

A normative model for age-dependent cognitive decline using the FLAME composite score

Jonas Radermacher¹, Abbie Palmer², Clive Ballard², Ellie Pickering², Anne Corbett², Dag Aarsland^{3,4}, Holger Fröhlich^{1,5,6,7}, and the PREDICTOM Consortium⁸

(1) Department of Biomedical AI & Data Science, Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Sankt Augustin, Germany, (2) University of Exeter, United Kingdom, (3) King’s College London, United Kingdom, (4) Stavanger University Hospital, Norway, (5) Bonn-Aachen International Center for IT, University of Bonn, Germany, (6) Institute for Digital Medicine, University Hospital Bonn, Bonn, Germany, (7) Lamarr Institute for Machine Learning and Artificial Intelligence, Dortmund, Germany, (8) <https://www.predictom.eu>



Learn more about
PREDICTOM



See all PREDICTOM
Posters at AAIC 2026

Introduction

- Normative models quantify individual deviations from a reference population via personalized z-scores, standardized against expected population means given covariates (e.g., age, sex)
- While cognition naturally decreases with age, cognitive tests are often evaluated based on thresholds which are not age-corrected.
- They are widely adopted in neuroimaging to characterize heterogeneity in psychiatric and neurological conditions (1).

We developed a normative model for outlier detection of cognitive performance using the FLAME composite score in the PROTECT study and apply it for risk stratification in PREDICTOM.

Methods

Study & Data

- PROTECT Study (2): UK-based online longitudinal cohort (launched 2014); adults ≥ 40 years, no dementia diagnosis; N > 50,000 participants.
- FLAME composite score: Computerized neuropsychological battery measuring attention, memory & executive function; composites normalized to [0, 100] and averaged into one global score. (3)
- IQCODE-S Informant Questionnaire on Cognitive Decline in the Elderly (4) (16-item self-report, threshold $\tau_{IQ} = 3.3$) used for labeling:

	All	CU	SCI
N (count)	20944	17136	3808
Age (mean (std))	62.01 (7.59)	61.64 (7.40)	63.67 (8.20)
Sex (female %)	73.20	74.59	66.91
Education (> 12 years %)	75.31	75.23	75.63
FLAME (mean (std))	84.34 (3.36)	84.48 (3.30)	83.75 (3.56)

Table 1. PROTECT Population statistics with a IQCODE-S threshold of $\tau_{IQ} = 3.3$ with CU: $IQCODE-S < \tau_{IQ}$

- The biggest, notable difference is in sex distribution (SCI has fewer females). SCI is also a bit older and has slightly lower FLAME scores, while education is largely the same across groups.

Models

- Models are fit on N = 17136 cognitively unimpaired (CU) and evaluated on N = 3808 subjectively cognitively impaired (SCI) participants as labeled by IQCODE-S threshold τ_{IQ} .
- Bayesian Linear Regression with B-Splines (BLR-B) implemented via pcntoolkit (5).
- Generalized Additive Models for Location Scale and Shape (GAMLSS (6)) implemented via PyNM (1).

Evaluation

- 5-fold cross-validation and model selection with Bayesian Information Criterion (BIC).
- Root Mean Square Error (RMSE): Goodness of fit metric on original FLAME score scale.
- Normative z-score distribution: Expect mean at 0 for CU and negative for SCI.
- Welch t-test (7): Tests whether z-score distributions of CU and SCI differ significantly.
- Cohen’s d (8): Effect size of CU and SCI z-score distributions separation in units of pooled standard deviation.

Outlier Detection

- One-sided z-test on normative z-scores (lower deviations only).
- Map z-scores to p-values via CDF.
- Correction for multiple testing via Benjamini-Hochberg correction (9).
- Subjects with corrected $p < 0.05$ labeled as outliers.

Results

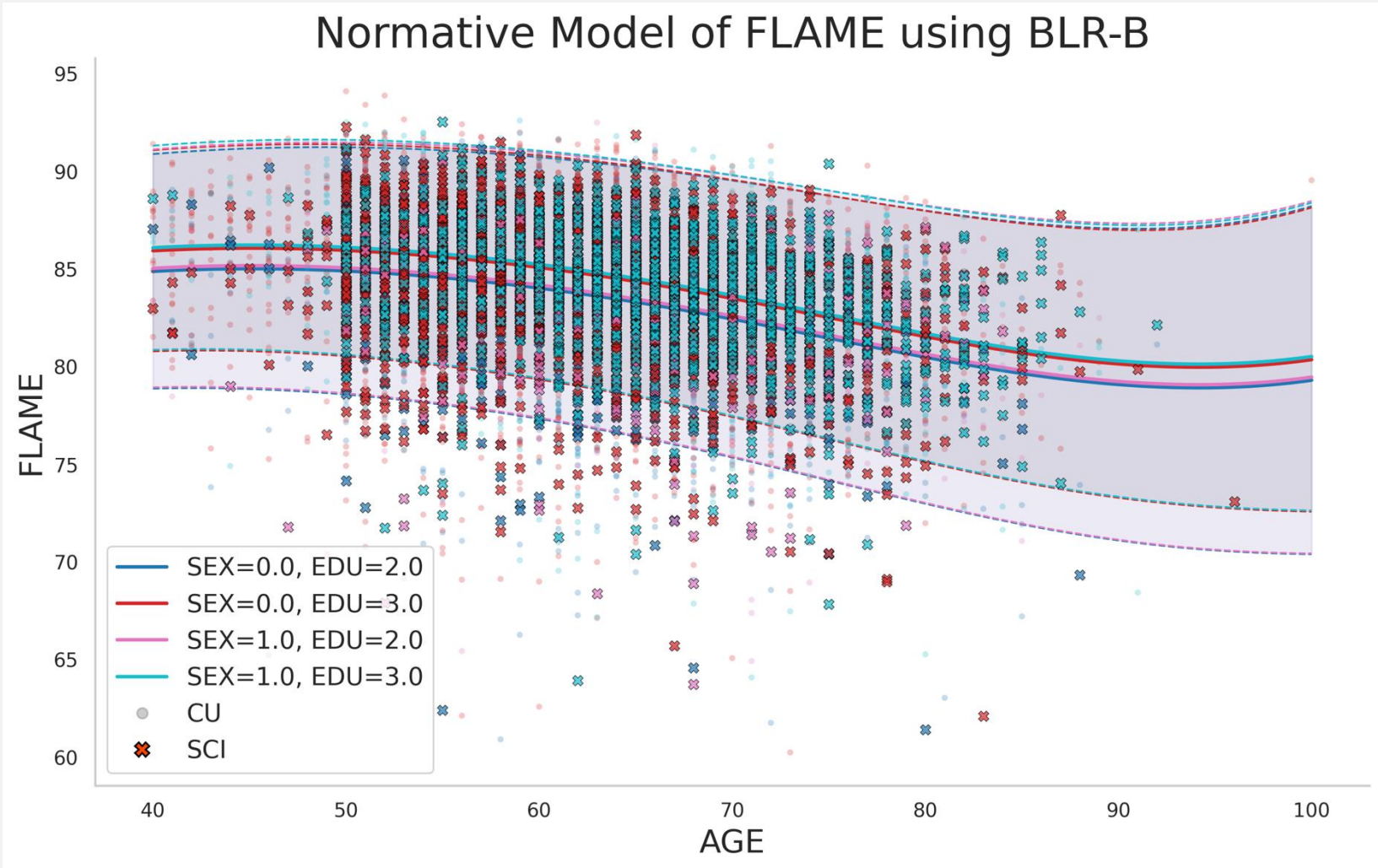


Figure 1. Normative Model of FLAME using BLR-B. SEX (0:female, 1: male), EDU (2: ≤ 12 years, 3: > 12 years)

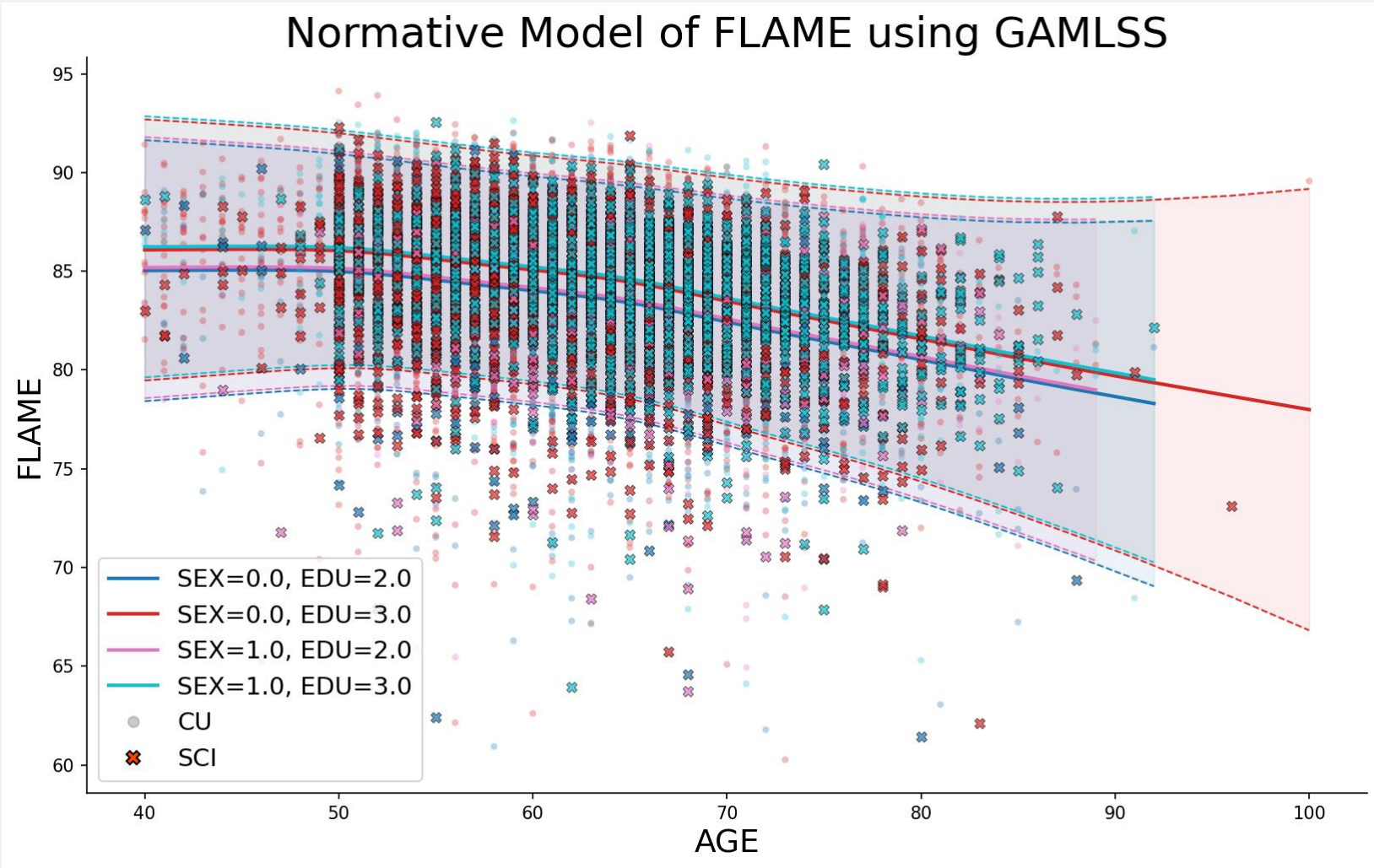


Figure 2. Normative Model of FLAME using GAMLSS. SEX (0:female, 1: male), EDU (2: ≤ 12 years, 3: > 12 years)

	RMSE	Coverage	p-ttest	Cohen’s d	Specificity	PPV
BLR-B	3.09	95.8 %	< 0.001	0.14	94.6%	0.23
GAMLSS	3.09	95.7 %	< 0.001	0.13	94.4%	0.23

Table 2. Normative Model results.

RMSE: Root Mean Squared Error, P-ttest: P-value of t-test, PPV: Positive Predictive Value

Summary

Both models lead to well-calibrated, similar fits, following the non-linear decrease of FLAME with age. Sex plays a minor role, while different education levels shift the normative fit by a few points. GAMLSS gives more realistic extrapolation towards higher ages and more flexible prediction intervals in the bulk region of 50-80 years.

Calibration

- Mean z-score of the CU close to 0 and a negative mean z-score of the SCI.
- Approximately normally distributed z-scores for CU.
- Small RMSE of around 3
- Coverage close to the expected 95% prediction interval.

Discriminative Power

- The z-score distributions are significantly different (significant t-tests) but only with a small effect size of 0.13.
- The specificity was found to be high at 94% with a small positive predictive value at 23% (prevalence of 18%)

Conclusion

- Our normative models allow for the correction for the effects of age, sex and education on the cognitive performance given by the at-home FLAME test.
- Individual scores can be mapped to deviations from the norm in terms of z-scores.
- Using these lead to an enrichment of 28%, improving risk stratification (see P1706).

References

- (1) <https://doi.org/10.1101/2021.06.14.448106> (2) www.protectstudy.org.uk (3) <https://doi.org/10.1002/dad2.12098>
- (4) <https://doi.org/10.1002/14651858.CD010079.pub3> (5) <https://doi.org/10.1038/s41596-022-00696-5> (6) <https://doi.org/10.1201/9780429298547>
- (7) <https://doi.org/10.2307/2332510> (8) <https://doi.org/10.4324/9780203771587> (9) <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

Abbreviations: BIC - Bayesian Information Criterion, BLR-B - Bayesian Linear Regression with B-Splines, CDF - Cumulative Distribution Function, CU - Cognitively Unimpaired, EDU - Education, FLAME - Factors of Longitudinal Attention Memory and Executive Function, GAMLSS - Generalized Additive Models for Location Scale and Shape, IQCODE-S - Informant Questionnaire on Cognitive Decline in the Elderly Self-report, PPV - Positive Predictive Value, RMSE - Root Mean Squared Error, SCI - Subjective Cognitive Impairment

#1702



Jonas Radermacher

Fraunhofer SCAI

Biomedical AI & Data Science

jonas.radermacher@scai.fraunhofer.de

