




Interpretable High-Dimensional Inference Via Score Projection With an Application in Neuroimaging

Simon N. Vandekar, Philip T. Reiss & Russell T. Shinohara


To cite this article: Simon N. Vandekar, Philip T. Reiss & Russell T. Shinohara (2018): Interpretable High-Dimensional Inference Via Score Projection With an Application in Neuroimaging, Journal of the American Statistical Association, DOI: [10.1080/01621459.2018.1448826](https://doi.org/10.1080/01621459.2018.1448826)

To link to this article: <https://doi.org/10.1080/01621459.2018.1448826>

 View supplementary material 

 Published online: 07 Aug 2018.

 Submit your article to this journal 

 Article views: 220

 View Crossmark data 



Interpretable High-Dimensional Inference Via Score Projection With an Application in Neuroimaging

Simon N. Vandekar^a, Philip T. Reiss^b, and Russell T. Shinohara,^a for the Alzheimer's Disease Neuroimaging Initiative

^aDepartment of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA; ^bDepartment of Statistics, University of Haifa, Haifa, Israel

ABSTRACT

In the fields of neuroimaging and genetics, a key goal is testing the association of a single outcome with a very high-dimensional imaging or genetic variable. Often, summary measures of the high-dimensional variable are created to sequentially test and localize the association with the outcome. In some cases, the associations between the outcome and summary measures are significant, but subsequent tests used to localize differences are underpowered and do not identify regions associated with the outcome. Here, we propose a generalization of Rao's score test based on projecting the score statistic onto a linear subspace of a high-dimensional parameter space. The approach provides a way to localize signal in the high-dimensional space by projecting the scores to the subspace where the score test was performed. This allows for inference in the high-dimensional space to be performed on the same degrees of freedom as the score test, effectively reducing the number of comparisons. Simulation results demonstrate the test has competitive power relative to others commonly used. We illustrate the method by analyzing a subset of the Alzheimer's Disease Neuroimaging Initiative dataset. Results suggest cortical thinning of the frontal and temporal lobes may be a useful biological marker of Alzheimer's disease risk. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received May 2016
Revised February 2018

KEYWORDS

Alzheimer's disease;
Association test; Cortical
thickness; Neuroimaging;
Post hoc inference; Score test

1. Introduction

In scientific fields in which high-dimensional data are prominent, it is often of interest to test the association of a single continuous or categorical outcome with a large number of predictors. A common approach in neuroimaging is to reduce the number of hypothesis tests by testing sequentially. For example, an investigator might first perform a test for the association of a phenotype with an imaging variable averaged across the entire brain. If the test rejects the null hypothesis of no association between brain and phenotype, then subsequent tests are conducted on regional averages of the data or on every voxel in the image using multiplicity correction to address the number of tests performed. Often, location-specific results yield few or no significant findings due to reduced signal and the necessary adjustment for the large number of tests, even though the whole brain average data show a significant association.

In this article, we propose a unified approach to test the association of an imaging or other high-dimensional predictor with an outcome and perform post hoc inference to localize signal. The framework is a modification of Rao's score test for models with a high- or infinite-dimensional parameter defined on a compact space such as the brain. Though the approach is designed for hypothesis testing in neuroimaging, it is applicable to a wide range of scientific domains.

Rao's score test assumes a model where $Y_i \in \mathbb{R}$ are independent and identically distributed observations from density

$f(y; \theta)$ and that the parameter $\theta = (\alpha, \beta) \in \Theta \subset \mathbb{R}^{m+p}$ where $\alpha \in \mathbb{R}^m$ is a nuisance parameter and $\beta \in \mathbb{R}^p$ is the parameter of interest. We seek to test the hypothesis $H_0: \beta = \beta_0$ for some $\beta_0 \in \mathbb{R}^p$. Define the score function $U = U(\theta) = n^{-1} \sum_{i=1}^n \frac{\partial \log f(Y_i|\theta)}{\partial \beta}(\theta)$ and let $\theta_0 = (\alpha, \beta_0)$ be the true value of the parameter under H_0 . Let $S = U(\hat{\alpha}, \beta_0)$ be the score function evaluated at the maximum likelihood estimate of α under the null hypothesis H_0 . Under the null and the conditions described in Section S.2, the covariance of S can be obtained from the Fisher information evaluated at the null parameter value,



$$\Omega(\theta_0) = \mathbb{E} \left\{ [(\partial/\partial\theta) \log f(Y_1|\theta)]^T [(\partial/\partial\theta) \log f(Y_1|\theta)] \Big|_{\theta_0} \right\}.$$

The sum of scores (Sum) test originally discussed by Rao (1948) has been used in genetics and neuroimaging (Pan 2009; Madsen and Browning 2009; Kim et al. 2014). The Sum test is based on the statistic


$$n \frac{(S^T \zeta)^2}{\zeta^T \hat{\Omega} \zeta}, \quad (1)$$

where $\zeta \in \mathbb{R}^p$ is a given vector of weights. The denominator is an estimate of the variance of $S^T \zeta$, so that the statistic is asymptotically χ_1^2 (Rao 1948). This test has low power when a large number of variables are not associated with the outcome (Pan et al. 2014).

In the case of unknown weights, when $p < n$, Rao (1948) proposed maximizing the Sum test statistic with respect to the

CONTACT Russell T. Shinohara  rshi@pennmedicine.upenn.edu  Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA 19104.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JASA.

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

weights,

$$\max_{\zeta \neq 0} n \frac{(S^T \zeta)^2}{\zeta^T \hat{\Omega} \zeta} = n S^T \hat{\Omega}^{-1} S. \quad (2)$$

When $n > p$ this statistic is approximately distributed as χ_p^2 under the null, however when $n < p$, it is not defined because $\hat{\Omega}$ is singular.

For finite-dimensional parameters, our proposed test can be thought of as a generalization of Rao's test in the case where the estimate of the information matrix is singular. When $p > n$, the test maximizes the statistic (1) with respect to the vector ζ over a subspace, \mathbb{L} , of \mathbb{R}^p . Maximization of the Sum test in the subspace \mathbb{L} is equivalent to projecting the scores for the original model to a lower-dimensional space within which the information matrix is invertible. For this reason, we call the test a projected score test (PST). The procedure does not assume sparsity, but attempts to conserve power by reducing the dimension of the data and performing inference in the lower-dimensional space.

In many cases, if a score test rejects H_0 , then it is of primary interest to perform post hoc inference to identify nonzero parameters. In neuroimaging, this amounts to a high-dimensional testing problem where the association is tested at each location in the image. The standard approach is to perform a hypothesis test at each parameter location and use a multiplicity correction procedure. Such methods in neuroimaging that control the family-wise error rate (FWER) have relied on Gaussian random field theory (Friston et al. 1994), but have recently been shown to have Type 1 error rates far from the nominal level in real data due to unmet assumptions (Silver, Montana, and Nichols 2011; Eklund, Nichols, and Knutsson 2016). Recently, considerable research activity has focused on leveraging the dependence of the tests to control the false discovery rate (FDR) in high-dimensional settings (Efron 2007). Sun et al. (2015) developed a procedure to control the FDR for spatial data as well as an approach for controlling the expected proportion of false clusters. Fan, Han, and Gu (2012) discussed estimation of the false discovery proportion (FDP) under dependence for normally distributed test statistics based on a factor approximation. In contrast, the PST post hoc inference procedure is performed by projecting the scores onto \mathbb{L} , and controlling the FWER of the projected scores.

Several recent studies have considered hypothesis tests for functional data, which is conceptually similar to our approach for an infinite-dimensional parameter. Reiss and Ogden (2010) proposed inverting simultaneous confidence bands for the parameter of a functional predictor to test which locations of the image are associated with the outcome. Smith and Fahrmeir (2007) used a binary Markov random field model to compute the joint probability that the marginal parameter estimates are equal to zero. Our post hoc inference is most similar to Smith and Fahrmeir (2007) as the interpretation of the contribution of the scores retains a marginal interpretation.

We derive the asymptotic null distribution of the PST statistic under some standard regularity conditions. For a normal linear model, we show how the finite sample distribution of our statistic can be calculated exactly for fixed n and p . For data that are measured on a compact space, such as brain images, we discuss

sufficient theoretical assumptions for characterizing test behaviors as both n and p approach infinity.

Our approach to asymptotics in p studies the growth of dimension of the grid at which the underlying stochastic process is observed. That is, as $p \rightarrow \infty$ we assume that the data are observed at increasing resolution. The rate that p increases is thus not dependent on n . In contrast, high-dimensional tests that do not make this assumption often have restrictions on the rates of growth of n and p . For example, Xu et al. (2016) bound the rate at which p increases by a function of n , and Cai, Liu, and Xia (2014) required that the maximum expected value of the false null statistics is larger than a given function of n and p .

To demonstrate how the test can be used in neuroimaging, we investigate the association of cortical thickness with mild cognitive impairment (MCI) in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, a dataset with $p = 18,715$ and $n = 628$. The outer surface of the brain (cortex) represents a highly folded sheet in three-dimensional space. The thickness of the cortex is known to be affected in individuals with psychopathology and neurological illness. MCI is a subtle pre-Alzheimer's disease decline in cognitive functioning. There is significant clinical interest in finding biological markers of MCI to identify those at risk for developing Alzheimer's disease, as prevention strategies and therapies for early disease are increasingly common. In this dataset, we seek to localize regions of the brain where cortical thinning provides additional information with regard to the diagnosis of MCI beyond what can be ascertained by neurocognitive scales alone.

For the remainder of the article, we denote matrices by uppercase italic letters (X), vectors by lowercase (x), and random vectors by uppercase Roman letters (X). Hilbert spaces are denoted with black-board letters (\mathbb{X}) and Greek letters denote model parameters. For the singular value decomposition (SVD) of any matrix, we will assume that the smallest dimensions of the matrices obtained are equal to the rank of the matrix. \xrightarrow{L} denotes convergence in law and \xrightarrow{P} denotes convergence in probability.

2. The Projected Score Test

In Section 2.1, we define the finite parameter PST statistic and give its asymptotic distribution for fixed p . In Section 2.2, we describe conditions sufficient for studying asymptotics in p . We discuss the PST for normal linear models in Section 2.3.

2.1. PST for Finite-Dimensional Parameters

We assume the observed data are finite-dimensional representations that are generated from an underlying stochastic process. Here, we informally define the finite-dimensional likelihood and refer the interested reader to Appendix A.1 for further details on deriving the finite-dimensional likelihood from the infinite-dimensional likelihood.

Let \mathbb{V} be a nonempty compact subset of \mathbb{R}^3 and $\mathcal{L}^2(\mathbb{V})$ be the space of square integrable functions from \mathbb{V} to \mathbb{R} . \mathbb{V} represents the space on which data can be observed; in neuroimaging this space is the volume of the brain. The underlying images G_i take values in $\mathcal{L}^2(\mathbb{V})$; however, the observed data are p -dimensional

discretizations of this image. Throughout this section, we model an outcome $Y_i \in \mathbb{R}$ as a function of observed image data $G_{ip} \in \mathbb{R}^p$ and a set of covariates $X_i \in \mathbb{R}^k$. Thus, the observed data can be described as independent and identically distributed observations (Y_i, X_i, G_{ip}) . We denote the collection of data by $Y = (Y_1, \dots, Y_n)$ and similarly define $X = (X_1, \dots, X_n)$ and $G_p = (G_{1p}, \dots, G_{np})$. We define a parameter $\theta_p = (\alpha, \beta_p) \in \Theta \subset \mathbb{R}^{m+p}$, where $\alpha \in \mathbb{R}^m$ is a nuisance parameter and $\beta_p \in \mathbb{R}^p$ is the parameter of interest. Together these parameters describe the conditional distribution of Y given the imaging data and covariates. We allow $m \geq k$ to flexibly model the relationship of the covariates and outcome, for example, with unpenalized splines.

Denote the finite-dimensional likelihood by $\ell(\theta_p; Y) = n^{-1} \sum_{i=1}^n \log f(Y_i | \theta_p, X, G_p) / \partial \beta_p$. Define the score function $U_{np} = \frac{\partial \ell}{\partial \beta_p} \{\theta_p; Y\}$ and let

$$S_{np} = U_{np}(\hat{\alpha}, \beta_{p0}) \in \mathbb{R}^p \quad (3)$$

denote the score function evaluated at the maximum likelihood estimate (MLE) under the null hypothesis

$$H_0 : \beta_p = \beta_{p0}. \quad (4)$$

Let the Fisher information for the full model be

$$\begin{aligned} \Omega_F(\theta_{p0}) &= \mathbb{E}_{\theta_{p0}} \left(\left\{ \frac{\partial}{\partial \theta_p} \log f(Y_1 | \theta_p, X, G_p) \right\} \right. \\ &\quad \times \left. \left\{ \frac{\partial}{\partial \theta_p} \log f(Y_1 | \theta_p, X, G_p) \right\}^T \Big|_{\theta_{p0}} \right) \\ &= \begin{bmatrix} \Omega_{\alpha\alpha} & \Omega_{\alpha\beta} \\ \Omega_{\beta\alpha} & \Omega_{\beta\beta} \end{bmatrix}, \end{aligned} \quad (5)$$

where $\theta_{p0} = (\alpha, \beta_{p0})$. Then the asymptotic variance of $\sqrt{n}S_{np}$ under H_0 is (see, e.g., Van der Vaart 2000)

$$\Omega(\theta_{p0}) = \Omega_{\beta\beta} - \Omega_{\beta\alpha} \Omega_{\alpha\alpha}^{-1} \Omega_{\alpha\beta}. \quad (6)$$

With the finite parameter scores defined, we can define the PST.

Definition 1. Let $P_{\mathbb{L}}$ be the orthogonal projection matrix onto a linear space $\mathbb{L} \subset \mathbb{R}^p$ with $r = \dim(\mathbb{L}) < n - m$. Let S_{np} be as defined in (3) and $\hat{\Omega}$ be the plug-in estimator of the covariance (6) obtained from

$$\begin{aligned} \hat{\Omega}_F &= n^{-1} \sum_{i=1}^n \left(\frac{\partial \log f(Y_i | \theta_p, X, G_p)}{\partial \theta_p} \right) \\ &\quad \times \left(\frac{\partial \log f(Y_i | \theta_p, X, G_p)}{\partial \theta_p} \right)^T \Big|_{\hat{\theta}_{p0}}, \end{aligned} \quad (7)$$

where $\hat{\theta}_{p0} = (\hat{\alpha}, \beta_{p0})$ denotes the maximum likelihood estimate of the parameter vector under the null hypothesis $H_0 : \beta = \beta_{p0}$. Then the PST statistic with respect to \mathbb{L} is defined as

$$R^{\mathbb{L}} = \max_{\zeta \in \mathbb{L} \setminus \{0\}} n \frac{(\zeta^T S_{np})^2}{\zeta^T \hat{\Omega}(\theta_{p0}) \zeta} = \max_{\gamma \in \mathbb{R}^p \setminus \{0\}} n \frac{(\gamma^T P_{\mathbb{L}} S_{np})^2}{\gamma^T P_{\mathbb{L}} \hat{\Omega}(\theta_{p0}) P_{\mathbb{L}} \gamma}. \quad (8)$$

The following theorem gives the asymptotic distribution (with respect to n) of the PST statistic provided the same regularity conditions required for the convergence of the scores to a multivariate normal random variable.

Theorem 1. Assume all objects are as described in Definition 1. Let $P_{\mathbb{L}} = QQ^T$ where the columns of the $p \times r$ matrix Q are any orthonormal basis for \mathbb{L} . Define

$$V_p = V(\theta_{p0}) = Q^T \Omega(\theta_{p0}) Q,$$

and assume the estimate $\hat{V}_{np} = Q^T \hat{\Omega}(\theta_{p0}) Q$ is invertible, and that the conditions given in Section S.2 are satisfied. Then, under the null (4), the rotated scores $S_{np}^{\mathbb{L}} \equiv Q^T S_{np}$ satisfy

$$n^{1/2} S_{np}^{\mathbb{L}} \xrightarrow{P} S_p^{\mathbb{L}} \sim N_r(0, V), \quad (9)$$

the PST statistic (8) is

$$R^{\mathbb{L}} = n \left(S_{np}^{\mathbb{L}} \right)^T \hat{V}_{np}^{-1} S_{np}^{\mathbb{L}}, \quad (10)$$

and $R^{\mathbb{L}} \xrightarrow{L} \chi_r^2$ as $n \rightarrow \infty$.

Theorem 1 requires that \hat{V}_{np} is nonsingular; however, in practice it is possible to ensure that Q is in the column space of $\hat{\Omega}(\theta_{p0})$, so that \hat{V}_{np}^{-1} exists. The proof of Theorem 1 is given in Section S.3. We also demonstrate there that the result of Theorem 1 does not depend on the choice of Q . We show how \mathbb{L} can be chosen for GLMs in Sections 3.1 and 3.2, and for imaging data in the analysis of the ADNI dataset in Section 5.

2.2. The PST as $p \rightarrow \infty$

We will show that as $p \rightarrow \infty$ the PST statistic converges to an integral over a stochastic process. The rate that p approaches infinity does not depend on the sample size. Here, we assume the data take values on the space $\mathbb{Y} = \mathbb{R} \times \mathbb{R}^k \times \mathcal{L}^2(\mathbb{V})$, where \mathbb{V} is a nonempty compact subset of \mathbb{R}^3 and $\mathcal{L}^2(\mathbb{V})$ is the space of square integrable functions, with respect to the Lebesgue measure, from \mathbb{V} to \mathbb{R} . Let $O_i = (Y_i, X_i, G_i)$, for $i = 1, \dots, n$, be independent and identically distributed with $Y_i \in \mathbb{R}$, $X_i \in \mathbb{R}^k$, and $G_i \in \mathcal{L}^2(\mathbb{V})$. Realizations of O_i are the outcome variable, a vector of k covariates and a function G_i . The infinite-dimensional score function, U_n , is defined in Section S.1 as the Fréchet derivative of the log likelihood with respect to the parameter $\beta \in \mathcal{L}^2(\mathbb{V})$, $U_n = U_n(v) = \frac{\partial \ell}{\partial \beta} \{(\alpha, \beta(v)); O(v)\}$. The score is defined for fixed β_0 as the stochastic process

$$S_n = U_n\{; (\hat{\alpha}, \beta_0)\} \in \mathcal{L}^2(\mathbb{V}). \quad (11)$$

Throughout Section 2.2, we assume that the infinite-dimensional scores converge in probability to a mean zero Gaussian process, S , that is,

$$n^{1/2} S_n \xrightarrow{P} S, \quad (12)$$

and that the dimension of the finite parameter α is fixed. Theorem S.1 in Section S.4 gives conditions under which this convergence holds (Van der Vaart 2000). The following definition of the PST statistic extends formula (10) to infinite-dimensional parameters.

Definition 2. Let $(q_1(v), \dots, q_r(v))$ be an orthonormal basis for the linear subspace $\mathbb{L} \subset \mathcal{L}^2(\mathbb{V})$ where q_j are continuous almost everywhere, and $r = \dim(\mathbb{L}) < n - m$. Also, assume that S_n and $\frac{\partial}{\partial \beta} \log f\{Y_i(v) | \hat{\alpha}, \beta_0(v)\}$, have continuous sample paths

with respect to v , where $\frac{\partial}{\partial \beta}$ denotes the Fréchet derivative. Define the column vector $S_n^Q \in \mathbb{R}^r$, with j th element

$$(S_n^Q)_j = \int_{\mathbb{V}} q_j(v) S_n(v) dv, \quad (13)$$

and let \hat{V}_n be the $r \times r$ matrix with (j, k) th element

$$\hat{V}_n^{j,k} = n^{-1} \sum_{i=1}^n \left(\int_{\mathbb{V}} q_j(v) \left[\frac{\partial}{\partial \beta} \log f\{Y_i | \hat{\alpha}, \beta_0(v)\} \right] dv \right) \times \left(\int_{\mathbb{V}} q_k(v) \left[\frac{\partial}{\partial \beta} \log f\{Y_i | \hat{\alpha}, \beta_0(v)\} \right] dv \right),$$

which is readily shown to be the covariance matrix of S_n^Q . Assume that \hat{V}_n is invertible. Then the PST statistic with respect to \mathbb{L} is defined as

$$R^{\mathbb{L}} = n (S_n^Q)^T \hat{V}_n^{-1} S_n^Q. \quad (14)$$

While we have given a definition of the PST statistic in infinite dimensions, in practice this statistic is not computable because it depends on functions which are only observed on a finite grid. The following theorem states that as the resolution of the grid is increased then the finite parameter PST statistic (10) converges to the infinite-parameter PST statistic (14). Moreover, as the sample size increases the statistic converges to a function of the Gaussian process S in (12).

Theorem 2. Let S_{np} be as defined in (3). Let Q_p be the $p \times r$ matrix with j th column $q_{jp} = (q_j(v_{1p})\nu(\mathbb{V}_{1p}), \dots, q_j(v_{pp})\nu(\mathbb{V}_{pp}))^T$, where v_{jp} and \mathbb{V}_{jp} are defined in Section S.1, $\mathcal{V}_p = \{\mathbb{V}_{1p}, \dots, \mathbb{V}_{pp}\}$, and ν denotes Lebesgue measure. Denote $S_{np}^Q = Q_p^T S_{np}$. Define $S^Q \in \mathbb{R}^r$ as the vector with j th element

$$(S^Q)_j = \int_{\mathbb{V}} q_j(v) S(v) dv.$$

Assume the conditions for Theorems 1 and S.1, and that S_n , $\frac{\partial}{\partial \beta} \log f\{Y_i(v) | \hat{\alpha}, \beta_0(v)\}$, and S have continuous sample paths with respect to v . Let $V = \mathbb{E}\hat{V}_n$. For $p_1 > p_2$, let \mathcal{V}_{p_1} be a refinement of \mathcal{V}_{p_2} , and assume that

$$\lim_{p \rightarrow \infty} \sup_k \nu(\mathbb{V}_{kp}) = 0. \quad (15)$$

Then as $n, p \rightarrow \infty$,

$$n (S_{np}^Q)^T \hat{V}_{np}^{-1} S_{np}^Q \xrightarrow{P} (S^Q)^T V^{-1} S^Q. \quad (16)$$

The proof is given in Section S.4.

2.3. The PST in Normal Linear Models

The finite-sample distribution of the PST statistic for a normal linear model can be found exactly. Define $X = [X_1, \dots, X_n]^T$ to be an $n \times m$ full-rank matrix of covariates for each observation, $G = [G_1, \dots, G_n]^T$ an $n \times p$ full-rank matrix of predictor variables of interest with $p > n$, and $Y = [Y_1, \dots, Y_n]^T$ an $n \times 1$ normal random vector with independent elements conditional on X and G . The score test with normal error is based on the model

$$Y_i = \alpha^T X_i + \beta^T G_i + E_i, \quad (17)$$

where $E_i \sim N(0, \sigma^2)$ are independent.

Theorem 3. Under model (17) and the null $H_0 : \beta = 0$

$$R^{\mathbb{L}} = L \frac{r(n-m)}{r + (n-m-r)F_{(n-m-r),r}}, \quad (18)$$

where $F_{(n-m-r),r}$ is F-distributed with $(n-m-r)$ and r degrees of freedom.

The proof can be found in Section S.5. The finite-sample distribution of $R^{\mathbb{L}}$ depends only on the sample size and the dimension of the basis, but not on the particular choice of \mathbb{L} .

3. Specifying the Linear Subspace \mathbb{L}

3.1. Specifying \mathbb{L} in Generalized Linear Models

Here, we discuss choices for the selection of \mathbb{L} in the context of GLMs with the canonical link function. We restrict attention to finite-dimensional parameters. Let X and G be as defined in Section 2.3. Assume the outcome $Y = [Y_1, \dots, Y_n]^T$ is from an exponential family where the expectation can be written as

$$h(\mathbb{E}Y_i) = \alpha^T X_i + \beta^T G_i,$$

where h is the canonical link function. For the GLM with canonical link, the scores are (McCullagh and Nelder 1989)

$$S_{np} = n^{-1} G^T (Y - \hat{Y}),$$

where

$$\hat{Y} = [\hat{Y}_1, \dots, \hat{Y}_n]^T$$

and $\hat{Y}_i = h^{-1}(x_i^T \hat{\alpha})$ is the i th fitted value under the null. Let Γ be the $n \times n$ diagonal matrix with i th diagonal element $\Gamma_{ii} = (Y_i - \hat{Y}_i)^2$. Then the estimate of the covariance (6) obtained using (7) is

$$\hat{\Omega} = n^{-1} \{G^T \Gamma G - G^T \Gamma X (X^T \Gamma X)^{-1} X^T \Gamma G\}. \quad (19)$$

The score statistic is obtained from the scores and the estimated information as in expression (2).

In this setup, the basis for \mathbb{L} can be constructed from the principal component analysis (PCA) of G . We write the PCA of G in terms of the SVD $G = T_* D Q^T$, where the principal scores are $T = T_* D = GQ$.

With this basis, the PST is equivalent to performing Rao's score test in a principal components regression model. To see this, first note that principal component regression is defined by

$$h(\mathbb{E}Y) = X\alpha + T\beta_T.$$

The scores for β_T are

$$S_{npT} = n^{-1} Q^T G^T (Y - \hat{Y}) = Q^T S_{np},$$

which are the same as the rotated scores in (9). The information estimate is also equivalent. Thus, the score test statistic, $n S_{npT}^T \hat{\Omega}_T^{-1} S_{npT}$, in principal component regression is equivalent to the PST statistic (10).

Another useful basis for \mathbb{L} can be constructed from vectors that are indicators of variables that are expected to have a similar relationship with the outcome. The anatomical basis we

use in Section 5 is an example. To define the basis vectors q_j , $j = 1, \dots, r$, we let $\mathcal{Q}_j \subset \{1, \dots, p\}$ such that $\mathcal{Q}_j \cap \mathcal{Q}_{j'} = \emptyset$ for $j \neq j'$, and then set the k th element of the j th basis vector to be $q_{kj} = \mathbb{1}(k \in \mathcal{Q}_j)$. These define orthogonal basis vectors since the sets \mathcal{Q}_j are disjoint. This basis is equivalent to averaging r subsets of the p predictor variables and performing a hypothesis test of the regression onto the r averaged variables.

The choice of the basis is a critical decision as it affects the power and interpretation of the post hoc inference. To clarify, under the alternative the scores have nonzero mean

$$\mathbb{E}S_{np} = \mu \in \mathbb{R}^p. \tag{20}$$

If the projection is orthogonal to μ , then the test will have power equal to the Type 1 error rate. The PCA basis assumes that μ has a spatial pattern similar to the covariance structure of the predictor variables. The anatomical basis assumes that all locations within a region have similar parameter values. We discuss the effect of the basis on the interpretation of the post hoc inference in Section 4.

3.2. Choosing a Dimension for the PCA Basis

To choose a dimension for the PCA basis, we propose an automatic procedure that sequentially tests bases of increasing dimension while controlling the Type 1 error rate. To do this, we first condition on the parameter estimate $\hat{\alpha}$ for the reduced model and perform the SVD $(\Gamma - \Gamma X(X^T \Gamma X)^{-1} X^T \Gamma)^{1/2} G = TDQ^T$. We use subsets of the columns of Q as the basis \mathbb{L} . For $j \neq k$

$$\begin{aligned} \text{cov}(q_j^T S_{np}, q_k^T S_{np} \mid \hat{\alpha}) &= q_j^T G^T (\Gamma - \Gamma X(X^T \Gamma X)^{-1} X^T \Gamma) G q_k \\ &= q_j^T Q D^2 Q^T q_k = 0. \end{aligned}$$

Thus, each rotated score $n^{1/2} q_j^T S_{np}$ is asymptotically independent, conditional on $\hat{\alpha}$, and can be tested by a separate chi-square test at level α^* . If this is done sequentially for $r = 1, \dots, n - m$, then, due to asymptotic independence, the probability of a Type 1 error under the global null is

$$\begin{aligned} \sum_{r=1}^{n-m} \mathbb{P}((n^{1/2} q_j^T S_{np})^2 > \chi_1^2(\alpha^*) \text{ for all } j \leq r) \\ \approx \sum_{r=1}^{n-m} (\alpha^*)^r \leq \sum_{r=1}^{\infty} (\alpha^*)^r = \frac{\alpha^*}{1 - \alpha^*}, \end{aligned}$$

where $\chi_1^2(\alpha^*)$ denotes the $1 - \alpha^*$ quantile of the chi-squared distribution with one degree of freedom. The approximate equality is due to the asymptotic approximation. To control the Type 1 error at level α , we choose $\alpha^* = \alpha / (1 + \alpha)$, then we sequentially test $r = 1, \dots, (n - m)$ until we fail to reject a test at level α^* . Note that the power depends critically on the first test in the sequence; subsequent tests serve only to increase the dimension of the basis. If the first component is orthogonal to μ in (20), the probability of reaching other components that are not orthogonal to μ is less than α^* .

A potentially more robust procedure is to test chunks of PCs by choosing $r_1 = 0, r_2, r_3, \dots, r_k = n - m$ and for the j th test perform a chi-square test of PCs $(r_j + 1), \dots, r_{j+1}$ on $r_{j+1} - r_j$ degrees of freedom. So long as the tests are independent, which

is true under the global null, the rejection threshold α^* will control the Type 1 error rate at level α . This automatic PCA (aPCA) procedure is implemented below by testing the first five components together and sequentially testing components $6, \dots, n - m$ one at a time. We demonstrate the procedure in the ADNI data analysis below and Type 1 error rates are assessed in Section 6.

4. Post hoc Inference for Localizing Signal

After performing the test of association using the PST, it is of primary interest to investigate the contribution of the scores to the statistic to identify which locations in the image are associated with the outcome and the direction of the effect. This can be done by projecting the scores onto \mathbb{L} and performing inference that controls the FWER for the projected scores. Because the projected scores are distributed in a linear subspace of \mathbb{R}^p , inference is much less conservative than performing inference on the original score vector.

Our aim is to construct a rejection region for each element of the projected score vector $(P_{\mathbb{L}} S_{np})_j$, for $j = 1, \dots, p$. Under the null, the projected scores are asymptotically normal,

$$P_{\mathbb{L}} S_{np} \sim N(0, P_{\mathbb{L}} \Omega P_{\mathbb{L}}).$$

The diagonal elements of $P_{\mathbb{L}} \Omega P_{\mathbb{L}}$ are not equal, so defining a single rejection threshold for all elements favors rejection for elements with larger variances. To resolve this issue, we scale by the inverse of the standard deviation of the projected scores. Let Δ be the diagonal matrix with j th diagonal element $\Delta_{jj} = 1/\sqrt{(P_{\mathbb{L}} \Omega P_{\mathbb{L}})_{jj}}$. Then the rejection threshold that controls the FWER for the standardized projected scores is defined by c that satisfies

$$\begin{aligned} 1 - \mathbb{P}(|(\Delta P_{\mathbb{L}} S_{np})_j| > c \text{ for some } j) \\ = \mathbb{P}(\max_j |(\Delta P_{\mathbb{L}} S_{np})_j| < c) = 1 - \alpha. \end{aligned} \tag{21}$$

Thus, the distribution of the infinity norm of $\Delta P_{\mathbb{L}} S_{np}$ can be used to compute a rejection threshold for the standardized projected scores that controls the FWER for the test of hypotheses about the projected scores

$$H_{0j}^{\mathbb{L}} : \mathbb{E}(\Delta P_{\mathbb{L}} S_{np})_j = 0. \tag{22}$$

We reject the null hypothesis (22) at location j if the observed projected score $|(\Delta P_{\mathbb{L}} S_{np})_j| > c$. This threshold corresponds to a single-step “maxT” joint multiple testing procedure (Westfall and Young 1993) and provides strong control of the FWER (Romano and Wolf 2005).

By (9) we have

$$\Delta P_{\mathbb{L}} S_{np} \xrightarrow{L} \Delta QV^{1/2}Z,$$

where $Z \sim N_r(0, I)$. Thus, we can approximate the region in (21) by finding c so that

$$\int_{|\Delta QV^{1/2}z|_{\infty} \leq c} \phi_r(z) dz = 1 - \alpha,$$

where ϕ_r denotes the PDF of Z . In practice, we approximate this integral by plugging in estimates for Δ and $V^{1/2}$.

The integral is difficult to calculate due to the large dimensions of Q , but can be approximated quickly and easily using Monte Carlo simulations. B simulations are used to estimate the CDF of the infinity norm, $\hat{F}_B(\cdot)$, which we use to obtain adjusted p -values for each observed standardized projected score, $(\Delta P_{\mathbb{L}S_{np}})_j$, by evaluating

$$p_j = 1 - \hat{F}_B((\Delta P_{\mathbb{L}S_{np}})_j), \quad (23)$$

or a rejection threshold can be obtained by using

$$c = \hat{F}_B^{-1}(1 - \alpha). \quad (24)$$

The p -value (23) for a given element of the standardized projected score vector is the probability of observing a j th projected score as large as $(\Delta P_{\mathbb{L}S_{np}})_j$ under the global null $H_0 : \beta = \beta_0$. The standard deviation of the Monte Carlo estimate (23) decreases at a \sqrt{B} rate and depends only on the volume of the space being integrated, so the procedure will perform well for computing adjusted p -values with a small error (Press et al. 2007). For example, with 10,000 simulations the standard deviation is on the order of $B^{-1/2} = 0.01$.

Rejection of the null hypothesis $H_0 : \beta = \beta_0$ is not strictly necessary to proceed with the post hoc inference procedure; the post hoc procedure can be used separately from the PST. In addition, it is important to note that the post hoc inference is restricted to the projected scores. When the alternative hypothesis is true, the rejection regions for the projected scores do not necessarily control the Type 1 error for the unprojected scores. This is demonstrated in the simulations in Section 6.

As mentioned above, the basis affects the interpretation of the inference on the projected scores. For the PCA basis, the interpretation is as follows: over repeated experiments, if the data are projected onto \mathbb{L} , then the probability of falsely rejecting one or more scores j with $(\Delta P_{\mathbb{L}\mu})_j = 0$ is at most α , where μ is as defined in (20). The anatomical basis assumes all locations in each basis vector have similar parameter values.

5. ADNI Data Analysis

We obtained data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The ADNI is a longitudinal observational study designed to investigate the early biomarkers of Alzheimer's disease; detailed MRI methods are given by Jack et al. (2008). Mild cognitive impairment (MCI) represents a subtle pre-Alzheimer's Disease decline in cognitive performance. The goal of our analysis is to identify whether a subset of the neuroimaging data from the ADNI can provide more information regarding diagnosis of MCI than the standardized memory tests obtained as part of the study. Moreover, we are interested in localizing areas of the cortex that differ on average between healthy controls (HC) and individuals with MCI. Three-dimensional T1-weighted structural images for 229 healthy controls and 399 subjects with MCI were obtained as part of the ADNI. This sample consists of subjects who had images and a composite memory score available at baseline.

Table 1. Results for the logistic regression of diagnosis onto covariates and whole-brain average cortical thickness. Results for average cortical thickness indicate a highly significant association between cortical thickness and diagnosis. SE denotes standard error.

	Estimate (SE)	p -Value
Age	−0.08 (0.02)	<0.001
Sex (Male)	−0.37 (0.26)	0.15
Memory score	−3.22 (0.27)	<0.001
Average cortical thickness	−4.23 (1.02)	<0.001

Cortical thickness was estimated using Freesurfer (Dale, Fischl, and Sereno 1999; Fischl and Dale 2000). Subjects' thickness data were registered to a standard template for analysis and smoothed at 10 mm FWHM to reduce noise related to pre-processing and registration. The template contains 18,715 vertex locations where cortical thickness is measured for each subject. Our goal is to identify whether the 18,715 cortical thickness measurements provide any additional information regarding the diagnosis of the individuals.

We perform the analysis in two ways: First, we proceed with standard analysis methods currently available for neuroimaging data in open access software (Fischl 2012). Second, we use the PST statistic and the high-dimensional inference procedure described above. For all analyses, we include age, sex, and the composite memory score as covariates (Crane et al. 2012).

5.1. Standard Neuroimaging Analysis Procedure: Average and Vertexwise Testing

Because neuroimaging studies typically collect many types of images with many covariates and possible outcomes, it is common to obtain a summary measure of a high-dimensional variable, and then proceed with further analysis if the summary measure appears to be associated with an endpoint of interest. In this analysis, we first take the average of all the cortical thickness measurements across the cortical surface for each subject and perform a regression with diagnosis as the outcome using logistic regression. Specifically, let C_i denote the average cortical thickness measurement for subject i , and X_i denote a vector with an intercept term, age, an indicator for sex, and the composite memory score for subject i . Then we fit the model

$$\text{logit}\{\mathbb{P}(Y_i = 1 \mid C_i, X_i)\} = X_i^T \alpha + C_i \beta_C.$$

If there is a significant relationship with the average cortical thickness measurements, that is, if we reject $H_0 : \beta_C = 0$, then we will proceed by performing mass-univariate vertexwise analyses by running a separate model at each point on the cortical surface.

The analysis using the average cortical thickness variable suggests a highly significant association of cortical thickness with diagnosis, indicating that subjects with thinner cortices are more likely to have MCI (Table 1). Based on these results we choose to investigate the relationship at each vertex to localize where in the cortex the association occurs.

For the vertexwise analyses, we use the software package Freesurfer to perform Benjamini–Hochberg (BH) correction separately across each hemisphere (Figure 1(a)). The spatial extent of the FDR-corrected results is more limited than what

we might expect given the very strong association between diagnosis and average cortical thickness. We conducted uncorrected exploratory analyses to further identify regions related to the whole-brain results (Figure 1(b)). The most significant results occur in left and right frontal lobes. These analyses suggest that distributed thinning in large portions of the frontal and temporal lobes is associated with increased risk of MCI; however, these results are not found using a method that guarantees control of the FWER or FDR.

5.2. PST and High-Dimensional Inference Procedures

To use the PST procedure, we perform the following steps:

1. Select a subspace \mathbb{L} .
2. Perform the PST for the association between the image and diagnosis.
3. If the test in Step 2 rejects, then perform post hoc inference as in Section 4.

We select a basis for \mathbb{L} in the two ways described in Section 3.1. For this analysis, we use the aPCA procedure described in Section 3.2 to choose the best PCA basis by testing the first five components together and sequentially testing components $6, \dots, n - m$. We also present results for the PCA basis fixed at several other dimensions ($r = 10, 20, 50$) to demonstrate how the basis dimension affects the results of the analysis. In addition, we consider a basis constructed from the $r = 148$ regions (74 per hemisphere) of the anatomical atlas of Destrieux et al. (2010). If we were unwilling to condition on the covariance structure of the scores or the anatomical atlas, a basis could be constructed that approximates a predetermined covariance

structure (e.g., a spatial AR(1)), or a covariance structure estimated from an independent sample, which can be used to construct the PCA basis. In addition to the PST, we perform the sequence kernel association test (SKAT; Wu et al. 2011), the sum of powered scores (SPU) test using the infinity norm, which corresponds to testing the max across the scores (Pan et al. 2014), and the adaptive sum of powered scores test (aSPU), which has competitive power to many other score tests (Pan et al. 2014). The SKAT is known to be powerful if there is a distributed signal, and the SPU infinity norm will be powerful for a sparse signal. The aSPU test combines multiple tests based on the norms $\|S\|_\gamma^\gamma$ for γ varying over a finite subset of \mathbb{N} by choosing one with the smallest p -value. Permutation testing is used to assess the significance of these statistics, however, recently, Xu et al. (2016) derived the asymptotic distributions of these tests under mild restrictions on the rate that p grows with respect to n .

The aPCA basis selected $r = 7$ by testing for $r = 5$ and then sequentially testing the next two PCs. With this and all other bases, we reject the null hypothesis using the PST (Table 2), indicating that there is an association between the image and diagnosis conditioning on the effects of age, sex, and composite memory score. The SKAT, SPU, and aSPU tests also reject the null.

Given the results of the PST, we are then interested in investigating how the scores contribute to the significant test statistic. To investigate the contributions of the scores to the PST statistic, we perform post hoc inference on the projected scores. We use 10,000 simulations to obtain rejection regions for each of the basis dimensions. The simulations ran for all bases in less than 2 min.

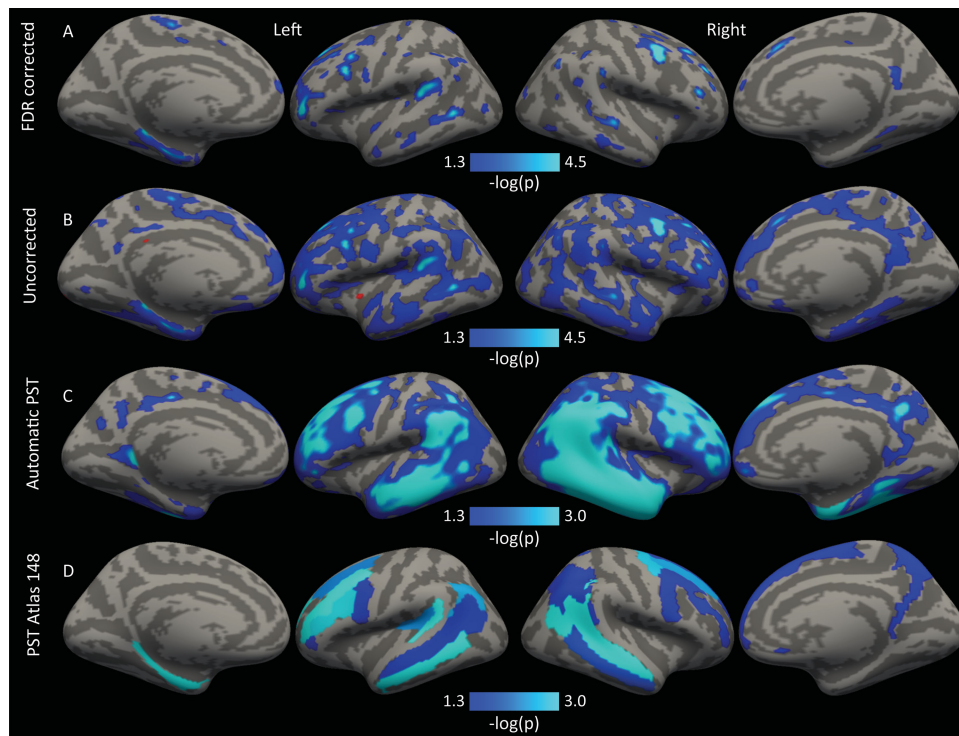


Figure 1. A comparison of inference procedures of the association between the imaging data and diagnosis. (a) Benjamini–Hochberg corrected vertexwise results, (b) uncorrected vertexwise results, and (c and d) results based on PST high-dimensional inference that control the FWER of the projected scores. (c) The dimension ($r = 7$) of the PCA basis for \mathbb{L} was selected using the automatic procedure. (d) The 148 dimensional basis constructed from the Destrieux anatomical atlas. Blue values show significant ($\alpha = 0.05$) negative association with diagnosis indicating that thinner cortex in these regions is associated with an MCI diagnosis.

Table 2. The χ^2 PST statistic and associated p -values for various basis dimensions; Automatic, 10, 20, and 50. “Anatomical” is a basis constructed from an anatomical atlas of dimension 148. The last column denotes the 5% family-wise error rejection thresholds for the projected scores, that is, the probability any projected score is above those values under the null is 5%. The thresholds are obtained using 10,000 simulations.

	Test statistic	p -Value	Rejection threshold
Automatic ($r = 7$)	38	<0.001	3.2
$r = 10$	41	<0.001	3.4
$r = 20$	50	<0.001	3.7
$r = 50$	91	<0.001	4
Anatomical basis	179	0.04	3.4
SKAT	5×10^6	<0.001	
SPU Inf	47	<0.001	
aSPU		0.001	

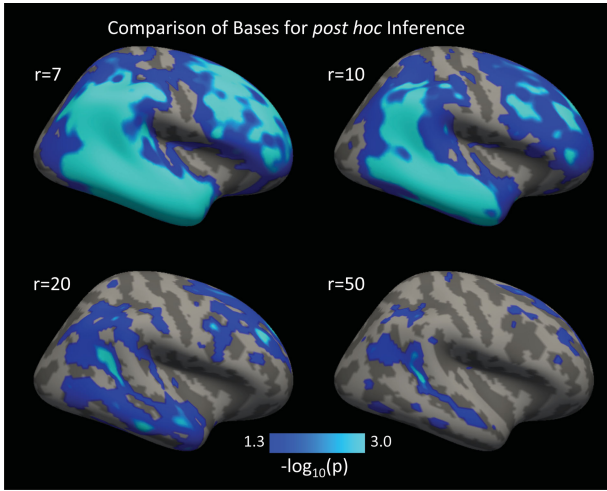


Figure 2. PST inference for PCA bases of various rank; Automatic (7), 10, 20, 50. Increasing the dimensionality of the basis increases spatial specificity, but comes with the cost of more conservative inference (see, e.g., Table 2).

The results suggest that thinner cortex in bilateral temporal and frontal lobes and right precuneus is associated with an increased risk of MCI (Figure 1(c) and 1(d)). Results are given as $-\log_{10}(p)$ where the p -value p is obtained using the simulated distribution (23). These locations are known to be thinner in AD versus HC as well as in AD versus MCI (Singh et al. 2006) and the results here demonstrate that there are significant differences between MCI and HC in the same region. The results indicate that the degree of frontal and temporal lobe thinning is correlated with diagnostic severity, and suggest that measurements of cortical thickness may provide useful information over and above neuropsychological scales in identifying people at risk for AD. Differences in these regions between MCI and HC were previously shown by Wang et al. (2009); however the authors did not control for multiple comparisons or adjust for covariates.

To reiterate, the blue areas in Figure 1(c) and 1(d) are based on low-rank inference and control the FWER of the projected scores. The procedure finds more significant locations over standard correction methods seen in Figure 1(a) and 1(b) by performing inference in a lower-dimensional space. The p -values obtained in Figures 1 and 2 use (23) and indicate the probability of observing a projected score statistic as or more extreme than the observed value under the global null $H_0 : \beta = 0$. Though interpretation is restricted to the projected scores, the results align with previous reports (Singh et al. 2006; Wang et al. 2009).

To demonstrate the impact of the choice of r , we performed post hoc inference on the scores for the PCA basis with four different dimensions (Figure 2). It is clear from Figure 2 that increasing the dimension of the basis increases the spatial specificity of the results. However, the larger bases also come with the cost of reduced power due to the larger degrees of freedom of the basis. This is also illustrated in Table 2, where the larger bases have a higher rejection threshold.

6. Neuroimaging-Based Simulation Study

In this section, we perform a simulation study using data generated for the right hemisphere of the cortical thickness data from the ADNI dataset measured at $p = 9361$ locations, called vertices. We simulate a binary outcome of interest, as in the ADNI analyses presented above. We select two anatomical regions (superior temporal sulcus and superior frontal sulcus) of 669 vertices total to have a negative association with the outcome and one region (anterior part of the cingulate gyrus and sulcus) of 191 vertices to have a positive association. The first two of these regions were selected because of their association in the ADNI dataset. The third region was selected to compare the performance of the tests when there are different locations with positive and negative associations with the outcome. To create a mean and covariance structure similar to real data within the regions of association, we create the mean vectors and covariance matrices for the simulations from the full sample of subjects used in the ADNI Freesurfer analysis above, yielding two full-rank covariance matrices, Σ_- and Σ_+ and mean vectors μ_+ and μ_- .

For each simulation, we select a random subset without replacement from the control subjects used in the ADNI neuroimaging analysis. Data within the negatively and positively associated regions are generated as independent multivariate normal distributions for each subject, with covariance structures $G_{i,-} \sim N(\mu_-, \Sigma_-)$ and $G_{i,+} \sim N(\mu_+, \Sigma_+)$, respectively. We centered the imaging data prior to analysis.

In each simulation, the outcome is generated under a logistic model

$$\text{logit}(\mathbb{E}Y_i) = \alpha_0 - \beta \mathbf{1}^T G_{i,-} + 2\beta \mathbf{1}^T G_{i,+}, \quad (25)$$

where α_0 is set to the log ratio of MCI to controls in the neuroimaging analyses section. $\mathbf{1}$, denotes a vector of ones, and β is an unknown parameter that we vary from 0 to 0.005. We multiply the values in the positive region by 2 to increase signal because it is a spatially smaller cluster than the two negative regions. In addition to simulations where the coefficients are constant across each region, in the supplement we perform simulations generating the parameters from a uniform distribution.

We construct the subspace \mathbb{L} in three ways. The first is to use the automatic procedure (Section 3.2) in each sample conditioning on the estimate \hat{a}_0 . The second basis type is constructed in each sample from the first $r = 10, 20, 50$ principal components from a PCA of $G(I - H)$, where H is the projection onto the intercept. The third basis is constructed from regions in the anatomical atlas of Destrieux et al. (2010), by randomly grouping the 74 regions into r groups and using normed indicator vectors for each group as the basis.

If we denote the set of indices with a nonzero association with the outcome by J , then the expectation of score j , μ_j , is nonzero only for $j \in J$, where μ is as defined in (20). Thus, for indices with $j \notin J$ we report Type 1 error and for indices with $j \in J$ we report power.

Similarly, the mean of the standardized projected scores, $\Delta P\mu$, determines Type 1 error and power for the projected scores ΔPS_{np} . The FWER and FDR of the projected scores are reported for the basis constructed from the anatomical atlas and the PCA bases. In general, no element of the standardized projected mean is exactly zero, so Type 1 error is assessed by thresholding the standardized projected parameter vector at the 0.2 quantile and reporting the rejection rate for vertices with projected parameter values below that threshold.

We perform 1000 simulations for sample sizes of $n = 100, 200$ and compare the PST for the automatic procedure and fixed bases with dimensions of $r = 10, 20, 50$. In addition, we compare the PST to the SKAT (Wu et al. 2011), the sum of powered scores (SPU) test using the infinity norm, and the aSPU (Pan et al. 2014). We assess pointwise power and Type 1 error of the PST inference procedure with uncorrected, Bonferroni-corrected, and BH-corrected results. We also compare FWER and FDR between methods. For the pointwise results, we assess the power and Type 1 error for the unprojected scores using inference designed for the projected scores.

Several types of bases for the PST demonstrate superior power to the other tests (Figure 3) due to their ability to remove the influence of unassociated scores from the test by maximizing over the basis. If the basis vectors are not informative about the structure of the signal, then the PST would not perform well. The PCA and aPCA PSTs leverage the spatial covariance of the data by using a basis constructed from the matrix G , which is an argument of the covariance of the scores as in Equation (19). The tests with rank $r = 50$ do not perform well with $n = 100$ due to the error in the asymptotic approximation; the error is evident in the conservative Type 1 error of these tests, which is well below the expected 0.05 threshold when $\beta = 0$. The aPCA has better

Table 3. Error rates for the projected scores for the automatic PCA bases and anatomical bases for $n = 200$ and $\beta = 0.002$.

	FWER	FDR
aPCA	0.06	<0.01
PCA 10	0.04	<0.01
PCA 20	0.03	<0.01
PCA 50	0.02	<0.01
Anatomical 5	0.03	0.02
Anatomical 10	0.04	0.02
Anatomical 20	0.04	0.02
Anatomical 50	0.06	0.02

power than the other PCA bases because a low rank basis suffices to capture the signal in the data, while the higher dimensional bases have low power due to the inclusion of basis vectors that do not capture signal. aSPU does not make any assumption about the structure of the signal but is adaptive to the sparsity of the signal, thus it performs better than the SKAT and PSTs with bases of higher dimension. In this simulation, the aSPU test does not perform as well as several of the PSTs because it does not leverage the spatial information in the covariance of the scores, which is informative in this case. However, if the covariance among predictors were uninformative about the signal, it is likely that the aSPU test would be superior.

As expected, the post hoc inference procedure controls the FWER of the projected scores for all basis dimensions (Table 3). In general, the post hoc inference procedure does not control the FWER or FDR of the unprojected scores (Table 4) as the inference is intended for the projected scores. However, for larger PCA bases our procedure does control the FDR (bold rows in Table 4). This is likely because the projection captures most of the variation in μ , so that the projection $\Delta P\mu$ is close to μ . Future investigation of whether inference for the projected scores will control any error rate for unprojected score vector is warranted.

The vertexwise error rate describes how effective a procedure is at controlling the error rate for the unprojected scores at each

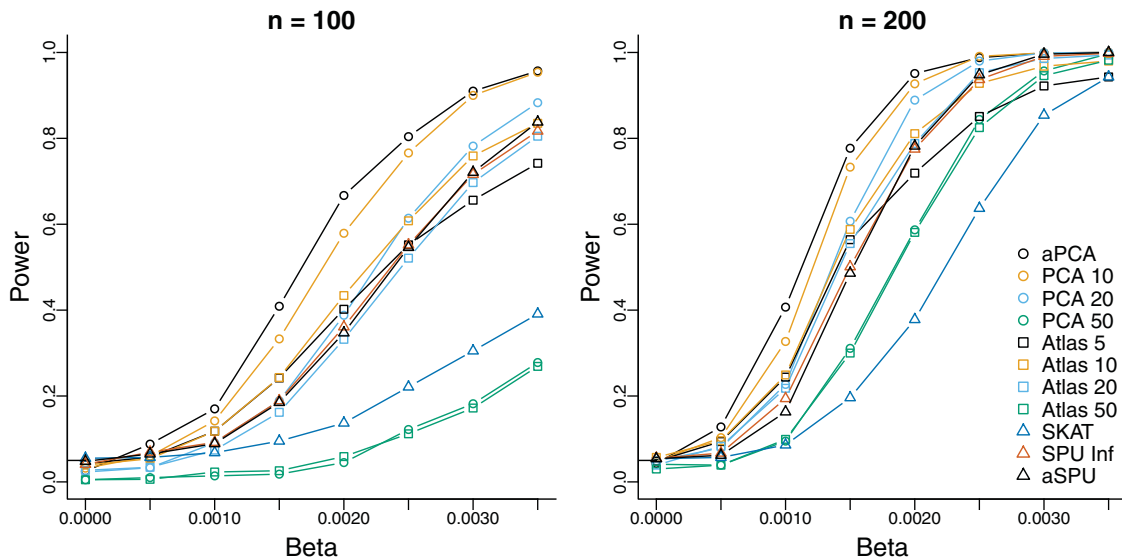


Figure 3. Power results for the PST with various bases compared to aSPU and the SKAT. “aPCA” is the automatic basis and “PCA r ” indicates the basis formed from the first r components of the PCA of the design matrix G . “Atlas r ” indicates the bases formed from the anatomical atlas with r regions covering the cortex.

Table 4. Hit (HR), false discovery (FDR), and family-wise error (FWER) rates for the unprojected scores. Higher dimensions of the PCA basis control the FDR while maintaining a higher hit rate for the unprojected scores. Note, however, that in practice it is not possible to know the dimension of the basis required to control the FDR of the unprojected scores. Bold rows control the FDR at $q = 0.05$. BH = Benjamini–Hochberg.

	HR	FDR	FWER
aPCA	0.69	0.13	0.70
PCA 10	0.55	0.05	0.45
PCA 20	0.34	0.02	0.23
PCA 50	0.12	0.01	0.10
Anatomical 5	0.16	0.79	0.27
Anatomical 10	0.34	0.64	0.57
Anatomical 20	0.53	0.45	0.84
Anatomical 50	0.63	0.17	0.70
Uncorrected	0.61	0.44	1.00
Holm	<0.01	<0.01	<0.01
BH	0.10	0.05	0.22

location. The vertexwise error rate of the PST inference procedure for the unprojected scores using the PCA 10 basis is low while maintaining better vertexwise power than BH (Figure 4; PCA 10). This is because in any given sample there may be a high false positive rate, but the errors across samples do not appear in the same locations. The BH and Bonferroni corrections both work well at controlling the vertexwise Type 1 error rate but have lower power than the PCA-based PST procedure (Figure 4). The bases constructed from the anatomical atlas tend to have large regions of vertexwise Type 1 error for the unprojected scores. At the largest basis dimension, the atlas allows for enough specificity to reduce the vertexwise error. All methods have lower power to detect the positive cluster than the two negative clusters. This may be due to the characteristics of the covariance structure in the positive cluster which overlaps gyral and sulcal regions.

7. Discussion

We have proposed the PST, a modification of the score test for high-dimensional data that works by projecting the scores to a lower-dimensional linear subspace. The procedure offers a novel post hoc inference on the projected scores by performing

inference in the subspace where the test statistic was estimated. Because the post hoc inference is based on the same model and degrees of freedom as the PST statistic, the interpretation of high-dimensional results agree closely with the results from the PST.

The ability to choose a subspace \mathbb{L} makes the procedure very flexible. For example, in medical imaging the basis for the space can be chosen based on anatomical or functional labels, or from data acquired in another imaging modality. Particular hypotheses can be targeted by selecting a basis that includes indicators of certain regions or weights particular locations to target specific spatial patterns. If orthogonal indicator vectors are used as the basis, then the approach can be seen as testing averages of subregions of the data as in Section 5. In this case, the PST procedure can be seen as a maxT multiple testing procedure of the regional averages that accounts for the correlation structure of the tests.

There are several limitations of the proposed procedure. First, the success of the procedure depends critically on the projection chosen. If a projection is chosen that is orthogonal to the mean vector, the PST will fail to capture any signal in the data. This is a limitation of any dimension-reducing procedure. Further research could investigate whether maximization of the score test with regularization can yield a test statistic whose distribution is tractable. Regularization may remove the subjectivity of selecting a basis and make the procedure more robust. Second, while the dimension reduction procedure preserves power and the results align closely with those from previous research, the inference does not guarantee control of the FWER or FDR of the original score vector. Future research will investigate how inference of the original score vector can be made by thresholding the projected score vector. This is similar in concept to the dependence-adjusted procedure discussed by Fan, Han, and Gu (2012) for controlling the FDP and may offer increased power by leveraging the covariance of the test statistics. These limitations notwithstanding, our procedure generalize Rao’s score test to the high- and infinite-dimensional settings and introduce a new inference approach based on projecting the test statistics to a lower-dimensional space where inference can be made on fewer degrees of freedom.

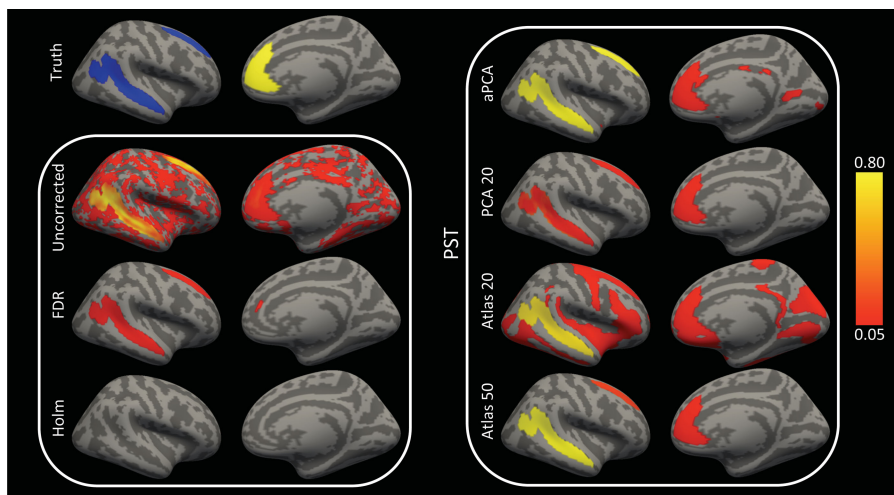


Figure 4. Vertexwise power or Type 1 error is measured as the proportion of simulated samples where each of the testing procedures rejects the null at a given location. Results are shown for $\beta = 0.002$. “Truth” indicates locations where signal was simulated according to the model (25).

Supplementary Materials

Proofs of all theorems and additional simulations are included in the Supplementary Material.

Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank Wei Pan and Haochang Shou for helpful discussions related to this work. Code to perform the analyses in this manuscript is provided at <https://bitbucket.org/simonvandekar/pst>. The investigators within the ADNI contributed to the design and implementation of ADNI and provided data, but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. The data used in this manuscript are publicly available at <http://adni.loni.usc.edu/>

Funding

SNV was supported by NIMH grant T32MH065218-11; PTR was supported by NIMH grant R01MH095836 and Israel Science Foundation grant 1777/16; RTS was supported by NINDS grant R01NS085211, R21NS093349, and R01NS060910, as well as NIMH grant R01MH112847.

References

- Cai, T. T., Liu, W., and Xia, Y. (2014), “Two-Sample Test of High Dimensional Means Under Dependence,” *Journal of the Royal Statistical Society, Series B*, 76, 349–372. [2]
- Crane, P. K., Carle, A., Gibbons, L. E., Insel, P., Mackin, R. S., Gross, A., Jones, R. N., Mukherjee, S., Curtis, S. M., Harvey, D., Weiner, M., and Mungas, D. (2012), “Development and Assessment of a Composite Score for Memory in the Alzheimer’s Disease Neuroimaging Initiative (ADNI),” *Brain Imaging and Behavior*, 6, 502–516. [6]
- Dale, A. M., Fischl, B., and Sereno, M. I. (1999), “Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction,” *Neuroimage*, 9, 179–194. [6]
- Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010), “Automatic Parcellation of Human Cortical Gyri and Sulci Using Standard Anatomical Nomenclature,” *NeuroImage*, 53, 1–15. [7,8]
- Efron, B. (2007), “Correlation and Large-Scale Simultaneous Significance Testing,” *Journal of the American Statistical Association*, 102, 93–103. [2]
- Eklund, A., Nichols, T. E., and Knutsson, H. (2016), “Cluster Failure: Why fMRI Inferences for Spatial Extent have Inflated False-Positive Rates,” *Proceedings of the National Academy of Sciences*, 113, 7900–7905. [2]
- Fan, J., Han, X., and Gu, W. (2012), “Estimating False Discovery Proportion Under Arbitrary Covariance Dependence,” *Journal of the American Statistical Association*, 107, 1019–1035. [2,10]
- Fischl, B. (2012), “FreeSurfer,” *NeuroImage*, 62, 774–781. [6]
- Fischl, B., and Dale, A. M. (2000), “Measuring the Thickness of the Human Cerebral Cortex from Magnetic Resonance Images,” *Proceedings of the National Academy of Sciences*, 97, 11050–11055. [6]
- Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., and Evans, A. C. (1994), “Assessing the Significance of Focal Activations Using Their Spatial Extent,” *Human Brain Mapping*, 1, 210–220. [2]
- Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., Whitwell, J. L., Ward, C., Dale, A. M., Felmlee, J. P., Gunter, J. L., Hill, D. L. G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C. S., Krueger, G., Ward, H. A., Metzger, G. J., Scott, K. T., Mallozzi, R., Blezek, D., Levy, J., Deb-bins, J. P., Fleisher, A. S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., and Weiner, M. W. (2008), “The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI Methods,” *Journal of Magnetic Resonance Imaging: JMRI*, 27, 685–691. [6]
- Kim, J., Wozniak, J. R., Mueller, B. A., Shen, X., and Pan, W. (2014), “Comparison of Statistical Tests for Group Differences in Brain Functional Networks,” *NeuroImage*, 101, 681–694. [1]
- Madsen, B. E., and Browning, S. R. (2009), “A Groupwise Association Test for Rare Mutations Using a Weighted Sum Statistic,” *PLoS Genetics*, 5, e1000384. [1]
- McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), Boca Raton, FL: Chapman and Hall/CRC. [4]
- Pan, W. (2009), “Asymptotic Tests of Association with Multiple SNPs in Linkage Disequilibrium,” *Genetic Epidemiology*, 33, 497–507. [1]
- Pan, W., Kim, J., Zhang, Y., Shen, X., and Wei, P. (2014), “A Powerful and Adaptive Association Test for Rare Variants,” *Genetics*, 197, 1081–1095. [1,7,9]
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007), *Numerical Recipes: The Art of Scientific Computing*, New York: Cambridge University Press. [6]
- Rao, C. R. (1948), “Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation,” *Mathematical Proceedings of the Cambridge Philosophical Society*, 44, 50–57. [1]
- Reiss, P. T., and Ogden, R. T. (2010), “Functional Generalized Linear Models with Images as Predictors,” *Biometrics*, 66, 61–69. [2]
- Romano, J. P., and Wolf, M. (2005), “Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing,” *Journal of the American Statistical Association*, 100, 94–108. [5]
- Silver, M., Montana, G., and Nichols, T. E. (2011), “False Positives in Neuroimaging Genetics Using Voxel-Based Morphometry Data,” *NeuroImage*, 54, 992–1000. [2]
- Singh, V., Chertkow, H., Lerch, J. P., Evans, A. C., Dorr, A. E., and Kabani, N. J. (2006), “Spatial Patterns of Cortical Thinning in Mild Cognitive Impairment and Alzheimer’s Disease,” *Brain*, 129, 2885–2893. [8]
- Smith, M., and Fahrmeir, L. (2007), “Spatial Bayesian Variable Selection With Application to Functional Magnetic Resonance Imaging,” *Journal of the American Statistical Association*, 102, 417–431. [2]
- Sun, W., Reich, B. J., Cai, T. T., Guindani, M., and Schwartzman, A. (2015), “False Discovery Control in Large-Scale Spatial Multiple Testing,” *Journal of the Royal Statistical Society, Series B*, 77, 59–83. [2]
- Van der Vaart, A. W. (2000), *Asymptotic Statistics* (Vol. 3), Cambridge: Cambridge University Press. [3]
- Wang, L., Goldstein, F. C., Veledar, E., Levey, A. I., Lah, J. J., Meltzer, C. C., Holder, C. A., and Mao, H. (2009), “Alterations in Cortical Thickness and White Matter Integrity in Mild Cognitive Impairment Measured by Whole-Brain Cortical Thickness Mapping and Diffusion Tensor Imaging,” *American Journal of Neuroradiology*, 30, 893–899. [8]
- Westfall, P. H., and Young, S. S. (1993), *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* (Vol. 279), New York: Wiley. [5]
- Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011), “Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test,” *American Journal of Human Genetics*, 89, 82–93. [7,9]
- Xu, G., Lin, L., Wei, P., and Pan, W. (2016), “An Adaptive Two-Sample Test for High-Dimensional Means,” *Biometrika*, 103, 609–624. [2,7]