Project Creation Plan

CLSA Dementia

# Intent for Use of CLSA Data in BRAIN Grant

**Rationale**

Dementia is a term used to describe a set of progressive symptoms affecting memory, thinking and social abilities that interfere with daily life and normal function. The goal of the BRAIN grant is to define dementia more clearly by looking at changes in cognition, behaviour and function of large cohorts of individuals over time. The research is based on the rationale that, the better we can define patterns of cognitive, functional and behavioural decline in these large cohorts, the better we will be able to understand the features and time course of different dementia phenotypes. The better we understand these dementia phenotypes, the better we will be able to predict and, hopefully, modify the course of dementia and the better we will be able to provide care for those with dementia. The research uses data obtained from the Canadian Longitudinal Study of Aging (CLSA) to address two research questions.

**Research Questions**

1. Can we identify dementia phenotypes using CLSA data on changes on cognition, behaviour and function? To answer this question, we use the extensive data on cognition, behaviour and function provided by CLSA participants at baseline and subsequent follow ups. We will then analyze these data to look for different patterns of decline and to define a set of dementia phenotypes based on the trajectories of rates of decline over time and nature of decline in terms of specific cognitive, functional and behavioral characteristics.
2. Can we identify predictors of dementia phenotypes in our CLSA cohort using data on demographics, behavioural risk factors, prior medical history and a range of clinical and genetic factors? To answer this question, we will use the extensive demographic, lifestyle and biological data collected in the CLSA to examine important factors that might be predictive of dementia phenotypes.

**Methods**

The study will use a range of methods. The dementia phenotype identification analysis (research question 1) will use a spectrum of unsupervised learning tools including regression, ensemble and deep learning techniques. The prediction of dementia phenotypes (research question 2) will use a spectrum of supervised learning tools that span regression, ensemble and deep learning techniques.

# Project Details

|  |  |
| --- | --- |
| **Project Title:** |  |
| **Project Purpose and Objectives (max 250 words):**  Please provide a short description of the project’s purpose and objectives/research questions. Indicate how they are consistent with the Statement of Intent for the use of CLSA in the CIHR-funded BRAIN grant.  Project approval is contingent on demonstration that project’s purpose and objectives/research questions are consistent with intent for the use of CLSA data as outlined in the approved CLSA data access application and in the relevant UofT REB (RIS # 37430) for the BRAIN grant. |  |

# Principal Investigator (PI):

Must be a faculty member at Canadian university and a Co-Investigator listed on the CLSA Access Agreement.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Project PI** | **Name** | **Organization** | **Email** | **Role** |
|  | Geoffrey Anderson | University of Toronto | geoff.anderson@utoronto.ca | Principal Investigator |
|  | Andrew P Costa | McMaster | acosta@mcmaster.ca | Co-Investigator |
|  | Colleen Maxwell | University of Waterloo | colleen.maxwell@uwaterloo.ca | Co-Investigator |
|  | David B Hogan | University of Calgary | dhogan@ucalgary.ca | Co-Investigator |
|  | David Naylor | University of Toronto | david.naylor@utoronto.ca | Co-investigator |
|  | Ken Rockwood | Dalhousie | Kenneth.Rockwood@Dal.Ca | Co-Investigator |
|  | Nathan Herrmann | University of Toronto | Nathan.Herrmann@sunnybrook.ca | Co-Investigator |
|  | Peter St. George-Hyslop | University of Toronto | p.hyslop@utoronto.ca | Co-investigator |

# Project Team Member(s)

Insert additional rows as needed.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Position/Title and Organization** | **Institution Email** | **Project Role**  Co-Investigator / Collaborator / Analyst / Programmer | **Data Access in secure RDEN environment**  Yes / No |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# PI Signature

|  |  |
| --- | --- |
| **PI Signature:** |  |
| **Name:** |  |
| **Date:** |  |

# CLSA Data Files Required for Project

|  |  |
| --- | --- |
|  |  |
| Comprehensive sample survey | COM (baseline), COF1 (first follow up) |
| Tracking sample survey | TRM (baseline), TRF1 (first follow up) |
| Status | CSTATUS, TSTATUS |
| Genomics and epigenomics |  |

# Modelling Approach for Project

|  |  |
| --- | --- |
| **Modelling Approach** | Supervised learning (e.g. predictive modelling, time to event analysis)  Unsupervised learning (e.g. disease trajectories, factor analysis) |
| **Required Software Packages** | Packages available in existing RStudio (see current listing)  Requires new packages to be installed in RDEN (provide details in description of modelling) |
| **Description of Modelling Approach (200 words)**  Provide an outline of modelling strategy and justification including reference to specific software packages |  |

# Project Variables Selection

You will need to provide a list of variables that will be used in the project:

1. All existing variables must be selected from CLSA website <https://datapreview.clsa-elcv.ca/mica/repository#/search?type=variable&display=list> and downloaded to an excel file and attached to this PCP
2. Any variables derived from the CLSA variables must be documented before use

|  |  |
| --- | --- |
| **CLSA Data Source** | Excel spread sheet with names of CLSA variables from that CLSA data source |
| **COM** |  |
| **COF1** |  |
| **CSTATUS** |  |
| **TRM** |  |
| **TRF1** |  |
| **TSTATUS** |  |

# Specific Analytical Methods Used in Project

|  |  |
| --- | --- |
| **Summary of specific methods including description of R software required** |  |

# Summary of Format of Expected Project Results

|  |  |
| --- | --- |
| **Summary of format of expected results to be exported from RDEN (max 200 words)** |  |
| **Detailed format for output tables that will be exported** |  |

# Project Approval (For office use only)

|  |  |
| --- | --- |
| **Project number:** |  |
| **Confirmation that PCP is consistent with Project Objectives:** |  |
| **Additional comments:** |  |

# Project Quality Assurance Activities (For office use only)

|  |  |
| --- | --- |
| **Review of proposed code by Director of Analytics:** |  |
| **Designated RDEN Research and Analysis Staff accountable for Project Documentation:**  Approval of programming details for projects. |  |
| **Review of data access:**  The final analytic dataset for each cohort includes all the data required to create the baseline tables and run all the models. |  |
| **Additional comments:** |  |

# Project Amendments and Reconciliation

|  |  |  |  |
| --- | --- | --- | --- |
| **Project Approval Amendment History including changes to objectives and staffing (add appendices as needed):** | *Privacy approval date* | *Person who submitted amendment* |  |
| ***Date*** | ***Name*** | ***Amendment*** |
| yyyy-mon-dd |  |  |
| **Data Access Amendment History including all new variables (add appendices as needed):** | *Date DCP amended* | *Person who made the PCP amendment* |  |
| ***Date*** | ***Name*** | ***Amendment*** |
| yyyy-mon-dd |  |  |
| **Date Programs/DCP reconciled** | *The person(s) creating the dataset and/or analyzing the data are responsible for ensuring that the final DCP reflects the final program(s) when the project is completed* | | |
| yyyy-mon-dd | | |