Brain Health Scores to Predict Neurological Outcomes from Electronic Health Records

Marta Fernandes, PhD[a,b,c], **Haoqi Sun, PhD**[a,b,c], **Zeina Chemali, MD, MPH**[a,b,d], **Shibani S. Mukerji, MD, PhD**[a,b], **Lidia M.V.R. Moura, MD, MPH, PhD**[a,b], **Sahar F. Zafar, MD**[a,b], **Akshata Sonni, PhD**[a,b,d], **Alessandro Biffi, MD**[a,b,d,e], **Jonathan Rosand, MD**[a,b,d,e], **M. Brandon Westover, MD, PhD**[a,b,c,d]

[a.] Department of Neurology, Massachusetts General Hospital (MGH), Boston, MA United States

[b.] Harvard Medical School, Boston, MA, United States

[c.] Clinical Data Animation Center (CDAC), MGH, Boston, MA, United States

[d.] Henry and Allison McCance Center for Brain Health, MGH, Boston MA, United States

[e.] Broad Institute of MIT and Harvard, Cambridge, MA, United States

Abstract

Background: Preserving brain health is a critical priority in primary care, yet screening for these risk factors in face-to-face primary care visits is challenging to scale to large populations. We aimed to develop automated brain health risk scores calculated from data in the electronic health record (EHR) enabling population-wide brain health screening in advance of patient care visits.

Methods: This retrospective cohort study included patients with visits to an outpatient neurology clinic at Massachusetts General Hospital, between January 2010 and March 2021. Survival analysis with an 11-year follow-up period was performed to predict the risk of intracranial hemorrhage, ischemic stroke, depression, death and composite outcome of dementia, Alzheimer's

Corresponding Author: Marta Fernandes, mbentofernandes@mgh.harvard.edu, Postal address: 55 Fruit Street, Boston MA 02114, United States. Telephone number: +1 857-316-9517.

disease, and mild cognitive impairment. Variables included age, sex, vital signs, laboratory values, employment status and social covariates pertaining to marital, tobacco and alcohol status. Random sampling was performed to create a training (70%) set for hyperparameter tuning in internal 5-fold cross validation and an external hold-out testing (30%) set of patients, both stratified by age. Risk ratios for high and low risk groups were evaluated in the hold-out test set, using 1000 bootstrapping iterations to calculate 95% confidence intervals (CI).

Results: The cohort comprised 17040 patients with an average age of 49 ± 15.6 years; majority were males (57%), White (78%) and non-Hispanic (80%). The low and high groups average risk ratios [95% CI] were: intracranial hemorrhage 0.46 [0.45-0.48] and 2.07 [1.95-2.20], ischemic stroke 0.57 [0.57-0.59] and 1.64 [1.52-1.69], depression 0.68 [0.39-0.74] and 1.29 [0.78-1.38], composite of dementia 0.27 [0.26-0.28] and 3.52 [3.18-3.81] and death 0.24 [0.24-0.24] and 3.96 [3.91-4.00].

Conclusions: Simple risk scores derived from routinely collected EHR accurately quantify the risk of developing common neurologic and psychiatric diseases. These scores can be computed automatically, prior to medical care visits, and may thus be useful for large-scale brain health screening.

1. Introduction

Brain disease affects 1 in 6 people [1]. Effective prevention is vital, especially in senior populations [2]. Brain health includes cognitive, motor, emotional, and sensory functions[3]. Pillars of brain health include lifestyle choices, diet and nutrition habits, physical and mental exercise, sleep and relaxation, engaging socially, learning new skills, and stress management [3–5].

The most common diseases affecting brain health in the aging population include ischemic and hemorrhage stroke [6–9], Alzheimer's disease and other dementias [3], and depression [10, 11]. Brain disease prevention should start during primary care visits. However, utilizing face-to-face visits to assess brain health in large patient populations is challenging. There is an unmet need for a scalable equitable approach to screening. We sought to develop a series of brain health scores that quantify the risk of developing intracranial hemorrhage, ischemic stroke, Alzheimer's disease, or mild cognitive impairment (MCI), depression, or death within the next 11 years. The risk scores leverage variables captured in electronic health records (EHR) during patient encounters. Our intention is for these risk scores to serve as screening tools to identify patients who may benefit from lifestyle risk mitigation and engage in face-to-face brain health visits to improve their modifiable risk factors to improve brain health.

2.   Methods

2.1.   Study Cohort

This study follows STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines [12]. EHR data was extracted under a protocol approved by Institutional Review Board under a waiver of informed consent. Data were analyzed for 17040 adults (age 18) who visited an outpatient neurology clinic (the Sleep Laboratory) at MGH from January 1st, 2010 to March 31st, 2021. These patients are generally referred by general practitioners but can also be referred by other doctors (e.g. a neurologist). Most are referred to evaluate possible sleep apnea. This cohort was selected because it is large and had a well-characterized set of neurologic and psychiatric outcomes from prior studies [13]. Patients with a first encounter less than 24 hours prior to outcome events were excluded. Deceased patients for whom date of death was not available were excluded.

2.2.   Neurological outcomes and comorbidities

Study outcomes included intracranial hemorrhage, ischemic stroke, depression, death and a composite of dementia, Alzheimer disease and MCI ("composite of dementia"). Outcomes and comorbidities were ascertained based on ICD codes and medications (Tables A.1, A.2), with additional information extracted from medical notes. Criteria to identify dementia and MCI were published previously [13]. Patients had to have at least one relevant ICD code to be assigned outcomes of intracranial hemorrhage or ischemic stroke. Patients had to have at least one relevant ICD code and one medication to be assigned outcomes of depression, diabetes or hypertension.

A patient diagnosed with any of the study outcomes, except death was assigned the corresponding comorbidity for all subsequent encounters. When predicting a given outcome (e.g. ischemic stroke), existing medical conditions were considered as possible predictors.

2.3.   Study covariates

Additional covariates were outpatient vital signs and laboratory values, including systolic and diastolic blood pressure, temperature, heart rate and respiration rate; hemoglobin A1C, alanine transaminase (ALT), aspartate aminotransferase (AST), high-density lipoprotein (HDL), low-density lipoprotein (LDL) and albumin. We created binary indicators for absence of vital sign and laboratory values. Body mass index (BMI), age, sex and employment status were included as covariates. Employment was coded as "Active = 1" for professional/student full/part-time, self-employed, homemaker, or active military duty; and "Active = 0" for disabled, retired, not employed, or unknown status. Social covariates pertaining to marital, tobacco and alcohol status were included. Marital status was coded as "Union = 1" for those married/in civil union; "Union = 0" otherwise. Tobacco status was coded "Smoker = 1" with active smoking status, otherwise as 0. Alcohol was recorded as number of drinks per week.

2.4.   Competing risk survival analysis

Time-to event was calculated as time between the encounter and outcome, loss to follow-up or death. Encounters were assigned one of the following status: non-censored ('1')

experienced the event of interest; (right) censored ('0') if lost to follow up or did not experience the event of interest during the study period; and, for competing risk analysis, patients who died without experiencing the event of interest were assigned a status of dead ('2'). The common encounters for all study outcomes problems were used as training data for the missing data imputation task. Covariate measurements in this dataset were similar to those in the full dataset with all encounters (Fig. A.1).

2.5. Outliers' processing and missing data imputation

Values outside physiological ranges were removed, followed by normalization (A.1. Data Preprocessing). Missing data was imputed using Multiple Imputation by Chained Equations (MICE) [14], where an estimator was developed for each covariate with missing data based on remaining covariates. Mean square error (MSE) was estimated on the training dataset using 5-fold cross validation. The estimator showing lowest MSE in the imputation task was selected for imputation of both train and test sets. The *ExtraTreesRegressor* [15] an estimator that fits a number of randomized decision trees on various subsamples of the dataset and uses averaging to improve the predictive accuracy, consistently achieved lowest MSE (Fig. A.2)[1].

2.6. Modeling design

Stratified random sampling was performed to create train (70%) and test (30%) sets. Patients were split into age strata by quantiles: 25% (18 to 47 years), 50% (48 to 60), 75% (61 to 71), and 99% (> 71 years) based on maximum age during the study. After obtaining train and test splits, outliers were removed, data was normalized [16], and missing values were imputed.

For model development, one random encounter was selected per patient. We trained a regularized Cox regression model with elastic net penalty [17] using 5-fold cross validation [18, 19] using Harrell's C-index [20] to evaluate model performance. Cox regression [21] was used because it is an interpretable model, designed for predicting time to an outcome while handling right censoring in a natural way. The full training set was used to train a final model using the hyperparameter values that yielded best training performance. With the covariates selected by the elastic net, we trained a Cox proportional hazards model [21] with competing risk. For prediction of death, we applied the same methodology with no competing risk.

2.7. Performance evaluation

We assessed risk ratios (RR) for higher and lower risk groups, by adding and subtracting the standard deviation (SD) to all patients' average risk probability, respectively, and dividing by average risk probability in the test set. RR of 1 signifies that two groups have the same risk, while results not equal to 1 indicate that one group is at more risk. Harrell's C-index [20] was used as complementary performance metric, as well as the cumulative dynamic area under the receiver operating curve (AUC) [22–24], true positive rate (TPR) and false positive rate (FPR). We performed 1000 bootstrapping iterations with a random selection of patient

---

[1]We made our code for data preprocessing and modeling publicly available in [https://github.com/mpriscila88/competing_risks_survival_analysis].

encounters in the hold-out test (30%) set, which was not used to make decisions about model training/validation or hyperparameter tuning, to calculate 95% confidence intervals (CI). We plotted both empirical cumulative incidence risk curves (Aalen-Johansen estimator [25]), and risk curves from the Cox proportional hazards model. Finally, we evaluated covariate importance for each outcome model by assessing Cox models coefficients.

3. Results

3.1. Patient characteristics

Our cohort comprised 17040 patients after applying inclusion and exclusion criteria (A.3. Supplementary results). Patients were predominantly male (57%), White (78%) non-smokers (87%) with average baseline age of 49 years (Table 1). The baseline average age for patients with depression (47 years) was below the cohort average, followed by diabetes, hypertension, ischemic stroke and intracranial hemorrhage, with 54, 56, 57 and 58 years, respectively. For patients with the composite of dementia and death outcomes, the average age at baseline was 64 years. the distributions of age at baseline according to positive or negative outcome is in Table A.3. Approximately equal numbers of patients were actively and non-actively employed (43% vs 40%). 57% were missing alcohol status; among those with data, 27% consumed alcohol with an average of 2.4 drinks per week (Table A.4). Most patients (63%) did not experience any study outcome events. Characteristics of training and test sets were comparable (Table A.5.).

3.2. Modeling performance

We randomly selected one encounter per patient for train (N=11928) and test (N=5112) sets. Survival times for each outcome are shown in Fig. A.3. The modeling average concordance index and risk ratios for lower and higher risk groups are presented in Table 2. Yearly risk ratios in 95% CI are shown in Table A.6.

The average C-index in train and test sets were comparable and CIs were narrow in the 3% range for the unbalanced outcome events, indicating the absence of any significant overfitting (see Table 2). The difference between lower and higher risk patients was more accentuated for the composite outcome of dementia and death. The difference between groups was more attenuated for depression. With selection of balanced outcome events in test we observed wider CIs overall. Since the model was trained with an unbalanced distribution of the data, the model is less confident when evaluated in a balanced set of outcome events. To balance outcomes, the number of patients from each class had to be the same, and one random encounter was selected in each bootstrapping from each class. For each outcome (number of patients in test for positive class): intracranial hemorrhage (n = 35), composite of dementia (n = 251), ischemic stroke (n = 604), depression (n = 774) and death (n = 272). For intracranial hemorrhage we observed a decrease of 18% in C-index. There was 16% and 14% decrease in C-index for the composite of dementia and death, respectively.

The RR were similar between balanced and unbalanced events, except for death, where the difference between groups was more accentuated for balanced events, however coupled with wider CIs.

We trained and evaluated a baseline model with only age and sex as covariates, presented in Table A.7., where we observed a 3% decrease in C-index for intracranial hemorrhage, composite of dementia and death, a 2% decrease for depression, and a 1% decrease for ischemic stroke, in the testing set.

We also assessed the cumulative dynamic AUC in Table 3, with the competing risk coded as "no event", since this dynamic AUC metric receives outcome events as binary. We performed 200 bootstrapping iterations to calculate 95% CI. We observed that the dynamic AUC exhibited slight oscillations in the range 2% to 5%, mean TPR from 2% to 4% and mean FPR from 2% to 4%, with the exception of the composite of dementia outcome, where there was an 8% increase of mean FPR from the 1st to the 10th year.

The cumulative risk for both the Cox proportional hazards and the Aalen-Johansen models is presented in Figs. 1 and 2 for patients in the test set. The cumulative incidence risk for patients in the lower and higher risk groups was well defined and separated for all outcomes. The empirical (Aalen-Johansen) risk curves were similar to those of the Cox proportional hazards model, and remained parallel, indicating acceptable fits and calibration of the Cox models.

We assessed the distribution of the cumulative incidence risk scores for each outcome (Fig. A.4). Patients who did not experience the outcome of interest tended to have overall lower risk scores compared to those who experienced the event or those who died.

3.3. Modeling covariates' importance

Covariate importance for each study outcome is presented in Figs. 3 and 4 and Table A.8. With the exception of depression, age was a positive risk factor for all study outcomes.. Female sex conferred higher risk for the composite of dementia, ischemic stroke ($p<0.1$) and depression ($p<0.001$), while male sex conferred higher risk for intracranial hemorrhage ($p<0.001$). Being active (positive employment status) was a consistent predictor for decreased risk across all outcomes, except depression.

For intracranial hemorrhage, risk factors included higher age ($p<0.01$) and respiratory rate, and not being active ($p<0.05$).

For the composite of dementia, Alzheimer's disease and MCI, risk factors included higher systolic blood pressure ($p<0.001$), A1C ($p<0.001$) and heart rate, older age ($p<0.001$), lower HDL ($p<0.01$) and ALT ($p<0.001$), and not being active ($p<0.001$).

For ischemic stroke, risk factors included older age ($p<0.001$) and higher A1C ($p<0.05$), not being active ($p<0.001$) and lower temperature ($p<0.001$).

For depression, risk factors included prior ischemic stroke, composite of dementia, Alzheimer's disease, MCI, female sex, younger age, and non-union. All covariates showed statistical significance with a p<0.001.

For death, risk factors included older age (p<0.001), higher heart rate (p<0.001), ischemic stroke (p<0.001), hypertension (p<0.05), intracranial hemorrhage, dementia (p<0.05), diabetes, lower HDL (p<0.01) and inactivity (p<0.001).

4. Discussion

4.1. Principal Findings

We found that the brain health risk scores obtained from routinely acquired data and easily calculated from existing electronic health records data were associated with neurological outcomes: higher risk scores corresponded with higher risk of developing brain disease.

Age and active employment played a significant role in prediction of nearly all outcomes. Older age was related with higher risk for intracranial hemorrhage, ischemic stroke, composite outcome of dementia, Alzheimer disease and MCI and death in our cohort. This is in keeping with previous studies [26–28] analyzing the global burden of major neurological disorders, including stroke, Alzheimer's disease and other dementias, which have shown that neurological disorders have been increasing in recent years, largely because of the aging of the population. Being employed usually requires learning new skills, maintaining a routine, and engaging socially, which may promote cognitive health, especially as one ages [29, 30]. Employment is also an activity where people are often faced with complex tasks. It is well established that engaging in complex activities in one's environment may help prevent age-associated cognitive decline and dementia [31, 32], by facilitating brain health and optimal cognitive functioning [33].

The identification of these risk indicators is vital to alert the population of important factors that might help keep their brains healthy and fit. Patients can keep track of their progress at each subsequent encounter in modifying their lifestyle choices and risk of brain disease. The brain health scores developed in this work are similar to the American Heart Association's (AHA) "Life's Simple 7" [34] premise, where patients are encouraged to adjust their lifestyle choices so that they can monitor their brain health scores and thus improve their brain health over time. "Life's Simple 7" [34] defines ideal cardiovascular health based on physical activity, healthy diet, smoking status, body mass index, total cholesterol, blood pressure and fasting blood glucose.

The brain health scores developed in this work can be calculated using the covariates which showed the most importance in the 11-year period survival prediction problems for each of the neurological outcomes. Those covariates were patients' age and sex, employment status, marital status, vital signs, including respiration rate, systolic blood pressure, heart rate, and temperature, laboratory values, including A1C, ALT and HDL, body mass index, diabetes and hypertension. The majority of these measures have been robustly associated with cognitive decline in epidemiological studies [35].

4.2. Comparison with prior work

Previous work has been done to create a brain care score, the McCance Brain Care Score (BCS) (see Fig. A.5) which can be calculated for an in-person encounter. The McCance BCS is based on a 17-point system where points are assigned to each of the measures in the physical, lifestyle and social and emotional categories, that sum up to give the total BCS for the patient. The McCance BCS was derived from evidence-supported interventions associated with reduced risk of brain diseases, building upon and expanding AHA "Life's Simple 7" and recommendations from the Alzheimer Association. The score has been piloted by the Henry and Allison McCance Center for Brain Health and iteratively refined and improved in outpatient practice. Our model performance is not directly comparable to the McCance BCS, because the EHR does not have information on several of the variables included in the score, such as social relationships or meaning of life.

4.3. Limitations

This study was performed for a cohort of patients who visited the Sleep Laboratory at MGH and may not be representative of other US and non-US populations, limiting the generalizability of the models across populations and hospital settings. The majority of patients were identified as White, and minorities are misrepresented on our dataset. This cohort was selected because it is large and had a well-characterized set of neurologic and psychiatric outcomes from prior studies by the authors [13], however there may be differences in insurance for patients who visit the Sleep Laboratory and those without insurance who visit other neurology clinics. This cohort may also have specific characteristics, such as low yearly incidence of hypertension, considering as reference the US population. Thus, we lack a study of generalizability to understand the importance of tailoring the models to different populations so that these may benefit from brain health scores assessment. Another limitation consisted of high missingness of laboratory values, which we tackled with a robust approach for missing data imputation. Also, our model only considers social covariates last registered in the system, thus a model that considers changes in social covariates each year should be considered in future studies.

4.4. Conclusions

A simple risk score derived from routinely collected data, easily acquired in a patient encounter, is associated with risk of neurological outcomes and death. This approach automatically uses EHR data, which makes it suitable for large-scale population screening. This enables the implementation of a very low-cost neurological screening tool for prevention of brain disease across healthcare systems. By adopting the approach of the chronic disease management model [36], the risk scores may empower especially younger to mid-life patients to make different lifestyle choices to improve their brain health and prepare for aging.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. American Brain Foundation. Brain Diseases. Accessed January 25, 2022. https://www.americanbrainfoundation.org/diseases/

2. Centers for Disease Control and Prevention. Promoting Health for Older Adults | CDC. Published January 24, 2022. Accessed January 25, 2022.https://www.cdc.gov/chronicdisease/resources/publications/factsheets/promoting-health-for-older-adults.htm

3. National Institute on Aging. Cognitive Health and Older Adults. Accessed January 25, 2022. http://www.nia.nih.gov/health/cognitive-health-and-older-adults

4. Cleveland Clinic. 6 Pillars of Brain Health. Healthy Brains by Cleveland Clinic. Accessed January 25, 2022. https://healthybrains.org/pillars/

5. Harvard Medical School and Harvard T.H. Chan School of Public Health. 12 ways to keep your brain young. Harvard Health. Published June 1, 2006. Accessed January 25, 2022. https://www.health.harvard.edu/mind-and-mood/12-ways-to-keep-your-brain-young

6. American Stroke Association. About Stroke, www.stroke.org. Accessed January 25, 2022. https://www.stroke.org/en/about-stroke

7. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart disease and stroke statistics—2021 update: a report from the American Heart Association. Circulation.2021;143:e254–743. [PubMed: 33501848]

8. Centers for Disease Control and Prevention,National Center for Health Statistics. Multiple Cause of Death, 1999-2019. CDC WONDER Online Database website. Atlanta, GA: Centers for Disease Control and Prevention; 2019. Accessed January 25, 2022. https://wonder.cdc.gov/mcd-icd10.html

9. American Heart Association. 2021 Heart Disease and Stroke Statistics Update Fact Sheet At-a-Glance. Published online 2021. Accessed January 25, 2022. https://www.heart.org/-/media/PHD-Files-2/Science-News/2/2021-Heart-and-Stroke-Stat-Update/2021_heart_disease_and_stroke_statistics_update_fact_sheet_at_a_glance.pdf

10. World Health Organization. Depression. Accessed January 25, 2022. https://www.who.int/news-room/fact-sheets/detail/depression

11. Institute of Health Metrics and Evaluation. GBD Results Tool | GHDx. Accessed March 5, 2022. http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/d780dffbe8a381b25e1416884959e88b

12. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. PLOS Med. 2007;4(10):e297. doi:10.1371/journal.pmed.0040297 [PubMed: 17941715]

13. Ye E, Sun H, Leone MJ, et al. Association of Sleep Electroencephalography-Based Brain Age Index With Dementia. JAMA Netw Open. 2020;3(9):e2017357. doi:10.1001/jamanetworkopen.2020.17357 [PubMed: 32986106]

14. Raghunathan T, Lepkowski J, Hoewyk J, Solenberger P. A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. Surv Methodol. 2001, 27.1: 85–96.

15. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Machine learning, 2006, 63.1: 3–42.

16. Box GEP, Cox DR. An Analysis of Transformations. J R Stat Soc Ser B Methodol. 1964;26(2):211–252.

17. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol. 2005;67(2):301–320. doi:10.1111/j.1467-9868.2005.00503.x

18. Adeoye J, Hui L, Koohi-Moghadam M, Tan JY, Choi SW, Thomson P. Comparison of time-to-event machine learning models in predicting oral cavity cancer prognosis. Int J Med Inf. 2022;157:104635. doi:10.1016/j.ijmedinf.202L104635

19. Murtojärvi M, Halkola AS, Airola A, et al. Cost-effective survival prediction for patients with advanced prostate cancer using clinical trial and real-world hospital registry datasets. Int J Med Inf. 2020;133:104014. doi:10.1016/j.ijmedinf.2019.104014

20. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA. 1982;247(18):2543–2546. [PubMed: 7069920]

21. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society: Series B (Methodological). 1972;34(2):187–220. doi:10.1007/978-1-4612-4380-9_37

22. Uno H, Cai T, Tian L, Wei LJ. Evaluating Prediction Rules for t-Year Survivors with Censored Regression Models. J Am Stat Assoc. 2007;102(478):527–537.

23. Hung H, Chiang CT. Estimation methods for time-dependent AUC models with survival data. Can J Stat Rev Can Stat. 2010;38(1):8–26.

24. Lambert J, Chevret S. Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves. Stat Methods Med Res. 2016;25(5):2088–2102. doi:10.1177/0962280213515571 [PubMed: 24395866]

25. Aalen OO, Johansen S. An Empirical Transition Matrix for Non-Homogeneous Markov Chains Based on Censored Observations. Scand J Stat. 1978;5(3):141–150.

26. GBD 2017 US Neurological Disorders Collaborators. Burden of Neurological Disorders Across the US From 1990-2017: A Global Burden of Disease Study. JAMA Neurol. 2021;78(2):165–176. doi:10.1001/jamaneurol.2020.4152 [PubMed: 33136137]

27. Feigin VL, Nichols E, Alam T, et al. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2019;18(5):459–480. doi:10.1016/S1474-4422(18)30499-X [PubMed: 30879893]

28. Hou Y, Dan X, Babbar M, et al. Ageing as a risk factor for neurodegenerative disease. Nat Rev Neurol. 2019;15(10):565–581. doi:10.1038/s41582-019-0244-7 [PubMed: 31501588]

29. Vance DE, Bail J, Enah CC, Palmer JJ, Hoenig AK. The impact of employment on cognition and cognitive reserve: implications across diseases and aging. Nurs Res Rev. 2016;6:61–71. doi:10.2147/NRR.S115625

30. Harada CN, Natelson Love MC, Triebel K. Normal Cognitive Aging. Clin Geriatr Med. 2013;29(4):737–752. doi:10.1016/j.cger.2013.07.002 [PubMed: 24094294]

31. Fratiglioni L, Paillard-Borg S, Winblad B. An active and socially integrated lifestyle in late life might protect against dementia. Lancet Neurol. 2004;3(6):343–353. doi:10.1016/S1474-4422(04)00767-7 [PubMed: 15157849]

32. Marioni RE, van den Hout A, Valenzuela MJ, Brayne C, Matthews FE, MRC Cognitive Function and Ageing Study. Active cognitive lifestyle associates with cognitive recovery and a reduced risk of cognitive decline. J Alzheimers Dis JAD. 2012;28(1):223–230. doi:10.3233/JAD-2011-110377 [PubMed: 21971400]

33. Park DC, Lodi-Smith J, Drew L, et al. The Impact of Sustained Engagement on Cognitive Function in Older Adults: The Synapse Project. Psychol Sci. 2014;25(1):103–112. doi:10.1177/0956797613499592 [PubMed: 24214244]

34. Lloyd-Jones DM, Hong Y, Labarthe D, et al. Defining and Setting National Goals for Cardiovascular Health Promotion and Disease Reduction. Circulation. 2010;121(4):586–613. doi:10.1161/CIRCULATIONAHA.109.192703 [PubMed: 20089546]

35. Lazar RM, Howard VJ, Kernan WN, et al. A Primary Care Agenda for Brain Health: A Scientific Statement From the American Heart Association. Stroke. 2021;52(6):e295–e308. doi:10.1161/STR.0000000000000367 [PubMed: 33719523]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

36. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, Part 2. JAMA. 2002;288(15):1909–1914. doi:10.1001/jama.288.15.1909 [PubMed: 12377092]

Highlights

- Preserving brain health is a key priority in primary care

- Screening for these risk factors is challenging to scale to large populations

- Survival prediction with an 11-year follow-up period was performed for 17040 patients

- Large-scale brain health screening can be performed using simple risk scores

- The scores quantify the risk of developing common neurologic and psychiatric diseases

8.

Summary Table

What was already known on the topic:

- Neurological disorders have been increasing in recent years largely because of the aging of the population

- Screening for brain health risk factors in face-to-face primary care visits is challenging to scale to large populations

What this study added to our knowledge:

- Simple risk scores derived from routinely collected EHR, including social covariates, accurately quantify the risk of developing common neurologic and psychiatric diseases.

- These scores can be computed automatically, prior to medical care visits, and may thus be useful for large-scale brain health screening
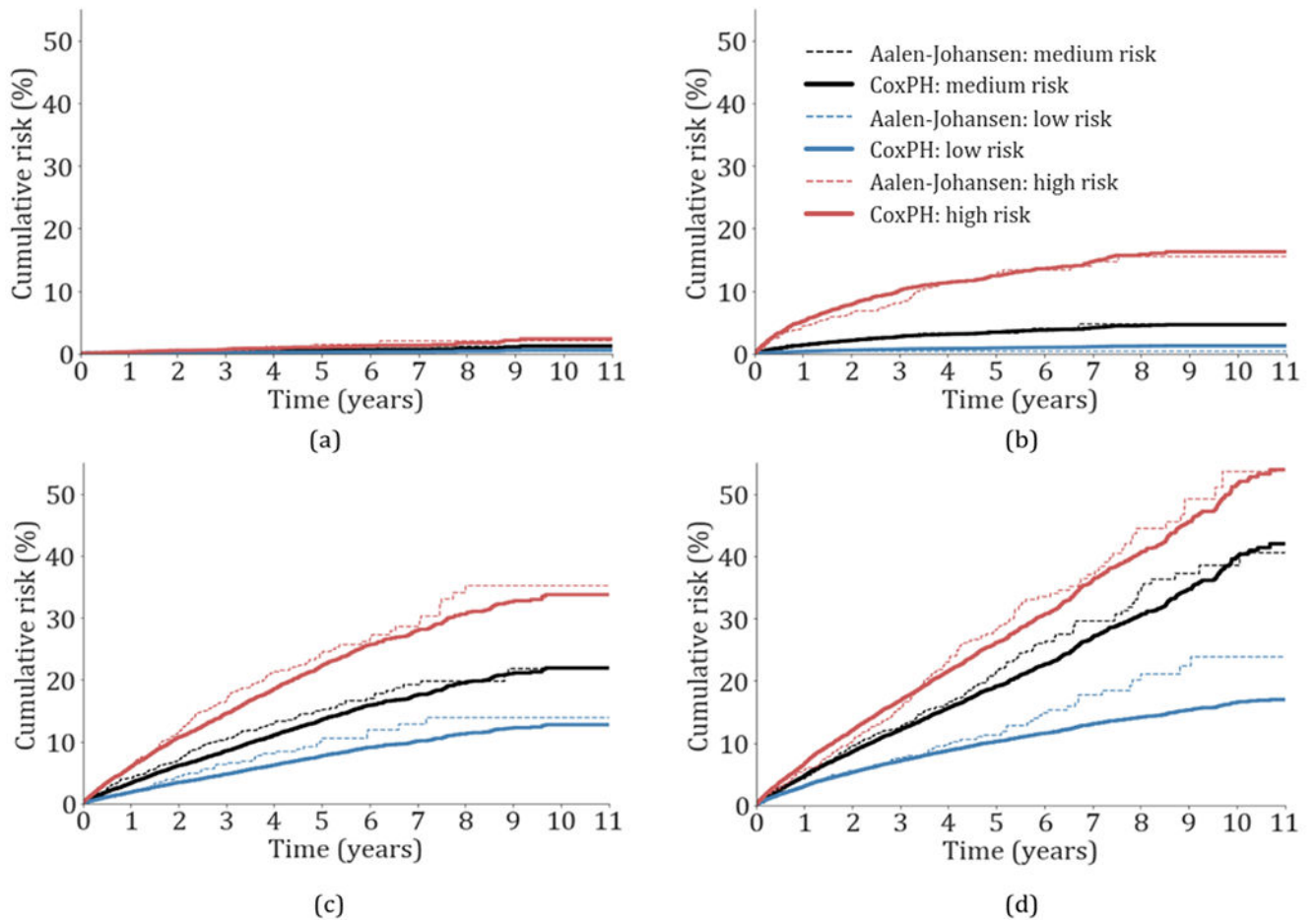
Fig. 1.
Cumulative incidence risk probability of patients in the test set during the study period
for (a) intracranial hemorrhage, (b) composite of dementia, (c) ischemic stroke, and (d)
depression. The low, medium and high-risk cumulative probabilities are presented in dotted
and continuous lines for the Aalen-Johansen estimator and the cox proportional hazards
model with competing risk (CoxPH), respectively. The medium cumulative risk corresponds
to the cumulative average risk probability.

Fig. 2.
Cumulative incidence risk probability of death for patients in the test set during the study period. The low, medium and high-risk cumulative probabilities are presented in dotted and continuous lines for the Aalen-Johansen estimator and the cox proportional hazards model with competing risk (CoxPH), respectively. The medium cumulative risk corresponds to the cumulative average risk probability.
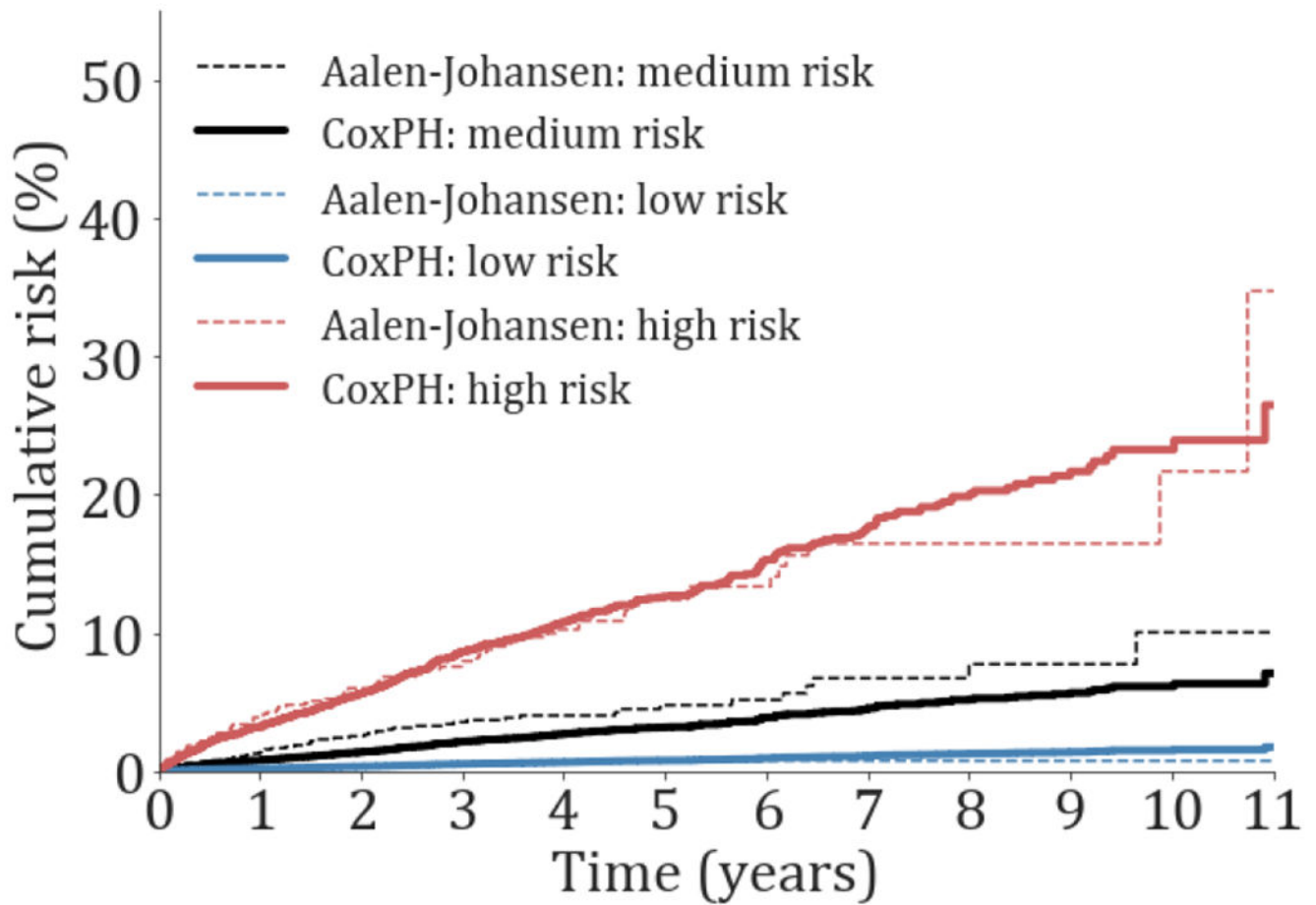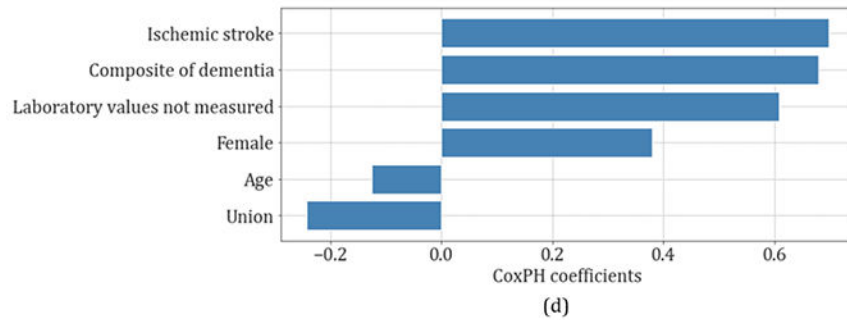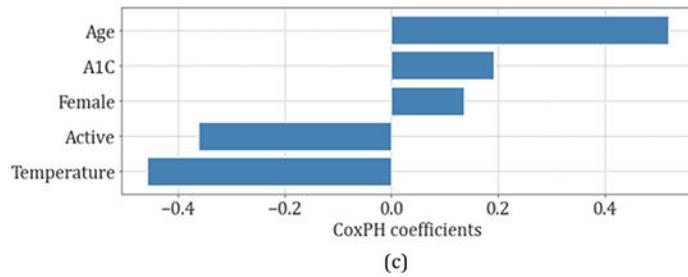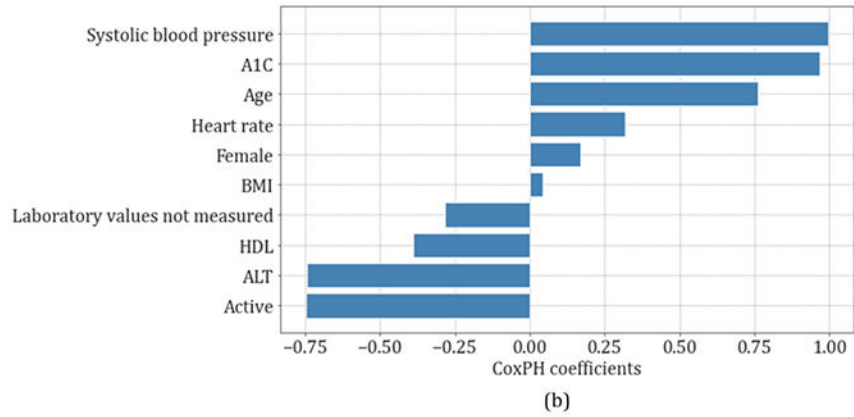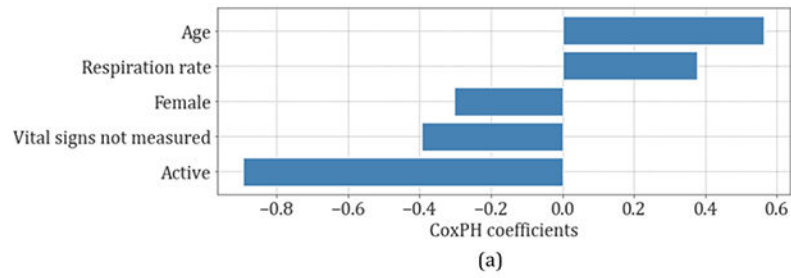
Fig. 3.
Covariates importance of the cox proportional hazards (CoxPH) competing risk model coefficients for (a) intracranial hemorrhage, (b) composite of dementia, (c) ischemic stroke, (d) depression.
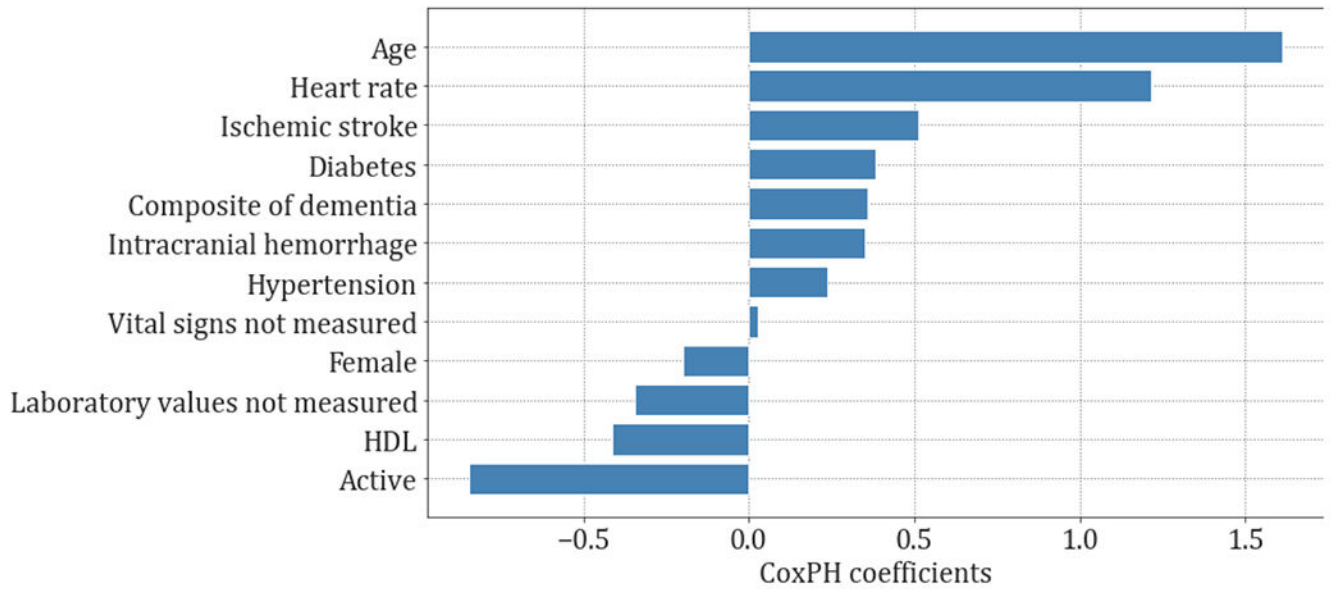
Fig. 4.
Covariates importance of the cox proportional hazards (CoxPH) coefficients for death.

Table 1.

Characteristics of the study cohort population.

Characteristic	Study cohort (n=17040)
Number of encounters, N	3786379
Age([a])(years, mean (SD))	49.0 (15.6)
Sex, n (%)	
Male	9724 (57.1)
Female	7316 (42.9)
Race, n (%)	
Black or African American	1069 (6.3)
Other ([b])	2725 (16.0)
White	13246 (77.7)
Ethnicity, n (%)	
Hispanic	1379 (8.1)
Non-Hispanic	13713 (80.5)
Unknown	1948 (11.4)
Marital status, n (%)	
Union	9055 (53.1)
Non-union	7644 (44.9)
Unknown	341 (2)
Tobacco status, n (%)	
Smoker	1288 (7.6)
Non-smoker	14850 (87.1)
Unknown	902 (5.3)
Alcohol status([c]), n (%)	
Consumption	4580 (26.9)
Non-consumption	2765 (16.2)
Unknown	9695 (56.9)
Employment status, n (%)	
Employed	7341 (43.1)
Non-employed	6833 (40.1)
Unknown	2866 (16.8)
Comorbidities, n (%), age (years, mean (SD))([a])	
Diabetes	1773 (10.4), 53.6 (12.2)
Hypertension	3573 (21.0), 56.1 (12.2)
Study outcomes, n (%), age (years, mean (SD))([a])	
Intracranial hemorrhage	96 (0.6), 58.4 (15.2)
Ischemic stroke	1964 (11.5), 57.1 (13.8)
Depression	2627 (15.4), 47.4 (15.2)
Composite of dementia	746 (4.4), 63.7 (9.5)
Death	896 (5.3), 64.1 (12.7)

Characteristic	Study cohort (n=17040)
None	10711 (62.9), 47.0 (15.2)

The number of patients and encounters is represented by n and N, respectively.

[a] Age at baseline for the first visit in the study period.

[b] 'Other' includes 'unknown', 'declined', 'American Indian or Alaska Native', 'Asian' and 'Native Hawaiian or other Pacific Islander'. Since these races represent less than 15% of the data, they were omitted to preserve patient privacy.

[c] Alcohol Consumption corresponds to consumption of at least one drink per week.

Table 2.

Average concordance index and risk ratios of the competing risk models for the study period, where event classes ('0', '1' and '2') are unbalanced or balanced in the test set for each of 1000 bootstrapping iterations to calculate 95% confidence intervals.

Outcome	Train C-index [SE]	Unbalanced outcome events			Balanced outcome events		
		Test C-index [95% CI]	RR lower [95% CI]	RR higher [95% CI]	Test C-index [95% CI]	RR lower [95% CI]	RR higher [95% CI]
Intracranial hemorrhage	0.78 [0.01]	0.78 [0.77-0.79]	0.46 [0.45, 0.48]	2.07 [1.95, 2.20]	0.60 [0.54-0.66]	0.47 [0.43, 0.52]	1.87 [1.67, 2.19]
Ischemic stroke	0.69 [0.01]	0.68 [0.67-0.69]	0.57 [0.57, 0.59]	1.64 [1.52, 1.69]	0.60 [0.58-0.62]	0.60 [0.58, 0.63]	1.57 [1.48, 1.65]
Depression	0.65 [0.01]	0.65 [0.64-0.66]	0.68 [0.39, 0.74]	1.29 [0.78, 1.38]	0.66 [0.64-0.67]	0.68 [0.61, 0.72]	1.36 [0.75, 1.44]
Composite of dementia	0.80 [0.01]	0.78 [0.76-0.79]	0.27 [0.26, 0.28]	3.52 [3.18, 3.81]	0.62 [0.60-0.65]	0.27 [0.25, 0.29]	3.22 [2.64, 3.75]
Death	0.80 [0.01]	0.79 [0.78-0.81]	0.24 [0.24, 0.24]	3.96 [3.91, 4.00]	0.65 [0.63-0.67]	0.79 [0.24, 1.73]	6.05 [1.91, 9.52]

SE – standard error. CI – confidence intervals. C-index – concordance index. RR – risk ratio.

Table 3.

Cumulative dynamic area under the receiver operating characteristic curve and average true positive and false positive rates [95% confidence intervals].

Outcome	Cumulative dynamic AUC			Mean TPR			Mean FPR		
	1st year	5th year	10th year	1st year	5th year	10th year	1st year	5th year	10th year
Intracranial hemorrhage	0.72 [0.60-0.86]	0.69 [0.64-0.76]	0.72 [0.59-0.80]	0.73 [0.60-0.86]	0.71 [0.65-0.78]	0.74 [0.63-0.80]	0.50 [0.50-0.51]	0.52 [0.51-0.53]	0.53 [0.49-0.59]
Ischemic stroke	0.63 [0.59-0.65]	0.66 [0.64-0.68]	0.64 [0.59-0.69]	0.62 [0.59-0.65]	0.64 [0.62-0.65]	0.62 [0.60-0.64]	0.50 [0.50-0.50]	0.48 [0.47-0.49]	0.49 [0.44-0.53]
Depression	0.62 [0.59-0.65]	0.62 [0.59-0.64]	0.62 [0.56-0.68]	0.57 [0.54-0.59]	0.55 [0.54-0.57]	0.53 [0.52-0.56]	0.48 [0.47-0.49]	0.46 [0.45-0.48]	0.44 [0.39-0.48]
Composite of dementia	0.74 [0.70-0.76]	0.72 [0.70-0.74]	0.69 [0.63-0.74]	0.74 [0.71-0.77]	0.76 [0.75-0.78]	0.77 [0.75-0.79]	0.51 [0.50-0.51]	0.55 [0.54-0.56]	0.59 [0.54-0.65]
Death	0.80 [0.77-0.83]	0.81 [0.79-0.83]	0.78 [0.73-0.84]	0.79 [0.76-0.82]	0.77 [0.76-0.79]	0.77 [0.73-0.80]	0.49 [0.49-0.50]	0.47 [0.46-0.48]	0.50 [0.45-0.55]

AUC – Area under the receiver operating characteristic. TPR – true positive rate; FPR – false positive rate.