mortality after stroke [2] means that post-stroke cognitive impairment will become more prevalent particularly since the risk of stroke [3] and cognitive impairment [4] rise exponentially with age. Studies have suggested that cognitive decline could be predictable after stroke [5-6].

A longitudinal follow-up and a patient-specific predictive models may be more appropriate to accurately capture health outcomes with the aim of planning immediate, mid and long-term care simultaneously for individual patients with poor physical and psychological outcomes. Preventive medication and rehabilitation programs are available for controlling risks but a patient-centred instrument to determine in advance when a poor health outcome might occur would assist management of care of these patients, and therefore, allow them to live a more normal life.

The Mini-Mental State Examination (MMSE) is the most widely used instrument for screening dementia [7] and it is significantly correlated with cognitive decline following stroke. Suzuki et al [8] used MMSE scores at baseline (stroke onset) to predict MMSE scores over time (model 1 R^2 = 67.6%, model 2 R^2 = 59.8%). However, this may not capture recovery accurately as MMSE scores at stroke onset are often much lower than at subsequent time points because many patients experience some improvement following the acute phase. Ross et al [9] used imaging metrics from proton magnetic resonance spectroscopy to predict cognitive decline in patients up to three years following stroke (R2=54.6%). Measures in the frontal white matter of the brain were associated with change in composite z-score of tests assigned to each cognitive domain three years post-stroke (R2=54.6%). Similarly, Saini et al [10] used metrics from computed tomography scans to predict cognitive decline three to six months following ischemic stroke. The presence of significant atrophy and white matter lesions were associated with cognitive decline with an odds ratio of 3.07 and 3.13 respectively. Whilst these models could be a useful assistive tool to predict cognitive decline, atrophy are not systematically measured in the practice.

Tang et al [11] reported that several models have been developed for people with stroke to predict dementia [12] [13] or cognitive impairment [14][15] and their predictive accuracy was found to be acceptable. Different variables including demographic, cognitive test scores and neuroimaging markers have been incorporated into different models, with predictive accuracy found to be moderate to high. However, their utility is challenging as they include neuroimaging variables that are not easily accessible.

A recent systematic review [16] shows that cognitive decline seems to become more apparent over a longer follow-up period, and thus new models could be developed to predict post stroke cognitive impairment and dementia over longer time periods.

In this study, we develop and validate a patient-specific predictive model to estimate risk for cognitive decline up to 5 years after ischemic stroke and assess deviations from observed and predicted recovery and differences in recovery trends.
The output from this research will be used to aid long-term monitoring and provide prognostic information to stroke survivors and their families, and to assist the development of more tailored long-term management and care plans.

2. Methods

2.1 Source of data

Data for this analysis were derived from the South London Stroke Register (SLSR), an ongoing population-based register that has prospectively recorded first ever strokes in patients of all age groups living within a geographically defined area of south London since 1995. In this analysis we used data collected between 1995 and 2018.
The methods of the SLSR have been described in detail by Wolfe et al [6-17] and are summarised here. All patients with a first ever stroke after 1st January 1995 and residing in a defined inner-city area of South London were eligible for inclusion. According to the 2011 Census, with annual predicted changes, the north Southwark and Lambeth (n=357,308) comprises a multi-ethnic population with a large proportion of black Caribbean and African residents (25.3%). Stroke is defined according to WHO definition of stroke [17]. Case ascertainment is estimated as 88% complete by a multinomial-logit capture-recapture model [17].

2.2 Participants

Patients admitted to hospitals serving the study area (2 teaching hospitals within and 3 hospitals outside the study area) were identified by regular reviews of acute wards admitting stroke patients, national data on patients admitted to any hospital in England and Wales with a diagnosis of stroke are screened for additional patients.
All general practitioners (N=699 (2011)) within and on the borders of the study area are contacted regularly and asked to notify the SLSR of stroke patients. Referral of non-hospitalized stroke patients to a neurovascular outpatient clinic (from 2003) or domiciliary visit to patients by the study team is also available to general practitioners. Community therapists are contacted every 3 months. Death certificates are checked regularly. Patients are assessed at the stroke onset, 3 months and annually after stroke.

2.3 Outcome and Predictors

The outcome of interest is cognitive impairment up to five years following stroke, measured by the Mini Mental State Examination (MMSE) or Abbreviated Mental Test (AMT) score [18]. Patients were assessed at seven days, 3 months, and annually after stroke. Before January 1, 2000, cognitive state was assessed with the Mini-Mental State Examination; after that date, the Abbreviated Mental Test was adopted. Subjects were defined as cognitively

impaired according to predefined cut-off points (Mini-Mental State Examination <24 or Abbreviated Mental Test <8). It has been shown that the Mini-Mental State Examination and Abbreviated Mental Test are insensitive to mild cognitive impairment and executive function [19][20]. The AMT shows good concordance with the MMSE (c-statistic from 0.83 to 0.87) [21].

The meta-analysis (73 studies) conducted by Pendlebury and Rothwell [22] was used to identify an initial list of candidate predictors for post-stroke cognitive decline.

These candidate predictors were subsequently screened for practicality based on clinical availability, ease of measurement, prevalence in academic literature, and on biological reasoning with experts (stroke physician, statistician and epidemiologist). This yielded an initial list of 93 candidate predictors available in the SLSR. Data are collected by SLSR field workers uninvolved in this study at baseline, 3months, one year and annually thereafter.

2.4 Missing data

Multiple imputation using Markov chain Monte Carlo methods was used to impute missing values, under a missing at random assumption, so as to reduce bias and avoid excluding participants from the analysis [23].

2.5 Statistical prediction methodology and analyses

2.5.1 Variables selection

Random forest method [24] was used to rank the candidate predictors in order of importance. Predictors considered to have great clinical relevance were forced back into the model. Penalized mixed models [25] were then adapted to develop trajectories of cognitive decline for a patient with the selected prognostic factors. Clinically meaningful interactions were included in the model. Their significance was tested as a group to avoid inflating type I error. All interaction terms were removed as a group, and the model was refitted if results were nonsignificant. Interactions with time were also examined.

2.5.2 Performance measures

We assessed internal validity with cross validation method for a realistic estimate of the performance of prediction model in similar future patients. Performance measures included the area under the Receiver Operating Curve (AUROC) curve, sensitivity and specificity, calibration slope, Brier score and Decision Curve Analysis (DCA) [26-27]. Discrimination refers to the ability of the risk score to differentiate between cognitively intact and cognitively impaired patients. DCA was performed in order to evaluate further the clinical usefulness of survival curves in prognostication of cognitive impairment at three months, one year and five years. DCA is a method to assess the added value of information provided by a prognostic test across a range of a patient's risks and benefits to facilitate clinical decisions, without the need for actually measuring these for individual patients. The DCA is expressed graphically as a

curve, with the clinical net benefit on the vertical axis and probability thresholds on the horizontal axis. The net benefit of prediction models was then evaluated by adding the benefits (true positives) and subtracting the harms (false positives). The weight assigned to true positives and false positives was derived from the threshold probability of the outcome. When the curve is at its highest over the range of probability thresholds, the associated intervention would be the best decision. Statistical analysis was performed using R-software.

2.5.3 Recovery curve trajectories

We plotted recovery curve trajectories to visually examine different well-defined at-risk subgroups. Average predicted patterns were analysed by age, stroke subtype, Glasgow coma scale and left- stroke occurrence. To assess the prognostic effectiveness and clinical utility of predicted recovery curves to estimate different cognitive outcomes at different time points, cognitive impairment was dichotomized using mild cognitive impairment (cut-off: 24/30 MMSE and 8/10 AMT) and severe cognitive impairment (cut-off: MMSE and 4/10 AMT) [28-29-30]. Clinical utility was also assessed at these thresholds of the predicted recovery curves at three months, one year and five years.

2.5.4 Model development and validation

A penalised mixed-effects linear model was developed and temporal-validated. Repeated random sub-sampling cross-validation methods were used to select best competing models and model parameter. Internal cross-validation was used to assess the performance of the developed prognostic recovery curve model. R2 and root-mean-square error (RMSE) were considered together to estimate the predictive error. Patient age, sex, ethnic group, cognition score at the onset of stroke, Barthel-index score at baseline, Glasgow coma scale (GCS), stroke subtype (LACI, PACI, POCI, TACI) [31], diabetes, left hemisphere stroke, dysphasia and interactions between predictor variables and the time in years were identified as good independent predictors.

Ethics

Patients, or for patients with communication problems their relatives, gave written informed consent to participate in stroke-related studies within the SLSR. The design was approved by the ethics committees of Guy's and St Thomas' NHS Foundation Trust, Kings College Hospital, Queens Square, and Westminster Hospitals (London).

3. Results

3.1 Participants' Characteristics

A total of 6,504 patients with their first-ever stroke between 1995 and 2018 were registered in the SLSR. Of whom n=3411 patients had cognitive function measured at seven days, of them n=1204 had cognitive impairment. A total of n = 1608 completed a follow-up interview at one

year, and n = 846 completed a follow-up interview at five years. A total of n = 2171 individuals died within three months. A total of n=2000 individuals did not have cognitive function measured at 7 days after stroke, due to medical reasons. At stroke onset, the medical reasons were communication impairment n = 992 and coma n = 737. The remaining number was due to late registration or because their date of follow-up was not reached n = 271.
The development cohort consisted of 2,468 participants from (1995-2010) and the validation cohort consisted of 940 stroke register participants recruited from (2011-2018).
Table 1 summaries the patients' characteristics in both development and validation cohorts. Missing data accounted for less than 15% of the data. Key characteristics were typically evenly distributed between both cohorts.

3.2 Model performance

Patient age, sex, ethnic group, cognition score at the onset of stroke, Barthel-index score at baseline, Glasgow coma scale (GCS), stroke subtype (LACI, PACI, POCI, TACI) [29], diabetes, left hemisphere stroke, dysphasia and interactions between predictor variables and the time in years were identified as good independent predictors.
The predictive recovery curves showed a good fit and prediction. In the internal-cross validation, predictive error RMSE over all time points was 0.12 and R2 was 73%. Average cognition score was characterized by an initial improvements over the first 3 months and then a gradual decline thereafter. Figure 1 presents the average predicted trajectories compared to the average observed cognition score after stroke up to 5 years.
The predictive curves show similarities between LACI and POCI stroke at baseline but large difference 1 year later, with LACI having the largest decline compared to POCI.
Dissimilarities were observed between TACI and PACI stroke at the baseline but was comparable after 3 years. For instance, we have shown that changes in cognition score vary between different age groups. We observed an improvement phase the first year in younger patients but a significant decrease in older stroke survivors up to 5 years. But despite the improvement phase in the younger patients, we would expect a small decline in the cognition score after 1 year. Sever stroke (moderate to severe Glasgow coma scale (GCS) or left-stroke occurrence at onset) showed significant association with cognitive decline.
Figure 2 presents average cognition score after stroke stratified by age group, stroke subtype and GCS.
The model was further evaluated to identify the prognostic accuracy, the sensitivity and specificity and the utility of the model at different cognitive decline cut-off scores that enable the discrimination between severe and mild cognitive impairment at 3 months, 1 year and 5 years following stroke. The validity of the model is good, at 3 months (sensitivity ranging from 52-71% and specificity 91- 94%) for severe cognition and (sensitivity ranging from 73-82% and specificity 68-75%) for mild cognition score.
The model has also shown potential utility, the negative predictive values were (96%, 95% CI [94- 97]), (96%, 95% CI [94- 97]), (97%, 95% CI [96- 98]) for severe cognitive impairment at 3 months, 1 year and 5 years respectively.

Table 2 summaries all the predictive values and likelihood rations for classifying each cognitive impairment score of interest.

The net benefit as a function of a threshold probability of cognitive impairment at 3 months 1 year and 5 years was illustrated in Figure 3. The grey line was drawn to reflect the strategy of assuming that all patients are cognitively impaired (i.e. recommend intervention for all), and the black line was drawn to reflect the strategy of assuming that all patients are not cognitively impaired (i.e. do not recommend any intervention).

The net benefit was maximized by the cognitive decline curve of the predictive model (red line) with threshold probabilities of 15–80% at 3 months, 15–79% at 1 year and 15-82% at 5 years. For higher thresholds (>80% for 3 months, >79% for 1 year and > 82% for 5 years) where the concerns are more about unnecessary interventions than missed prognosis, the option to not intervene was preferred.

4. Discussion

In this study, we have developed and validated a patient-specific prognostic tool for cognitive decline post-stroke in a population-based cohort. The proposed model is patient-specific and enables cognitive impairment to be predicted using a continuous score. It has additionally provided the ability to accurately predict trajectories up to 5 years post-stroke.

A recent systematic review reported that several models have been used to predict dementia and cognitive impairment [16]. Regarding global cognitive function, the majority of studies reported decline [32-33-34-35-36-37] whereas [38-39-40-41] reported no change. Most models have a relatively short predictive period and don't asses the risk of cognitive decline over longer periods particularly in those who have a stroke at a younger age. Furthermore, they predict risk of cognitive impairment at predefined time points only. At predefined time points, the accuracy of the proposed model has been shown to be superior compared to other exiting prediction models cited earlier. The general pattern of cognitive decline from stroke has already been discussed and illustrated at population level in previous studies [6][42]. This could be useful for early rehabilitation and discharge planning, by predicting whether a patient is likely to be dependent, require some assistance, or be independent, at a certain time post-stroke. Factors influencing recovery were the laterality of the stroke and lowered consciousness on admission. Patients with right-side brain damage performed better than those with left-side brain damage and showed more improvement in cognition score over time. Cognitive impairment progression in patients with lowered consciousness on admission was worse than patients without lowered consciousness overtime. Specialized stroke rehabilitation may be beneficial for all ages but important for over 65. It also confirms that older patients may need longer rehabilitation and are less likely to be discharged earlier. Apart from old age, factors such as onset stroke severity should also be taken into consideration when planning interventions and rehabilitation after stroke. We have shown that using a multivariate patient-specific predictive model, we can make individual recovery profiles and accurately classify future risks of cognitive decline. This model makes predictions of continuous outcome rather than restricting to binary abstractions. Furthermore, the predictions are not restricted by specific time points, as demonstrated by the average recovery patterns

and subgroup analyses up to 23 years post-stroke. The final model parameters were selected using k-fold cross validation.

This signifies that the final coefficients reflect averages of many models built on random subpopulation permutations to ensure that the final model parameters reflect real associations rather than being subject to overfitting. In addition, the sample size is large in relation to the number of prognostic variables, increasing the power of the study. The variables incorporated into the model were selected for their association with cognitive decline following stroke, and tested using several, robust methods, thus ensuring complete confidence in the predictive abilities of the variables. In addition, these variables are routinely collected during acute stroke care and follow-up assessments, thus increasing the ease of use of the model. Patient-specific recovery curves predictions could allow more insight into both spontaneous and directed neurological recovery after stroke. This prognostic information is important for clinicians, stroke survivors and their careers. In clinical research, this could also be applied as an aid in assessing the beneficial effects of evidence-based interventions and care settings. As a research tool, this could be used to test novel interventions or to identify enriched samples, reducing the reliance on the need for expensive and often impractical randomized controlled trials. This predictive enrichment strategy is of importance for designing future trials as it enables the enrolment of the most suitable patients thereby permitting the use of a smaller study population. Another potential application could be to derive a set of preliminary cost weights on resource uses which could help commissioners build personalised patient care funding models.

A key strength of the current study is that the model was built using a prospective, non-selected population-based cohort of first ever stroke. This is preferable to hospital based populations, which may result in case-mix specific models, or aggregate data from clinical trials, which usually represent heavily selected and thus non-representative populations. Our data sample is truly reflective of the geographic population of interest and therefore optimal for deriving a representative model.

Appropriate vascular risk management was associated with a long-term reduced risk of cognitive impairment. Focus on optimal preventive drug therapy of vascular risk factors and management should be supported [43]. This model can potentially assist clinicians in organizing a program of care for patients following stroke, which is tailored to their predicted pattern of cognitive development. It can also aid in communicating risk to patients and their families and careers, in a straightforward and clear manner, especially through the use of the graphical representations of cognition score as a function of time.

Limitations:

Notwithstanding the strengths, the following limitations of this study must be acknowledged. Firstly, the study could be improved further if the model were to be validated in a completely independent population, preferably from another country and by independent researchers. Secondly, we used a penalized mixed effect model that automatically selects and subsequently shrinks effect sizes of important predictors. This regularization strategy may have led to some underestimation of predictor effects in the development sample, but it increases the likelihood of replication in validation studies.

Thirdly, an impact study needs to be conducted in a randomized control trial (RCT) setting to confirm whether being able to predict recovery and the resulting intervention, could make a difference to the patient. Fourthly, mild cognitive decline is measured by executive function, the MMSE and AMT scores do not measure this. Perhaps the accuracy of the model would be increased by using measures that take into account executive function such as the Montreal Cognitive Assessment tool (MoCA)[44]. MoCA was feasible and reliable, however, examination of Visio executive and complex language tasks was limited compared with face-to-face assessment. MMSE scores were predictive of cognitive impairment and dementia on follow-up, and most widely used by clinicians in routine care.

It can be noted that a key limitation of the current study was including only routinely collected data and no other important variables that could potentially refine the model further, for instance genomic data. However, the objective of this work was to produce simple models that can be used in a wider clinical setting. Finally, cognitive impairment prediction without an effective therapy at hand raises ethical concerns. Such models are unlikely to be rolled out into clinical practice before further validation and assessment are undertaken.

Future directions:

Future work may wish to consider further evaluation of the proposed model, impact study to bring such models into clinical practice and application of the recovery curves methods to other outcomes.

5. Implications

The research implications for this study lie in the increased understanding of patient-specific post-stroke cognitive decline patterns. The model can predict long-term risk up to 5 years post-stroke, which to the knowledge of the authors, has not been performed for cognitive decline in stroke patients. The model may be used as a research tool to test the influence of novel interventions and drugs on cognitive decline.

The proposed model yields personalized patterns of cognitive decline scores, up to 5 years post-stroke by taking into account clinically available predictors for post-stroke cognition.

The predictions can be altered in light of observed recovery, thus refining the model and allowing for more precise predictions. The cut-off used by the model can be adjusted to emphasize detection of mild or severe cognitive impairment, thus enabling flexibility depending on the patient's characteristics. The model can detect those at higher risk of cognitive decline, as demonstrated by our subgroup analyses, thus identifying patients who will most benefit from treatment and improving cost-effectiveness in stroke care.

The implications of this study are wide ranging, providing a way to effectively organize post-stroke care by determining groups at higher risk, communicating risk to patient and families and predicting drug treatment outcomes on cognition for research purposes.

6. Conclusion

The derived prognostic model seems to accurately predict the risk of post-stroke cognitive decline. This confirm that the recovery curves models applied to patient specific stroke data can predict trajectories accurately. Longitudinal measurement adds greater dimension to predictions and more accurate than measurement at an isolated time-point . The predictors of post stroke cognition used in the proposed model, were typically common routinely collected information. Therefore, could be used in post-stroke care, particularly in early detection and prevention to support clinical decisions.

DATA SHARING STATEMENT

The study and its consent procedure were approved by the ethics committees of Guy's and St Thomas' Hospital Trust, King's College Hospital, Queen's Square, and Westminster Hospital. Consent for data sharing was not obtained from study participants. The research team will consider reasonable requests for sharing of anonymised patient level data.

Conflict of Interest

We declare that we have no conflicts of interest.

Funding

The study forms part of a wider PhD thesis. We thank patients, their families, and the fieldworkers who have collected data for the South London Stroke Register since 1995. This work was supported by funding from the National Institute for Health Research (NIHR) Applied Research Collaboration (ARC) South London at King's College Hospital National Health Service Foundation Trust and the Royal College of Physicians, as well as the support from the NIHR Biomedical Research Centre based at Guy's and St Thomas' National Health Service Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the Kings College London, NHS, the NIHR or the Department of Health.

Authors' contribution

CDAW, AD conceived the study and provided study design guidance. YH performed statistical work for the study and drafted the manuscript. MF Marion Fahey worked in post stroke cognitive decline during her PhD and have thoroughly revised the manuscript and checked all analyses. MO reviewed the manuscript.
All authors contributed substantially to its revision.

References

1. Allan LM, Rowan EN, Firbank MJ, et al. (2011). Long term incidence of dementia, predictors of mortality and pathological diagnosis in older stroke survivors. Brain.134(pt 12):3716–3727.

2. Rothwell PM, Coull AJ, Giles MF, et al. (2004). Change in stroke incidence, mortality, case-fatality, severity, and risk factors in Oxfordshire, UK from 1981 to 2004 (Oxford Vascular Study). Lancet. ;363(9425):1925-1933. doi:10.1016/S0140-6736(04)16405-2

3. Rothwell PM, Coull AJ, Silver LE, et al. (2005) Population-based study of event-rate, incidence, case fatality, and mortality for all acute vascular events in all arterial territories (Oxford Vascular Study). Lancet.;366(9499):1773-1783. doi:10.1016/S0140-6736(05)67702-1

4. Fratiglioni L, Launer LJ, Andersen K, et al. (2000) Incidence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology.; 54(11 Suppl 5):S10-S15.

5. Tilling K, Sterne JA, Wolfe CD. (2001). Estimation of the incidence of stroke using a capture-recapture model including covariates. Int J Epidemiol.;30:1351–1359; discussion 1359.

6. Wolfe C, Crichton S, Heuschmann P, et al. (2011). Estimates of outcomes up to ten years after stroke: analysis from the prospective south london stroke register. PLoS Med; 8(5): e1001033.

7. Folstein MF, Folstein SE, McHugh PR. (1975). “Mini-mental state.” A p cal method for grading the cognitive state of patients for the clinician. J Psychiatr Res.;12:189–198.

8. Suzuki M, Sugimura Y, Yamada S, et al. (2013). Predicting recovery of cognitive function soon after stroke: differential modeling of logarithmic and linear regression. PLoS One.;8(1):e53488. doi:10.1371/journal.pone.0053488

9. Ross, Amy J. et al. (2006). Prediction of cognitive decline after stroke using proton magnetic resonance spectroscopy. Journal of the Neurological Sciences, Volume 251, Issue 1, 62 - 69

10. Saini et al. (2014). Computer tomography for prediction of cognitive outcomes after ischemic cerebrovascular events. Journal of stroke and cerebrovascular diseases : the official journal of National Stroke Association, 23 7, 1921-7

11. Tang EYH et al. Risk Prediction Models for Post-Stroke Dementia. Geriatrics (Basel). 2017 Jun 22;2(3):19. doi: 10.3390/geriatrics2030019. PMID: 31011029; PMCID: PMC6371182.

12. Stephan BCM, Kurth T, Matthews FE, Brayne C, Dufouil C. Dementia risk prediction in the population: are screening models accurate? Nat Rev Neurol. 2010;6:318–326.

13. Lim JS, Oh MS, Lee JH, Jung S, Kim C, Jang MU, Lee SH, Kim YJ, Kim Y, Park J, Kang Y, Yu KH, Lee BC. Prediction of post-stroke dementia using ninds-csn 5minute neuropsychology protocol in acute stroke. Int Psychogeriatr. 2017;29:777–784.

14. Kandiah N, Chander RJ, Lin X, Ng A, Poh YY, Cheong CY, Rae Cenina A, Nkouibert Assam P. Cognitive impairment after mild stroke: development and validation of the SIGNAL2 risk score.**J Alzheimers Dis**. 2016; 49:1169–1177. doi: 10.3233/JAD-150736.

15. Chander RJ, Lam BYK, Lin X, Ng AYT, Wong APL, Mok VCT, Kandiah N. Development and validation of a risk score (CHANGE) for cognitive impairment after ischemic stroke.**Sci Rep**. 2017; 7:12441. doi: 10.1038/s41598-017-12755-z

16. Tang EY, Amiesimaka O, Harrison SL, Green E, Price C, Robinson L, Siervo M, Stephan BC. Longitudinal Effect of Stroke on Cognition: A Systematic Review. J Am Heart Assoc. 2018 Jan 15;7(2):e006443. doi: 10.1161/JAHA.117.006443. PMID: 29335318; PMCID: PMC5850140.

17. Tilling K, Sterne JA, Wolfe CD. (2001). Estimation of the incidence of stroke using a capture-recapture model including covariates. Int J Epidemiol.;30:1351–1359; discussion 1359.

18. Hodkinson HM. Evaluation of a Mental Test Score For Assessment of Mental Impairment in the Elderly. Age Ageing. 1972;1:35–40.

19. Pendlebury ST, Cuthbertson FC, Welch SJ, Mehta Z, Rothwell PM.Underestimation of cognitive impairment by Mini-Mental State Examination versus the Montreal Cognitive Assessment in patients with transient ischemic attack and stroke: a population-based study. Stroke. 2010;41:1290–1293.

20. Elias MF, Sullivan LM, D'Agostino RB, Elias PK, Beiser A, Au R, Seshadri S, DeCarli C, Wolf PA. Framingham stroke risk profile and lowered cognitive performance. Stroke. 2004;35:404–409.

21. Piotrowicz, K., Romanik, W. et al. (2019). The comparison of the 1972 Hodkinson's Abbreviated Mental Test Score (AMTS) and its variants in screening for cognitive impairment. Aging clinical and experimental research, 31(4), 561–566. doi:10.1007/s40520-018-1009-7

22. Pendlebury ST, Rothwell PM. (2009). Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis. Lancet Neurol.;8:1006–1018.

23. Crichton S. (2016). Methods for handling missing data in a population-based cohort study. PhD Thesis, King's College, London (University of London).

24. Breiman, L. Random Forests. *Machine Learning* **45,** 5–32 (2001). https://doi.org/10.1023/A:1010933404324

25. Wood, S. (2017). Generalized Additive Models. New York: Chapman and Hall/CRC, https://doi.org/10.1201/9781315370279

26. Vickers AJ and Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006; 26: 565–574.

27. Fitzgerald M, Saville BR, Lewis RJ. (2015). Decision curve analysis. JAMA 313: 409–410.

28. Jitapunkul S, Pillay I, Ebrahim S. (1991). The abbreviated mental test: its use and validity. Age and ageing 20(5): 332-336.

29. Harvan JR, . Cotter VT. (2006). An evaluation of dementia screening in the primary care setting. J Am Acad Nurse Pract 18(8): 351-360.

30. Perneczky R, et al. (2006). Mapping scores onto stages: mini-mental state examination and clinical dementia rating. Am J geriatric psych 14(2): 139-144.

31. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. Lancet. 1991;337:1521–1526.

32. Rajan KB, Aggarwal NT, Wilson RS, Everson-Rose SA, Evans DA. Association of cognitive functioning, incident stroke, and mortality in older adults. Stroke. 2014;45:2563–2567.

33. Toole JF, Bhadelia R, Williamson JD, Veltkamp R. Progressive cognitive impairment after stroke. J Stroke Cerebrovasc Dis. 2004;13:99–103. Douiri A, Grace J, Sarker S, Tilling K, et al. (2017). Patient-specific prediction of functional recovery after stroke. International Journal of Stroke 12(5): 539-548.

34. Ben Assayag E, Shenhar-Tsarfaty S, Korczyn AD, Kliper E, Hallevi H, Shopin L, Auriel E, Giladi N, Mike A, Halevy A, Weiss A, Mirelman A, Bornstein NM, Hausdorff JM. Gait measures as predictors of poststroke cognitive function: evidence from the TABASCO study. Stroke. 2015;46:1077–1083.

35. Tene O, Shenhar-Tsarfaty S, Korczyn AD, Kliper E, Hallevi H, Shopin L, Auriel E, Mike A, Bornstein NM, Assayag EB. Depressive symptoms following stroke and transient ischemic attack: is it time for a more intensive treatment approach? Results from the TABASCO cohort study. J Clin Psychiatry. 2016;77:673–680.

36. Levine DA, Haan MN, Langa KM, Morgenstern LB, Neuhaus J, Lee A, Lisabeth LD. Impact of gender and blood pressure on poststroke cognitive decline among older Latinos. J Stroke Cerebrovasc Dis. 2013;22:1038–1045.

37. Ghosal MK, Burman P, Singh V, Das S, Paul N, Ray BK, Hazra A, Banerjee TK, Basu A, Chaudhuri A, Das SK. Correlates of functional outcome among stroke survivors in a developing country—a prospective community-based study from India. J Stroke Cerebrovasc Dis. 2014;23:2614–2621.

38. Comijs HC, Kriegsman DM, Dik MG, Deeg DJ, Jonker C, Stalman WA. Somatic chronic diseases and 6-year change in cognitive functioning among older persons. Arch Gerontol Geriatr. 2009;48:191–196.

39. Wagle J, Farner L, Flekkoy K, Wyller TB, Sandvik L, Eiklid KL, Fure B, Stensrod B, Engedal K. Cognitive impairment and the role of the apoe epsilon4-allele after stroke–a 13 months follow-up study. Int J Geriatr Psychiatry. 2010;25:833–842.

40. Dik MG, Deeg DJ, Bouter LM, Corder EH, Kok A, Jonker C. Stroke and apolipoprotein E epsilon4 are independent risk factors for cognitive decline: a population-based study. Stroke. 2000;31:2431–2436.

41. Rowan EN, Dickinson HO, Stephens S, Ballard C, Kalaria R, Kenny RA. Homocysteine and post-stroke cognitive decline. Age Ageing. 2007;36:339–343.

42. Douiri A, Grace J, Sarker SJ, Tilling K, McKevitt C, Wolfe CD, Rudd AG. Patient-specific prediction of functional recovery after stroke. Int J Stroke. 2017 Jul;12(5):539-548. doi: 10.1177/1747493017706241. Epub 2017 Apr 25. PMID: 28440112.

43. Abdel Douiri , Christopher McKevitt, et al. (2013). Long-Term Effects of Secondary Prevention on Cognitive Function in Stroke Patients. Circulation, 128(12):1341-1348 10.1161/CIRCULATIONAHA.113.002236
44. Nasreddine, Z.S., Phillips, N.A., Bédirian, V., et al. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. Journal of the American Geriatrics Society, 53: 695-699.

Table1: Baseline post stroke characteristics of patients including sociodemographic, past medical history, case mix, and stroke subtypes

	Development cohort (1995-2010)		**Validation cohort (2011-2018)**	
Cognitive Impairment	**Intact (%)** 1468	**Impaired (%)** 1000	**Intact (%)** 736	**Impaired (%)** 204
Age, mean (SD)	66.84 (14.61)	74.10 (12.90)	69.60 (15.40)	70.50 (15.31)
Sex				
female	824 (56.13%)	540 (54%)	279 (38%)	98 (48.04%)
Male	644 (43.87%)	460 (46%)	457 (62%)	106 (51.96%)
Ethnicity				
White	1044 (71.12%)	732 (73.2%)	412 (56%)	95 (46.57%)
Black	346 (23.57%)	219 (21.9%)	278 (38%)	95 (6.37%)
Other	63 (4.29 %)	44 (4.4%)	43 (0.6%)	13 (6.37%)
Missing	15 (1.02%)	5 (0.5%)	3 (0.41%)	1 (0.5%)
Socioeconomic group				
Manual	831 (56.61%)	621 (62.1%)	218 (29.62%)	59 (28.92%)
Non-manual	533 (36.31%)	207 (20.7%)	222 (30.16%)	47 (23.04%)
Unknown	2 (0.14%)	2 (0.2%)	1 (0.14%)	0 (0%)
Missing	102 (6.95%)	170 (17%)	295 (40.08%)	98 (48.04%)
Pre-stroke vascular risk factors				
Transient ischemic attack				
No	1285 (87.53%)	858 (85.8%)	666 (90.5%)	179 (87.75%)
Yes	173 (11.78%)	129 (12.9%)	57 (7.74%)	22 (10.78%)
Missing	10 (0.68%)	13 (0.13%)	13 (1.77%)	3 (1.47%)
Atrial fibrillation				

No	1310 (89.24%)	787 (78.7%)	602 (81.79%)	156 (76.47%)
Yes	148 (10.08%)	198 (19.8%)	115 (15.63%)	44 (21.57%)
Missing	10 (0.68%)	15 (0.15%)	19 (2.58%)	4 (1.96%)
Hypertension				
No	517 (35.22 %)	301 (30.1%)	245 (33.29%)	62 (30.39%)
Yes	944 (64.31 %)	691 (69.1%)	484 (65.76%)	140 (68.63%)
Missing	7 (0.48%)	8 (0.8%)	7 (0.95%)	2 (1%)
Diabetes mellitus				
No	1188 (80.93%)	775 (77.5%)	540 (73.37%)	133 (65.20%)
Yes	271 (18.46%)	217 (21.7%)	189 (25.68%)	67 (32.84%)
Missing	9 (0.61%)	8 (0.8%)	7 (0.95%)	4 (2%)
Hypercholesterolemia				
No	895 (60.97%)	512 (51.2%)	428 (58,15%)	113 (55.4%)
Yes	310 (21.12%)	163 (16.3%)	298 (40.5%)	86 (42.16%)
Unknow	263 (17.92%)	325 (32.5%)	10 (1.36%)	5 (2.45%)
Current Smoker				
No	491 (33.45%)	365 (36.5%)	306 (41.60%)	90 (44.12%)
Yes	481 (32.77%)	308 (30.8%)	246 (33.42%)	59 (28.92%)
Unknown	473 (32.22%)	270 (27%)	180 (24.46%)	48 (23.53%)
Missing	23 (1.57%)	57 (0.57%)	4 (0.54%)	7 (3.43%)
Drinker				
No	504 (34.33%)	399 (39.90%)	336 (45.65%)	120 (58.82%)
Yes	931 (63.41%)	531 (53.10%)	392 (53.26%)	79 (38.72%)
Missing	33 (2.24%)	70 (7.00%)	8 (1.08%)	5 (2.45%)
Antiplatelet prior to stroke				
No	1089 (74.18%)	735 (73.50%)	686 (93.20%)	177 (86.76%)
Yes	209 (14.23%)	181 (18.10%)	38 (5.16%)	23 (11.27%)
Missing	170 (11.58%)	84 (8.40%)	12 (1.63%)	4 (1.96%)
Family history of stroke				
No	161 (10.96%)	76 (7.60%)	430 (58.42%)	133 (65.19%)
Yes	114 (7.76%)	30 (3.00%)	242 (32.88%)	33 (16.17%)
Missing	1193 (81.26%)	894 (89.40%)	64 (8.69%)	38 (18.63%)
Stroke severity (Case-mix)				
Glasgow coma scale (GCS)				
Severe (<8)	21 (1.43%)	69 (0.69%)	6 (0.82%)	8 (3.92%)
Moderate (9-12)	38 (2.59%)	173 (17.3%)	31 (4.21%)	25 (12.25%)
Mild (13-15)	1375 (93.66%)	739 (73.9%)	678 (92.12%)	162 (79.41%)
Missing	34 (2.32%)	19 (0.19%)	21 (2.85%)	9 (4.41%)
Urinary incontinence				
No	1180 (80.38%)	424 (42.4%)	633 (86%)	130 (63.73%)
Yes	251 (17.10%)	555 (55.5%)	82 (11.14%)	67 (32.84%)
Missing	37 (2.52%)	21 (0.21%)	21 (2.85%)	7 (3.43%)
Stroke subtype				

Infarct	1286 (87.60%)	817 (81.7%)	629 (85.46%)	179 (87.75%)
Haemorrhagic	169 (11.51%)	161 (16.1%)	106 (14.40%)	25 (12.25%)
Missing	13 (0.89%)	22 (0.22%)	1 (0.14%)	0 (0%)

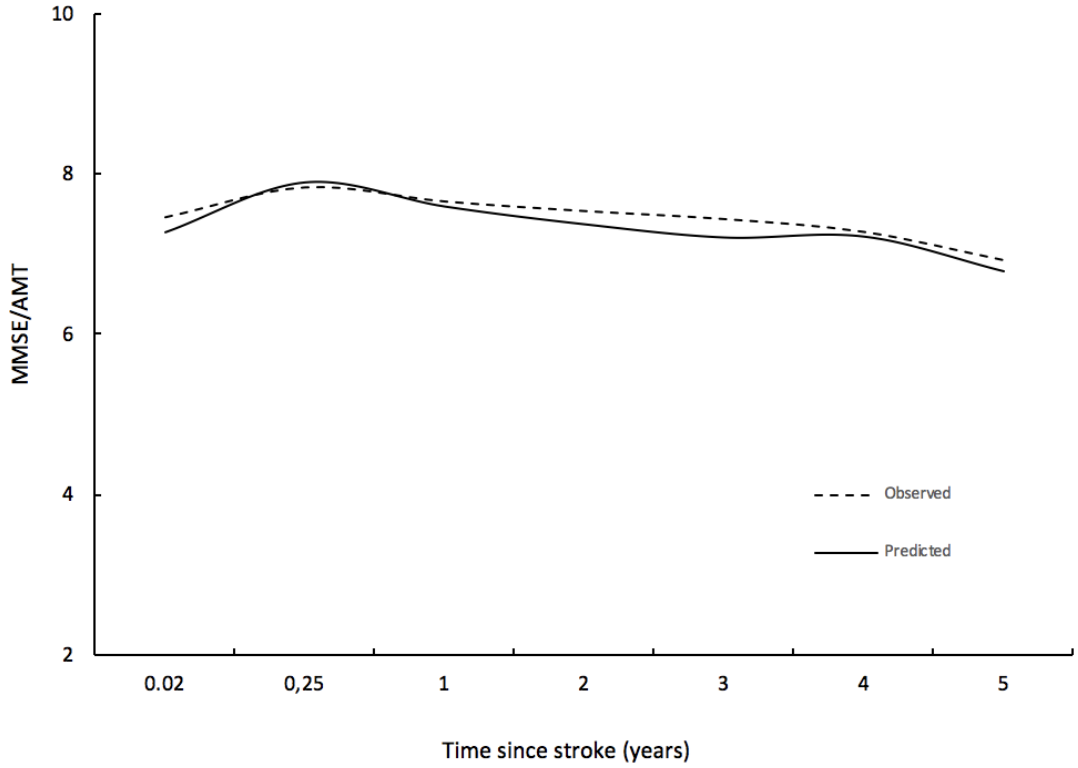
Table 2: predictive values and likelihood rations for classifying each cognitive impairment score of interest.

Measure	**3 months**	**1 year**	**5years**
(cut-off=4)			
Prevalence	10% [8- 12]	9% [7- 11]	6% [4-7]
Overall prognostic performance			
Overall performance (Brier)	7%	7%	8%
Discrimination (AUC)	88.5% [85-90]	89.6% [86-92]	87% [85-91]
Prognostic performance at a cut-off			
Sensitivity	62% [52-71]	58% [48-68]	59% [46-71]
Specificity	93% [91- 94]	92% [90-94]	90% [88-92]
Clinical utility at cut-off			
PPV	49% [41- 58]	42% [33-50]	27% [20- 35]
NPV	96% [94-97]	96% [94- 97]	97% [96 -98]
LR+	8.75 [6.68 -11.46]	7.29 [5.57-9.54]	6.10 [4.61-8.05]
LR-	0.41 [0.32- 0.52]	0.45 [0.36 - 0.57]	0.46 [0.34- 0.61]
DOR	21.30 [13.51- 33.57]	16.03 [10.10 - 25.45]	13.35 [7.76- 22.96]
Youden	0.54 [0.43-0.65]	0.50 [0.38 - 0.62]	0.50 [0.34 - 0.63]
(cut-off=8)			
Prevalence	32% [29-35]	39% [36-42]	42% [39- 45]
Overall prognostic performance			

Overall performance (Brier)	17%	19%	20%
Discrimination (AUC)	80% [76-81]	77% [73-78]	75% [72-78]
Prognostic performance at a cut-off			
Sensitivity	78% [73-82]	74% [70-78]	72% [68-76]
Specificity	72% [68-75]	65% [61-68]	65% [61-69]
Clinical utility at cut-off			
PPV	56% [52-61]	58% [53 - 62]	60% [55-64]
NPV	87% [85-90]	79% [76 - 83]	76% [73- 80]
LR+	2.75 [2.42 - 3.12]	2.10 [1.87 - 2.36]	2.05 [1.82 - 2.31]
LR-	0.31 [0.25 - 0.37]	0.40 [0.34 - 0.47]	0.43 [0.37 - 0.51]
DOR	8.99 [6.67-12.11]	5.24 [4.01 - 6.84]	4.74 [3.66 - 6.15]
Youden	0.50 [0.41- 0.57]	0.39 [0.31- 0.46]	0.37[0.29 -0.47]

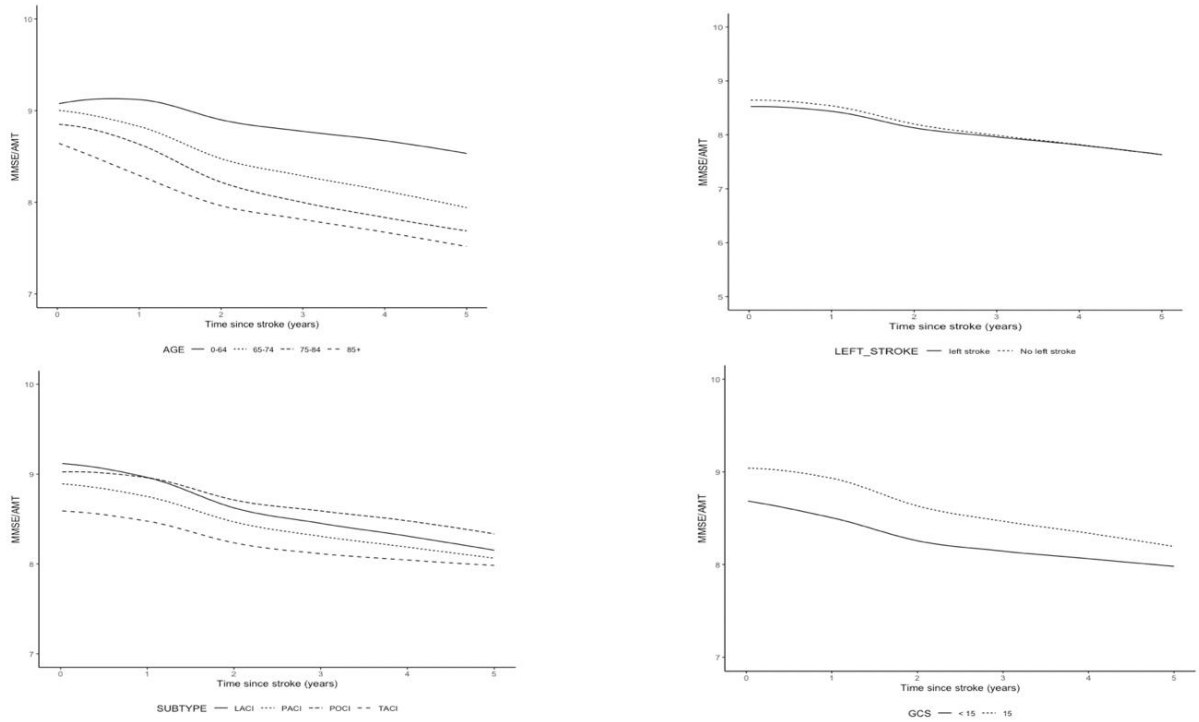
AUC : area under the curve; LR : likelihood ratio; DOR : diagnostic odds ratio; NPV : negative predictive value; PPV : positive predictive value.

Figure 1. Average predicted trajectories compared to the average observed cognition score after stroke up to 5 years



705
706
707

Figure 2. Average predicted recovery patterns after stroke stratified by age, stroke subtype, GCS, and left-stroke. GCS: Glasgow coma score.



708

Figure 3. Decision curves for recovery curves to predict mild cognitive impairment in stroke survivors at three months, 1 and 5 years. Red line: Prediction model. Grey line: assume all are cognitively impaired. Black line: assume all are not cognitively impaired.

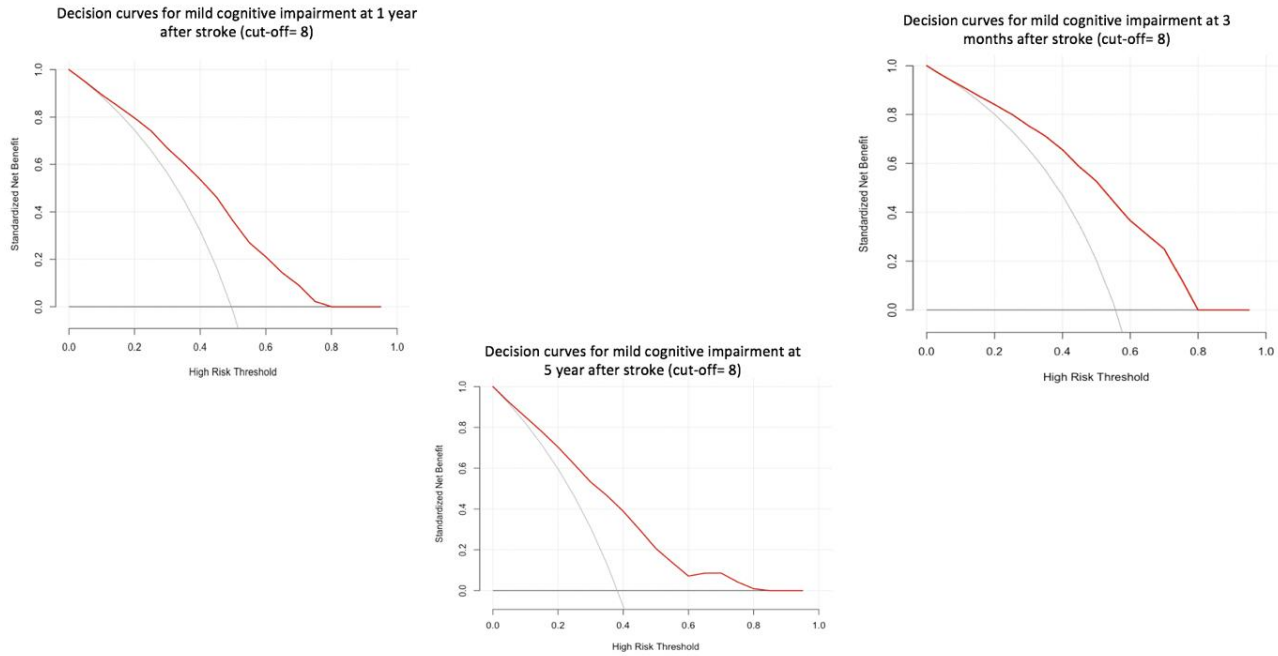


Figure legends

Figure1. Average predicted trajectories compared to the average observed cognition score after stroke up to 5 years

Figure 2. Average predicted recovery patterns after stroke stratified by age, stroke subtypes, GCs, and left-stroke. GCS: Glasgow coma score.

Figure 3. Decision curves for recovery curves to predict mild cognitive impairment in stroke survivors at three months, 1 and 5 years. Red line: Prediction model. Grey line: Assume all are cognitively impaired. Black line: assume all are not cognitively impaired.

List of Tables:

Table1: Baseline post stroke characteristics of patients including sociodemographic, past medical history, case mix, and stroke subtypes

Table 2: predictive values and likelihood rations for classifying each cognitive impairment score of interest.