

What Post-Retirement Careers Can Protect Seniors' Cognitive Abilities?

Jingxin Wang, Qi Chen, Erwen Wu, Yuhan Pan, Ruijia Zhang

Introduction

In Canada, the aging population is expanding rapidly. By 2040, nearly one in four Canadians will be age 65 and above, with the population aged 80 and above growing even more rapidly [5]. Among age-related health concerns, dementia refers to progressive impairment of behavioural and cognitive functions, and the CDC has ranked Alzheimer's Disease, a type of dementia, the seventh leading cause of death in the U.S in 2022 [4]. As the complications deepen, the patient increasingly relies on caregivers. In Ontario, 24/7 in-home care range from \$10,000 to \$20,000 per month [2], presenting a significant financial burden to families and the government. In midst of continuing shortages in medical professionals [3] and the long training cycles [1] required to expand the healthcare workforce, preventative strategies are becoming increasingly important. Slowing cognitive decline may not only enhance the quality of life for older adults but also reduce the economic and caregiving burdens on families and the healthcare system [2]. Post-retirement activities may serve as accessible and modifiable protective measures. For example, Lee et al. (2019) found that older adults who remained engaged in career and leisure activities after retirement experienced better cognitive performance compared to those who remained inactive [6]. However, existing research in this area remains fragmented and is often limited to specific types of activity. There is still limited understanding of what activities are most beneficial for cognitive health and cost-saving, particularly among older adults in Canada.

This study aims to evaluate effect of different post-retirement occupation on cognitive performance of elderly people, and to understand similarities in the population that experience no or mild cognitive decline. This may involve deploying a prediction model on the Shinny app that forecasts the cognitive scores of an aged person in 5 years following retirement, based on post-retirement career of their choices, demographics, health conditions, and employment history. This data-driven solution can support older adults to make informative plan and relieve potential healthcare financial burden.

Research Question

Question: what post-retirement careers can protect seniors' cognitive abilities and reduce financial burden in health care?

Rationale: Lee et al. (2019) have shown that older adults who are still working after retirement have significantly better cognition than those who are not [3]. A further question to ask is: how should elderly people plan their career after retirement based on their previous experience? By identifying occupations that best suit old adults based on their work experience and health

conditions, old adults can find new careers meaningful to their later years, retain and utilize their intelligence to create value. Consequently, the burden to the healthcare system is relieved.

Method

Due to long data request period on Canadian Longitudinal Study on Aging that provides desired datasets about seniors aged 65 and over, retirement data in the U.S is used instead for the study population since both countries are facing similar issues in aging and healthcare. The data source is [RAND HRS Longitudinal File 2022](#) derived from the University of Michigan Health and Retirement Study (HRS), a national panel survey of individuals over age 50 and their spouses. HRS' main goal is to provide panel data that enables research and analysis in support of policies on retirement, health insurance, saving, and economic well-being. The survey elicits comprehensive information such as demographics, health, cognition, health care, job status and history. The file consists of 8 cohorts that are identified by their birth years and the first year of interview.

Our interested variables of outcomes are mostly positive and continuous, including 1) the mental status index that summarizes scores from counting, naming, and vocabulary tasks, 2) the recall index that summarizes the immediate and delayed word recall tasks, 3) cognitive impairment (categorical) such as getting lost, wandering, and hallucination.

Primary predictor variables of interests include 1) whether they are working for pay (categorical), 2) the current occupation categories, such as managerial specialties, sales, tech, food preparation, etc. 3) if they are working, hours per week the current job (positive, continuous) will be assessed, 4) hourly wage rate (positive, continuous), 5) time before and after retirement, 6) types of effort and required and frequency in the job (categorical), such as heavy lifting, eye sight, stress levels, etc. For participants who have retired, our model will include their retirement age and years after retirement (positive, continuous). For participant who have not retired, our model will include their planned retirement ages and retirement plans.

Mediator variables include, 1) demographic: age at interview, sex, birth year, census region, race, years of education, marital status, number of people living in the household, number of children, number of living siblings; 2) socioeconomic: individual earnings, household capital income, number of pensions currently receiving; 3) health behaviors: smoking status and drinking habits, 4) medical information: BMI, disabilities, depressive symptoms, function limitations such as difficulties in shopping, cooking, phone usage, finance, and medication.

In the exploration stage, 80% of all individuals will be randomly sampled and allocated to the training set, whereas 20% will be allocated to the testing set. Descriptive analysis will be performed on the training set. For continuous metrics, their five-number summaries, standard deviations, and Pearson coefficients will be computed. If the variable has a skewed distribution, power transformation will be applied make the distribution symmetric. Centering and scaling will

be applied to variates with dissimilar scales. Years of interviews will be standardized into years before and after retirement. To visualize the relationships between primary predictors and outcomes, scatter plot will be created. For categorical covariates, bar plots and box plots will be created visualize relationships with continuous outcome variables. Odds ratio will be calculated to quantify association with discrete outcomes. For variables with less than 50% missing values at random, NA values will be linearly interpolated by adjacent values in time of the same person. Then, a Principal Component Analysis and UMAP will be performed on continuous variables. By coloring points by cognitive scores, employment status, occupation types, and types of effort required at workplace, we can visualize the distribution and clustering of data in a lower dimensional space, identify outliers, and inform model selection.

Generalized mixed effect models is suitable for modelling the longitudinal data. The model will be a linear regression for recall and mental state index as the outcome and logistic regression for cognitive impairment. The reason for choosing generalized mixed effect models is because responses collected from the same individual are correlated. The mixed effect model allows us to separate sources of variations and quantifies non-noise variations, which allows us to generalize our findings. Additionally, as random effects use fewer degrees of freedom than fixed effects, the power of hypothesis test will also increase.

To implement, intercept, time, census region will be treated as random effects, while all others are fixed effects. In a linear model, the interpretation of coefficient of fixed effects is the change in cognitive performance score for every unit increase in the target covariate. In a logistic regression, the coefficient represents the change in log odds ratio of having cognitive impairment for every unit increase in the target covariate. Random effects are deviations from the baseline effect among participants. The significance of fixed effects can be tested via ANOVA table, and that of random effects can be tested via likelihood ratio test. Finally, the model performance can be evaluated by the sum of squares. A feature selection with Lasso will be performed on the training set to select important variables. Selected predictors will be used to predict outcomes on the testing set and evaluate the final model performance.

If time permits, several secondary analyses may be conducted. First, propensity score matching will be used to create comparable groups of working and non-working participants, helping to reduce confounding in observational data. Second, a causal forest model can estimate heterogeneous treatment effects and explore how different post-retirement careers impact cognitive outcomes across subgroups. Finally, a 5-fold cross-validation can assess the stability and uncertainty of our primary predictors across the full dataset.

If the mixed effect model suffers from over parametrization or overfitting, Long Short-Term Memory (LSTM) networks are well-suited for modeling temporal patterns in longitudinal data. LSTMs can capture complex, nonlinear dependencies across time, making them valuable for predicting how occupational engagement influences cognitive decline and healthcare expenditures.

Recent work by Faghihpirayesh et al. (2025) highlights the effectiveness of LSTM-based architectures in detecting meaningful longitudinal changes, principles that are highly transferable to our survey-based cognitive health data.

Timeline and Expected Outcomes

5/17—5/25: exploratory data analysis, data extraction and cleaning

5/26—6/01: initial modelling pipeline, hypothesis testing, and initial modelling result

6/01—6/08: draft initial paper, peer review to refine research questions and statistical methods

6/09—6/15: revise modelling pipeline; create Shiny app in sketch; revise initial draft on an ongoing basis

6/16--6/22: perform secondary analyses

6/22—6/27: finalize the paper; summarize the paper into an abstract, submit on 6/27

6/28—7/4: create presentation slides and video, submit on 7/4

7/4—7/10: presentation rehearsals for conference on 7/10

References

1. Canadian Medical Association (CMA). (n.d.). *How do I become a doctor in Canada?* <https://www.cma.ca/healthcare-for-real/how-do-i-become-doctor-canada>
2. ConsidraCare. (2023). *How much does home care cost for seniors in Ontario?* <https://www.considrare.com/how-much-does-home-care-cost-for-seniors-in-ontario/#:~:text=Private%20home%20care%20costs%20for,complex%20medical%20conditi ons%20or%20disabilities>
3. Government of Canada. (2024, July 5). *Supporting Canada's health workers by improving health workforce research, planning, and data* [News release]. Health Canada. <https://www.canada.ca/en/health-canada/news/2024/07/supporting-canadas-health-workers-by-improving-health-workforce-research-planning-and-data0.html>
4. Kumar A, Sidhu J, Lui F, et al. Alzheimer Disease. [Updated 2024 Feb 12]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK499922/>
5. Public Health Agency of Canada. Aging and chronic diseases: A profile of Canadian seniors [Internet]. www.canada.ca. 2020. Available from: <https://www.canada.ca/en/public-health/services/publications/diseases-conditions/aging-chronic-diseases-profile-canadian-seniors-report.html>
6. Lee Y, Chi I, Palinkas LA. Retirement, Leisure Activity Engagement, and Cognition Among Older Adults in the United States. *Journal of Aging and Health*. 2018 Apr 21;31(7):089826431876703.
7. Athey, S., & Imbens, G. W. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27), 7353–7360. <https://doi.org/10.1073/pnas.1510489113>
8. Faghihpirayesh, R., Guo, X., Wolf, M. M., Chung, K., & Abdi, M. (2025). Deep-learning framework for analysis of longitudinal MRI studies. *Medical Imaging 2025: Image Processing, SPIE*, 13406, 1340634. <https://doi.org/10.1117/12.3002583>