



## ORIGINAL RESEARCH

# Analyzing the Covariance Structure of Plasma Signaling Proteins in Relation to the Diagnosis of Dementia

Calvin Guan<sup>1\*</sup>, Rhoda Au<sup>2</sup>, Alvin Ang<sup>2</sup> and Ashis Gangopadhyay<sup>1</sup>

<sup>1</sup>Department of Mathematics & Statistics, Boston University, 665 Commonwealth Ave, Boston, MA, 02215, USA

<sup>2</sup>Department of Anatomy & Neurobiology, Boston University School of Medicine, 72 East Concord St (L 1004), Boston, MA, 02118, USA

\*Corresponding author: Calvin Guan, Department of Mathematics & Statistics, Boston University, 665 Commonwealth Ave, Boston, MA, 02215, USA



## Abstract

Numerous studies have shown that individuals with dementia have exhibited activation of inflammatory pathways in their brains. Typically, these studies use traditional and well-established regression methods for data analysis. In this paper, a new approach is introduced that utilizes the analysis of the covariance structure using methods related to the principal component analysis (PCA) theory. Eleven biomarkers related to neuroinflammation were used to determine the association with the onset of dementia. Various demographic covariates were adjusted to account for possible confounding effects of the covariance structure. Three hypothesis testing methods were considered to discern differences between partial covariance matrices for comparing power and Type I errors through simulation studies. Application of hypothesis testing methods using data from Framingham Heart Study (FHS) found significant differences in covariance matrices between the non-dementia and dementia groups.

## Keywords

Inflammatory biomarkers, Dementia, Comparison of covariance matrices, Principal component analysis (PCA)

## Introduction

There is general research consensus that inflammation in pathologically vulnerable regions of the brain is associated with dementia, and specifically with the subtype of Alzheimer's disease [1]. For example, [2] surmises that microglial activation neuroinflammation plays a significant role in the development of dementia.

Therefore, it has been suggested that treating inflammatory conditions could inhibit the development of dementia. Given that inflammatory molecules have been found in autopsy [3], it raises the question at what antemortem time point is neuroinflammation a risk factor for the onset of dementia later in life. Whether inflammation is a causal factor of dementia or simply a byproduct is still widely debated among the scientific and clinical communities [4]. Epidemiological studies and treatment trials using anti-inflammatory drugs have been disappointing, although concerns have been raised about the methodological fidelity of the trials [4]. As there is yet a proven and effective cure for dementia, it is crucial to understand the underlying mechanisms further to develop such treatments.

Protein biomarkers are used to quantify the level of inflammation in the brain. For this study, eleven measures that are recognized as inflammatory biomarkers are assessed as risk factors for dementia. The most common inflammatory biomarker is C-reactive protein (CRP), which is found in higher concentrations after an aneurysmal subarachnoid hemorrhage [5]. Other biomarkers include interleukin-6 (IL6) and osteoprotegerin (OPG). IL6 is a pleiotropic cytokine that contributes to host defense during infection and tissue injury [6], while OPG is a soluble secreted protein and decoy receptor that is associated with inflammation after ischemic stroke [7]. These protein biomarkers

are gathered by venous blood samples drawn from participants and then the proteins of interest are isolated, frozen, and measured [8]. The data presented in this study is from the Framingham Heart Study (FHS). The FHS has been collecting longitudinal data from three generations of participants for over 7 decades that includes demographics, inflammatory biomarkers, incident dementia and clinical comorbidities.

Traditional hypothesis testing methods are adequate for assessing the effects of biomarkers' individual effects on dementia. However, multiple-hypothesis testing procedures can produce false positives at an unacceptably high rate [9]. Smart manipulation of the rejection criteria can result in a good balance of power and false positives, but this study circumvents the need by offering an alternative approach. Principal component analysis (PCA) and partial covariance analysis are deployed to condense the problem into a single hypothesis testing procedure. PCA approximates data by the product of the "object patterns" and the "variable patterns" [10], while partial covariance adjusts the scale measures for a set of variables [11]. These methods can summarize a set of variables and identify the fundamental structure of the set, which can be invaluable for understanding the underlying factors that affect the combined impact of these variables. A useful application of the group structure is to facilitate the comparison of a set of variables for different population groups.

For this study, three distinct methods for the hypothesis testing of equivalence of covariance matrices are applied to determine potential inflammation-related risk factors for dementia: Principal Component Group Comparison [12], Forkman's Test [13], and the Tracy-Widom Statistic [14]. Primary objectives are to compare the test statistics and p-values via various visualization tools, as well as to assess the Type I error rate and power of the hypothesis testing procedure via simulations. Given the novelty of the approaches utilized in this paper, it is imperative to perform simulation analysis for a comprehensive understanding of these inferential tools. For example, the simulation analysis allows visualizing and evaluating each method's unique advantages and disadvantages within controlled settings. The PCA comparison method is intuitive, and the non-parametric framework requires fewer assumptions about the underlying data generating process but is untested in the context of hypothesis testing. The Tracy-Widom distribution is parametric; therefore, the p-values can be derived from the distribution by theoretical analysis. However, there are strict assumptions for the Tracy-Widom distribution, so the results may be questionable if the assumptions are not met. The Forkman's test method allows the computation of the p-values without the restrictive distributional assumptions. The results are further summarized into test statistics to test the differences in partial covariance matrices.

In summary, the paper discusses three novel approaches to understanding the differences between covariance structures of multivariable data and provides guidelines for using these methods using extensive simulation studies. The potential clinical contribution of these analyses is their application to confirm the role of inflammatory biomarkers as a risk factor for dementia.

## Methodology and Materials

### Design

In this section, we briefly describe the details of the study design.

**Participants:** The Framingham Heart Study (FHS) is a community-based multi-generational prospective cohort study that began in 1948 to identify risk factors for cardiovascular disease and subsequently expanded to incorporate study of many common chronic diseases including dementia. For the current study, participants are members of the offspring cohort, which includes the biological children of the original FHS cohort and the spouses of the children [15]. These participants had regular health examinations on average every four years. Data used for this analysis were collected at the seventh and eighth health examinations performed during 1998-2001 and 2005-2008, respectively (n = 2684).

**Inflammatory biomarkers:** Inflammatory biomarkers were measured from venous blood samples, as previously described [8]. Table 1 lists the 12 different biomarkers of inflammation [16,17].

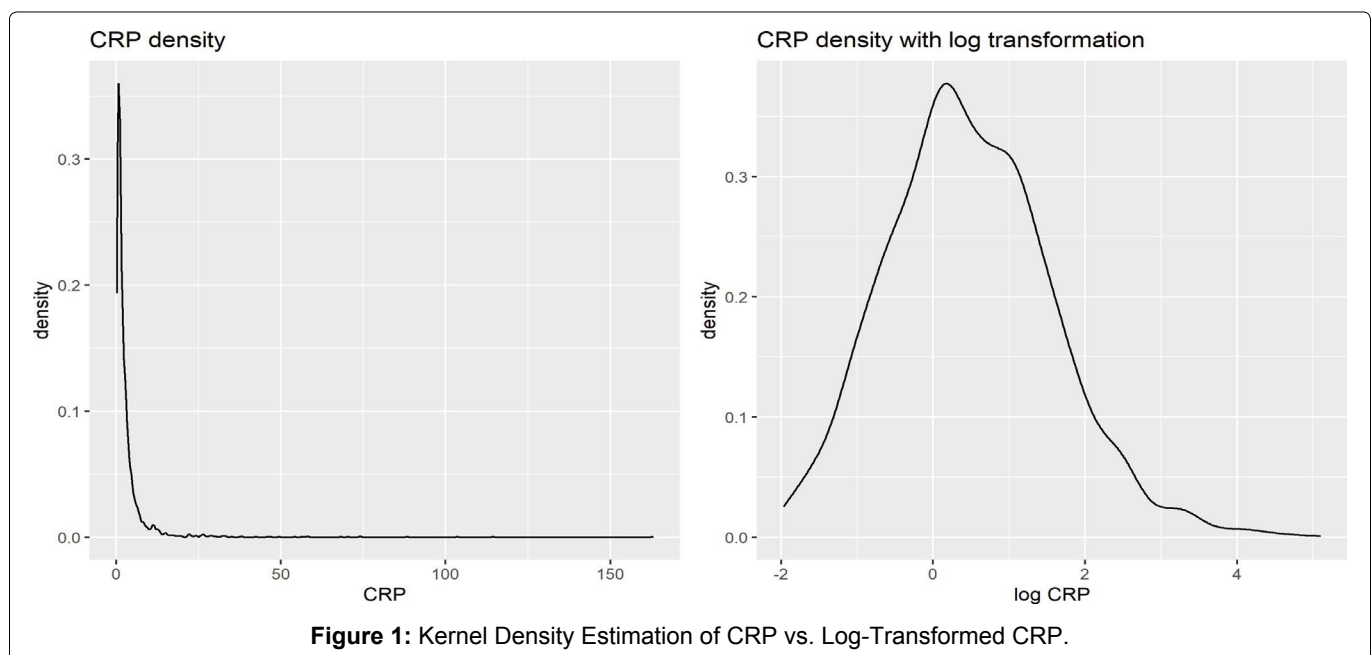
The biomarkers' blood concentration levels are typically right-skewed, as there is a lower limit on the levels but no upper limit (Figure 1). Therefore, a logarithmic transformation is applied to biomarkers to reduce skewness and improve the performance of PCA decomposition.

**Dementia review:** The dementia diagnosis of each participant in FHS is evaluated and verified through an adjudication panel that includes at least one neurologist and one neuropsychologist, who determined the diagnosis based on information from clinical examinations, medical records, and, when available, neuropsychological and neurological assessments and family interviews [18,19]. Ongoing surveillance for incident dementia that has been underway since 1976 identifies participants for diagnostic consideration. Therefore, those who do not undergo a dementia review assessment are presumed not to have dementia. Participants identified for possible mild cognitive impairment (MCI) were also excluded from the analysis (n = 158), to create a distinct separation between the dementia and non-dementia groups. Dementia is flagged whether the onset is before or after the biomarkers sampling.

**Adjusting for covariates:** Covariates that could

**Table 1:** Biomarker sampling method.

Biomarker	Abbreviation	Sample Type	Units
C-reactive protein	CRP	Serum	mg/L
Soluble intercellular adhesion molecule-1	sICAM-1	Serum	ng/mL
Interleukin-6	IL6	Serum	pg/mL
Interleukin-18	Il18	Serum	pg/mL
8-EPI-Isoprostane	ISO	Urine	pg/mL
Lipoprotein-associated phospholipase-A	Lp-PLA2-mass	Plasma	ng/mL
Lipoprotein-associated phospholipase-A	Lp-PLA2-activity	Plasma	nmol/min/mL
Monocyte chemoattractant protein-1	MCP-1	Serum	pg/mL
Osteoprotegerin	OPG	Serum	pmol/L
P-selectin		Plasma	ng/mL
Tumor necrosis receptor II	TNFRII	Plasma	pg/mL



significantly affect protein biomarker concentration are identified using rudimentary linear models and their mutual correlation. These covariates are age, sex, BMI, current smoking status, total cholesterol, ventricular rate, serum creatinine concentration and indicator of treatment for lipid disorders. Furthermore, the models are adjusted for the time difference between biomarker sampling and dementia diagnosis. Participants with missing covariates are excluded ( $n = 1021$ ), these exclusions were already applied in Section 4.1.1.

**Data imputation:** Missing data, be it due to non-attendance at the time of data collection, inconclusive results, unreadable notation, or something else, pose a significant challenge in a clinical setting. However, in traditional statistical analysis, a sample would be discarded if its variables had missing values. Instead, the PCA imputation method is applied on the logarithm of biomarker concentrations to preserve the sample size with minimal effect on the data quality. This method uses a regularized iterative PCA algorithm that fits an expectation-maximization (EM) algorithm

to remove noise during the imputation step [20]. Note that imputation works well when the proportion of missing values is not too high; as such, the inflammatory biomarker lysophosphatidic acid (lpa) was not considered in this analysis due to too many missing values ( $n = 1270$ ). Participants with missing values in the other biomarkers had their missing values imputed ( $n = 433$ ).

### Statistical analyses

This section describes the various statistical techniques implemented to analyze the data along with a detailed discussion of the various characteristics of the methodologies.

**Participant characteristics:** After screening out exclusions and implementing imputation for missing observation, the data included 2684 participants, of which 2536 were in the non-dementia group and 148 in the dementia group. The study sample was characterized using means and percentages along with rudimentary univariate analysis to examine potential differences

**Table 2:** Quantitative covariates descriptions.

Demographics	$\hat{\mu}_n$	$\hat{\sigma}_n$	$\hat{\mu}_d$	$\hat{\sigma}_d$	t	p-value
Age	65.285	8.516	77.453	6.193	22.681	< 0.001
BMI (kg/m <sup>2</sup> )	28.406	5.443	26.768	4.597	-4.167	< 0.001
Creatinine (mg/dL)	0.908	0.299	0.986	0.332	2.817	0.005
Total Cholesterol (mg/dL)	186.432	37.576	179.115	35.855	-2.407	0.017
Ventricular Rate (beats/min.)	62.336	10.415	63.101	11.380	0.799	0.426

$\hat{\mu}$  is the sample mean,  $\hat{\sigma}$  is the sample standard deviation, n denotes the non-dementia group, and d the dementia group

**Table 3:** Covariates variables descriptions (sex).

Sex	N	N <sub>n</sub>	% <sub>n</sub>	N <sub>d</sub>	% <sub>d</sub>
Male	1228	1164	0.459	64	0.432
Female	1456	1372	0.541	84	0.568

$\chi^2$  test of independence:  $\chi^2 = 0.298$ ,  $p$ -value = 0.585

N is the sample size of the respective group, while %<sub>n</sub> and %<sub>d</sub> are the proportion of the non-dementia and dementia group with given characteristics.

**Table 4:** Covariates variables descriptions (current smoking status).

Current Smoking Status	N	N <sub>n</sub>	% <sub>n</sub>	N <sub>d</sub>	% <sub>d</sub>
No	2444	2302	0.908	142	0.959
Yes	240	234	0.092	6	0.041

$\chi^2$  test of independence:  $\chi^2 = 3.983$ ,  $p$ -value = 0.046

**Table 5:** Covariates variables descriptions (treated for lipids).

Lipids	N	N <sub>n</sub>	% <sub>n</sub>	N <sub>d</sub>	% <sub>d</sub>
No	1535	1476	0.582	59	0.399
Yes	1149	1060	0.418	89	0.601

$\chi^2$  test of independence:  $\chi^2 = 18.464$ ,  $p$ -value < 0.001

**Table 6:** Log predictor descriptions.

Biomarker (log concentration)	$\hat{\mu}_n$	$\hat{\sigma}_n$	$\hat{\mu}_d$	$\hat{\sigma}_d$	t	p-value
CRP	0.479	1.081	0.540	1.252	0.623	0.533
sICAM-1	5.657	0.305	5.714	0.315	2.271	0.024
IL6	0.622	0.728	0.993	0.773	6.031	< 0.001
IL18	5.446	0.386	5.569	0.405	3.825	< 0.001
ISO	6.685	0.841	6.710	0.713	0.424	0.672
Lp-PLA2-mass	5.277	0.261	5.309	0.246	1.587	0.114
Lp-PLA2-activity	4.904	0.257	4.951	0.233	2.487	0.014
MCP-1	5.901	0.307	6.014	0.341	4.181	< 0.001
OPG	1.532	0.302	1.816	0.326	10.952	< 0.001
P-selectin	3.666	0.318	3.687	0.327	0.771	0.442
TNFRII	7.796	0.396	7.954	0.813	2.492	0.014

between the two groups. Table 2 provides the mean and standard deviation of continuous variables. The two-sample t-test of the mean difference between the dementia group and the non-dementia group is included for each variable. Table 3, Table 4 and Table 5 show the counts and proportions for categorical variables, along

with the  $\chi^2$  test of independence. Age, BMI, creatinine concentration, total cholesterol, current smoking status, and lipid disorder treatment indicator show significant mean differences between the dementia and non-dementia groups whereas sex and ventricular rate were not significantly different. Interestingly, the dementia

rate is slightly higher for females than males, which goes against the findings of contemporary studies [21].

**Biomarker characteristics:** Table 6 provides the characterization of inflammatory biomarkers using the mean and standard deviation with two samples t-test of the mean difference as a comparison between the dementia and non-dementia groups. The results are mixed; four biomarkers have significant differences, while seven did not. However, this method evaluates biomarkers individually and is not able to detect potential inter-biomarker covariance differences. Furthermore, there may be demographic effects that cannot be captured using simple summary statistics. Therefore, it is necessary to use more sophisticated statistical methods to sufficiently understand the underlying covariance structure between the biomarkers.

**Partial covariance:** A multivariable data covariance matrix measures the joint structural association among a set of variables. Comparison of covariance matrices is a key tool for understanding differences between multivariable structures of multiple groups of data. This approach has been extensively used in the study of a wide range of problems involving multidimensional data, such as understanding the role of genetic constraints in the determination of evolutionary trajectories in adaptive radiation [22], the response of the genetic architecture to environmental heterogeneity [23], the evolution of phenotypic integration [24,25], multicharacter phenotypic plasticity [26] and sexual dimorphism [27,28] among others.

The partial covariance allows exploration of the structural association between the variables of interest after adjusting for exogenous variables. The logarithmic transformed biomarkers are adjusted by controlling for covariates. This is done by fitting a linear model for each biomarker against the covariates and estimating the covariance using the resulting residuals. Therefore, data analysis uses partial covariance matrices instead of traditional covariance matrices to account for demographic effects.

**Preliminary screening for the association of inflammatory biomarkers with dementia:** To explore the relationship of inflammatory biomarkers with dementia status, the partial covariance structure of biomarkers is compared between the two groups. Visual comparisons of the partial correlation matrices between the two groups were made using heat maps and scree plots. The heat map visualizes high values using reddish colors and low values using blueish colors, akin to a temperature map. A scree plot is a graphical tool that plots the eigen-values, i.e. amount of variance explained by each principal component, of the covariance matrix by decreasing orders of magnitude. The scree plot can be used to find the number of significant components to keep in a principal component decomposition or the number of dominant factors in factor analysis [29].

These descriptive tools were utilized to identify potential structural differences between the partial covariance matrices of the non-dementia and dementia groups.

**Hypothesis testing:** Hypothesis testing methods served as a reference analytical tool to assess the difference between the partial covariance matrices of the two groups.

$$H_0: P_0 = P_1, H_A: P_0 \neq P_1$$

where  $P_0$  and  $P_1$  are the population partial covariance matrices of the non-dementia and dementia groups, respectively. A hypothesis testing method calculates a test statistic and the corresponding p-value from the data that are used to make a decision. Three novel approaches to the hypothesis testing problem were considered, namely, Principal Component Group Comparisons, Forkman's test, and Tracy-Widom statistic.

**Principal components group comparison:** The first method to assess the differences between two covariance matrices was the Principal Components Group Comparison (PCGC) method based on the comparison of principal components of different groups [12]. Let  $L$  and  $M$  be the principal component loading matrices of the non-dementia and dementia groups, respectively, then the test statistic is the minimum angle between the space of the first component, given by

$$T = \cos^{-1} \left\{ (\lambda_1)^{-\frac{1}{2}} \right\}$$

where  $\lambda_1$  is the first eigenvalue of  $LM'ML'$

A small minimum angle implies that the two groups are similar with respect to the first principal component. As the minimum angle does not follow a known probability distribution, a bootstrap method is used to calculate the p-value. A bootstrap method creates  $B$  bootstrap data by randomly shuffling the group labels. Then, a bootstrap test statistic  $T_b$  is calculated for each bootstrap data ( $b = 1, \dots, B$ ). Finally, the p-value is calculated as the proportion of bootstrap samples in which the observed test statistic is smaller than the bootstrap test statistics.

**Forkman's test:** The Forkman's test takes inspiration from [13], where the authors propose a hypothesis testing method for the number of significant principal components in a standardized data matrix. First, the authors simulate simple parametric bootstrap samples and calculate the test statistic for each sample. The bootstrap is parametric because it assumes a normal distribution. It is a relatively simple approach because it is based on a standard distribution with no estimation of parameters. Then, the bootstrap test statistics are compared to the observed test statistic to calculate the bootstrap p-value. Estimation of parameters are not needed because the observed data are already standardized to zero mean and unit variance.

The methods in Forkman, et al. [13] are then modified to test the hypothesis of the difference between two covariance matrices. First the test statistic is calculated as such:

$$Z = (n_0 S_0 + n_1 S_1)^{-1} n_1 S_1$$

$$T = \lambda_1 / \sum_{k=1}^p \lambda_k$$

Where  $S_0$  and  $S_1$  are the sample partial covariance matrices of the standardized data from the dementia group and the non-dementia group, and  $n_0$  and  $n_1$  are the respective sample sizes. Furthermore,  $T$  is the test statistic and  $\lambda_k$  is the  $k^{\text{th}}$  eigenvalue of  $Z$ .

In summary, the test statistic is calculated by combining the covariance matrices of two groups in such a way that if the two matrices are equal, the largest eigenvalue of the combined matrices follows the greatest root statistic distribution [14].

The p-value is then calculated using the following boot-strap algorithm:

**Algorithm 1** Forkman's Test P-Value

- 1: **for**  $b \in 1, 2, \dots, B$  **do**
- 2: Generate  $Z_b = (n_0 S_{b0} + n_1 S_{b1})^{-1} n_1 S_{b1}$
- 3:  $z_b = (n_0 S_{b0} + n_1 S_{b1})^{-1} n_1 S_{b1}$  ( $\Sigma$  is the covariance matrix)
- 4:  $T_b = \lambda_{b1} / \sum_{k=1}^p \lambda_{bk}$  ( $\lambda_{bk}$  is the  $k^{\text{th}}$  eigenvalue of  $Z_b$ )
- 5: **end for**
- 6: P-value =  $\sum_{b=1}^B I(T_b > T) / B$  ( $I$  is the indicator function)

**Tracy-widom statistic:** The Tracy-Widom distribution was introduced in [30], as the probability distribution of the normalized eigenvalue of a random Hermitian matrix. Johnstone [31] established the use of the Tracy-Widom distribution of order 1 as the asymptotic distribution of the largest eigenvalue in the covariance matrix of independent Gaussian variables. A follow-up paper [14] explores the application of the Tracy-Widom distribution of order 1 to multivariable analysis, of relevance is the group comparisons of covariance matrices.

The hypothesis tested considered  $S_0$  and  $S_1$ , the sample covariance matrices for both groups. Let  $n_0$  be the number of participants in the non-dementia group, and  $n_1$  for the dementia group. And let  $\lambda_1$  be the largest eigenvalue of  $S = (n_0 S_0 + n_1 S_1)^{-1} n_1 S_1$ . Finally, define  $F_1$  as the Tracy-Widom distribution of order 1 and  $p$  the number of predictors, then under  $H_0$ :

$$T = \frac{\log it(\lambda_1) - \mu(p, n_0, n_1)}{\sigma(p, n_0, n_1)} \sim F_1$$

Where  $\mu$  is the centering term, and  $\sigma$  the scaling term, defined as:

$$\mu(p, n_0, n_1) = 2 \log \left( \tan \left( \frac{\gamma + \phi}{2} \right) \right)$$

$$\sigma(p, n_0, n_1)^3 = 16 \left( (n_0 + n_1 - 1) \sin^2(\gamma + \phi) \sin \gamma \sin \phi \right)^{-1}$$

Where

$$\sin^2 \left( \frac{\gamma}{2} \right) = \frac{\min(p, n_1) - 0.5}{n_0 + n_1 - 1}$$

$$\sin^2 \left( \frac{\phi}{2} \right) = \frac{\max(p, n_1) - 0.5}{n_0 + n_1 - 1}$$

Therefore,  $T$  is the test statistic, and the p-value is given by  $Pr(T > t | H_0) = 1 - F_1(t)$ .

In practice, we approximate the Tracy-Widom p-value using a Gamma distribution as outlined by [32], as the Tracy-Widom p-value is difficult to calculate directly.

**Simulation study:** Before addressing the core question related to differences in inflammatory biomarkers between the non-dementia group and the dementia group, it is important to first compare the performances of the three hypothesis testing methods discussed in the last section using simulation studies. Simulation is necessary to evaluate and compare the results of the three methods in a controlled setting, as established theoretical results for these methods are limited. Two groups of random variables are generated from a parametric distribution in the simulations. The first group had a fixed parameter, while the second group varied the parameter in a range where the fixed parameter of the first group is the center. The goal of these simulations is to visualize the change in statistic and p-values as the distributions of the two groups differ. The Type I error and power of these tests are assessed below. The following describes the data-generating process (DGP) for each simulation below.

**DGP1:** Data is generated from bivariate normal distributions with zero-mean and unit variance,  $\mathbf{X}_0, \mathbf{X}_1$ , where the correlation is defined as  $\rho_0 = 0, \rho_1 \in (-0.5, 0.5)$

**DGP2:** Following the work of [33], let  $\mathbf{x}_{ki} = (x_{kli}, \dots, x_{kpi})$  represent the predictor vector for the  $i^{\text{th}}$  observation in group  $k$ , in this case  $k=0,1$  and generate  $x_{kji} = w_{kji} + 2w_{k,j+1,i} + \theta_k w_{k,j+2,i}$ , where  $w_{kji} \sim N(0,1)$ . Parameters are defined as  $\vartheta_0 = 3, \vartheta_1 \in (1, 5)$ . This simulation creates a complex dependency based on a moving average process that is not directly obvious.

**DGP3:** Designed as a multivariable normal simulation with  $X_k \sim N(0, \Sigma_k) k=0,1$ . Where the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of  $\Sigma_k, \sigma_{ijk}$  is defined as  $\sigma_{ijk} = \phi_k^{|i-j|}$ , which follows the correlation structure of an autoregressive process of order 1 (AR(1)). Specifically, the parameters are defined as  $\phi_0 = 0.5, \phi_1 \in (-0.5, 0.5)$ . This simulation can be described as a multivariable variation of **DGP1**.

In each simulation, the test statistic and the p-value are graphically visualized. A fine grid of discrete values is set for  $\rho_1$ ,  $\vartheta_1$ , and  $\phi_1$  respectively. Each DGP is simulated 10 times, and the average of each set of test statistic or p-values against their respective value of  $\rho_1$ ,  $\vartheta_1$ , or  $\phi_1$  are then plotted.

Furthermore, Type I error and power for each DGP and each statistical method are calculated by estimating the Type I error by setting the parameter of interest equal to (**DGP1**:  $\rho_0 = \rho_1 = 0$ , **DGP2**:  $\vartheta_0 = \vartheta_1 = 3$ , **DGP3**:  $\phi_0 = \phi_1 = 0$ ). To assess power, a reference point for group 0 ( $\rho_0 = \phi_0 = 0$ ,  $\vartheta_0 = 3$ ) is set, while setting the parameter of interest of group 1 to values surrounding group 0's ( $\rho_1 = \phi_1 = (-.4, -.2, .2, .4)$ ,  $\vartheta_1 = (1, 2, 4, 5)$ ). The Type I error and power are evaluated in a range of sample sizes ( $n_0 = n_1 = 500, 1000, 1500$ ) and a fixed number of predictors ( $p = 10$ ) to evaluate changes in the estimator as sample sizes increase.

**Inflammatory biomarkers and effect on dementia:** Finally, the effect that inflammatory biomarkers collectively had on dementia risk is analyzed by performing the three aforementioned methods on the partial covariance structure of non-dementia and dementia group.

## Results

### Simulation analysis

This section summarizes the results of the simulation analysis as described in Section 4.2.9.

**Test statistics and P-values Results:** Figure 2 depicts the graphs of the three test statistics and the respective p-values (Figure 3) for the three different simulations. Note that the parameters for **DGP2** ( $\vartheta_1$ ) are scaled to match the other simulations in the overlying illustration. As a reminder, in the original simulation ( $\vartheta_1 \in (1, 5)$ ,  $\vartheta_0$

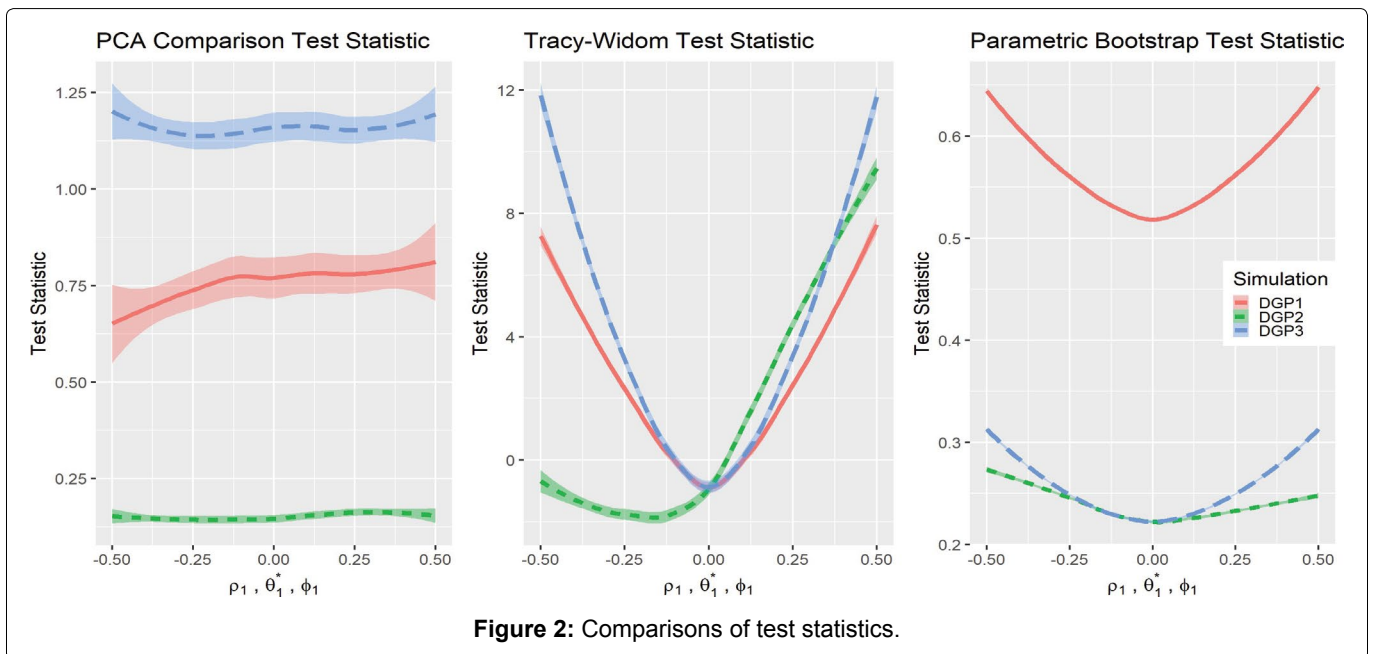


Figure 2: Comparisons of test statistics.

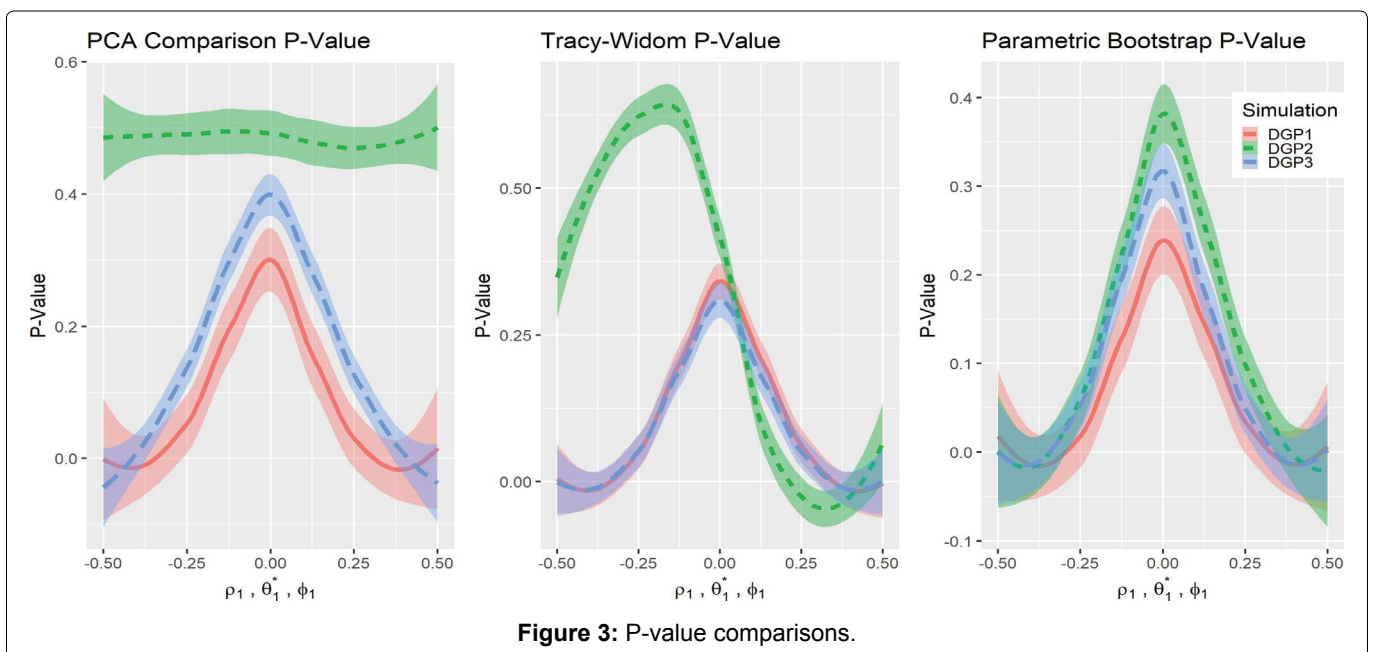


Figure 3: P-value comparisons.

= 3), this is rescaled to  $(\theta_1^* \in (-0.5, 0.5), \theta_0^* = 0)$  in graphical presentation. The test statistics in **DGP1** do not change noticeably with the parameters, although their corresponding p-values changed in a predictable way except for **DGP2**. The Tracy-Widom distribution performed well under **DGP1** and **DGP3** as it exhibits symmetry in the test statistic and the p-value graph. However, the **DGP2** simulation was not symmetric but was more significant on the right side. The Tracy-Widom p-value did not reach significance level ( $\alpha = 0.05$ ) even at  $\theta_1 = 1(\theta_1^* = -0.5)$ , the lowest value in **DGP2** simulation. In fact, the minimum of the Tracy-Widom test statistic (the maximum p-value) was reached when  $\vartheta_1 < 3$ , lower than the true null case. Lastly, the scaled Forkman's test statistic had good results in all three simulation methods.

The one-sided peculiarity of Tracy-Widom's distribution in **DGP2** is worth exploring. A compelling argument for this phenomenon is that the size or determinant of  $\Sigma_1$  has an one-sided effect on the Tracy-

Widom distribution, as  $S_1$  is used in the numerator to make the combined co-variance matrix. Indeed, a simulation that varies the matrix size clearly shows this effect, we call this simulation **DGP4**.

**DGP4:** Data was generated from independent bivariate normal distributions with zero-mean and variance  $\sigma^2, X_k \sim N_2(0, \sigma_k^2)$ . Where  $k = (0,1)$  and  $\sigma_0^2 = 3, \sigma_1^2 \in (2.5, 3.5)$ .

**DGP4** simulation result (Figure 4) clearly showed the one-sided relationship between the Tracy-Widom statistic and the difference between the variances  $\sigma_1^2 - \sigma_0^2$ . Therefore, the appropriate way to test the equality of the co-variance matrices would be by means of a two-sided p-value. However, this contradicts the results of **DGP1** and **DGP3**, which showed that the Tracy-Widom statistic is two-sided with respect to the difference in covariance between variables. This contradiction creates a difficult dilemma. On the one hand, there is the option of looking for differences in magnitude in variances, where a one-sided p-value

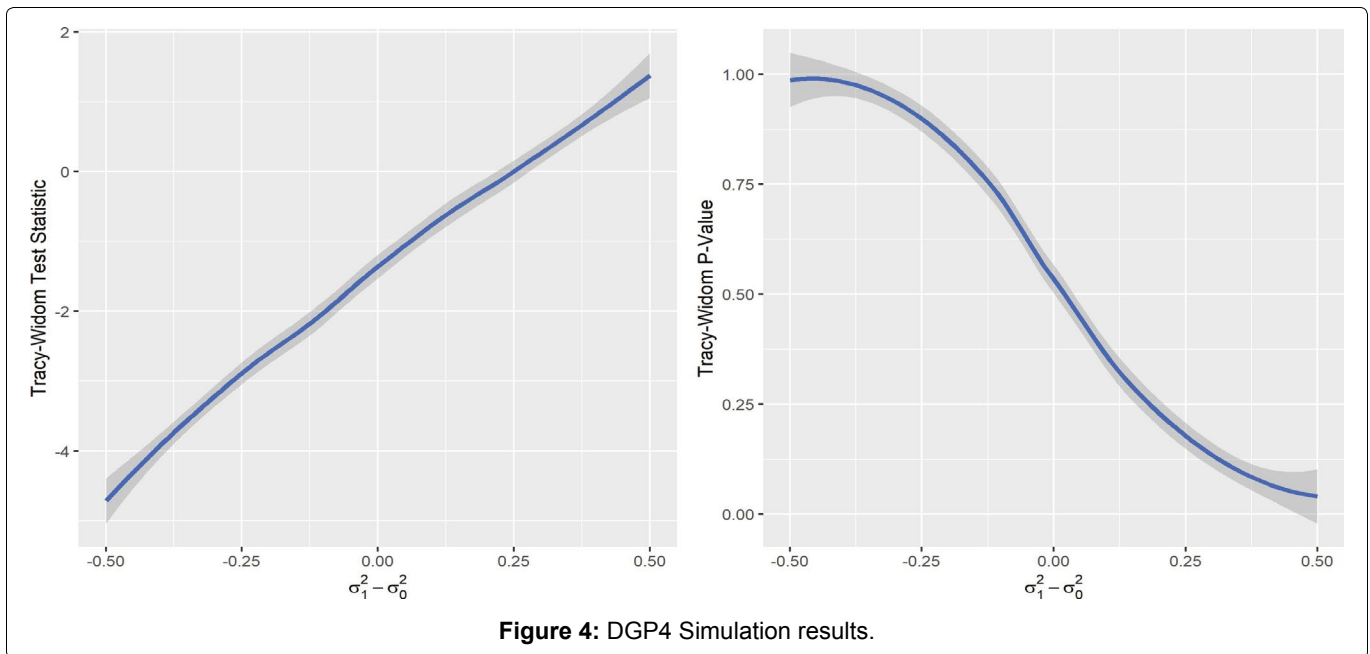


Figure 4: DGP4 Simulation results.

Table 7: Type I error and power under **DGP1**.

Method <sup>1</sup>	$n_0$	$n_1$	$\rho_1 = -0.4$	$\rho_1 = -0.2$	$\rho_1 = 0$	$\rho_1 = 0.2$	$\rho_1 = 0.4$
PCGC	500	500	0.8	0.61	0.052	0.63	0.816
PCGC	1000	1000	0.87	0.762	0.03	0.754	0.876
PCGC	1500	1500	0.892	0.804	0.044	0.79	0.89
Fork	500	500	1	0.79	0.034	0.83	1
Fork	1000	1000	1	0.986	0.048	0.98	1
Fork	1500	1500	1	1	0.046	0.998	1
TW	500	500	0.992	0.334	0.004	0.342	0.994
TW	1000	1000	1	0.816	0.002	0.824	1
TW	1500	1500	1	0.98	0.004	0.974	1

$\rho_1 = 0$  constitutes the null distribution

<sup>1</sup>PCGC: Principal Components Group Comparison; Fork: Forkman's Test; TW: Tracy-Widom



would be appropriate. On the other hand, if the objective is to determine differences in correlation between variables, the two-sided p-value would be the appropriate choice. For sake of simplicity, the data analysis focuses on the latter by standardizing the variables of interest. The standardization to zero mean and unit variance is consistent with similar analyses, such as in [34].

**Type I error and power results:** Next, the three methods using Type I error and power of the tests were assessed (Table 7, Table 8 and Table 9). Type I error is represented in the columns where the parameters for both groups are equal to ( $\rho_1 = 0$ ,  $\vartheta_1 = 3$ , and  $\phi_1 = 0$ ). Although the PCGC method captures the true Type I error ( $\alpha = 0.05$ ) quite well, the power is relatively low in all simulations; in fact, the power does not appear to change at all in **DGP2**. The Forkman's test method accurately estimates Type I error, while also having high power for all simulations. Meanwhile, the Tracy-Widom method was consistently conservative in estimating the Type I error in all simulations, with the **DGP1** simulation being the most conservative. As for power, the Tracy-Widom method had good power distribution in **DGP1** and **DGP3**, although is in general less powerful than the Forkman's test method. However, the Tracy-Widom method was extremely under-powered in **DGP2** when

$\vartheta_1 < \vartheta_0$ , this phenomenon is a consequence of Tracy-Widom method's non-symmetric interaction with **DGP2**.

In general, power increases as the sample size increases, whereas the Type I error does not have any discernible changes. Additionally, the power increases along with sample sizes given a fixed number of variables; this is true for all three methods.

In summary, the Forkman's test method had the most impressive simulation results; not only did it reasonably estimate Type I error, but it also demonstrated high power in all the simulations. The least impressive method was PCGC. While it had reasonable Type I power estimation, it was under-powered in comparison. In addition, PCGC was completely unable to detect the parameter changes in **DGP2**. The Tracy-Widom method was conservative in estimating the Type I error in all simulations. It had high power in **DGP1** and **DGP3**, although not as much as the Forkman's test. However, the Tracy-Widom method was extremely underpowered in **DGP2** when ( $\vartheta_1 < \vartheta_0$ ) as mentioned previously.

### Analysis of inflammatory biomarker data

This section summarizes the results of the analysis of the data on the FHS inflammatory biomarkers and dementia.

**Table 8:** Type I error and power under **DGP2**.

Method <sup>1</sup>	$n_0$	$n_1$	$p$	$\theta_1 = 1$	$\theta_1 = 2$	$\theta_1 = 3$	$\theta_1 = 4$	$\theta_1 = 5$
PCGC	500	500	10	0.05	0.056	0.052	0.05	0.06
PCGC	1000	1000	10	0.046	0.046	0.046	0.046	0.042
PCGC	1500	1500	10	0.056	0.046	0.044	0.048	0.05
Fork	500	500	10	1	1	0.05	0.762	1
Fork	1000	1000	10	1	1	0.074	1	1
Fork	1500	1500	10	1	1	0.064	1	1
TW	500	500	10	0.074	0.008	0.046	1	1
TW	1000	1000	10	0.192	0.026	0.062	1	1
TW	1500	1500	10	0.33	0.036	0.052	1	1

$\theta_1 = 3$  constitutes the null distribution

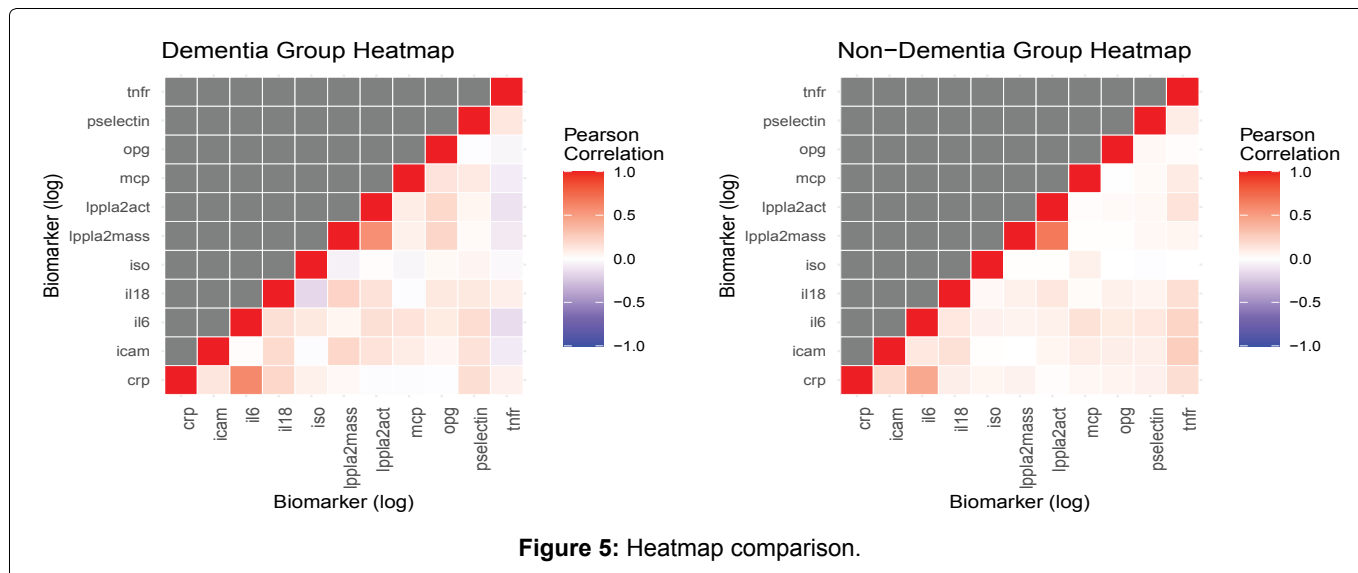
<sup>1</sup>PCGC: Principal Components Group Comparison; Fork: Forkman's Test; TW: Tracy-Widom

**Table 9:** Type I error and power under **DGP3**.

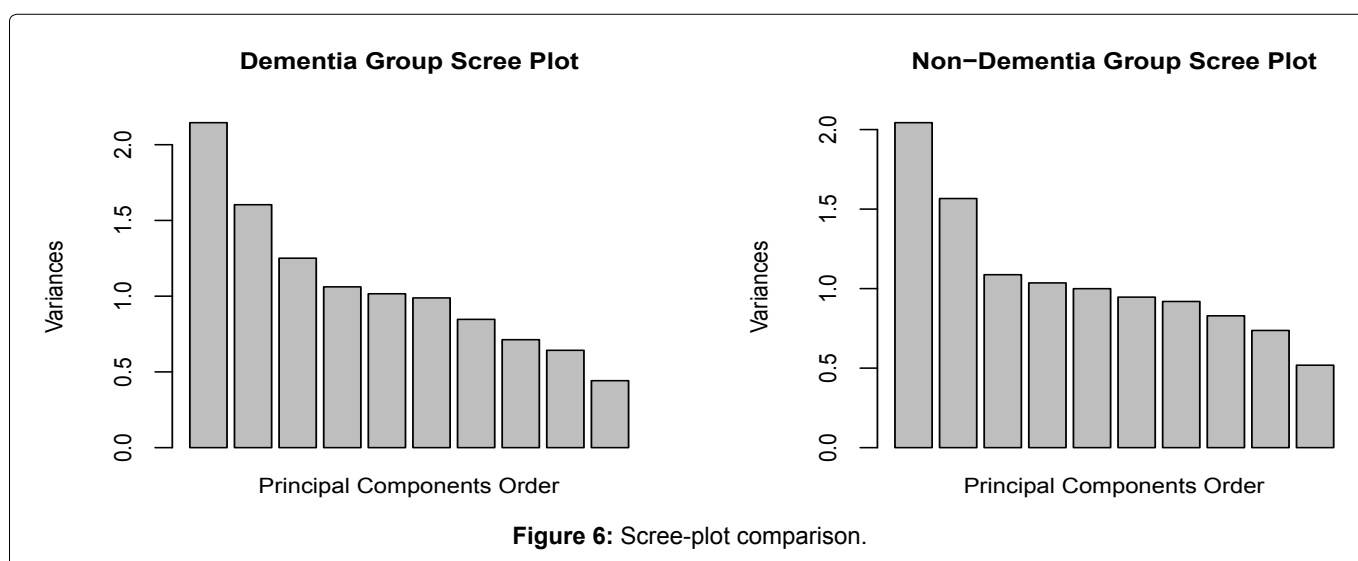
Method <sup>1</sup>	$n_0$	$n_1$	$p$	$\varphi_1 = -0.4$	$\varphi_1 = -0.2$	$\varphi_1 = 0$	$\varphi_1 = 0.2$	$\varphi_1 = 0.4$
PCGC	500	500	10	0.824	0.282	0.04	0.28	0.82
PCGC	1000	1000	10	0.936	0.382	0.05	0.416	0.944
PCGC	1500	1500	10	0.986	0.466	0.034	0.508	0.992
Fork	500	500	10	1	0.984	0.06	0.984	1
Fork	1000	1000	10	1	1	0.052	1	1
Fork	1500	1500	10	1	1	0.052	1	1
TW	500	500	10	1	0.916	0.03	0.892	1
TW	1000	1000	10	1	1	0.042	1	1
TW	1500	1500	10	1	1	0.026	1	1

$\varphi_1 = 0$  constitutes the null distribution

<sup>1</sup>PCGC: Principal Components Group Comparison; Fork: Forkman's Test; TW: Tracy-Widom



**Figure 5:** Heatmap comparison.



**Figure 6:** Scree-plot comparison.

**Preliminary analysis:** These analyzes sought evidence suggesting structural differences in biomarkers between participants in the dementia and non-dementia group. [Figure 5](#) illustrates the comparison between the partial correlation of the two groups using a heat map visualization. There seems to be some negative correlation among some variables in the dementia group, which does not appear in the non-dementia group.

Another possible way to visualize the structural differences in the partial covariance matrices between the non-dementia group and the dementia group is by comparison of the scree plots ([Figure 6](#)). In both groups the eigen values of the first two components appear to be very similar, while there exists minor differences in the latter components.

**Results of the inferential procedures:** The previous sections considered inflammatory biomarkers and the difference in the structure of the covariance between the two groups by generating a simple covariance matrix and the partial covariance matrix. The simple covariance matrix was calculated directly from the log-

**Table 10:** Data analysis, test statistics and P-values.

	Simple Covariance	Partial Covariance
PCGC	0.038 (0.64)	1.531 (0.01)
Forkman	0.947 (< 0.001)	0.953 (< 0.001)
Tracy-Widom	86.333 (< 0.001)	91.248 (< 0.001)

transformed biomarkers, while the partial covariance was calculated from the residuals after adjusting for the covariates. The results from analyzing the simple covariance would show whether there is a difference in covariance structure by itself, while the results from the partial covariance would show whether differences exist after adjusting for confounding factors.

The results show significant differences between the co-variance matrices in all cases except for the PCGC method on the simple covariance matrix ([Table 10](#)). Given the lackluster performance of PCGC in the simulations, the results of Tracy-Widom and Forkman's test seem more reasonable. Therefore, it appears that there is a significant difference between the simple covariance and partial covariance of the log-adjusted biomarkers of the dementia group and the non-dementia groups.

Partial covariance appears to be more significant than simple covariance, which gives credence to covariate adjustments. This result explicitly reveals differences in the dementia groups that the preliminary visualizations did not clearly show.

## Discussion

Despite extensive research on individual risk factors for dementia, there are far fewer studies that consider how these risk factors synergistically influence dementia. Current solutions to multivariable problems (such as machine learning) are complex given the large and varied data involved. On the other hand, statistical methods based on comparison of covariance matrices have been applied to a variety of problems, but not to biological risk factors of dementia. The purpose of this study was to test simpler covariance-based models that summarize multiple variables into a single test and score by efficiently examining the combined effect of many variables.

Three methods were investigated that combined the covariance matrices of two groups and then were summarized as a test statistic to test whether the matrices are structurally different. The validity of these methods via simulation, Type I error, and power analysis suggests that the Forkman's test method is effective in identifying both significant and insignificant differences between covariance matrices. By comparison, the PCGC method has low power and was unreliable in some cases. The Tracy-Widom method was conservative in estimating the Type I error but was accurate in estimating power. However, for non-normal distributed data, there were conflicting effects of variable variances and inter-variable covariances using the Tracy-Widom Statistics. These results indicate that the Forkman's test and Tracy-Widom methods were reliable methods for analyzing approximately normally distributed data.

The application of these methods to the clinical question of inflammatory biomarkers and their relationship with dementia risk found that these tests detected significant differences in the covariance structure of biomarkers between the dementia group and the non-dementia group. This result reinforces the established consensus that inflammation biomarkers are indeed risk factors for dementia. Furthermore, the 11 biomarkers analyzed in this paper can act as a collective risk factor for dementia. The collective covariance effect is certainly significant for dementia regardless of the significance of the individual mean effects.

There are limitations both in the methods and with the data. These methods are not well tested and thus are not established theoretically, which leads to a reliance on the use of simulation to assess the validity of these methods. Although these methods did determine the existence of a significant difference between two groups collectively among biomarkers, they could not

determine which variables are different. Furthermore, the data was not collected from the same period. Interleukin-18 (IL18) samples were measured from blood from the 7<sup>th</sup> health exam (1998- 2001), while the rest were measured from health exam 8 (2005-2008).

Furthermore, the covariate information came mainly from exam 8 values; exam 7 values were used only when exam 8 values are unavailable. These compromises could introduce additional noise that could not be controlled in order to maximize the number of biomarkers in the study. Furthermore, there was an imbalance in sample sizes between the non-dementia group ( $n = 2536$ ) and the dementia group ( $n = 168$ ). Imbalanced data introduces bias [35] and 'wastes' the extra samples in the non-dementia group. Future research is warranted to determine theoretical justifications for the Forkman's test and Tracy-Widom methods using more balanced study groups.

In summary, Forkman's test and Tracy-Widom methods are valid candidates for testing the significance of co-variance differences among standardized variables between groups, while the PCGC method is unreliable. Forkman's test appears to have suitable Type I error estimates and strong power, while Tracy-Widom has conservative Type I error estimates but good power. In a clinical research application, data analysis showed significant differences in protein biomarker covariance matrices between the non-dementia group and the dementia group. This result is expected, given the literature on inflammatory biomarkers as a risk factor for dementia. Utilizing new and cutting-edge analytic approaches can further enrich the understanding of inflammatory biomarkers and their effects on dementia.

## Statements and Declarations

### Acknowledgements

We would like to thank the participants of the Framingham Heart Studies for their contributions to the study.

### Funding

This paper is generously supported by Framingham Heart Study's National Heart, Lung, and Blood Institute contract (N01-HC-25195), and the National Institute on Aging (AG016495, AG008122, AG062109, AG068753).

### Conflict of interest

Rhoda Au is a scientific advisor to Signant Health and a scientific consultant to Biogen and the Davos Alzheimer's Collaborative (DAC). She also serves as Director of the Global Cohort Program for DAC. The other authors have no conflict of interest to declare.

### Ethical approval

All participants in the FHS study referenced in the data analysis have provided informed consent. The

study protocol received approval from the Institutional Review Board of the Boston University Medical Center. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement was followed.

### Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### References

- Morgan AR, Touchard S, Leckey C, Hagan CO, Nevado-Holgado AJ, et al. (2019) Inflammatory biomarkers in Alzheimer's disease plasma. *Alzheimer's Dement* 15: 776-787.
- Pasqualetti G, Brooks DJ, Edison P (2015) The role of neuroinflammation in dementias. *Curr Neurol Neurosci Rep* 15: 17.
- Hopperton KE, Mohammad D, Trepanier MO, Giuliano V, Bazinet RP (2018) Markers of microglia in post-mortem brain samples from patients with Alzheimer's disease: A systematic review. *Mol Psychiatry* 23: 177-198.
- Wyss-Coray T, Rogers J (2012) Inflammation in Alzheimer disease - a brief review of the basic science and clinical literature. *Cold Spring Harb Perspect Med* 2: a006346.
- Frontera JA, Provencio JJ, Sehba FA, McIntyre TM, Nowacki AS, et al. (2017) The role of platelet activation and inflammation in early brain injury following subarachnoid hemorrhage. *Neurocrit Care* 26: 48-57.
- Kang S, Tanaka T, Narazaki M, Kishimoto T (2019) Targeting interleukin-6 signaling in clinic. *Immunity* 50: 1007-1023.
- Shimamura M, Nakagami H, Osako MK, Kurinami H, Koriyama H, et al. (2014) OPG/RANKL/RANK axis is a critical inflammatory signaling system in ischemic brain in mice. *Proc Natl Acad Sci USA* 111: 8191-8196.
- Staerk L, Preis SR, Lin H, Lubitz SA, Ellinor PT, et al. (2020) Protein biomarkers and risk of atrial fibrillation: The FHS. *Circ Arrhythm Electrophysiol* 13: e007607.
- Storey JD (2002) A direct approach to false discovery rates. *J. R. Statist. Soc Series B (Statistical Methodology)* 64: 479-498.
- Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 2: 37-52.
- Baba K, Shibata R, Sibuya M (2004) Partial correlation and conditional correlation as measures of conditional independence. *Aust N Z J Stat* 46: 657-664.
- Krzanowski WJ (1979) Between-groups comparison of principal components. *J Stat Data Sci Educ* 74: 703-707.
- Forkman J, Josse J, Piepho HP (2019) Hypothesis tests for principal component analysis when variables are standardized. *J Agric Biol Environ Stat* 24: 289-308.
- Johnstone IM (2009) Approximate null distribution of the largest root in multivariate analysis. *Ann Appl Stat* 3: 1616-1633.
- Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP (1975) The framingham offspring study. Design and preliminary data. *Prev Med* 4: 518-525.
- Fricker ZP, Pedley A, Massaro JM, Vasan RS, Hoffmann U, et al. (2019) Liver fat is associated with markers of inflammation and oxidative stress in analysis of data from the framingham heart study. *Clin Gastroenterol Hepatol* 17: 1157-1164.e4.
- Li W, Dorans KS, Wilker EH, Rice MB, Ljungman PL, et al. (2019) Short-term exposure to ambient air pollution and circulating biomarkers of endothelial cell activation: The framingham heart study. *Environ Res* 171: 36-43.
- Satizabal CL, Beiser AS, Chouraki V, Chêne G, Dufouil C, et al. (2016) Incidence of dementia over three decades in the framingham heart study. *N Engl J Med* 374: 523-532.
- Yuan J, Maserejian N, Liu Y, Devine S, Gillis C, et al. (2021) Severity distribution of Alzheimer's disease dementia and mild cognitive impairment in the framingham heart study. *J Alzheimers Dis* 79: 807-817.
- Julie J, François H (2012) Handling missing values in exploratory multivariate data analysis methods. *Journal de la société Française de Statistique* 153: 79-99.
- Podcasy JL, Epperson CN (2016) Considering sex and gender in Alzheimer disease and other dementias. *Dialogues Clin Neurosci* 18: 437-446.
- Colautti RI, Barrett SCH (2011) Population divergence along lines of genetic variance and covariance in the invasive plant *Lythrum salicaria* in eastern North America. *Evolution* 65: 2514-2529.
- Robinson MR, Wilson AJ, Pilkington JG, Clutton-Brock TH, Pemberton JM, et al. (2009) The impact of environmental heterogeneity on genetic architecture in a wild population of soay sheep. *Genetics* 181: 1639-1648.
- Kolbe JJ, Revell LJ, Szekely B, Brodie ED III, Losos JB (2011) Convergent evolution of phenotypic integration and its alignment with morphological diversification in Caribbean *Anolis* ecomorphs. *Evolution* 65: 3608-3624.
- Monteiro LR, Nogueira MR (2010) Adaptive radiations, ecological specialization, and the evolutionary integration of complex morphological structures. *Evolution* 64: 724-744.
- Mallitt KL, Bonser SP, Hunt J (2010) The plasticity of phenotypic integration in response to light and water availability in the pepper grass, *Lepidium bonariense*. *Evolutionary Ecology* 24: 1321-1337.
- Barker BS, Phillips PC, Arnold SJ (2010) A test of the conjecture that g-matrices are more stable than b-matrices. *Evolution* 64: 2601-2613.
- Campbell DR, Weller SG, Sakai AK, Culley TM, Dang PN, et al. (2011) Genetic variation and covariation in floral allocation of two species of *Schiedea* with contrasting levels of sexual dimorphism. *Evolution* 65: 757-770.
- Cattell RB (1966) The scree test for the number of factors. *Multivariate Behav Res* 1: 245-276.
- Tracy CA, Widom H (1994) Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics* 159: 151-174.
- Johnstone IM (2001) On the distribution of the largest eigenvalue in principal components analysis. *Ann Stat* 29: 295-327.
- Chiani M (2014) Distribution of the largest eigenvalue for real wishart and gaussian random matrices and a simple approximation for the tracy-widom distribution. *J Multivar Anal* 129: 69-81.
- Zheng S, Lin R, Guo J, Yin G (2017) Testing homogeneity of high-dimensional covariance matrices. *Statistica Sinica* 30: 35-53.
- Frost HR, Amos CI, Moore JH (2016) A global test for gene-gene interactions based on random matrix theory. *Genet Epidemiol* 40: 689-701.
- King G, Zeng L (2001) Logistic regression in rare events data. *Political Analysis* 9: 137-163.