

Using Weights to Analyze Complex Survey Data

6/1/2018

Presented by: Mark Oremus

School of Public Health and Health Systems



UNIVERSITY OF WATERLOO
FACULTY OF APPLIED HEALTH SCIENCES



Background

- **Population aging and the increased incidence of dementia have called attention to the importance of cognitive function (CF) in healthy aging, and led to the search for factors that can affect CF**
- **CF: intellectual activity involving mental processes, including attention, processing speed, learning and memory, executive function, verbal fluency, and working memory**
- **Evidence suggests social support availability (SSA) may buttress CF through the supportive nature of social mechanisms such as emotional encouragement and interpersonal activity**
- **SSA: extent to which individuals can draw upon people and communities for help, care, and comfort in times of need**



Background

- **Hypothesis: SSA is positively associated with CF**
- **Importance: SSA is modifiable → public health authorities can fund programs such as seniors' centres that provide a base for social support**
- **Previous studies have typically been conducted in persons aged 65 years or older**
 - **Informative yet limited because they do not tell us how relations between SSA and CF across the life course / what happens in middle-age affects older-age**
- **Previous studies often use single measures of cognition, which do not capture the multifaceted nature of CF**



Background

- **Canadian Longitudinal Study on Aging (CLSA):** designed to collect data on the changing biological, clinical, sociological, economic, and psychological aspects of participants as they age
 - Participants aged 45-85 years at baseline AND up to seven different measures of CF, depending on the cohort
- Ideal study to address our research question
- CIHR recently funded a Catalyst Grant to analyze CLSA data and we got funded



CLSA

- **CLSA - two cohorts:**
 - **Tracking: 21,241 participants recruited randomly from the 10 provinces and interviewed by telephone at three-year intervals**
 - **Comprehensive: 30,097 participants recruited randomly from within 25-50 km of 11 data collection sites spread across Canada and interviewed at home and at their local site once every three years**
- **Recruitment undertaken through random digit dialing and targeted mailings using public health administration databases**
- **Some Tracking participants were recruited from CCHS 4.2 Healthy Aging and some Comprehensive participants from Québec's NuAGE study**



CLSA

- Participants were recruited into strata based on ...
 - Province
 - Sex
 - Age group (45-54y, 55-64y, 65-74y, 75+y)
 - Reside within data collection site catchment area (yes/no) – Tracking only
- 136 strata for Tracking and 56 strata for Comprehensive
- Intensive process to calculate sample weights
 - Well-described in a CLSA technical document: <https://www.clsa-elcv.ca/doc/1041>



Sample Weights

- **In surveys involving complex sampling, individuals in the population often have unequal probabilities of participation**
- **Sample weights can help address the issue in analyses**
- **Each survey participant's numeric sample weight is an estimate of the number of people in the source population who are represented by that participant**
- **Prevalence estimates, means, and regression coefficients obtained from analyses adjusted by sample weights apply to all of the individuals in a given source population, not just to survey participants**



Sample Weights

- **Weights are sometimes ignored in analyses**
 - Researchers may not understand the use of weights
 - Method of calculating weights is opaque
- **Evidence on whether the use of weights makes a difference is equivocal**
 - Some comparative analyses show no differences between weighted and unweighted analyses
 - Other analyses show the weights do make a difference
- **CLSA: Psychology working group found the use of weights did not impact their results during the development of standardized scores for the neurocognitive test battery**



Sample Weights

- We looked at whether the use of weights would affect our examination of the association between SSA and CF
- Important foundational work for our research program, as well as for other researchers' programs
- Variables
 - SSA: average score (range: 1-5) on all 19 questions of the MOS-SSS (perceived support in 19 domains)
 - CF: two domains, memory and executive function
 - Memory: Rey Auditory Verbal Learning Test immediate and delayed recall
 - Executive function: Mental Alternation Test, Animal Fluency Test, and Controlled Oral Word Association Test



Sample Weights

- For each cognitive test, we obtained z-scores separately for English and French speakers
- Z-scores were added together to get scores for each CF domain
- Analyses adjusted for province of residence, age group, sex, education, cigarette smoking, alcohol consumption, hypertension, diabetes/borderline diabetes/high blood sugar, depressive symptoms, any help required on ≥ 1 activity of daily living, and any help required on ≥ 1 instrumental activity of daily living
- Comprehensive data only



SAS Code

SAS procedure to handle the analysis of complex survey data

```
proc surveyreg data=cop2 order=data;  
  weight wghts_analytic_com; strata wghts_geostrat_com; cluster entity_id;  
  class educ agr prov smoke alc hyp dia adl iadl;  
  model mem=osi agr educ Female prov smoke alc hyp dia adl iadl cesdtotal/solution;  
  title 'Regression Analysis MEM & OSI: Weighted';  
run;
```

Weight: CLSA weight that should be used to study relationships between variables

Strata: variable that specifies the 56 Comprehensive strata

Cluster: entity_id is the variable identifying each participant (omitting it does not change the results)



Results

Unweighted and weighted multiple regression analyses		
	Unweighted analysis^b	Weighted analysis
Memory^a	0.1600 (0.1298,	0.1548 (0.1193,
	0.1901)	0.1903)
Standard error	0.0154	0.0181
n	27,954	27,954
Executive function^a	0.3607 (0.2956,	0.3613 (0.2860,
	0.4258)	0.4366)
Standard error	0.0332	0.0384
n	26,765	26,765

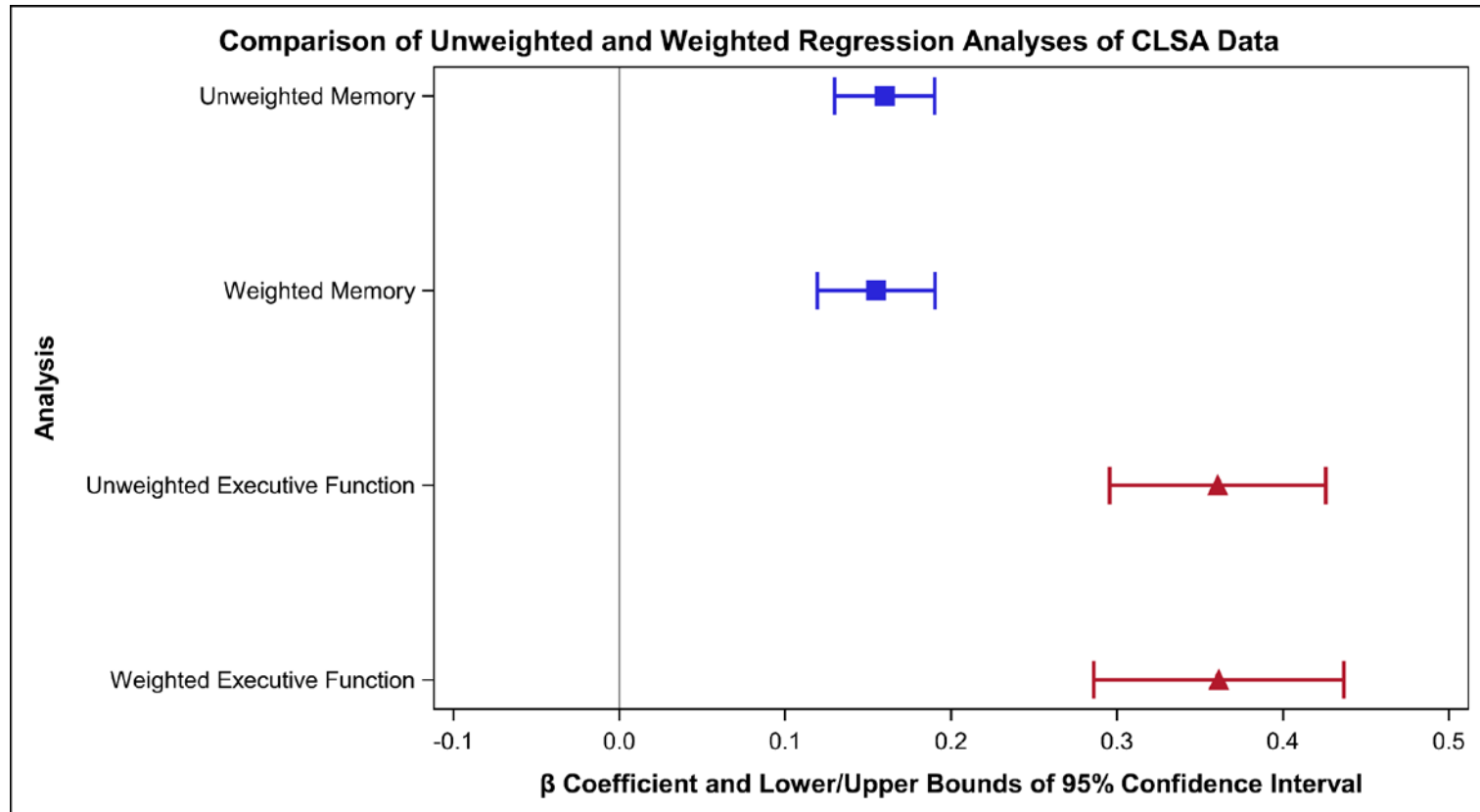
^aRegression coefficient (95% confidence interval) representing the change in cognitive function z-score per one-unit change in the overall social support availability index score, controlling for province of residence, age group, sex, education, cigarette smoking, alcohol consumption, hypertension, diabetes/borderline diabetes/high blood sugar, depressive symptoms, any help on ≥ 1 activity of daily living, and any help on ≥ 1 instrumental activity of daily living.

^bProc Genmod, distribution = normal, link = identity.

Vs = versus.



Results



Wrap-up

- **Achievements:**

- Found little difference between the unweighted and weighted analyses (differences may exist when analyzing other sets of variables)
- Learned how to use proc surveyreg, surveymeans, surveyfreq, surveylogistic, with help from others

- **Advice:**

- Involve a biostatistician in grant planning from the start
- Use weights with CLSA or other complex surveys (technically correct, practicalities of peer review)



Acknowledgements

- **Research Team**

- Candace Konnert (U. Calgary)
- Jane Law (U. Waterloo)
- Colleen Maxwell (U. Waterloo)
- Megan O'Connell (U. Saskatchewan)
- Holly Tuokko (U. Victoria)
- Suzanne Tyas (U. Waterloo)

- **The researchers have no conflicts of interest to declare**



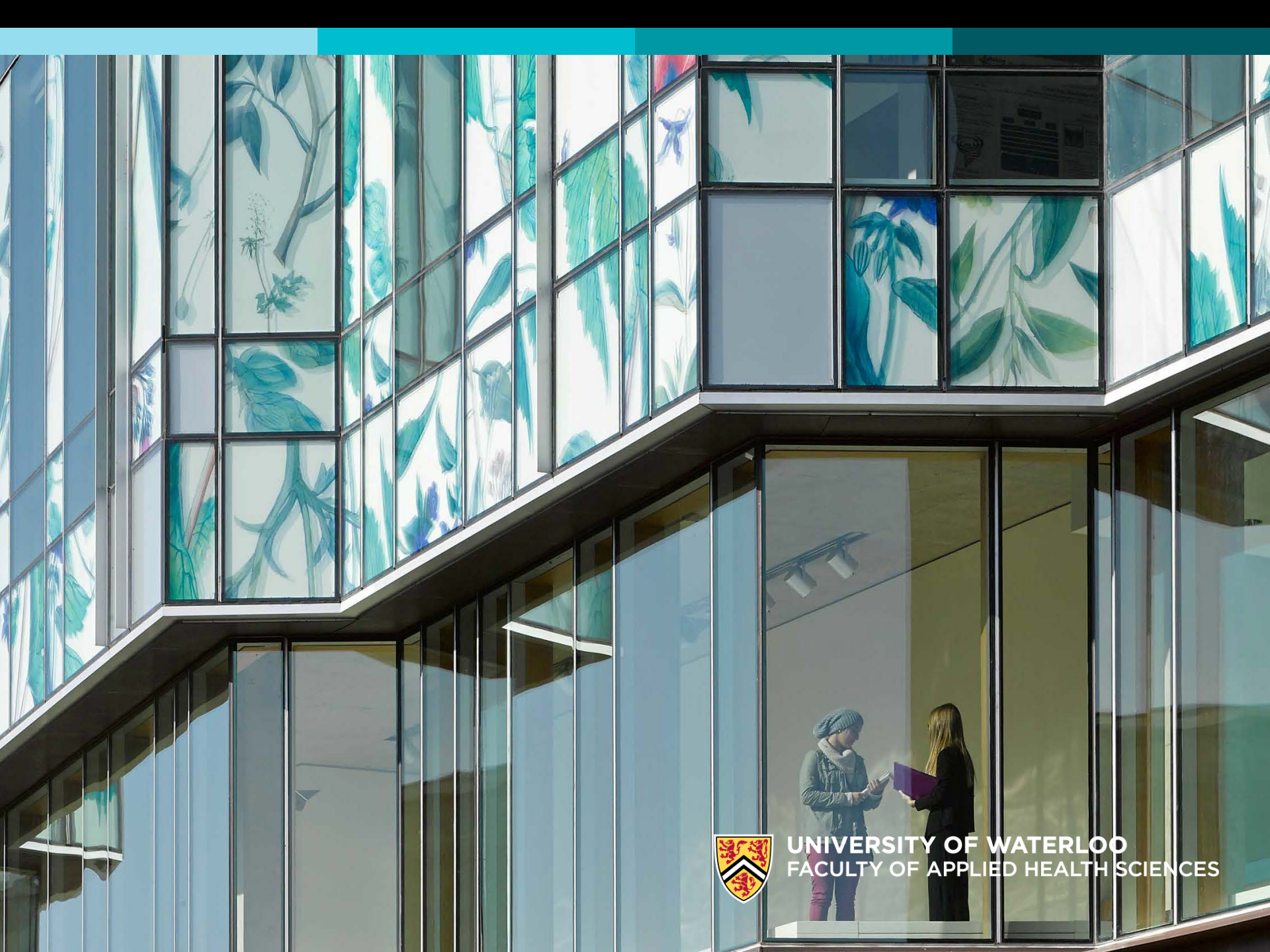
Funding

- **Analysis: CIHR Catalyst Grant to Analyze CLSA Data (ACD-151265)**
- **Travel: Statistical Society of Canada, University of Waterloo**
- **CLSA: Government of Canada through the Canadian Institutes of Health Research under grant reference number LSA 9447 and the Canada Foundation for Innovation. The analyses reported today utilized the CLSA Comprehensive data v.1.0 (Baseline plus Cognition).**



Questions?





UNIVERSITY OF WATERLOO
FACULTY OF APPLIED HEALTH SCIENCES