

Canonical Correlation Analysis and Partial Least Squares for Identifying Brain–Behavior Associations: A Tutorial and a Comparative Study

Agoston Mihalik, James Chapman, Rick A. Adams, Nils R. Winter, Fabio S. Ferreira, John Shawe-Taylor, and Janaina Mourão-Miranda, for the Alzheimer's Disease Neuroimaging Initiative

ABSTRACT

Canonical correlation analysis (CCA) and partial least squares (PLS) are powerful multivariate methods for capturing associations across 2 modalities of data (e.g., brain and behavior). However, when the sample size is similar to or smaller than the number of variables in the data, standard CCA and PLS models may overfit, i.e., find spurious associations that generalize poorly to new data. Dimensionality reduction and regularized extensions of CCA and PLS have been proposed to address this problem, yet most studies using these approaches have some limitations. This work gives a theoretical and practical introduction into the most common CCA/PLS models and their regularized variants. We examine the limitations of standard CCA and PLS when the sample size is similar to or smaller than the number of variables. We discuss how dimensionality reduction and regularization techniques address this problem and explain their main advantages and disadvantages. We highlight crucial aspects of the CCA/PLS analysis framework, including optimizing the hyperparameters of the model and testing the identified associations for statistical significance. We apply the described CCA/PLS models to simulated data and real data from the Human Connectome Project and Alzheimer's Disease Neuroimaging Initiative (both of $n > 500$). We use both low- and high-dimensionality versions of these data (i.e., ratios between sample size and variables in the range of ~ 1 – 10 and ~ 0.1 – 0.01 , respectively) to demonstrate the impact of data dimensionality on the models. Finally, we summarize the key lessons of the tutorial.

<https://doi.org/10.1016/j.bpsc.2022.07.012>

Neuroimaging datasets with sample sizes of $n > 1000$ (e.g., UK Biobank, Human Connectome Project [HCP], Alzheimer's Disease Neuroimaging Initiative [ADNI]) represent a unique opportunity to advance population neuroscience and mental health (1–3). These datasets comprise multiple data modalities (e.g., structural magnetic resonance imaging, resting-state functional magnetic resonance imaging, mental health, cognition, environmental factors and genetics), several of which can be high-dimensional, meaning that there are hundreds or thousands of variables per subject. Understanding the links across these different modalities is fundamental for enabling new discoveries; however, analyzing multimodal datasets with more variables than samples poses technical challenges.

The most established methods to find associations across multiple modalities of multivariate data are canonical correlation analysis (CCA) (4) and partial least squares (PLS) (5). CCA and PLS have recently become very popular, with numerous applications linking brain imaging to behavior or genetics [e.g., (6–26)]. However, when the variables in at least one modality (e.g., brain) outnumber the sample size, standard CCA and PLS models may overfit, i.e., they are more likely to find

spurious associations that generalize poorly to independent samples [e.g., (26–28)]. Moreover, there is no unique standard CCA solution when the number of variables exceeds the sample size. Two approaches have been proposed to address this problem: 1) reducing the dimensionality of the data with principal component analysis (PCA) (9,10,12,22,24,26) and 2) using regularized extensions of CCA and PLS (11,20,23,27). However, most studies using these approaches have potential limitations. For instance, 1) they usually do not optimize the hyperparameters (e.g., the number of principal components [PCs] or amount of regularization) (9,10,12,15,22,24,26); 2) many studies do not test the significance of the associations using hold-out data (e.g., out-of-sample correlation) (7,9–11,22); and 3) they often do not assess the stability of the CCA/PLS model (7,9,18,21–25). Finally, few studies compare different CCA/PLS models and analytic frameworks across different datasets with different dimensionalities [e.g., (25–27)].

Several tutorial papers were recently published on CCA and PLS (29–32). Here, we complement these tutorials by discussing some important conceptual and practical aspects of these methods. These comprise 1) the advantages and disadvantages of the various CCA/PLS models, 2) the impact of PCA

and regularization on these models (e.g., on overfitting and stability), and 3) the importance of the analytic framework in optimizing the models' hyperparameters and performing statistical inference.

In Part 1, we present the theoretical background of these models and discuss the most common strategies to mitigate the problems caused when the ratio between sample size and number of variables is small (e.g., around ~ 0.1 – 0.01). We also examine the most prevalent analytical frameworks used with CCA/PLS models. In Part 2, we apply the models introduced in Part 1 to simulated data and real data from the HCP and ADNI ($n > 500$ in all). We illustrate how the different CCA/PLS models perform with data dimensionalities often used in practice (i.e., ratios between sample size and number of variables in the ranges of ~ 1 – 10 or ~ 0.1 – 0.01). Moreover, we show that regularization can be helpful even when the number of variables in both data modalities is smaller than the sample size. Mathematical details of the CCA/PLS models and their connections are provided in the [Supplement](#).

PART 1: TECHNICAL BACKGROUND OF CCA AND PLS

CCA/PLS Optimization and Nomenclature

CCA (4) and PLS (5) are multivariate latent variable models that capture associations across 2 modalities of data (e.g., brain and behavior). For example (Figure 1), \mathbf{X} contains voxel-level brain variables and \mathbf{Y} contains behavioral variables from item-level self-report questionnaires (and are matrices with rows and columns representing subjects and variables, respectively). Standard CCA/PLS models find pairs of brain and behavioral weights \mathbf{w}_x and \mathbf{w}_y (column vectors) such that the linear combination (weighted sum) of the brain and behavioral variables maximizes the correlation (CCA) or covariance (PLS) between the resulting latent variables, i.e., between $\xi = \mathbf{X}\mathbf{w}_x$ and $\omega = \mathbf{Y}\mathbf{w}_y$, respectively.

In the PLS literature, the weights are often referred to as saliences and the latent variables as scores. In the CCA literature, the weights are often referred to as canonical vectors, the latent variables as canonical variates, and the correlation between the latent variables as canonical correlations. The brain and behavior weights have the same dimensionality as their respective data modality (e.g., number of brain/behavioral variables) and quantify each brain and behavioral variable's contribution to the identified association. Sometimes, Pearson's correlations between the brain and behavioral variables and their respective latent variable are presented instead of the

model's weights and are called structure correlations (CCA) (33) or loadings (PLS) (34) (for details, see the [Supplement](#)). The latent variables (one latent variable score per data modality and subject) quantify how the associative effect is expressed across the sample. Table 1 summarizes the different nomenclatures used in the CCA and PLS literature.

While standard CCA refers to a single method, standard PLS refers to a family of methods with different modeling aims (e.g., assuming a symmetric or asymmetric relationship between the 2 data modalities; for details, see the [Supplement](#)). Both standard CCA and PLS can be solved by iterative [e.g., alternating least squares (35), nonlinear iterative PLS (36)] and noniterative [e.g., eigenvalue problem (29,34)] methods. In the case of iterative methods, once a pair of weights is obtained, the corresponding associative effect is removed from the data (by a process called deflation) and new associations are sought.

Because standard CCA maximizes the correlation between the latent variables, it is more sensitive to the direction of the relationships across modalities, and it is not driven by within-modality variances. On the other hand, standard PLS—which maximizes covariance—is less sensitive to the direction of the across-modality relationships, as it is also driven by within-modality variances. Formally, we can see this from the optimization of these models. Standard CCA optimizes correlation across modalities:

$$\max_{\mathbf{w}_x, \mathbf{w}_y} \text{corr}(\mathbf{X}\mathbf{w}_x, \mathbf{Y}\mathbf{w}_y) \quad (1)$$

Standard PLS optimizes covariance across modalities—the product of correlation and standard deviations (i.e., square root of variance):

$$\max_{\mathbf{w}_x, \mathbf{w}_y} \text{cov}(\mathbf{X}\mathbf{w}_x, \mathbf{Y}\mathbf{w}_y) = \text{corr}(\mathbf{X}\mathbf{w}_x, \mathbf{Y}\mathbf{w}_y) \sqrt{\text{var}(\mathbf{X}\mathbf{w}_x)} \times \sqrt{\text{var}(\mathbf{Y}\mathbf{w}_y)} \quad (2)$$

This also means that standard CCA and PLS are equivalent optimization problems when $\text{var}(\mathbf{X}\mathbf{w}_x) = \text{var}(\mathbf{Y}\mathbf{w}_y) = 1$, which is true when the within-modality variances are identity matrices, i.e., $\mathbf{X}^T\mathbf{X} = \mathbf{Y}^T\mathbf{Y} = \mathbf{I}$.

Limitations of Standard CCA/PLS Models

When the ratio between the sample size and the number of variables is similar to or smaller than 1, standard CCA/PLS models present limitations. These limitations can exist irrespective of sample size if the number of variables is large or if the variables are highly correlated. In the case of standard

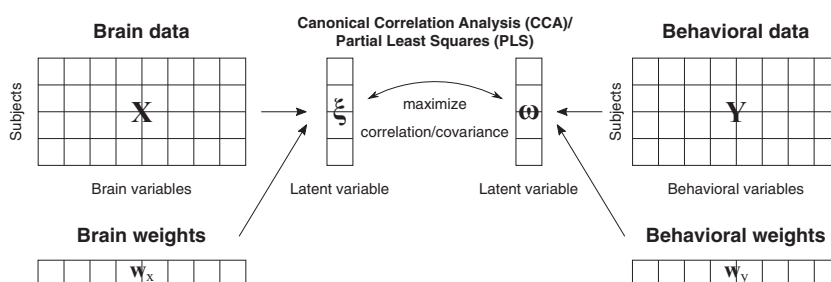


Figure 1. Overview of canonical correlation analysis/partial least squares (CCA/PLS) models for investigating brain–behavior associations. CCA/PLS models maximize the correlation (CCA) or covariance (PLS) between latent variables extracted as weighted linear combinations of the brain and behavioral variables (see formulae in text). Note that the weights are column vectors but are represented as rows to highlight that they have the same dimensionality as their respective data modality.

Table 1. Different Nomenclatures in CCA and PLS Literature and Summary of the Corresponding Terms

Model	Relationship	Model Weights	Latent Variable	Correlation Between Original Variables and Latent Variable
CCA	Mode/association	Canonical vector/coefficient	Canonical variable/ variate	Structure correlation
PLS	Association	Saliency	Score	Loading

CCA, canonical correlation analysis; PLS, partial least squares.

CCA, the key limitations are that 1) the optimization is ill-posed (i.e., there is no unique solution) when the number of variables in at least one of the modalities exceeds the sample size and 2) the CCA weights \mathbf{w}_x and \mathbf{w}_y are unstable when the variables within one or both modalities are highly correlated, known as the multicollinearity problem (37). Not surprisingly, these limitations might sound familiar, as standard CCA can be viewed as a multivariate extension of the univariate general linear model (38,39). The standard PLS optimization is never ill-posed and copes with multicollinearity [i.e., standard PLS weights are stable (36)]; however, standard PLS and CCA cannot perform feature selection (i.e., setting the weights of some variables to zero) and may therefore have low performance in cases in which the effects are sparse.

These limitations can be addressed by dimensionality reduction (i.e., PCA) or regularization. Regularization adds further constraints to the optimization to solve an ill-posed problem or prevent overfitting. For CCA/PLS models, the most common forms of regularization are L1-norm (lasso) (40), L2-norm (ridge) (41), and combinations of L1-norm and L2-norm regularization (elastic-net) (42).

CCA With PCA Dimensionality Reduction

PCA transforms one modality of multivariate data into uncorrelated PCs (it is also related to whitening, see [Effects of Prewhitening on CCA/PLS Models](#)). PCA is often used as a naïve dimensionality reduction technique, as PCs explaining little variance are assumed to be noise and discarded, and the remaining PCs are entered into standard CCA. However, PCA when applied before CCA (PCA-CCA) can be also seen as a technique similar to regularization: It makes the CCA model well posed and addresses the multicollinearity problem.

The number of retained PCs can be selected based on their explained variance, e.g., 99% of total variance. In PCA-CCA applications, often the same number of PCs are chosen for both data modalities, based on the lower-dimensional data, usually behavior [e.g., (9,10,22,24)]. Sometimes, the same proportion of explained variance—rather than numbers of PCs—is used for both data modalities [e.g., (12,26)]. One problem with discarding PCs with low variance is that there is no guarantee that PCs with high variance in either modality are best to link the different data modalities, while some discarded PCs might contain useful information. To address this problem, we can use a data-driven approach, by selecting the number of PCs that maximize the correlation across modalities (see [CCA With PCA Dimensionality Reduction Versus RCCA in High-Dimensional Data](#) in Part 2).

Regularized CCA

L2-norm regularization is a popular form of regularization for ill-posed problems or for mitigating the effects of multicollinearity, originally used in ridge regression (41). In L2-norm regularization, the added constraint corresponds to the sum of squares of all weight values,¹ which forces the weights to be small but does not make them zero. L2-norm regularization has been proposed for CCA (43), commonly referred to as regularized CCA (RCCA) (34,44–46). Interestingly, in RCCA, the regularization terms added to the CCA problem lead to a mixture of standard CCA and standard PLS optimization. We can see this from the RCCA optimization problem:

$$\max_{\mathbf{w}_x, \mathbf{w}_y} \frac{\text{corr}(\mathbf{X}\mathbf{w}_x, \mathbf{Y}\mathbf{w}_y) \sqrt{\text{var}(\mathbf{X}\mathbf{w}_x)} \sqrt{\text{var}(\mathbf{Y}\mathbf{w}_y)}}{\sqrt{(1-c_x)\text{var}(\mathbf{X}\mathbf{w}_x) + c_x} \sqrt{(1-c_y)\text{var}(\mathbf{Y}\mathbf{w}_y) + c_y}} \quad (3)$$

where the 2 hyperparameters (c_x , c_y) control the amount of regularization and provide a smooth transition between standard CCA ($c_x = c_y = 0$, not regularized) and standard PLS ($c_x = c_y = 1$, most regularized) (34,44). Importantly, as L2-norm regularization mitigates multicollinearity, it increases the stability of the RCCA weights. However, it also means that similar to standard PLS, RCCA can be driven by within-modality variances. For additional connections between standard CCA, RCCA, standard PLS, and how they are related to PCA-CCA, see the [Supplement](#).

Sparse PLS

L1-norm regularization was originally proposed in lasso regression (40). In L1-norm regularization, the added constraint corresponds to the absolute sum of weight values,² which sets some of the weight values to zero, resulting in variable selection and promoting sparsity. Sparse solutions facilitate the interpretability of the model and may improve performance when only a subset of variables is relevant (40). However, sparsity can also introduce instability to the model if different sets of variables provide similar performance. Elastic-net regularization is a mixture of L1-norm and L2-norm regularization that combines the properties of both forms of regularization and can mitigate the instability of L1-norm regularization (42). In one popular algorithm (17), which we will refer to as sparse PLS (SPLS), hyperparameters control the amount of L1-norm regularization or sparsity. Because standard PLS can be seen as CCA with maximal L2-norm regularization (see previous section), SPLS can also be viewed as an elastic-net regularized CCA (for details, see the [Supplement](#)).

Effects of Prewhitening on CCA/PLS Models

In machine learning, data are often whitened as a pre-processing step. Whitening transforms the original variables into new, uncorrelated features, which are normalized to have unit length (i.e., the L2-norm of each feature equals 1).

¹L2-norm: $\|\mathbf{w}\|_2 = \sqrt{\sum_i w_i^2}$, where $\mathbf{w} = (w_1, w_2, \dots, w_n)$ is a vector of size n .

²L1-norm: $\|\mathbf{w}\|_1 = \sum_i |w_i|$, where $\mathbf{w} = (w_1, w_2, \dots, w_n)$ is a vector of size n .

Whitening is not a unique transformation, and the most commonly used forms are PCA, Mahalanobis whitening, and Cholesky whitening (47). The critical difference between PCA and PCA whitening is that PCA retains the variance of the original data, i.e., the PCs are not normalized to have unit length.

Whitening as a preprocessing step has a major drawback in CCA/PLS models: The beneficial effects of L1-norm and L2-norm regularization on the original variables cannot be achieved anymore, as the whitened data are the new inputs of the model. In the case of SPLS, L1-norm regularization will result in sparsity on the whitened variables (instead of on the original variables); thus, the interpretability of the results will not be facilitated. In the case of RCCA, L2-norm regularization is not active on whitened data, which means that standard CCA, RCCA, and standard PLS will yield the same results. For additional details on whitening, see the Supplement.

Analytic Frameworks for CCA/PLS Models

The statistical significance of the CCA/PLS model (i.e., the number of significant associative effects) can be evaluated using either a descriptive or a predictive (also referred to as a machine learning) framework. The 2 frameworks have distinct goals: The aim of the descriptive framework is to detect above-chance associations in the current dataset, whereas the aim of the predictive framework is to test whether such associations generalize to new data (48–51).

In the descriptive framework (Figure 2A), the CCA/PLS model is fitted on the entire sample; thus, the statistical inference is based on in-sample correlation. In this framework, there is usually no hyperparameter optimization (i.e., the number of PCs or regularization parameter is fixed a priori). In the predictive framework (Figure 2B), the CCA/PLS model is fitted on a training/optimization set and evaluated on a test/holdout set; thus, the statistical inference is based on out-of-sample correlation. This procedure assesses the generalizability of the model, i.e., how

well the association found in the training set generalizes to an independent test set. In the predictive framework, the hyperparameters are usually optimized; therefore, the training/optimization set is further divided into a training set and a validation set, and the best hyperparameters are selected based on out-of-sample correlation in the validation set. In both descriptive and predictive frameworks, permutation inference (based on in-sample or out-of-sample correlation) is often used to assess the number of significant associative effects (51,52).

Last, an important component of any CCA/PLS framework is testing the stability of the model. Usually a bootstrapping procedure is applied to provide confidence intervals on the model's weights (51). Recently, stability selection (19,20,53–55) has been proposed with the aim of selecting the most stable CCA/PLS model in the first place, rather than evaluating the stability of the model post hoc. Alternatively, the stability of the CCA/PLS models can be measured as the average similarity of weights across different splits of training data, which avoids the additional computational costs of the previous 2 approaches (27). For more details on analytic frameworks, see for example (22,27,51,56).

PART 2: DEMONSTRATIONS OF CCA AND PLS ANALYSES

Description of Experiments

In order to demonstrate the properties of different CCA and PLS approaches, we applied the models introduced in Part 1 to real and simulated datasets with different dimensionalities and sample sizes. Table 2 gives an overview of all experiments.

We chose the HCP and the ADNI datasets based on 2 recent landmark studies (22,56). In the HCP dataset, we used resting-state functional magnetic resonance imaging connectivity data (19,900 and 300 brain variables in the high- and low-dimensional data, respectively) and 145 nonimaging subject measures (e.g., behavioral, demographic, lifestyle measures) of

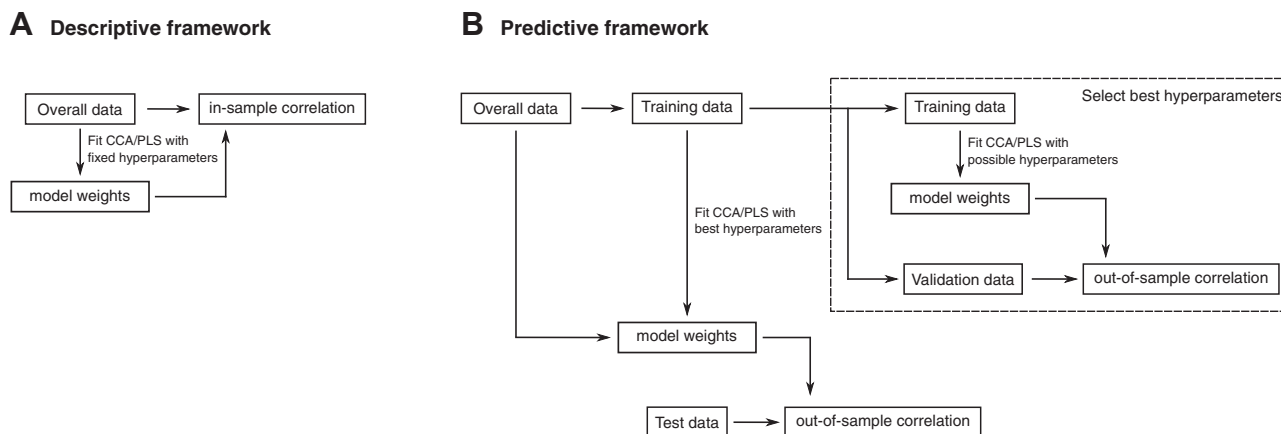


Figure 2. Descriptive and predictive (or machine learning) frameworks. **(A)** The descriptive framework fits canonical correlation analysis/partial least squares (CCA/PLS) model with fixed hyperparameters (i.e., the number of principal components or regularization parameter) on the entire sample; thus, the statistical inference is based on in-sample correlation. **(B)** The predictive (or machine learning) framework fits CCA/PLS model on a training set and evaluates the model on a test set; thus, the statistical inference is based on out-of-sample correlation. The hyperparameters are usually optimized: The training set is further divided into a training set and a validation set, and the best hyperparameters are selected based on out-of-sample correlation in the validation set. We note that although not all models maximize correlation, typically all CCA/PLS models are evaluated based on the correlation between the latent variables (see Figure 1).

Table 2. Summary of CCA/PLS Models on High- and Low-Dimensional Real and Simulated Data

Model	Analytical Framework	Hyperparameter Optimization	Model Hyperparameter
High-Dimensional Data			
PCA-CCA	Descriptive	None (fixed)	Number of PCs
PCA-CCA	Predictive	None (fixed)	Number of PCs
PCA-CCA	Predictive	Data-driven	Number of PCs
RCCA	Predictive	Data-driven	Amount of L2-norm regularization
Standard PLS	Predictive	None	None
SPLS	Predictive	Data-driven	Amount of L1-norm regularization
Low-Dimensional Data			
Standard CCA	Predictive	None	None
RCCA	Predictive	Data-driven	Amount of L2-norm regularization
Standard PLS	Predictive	None	None
SPLS	Predictive	Data-driven	Amount of L1-norm regularization

CCA, canonical correlation analysis; PC, principal component; PCA, principal component analysis; PLS, partial least squares; RCCA, regularized canonical correlation analysis; SPLS, sparse partial least squares.

1003 healthy subjects. In the ADNI dataset, we used whole-brain gray matter volumes (168,130 and 120 brain variables in the high- and low-dimensional data, respectively) and 31 item-level measures of the Mini-Mental State Examination of 592 elderly subjects. We generated the simulated data with a sparse signal (i.e., 10% of the variables in each modality were relevant to capture the association across modalities) and properties similar to the HCP dataset (in terms of sample size, dimensionality, and correlation between latent variables). [Table 3](#) displays the characteristics of the real and simulated

datasets. For further details of the datasets and the simulated data generation, see the [Supplement](#).

The PCA-CCA model was used both with fixed numbers of PCs within a descriptive framework and with optimized number of PCs within a predictive framework. All the other CCA/PLS models were used within a predictive framework. The predictive framework was based on Monteiro *et al.* (56), who used multiple test/holdout sets to assess the generalizability and robustness of the CCA/PLS models (detailed in the [Supplement](#)). In both frameworks, permutation testing was used to assess the number of statistically significant associative effects based on in-sample and out-of-sample correlations between the latent variables, respectively. Importantly, the family structure of the HCP dataset was respected during the different data splits (training, validation, test/holdout sets) and permutations (57). We used iterative methods to solve the CCA/PLS model and applied mode-A deflation for standard PLS and SPLS and generalized deflation for standard CCA, PCA-CCA, and RCCA (for details, see the [Supplement](#)). For simplicity, we present the results for the first associative effect in most CCA/PLS experiments (for a summary of all associative effects, see [Table S1](#)). Throughout the article, we present the weights (canonical vector for CCA models, salience for PLS models) and latent variables obtained by the model.

We used linear mixed-effects models to compare the different CCA/PLS models on the following measures across the outer training or test sets: 1) in-sample correlation, 2) out-of-sample correlation, 3) similarity of the model weights (measured by Pearson's correlation), and 4) variance explained by the model. In addition, we compared the number of PCs between PCA-CCA models with fixed versus data-driven numbers of PCs. We report significance at $p < .005$ in all linear mixed-effects models. For further details of the linear mixed-effects analyses, see the [Supplement](#). We also quantified the rank similarity of the weights (measured by Spearman's correlation) across the different CCA/PLS models in the real datasets.

Table 3. Characteristics of Real and Simulated Datasets

Data	HCP		ADNI		Simulation	
	Low Dimensional	High Dimensional	Low Dimensional	High Dimensional	Low Dimensional	High Dimensional
Subjects	Healthy ($n = 1001$)	Healthy ($n = 1001$)	Healthy + clinical ($n = 592$)	Healthy + clinical ($n = 592$)	Not applicable ($n = 1000$)	Not applicable ($n = 1000$)
Brain Variables	Connectivity of 25 ICA ^a components ($d = 300$)	Connectivity of 200 ICA ^a components ($d = 19900$)	ROI-wise ^b gray matter volume ($d = 120$)	Voxelwise gray matter volume ($d = 168130$)	Not applicable ($d = 100$)	Not applicable ($d = 20000$)
Behavioral Variables	Behavior, psychometrics, demographics ($d = 145$)	Behavior, psychometrics, demographics ($d = 145$)	Items of MMSE ^c questionnaire ($d = 31$)	Items of MMSE ^c questionnaire ($d = 31$)	Not applicable ($d = 100$)	Not applicable ($d = 100$)

ADNI, Alzheimer's Disease Neuroimaging Initiative; d, number of variables; HCP, Human Connectome Project; ICA, independent component analysis; ROI, region of interest; MMSE, Mini-Mental State Examination.

^aData-driven brain parcellation.

^bBrain parcellation using the Automated Anatomical Labeling 2 atlas (62).

^cScreening questionnaire for dementia (63).

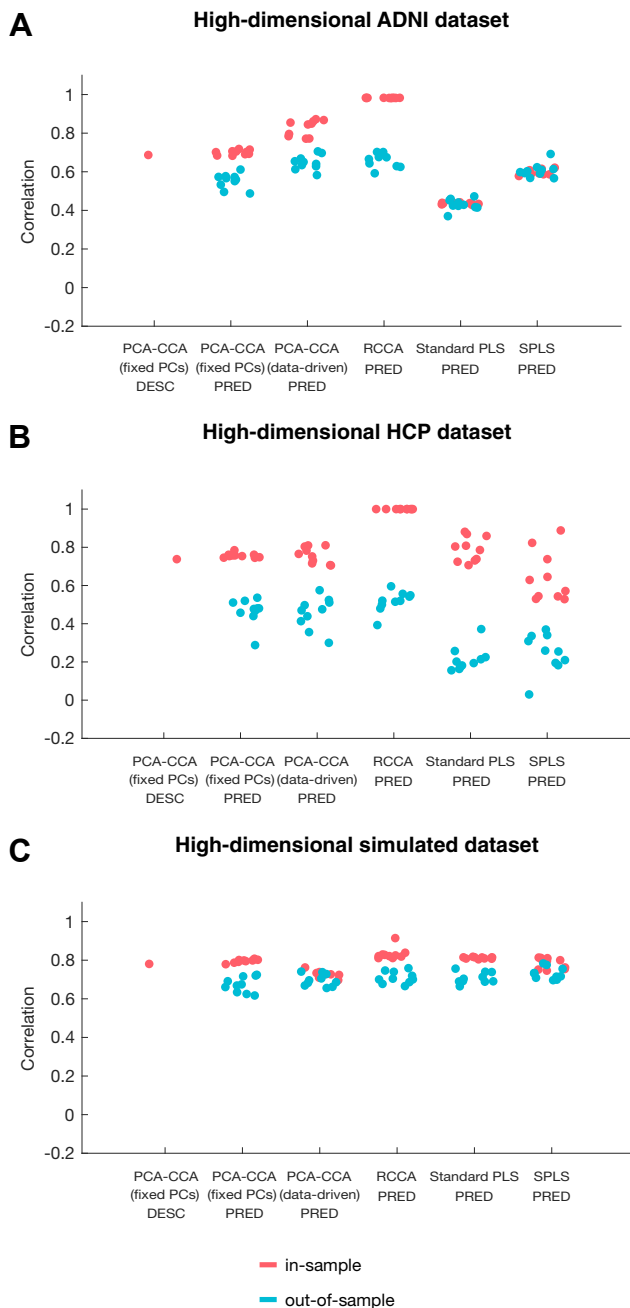


Figure 3. Dot plot of in-sample and out-of-sample correlations for the first associative effects of all experiments in all 3 high-dimensional datasets. Each dot represents a model trained on the overall data (descriptive framework) or on 10 random subsets of the data (predictive framework). The horizontal jitter is for visualization purposes. **(A)** High-dimensional Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset. **(B)** High-dimensional Human Connectome Project (HCP) dataset. Note that we display the second associative effect for standard partial least squares (PLS) and sparse PLS (SPLS), as it is the most similar to the first associative effects identified by the other models. **(C)** High-dimensional simulated dataset. CCA, canonical correlation analysis; data-driven, data-driven number of principal components; DESC, descriptive framework; PC, principal component; PCA, principal component analysis; PRED, predictive framework; RCCA, regularized canonical correlation analysis.

In-Sample Versus Out-of-Sample Correlation in High-Dimensional Data

Figure 3 and Table 4 display the in-sample and out-of-sample correlations for all experiments using all 3 high-dimensional datasets. On average the out-of-sample correlations are lower than the in-sample correlations ($t_{14} = 4.51, p = .0005$). In real datasets, CCA/PLS models with dimensionality reduction or regularization provide high out-of-sample correlations in most cases, underlining that these models generalize well to unseen data. The only notable exceptions are standard PLS and SPLS, which presented significantly lower out-of-sample correlations in the HCP dataset ($F_{2,56} = 289.30, p < .0001$) (Figure 3B). This can be attributed to the different properties of the HCP dataset (e.g., higher noise level and nonsparse associative effect) and the fact that standard PLS and SPLS are especially dominated by within-modality variance in this dataset (Table 4).

In conclusion, we recommend embedding all models in a predictive framework that splits the data into training and test sets to assess the model’s out-of-sample generalizability.

CCA With PCA Dimensionality Reduction Versus RCCA in High-Dimensional Data

In this section, we present the results of applying PCA-CCA and RCCA to all 3 high-dimensional datasets. We focus on experiments using the predictive framework, compare PCA-CCA with fixed versus data-driven numbers of PCs, and compare both of these models with RCCA.

Figures 4A–C and 5A–C display the brain and behavioral weights and corresponding latent variables for the 3 models (note that for the HCP dataset, the brain weights were transformed into brain connection strength increases/decreases). Figure 6 compares the brain and behavioral weights using rank similarity across the models, which indicates that although the weights are similar across the 3 models, data-driven PCA-CCA and RCCA are more similar to each other. The model weights and latent variables for the simulated dataset can be found in Figures 7A–C, which suggest that all 3 models recovered sufficiently the true weights of the generative model. Nevertheless, the nonsparse models attributed nonzero weights for many nonrelevant variables (for details, see Table S2).

To further investigate the characteristics of the 3 models, Table 4 shows the stability of weights and the explained variance by the models. The stability of weights varied significantly across brain and behavior modalities ($F_{1,804} = 84.51, p < .0001$) and models ($F_{2,804} = 91.63, p < .0001$). Notably, the stability of RCCA weights was consistently high. The explained variance varied significantly only across modalities ($F_{1,174} = 241.55, p < .0001$) but not across models ($F_{2,174} = 0.31, p = .7303$).

Next, we examined the number of PCs in the 2 PCA-CCA models. We found a significant interaction between the effect of data modality and model on the number of PCs ($F_{1,114} = 22.63, p < .0001$). Data-driven PCA-CCA yielded more brain PCs and fewer behavioral PCs than PCA-CCA with fixed number of PCs (Table S3). These results confirm that lower-ranked brain PCs might also carry information that links brain and behavior and should

Table 4. Main Characteristics of the First Associative Effects in the High-Dimensional Datasets Obtained With the Different CCA/PLS Models Using the Predictive Framework

Model	Brain		Behavior		Across-Modality Relationship	
	Stability of Weights ^a	Explained Variance ^b	Stability of Weights ^a	Explained Variance ^a	In-Sample Correlation ^c	Out-of-Sample Correlation ^d
ADNI Dataset						
PCA-CCA (Fixed PCs)	0.86 ± 0.00	8.47 ± 0.16	0.85 ± 0.01	14.91 ± 0.23	0.70 ± 0.00	0.55 ± 0.01
PCA-CCA (Data-Driven PCs)	0.70 ± 0.01	5.26 ± 0.25	0.93 ± 0.00	15.73 ± 0.13	0.83 ± 0.01	0.65 ± 0.01
RCCA (L2-Reg. Opt.)	0.82 ± 0.00	5.47 ± 0.06	0.94 ± 0.00	16.63 ± 0.26	0.98 ± 0.00	0.66 ± 0.01
Standard PLS	0.96 ± 0.00	21.54 ± 0.16	0.94 ± 0.00	18.64 ± 0.21	0.44 ± 0.00	0.43 ± 0.01
SPLS (L1-Reg. Opt.)	0.83 ± 0.02	14.05 ± 0.13	0.96 ± 0.01	15.86 ± 0.42	0.60 ± 0.00	0.61 ± 0.01
HCP Dataset						
PCA-CCA (Fixed PCs)	0.72 ± 0.01	0.42 ± 0.01	0.78 ± 0.01	2.67 ± 0.10	0.76 ± 0.00	0.47 ± 0.02
PCA-CCA (Data-Driven PCs)	0.56 ± 0.02	0.35 ± 0.03	0.53 ± 0.04	3.73 ± 0.39	0.76 ± 0.01	0.45 ± 0.03
RCCA (L2-Reg. Opt.)	0.78 ± 0.01	0.29 ± 0.01	0.88 ± 0.01	4.39 ± 0.18	1.00 ± 0.00	0.52 ± 0.02
Standard PLS-2	0.52 ± 0.04	0.50 ± 0.05	0.62 ± 0.05	8.07 ± 0.30	0.79 ± 0.02	0.21 ± 0.02
SPLS-2 (L1-Reg. Opt.)	0.25 ± 0.04	0.48 ± 0.07	0.51 ± 0.05	7.23 ± 0.37	0.64 ± 0.04	0.25 ± 0.03
Simulated Dataset						
PCA-CCA (Fixed PCs)	0.74 ± 0.01	0.76 ± 0.01	0.90 ± 0.00	1.82 ± 0.01	0.80 ± 0.00	0.67 ± 0.01
PCA-CCA (Data-Driven PCs)	0.96 ± 0.00	0.85 ± 0.00	0.91 ± 0.00	1.95 ± 0.02	0.73 ± 0.01	0.70 ± 0.01
RCCA (L2-Reg. Opt.)	0.93 ± 0.00	0.77 ± 0.00	0.97 ± 0.00	1.99 ± 0.01	0.83 ± 0.01	0.71 ± 0.01
Standard PLS	0.94 ± 0.00	0.84 ± 0.00	0.97 ± 0.00	2.07 ± 0.01	0.81 ± 0.00	0.71 ± 0.01
SPLS (L1-Reg. Opt.)	0.78 ± 0.03	0.84 ± 0.00	1.00 ± 0.00	1.94 ± 0.01	0.79 ± 0.01	0.73 ± 0.01

Values are mean ± SEM. Note that we display the second associative effect for standard PLS (PLS-2) and SPLS (SPLS-2) in the HCP dataset, as it is the most similar to the first associative effects identified by the other models.

ADNI, Alzheimer's Disease Neuroimaging Initiative; CCA, canonical correlation analysis; HCP, Human Connectome Project; L1-reg., L1-norm regularization; L2-reg., L2-norm regularization; opt., optimized; PC, principal component; PCA, principal component analysis; PLS, partial least squares; RCCA, regularized canonical correlation analysis; SPLS, sparse partial least squares;

^aSimilarity of model weights measured by Pearson's correlation between each pair of training sets of the outer data splits.

^bPercent variance explained by the model relative to all within-modality variance in the training sets of the outer data splits.

^cCorrelation between the latent variables in the training sets of the outer data splits.

^dCorrelation between the latent variables in the test sets of the outer data splits.

not necessarily be discarded. Moreover, fixing the same number of PCs for both modalities might not be a good choice.

Based on these results, and as the optimal numbers of PCs can vary even across different brain-behavior associations in the same dataset, we recommend data-driven PCA-CCA over PCA-CCA with fixed numbers of PCs. Furthermore, we found that data-driven PCA-CCA and RCCA gave similar results, both having a similar regularizing effect on the CCA model.

Sparse Versus Nonsparse CCA/PLS Models in High-Dimensional Data

In this section, we show how SPLS found associations between subsets of features in all 3 high-dimensional datasets, and we compare the SPLS results with standard PLS and RCCA.

Figures 4C–E and 5C–E display the models' weights and latent variables (note that for the HCP dataset, the brain weights were transformed into brain connection strength

increases/decreases). The first associative effect found by standard PLS and SPLS was similar to the first found by RCCA in both the ADNI and simulated datasets, but in the HCP dataset, the first associative effect identified by RCCA was more similar to the second effect found by standard PLS and SPLS (Figure 6). This is likely because the within-modality variances in the HCP dataset differ substantially from the identity matrix, and therefore the difference between the objectives of CCA and PLS models is more pronounced (see equations 1 and 2). The brain and behavioral weights were similar across the 3 models in both real datasets, especially the top-ranked variables (i.e., the variables with the highest weights). Similar to RCCA, standard PLS and SPLS recovered sufficiently the true weights of the generative model; however, the SPLS model assigned fewer nonzero weights to nonrelevant variables (Figure 7C–E). These results demonstrate that when the signal is sparse, SPLS can lead to high true positive and high true negative rates of weight recovery (Table S2). Table S4 shows the sparsity of the associative effects identified by SPLS.

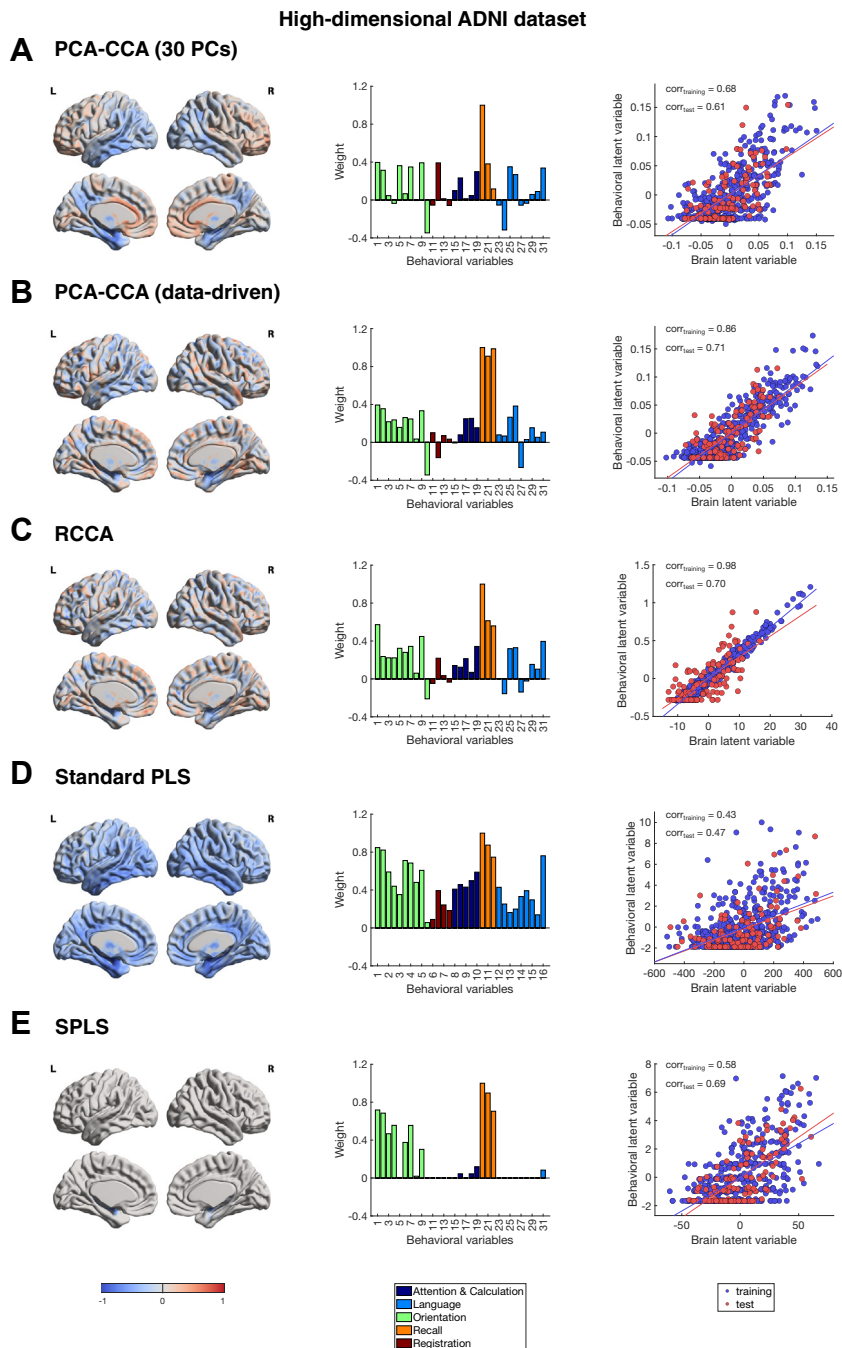


Figure 4. Brain weights (left), behavioral weights (middle), and latent variables (right) for the high-dimensional Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset. For visualization purposes, the model weights are normalized (divided by largest absolute value). The scatter plot between the brain and behavioral latent variables is overlaid by a least-squares regression line separately for the training and test data. **(A)** Principal component analysis–canonical correlation analysis (PCA-CCA) with fixed number of principal components (PCs). **(B)** PCA-CCA with data-driven number of PCs. **(C)** Regularized CCA (RCCA). **(D)** Standard partial least squares (PLS). **(E)** Sparse PLS (SPLS). $\text{corr}_{\text{test}}$, out-of-sample correlation in the test data; $\text{corr}_{\text{training}}$, in-sample correlation in the training data; L, left hemisphere; R, right hemisphere.

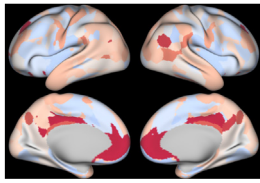
The stability of the weights differed significantly between the brain and behavioral modalities ($F_{1,804} = 75.26, p < .0001$) and the 3 models ($F_{2,804} = 61.77, p < .0001$) (Table 4). The stability of the SPLS weights was lowest in the HCP dataset, which is likely due to the model’s sparsity and that different sets of variables might provide similar performance. The instability of SPLS could be mitigated by stability selection (20) or a stability criterion during

hyperparameter optimization (27). The explained variance varied significantly across modalities ($F_{1,174} = 80.00, p < .0001$) and the 3 models ($F_{2,174} = 28.60, p < .0001$).

In summary, while RCCA is likely to yield similar or higher out-of-sample correlations than standard PLS and SPLS, SPLS can perform variable selection and may improve the interpretability of the results; however, it can also present instabilities. In practice, the 3 models

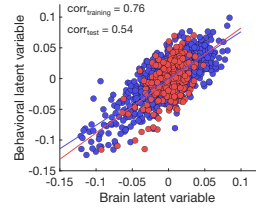
High-dimensional HCP dataset

A PCA-CCA (100 PCs)

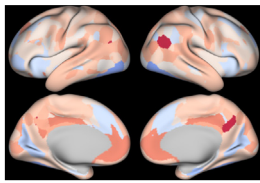


0.83 ↑ **OSM Hyperactivity Problems**
Global feelings and attitudes about one's life
Picture Vocabulary Test
Delay Discounting (AUC for Discounting of \$40K)
Fluid intelligence
ASR Other Problems
 Smoking history
Intrusive
 Avg total weekday **alcoholic** drinks/day in past 7 days
 DSM **Avoidant Personality Problems**
 Non-diagnostic screen of **agoraphobia**
 Frequency of drinking 5+ **drinks** in past 12 months
Father had depression
 Number of symptoms of DSM4 **Alcohol Abuse** over lifetime

-0.39 ↓ **Total score of Pittsburgh Sleep Quality Index**
Attention Problems
Penn Line Orientation (Median Reaction Time)
Total alcoholic drinks in past 7 days
 Any positive test for THC (**cannabis**)
 DSM4 Major **Depressive Episode** over lifetime
Externalizing
Aggressive Behavior
Aggression
Life matters or makes sense
 Perception of **stress**
Conscientiousness Scale Score
 Perceptions of **loneliness**
 Number days smoked/used ANY **TOBACCO** in past 7 days

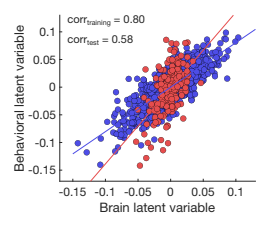


B PCA-CCA (data-driven)

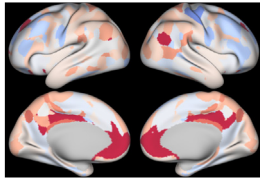


1.00 ↑ **Reading Test**
Picture Vocabulary Test
Fluid intelligence
Penn Line Orientation (Total Number Correct)
List sorting
 Openness Scale Score
Picture Sequence Memory Test
Delay Discounting (AUC for Discounting of \$200)
Delay Discounting (AUC for Discounting of \$40K)
Penn Word Memory (Total Number of Correct Responses)
Manual dexterity
 Total **alcoholic** drinks in past 7 days
 Number days drank **alcohol** in past 7 days
Global feelings and attitudes about one's life
Short Penn CPT Specificity

-0.20 ↓ **Perception of stress**
Hostility and cynicism
Drinks consumed per drinking day in past 12 months
 Times used/smoked ANY **TOBACCO** TODAY
 Avg weekday **CIGARETTES** per day in past 7 days
 Avg weekend **CIGARETTES** per day in past 7 days
 Any positive test for THC (**cannabis**)
 Total **CIGARETTES** in past 7 days
 Avg total weekday ANY **TOBACCO** per day in past 7 days
 Avg total weekend ANY **TOBACCO** per day in past 7 days
 Total times used/smoked ANY **TOBACCO** in past 7 days
 Perceptions of **rejection** in daily social interactions
 Participant still **smoking**
 Number days smoked/used ANY **TOBACCO** in past 7 days

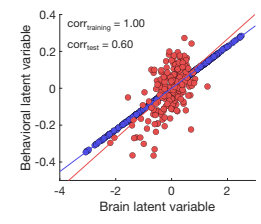


C RCCA

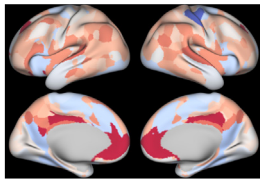


1.00 ↑ **Picture Vocabulary Test**
Global feelings and attitudes about one's life
Delay Discounting (AUC for Discounting of \$40K)
Delay Discounting (AUC for Discounting of \$200)
Fluid intelligence
 DSM **Hyperactivity Problems**
Father had depression
Penn Emotion Recognition (Correct Anger Identifications)
Intrusive
Mother had depression
Pattern Comparison Processing Test
 DSM **Avoidant Personality Problems**
 Avg total weekday **alcoholic** drinks/day in past 7 days
Penn Line Orientation (Total Number Correct)
List sorting

-0.31 ↓ **Short Penn CPT** Sensitivity
Conscientiousness Scale Score
Withdrawn
 Avg total weekend **alcoholic** drinks/day in past 7 days
Penn Line Orientation (Total Positions Off for All Trials)
Times used marijuana
 Total **alcoholic** drinks in past 7 days
 Number of Childhood **Conduct problems**
 Number days smoked/used ANY **TOBACCO** in past 7 days
Penn Emotion Recognition (Correct Neutral Identifications)
 Age at **first alcohol** use
 Any positive test for THC (**cannabis**)
Thought Problems
Aggression
Total score of Pittsburgh Sleep Quality Index

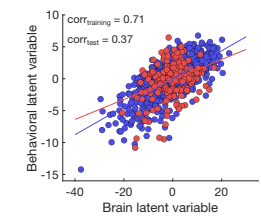


D Standard PLS

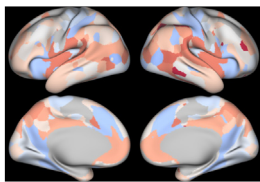


1.00 ↑ **Picture Vocabulary Test**
Reading Test
Fluid intelligence
Penn Line Orientation (Total Number Correct)
Delay Discounting (AUC for Discounting of \$40K)
Dimensional Change Card Sort Test
Manual dexterity
Flanker test
List sorting
Delay Discounting (AUC for Discounting of \$200)
Pattern Comparison Processing Test
Picture Sequence Memory Test
 Total **alcoholic** drinks in past 7 days
Mini Mental Status Exam Total Score
Penn Emotion Recognition (Correct Responses)

-0.38 ↓ **Frequency of any alcohol use in heaviest 12-month drinking period**
Aggression
Frequency drunk in heaviest 12-month drinking period
Frequency of any alcohol use in past 12 months
 Participant still **smoking**
 Times used/smoked ANY **TOBACCO** TODAY
 Any positive test for THC (**cannabis**)
 Avg weekend **CIGARETTES** per day in past 7 days
 Avg weekday **CIGARETTES** per day in past 7 days
 Total **CIGARETTES** in past 7 days
 Avg total weekday ANY **TOBACCO** per day in past 7 days
 Avg total weekend ANY **TOBACCO** per day in past 7 days
Penn Line Orientation (Total Positions Off for All Trials)
 Total times used/smoked ANY **TOBACCO** in past 7 days



E SPLS



1.00 ↑ **Reading Test**
Picture Vocabulary Test
Dimensional Change Card Sort Test
Fluid intelligence
Penn Line Orientation (Total Number Correct)
Delay Discounting (AUC for Discounting of \$40K)
List sorting
 Number days drank **alcohol** in past 7 days
Flanker test
Mini Mental Status Exam Total Score
Picture Sequence Memory Test
Pattern Comparison Processing Test
 Total **drinks** in past 7 days
Manual dexterity
Delay Discounting (AUC for Discounting of \$200)

-0.32 ↓ **Avg total weekday ANY TOBACCO** per day in past 7 days
 Total times used/smoked ANY **TOBACCO** in past 7 days
 Avg total weekend ANY **TOBACCO** per day in past 7 days
 Any positive test for THC (**cannabis**)
Aggression
Frequency of drinking 5+ drinks in past 12 months
 Number days smoked/used ANY **TOBACCO** in past 7 days
Frequency of drinking 5+ drinks during heaviest 12-month drinking period
Frequency drunk in past 12 months
 Electronic **Visual Acuity** Denominator
 Age at **first alcohol** use
Frequency of any alcohol use in past 12 months
Frequency of any alcohol use in heaviest 12-month drinking period
Penn Line Orientation (Total Positions Off for All Trials)
Frequency drunk in heaviest 12-month drinking period

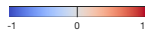
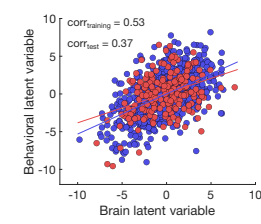


Figure 5. Brain connection strengths (left), behavioral weights (text in the middle), and latent variables (right) for the high-dimensional Human Connectome Project (HCP) dataset. For visualization purposes, the brain weights were transformed into brain connection strength (i.e., brain weights multiplied by the sign of the population mean connectivity) increases (red) and decreases (blue), summed across the brain nodes (i.e., independent component analysis [ICA] components in which each brain vertex is assigned to an ICA component it is most likely to belong and normalized (divided by largest absolute value). Only the top 15 positive (red) and top 15 negative (blue) behavioral weights are shown (secondary [e.g., age adjusted] measures that are highly redundant with those

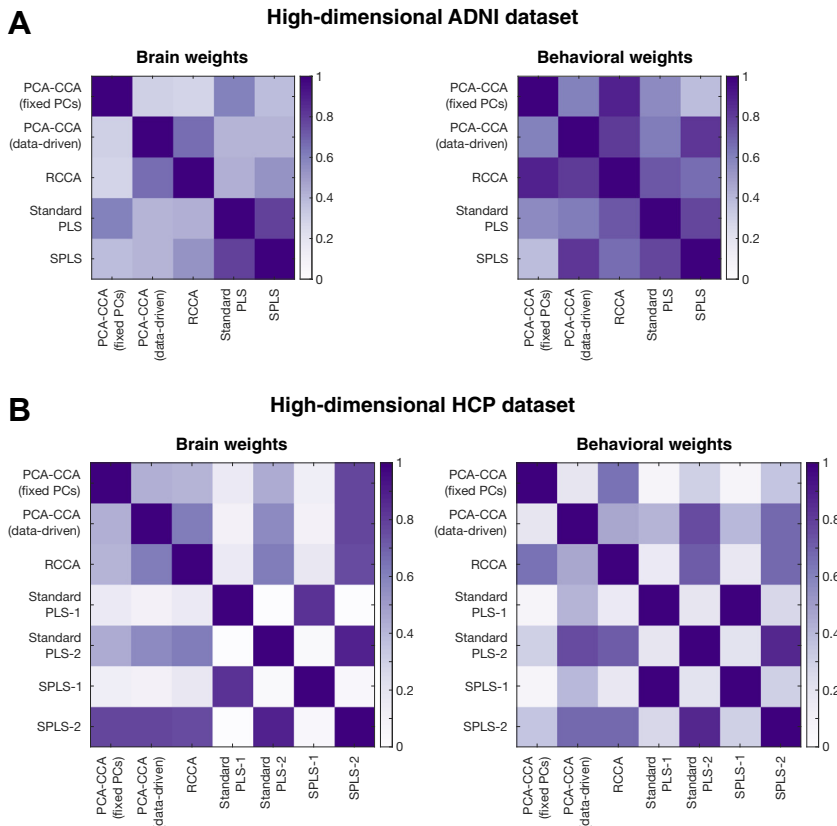


Figure 6. Comparison of brain weights (left) and behavioral weights (right) across canonical correlation analysis/partial least squares (CCA/PLS) models for the high-dimensional Alzheimer's Disease Neuroimaging Initiative (ADNI) and Human Connectome Project (HCP) datasets obtained by the predictive framework. The similarity between the model weights was measured by Spearman's correlation. The similarity between sparse PLS (SPLS) and the other models was measured only for the subset of variables identified by SPLS (the similarity between the 2 SPLS models was measured for the subset of variables that were present in both models). **(A)** High-dimensional ADNI dataset. **(B)** High-dimensional HCP dataset. Note that the second associative effect identified by standard PLS (PLS-2) and SPLS (SPLS-2) was similar to the first associative effects identified by the other models. PC, principal component; PCA, principal component analysis; RCCA, regularized canonical correlation analysis.

often provide similar weights for the top-ranked variables.

Standard Versus Regularized Extension of CCA/PLS Models in Low-Dimensional Data

To investigate the effects of regularization in all 3 low-dimensional datasets, we compared standard CCA, RCCA, standard PLS, and SPLS. The regularized models (RCCA, SPLS) were more stable ($F_{3,1075} = 80.54, p < .0001$) (Table S5) and showed a trend toward higher out-of-sample correlations ($F_{1,10} = 3.35, p = .0972$) (Figure S1) than their nonregularized variants (standard CCA and PLS). The stability of standard PLS and RCCA weights was consistently high, the stability of SPLS varied across datasets, and standard CCA was rather unstable (Table S5). SPLS provided sparse results, similar to the high-dimensional datasets (Table S4). As expected, RCCA and standard PLS explained increasingly more within-modality variance than standard CCA. For a detailed description of

these results, see the Supplement. Taken together, these results suggest that RCCA/SPLS models should be preferred even for low-dimensional data.

Conclusions

This tutorial compared standard and regularized extensions of CCA and PLS models and highlighted the benefits of regularization. Here, we outline the key lessons.

First, we showed that regularized extensions of CCA/PLS models give similar out-of-sample correlations in large datasets (with the exception of standard PLS and SPLS in the high-dimensional HCP dataset) when the sample size is similar to or much smaller than the number of variables (i.e., when the ratio between examples and variables is $\sim 1-10$ or $\sim 0.1-0.01$). Importantly, RCCA and SPLS outperformed standard CCA and PLS even when the ratio between examples and variables was ~ 1 to 10. Second, we emphasized that it is important to use a predictive framework, as high in-sample correlations do not necessarily imply generalizability to unseen data.

shown here are not displayed). The behavioral model weights are normalized (divided by largest absolute value). The scatter plot between the brain and behavioral latent variables is overlaid by a least-squares regression line separately for the training and test data. **(A)** Principal component analysis–canonical correlation analysis (PCA-CCA) with fixed number of principal components (PCs). **(B)** PCA-CCA with data-driven number of PCs. **(C)** Regularized CCA (RCCA). **(D)** Standard partial least squares (PLS). **(E)** Sparse PLS (SPLS). ASR, Achenbach Adult Self Report; AUC, area under curve; $corr_{test}$, out-of-sample correlation in the test data; $corr_{training}$, in-sample correlation in the training data; CPT, continuous performance test; L, left hemisphere; R, right hemisphere; THC, Δ^9 -tetrahydrocannabinol.

High-dimensional simulated dataset

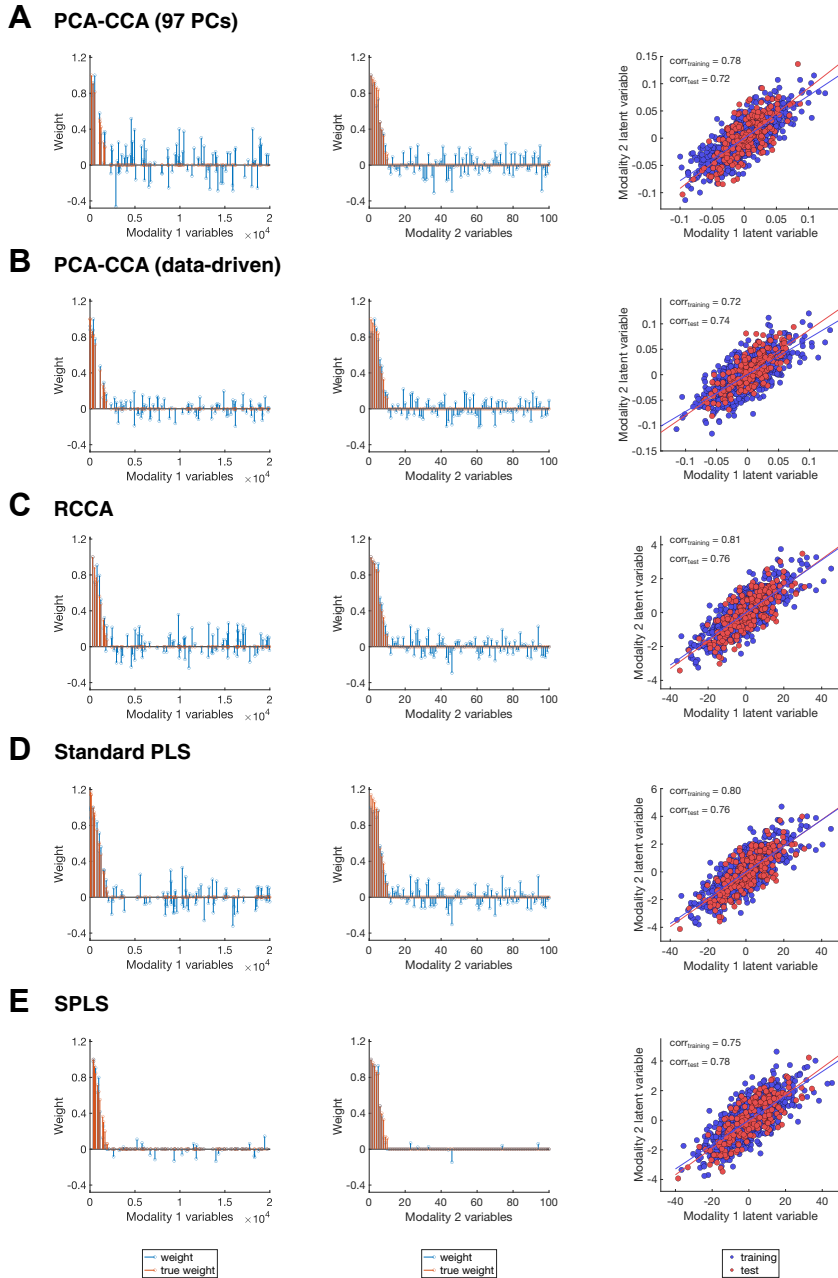


Figure 7. Model weights (left: high-dimensional modality; middle: low-dimensional modality) and latent variables (right) for the high-dimensional simulated dataset. For comparison, the true weights (red) of the generative model are overlaid on the model weights (blue). For visualization purposes, the model weights are normalized (divided by largest value), and only a subset of 100 random weights (out of the total 20,000) is displayed for the high-dimensional modality. The scatter plot between the brain and behavioral latent variables is overlaid by a least-squares regression line separately for the training and test data. **(A)** Principal component analysis–canonical correlation analysis (PCA-CCA) with fixed number of principal components (PCs). **(B)** PCA-CCA with data-driven number of PCs. **(C)** Regularized CCA (RCCA). **(D)** Standard partial least squares (PLS). **(E)** Sparse PLS (SPLS). $corr_{test}$, out-of-sample correlation in the test data; $corr_{training}$, in-sample correlation in the training data.

Going beyond model performance, we demonstrated both in theory and in practice that standard CCA is prone to instability (Table S3). L2-norm regularization improves stability, which comes at a cost of the models (RCCA, standard PLS, SPLS) being driven by within-modality variances. PCA-CCA with data-driven selection of PCs improves on a priori selection. Data-driven PCA-CCA has a comparable regularizing effect to RCCA. Sparsity (i.e., L1-norm regularization) can facilitate the interpretability and the generalizability of the

models, but it can also introduce instability. Sparsity is most useful when the associative effect itself is sparse (e.g., in the ADNI and simulated datasets). Data-driven PCA-CCA, RCCA, and SPLS yielded similar model weights and accounted for similar variances.

We hope that this work, together with recent efforts [e.g., (26,27,30,31,52)] and critical exchanges [e.g., (28,58–61)], illuminates these complex methods and facilitates their application to the brain and its disorders.

ACKNOWLEDGMENTS AND DISCLOSURES

AM was funded by the Wellcome Trust (Grant No. WT102845/Z/13/Z) and by MQ: Transforming Mental Health (Grant No. MQF17_24). JC was supported by the Engineering and Physical Sciences Research Council-funded University College London Centre for Doctoral Training in Intelligent, Integrated Imaging in Healthcare (Grant No. EP/S021930/1) and the Department of Health's National Institute for Health and Care Research-funded Biomedical Research Centre at University College London Hospitals. RAA was supported by a Medical Research Council Skills Development Fellowship (Grant No. MR/S007806/1). NRW was supported by grants from the German Research Foundation (Grant Nos. HA7070/2-2, HA7070/3, and HA7070/4). FSF was funded by a PhD scholarship awarded by Fundação para a Ciência e a Tecnologia (Grant No. SFRH/BD/120640/2016). JM-M was funded by the Wellcome Trust (Grant No. WT102845/Z/13/Z).

A complete listing of Alzheimer's Disease Neuroimaging Initiative (ADNI) investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Some data used in preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. The code used for the different CCA/PLS analyses is implemented in a CCA/PLS toolkit that is available at http://www.mnrl.cs.ucl.ac.uk/resources/cca_pls_toolkit together with a demo demonstrating how to use the toolkit for generating the SPLS results for the low-dimensional simulated dataset. The CCA/PLS toolkit is also available on Zenodo (<https://doi.org/10.5281/zenodo.7153571>).

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Centre for Medical Image Computing (AM, JC, RAA, FSF, JM-M), Department of Computer Science, University College London; and Max Planck University College London Centre for Computational Psychiatry and Ageing Research (AM, JC, RAA, FSF, JM-M), Wellcome Centre for Human Neuroimaging (RAA), and Department of Computer Science (JS-T), University College London, London; and Department of Psychiatry (AM), University of Cambridge, Cambridge, United Kingdom; and the Institute of Translational Psychiatry (NRW), University of Münster, Münster, Germany.

Address correspondence to Agoston Mihalik, Ph.D., at am3022@cam.ac.uk.

Received Jun 17, 2021; revised Jun 30, 2022; accepted Jul 22, 2022.

Supplementary material cited in this article is available online at <https://doi.org/10.1016/j.bpsc.2022.07.012>.

REFERENCES

- Smith SM, Nichols TE (2018): Statistical challenges in "big data" human neuroimaging. *Neuron* 97:263–268.
- Bzdok D, Yeo BTT (2017): Inference in the age of big data: Future perspectives on neuroscience. *Neuroimage* 155:549–564.
- Bzdok D, Nichols TE, Smith SM (2019): Towards algorithmic analytics for large-scale datasets. *Nat Mach Intell* 1:296–306.
- Hotelling H (1936): Relations between two sets of variates. *Biometrika* 28:321–377.
- Wold H (1985): Partial least squares. In: Kotz S, Johnson N, editors. *Encyclopedia of Statistical Sciences*. New York: Wiley, 581–591.
- Kebets V, Holmes AJ, Orban C, Tang S, Li J, Sun N, et al. (2019): Somatosensory-motor dysconnectivity spans multiple transdiagnostic dimensions of psychopathology. *Biol Psychiatry* 86:779–791.
- Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. (2017): Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med* 23:28–38.
- Moser DA, Doucet GE, Lee WH, Rasgon A, Krinsky H, Leibu E, et al. (2018): Multivariate associations among behavioral, clinical, and multimodal imaging phenotypes in patients with psychosis. *JAMA Psychiatry* 75:386–395.
- Li J, Bolt T, Bzdok D, Nomi JS, Yeo BTT, Spreng RN, Uddin LQ (2019): Topography and behavioral relevance of the global signal in the human brain. *Sci Rep* 9:14286.
- Bijsterbosch JD, Woolrich MW, Glasser MF, Robinson EC, Beckmann CF, Van Essen DC, et al. (2018): The relationship between spatial configuration and functional connectivity of brain regions. *Elife* 7:e32992.
- Xia CH, Ma Z, Ciric R, Gu S, Betzel RF, Kaczkurkin AN, et al. (2018): Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat Commun* 9:3003.
- Modabbernia A, Janiri D, Doucet GE, Reichenberg A, Frangou S (2021): Multivariate patterns of brain-behavior-environment associations in the Adolescent Brain and Cognitive Development Study. *Biol Psychiatry* 89:510–520.
- Avants BB, Libon DJ, Rascovsky K, Boller A, McMillan CT, Massimo L, et al. (2014): Sparse canonical correlation analysis relates network-level atrophy to multivariate cognitive measures in a neurodegenerative population. *Neuroimage* 84:698–711.
- Ziegler G, Dahnke R, Winkler AD, Gaser C (2013): Partial least squares correlation of multivariate cognitive abilities and local brain structure in children and adolescents. *Neuroimage* 82:284–294.
- Jia T, Ing A, Quinlan EB, Tay N, Luo Q, Francesca B, et al. (2020): Neurobehavioural characterisation and stratification of reinforcement-related behaviour. *Nat Hum Behav* 4:544–558.
- Le Floch E, Guillemot V, Frouin V, Pinel P, Lalanne C, Trinchera L, et al. (2012): Significant correlation between a set of genetic polymorphisms and a functional brain network revealed by feature selection and sparse partial least squares. *Neuroimage* 63:11–24.
- Witten DM, Tibshirani R, Hastie T (2009): A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10:515–534.
- Marquand AF, Haak KV, Beckmann CF (2017): Functional cortico-striatal connection topographies predict goal-directed behaviour in humans. *Nat Hum Behav* 1:0146.
- Lin D, Calhoun VD, Wang YP (2014): Correspondence between fMRI and SNP data by group sparse canonical correlation analysis. *Med Image Anal* 18:891–902.
- Ing A, Sämann PG, Chu C, Tay N, Biondo F, Robert G, et al. (2019): Identification of neurobehavioural symptom groups based on shared brain mechanisms. *Nat Hum Behav* 3:1306–1318.
- Wang HT, Bzdok D, Margulies D, Craddock C, Milham M, Jefferies E, Smallwood J (2018): Patterns of thought: Population variation in the associations between large-scale network organisation and self-reported experiences at rest. *Neuroimage* 176:518–527.
- Smith SM, Nichols TE, Vidaurre D, Winkler AM, Behrens TEJ, Glasser MF, et al. (2015): A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat Neurosci* 18:1565–1567.
- Popovic D, Ruef A, Dwyer DB, Antonucci LA, Eder J, Sanfelici R, et al. (2020): Traces of trauma: A multivariate pattern analysis of childhood trauma, brain structure, and clinical phenotypes. *Biol Psychiatry* 88:829–842.
- Alnæs D, Kaufmann T, Marquand AF, Smith SM, Westlye LT (2020): Patterns of sociocognitive stratification and perinatal risk in the child brain. *Proc Natl Acad Sci U S A* 117:12419–12427.
- Mihalik A, Ferreira FS, Rosa MJ, Moutoussis M, Ziegler G, Monteiro JM, et al. (2019): Brain-behaviour modes of covariation in healthy and clinically depressed young people. *Sci Rep* 9:11536.
- Helmer M, Warrington S, Mohammadi-Nejad A-R, Lisa J, Howell A, Rosand B, et al. (2020): On stability of canonical correlation analysis and partial least squares with application to brain-behavior associations. *bioRxiv*. <https://doi.org/10.1101/2020.08.25.265546>.
- Mihalik A, Ferreira FS, Moutoussis M, Ziegler G, Adams RA, Rosa MJ, et al. (2020): Multiple holdouts with stability: Improving the generalizability of machine learning analyses of brain-behavior relationships. *Biol Psychiatry* 87:368–376.
- Dinga R, Schmaal L, Penninx BWJH, van Tol MJ, Veltman DJ, van Velzen L, et al. (2019): Evaluating the evidence for biotypes of depression: Methodological replication and extension of. *Neuroimage Clin* 22:101796.

Tutorial on CCA and PLS for Neuroscience

29. Uurtio V, Monteiro JM, Kandola J, Shawe-Taylor J, Fernandez-Reyes D, Rousu J (2017): A tutorial on canonical correlation methods. *ACM Comput Surv* 50:1–33.
30. Zhuang X, Yang Z, Cordes D (2020): A technical review of canonical correlation analysis for neuroscience applications. *Hum Brain Mapp* 41:3807–3833.
31. Wang H-T, Smallwood J, Mourao-Miranda J, Xia CH, Satterthwaite TD, Bassett DS, Bzdok D (2020): Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists. *Neuroimage* 216:116745.
32. Krishnan A, Williams LJ, McIntosh AR, Abdi H (2011): Partial least squares (PLS) methods for neuroimaging: A tutorial and review. *Neuroimage* 56:455–475.
33. Meredith W (1964): Canonical correlations with fallible data. *Psychometrika* 29:55–65.
34. Rosipal R, Krämer N (2006): Overview and recent advances in partial least squares. In: Saunders C, Grobelnik M, Gunn S, Shawe-Taylor J, editors. *Subspace, Latent Structure and Feature Selection*. Berlin: Springer, 34–51.
35. Golub GH, Zha H (1994): Perturbation analysis of the canonical correlations of matrix pairs. *Linear Algebra Appl* 210:3–28.
36. Wegelin JA: A survey on partial least squares (PLS) methods, with emphasis on the two-block case. Technical Report No. 371. <https://stat.uw.edu/sites/default/files/files/reports/2000/tr371.pdf>. Accessed November 23, 2017.
37. Vounou M, Nichols TE, Montana G (2010): Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *Neuroimage* 53:1147–1159.
38. Knapp TR (1978): Canonical correlation analysis: A general parametric significance-testing system. *Psychol Bull* 85:410–416.
39. Izenman AJ (1975): Reduced-rank regression for the multivariate linear model. *J Multivar Anal* 5:248–264.
40. Tibshirani R (1996): Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B* 58:267–288.
41. Hoerl AE, Kennard RW (1970): Ridge regression: Applications to nonorthogonal problems. *Technometrics* 12:69–82.
42. Zou H, Hastie T (2005): Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol* 67:301–320.
43. Vinod HD (1976): Canonical ridge and econometrics of joint production. *J Econom* 4:147–166.
44. Haroon DR, Szedmak S, Shawe-Taylor J (2004): Canonical correlation analysis: An overview with application to learning methods. *Neural Comput* 16:2639–2664.
45. Tenenhaus A, Tenenhaus M (2011): Regularized generalized canonical correlation analysis. *Psychometrika* 76:257–284.
46. Tuzhilina E, Tozzi L, Hastie T (2020): Canonical correlation analysis in high dimensions with structured regularization. *arXiv*. <https://doi.org/10.48550/arXiv.2011.01650>.
47. Kessy A, Lewin A, Strimmer K (2018): Optimal whitening and decorrelation. *Am Stat* 72:309–314.
48. Shmueli G (2010): To explain or to predict? *Stat Sci* 25:289–310.
49. Bzdok D, Engemann D, Thirion B (2020): Inference and prediction diverge in biomedicine. *Patterns (New York, NY)* 1: 100119.
50. Arbabshirani MR, Plis S, Sui J, Calhoun VD (2017): Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage* 145:137–165.
51. Abdi H (2010). Partial least squares regression and projection on latent structure regression (PLS regression). *Wiley Interdiscip Rev Comput Stat* 2:97–106.
52. Winkler AM, Renaud O, Smith SM, Nichols TE (2020): Permutation inference for canonical correlation analysis. *Neuroimage* 220: 117065.
53. Lê Cao K-A, Boitard S, Besse P (2011): Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* 12:253.
54. Labus JS, Van Horn JD, Gupta A, Alaverdyan M, Torgerson C, Asher-McNailey C, et al. (2015): Multivariate morphological brain signatures predict patients with chronic abdominal pain from healthy control subjects. *Pain* 156:1545–1554.
55. Olson Hunt MJ, Weissfeld L, Boudreau RM, Aizenstein H, Newman AB, Simonsick EM, et al. (2014): A variant of sparse partial least squares for variable selection and data exploration. *Front Neuroinform* 8:18.
56. Monteiro JM, Rao A, Shawe-Taylor J, Mourão-Miranda J (2016): A multiple hold-out framework for Sparse partial least squares. *J Neurosci Methods* 271:182–194.
57. Winkler AM, Webster MA, Vidaurre D, Nichols TE, Smith SM (2015): Multi-level block permutation. *Neuroimage* 123:253–268.
58. Mihalik A, Adams RA, Huys Q (2020): Canonical correlation analysis for identifying biotypes of depression. *Biol Psychiatry Cogn Neurosci Neuroimaging* 5:478–480.
59. Grosenick L, Shi TC, Gunning FM, Dubin MJ, Downar J, Liston C (2019): Functional and optogenetic approaches to discovering stable subtype-specific circuit mechanisms in depression. *Biol Psychiatry Cogn Neurosci Neuroimaging* 4:554–566.
60. Grosenick L, Liston C (2020): Reply to: A closer look at depression biotypes: Correspondence relating to Grosenick et al. (2019). *Biol Psychiatry Cogn Neurosci Neuroimaging* 5:556.
61. Dinga R, Schmaal L, Marquand AF (2020): A closer look at depression biotypes: Correspondence relating to Grosenick et al. (2019). *Biol Psychiatry Cogn Neurosci Neuroimaging* 5:554–555.
62. Rolls ET, Joliot M, Tzourio-Mazoyer N (2015): Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *Neuroimage* 122:1–5.
63. Folstein MF, Folstein SE, McHugh PR (1975): “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12:189–198.