Project 2 Report

Analyst: Max McGrath

Investigator: Nichole Carlson

Report generated: 11/01/2021

Introduction

This report will assist in preparing a grant application by developing an analysis plan, sample size estimation, and a budget with justification for a study by Dr. Brianne Bettcher seeking to understand the relationship between inflammation, Alzheimer's disease (AD), and cognitive decline. This report will contribute to Aims 1 and 2 of the grant.

Aim 1 Analysis

For the first aim, we will use linear regression models to analyze the associations, with changes in memory and cortical thickness in one year as the outcomes and baseline cytokines and chemokines as predictors for the first part and changes in cytokines and chemokines in one year as predictors for the second part. In both models we will control for age and sex as confounders.

We have also been given a large number of potentially confounding factors such as inflammation, cardiovascular risk, health history, and APOE genotype. Given the number of potential confounders coupled with the foreseeable collinearities between confounders, we will use elastic net regressions with potential confounders as predictors and each aforementioned outcome as responses to identify a final set of confounders which will be included in the above linear regression models.

We will then use coefficient estimates and t-tests to evaluate each variable as a predictor, using a Bonferroni corrected p-value to account for the number of t-tests performed.

Aim 2 Analysis

For both parts of the second aim, we will use linear models to analyze the relevant predictive power and associations. For the first part, we will use amyloid deposition and cortical thickness as the outcomes with inflammatory markers as predictors. For the second part, we will use changes in memory and cortical thickness in one year as outcomes and amyloid

deposition, inflammatory markers, and cytokines and chemokines as variables of interest. We will also include an interaction term between amyloid deposition and inflammatory markers to test the hypothesis that inflammatory markers differentially impact the outcomes depending on amyloid deposition, where amyloid deposition is treated as a dichotomous variable (high/low levels).

As with Aim 1, we will control for age and sex and use elastic net regression to expand the set of confounders to include in the above linear models. We will then use coefficient estimates and t-tests to evaluate each variable as a predictor, using a Bonferroni corrected p-value to account for the number of t-tests performed.

Sample Size Estimation

Using change in recall on a episodic memory list-learning test (CVLT) as the primary outcome of interest, to achieve 80% power with a Bonferroni corrected p-value of 0.0083 (based on a desired type 1 error rate of 0.05 and tests for 6 cytokines and chemokines of interest), assuming a difference in correlation of 0.4, a sample size of 276 (138 with high amyloid deposition and 138 with low deposition) is estimated to be necessary. Then, accounting for 10% one-year dropout, a total sample size of 304 is estimated to be sufficient to achieve 80% power.

This estimate was powered for the second part of Aim 2, where the linear model described above was simplified to a two independent correlation z-test. This sample size estimate leads to an estimated power of just under 1 for Aim 1 based on a simplification of the linear models to a correlation z-test. As such, a sample size of 304 is estimated to be sufficient for both aims with allowance for 10% attrition. This is notably higher than Dr. Bettcher's initial estimate of 194 subjects, so more enrollment is necessary. Note that this is an anti-conservative estimate, as we plan to add confounders to each model which may reduce power.

Budget

Tables 1 and 2 provide a budget estimate and budget justification for the statisticians and data managers involved in the 5-year project. The budget will fund one PhD statistician and one MS statistician who will be involved with study design, implementation, data collection, analysis, and regular meeting attendance throughout the duration of the grant. The PhD statistician will primarily be responsible for assembling the rest of the statistical personnel and supervising and guiding their work, while the MS statistician will be primarily responsible for analysis and manuscripts. The budget will also fund a data manager who will primarily be tasked with the creation of a REDCap data set, data management, and data documentation in the first two years of the grant, with some data set maintenance in the latter years. Lastly, the budget will fund a student researcher who will assist with data management and analysis.

Estimated personnel salaries are provided in Table 1, and yearly personnel effort estimates along with associated costs are provided in Tables 1 and 2. The total statistical budget for the grant is $92,670 for the first year, $70,770 for the second, $52,350 for the third, $88,700 for the fourth, and $92,860 for the fifth year. The total budget of the 5-year duration of the grant is $397,350.

Table 1: Budget Justification

Position	Base Salary	Salary w/ Benefits	Y1 Effort	Y1 Cost	Y2 Effort	Y2 Cost
PhD Statistician	$115,000	$147,200	10%	$14,720	5%	$7,360
MS Statistician	$80,000	$102,400	25%	$25,600	15%	$15,360
Data Manager	$65,000	$83,200	50%	$41,600	50%	$41,600
Research Assistant	$43,000	$43,000	25%	$10,750	15%	$6,450
Totals				$92,670		$70,770

Table 2: Budget Justification cont'd

Position	Y3 Effort	Y3 Cost	Y4 Effort	Y4 Cost	Y5 Effort	Y5 Cost	Totals
PhD Statistician	15%	$22,080	15%	$22,080	15%	$22,080	$88,320
MS Statistician	15%	$15,360	40%	$40,960	40%	$40,960	$138,240
Data Manager	5%	$4,160	5%	$4,160	10%	$8,320	$99,840
Research Assistant	25%	$10,750	50%	$21,500	50%	$21,500	$70,950
Totals		$52,350		$88,700		$92,860	$397,350