Performance of a Chinese Cognitive Decline Risk Model in a Japanese Cohort: A Validation Study

Lihui Tu

Social Cognition and Behavior Collaboration Unit, CBS-TOYOTA Collaboration Center (BTCC)

RIKEN Center for Brain Science, Wako, Saitama, 351-0198, Japan

Introduction

Cognitive decline is a major health problem in aging societies. Building models that can predict decline and testing them in different countries is important for future clinical use[1, 2]. Our study is **to develop and validate a risk prediction model for cognitive decline in Chinese older adults, with temporal and cross-national validation**.

Methods

We compared sevenl machine learning algorithms, including survival tree–based methods and ensemble models, using cognitively normal participants from the **China Health and Retirement Longitudinal Study** (CHARLS, 2011–2015). The **Cox proportional hazards model** was selected for its balance of performance and interpretability. This model was trained on CHARLS and validated temporally on a later CHARLS wave (2015–2018) and externally on the **Japanese Study of Aging and Retirement** (JSTAR, 2007–2009). Predictors were harmonized across cohorts, and performance was evaluated by discrimination, calibration, risk stratification, and decision curve analysis (data not shown).

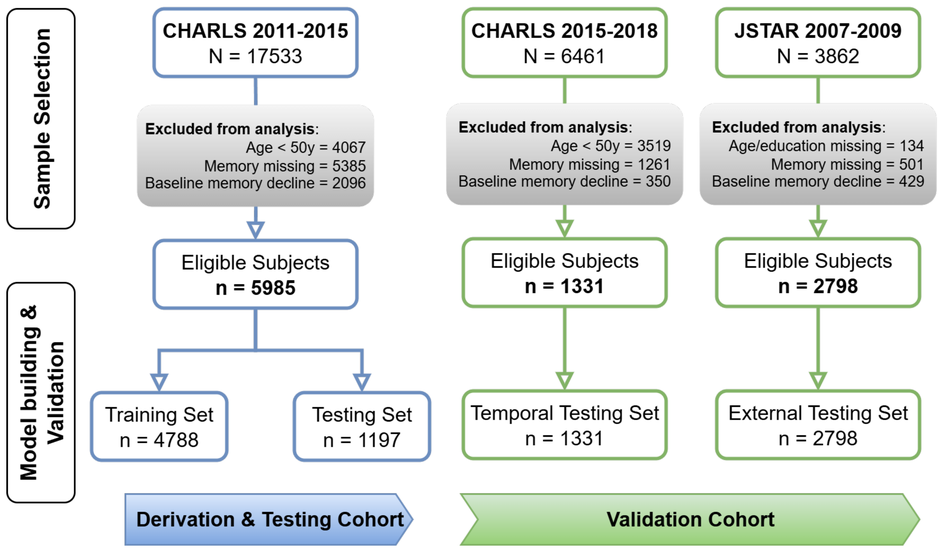

Figure 1. Cohort and subject selection

Table 1. Baseline Characteristics by Cohort

Variable	Training Set CHARLS (n=5985)	Validation Set CHARLS (n=1331)	*p* *	Validation Set JSTAR (n=2798)	*p* **
Age and group	59.0(55.0,65.0)	58.0(52.0,64.0)	<0.001	62.0(57.0,68.0)	<0.001
<65	4457 (74.5)	1006 (75.6)		1693 (60.5)	
65–74	1288 (21.5)	264 (19.8)	0.285	1062 (38.0)	<0.001
≥75	240 (4.0)	61 (4.6)		43 (1.5)	
Female	3008 (50.3)	636 (47.8)	0.108	1478 (52.8)	0.026
Education group					
No formal schooling	1678 (28.0)	235 (17.7)	<0.001	0 (0.0)	<0.001
Primary/Middle	3585 (59.9)	858 (64.5)		810 (28.9)	
High School and above	722 (12.1)	238 (17.9)		1988 (71.1)	
Grip strength§	437 (8.3)	81 (7.6)	0.5	212 (7.7)	0.412
Memory score	8.0(6.0,10.0)	8.0(6.0,10.0)	<0.001	10.0(9.0,12.0)	<0.001
Orient score§§	3.0(3.0,4.0)	4.0(3.0,4.0)	<0.001	6.0(6.0,6.0)	<0.001

Data are presented as median (IQR) or n (%).
Mann–Whitney U test for continuous variables, and chi-square for categorical variables.
p * Validation set 1 (CHARLS) vs. training set; *p* ** Validation set 2 (JSTAR) vs. training set.
§Missing number is 713 (11.9%), 266 (20.0%), and 56 (2.0%), respectively.
§§ Missing number is 225 (3.8%), 37 (2.8%) and 0, respectively.

Table 2. Performance of the Final Model in Validation Cohorts

Model	C-index	CD-AUC	IBS	RMSE
Cox Proportional Hazards Model without elastic net penalty	**0.627 (0.597–0.657)**	**0.656 (0.632–0.680)**	**0.120 (0.115–0.124)**	**0.016**
Final model – Temporal Validation (CHARLS)	0.722 (0.677–0.764)	0.701 (0.641–0.756)	0.066 (0.059–0.074)	0.072
Final model – External Validation (JSTAR)	0.682 (0.654–0.713)	0.689 (0.660–0.724)	0.032 (0.028–0.037)	0.013

Values represent the mean and 95% confidence interval.
C-index, concordance index based on inverse probability of censoring weights; CD-AUC, cumulative dynamic area under the curve; IBS, integrated Brier score; RMSE, root mean

Table 3. Key Predictors of Cognitive Decline in the Development Cohort (CoxPH Model)

Variable	β (Coef)	HR (exp(β))	p-value
Age: <65	-0.53	0.59	<0.001
Age: ≥75	0.26	1.3	0.016
Sex: Male	0.38	1.47	<0.001
Education: Primary/Middle	0.45	1.56	<0.001
Grip strength (kg)	-0.16	0.85	<0.001
Memory score	-0.33	0.72	<0.001
Orientation score	-0.15	0.86	<0.001

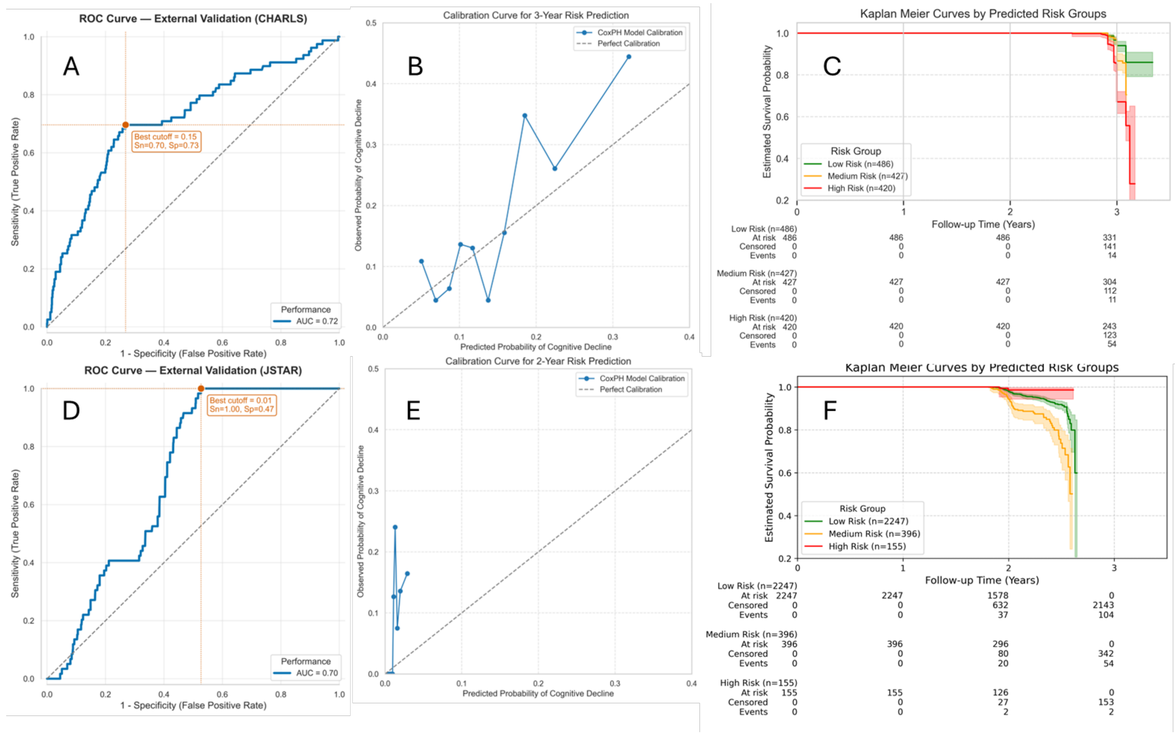

Figure 2. External Validation of the CoxPH Risk Prediction Model in CHARLS and JSTAR Cohorts

Panels A–C show external validation results in the CHARLS cohort (2015–2018)

Panels D–F show results in the JSTAR cohort.

The CoxPH model demonstrated good discrimination (AUC 0.70–0.72), acceptable calibration, and clear separation of survival curves across risk groups, indicating consistent predictive performance across cohorts.

Results

In temporal validation, the model showed good discrimination and calibration. In the external validation, it retained fair discrimination (C-index ≈0.68) and separated participants into distinct risk groups, but calibration was poor. Decision curve analysis (data now shown) indicated limited net benefit in the external JSTAR cohort.

Discussions

1. Model performance and robustness
The CoxPH model developed in CHARLS showed good discrimination and reasonable calibration; External validation confirmed consistent ranking ability for cognitive decline risk.

2. Calibration and risk transferability
The model maintained relative risk ordering but underestimated absolute risks in the Japanese cohort, indicating the need for population-specific recalibration.

3. Risk stratification utility
Risk grouping (low, medium, high) showed clear survival separation in both cohorts, supporting its potential for population screening and risk communication.

4. Cross-cultural implications
Predictive models may not directly generalize across populations, even with similar ethnicity or region. Rigorous external validation and recalibration are crucial.

Conclusion

A cognitive decline risk model developed in a Chinese cohort demonstrates partial transportability to a Japanese cohort, underscoring both the promise and limits of transferring interpretable models across populations.

References

1. 1. Sakal C, Li T, Li J, Li X. Identifying Predictive Risk Factors for Future Cognitive Impairment Among Chinese Older Adults: Longitudinal Prediction Study. JMIR Aging. 2024 Mar 22;7:e53240.
2. 2. Finney CA, Brown DA, Shvetcov A. Developing multifactorial dementia prediction models using clinical variables from cohorts in the US and Australia. Transl Psychiatry. 2025 Jan 21;15(1):1–10.

Acknowledgements

1. The RIKEN Center for Brain Science (CBS) - Toyota Collaboration Center (BTCC) funded this research.
2. JSTAR was conducted by the Research Institute of Economy, Trade and Industry (RIETI), Hitotsubashi University, and the University of Tokyo.