# Supplementary Materials

## Procedures for administering the Cognitive Measures from the CLSA Comprehensive Cohort (at baseline)

The cognitive measures were administered at two different times during the data collection process: 1) at the participants’ home, and 2) at a CLSA Data Collection Site (DCS). The Rey Auditory Verbal Learning Test (RAVLT), Animal Fluency (AF) and the Mental Alternations Test (MAT) were administered in the participant’s home, whereas the Miami Prospective Memory Test (MPMT), the Stroop Test (Victoria Version), the Controlled Oral Word Association Test (COWAT), and a Choice Reaction Time (CRT) task were administered at the DCS. In-person interviewer training, supplemented by a video training aid illustrating the administration methods, took place at the DCSs. For each task, the examiner made note of any difficulties in administration that could potentially confound performance (e.g., not understanding due to language difficulties, a distracting environment, technical difficulties) and this information is available in the CLSA data.

For all participants, the order of administration of the cognitive measures was consistent, with the in-home interview occurring prior to the participant attending the DCS. During the in-home interview, each participant provided signed informed consent for taking part in the interview and for the audio-recording of the session. Audio-recordings of participant responses facilitated accuracy of scoring. The interviewers noted information concerning the quality of the audio-recordings for each cognitive measure and this information is available in the CLSA data as meta-variables.

 The three cognitive measures administered during the in-home interview (RAVLT, AF, MAT) were administered via standardized recordings using Computer Assisted Personal Interviews (CAPI). The recordings were available in French or English, and they were developed by Statistics Canada for the Canadian Community Health Survey – Healthy Aging (Statistics Canada, 2009). These recordings are the same as the CLSA Tracking Cohort cognitive measures (RAVLT, AF, MAT) that were administered over the telephone.

The remaining four cognitive measures were administered at a CLSA DCS in a standardized order, starting with the first MPMT task (the Event-based task), which requires a response after 30 minutes of controlled distraction. During the 30-minute interval, the Stroop Test, COWAT, and CRT were administered, in that order. After completion of the MPMT Event-Based task, three data collection modules (i.e., Social Network, Social Support and Social Participation) were administered. Subsequently, the final cognitive measure, the second MPMT task (the Time-based task) was administered. This task requires a response after 15 minutes of controlled distraction and during this 15-minute interval, the Disease Symptoms module was administered.

Administration and participant responses were in either French or English. During some tasks, some participants switched between French and English. For this reason, variables indicating language of response may include English, French, or “bilingual” which means that some responses were provided in French and some responses were provided in English. These responses were coded as bilingual for Animal Fluency and COWAT, but for all other measures, participants were requested to respond in the language in which the measure was administered. Scores on the measure may be affected by the language used by the participant. It is of utmost importance to note the language of administration when evaluating these scores. If participants switch between languages, this changes the nature of the task. A number of language variables are available in the CLSA dataset for each measure in the cognitive module. All responses were recorded electronically, with meta-notes indicating if there were anomalies or if a participant chose not to complete the measure.

## Administration and Scoring

For further information on the administrative and scoring procedures of the RAVLT, AF and MAT, see Tuokko, Griffith, Simard & Taler (2017). The administration script used for the Animal Fluency task in the CLSA was derived from the Canadian Community Health Survey (CCHS) (2008), which was adapted from the Study of Health, Ageing and Retirement in Europe (SHARE) (2004). Unlike other versions of the AF, this script communicates to the participant the time limit of the task (i.e. “You have one minute to do this”). Details of the CLSA data collection software and entry/storage procedures are provided in the CLSA combined-protocol (version 3) and can be found at <https://clsa-elcv.ca/doc/511>.

The MPMT comprised two measures of Prospective Memory: 1) an Event-based prospective memory task, and 2) a Time-based prospective memory task. In the Event-based task, participants are shown an envelope containing currency (three one-dollar coins, a single five-dollar bill, a single ten-dollar bill, a single twenty-dollar bill, a quarter and a nickel) and are instructed that, when a timer sounds after thirty-minutes, they are to give ten dollars to themselves, and five dollars to the examiner. In the Time-based task, the participants are shown a large clock with hands set at 8:00 and an envelope with numbered cards (“28”, “14”, “17” and “11”). Participants are instructed, at 8:15, to open the envelope and hand the card numbered “17” to the examiner.

Scores for both the Event-based MPMT and the Time-based MPMT consisted of the sum of three scoring components (with a maximum of 9): Intention to perform (0-3), Accuracy of response (0-3) and Need for Reminders (0-3). For the Event-based MPMT, intention to perform was scored as follows: 3 = spontaneously takes the envelope when the timer sounds; 2 = does not take the envelope, but verbally indicates that a response is required 1 = provides a nonspecific, non-verbal response; 0 = no response. Accuracy of response was scored as follows: 3 = task is completed correctly; 2 = the correct bills were selected but they were used incorrectly; 1 = only one of the bills were selected and given to either self or examiner; 0 = some other response not indicated above. Need of Reminders was scored based on response to a maximum of 3 controlled cues given if a participant did not respond within 60 seconds of the timer sounding: 3 = no reminder needed; 2 = needed only one reminder; 1 = needed two reminders; 0 = needed all three reminders. For the Time-based MPMT, Intention to perform was scored as follows: 3 = participant interrupts at exactly 8:15; 2 = participant interrupts two minutes before or after 8:15; 1 = participant interrupts the examiner within 4 minutes before or after 8:15; 0 = the participant does not interrupt the examiner before 8:19, or interrupts more than 4 minutes earlier than the target 8:15. Accuracy of response was scored based on the completion of three target responses (interrupts at the right time, gives examiner a card, the card is the “17” card): 3 = all tasks completed accurately; 2 = two of the three tasks were completed accurately; 1 = one of the three tasks were performed accurately; 0 = none of the target tasks were performed. Need of Reminders was scored based in response to a maximum of 3 controlled cues given if a participant did not respond by the time the clock read 8:19: 3 = no reminder needed; 2 = needed only one reminder; 1 = needed two reminders; 0 = needed all three reminders.

The Victoria version of the Stroop Test (Regard, 1981; available from the Psychology Clinic of the University of Victoria, BC, Canada) consists of three off-white cards, with six rows of four items. In the CLSA, a demonstration/practice trial was administered with participants being asked to say the color of each item on the first row of each card. This is consistent with the administration procedures of the French Victoria-version of the Stroop Test (Bayard, Erkes, Moroni & CPCN-LR, 2009; available from Univ-lille site, France) that is used in the CLSA.

The first card administered (Dots) is of colored dots in blue, green, red, and yellow ink. Each color appears once in each row and is arranged in a pseudorandom order within the full array. Participants are instructed to say the color of each of the 24 dots on the card as quickly as they can and the examiner times their performance in seconds. French-language modifications of the Victoria version of the Stroop Test (Bayard et al., 2009) require participants to use the French names of the colored dots (bleu, vert, rouge, jaune).

The second card administered (Words) is of four neutral “common” words not related to color (English: and, hard, when, over; French: quand, donc, mais, pour) in the colored ink used in the first card. Again, each color appears once in each row and is arranged in a pseudorandom order within the full array. The participant names the colors of the ink as quickly as they can while disregarding the verbal content of the words (English: blue, green, red, yellow; French: bleu, vert, rouge, jaune). The examiner records the time to completion in seconds.

The third card (Color) has color names in colored ink (blue, green, red, yellow), with ink colors incongruent to the color words, so that the print color never corresponds to the color word (e.g., the word “blue” printed in red ink). Again, each color appears once in each row and is arranged in a pseudorandom order within the full array. Participants are asked to name the colors of the ink and disregard the word content. Performance is measured in seconds. The French-language modification of the Victoria-version of the Stroop Test (Bayard et al., 2009; Tremblay et al., 2016) requires the French words for the colors presented.

For the COWAT, three subtasks were administered using the letters “F”, “A”, and “S”. For each letter, participants were asked to generate as many words as possible excluding proper nouns (e.g., George, Henry) and the same word with different endings (e.g., add, added, adding). The responses were audio-recorded to facilitate scoring. Each subtask was discontinued after 60 seconds. For scoring purposes, dictionary sources were selected and master dictionaries created for each language. Electronic algorithms were developed to identify homophones, sister words (same root word with different suffix), and subordinate or repeated words. Each admissible word for each letter was awarded one point and tallied.

The CRT, used in the CLSA (Gallacher, et al., 2013), is a simple choice reaction measure with 56 trials, organized into seven blocks comprising eight sets of stimulus presentations differing in *ms*. There is a “warm-up” of four trials at the beginning of the test. Variables available in the CLSA baseline dataset include the mean time of completing the overall test (the 4 warm-up trials + 56 trials = 60 trials) and percentage of correct answers for the whole test. Because the CRT is a touch screen computer-administered measure, scoring is automatically generated by the computer software.

## Selection of Studies for Comparison

For comparison purposes, we selected studies from the extant literature with sample characteristics as similar as possible to the CLSA Comprehensive Cohort. In selecting these studies, we used the following criteria: an age range of 45-85 years to be comparable to CLSA Comprehensive Cohort baseline; administered in French or English; means and standard deviations presented for a healthy sample of adults; and an overall sample size of about 100 participants. For each comparison, subsets of the CLSA Comprehensive Cohort data were selected to reflect the characteristics of the previous research. Age groupings were selected to correspond to the previous research; on occasion, some age groups were not available in the CLSA (which only included ages 45-85). Some of the cognitive measures (i.e., RAVLT, AF, COWAT, and Stroop Test) have been used extensively in previous research (e.g., Loonstra, Tarlow, & Sellers, 2001; Rodriguez-Aranda & Martinussen, 2006) and clinical practice, whereas others (i.e., MAT, MPMT, and CRT) have not, because they are relatively new instruments. Since only one comparable study was available for the MAT, and no comparable studies were available for the MPMT and the CRT, tables are not included here for those measures.

As the Stroop Test has been used extensively in research and clinical practice, there are many different versions and variations available. The CLSA administered the Victoria-version of the Stroop test (Regard, 1981), where participants are to name the color of ink, on all three cards, while disregarding the stimulus item (words). Previous studies with sample characteristics as similar as possible to the CLSA Comprehensive Cohort are described in Table A. All of these studies were conducted in English. Studies of the Victoria version in French (Bayard, Erkes, Moroni, & CPCN-LR, 2011; Tremblay et al., 2016) do not have comparable age groupings to the CLSA (45-80 year olds).

 Given the extensive literature on the COWAT, we selected only those previous studies that included the stimulus letters F, A and S and had an overall sample size greater than 100. A summary of the sample composition and study selection criteria are reported in Table B.

# References

Bayard, S. Erkes, J., Moroni, C. & les membres du Collége des Psychologues Cliniciens spécialisés en Neuropsychologie du Languedoc Roussillion (CPCN-LR). (2009). Test du Stroop Victoria - Adaptation francophone. Matérial, consignes, procédure de cotation et données normatives. France: gratuitement téléchargeable du Moroni á Univ-lille. <https://psitec.univ-lille3.fr/en/presentation/membres/membres-titulaires/moroni-c/>

Bayard, S., Erkes, J., & Moroni, C., & les membres du Collége des Psychologues Cliniciens spécialisés en Neuropsychologie du Languedoc Roussillion (CPCN-LR) (2011). Victoria Stroop Test: Normative data in a sample group of older people and the study of their clinical applications in the assessment of inhibition in Alzheimer’s disease. *Archives of Clinical Neuropsychology, 26*(7)*,* 653–661. doi:10.1093/arclin/acr053

Bolla, K.I., Gray, S., Resnick, S.M., Galante, R., & Kawas, C. (1998). Category and letter fluency in highly educated older adults. *The Clinical Neuropsychologist, 12*(3), 330-338. doi: 10.1076/clin.12.3.330.1986

Bullock, L., Brulot, M., & Strauss, E. (1996). Unpublished Stroop Test data. In O. Spreen & E. Strauss (1998). *A Compendium of neuropsychological tests: Administration, norms, and commentary.* (2nd ed.) (pp. 217). New York: Oxford University Press.

Cerhan, J. R., Folsom, A. R., Mortimer, J. A., Shahar, E., Knopman, D. S., McGovern, P. G., … Heiss, G. (1998). Correlates of cognitive function in middle-aged adults. Atherosclerosis Risk In Communities (ARIC) study investigators. *Gerontology, 44*(2), 95-105. doi: 10.1159/000021991

Gallacher, J., Collins, R., Elliott, P., Palmer, S., Burton, P., Mitchell, C., . . . Lyons, R. (2013). A platform for the remote conduct of gene-environment interaction studies. *PLoS One*, *8*(1), e54331. doi: 10.1371/journal.pone.0054331

Loonstra, A. S., Tarlow, A. R., & Sellers, A. H. (2001). COWAT metanorms across age, education and gender. *Applied Neuropsychology, 8*(3), 161-166. doi: 10.1207/S15324826AN0803\_5

Regard, M. (1981). Stroop Test: Victoria Version. Manual of Instructions and Norms. British Columbia, Canada: University of Victoria, Department of Psychology. <https://www.uvic.ca/socialsciences/psychology/research/clinic/stroop.php>

Rodríguez-Aranda, C., & Martinussen, M. (2006). Age-related differences in performance of phonemic verbal fluency measured by Controlled Oral Words Association Task (COWAT): a meta-analytic study. *Developmental Neuropsychology, 30*(2), 697-717. doi: 10.1207/s15326942dn3002\_3

Statistics Canada (2009). Canadian Community Health Survey (CCHS) – Healthy Aging. User Guide. Retrived from Statistics Canada website http://www23.statcan.gc.ca/imdb-bmdi/document/5146\_D1\_T1\_V1-eng.pdf

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and Animal Naming. *Archives of Clinical Neuropsychology,* *14*(2), 167-177. doi: 10.1016/S0887-6177(97)00095-4

[Tremblay, M. P](https://www.ncbi.nlm.nih.gov/pubmed/?term=Tremblay%20MP%5BAuthor%5D&cauthor=true&cauthor_uid=27246959)., [Potvin, O](https://www.ncbi.nlm.nih.gov/pubmed/?term=Potvin%20O%5BAuthor%5D&cauthor=true&cauthor_uid=27246959)., [Belleville, S](https://www.ncbi.nlm.nih.gov/pubmed/?term=Belleville%20S%5BAuthor%5D&cauthor=true&cauthor_uid=27246959)., [Bier, N](https://www.ncbi.nlm.nih.gov/pubmed/?term=Bier%20N%5BAuthor%5D&cauthor=true&cauthor_uid=27246959)., [Gagnon, L](https://www.ncbi.nlm.nih.gov/pubmed/?term=Gagnon%20L%5BAuthor%5D&cauthor=true&cauthor_uid=27246959)., [Blanchet, S](https://www.ncbi.nlm.nih.gov/pubmed/?term=Blanchet%20S%5BAuthor%5D&cauthor=true&cauthor_uid=27246959)., … [Hudon, C](https://www.ncbi.nlm.nih.gov/pubmed/?term=Hudon%20C%5BAuthor%5D&cauthor=true&cauthor_uid=27246959). (2016). The Victoria Stroop Test: Normative data in Québec-French adults and elderly. *Archives of Clinical Neuropsychology, 31*, 926–933. doi:10.1093/arclin/acw029

Troyer, A. K., Leach, L., & Strauss, E. (2006). Aging and response inhibition: Normative data for the Victoria Stroop Test. *Aging, Neuropsychology, and Cognition, 13*(1)*,* 20-35. doi: 10.1080/138255890968187

Tuokko, H., Griffith, L. E., Simard, M., & Taler, V. (2017). Cognitive measures in the Canadian Longitudinal Study on Aging. *The Clinical Neuropsychologist, 31*(1), 233–250. doi: 10.1080/13854046.2016.1254279

Table A. Normative Studies for the Stroop Test (Victoria-version)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Authors** | ***N*** | **Age in years comparable to CLSA** | **Education / IQ** | **Sex** | **Sample composition** | **Normative data** | **Reliability/ Validity** |
| Regard (1981) | 86 | 3 groupings: 50-59 60-69 70-79 | Described as “well-educated” *M* = 13.2 yrs*SD* = 3.1 | \_ | Language: EnglishVictoria, BCCanadaReferred to as “healthy” elderly | Means and *SD* across age groupings | Test-retest reliability: one-month intervalReliability estimates Dots .90, Words .83, Color .91;significant practice effects (*p* <. 001) |
|  |  |  |  |  |  |  |  |
| Bullock, Brulot & Strauss (1996)  | 188 | 3 groupings: 50-5960-6970-79 | Described as having “above-average educational achievement”*M* = 14.28 yrs*SD* = 2.29 | \_ | Language: EnglishVictoria, BCCanadaReferred to as“healthy adults” | Means and *SD* across age groupings | \_ |
|  |  |  |  |  |  |  |  |
| Troyer, Leach & Strauss (2006) | 272 | 4 groupings (overlapping)50-64 60-69 65-7570-7975-84 | Education*M* = 12.9 yrs*SD* = 2.7 | 55% - 78% women across age groups | Language: EnglishVictoria, BC, West Canada Toronto, Ont., East CanadaScreened for neurological disorders. Healthy, community-dwelling adults. | Means, *SD* and percentile ranks (*PR*) across age groupings | **\_** |

Table B. Normative Studies for the COWAT (F-A-S)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  **Authors** | ***N*** | **Age in years comparable to CLSA** | **Education / IQ** | **Sex** | **Sample composition** | **Normative data** |
| Bolla et al. (1998) | 478 | 2 age groupings55-69 70-79 | Education:4-20 years*Median* = 18;Less than high school = 9%; 1-4 years of college = 35%; graduate degrees = 56% | 23.2% women | Language: EnglishUnited StatesFrom the Baltimore Longitudinal Study on Aging (BLSA). Healthy, community dwelling adults; 95-99% “Caucasian” | Means, *SD* Across sex, age, education groupings. Regression estimates for 2 age groups, sex & education |
|  |  |  |  |  |  |  |
| Cerhan et al. (1998) | 13,913 | 4 age groupings45-4950-5455-5960-6465-69 | Education: 5% had not achieved a ninth grade education; 47% had at least some college.  | 55% women | Language: EnglishFrom the Atherosclerosis Risk in Communities (ARIC) Study. From 4 US communities:North Carolina, Mississippi, Minnesota, Maryland.Screened for neurological disorders. Healthy, community dwelling adults; 25% “African American”, 75% “White” | Means, *SD* and percentile rank scores (10th, 50th, 90th) |
|  |  |  |  |  |  |  |
| Tombaugh, Kozak, & Rees (1999) | 895 | 3 age groupings50-59 60-69 70-79 | Education:4-21 years*M* = 12.9;*SD*=2.6 | 55.6% women | Language: EnglishCanadaHealthy, community dwelling adults, scoring > 23 on MMSE, < 12 on Geriatric Depression Scale.  | Means, *SD* for age & education groups  |