

Predicting Cognitive Impairment from Language

Carnegie Mellon University
Statistics & Data Science

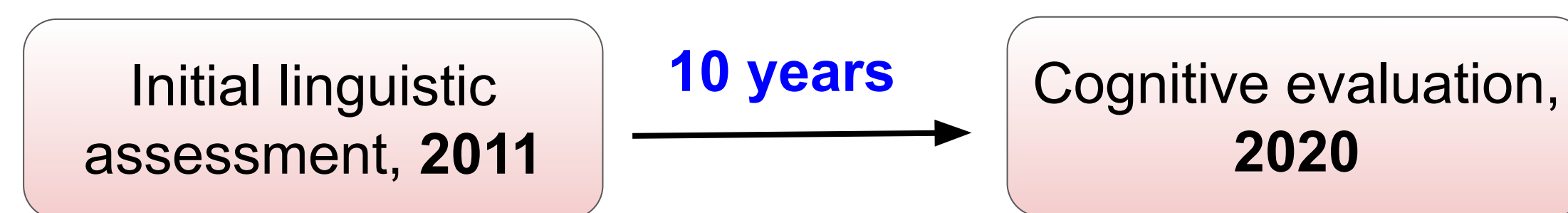
Authors: Xinfei Cen, Jin Yu Kim, Yuntian Shen, Ziyang Wang
Advisor: Dr. Joel Greenhouse Client: Dr. Davida Fromm



Introduction

Research Question: Can past linguistic abilities indicate cognitive impairment occurrence a decade later?

- Data source: Wisconsin Longitudinal Study (WLS)



Data

Data Description

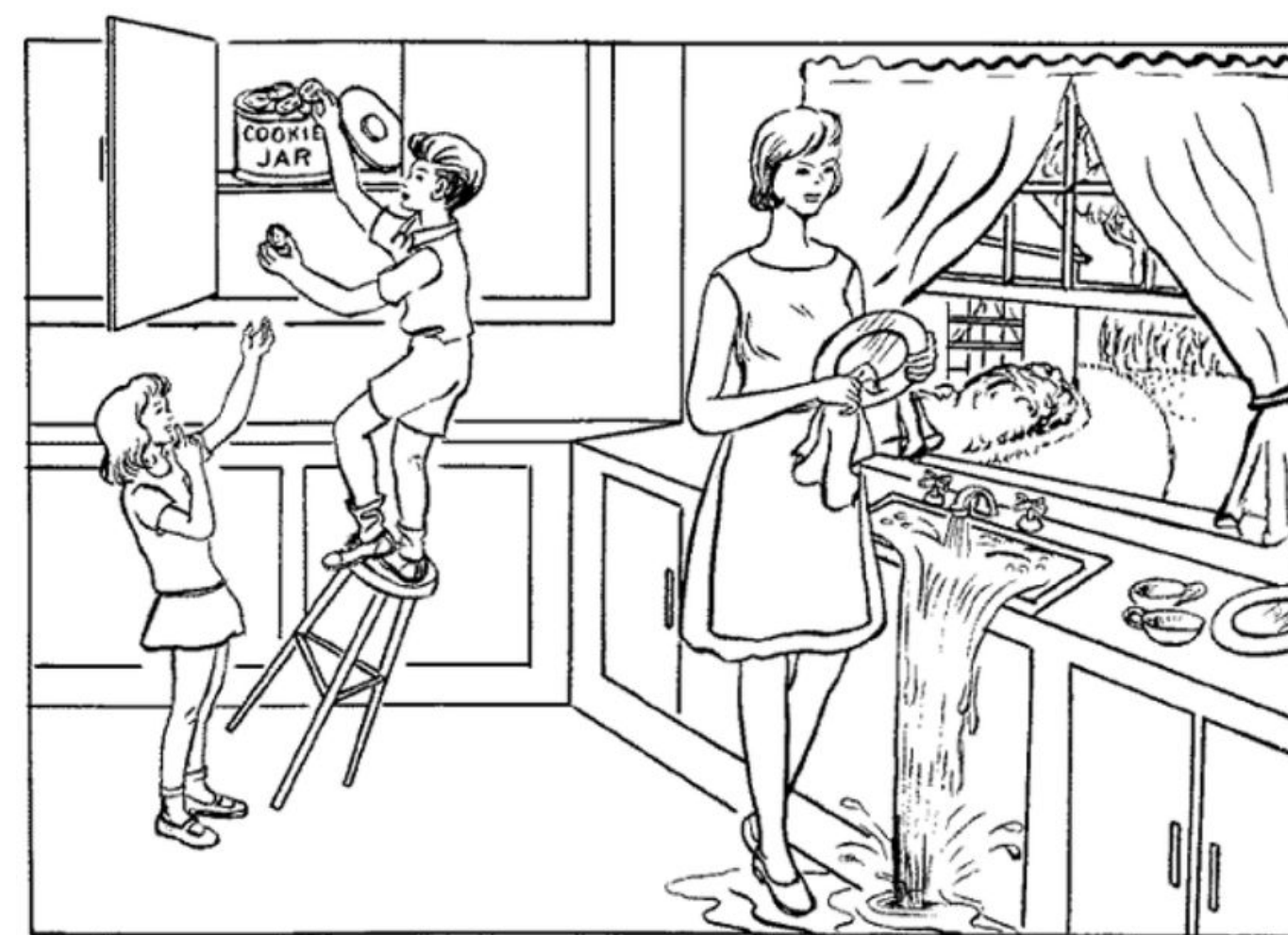
- 801 participants of the WLS, completed Cookie Theft Task in **2011**
- Cognitive status assessed in **2020**:

- Normal Cognition: 698**
- Cognitively Impaired: 103**

Predictors:

Linguistic abilities assessed via the Cookie Theft Task, categorized into three groups:

- General Linguistic Ability
- Linguistic Disfluency
- Core Lexicon Term Usage



Cookie Theft Picture, an assessment tool to investigate linguistic abilities

Methods

Approach

Logistic Regression

- Stepwise; separate models from each of three sub-categories; combined model

(i) Stepwise Logistic Regression, adjusted for Age and Sex



(ii) Assess Multicollinearity using VIF and Correlation Matrix



(iii) Assess Goodness of Fit (Hosmer Lemeshow Test)

Results

Summary of logistic regression models built using selected variables, with age (statistically significant in all models) and sex as control variables. Only statistically significant variables $p < .10$ are reported, except for the final model

	Variable	Coefficient (SE)	p-value	Goodness of Fit (HL test) p-value
General linguistic	Grammatical complexity index	0.05 (0.03)	0.081	0.29
	log(Duration)	-1.32 (0.40)	0.001	
	log(Lexical diversity)	-3.38 (1.24)	0.006	
	log(Words/minute)	-1.26 (0.50)	0.012	
Disfluency	sqrt(Total words, without repetition and revision)	-0.09 (0.04)	0.039	0.07
Corelex	# of omitted words	0.12 (0.06)	0.035	0.23
Final Model (all variables used are reported)	Grammatical complexity index	0.05 (0.03)	0.089	0.21
	log(Duration)	-1.19 (0.42)	0.005	
	log(Mean length of utterance in words)	-0.21 (0.32)	0.506	
	log(Lexical Diversity)	-3.37 (1.24)	0.006	
	log(Words/minute)	-1.14 (0.52)	0.027	
	# of omitted words	0.06 (0.07)	0.369	

Conclusion

Main Findings

- Measurements of linguistic ability are able to predict cognitive impairment **10 years later**.
- Other analytic methods used: random forest, survival analysis (no significant findings).

Implication

- Assessment of linguistic ability help with early detection of cognitive impairment.
- Contribute to the broader field of aging research through analysis of longitudinal data.

References

TalkBank Repository. Retrieved from <https://talkbank.org/>

Lanzi, Alyssa M., et al. "DementiaBank: Theoretical Rationale, Protocol, and Illustrative Analyses." 2022.
Williams, Victoria J., et al. "Assessing Dementia Prevalence in the Wisconsin Longitudinal Study: Cohort Profile, Protocol, and Preliminary Findings." 2021.