Development and validation of domain-specific clinical prediction models of post-stroke cognitive impairment

Andrea Kusec, PhD[1], Kym IE Snell, PhD[2,3], & Nele Demeyere, PhD*[1]

[1]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

[2]Department of Applied Health Sciences, University of Birmingham, Birmingham, UK

[3]National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre, UK

Author Emails

andrea.kusec@ndcn.ox.ac.uk

k.snell@bham.ac.uk

*Corresponding author: nele.demeyere@ndcn.ox.ac.uk

Word count: 5999

Background: Post-stroke cognitive impairment (PSCI) is highly prevalent across multiple cognitive domains. Individualised PSCI prognosis has mainly been researched using global cognitive outcomes. Here, we develop and externally validate clinical prediction models for overall and domain-specific PSCI, including language, memory, attention, executive function, numeracy, and praxis.

Methods: *N*=430 stroke survivors completed the Oxford Cognitive Screen (OCS) in acute care and at 6-month follow-up (binarized outcome; impaired vs unimpaired). Logistic regression models were fitted comprising both mandatory clinically-relevant (age, sex, stroke severity, education, stroke hemisphere, acute PSCI) and data-driven (acute mood difficulties, length of stay in acute care, multimorbidity) predictors using backward elimination ($p < 0.10$) on multiply imputed data. Internal validation used bootstrapping to obtain optimism-adjusted performance estimates. External validation used the optimism-adjusted C-Slope as a uniform shrinkage factor.

Results: Compared to the overall PSCI model (C-Statistic=0.76 [95% CI=0.71–0.80]), comparable or improved optimism-adjusted performance was observed in models of language (C-Statistic=0.77 [95% CI=0.72–0.81]) memory (C-Statistic=0.72 [95% CI=0.65–0.75]), and attention (C-Statistic=0.74 [0.69–0.78]). Numeracy (C-Statistic=0.69 [95% CI=0.63–0.74]), executive function (C-Statistic=0.71 [95% CI=0.65–0.76]), and praxis (C-Statistic=0.60 [95% CI=0.53–0.65]) models showed weaker performance. In external validation, the overall PSCI model was comparable to development data (C-Statistic=0.74 [95% CI=0.67–0.79]).

Conclusions: Domain-specific prediction models have the potential to offer more meaningful PSCI prognoses compared to overall PSCI models. External performance of overall PSCI models show promise in different stroke severity cohorts. Future recalibration of memory, numeracy, executive function, and praxis models would be beneficial.

Keywords: stroke; cognitive dysfunction, cognitive impairment, prediction modelling, risk prediction

Introduction

Stroke is the leading cause of long-term physical and cognitive disability worldwide[1]. One-year post-stroke cognitive impairment (PSCI) prevalence estimates range from 40%[2] to 98%[3], with rates of 45%[4] to 80%[5] in chronic stroke. PSCI negatively impacts patients[6], caregivers/families[7], and has considerable economic costs[8].

Clinical prediction models have been developed to improve PSCI prognostication, chiefly post-stroke cognitive decline and dementia[9,10]. However, PSCI does not necessarily cause cognitive decline or dementia. Research demonstrates that whilst some patients exhibit decline, others have a stable, chronic cognitive impairment or even demonstrate continued improvement[11]. New PSCI definitions acknowledge the complex interplay of declining brain health, focal brain injury and cognitive recovery, with outcomes including decline, stability, and improvement[12]. PSCI is highly prevalent across multiple cognitive domains of language, memory, attention, numeracy, executive function, and praxis[13]. These impairments have previously been studied in isolation, despite research demonstrating differing recovery rates across domains (e.g., hemispatial neglect[14]).

Existing prediction models of post-stroke dementia[15–17] perform poorly in PSCI[18,19], possibly because specific cognitive domains have different relationships to functional outcomes[20]. Newer models have attempted to improve PSCI prediction, but remain focused on traditional clinical and demographic predictors (e.g., age, stroke severity). A recent meta-analysis (*N*=160,783) of PSCI predictors reported that, by far, the strongest predictor of chronic PSCI was acute cognitive functioning[21], demonstrating the importance of baseline cognitive performance in developing accurate prediction models[22]. However, existing PSCI prediction models do not routinely include acute cognition as a predictor. With early PSCI assessment now recommended by national and international guidelines[23–25] acute cognitive data should be routinely available and used in PSCI prognostication.

Existing PSCI prediction models often assess acute and long-term cognition via dementia screening tools (e.g., Mini Mental State Examination). These tools are not suitable for left hemisphere stroke due to overreliance on verbal abilities[26]. Utilizing a stroke-specific cognitive screen for PSCI clinical prediction models would avoid assessment confounds and is more strongly associated with 6-month cognitive recovery[13,27].

Study Aims

To develop and externally validate clinical prediction models of 6-month PSCI outcomes both in overall cognition and across language, memory, attention, numeracy, praxis, and executive function domains.

Methods

The study is a secondary analysis of data collected from the Oxford Cognitive Screening Programme. All participants provided informed consent to take part (REC Reference: 18/SC/0550).

Participants

Participants comprised a single, consecutively recruited cohort from the John Radcliffe Hospital acute stroke ward between March 2012 (first consented participant) to March 2020 (final follow-up participant). Programme inclusion criteria were 1) stroke diagnosis (first ever or recurrent); 2) ≥18 years; 3) ability to remain alert for ≥20 minutes, and 4) ability to provide informed consent. Participants completed a brief stroke-specific cognitive assessment acutely (N=866) and at 6-months post-stroke (N=430). Stroke severity (National Institute of Health Stroke Severity; NIHSS), acute function (Modified Rankin Scale; mRS) and other stroke-related details were obtained from electronic health records.

Study Outcome Measure

Domain-specific PSCI was assessed using the ***Oxford Cognitive Screen*** (OCS[28]).

The OCS comprises 12 subtasks forming 6 cognitive domains: Language (picture naming, semantics, sentence reading), Memory (orientation, verbal recall, episodic recognition), Spatial attention (broken hearts cancellation task), Numeracy (number writing, calculation), Praxis (gesture imitation), and Executive Function (mixed trails). Subtask scores are binarized as impaired (1) or unimpaired (0) relative to cut-off scores from a normative sample. A domain impairment was defined as the presence of any impairment in any subtask within that domain.

Models developed included overall PSCI severity (total proportion of OCS subtasks impaired at 6-months post-stroke; continuous outcome model), and binarized PSCI presence in any domain (logistic outcome model). Binary domain-specific impairment scores were used in Language, Memory, Spatial Attention, Numeracy, Praxis, and Executive Function prediction models.

Analysis

Analyses were performed in R Version 4.4.0[29]. Baseline descriptive statistics were first summarised. R packages used included *rms*[30], *psfmi*[31], *mice*[32], and *pmvalsampsize*[33]. Data is freely available at https://www.dementiasplatform.uk/ and analysis code at https://osf.io/3pc5k/

Predictor Selection

For all models, we selected predictors likely to be available in electronic health records upon deployment. This included clinically relevant predictors (age at stroke, sex, NIHSS scores, education years, first vs recurrent stroke, type of stroke [ischaemic vs haemorrhagic], stroke hemisphere, acute mRS score, and acute cognitive impairment[21]). We additionally included "data-driven predictors" that are are available in electronic health records (length of stay in acute care, independence prior to admission, presence of mood difficulties during acute care, and Charlson multimorbidity index). Each model therefore had 13 potential predictors forming an "initial model." Following predictor selection, performance was estimated with clinically relevant predictors and only significant data-driven predictors retained, labelled throughout the manuscript as the "final model." Predictor selection per model was repeated using bootstrapping across 1000 iterations.

Sample Size Justification

Sample size sufficiency was evaluated based on the events fraction, total sample size, number of predictor parameters, and a target shrinkage factor of >0.90 to minimise overfitting[34,35]. Event fraction rates and assumed apparent Nagelkerke's R^2 performance (0.30) were based on previous Oxford Cognitive Screening programme analyses[13]. For the overall PSCI model (*N* Events=295 of 430), the sample size required was 393 participants. Estimates were larger for the Language (1000 participants; *N* Events=138 of 428), Memory (1016 participants; *N* Events=137 of 430), Attention (638 participants; *N* Events=187 of 416), Praxis (2912 participants; *N* Events=79 of 403), Number (2524 participants; *N* Events=85 of 420), and Executive Function (1724 participants; *N* Events=98 of 400) models. Given all models are intended to be deployed using the OCS, the same predictor selection and development process was used across all domains.

Missing Data Management

Those with complete vs incomplete data at 6-months post-stroke were compared on predictor variables. To increase statistical power and reduce bias, multiple imputation was conducted across 20 imputed datasets (due to 28.8% and 42.3% missingness in acute NIHSS and mRS scores, respectively) with 50 iterations. Data were assumed missing at random given that variables with the highest missingness rates (NIHSS and mRS) were historically unavailable in electronic health records in earlier recruitment periods. Only predictor variables were imputed[36]. Upon model deployment, missingness is likely to occur (e.g., missing stroke severity information) and therefore imputation would be necessary[36]. Sensitivity analyses were conducted investigating the influence of missing information. A detailed account of participant attrition in this cohort is elsewhere[13].

Model Development and Internal Validation

Across all models, clinically relevant predictors were retained irrespective of statistical significance. Backward stepwise elimination was used to remove only non-significant (p>0.10) data-driven predictors. This approach was taken given criticisms around removing clinically relevant (though statistically insignificant) predictors[37]. Models developed across 20 imputed datasets were compared to complete case data.

The same model development approach was used for each OCS domain, resulting in 8 prediction models (2 overall PSCI models, 6 domain-specific).

Apparent final model performance (i.e., non-significant data-driven predictors removed) was evaluated using discrimination (model's ability to correctly identify

individuals with and without 6-month PSCI; estimated via the Area Under the Curve [AUC; binary outcome models only], C-Statistic), calibration measures (calibration-in-the-large [CITL], calibration slope [C-Slope], calibration plots, Brier scores, observed- expected ratio) and goodness-of-fit measures (adjusted R^2, continuous PSCI model; Nagelkerke's R^2, binary PSCI models). Pooled (across imputed datasets) b-values, odds ratios (ORs), and performance statistics are reported per model.

Optimism-adjusted performance estimates were obtained via bootstrapping each model on multiply imputed data across 1000 iterations. The model-specific optimism-adjusted C-Slope was used as a uniform shrinkage factor and was multiplied with model regression coefficients to correct for potential overfitting[38,39]. Model intercepts were re-estimated using the shrunken regression coefficients to obtain an accurate CITL.

Risk groups were created using 10th decile groups on prediction model estimates for visualisation purposes via calibration plots.

External Validation

The OCS-Care dataset[11] was used for external validation. In parallel to OCS-Recovery data, the OCS-Care dataset (*N*=264, *M* age = 68.9) assessed acute PSCI using the OCS and 6-months later, comprising a mild severity cohort (*Mean* NIHSS=2.8). Model predictors were collected from electronic health records, except acute mRS scores which were not available in the OCS-Care dataset. Many OCS-Care participants had at least one cognitive impairment at 6-months (*N* events=147; 55.6%), versus 70.8% acutely (*N* events=187). When estimating minimum sample size requirements[40], this dataset was sufficient to estimate a C-Statistic of 0.80 (CI width=0.20), though precise C-Slope estimates require a much larger dataset.

Shrunken model coefficients obtained through internal validation were applied to the OCS-Care data to estimate performance. Performance measures described above (C-Slope, CITL, C-Statistic, R^2 and Nagelkerke's R^2) were estimated. Overall binary PSCI model performance was further evaluated within subgroups by age range, sex, and acute PSCI severity.

Results

Demographics are in Table 1. At 6-months, all participants provided outcome data, though there was variation across PSCI domains (Language *N*=428; Memory *N*=430; Attention *N*=416; Praxis *N*=403; Numeracy *N*=420; Executive Function *N*=400).

Participants with missing 6-month PSCI data were more likely to have higher acute PSCI ($p<0.0001$), be older in age ($p=0.02$), and have acute language ($p<0.01$), memory ($p<0.001$), or numeracy impairments ($p<0.001$). NIHSS and mrS missingness was not related to demographic factors (*p*s>0.15; see *Supplemental Materials*).

[Table 1]

Overall PSCI Models

Pooled shrunken coefficients of the final overall continuous and binary PSCI models are in Table 2. Shrunken domain-specific coefficients are in *Tables S3–S8.*

[Table 2]

In the multivariable continuous model of proportion of 6-months OCS tasks impaired, the strongest clinically relevant predictors included higher age (pooled *B*=0.001 [95% CI=0.001–

0.003]) and a greater proportion of acute OCS tasks impaired (pooled *B*=0.31 [95% CI=0.26–0.37]). In bootstrapped and complete case data, the only data-driven predictor retained was requiring carer support prior to admission (pooled *B*=0.09 [95% CI=0.04–0.13]) and improved model fit in complete case data (*F*=3.77, *p*=0.02). The optimism-adjusted performance of the continuous overall PSCI model was good to excellent (C-Slope=0.95, CITL=-0.04; MSE=0.02; Adj R^2=0.32).

In the multivariable binary model, higher age (pooled OR=0.93 [95% CI=0.92–0.93]), bilateral hemisphere lesions (pooled OR=0.75 [95% CI=0.64–0.88]), fewer years of education (pooled OR=0.91 [95% CI=0.89–0.92]), and a greater proportion of acute OCS tasks impaired (pooled OR=1.62 [95% CI=1.35–1.95]) were associated with an increased risk of 6-month PSCI. No data-driven predictors were retained for the binary PSCI model. The final optimism-adjusted performance showed good performance (C-Statistic=0.76 [95% CI=0.71–0.80]; C-Slope=0.92 [95% CI=0.74–1.09]; CITL= -1.84 [95% CI= -2.08– -1.62]; Brier Score=0.12 [95% CI=0.10–0.14]; Nagelkerke's R^2=0.21).

In sensitivity analyses using complete case data, there were no notable differences in predictor selection for either the continuous or binary overall PSCI models.

[Figure 1]

Language Model

In the final multivariable language model, higher age (pooled OR=1.00 [95% CI=0.99–1.01]), years of education (pooled OR=0.99 [95% CI=0.98–1.00]), left hemisphere stroke (pooled OR=0.90 [95% CI=0.83–0.99]) and acute language impairments (pooled OR = 1.38

[95% CI=1.28–1.49]) were most strongly associated with 6-month language impairment. Of the data-driven predictors, only requiring carer support prior to admission was retained (pooled OR=1.20 [95% CI=1.04–1.39]).

The final optimism-adjusted language model showed good performance (C-Statistic=0.77 [95% CI=0.72–0.81]; C-Slope=0.90 [95% CI=0.74–1.05]; CITL= -0.99 [95% CI= -1.19– -0.79]; Brier Score=0.10 [95% CI=0.08–0.12]; Nagelkerke's R^2=0.25), with no notable predictor selection differences in complete case data.

Memory Model

In the final multivariable memory model, acute memory impairments (pooled OR=1.30 [95% CI=1.19–1.40]) predicted likelihood of 6-month memory impairment *(p*<0.10). Of the data-driven predictors, only requiring carer support prior to admission was retained OR=1.36 [95% CI=1.18–1.57]).

The final optimism-adjusted memory model showed acceptable to good performance (C-Statistic=0.71 [95% CI=0.65–0.75]; C-Slope=0.88 [95% CI=0.70–1.06]; CITL= -0.92 [95% CI= -1.13– -0.72]; Brier Score =0.09 [95% CI=0.07–0.11]; Nagelkerke's R^2=0.16), with no differences in predictor selection in complete case data.

Attention Model

For the final multivariable attention model, higher age (pooled OR=1.01 [95% CI=1.00–1.01]), lower number of years of education (pooled OR=0.99 [95% CI=0.98–0.99]), right hemisphere lesions (pooled OR=1.13 [95% CI=1.03–1.23]), recurrent stroke (pooled OR=1.08 [95% CI=0.99–1.18]) and acute attention impairments (pooled OR=1.20 [95%

CI=1.11–1.30]) had the strongest associations with likelihood of 6-month attention impairment, with only greater CCI scores retained from the data-driven predictors (pooled OR=1.03 [95% CI=0.99–1.07]).

The final optimism-adjusted attention model showed acceptable to good performance (C-Statistic=0.74 [95% CI=0.69–0.78]; C-Slope=0.88 [95% CI=0.70–1.06]; CITL= -1.28 [95% CI= -1.48– -1.07]; Brier Score=0.10 [95% CI=0.08–0.11]; Nagelkerke's R^2=0.20), with no predictor selection differences in complete cases.

Numeracy Model

In the final multivariable numeracy model, non-haemorrhagic stroke (pooled OR=0.92 [95% CI=0.83–1.02]) and acute numeracy impairments (pooled OR=1.24 [95% CI=1.14–1.34]) demonstrated the strongest predictive value toward likelihood of 6-month numeracy impairments. No data-driven predictors were retained (*p*s>0.10).

The final optimism-adjusted numeracy model showed acceptable performance (C-Statistic=0.69 [95% CI=0.63–0.74]; C-Slope=0.83 [95% CI=0.58–1.08]; CITL= -1.37 [95% CI= -1.61– -1.13]; Brier Score=0.07 [95% CI=0.06–0.09]; Nagelkerke's R^2=0.10), with no differences in predictor selection in complete cases.

Executive Function Model

In the final multivariable executive function model, female sex (pooled OR=1.11 [95% CI=1.02–1.21]), non-haemorrhagic stroke (pooled OR=0.89 [95% CI=0.80–1.00]), and acute executive function impairments (pooled OR=1.19 [95% CI=1.09–1.31]) demonstrated the

strongest associations with 6-month executive function impairments. Of the data-driven predictors, only greater CCI scores were retained (pooled OR=1.03 [95% CI=0.99–1.07]).

The final optimism-adjusted executive function model showed acceptable performance (C-Statistic=0.71 [95% CI=0.65–0.76]; C-Slope=0.81 [95% CI=0.57–1.05]; CITL= -1.55 [95% CI= -1.78– -1.32]; Brier Score=0.09 [95% CI=0.07–0.11]; Nagelkerke's R^2=0.13), with no differences in predictor selection in complete case data.

Praxis Model

In the final multivariable praxis model, greater age at time of stroke (pooled OR=1.00 [95% CI=1.00–1.01]), male sex (pooled OR=0.92 [95% CI=0.85–1.00]), bilateral hemisphere lesions (pooled OR=0.84 [95% CI=0.72–0.98]), and acute praxis impairments (pooled OR=1.11 [95% CI=1.01–1.21]) were most strongly associated with likelihood of 6-month praxis impairments. Of the data-driven predictors, only requiring carer support prior to admission was retained (pooled OR=0.88 [95% CI=0.74–1.03]).

The final optimism-adjusted praxis model showed poor to acceptable performance (C-Statistic=0.60 [95% CI=0.53–0.65]; C-Slope=0.66 [95% CI=0.34–0.97]; CITL= -1.61 [95% CI= -1.85– -1.36]; Brier Score=0.06 [95% CI=0.03–0.07]; Nagelkerke's R^2=0.01). No predictor selection differences were observed in complete cases.

Unadjusted relationships between predictor and outcome variables are in *Tables S9–S15.* All model performance measures are in Table 3. Domain-specific calibration plots are in Figure 2 (see *Figure S1* for complete case plots).

[Figure 2]

[Table 3]

External Validation

External validation estimates are in Table 4. Model discrimination, calibration, and goodness-of-fit was not largely discrepant in the OCS-Care dataset, suggesting good model performance across cohorts. By contrast, Language and Attention models demonstrated good discrimination and goodness-of-fit, however calibration was poor across all domains, particularly in Memory, Numeracy, Executive Function, and Praxis.

In the binary overall PSCI model subgroup analyses (*Table S16*), performance did not vary by sex (Male C-Statistic=0.76 [0.67–0.83], Female C-Statistic=0.76 [0.66–0.84]). Model performance varied by age group (<60 years C-Statistic=0.76 [0.62–0.86], >60 years C-Statistic=0.65 [0.48–0.78], >70 years C-Statistic=0.65 [0.49–0.78], >80 years C-Statistic=0.71 [0.52–0.72]), and by level of acute PSCI (Mild acute PSCI C-Statistic=0.62 [0.52–0.72]; Moderate-severe acute PSCI=0.72 [0.61–0.81]). We developed an online risk calculator of the binary overall PSCI model given its promising performance: https://ocs-strokecogpredictor.shinyapps.io/OCS-StrokeCog-Predictor/

[Table 4 here]

Discussion

We developed and externally validated overall and domain-specific PSCI prediction models, utilising domain-specific acute cognitive information from a stroke-specific screen (OCS) alongside established PSCI predictors. To our knowledge, this is the first study to develop domain-specific PSCI clinical prediction models and use acute cognitive data as a predictor.

Overall and Domain-Specific Model Performance

Our models provided good explanatory power, with optimism-adjusted C-Slopes of 0.95 and 0.76 for continuous and binary overall PSCI models, respectively. Compared to models of post-stroke decline and dementia (C-Statistic range=0.53–0.66[19]), our models comparatively perform better even when considering domain-specific models (C-Statistic range=0.60–0.77). Promisingly, in external data, overall PSCI model performance was comparable to internal validation, suggesting it could be used across different stroke cohorts.

In domain-specific model performance, language, memory, attention and executive function models were at least comparable to overall PSCI models (C-Statistics=0.71–0.77). However, numeracy and praxis domain models showed poorer performance (C-Statistics=0.60–0.69; Nagelkerke's $R^2 \leq 0.10$), likely due to the lower prevalence of these impairments at 6-months[41] and underpowered models.

The optimism-adjusted C-Slopes for the majority of domain-specific models presented here (<0.90) suggests overfitting may have occurred. We applied penalization and shrinkage methods to account for the modest sample size[42]. Given the low explained variance in these PSCI domains (e.g., praxis), predictive accuracy will likely improve in recalibration in a larger sample.

In external data, CITL estimates were consistently negative, suggesting systematic overprediction of 6-month PSCI risk. This is likely due to development data comprising a

more moderate-severe stroke cohort, whilst external data comprised a more mild stroke cohort. Additionally, C-Slopes in external data are larger, potentially indicating overshrinkage. Prediction models should be recalibrated across a range of stroke severities.

Clinical Implementation of Overall and Domain-Specific PSCI Models

Age, sex, years of education, NIHSS scores, recurrent stroke and stroke type contribute to PSCI[21] and should be included in recalibrations of our models. Novel predictors should also be considered; our modelling approach includes data-driven predictors to allow for routine model updating. This approach identified potential predictors for recalibration (i.e., requiring carer support pre-stroke and multimorbidity) that are excluded from PSCI prediction modelling[9,19,43,44]. Crucially, we selected predictors available in electronic health records. PSCI models often include predictors not routinely available at deployment[9,10,19]. In the UK specifically, predictor selection should guided by the National Clinical Guideline for Stroke[25] such as including acute cognitive assessment, as these are more likely to available. Other biopsychosocial (e.g., white matter hyperintensities, socioeconomic status) and clinical (e.g., amount/intensity of neurorehabilitation offered) predictors may improve model performance, however these may be less available or have significant economic considerations. For example, imaging-based data improves PSCI prediction models[45], however behavioural data (e.g., cognitive assessments) is considerably more affordable and feasible to implement[46].

As typical for electronic health record data, NIHSS and mRS scores had large amounts of missingness. Imputation methods should be considered at deployment[36]. Collecting feedback on model usability given predictor missingness (e.g., Archer et al.[47]) would aid implementation.

Role of Overall and Domain-Specific Clinical Prediction Models in PSCI Prognoses

Domain-specific clinical prediction models may offer more meaningful PSCI prognoses. Though PSCI rates are highest during acute stroke, early PSCI may be reversible[48] and information about likely 6-month outcomes is valuable to stroke patients[49]. Qualitative research suggests that focusing solely on cognitive decline as a possible PSCI outcome (e.g., Hbid et al.[9]) may cause undue concern or at best be irrelevant[50]. Stroke survivors and families commonly report wanting personalized information about managing cognitive changes[49]. Our models are an essential first step to providing person-specific and cognitive domain-specific trajectories.

Strengths and Limitations

A notable strength is using the OCS, a stroke-specific PSCI measure rather than cognitive decline or dementia. The OCS' minimisation of confounds, brief administration time, and information on Language, Memory, Attention, Numeracy, Executive Function and Praxis make it a credible candidate for PSCI model development.

Low prevalence of certain domain-specific outcomes (numeracy, executive function, and praxis) restricted model accuracy. Less prevalent impairments require substantial sample sizes for sufficient development. Finally, as typical for new prediction models, our domain-specific models require recalibration.

Future Domain-Specific PSCI models

Future domain-specific PSCI models could consider how combinations of specific cognitive impairments influence performance. Overall PSCI models assume multiple within-

domain and multidomain impairments have equal additive contributions to outcomes. Specific impairment combinations (e.g., language and executive function) could affect outcomes, given differential correlations between cognitive domains[13]. Developing within-domain models (e.g., sentence reading model vs language impairment model) in sufficiently large datasets may be helpful, given varying recovery within domains[4,13]. Finally, predictor selection should be carefully considered. Acute cognition best explains long-term PSCI[21] with established predictors other than age explaining little variance[13,44]. Less frequently researched PSCI domains (i.e., numeracy, executive function, and praxis) may particularly benefit from data-driven predictors.

Conclusion

We demonstrate that acute cognitive information improves prediction of overall and domain-specific 6-month PSCI. Overall 6-month PSCI models show promise in external data. Our model development process allows for future inclusion of novel data-driven predictors. Domain-specific models should be recalibrated to better inform PSCI outcomes.

Acknowledgements

This study has been delivered through the National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the [name of the research funder], the NIHR or the Department of Health and Social Care. The authors thank all participants and members of the Oxford Translational Neuropsychology Group for contributions to recruitment/testing.

Sources of Funding

ND (Advanced Fellowship NIHR302224) and AK (DSE Award NIHR305153) are funded by NIHR.

Disclosures

ND is a developer of the OCS but does not receive renumeration for its use.

References

1. Feigin VL, Stark BA, Johnson CO, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol*. 2021;20(10):795-820. doi:10.1016/S1474-4422(21)00252-0

2. Pendlebury ST, Rothwell PM. Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis. *Lancet Neurol*. 2009;8(11):1006-1018. doi:10.1016/S1474-4422(09)70236-4

3. Sexton E, McLoughlin A, Williams DJ, et al. Systematic review and meta-analysis of the prevalence of cognitive impairment no dementia in the first year post-stroke. *Eur Stroke J*. 2019;4(2):160-171. doi:10.1177/2396987318825484

4. Kusec A, Milosevich E, Williams OA, et al. Long-term psychological outcomes following stroke: the OX-CHRONIC study. *BMC Neurol*. 2023;23(1):426. doi:10.1186/s12883-023-03463-5

5. Mahon S, Parmar P, Barker-Collo S, et al. Determinants, Prevalence, and Trajectory of Long-Term Post-Stroke Cognitive Impairment: Results from a 4-Year Follow-Up of the ARCOS-IV Study. *Neuroepidemiology*. 2017;49(3-4):129-134. doi:10.1159/000484606

6. Cumming TB, Brodtmann A, Darby D, Bernhardt J. The importance of cognition to quality of life after stroke. *J Psychosom Res*. 2014;77(5):374-379. doi:10.1016/j.jpsychores.2014.08.009

7. Stolwyk RJ, Mihaljcic T, Wong DK, Hernandez DR, Wolff B, Rogers JM. Post-stroke Cognition is Associated with Stroke Survivor Quality of Life and Caregiver Outcomes: A Systematic Review and Meta-analysis. *Neuropsychol Rev*. Published online March 11, 2024. doi:10.1007/s11065-024-09635-5

8. Luengo-Fernandez R, Violato M, Candio P, Leal J. Economic burden of stroke across Europe: A population-based cost analysis. *Eur Stroke J*. 2020;5(1):17-25. doi:10.1177/2396987319883160

9. Hbid Y, Fahey M, Wolfe CDA, Obaid M, Douiri A. Risk Prediction of Cognitive Decline after Stroke. *J Stroke Cerebrovasc Dis*. 2021;30(8):105849. doi:10.1016/j.jstrokecerebrovasdis.2021.105849

10. Tang EYH, Robinson L, Stephan BCM. Risk Prediction Models for Post-Stroke Dementia. *Geriatrics*. 2017;2(3):19. doi:10.3390/geriatrics2030019

11. Demeyere N, Sun S, Milosevich E, Vancleef K. Post-stroke cognition with the Oxford Cognitive Screen vs Montreal Cognitive Assessment: a multi-site randomized controlled study (OCS-CARE). *AMRC Open Res*. 2019;1:12. doi:10.12688/amrcopenres.12882.1

12. Rost NS, Brodtmann A, Pase MP, et al. Post-stroke cognitive impairment and dementia. *Circulation research*. 2022;130(8):1252-1271.

13. Milosevich ET, Moore MJ, Pendlebury ST, Demeyere N. Domain-specific cognitive impairment 6 months after stroke: The value of early cognitive screening. *Int J Stroke*. Published online September 25, 2023:17474930231205787. doi:10.1177/17474930231205787

14. Moore MJ, Vancleef K, Riddoch MJ, Gillebert CR, Demeyere N. Recovery of Visuospatial Neglect Subtypes and Relationship to Functional Outcome Six Months After Stroke. *Neurorehabil Neural Repair*. 2021;35(9):823-835. doi:10.1177/15459683211032977

15. Fahey M, Rudd A, Béjot Y, Wolfe C, Douiri A. Development and validation of clinical prediction models for mortality, functional outcome and cognitive impairment after stroke: a study protocol. *BMJ Open*. 2017;7(8):e014607. doi:10.1136/bmjopen-2016-014607

16. Lee M, Yeo NY, Ahn HJ, et al. Prediction of post-stroke cognitive impairment after acute ischemic stroke using machine learning. *Alzheimers Res Ther*. 2023;15(1):147. doi:10.1186/s13195-023-01289-4

17. Li X, Chen Z, Jiao H, et al. Machine learning in the prediction of post-stroke cognitive impairment: a systematic review and meta-analysis. *Front Neurol*. 2023;14:1211733. doi:10.3389/fneur.2023.1211733

18. Stephan BCM, Minett T, Muniz Terrera G, Matthews FE, Brayne C. Dementia prediction for people with stroke in populations: is mild cognitive impairment a useful concept? *Age Ageing*. 2015;44(1):78-83. doi:10.1093/ageing/afu085

19. Tang EYH, Price CI, Robinson L, et al. Assessing the Predictive Validity of Simple Dementia Risk Models in Harmonized Stroke Cohorts. *Stroke*. 2020;51(7):2095-2102. doi:10.1161/STROKEAHA.120.027473

20. Mole JA, Demeyere N. The relationship between early post-stroke cognition and longer term activities and participation: A systematic review. *Neuropsychol Rehabil*. 2020;30(2):346-370. doi:10.1080/09602011.2018.1464934

21. Filler J, Georgakis MK, Dichgans M. Risk factors for cognitive impairment and dementia after stroke: a systematic review and meta-analysis. *Lancet Healthy Longev*. 2024;5(1):e31-e44. doi:10.1016/S2666-7568(23)00217-9

22. Demeyere N. Acute post-stroke screening for a cognitive care pathway. *Lancet Healthy Longev*. 2024;5(1):e4-e5. doi:10.1016/S2666-7568(23)00257-X

23. Lindsay P, Furie KL, Davis SM, Donnan GA, Norrving B. World Stroke Organization Global Stroke Services Guidelines and Action Plan. *Int J Stroke*. 2014;9(SA100):4-13. doi:10.1111/ijs.12371

24. Quinn TJ, Richard E, Teuschl Y, et al. European Stroke Organisation and European Academy of Neurology joint guidelines on post-stroke cognitive impairment. *Eur Stroke J*. 2021;6(3):I-XXXVIII. doi:10.1177/23969873211042192

25. National Clinical Guideline for Stroke. Accessed February 9, 2024. https://www.strokeguideline.org/contents/

26. Demeyere N, Riddoch MJ, Slavkova ED, et al. Domain-specific versus generalized cognitive screening in acute stroke. *J Neurol*. 2016;263(2):306-315. doi:10.1007/s00415-015-7964-4

27. Bisogno AL, Franco Novelletto L, Zangrossi A, et al. The Oxford cognitive screen (OCS) as an acute predictor of long-term functional outcome in a prospective sample of stroke patients. *Cortex*. 2023;166:33-42. doi:10.1016/j.cortex.2023.04.015

28. Demeyere N, Riddoch MJ, Slavkova ED, Bickerton WL, Humphreys GW. The Oxford Cognitive Screen (OCS): Validation of a stroke-specific short cognitive screening tool. *Psychol Assess*. 2015;27(3):883-894. doi:10.1037/pas0000082

29. R Core Team. R Studio. Published online 2024.

30. Frank E Harrell Jr <fh@fharrell.com>. rms: Regression Modeling Strategies. Published online September 8, 2009:6.8-1. doi:10.32614/CRAN.package.rms

31. Heymans M. psfmi: Prediction Model Pooling, Selection and Performance Evaluation Across Multiply Imputed Datasets. Published online May 16, 2019:1.4.0. doi:10.32614/CRAN.package.psfmi

32. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw*. 2011;45:1-67. doi:10.18637/jss.v045.i03

33. Ensor J. pmvalsampsize: Sample Size for External Validation of a Prediction Model. Published online November 16, 2023:0.1.0. doi:10.32614/CRAN.package.pmvalsampsize

34. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med*. 2019;38(7):1276-1296. doi:10.1002/sim.7992

35. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441. doi:10.1136/bmj.m441

36. Sisk R, Sperrin M, Peek N, van Smeden M, Martin GP. Imputation and missing indicators for handling missing data in the development and deployment of clinical prediction models: A simulation study. *Stat Methods Med Res*. 2023;32(8):1461-1477. doi:10.1177/09622802231165001

37. Smith G. Step away from stepwise. *J Big Data*. 2018;5(1):32. doi:10.1186/s40537-018-0143-6

38. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiol Camb Mass*. 2010;21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2

39. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD). *Circulation*. 2015;131(2):211-219. doi:10.1161/CIRCULATIONAHA.114.014508

40. Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med*. 2021;40(19):4230-4251. doi:10.1002/sim.9025

41. Sor R, Park YS, Boets P, Goethals PLM, Lek S. Effects of species prevalence on the performance of predictive models. *Ecol Model*. 2017;354:11-19. doi:10.1016/j.ecolmodel.2017.03.006

42. Minimum sample size for external validation of a clinical prediction model with a binary outcome - Riley - 2021 - Statistics in Medicine - Wiley Online Library. Accessed August 5, 2024. https://onlinelibrary.wiley.com/doi/full/10.1002/sim.9025

43. Tang EY, Amiesimaka O, Harrison SL, et al. Longitudinal Effect of Stroke on Cognition: A Systematic Review. *J Am Heart Assoc*. 7(2):e006443. doi:10.1161/JAHA.117.006443

44. Predicting post-stroke cognitive impairment using electronic health record data - Jeffrey M Ashburner, Yuchiao Chang, Bianca Porneala, Sanjula D Singh, Nirupama Yechoor, Jonathan M Rosand, Daniel E Singer, Christopher D Anderson, Steven J Atlas, 2024. Accessed August 5, 2024. https://journals.sagepub.com/doi/full/10.1177/17474930241246156

45. Yuan X, Zhang L, Sui R, Wang Z. A risk prediction model of post-stroke cognitive impairment based on magnetic resonance spectroscopy imaging. *Neurol Res*. 2021;43(8):642-652. doi:10.1080/01616412.2021.1908659

46. Simpkins AN, Janowski M, Oz HS, et al. Biomarker Application for Precision Medicine in Stroke. *Transl Stroke Res*. 2020;11(4):615-627. doi:10.1007/s12975-019-00762-3

47. Archer L, Snell KIE, Stynes S, et al. Development and External Validation of Individualized Prediction Models for Pain Intensity Outcomes in Patients With Neck Pain, Low Back Pain, or Both in Primary Care Settings. *Phys Ther*. 2023;103(11):pzad128. doi:10.1093/ptj/pzad128

48. Rasquin SMC, Lodder J, Verhey FRJ. Predictors of reversible mild cognitive impairment after stroke: a 2-year follow-up study. *J Neurol Sci*. 2005;229-230:21-25. doi:10.1016/j.jns.2004.11.015

49. Lin B lei, Mei Y xia, Wang W na, et al. Unmet care needs of community-dwelling stroke survivors: a systematic review of quantitative studies. *BMJ Open*. 2021;11(4):e045560. doi:10.1136/bmjopen-2020-045560

50. Hobden G, Tang E, Demeyere N. Cognitive assessment after stroke: A qualitative study of patients' experiences. *BMJ Open*. 2023;13(6):e072501. doi:10.1136/bmjopen-2023-072501

Participants (*N*=430)		Min-Max
Sex–*n* (%)		
Male	230 (53.5%)	
Female	200 (46.5%)	
Age–Mean (*SD*)	73.8 (12.5)	18–95
Education Years–Mean (*SD*)	12.3 (3.6)	0–23
Stroke Type–*n* (%)		
Ischaemic	362 (84.2%)	
Haemorrhagic	68 (15.8%)	
Lesion Hemisphere–*n* (%)		
Left	153 (35.6%)	
Right	168 (39.1%)	
Bilateral	34 (7.9%)	
Undetermined from scan	75 (17.4%)	
First or Recurrent Stroke–*n* (%)		
First	292 (67.9%)	
Recurrent	138 (32.1%)	
Acute NIHSS–Mean (*SD*)	6.8 (6.1)	0–30

Table 1. Participant demographics.
NIHSS=National Institute of Health Stroke Severity

	Overall Score (severity-	**Any cognitive impairment**

	continuous)		(binary)	
	***B* [95% CI]**	***p*-value**	**OR [95% CI]**	***p*-value**
Clinically Relevant Predictors				
Intercept	-0.05 [-0.16–0.06]	0.33	0.12 [0.02–0.72]	<0.05
Stroke Age	0.001 [0.001–0.003]	<0.01	0.93 [0.92–0.93]	<0.001
Sex (Female)	0.01 [-0.01–0.03]	0.38	0.94 [0.87–1.03]	0.57
Education Years	-0.002 [-0.01–0.001]	0.19	0.91 [0.89–0.92]	0.08
NIHSS	0.001 [-0.005–0.001]	0.55	0.92 [0.91–0.93]	0.59
Acute Rankin	-0.002 [-0.02–0.01]	0.74	0.93 [0.89–0.98]	0.57
Hemisphere				
Right	0.01 [-0.02–0.03]	0.63	0.93 [0.84–1.02]	0.89
Bilateral	-0.04 [-0.09–0.002]	0.06	0.75 [0.64–0.88]	<0.05
Undetermined from Scan	0.01 [-0.03–0.04]	0.73	0.86 [0.76–0.97]	0.32
First/Recurrent Stroke				
Recurrent Stroke	0.02 [-0.01–0.05]	0.14	1.01 [0.92–1.10]	<0.05
Stroke Type				
Haemorrhagic	-0.02 [-0.05–0.01]	0.22	0.89 [0.80–1.01]	0.67
Acute Proportion OCS Tasks Impaired	0.31 [0.26–0.37]	<0.001	1.62 [1.35–1.95]	<0.001
Data Driven Predictors				
Independence Pre-Admission				
Carer Support	0.09 [0.04–0.13]	<0.001	--	--
Family Support	0.03 [-0.02–0.09]	0.24	--	--

Table 2. Pooled model coefficients of continuous and binary overall PSCI models. Shrinkage was applied to coefficients using the optimism-adjusted C-Slope. For categorical variables stroke hemisphere and independence pre-admission, reference categories were left-hemisphere stroke and independent pre-admission, respectively.
NIHSS=National Institute of Health Stroke Severity; OCS=Oxford Cognitive Screen

Model	Estimate	Model Performance [95% CI]	Average Optimism [95% CI]	Optimism-Adjusted Performance [95% CI]
Overall PSCI (Continuous)	C-Slope	1 [0.88–1.12]	0.05	0.95
	CITL	0.00 [-0.0001–0.0001]	-0.007	-0.04
	MSE	0.02 [-0.05–0.08]	0.00	0.02
	Adjusted R^2	0.38 [0.31–0.45]	0.05	0.32
Any domain impairment (Binary)	C-Statistic	0.78 [0.73–0.83]	0.02 [0.02–0.03]	0.76 [0.71–0.80]
	C-Slope	1.04 [0.85–1.22]	0.12 [0.10–0.12]	0.92 [0.74–1.09]
	CITL	0.0001 [-0.04–0.04]	1.83 [1.58–2.04]	-1.84 [-2.08– -1.62]
	Brier Score	0.17 [0.15–0.19]	0.05 [0.04–0.05]	0.12 [0.10–0.14]
	E/O	1 [1–1]	0.13 [0.06–0.18]	0.87 [0.82–0.94]
	Nagelkerke's R^2	0.28	0.06 [0.05–0.07]	0.21
Language	C-Statistic	0.80 [0.80–0.81]	0.03 [0.03–0.04]	0.77 [0.72–0.81]
	C-Slope	1.01 [1.00–1.01]	0.11 [0.10–0.11]	0.90 [0.74–1.05]
	CITL	0.0001 [0.0001–0.0001]	0.99 [0.79–1.19]	-0.99 [-1.19– -0.79]
	Brier Score	0.16 [0.16–0.17]	0.06 [0.06–0.07]	0.10 [0.08–0.12]
	E/O	1 [1–1]	0.00 [-0.12–0.12]	1.00 [0.88–1.12]
	Nagelkerke's R^2	0.32	0.07 [0.07–0.08]	0.25
Memory	C-Statistic	0.76 [0.75–0.77]	0.05 [0.05–0.05]	0.71 [0.65–0.75]
	C-Slope	1.00 [0.99–1.01]	0.13 [0.11–0.13]	0.88 [0.70–1.06]
	CITL	-0.0001 [-0.0001- -0.0001]	0.92 [1.13–0.72]	-0.92 [-1.13– -0.72]
	Brier Score	0.17 [0.17–0.17]	0.08 [0.07–0.08]	0.09 [0.07–0.11]
	E/O	1 [1–1]	0.001 [-0.13–0.13]	0.99 [0.87–1.13]
	Nagelkerke's R^2	0.26	0.09 [0.09–0.10]	0.16
Attention	C-Statistic	0.78 [0.77–0.79]	0.04 [0.03–0.04]	0.74 [0.69–0.78]
	C-Slope	1.01 [1.01–1.01]	0.13 [0.12–0.14]	0.88 [0.70–1.06]
	CITL	-0.0001 [-0.0001– -0.0001]	1.28 [1.07–1.48]	-1.28 [-1.48– -1.07]
	Brier Score	0.17 [0.17–0.18]	0.07 [0.07–0.08]	0.10 [0.08–0.11]
	E/O	1 [1–1]	0.00 [-0.12–0.12]	1.00 [0.88–1.12]
	Nagelkerke's R^2	0.28	0.08 [0.08–0.09]	0.20
Numeracy	C-Statistic	0.74 [0.73–0.75]	0.05 [0.05–0.05]	0.69 [0.63–0.74]
	C-Slope	1.01 [0.99–1.02]	0.18 [0.17–0.19]	0.83 [0.58–1.08]
	CITL	-0.0001 [-0.0001- -0.0001]	1.37 [1.13–1.61]	-1.37 [-1.61– -1.13]
	Brier Score	0.14 [0.14–0.14]	0.07 [0.06–0.07]	0.07 [0.06–0.09]
	E/O	1 [1–1]	0.00 [-0.18–0.18]	1.00 [0.82–1.18]
	Nagelkerke's R^2	0.17	0.07 [0.07–0.08]	0.10
Executive Function	C-Statistic	0.75 [0.74–0.75]	0.04 [0.04–0.04]	0.71 [0.65–0.76]
	C-Slope	0.99 [0.98–1.00]	0.18 [0.16–0.19]	0.81 [0.57–1.05]
	CITL	-0.0001 [-0.0001- -0.0001]	0.98 [0.97–0.99]	-1.55 [-1.78– -1.32]
	Brier Score	0.16 [0.16–0.16]	0.06 [0.06–0.07]	0.09 [0.07–0.11]
	E/O	1 [1–1]	0.24 [0.04–0.44]	1.24 [1.04–1.44]
	Nagelkerke's R^2	0.19	0.06 [0.06–0.07]	0.13
Praxis	C-Statistic	0.69 [0.63–0.75]	0.09 [0.09–0.10]	0.60 [0.53–0.65]
	C-Slope	0.99 [0.97–1.01]	0.33 [0.32–0.34]	0.66 [0.34–0.97]
	CITL	-0.0001 [-0.03–0.03]	1.61 [1.82–1.33]	-1.61 [-1.85– -1.36]
	Brier Score	0.14 [0.14–0.15]	0.09 [0.09–0.10]	0.06 [0.03–0.07]
	E/O	1 [1–1]	0.00 [-0.19–0.19]	1.00 [0.81–1.19]
	Nagelkerke's R^2	0.13	0.12 [0.11–0.12]	0.01

Table 3. Performance metrics all final models pooled across 20 imputed datasets.
CITL=Calibration-in-the-large; E/O=Expected:Observed Ratio; C-Slope=Calibration Slope

Model	Estimate	Model Performance [95% CI]
Overall PSCI (Continuous)	C-Slope CITL MSE Adjusted R^2	1.26 [1.03–1.49] -0.08 [-0.04– -0.12] 0.01 [0.009–0.017] 0.46
Any domain impairment (Binary)	C-Statistic C-Slope CITL Nagelkerke's R^2	0.74 [0.67–0.79] 1.01 [0.67–1.35] -0.06 [-0.33–0.22] 0.22
Language	C-Statistic C-Slope CITL Nagelkerke's R^2	0.77 [0.69–0.84] 6.10 [4.09–8.13] -2.69 [-3.31– -2.08] 0.23
Memory	C-Statistic C-Slope CITL Nagelkerke's R^2	0.62 [0.52–0.71] 3.92 [1.59–6.25] -2.31 [-2.98–1.64] 0.12
Attention	C-Statistic C-Slope CITL Nagelkerke's R^2	0.75 [0.66–0.82] 5.97 [3.62–8.33] -2.78 [-3.50– -2.07] 0.20
Numeracy	C-Statistic C-Slope CITL Nagelkerke's R^2	0.61 [0.51–0.69] 12.39 [8.26–16.53] -3.46 [-4.30– -2.62] 0.03
Executive Function	C-Statistic C-Slope CITL Nagelkerke's R^2	0.74 [0.62–0.83] 8.67 [4.48– 12.86] -3.37 [-4.25– -2.50] 0.03
Praxis	C-Statistic C-Slope CITL Nagelkerke's R^2	0.65 [0.55–0.75] 12.33 [3.90–20.76] -3.42 [-4.55– -2.29] 0.08

Table 4. Model performance following external validation across 20 imputed datasets using OCS-Care data. CITL=Calibration-in-the-large; C-Slope=Calibration Slope

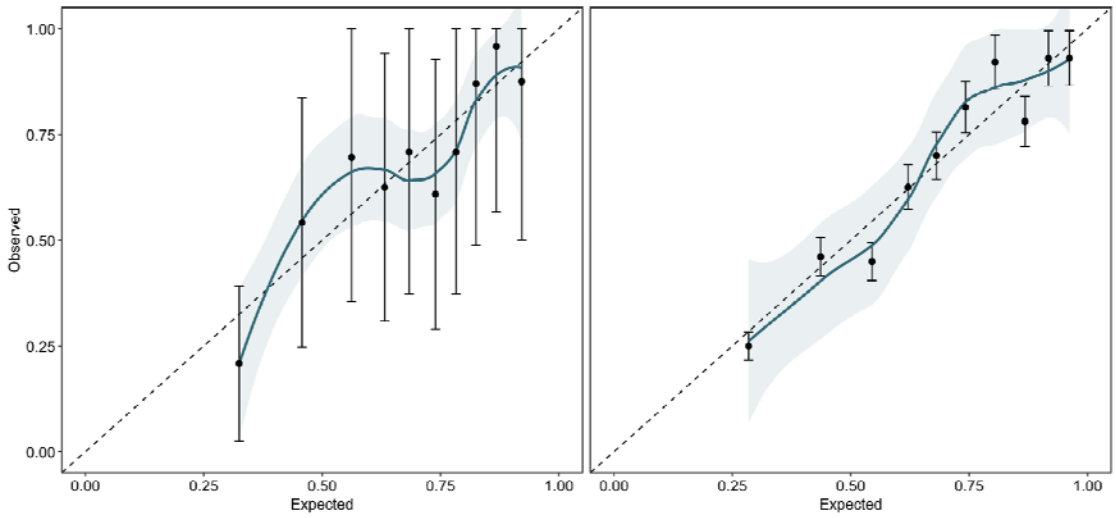

Figure 1. Calibration plot of any 6-month PSCI (0=no impairment, 1=any impairment) in complete cases (left; *N*=237) and across 20 imputed datasets.

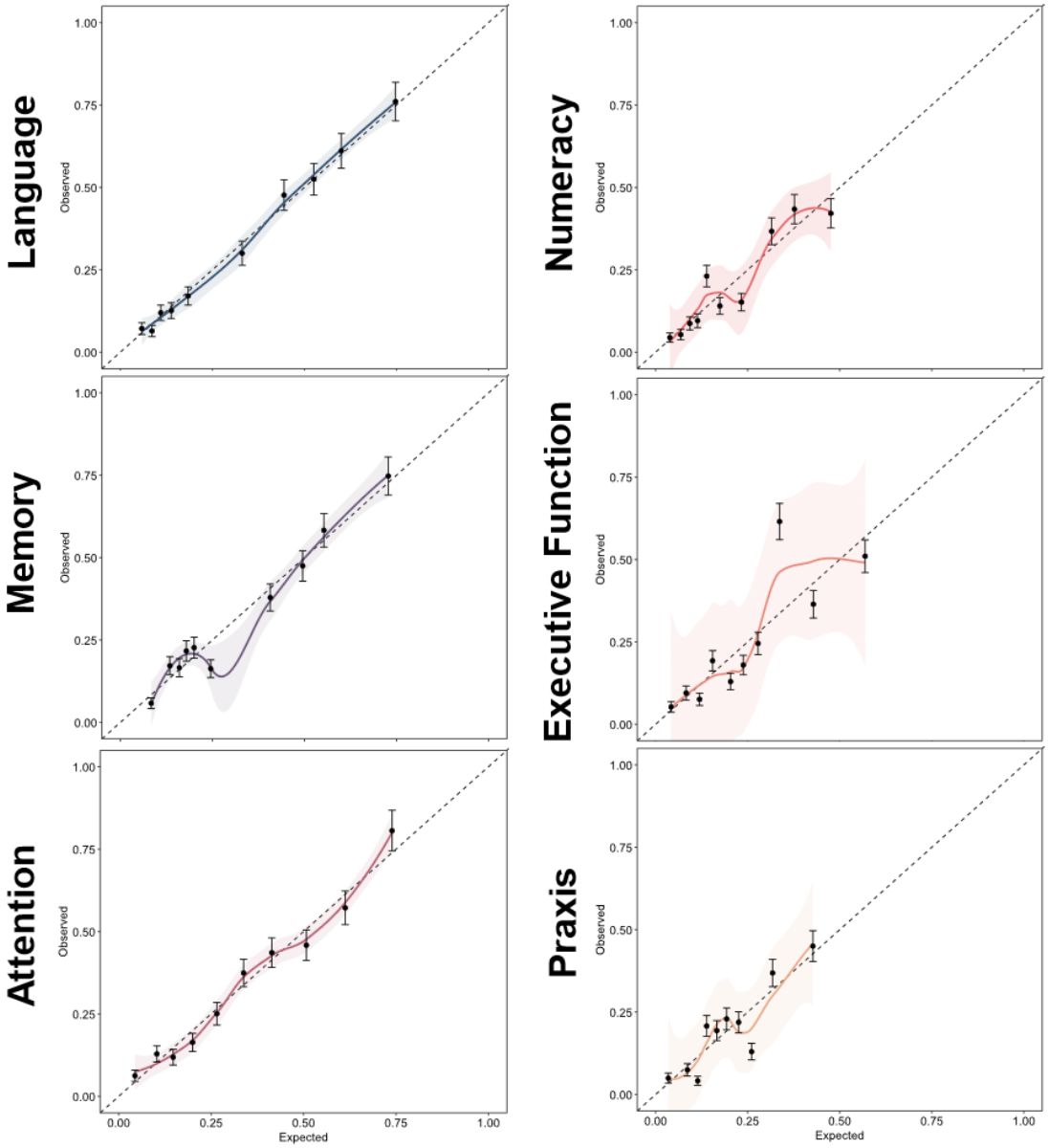

Figure 2. Calibration plots of 6-month domain-specific impairments (0 = no impairment, 1 = any impairment) across 20 imputed datasets.