Frontiers in NEUROINFORMATICS

TECHNOLOGY REPORT ARTICLE

published: 03 March 2014
doi: 10.3389/fninf.2014.00018

A variant of sparse partial least squares for variable selection and data exploration

Megan J. Olson Hunt[1], Lisa Weissfeld[1], Robert M. Boudreau[2], Howard Aizenstein[3], Anne B. Newman[4], Eleanor M. Simonsick[5], Dane R. Van Domelen[6], Fridtjof Thomas[7], Kristine Yaffe[8] and Caterina Rosano[2]*

[1] *Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA*
[2] *Department of Epidemiology, Center for Aging and Population Health, University of Pittsburgh, Pittsburgh, PA, USA*
[3] *Departments of Psychiatry, Bioengineering and Clinical and Translational Science, University of Pittsburgh, Pittsburgh, PA, USA*
[4] *Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA*
[5] *Intramural Research Program, National Institute on Aging, Baltimore, MD, USA*
[6] *Department of Biostatistics, Emory University, Atlanta, GA, USA*
[7] *Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN, USA*
[8] *Department of Psychiatry, Neurology and Epidemiology, University of California, San Francisco, San Francisco, CA, USA*

Edited by:
Marc-Oliver Gewaltig, Ecole Polytechnique Federale de Lausanne, Switzerland

Reviewed by:
Graham J. Galloway, The University of Queensland, Australia

****Correspondence:***
Caterina Rosano, Department of Epidemiology, Center for Aging and Population Health, University of Pittsburgh, 130 N. Bellefield Street, Pittsburgh, PA 15213, USA
e-mail: rosanoc@edc.pitt.edu

When data are sparse and/or predictors multicollinear, current implementation of sparse partial least squares (SPLS) does not give estimates for non-selected predictors nor provide a measure of inference. In response, an approach termed "all-possible" SPLS is proposed, which fits a SPLS model for all tuning parameter values across a set grid. Noted is the percentage of time a given predictor is chosen, as well as the average non-zero parameter estimate. Using a "large" number of multicollinear predictors, simulation confirmed variables not associated with the outcome were least likely to be chosen as sparsity increased across the grid of tuning parameters, while the opposite was true for those strongly associated. Lastly, variables with a weak association were chosen more often than those with no association, but less often than those with a strong relationship to the outcome. Similarly, predictors most strongly related to the outcome had the largest average parameter estimate magnitude, followed by those with a weak relationship, followed by those with no relationship. Across two independent studies regarding the relationship between volumetric MRI measures and a cognitive test score, this method confirmed *a priori* hypotheses about which brain regions would be selected most often and have the largest average parameter estimates. In conclusion, the percentage of time a predictor is chosen is a useful measure for ordering the strength of the relationship between the independent and dependent variables, serving as a form of inference. The average parameter estimates give further insight regarding the direction and strength of association. As a result, all-possible SPLS gives more information than the dichotomous output of traditional SPLS, making it useful when undertaking data exploration and hypothesis generation for a large number of potential predictors.

Keywords: high-dimensional, multicollinearity, over-fitting, SPLS, inference, tuning parameters, network, MRI

INTRODUCTION

In fields such as neuroscience, chemometrics, and genetics, data is often collected on a large number of variables but with a relatively small sample size, and predictors may also be highly collinear. Statistical methods used in this setting include regression models, cluster analysis and/or tree-based methods, ridge regression and dimension-reduction techniques such as partial least squares (PLS). However, when variable selection is the goal, these may prove inadequate or difficult to interpret.

In the realm of ordinary least squares (OLS), multicollinearity affects both the stability of the estimated coefficients (Wold et al., 1984) and inference on these estimates (Farrar and Glauber, 1967). Essentially, model prediction ability is poor when estimates are unstable (Wold et al., 1984), and one cannot trust conclusions drawn from test statistics, *p*-values or confidence intervals due to artificially inflated standard errors (Farrar and Glauber, 1967). As an alternative to OLS, ridge regression (Hoerl and Kennard, 2000; McDonald, 2009) and PLS account for multicollinearity and/or over-fitting. However, they are not intended for variable selection without additional computation such as bootstrapping (Abdi, 2010).

In PLS, latent variables (linear combinations of the predictors) are formed using both the outcome(s) and predictors such that all pairs of latent variables are orthogonal and have a sample correlation of zero (Garthwaite, 1994). Regression models are then fit using these latent variables rather than the original predictors and multicollinearity is no longer a concern. In addition, the number of latent variables is often smaller than the number of predictors, so that PLS reduces the dimensionality of the data and the likelihood of over-fitting. However, all predictors are assigned a

non-zero weight and inference is not provided, so that variable selection is not readily achieved (Tobias, 1997; Chun and Keleş, 2010). Further detail on the theory underlying PLS regression is available elsewhere (Garthwaite, 1994; Wold et al., 2001; Krishnan et al., 2011).

Given standard PLS is not intended for variable selection but rather prediction, sparse methods such as sparse partial least squares (SPLS) were developed. Variable selection is accomplished by using tuning parameters in the modeling process, which drive both the latent variable selection and computation of predictors' weights (Chun and Keleş, 2010). Here, estimates may be set to zero, indicating a predictor is not significantly associated with the outcome.

Although some weights are zero so as to provide variable selection, this can also be viewed as a weakness of SPLS. In data exploration and hypothesis generation, effect size and p-values, despite insignificance, are often of interest. During exploratory analyses, one may wish to increase the type-I error rate and allow variables that would otherwise be borderline significant or insignificant into the set of selected predictors. Also, one may wish to compare standardized estimates of various predictors despite insignificance. None of this information is provided by executing SPLS in its traditional manner.

To address these shortcomings, an alternative approach, referred to here as "all-possible" SPLS, is proposed. Briefly, a SPLS model is fit for "all possible" values of the model's tuning parameters, as opposed to fitting only one model based on the "optimal" parameters (this latter approach will be referred to as "traditional" SPLS). Predictors are ranked by the percentage of time they are chosen across all models, and the average of non-zero standardized parameter estimates is given for all predictors, even those not chosen by traditional SPLS. Although not formal inference such as a p-value, the former gives the relative ranking of predictors, allowing one to identify potentially borderline significant variables, as well as those least likely to be predictive of the outcome. Simulation confirms predictors most strongly associated with the outcome are robust to changes in the tuning parameters and continue to be selected as sparsity increases, while those with the weakest association are less likely to be chosen under high levels of sparsity. This approach yields supplementary information lost in the traditional application of SPLS, providing increased insight into one's data.

METHODS

TRADITIONAL SPLS

The spls package (version 2.1-0) in R (version 2.13.2) based on the theory presented by Chun and Keleş (2010) is considered here. The algorithm requires the specification of two tuning parameters, K and η. K (an integer between 1 and $\min\{p, (v - 1)n/v\}$, where v is the number of folds for the cross-validation (CV), p is the number of predictors and n is the sample size (Chung et al., 2009)) is the number of latent variables and η (a continuous value on the interval $[0, 1)$) determines the amount of sparsity in the algorithm. In general, lower values of η represent less sparsity (and thus more variables tend to be selected), whereas higher values imply more sparsity. However, the choice of K also affects variable selection in conjunction with η (lower values of K tend to result in fewer chosen variables).

To facilitate the choice of K and η, the package includes a CV function, where the "optimal" K and η are those with the lowest mean squared prediction error. For the purposes of this paper, "traditional" SPLS refers to the use of this CV to choose one pair of "optimal" tuning parameters. Once determined, the SPLS model is fit and selected predictors are noted.

While using traditional SPLS, it was discovered the selection of optimal tuning parameters was affected by the seed if CV other than leave-one-out (LOO) was used. For example, for 1000 randomly-chosen seeds, the optimal values of the tuning parameters chosen most often by a 10-fold CV in the real data used in Section Data Application: Volumetric MRI Regions as Predictors of Cognitive Test Results were $K = 2$, $\eta = 0.7$. However, they were only chosen for 171 seeds out of 1000—about 17% of the time. The next pair chosen most often was $K = 3$, $\eta = 0.6$, at 106 times. All of the remaining pairings were chosen less than 10% of the time, so that no one pair was selected notably more than the others. Note that if K and/or η differ only by one unit, this can mean the addition or exclusion of one or more variables from the results. Here, eight predictors were chosen by the first set of tuning parameters, whereas 17 were chosen by the second, indicating instability in the tuning parameter values can cause instability in the variable selection process, affecting conclusions. Because of the unreliability of the 10-fold CV with these data, LOO CV is recommended for traditional SPLS.

Another consideration with the CV is how fine of a grid to use when searching for the optimal value of η, since, again, it is continuous. In the examples provided by the authors of the spls package, η may be one of 0.1, 0.2, 0.3, . . . , 0.9 (Chung et al., 2009; Chun and Keleş, 2010). Given this, and also the fact that considering more η-values results in significantly more computational time, η-values of 0.1, 0.2, 0.3, . . . , 0.9 were used in this paper as well.

"ALL-POSSIBLE" SPLS

"All-possible" is quoted because, given η is continuous, one cannot actually achieve every possible combination of tuning parameters. Given a discrete subset of η (here, $\{0.1, 0.2, \ldots, 0.9\}$), however, one considers "all possible" combinations of the parameters. Specifically, there will be $K \times \eta$ total models fit, one for each combination of K and η, with standardized estimates recorded in each instance. The results are the percentage of time chosen (i.e., the parameter estimate was non-zero), as well as the average non-zero standardized parameter estimate.

It should be noted that with this method it is expected all predictors will be chosen a reasonable number of times (usually in at least 70% of the models). This is because once a large enough K- and/or small enough η-value is used, the method no longer induces enough sparsity to allow for variable selection—it essentially acts like PLS and chooses all variables. Since *all* pairings of K and η were considered here, many of them resulted in all variables being selected.

There are two advantages to all-possible SPLS. First, by ranking the variables based on how often they are chosen across all models, one has a relative way to compare them, as opposed to

"chosen" or "not chosen." Specifically, one can see those variables selected most and least frequently, as well as those that were somewhere in between. In this way, one obtains a continuum of information instead of a dichotomy. Second, an effect size for all predictors—not just those chosen by traditional SPLS—is provided. Thus, even if a predictor was only selected 75% of the time, one still has information on its estimate whenever it *was* selected.

SIMULATION

SIMULATION STRUCTURE

A design analogous to that in Chun and Keleş (2010) was used to create collinear predictors of varying association with the outcome—one set of predictors was strongly associated, another weakly and a third not at all. For $j = 1, 2, 3$ and $c_{j-1} + 1 \leq i \leq c_j$, where $(c_0, c_1, c_2, c_3) = (0, 7, 17, 27)$, predictors were of the form $\mathbf{x_i} = \mathbf{m_j} + \boldsymbol{\varepsilon_i}$. Given a sample size of $n = 100$, $\mathbf{m_j}$ were each vectors of length 100 from $N(\mathbf{0}, 20\mathbf{I}_{100})$ and $\boldsymbol{\varepsilon_i} \sim N(\mathbf{0}, \mathbf{I}_{100})$. Lastly, $\mathbf{y} = 2\mathbf{m}_1 - 0.2\mathbf{m}_2 + \boldsymbol{\tau}$, where $\boldsymbol{\tau} \sim N(\mathbf{0}, \mathbf{I}_{100})$. All variables were standardized while other settings for the SPLS function were kept at default.

PREDICTORS WITH WEAKER ASSOCIATION ARE LESS LIKELY TO BE CHOSEN WITH INCREASED SPARSITY

This simulation demonstrated how predictors with varying levels of association with $\mathbf{y}$ are affected by changes in the tuning parameter pair, (K, η). The general pattern is that for lower values of K and higher values of η, sparsity increases and fewer variables are selected. Here, $K = \{1, \ldots, 27\}$ and again $\eta = \{0.1, \ldots, 0.9\}$.

Consider three sets of predictors: $\mathbf{S}_1 = \{\mathbf{x}_1, \ldots, \mathbf{x}_7\}$ (strongly associated with $\mathbf{y}$), $\mathbf{S}_2 = \{\mathbf{x}_8, \ldots, \mathbf{x}_{17}\}$ (weakly associated) and $\mathbf{S}_3 = \{\mathbf{x}_{18}, \ldots, \mathbf{x}_{27}\}$ (not associated). For each $d = 1, \ldots, D = 1000$ samples drawn randomly from the distribution as outlined in Section Simulation Structure, a SPLS model was run for all pairs of K and η. The percentage of predictors chosen from each set was noted for each pair and the average across all 1000 data sets is shown in **Figures 1A,B** for $\mathbf{S}_2$ and $\mathbf{S}_3$. Note that K only ranges from 1 to 15, as after $K = 15$, the average was 100% for all pairs of tuning parameters. For $\mathbf{S}_1$, all seven predictors were always chosen (i.e., the average was always 100%).

These results confirm variables in set $\mathbf{S}_3$ (not associated with $\mathbf{y}$) were less likely to be chosen as K decreased and η increased (i.e., sparsity increased). Variables in $\mathbf{S}_2$ showed a similar pattern due to their weak association, although their rate of selection was notably higher than those in $\mathbf{S}_3$. The fact that all variables in $\mathbf{S}_1$ were chosen for 100% of the (K, η) pairs across all D data sets shows strongly associated variables are robust to changes in the tuning parameters. Subsequently, calculating the percentage of time a variable is selected over all pairs of tuning parameters (i.e., conducting all-possible SPLS) will result in those with the strongest association having the highest percentage of time chosen, while the opposite will be true for those with the weakest. This is shown via simulation in the next section.

PERCENTAGE OF TIME CHOSEN AND AVERAGE NON-ZERO STANDARDIZED ESTIMATES

For each of $d = 1, \ldots, D = 1000$ samples from the distribution as described in Section Simulation Structure, all-possible SPLS was conducted: For a given data set, a SPLS model was run for all pairs of $K = \{1, \ldots, 27\}$ and $\eta = \{0.1, \ldots, 0.9\}$. Recorded was the percentage of time each variable was chosen, as well as the mean non-zero standardized parameter estimates. **Table 1** reports the average of these percentages and mean estimates across all 1000 samples, in order to assess the method's behavior in the long run.

The average percentage of time chosen for all predictors in $\mathbf{S}_1$ was 100, while those in $\mathbf{S}_2$ and $\mathbf{S}_3$ were all chosen around 96% and 90% of the time on average, respectively, resulting in three distinct groups. The average mean non-zero standardized estimates for those in $\mathbf{S}_1$ were all around 0.15, while those in $\mathbf{S}_2$ were

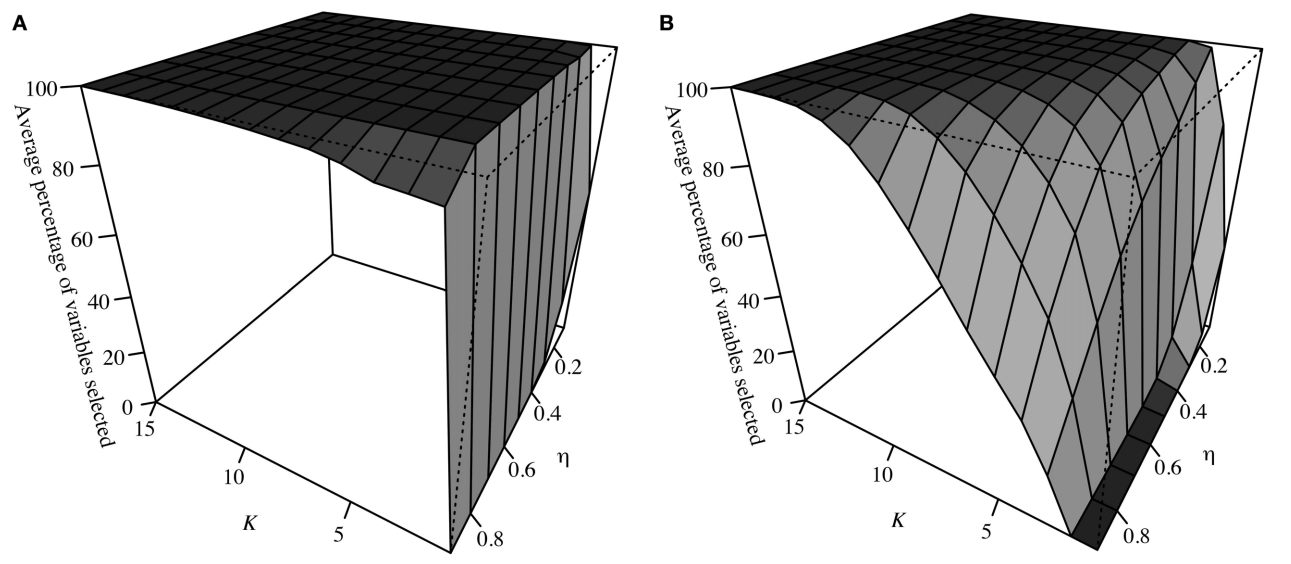

FIGURE 1 | (A) shows the average percentage of variables in $\mathbf{S}_2$ selected for each pair of tuning parameters across $D = 1000$ simulated data sets, while **(B)** shows this for $\mathbf{S}_3$.

Table 1 | From all-possible SPLS conducted on $D = 1000$ samples from the same distribution.

Predictor		Average percentage of time chosen	Average of mean non-zero standardized $\hat{\beta}$
$\mathbf{S}_1$	$\mathbf{x}_1$	100	0.144
	$\mathbf{x}_2$	100	0.147
	$\mathbf{x}_3$	100	0.145
	$\mathbf{x}_4$	100	0.142
	$\mathbf{x}_5$	100	0.144
	$\mathbf{x}_6$	100	0.147
	$\mathbf{x}_7$	100	0.145
$\mathbf{S}_2$	$\mathbf{x}_{12}$	96.5	−0.011
	$\mathbf{x}_{11}$	96.482	−0.011
	$\mathbf{x}_{17}$	96.476	−0.010
	$\mathbf{x}_8$	96.474	−0.011
	$\mathbf{x}_{10}$	96.472	−0.011
	$\mathbf{x}_{15}$	96.466	−0.012
	$\mathbf{x}_{16}$	96.465	−0.010
	$\mathbf{x}_{14}$	96.460	−0.007
	$\mathbf{x}_{14}$	96.453	−0.009
	$\mathbf{x}_9$	96.451	−0.011
$\mathbf{S}_3$	$\mathbf{x}_{20}$	89.912	0.0011
	$\mathbf{x}_{21}$	89.886	−0.0003
	$\mathbf{x}_{24}$	89.851	−0.0051
	$\mathbf{x}_{18}$	89.839	−0.0026
	$\mathbf{x}_{27}$	89.786	0.0021
	$\mathbf{x}_{23}$	89.778	0.0038
	$\mathbf{x}_{26}$	89.770	0.0007
	$\mathbf{x}_{22}$	89.734	−0.0023
	$\mathbf{x}_{25}$	89.730	0.0001
	$\mathbf{x}_{19}$	89.710	0.0035

Average percentage of time a variable was chosen and its average mean non-zero standardized parameter estimate across D data sets. Variables are ordered by average percentage.

about −0.01, and those in $\mathbf{S}_3$ were always smaller than those in $\mathbf{S}_2$ (and $\mathbf{S}_1$). Both the magnitudes and directions of the estimates for $\mathbf{S}_1$ and $\mathbf{S}_2$ were as expected given the structure of the data outlined in Section Simulation Structure and the fact that estimates were standardized. The small magnitudes and varying directions of predictors in $\mathbf{S}_3$ were reasonable, as they should have estimates that hover around zero.

DATA APPLICATION: VOLUMETRIC MRI REGIONS AS PREDICTORS OF COGNITIVE TEST RESULTS

In neuroimaging, brain regions tend to be numerous and highly correlated, so that over-fitting and multicollinearity are of concern. Here, a well-established predictor-outcome relationship is used to illustrate the proposed SPLS method.

DATA COLLECTION

Participants

Data were obtained from the Cardiovascular Health Study (CHS), which is an ongoing, population-based, longitudinal study, and the Healthy Brain Project (HBP), a sub-study of the Health, Aging and Body Composition (Health ABC) Study, which is also longitudinal and population-based.

The CHS is a study of coronary heart disease and stroke risk in older adults. Briefly, 5888 community-dwelling older adults were identified between 1987 and 1993 from Medicare eligibility lists in four clinical centers (Forsyth County, NC; Sacramento County, CA; Washington County, MD and Pittsburgh, PA) (Fried and Borhani, 1991). Participants were recruited if they were age 65 or older at time of recruitment, non-institutionalized, not wheelchair-bound or undergoing active cancer treatment, able to give informed consent and expected to remain in the area for at least 3 years. The participants had annual clinic examinations through 1998–1999.

Brain MRIs were acquired for 523 participants in Pittsburgh in 1997–1999 (Lopez et al., 2003). Compared to the participants who did not have a brain MRI, these participants were younger, more likely to have more years of education and had a lower prevalence of cardiovascular diseases and cerebrovascular findings (Rosano et al., 2006, 2007a). In 2003–2004, a random sample of 327 brain MRIs from the 523 were re-read (Rosano et al., 2005, 2007a,b, 2008). No significant differences were observed with regard to demographics or health-related factors between these 327 participants and the 523 total subjects.

The Health ABC study began in 1997–1998 as a longitudinal, observational cohort study of 3075 well-functioning older adults from Pittsburgh, PA and Memphis, TN (Simonsick et al., 2001). Participants were enrolled if they were 70–79 years old and reported no difficulty walking a quarter of a mile (400 meters), climbing 10 steps or performing activities of daily living; were free of life-threatening cancers with no active treatment within the prior 3 years and had planned to remain within the study area for at least 3 years. In 2006–2007, 314 Health ABC participants from the Pittsburgh site who were interested in and eligible for a brain 3T MRI received a MRI in addition to in-person Health ABC assessments. This ancillary study of the Health ABC is referred to as the HBP.

Both studies have been approved by the institutional review boards of the University of Pittsburgh.

Magnetic Resonance Imaging (MRI) Measures

In both the CHS and HBP, brain MRI assessments included volumetric measures of gray matter for both individual regions and the whole brain.

The brain MRI protocol for the CHS carried out in 1997–1999 has been described elsewhere (Yue et al., 1997). Briefly, sagittal T1-weighted localizer sequences and axial spin-echo spin-density-weighted, spin-echo T2-weighted and T1-weighted images were acquired using a 1.5T scanner. A volumetric Spoiled Gradient Recalled Acquisition (SPGR) sequence with parameters optimized for maximal contrast among gray matter, white matter and cerebrospinal fluid (CSF) was acquired in the coronal plane (echo time/repetition time (TE/TR) = 5/25, flip angle = 40 deg., NEX = 1, slice thickness = 1.5/0 mm interslice). All MRI data were interpreted at a central MRI Reading Center using a standardized protocol (Bryan et al., 1997; Yue et al., 1997).

The protocol for the HBP study was performed with a Siemens 12-channel head coil and 3T Siemens Tim Trio MR

scanner at the Magnetic Resonance Research Center, University of Pittsburgh (Venkatraman et al., 2011; Rosano et al., 2012a,b). Magnetization-prepared rapid gradient echo T1-weighted images (MPRAGE) were acquired in the axial plane ($TR = 2300$ ms, $TE = 3.43$ ms, imaging time (TI) $= 900$ ms, 9° flip angle, 256 × 224 mm field of view (FOV), 1 × 1 mm voxel size, 256 × 224 matrix size, 176 slices and 1 mm thick). Fluid-attenuated inversion recovery (FLAIR) images were acquired in axial plane ($TR = 9160$ ms, $TE = 89$ ms, $TI = 2500$ ms, 150° flip angle, 256 × 212 mm FOV, 256 × 240 matrix size, 48 slices, 3 mm thick and 1 × 1 mm voxel size). Diffusion-weighted images were acquired using a single short spin-echo sequence ($TR = 5300$ ms, $TE = 88$ ms, $TI = 2500$ ms, 90° flip angle, 256 × 256 mm FOV, two diffusion values of $b = 0$ and 1000 s/mm, 12 diffusion directions, four repetitions, 40 slices, 3 mm thick, 128 × 128 matrix size, 2 × 2 mm voxel size and GRAPPA = 2). A neuroradiologist examined each MRI for neurologic abnormalities. A radiologist verified the presence of abnormalities with potential clinical relevance. No images were excluded because of unexpected findings.

Voxel counts of the gray matter were obtained for individual regions of interest and for the whole brain using a procedure previously described (Zhang et al., 2001; Tzourio-Mazoyer et al., 2002; Rosano et al., 2005; Wu et al., 2006). After skull and scalp stripping (Smith, 2002), and after segmentation of gray matter, white matter and CSF, the brain atlas and the individual subject brain were aligned and intensity normalization was done on each subject's structural image (SPGR for the CHS and MPRAGE for the HBP images), as well as on the template colin27, to give each subject the same orientation and image intensity distribution as the template and to improve the registration accuracy. For both the CHS and HBP, FMRIB-FAST was applied to segment the image into gray matter, white matter and CSF, while also correcting for spatial intensity variations such as bias field or radio-frequency inhomogeneities (Rosano et al., 2005; Wu et al., 2006). The registration procedure used a fully-deformable automatic algorithm (Thirion, 1998) that does not warp or stretch the individual brain, and thus minimizes measurement inaccuracies (Wu et al., 2006). Volumes were converted from number of voxels to cubic millimeters.

DATA DESCRIPTION AND PREPARATION

Dependent variable

Scores from the Modified Mini-Mental State Examination (3MS) were used as the dependent variable, as it is a highly studied outcome with regard to memory. The 3MS is a brief, general cognitive battery with components for orientation, concentration, language, praxis and immediate and delayed memory (Teng and Chui, 1987). Because scores tend to be clustered at the high end of the scale, a transformation for left-skewed data was used:-ln(101 - 3MS), where 3MS represents the test score for a given individual (Shackman et al., 2006).

Regions of interest and confounding variables

A tiered hypothesis was formed based on the strength of current findings, with the expectation that primary regions would have the strongest association with 3MS, followed by secondary regions. A third set of regions referred to as "non-hypothesized" were not expected to be associated with the outcome.

The primary hypothesized regions were the hippocampus, parahippocampus and entorhinal cortex (Zola-Morgan and Squire, 1993; Dickerson et al., 2001). The secondary hypothesis included additional memory-related regions: amygdala, caudate and medial parietal, lateral parietal and posterior cingulate cortices (Packard and Knowlton, 2002; Koivunen et al., 2011; Squire and Wixted, 2011). Lastly, non-hypothesized regions were those traditionally related to motor tasks and performance (not memory): putamen, pallidum, thalamus, supplementary motor cortex, cerebellum, and post-central and pre-central gyri (Rosano et al., 2007a). Because the pallidum measurements were highly skewed right, the natural logarithm of these values was used. Regions were not normalized, as total gray matter parenchyma was included as a covariate.

The following variables were included as predictors in all models because of prior work indicating an association with 3MS and/or brain structure (Brickman et al., 2008; Raji et al., 2010): race (coded as white and all other races), sex, age, obesity (indicated by a BMI greater than 30) and total brain parenchyma volume (here, represented by total gray matter volume). The treatment of confounding variables here is analogous to that in the OLS regression framework: They were included in all models and never removed, even if they were ultimately not significant. Thus, the interpretation of a set of selected variables is that they are significantly related to the outcome, controlling for confounding variables and all other brain regions.

Influential points

Before the analysis commenced, potentially influential data points were determined by modeling each predictor against each outcome individually and calculating externally studentized residuals in each case (SAS Institute Inc, 2008). Any observation with a residual greater than 2.5 in absolute value was removed from the analysis (this value is slightly less conservative than the cut-off of 2 suggested by the SAS documentation).

Three observations were removed from the HBP data based on the above criterion, while 11 were removed from the CHS. In both data sets, influential points were those with a notably small/large 3MS value paired with a large/small regional volume. The only exception was one observation in the HBP data, which had a very large total brain volume relative to the other subjects. For each data set, there were some subjects with invalid MRIs and/or missing covariate values, so that after removing these subjects and also the influential observations, the final sample size for the CHS was $n = 286$, while $n = 302$ for the HBP. In **Table 2**, p-values for differences in demographic measures between the CHS and HBP cohorts were obtained either by a chi-square test, two-sample t-test or the Kruskal-Wallis Test when normality was suspect.

Analyses were conducted using R version 2.13.2 (spls package 2.1-0) and SAS version 9.2 (SAS Institute Inc, 2008). Both the dependent and continuous independent variables were standardized, and, unless otherwise mentioned, all other settings were kept at default for all functions/procedures used. Run-time for the

Table 2 | Demographic and MRI volumetric summaries for Cardiovascular Health Study and Healthy Brain Project participants.

	Cardiovascular health study	Healthy brain project	*p*-value
Sample size	$n = 286$	$n = 302$	
Female (n, %)	177 (62%)	174 (58%)	0.33
White (n, %)	224 (78%)	181 (60%)	<0.001[a]
Obese (n, %)	46 (16%)	79 (26%)	0.004[a]
Age [mean (*SD*)]	78 (4.0)	83 (2.8)	<0.001[a]
3MS score [mean (*SD*)]	93.6 (5.2)	92.9 (6.7)	0.92
MRI volumes [mean mm^3 (*SD*)][b]			
Amygdala	2786 (605)	2934 (419)	
Anterior cingulate cortex	10554 (1587)	9615 (1517)	
Caudate	7704 (1835)	8586 (2047)	
Cerebellum	68220 (24799)	99264 (12788)	
Dorsolateral prefrontal cortex	67790 (9706)	26837 (3240)	
Entorhinal cortex	3922 (808)	3702 (664)	
Hippocampus	9649 (1296)	9452 (1205)	
Lateral parietal inferior cortex	10368 (2061)	10990 (1527)	
Lateral parietal superior cortex	9719 (2342)	11787 (1807)	
Medial parietal cortex	18483 (3408)	20611 (3031)	
Pallidum (natural logarithm)	5.49 (1.11)	5.90 (0.93)	
Parahippocampus	10144 (1570)	10663 (1608)	
Parenchyma (total gray matter)	466482 (66738)	527997 (55062)	
Post-central gyrus	15255 (3132)	18972 (2730)	
Posterior cingulate cortex	2557 (458)	2816 (695)	
Pre-central gyrus	13485 (2579)	16741 (2538)	
Putamen	2192 (1946)	3002 (2443)	
Supplementary motor cortex	9260 (2168)	128218 (2220)	
Thalamus	1872 (1016)	1856 (460)	

[a] *Significant with $\alpha = 0.05$.*

[b] *Mean volumes are expected to differ since the CHS and HBP used different MR scanners, thus p-values are not reported.*

SPLS analyses of interest was less than 5 minutes on a machine with the Windows 7 operating system (64 bit) and a 2.16 GHz Intel Core i7 processor.

MULTICOLLINEARITY AND OVER-FITTING

Multicollinearity was assessed using the condition number by fitting an OLS regression model that included all regions of interest and *a priori* confounders, where a value greater than 100 indicated significant multicollinearity (Belsley et al., 1980).

The CHS cohort had a condition number of 190, while the HBP group had a value of 227. Since both are notably larger than 100, multicollinearity is likely present in these data when all MRI regions are considered simultaneously in the same model (Belsley et al., 1980).

While the number of predictors (23) was not larger than the sample sizes (297 and 302), various rules of thumb indicate there should be 10–20 observations for each predictor in a model (Harrell, 2001). This suggests one should have at least 230 observations, and potentially as many as 460, which could indicate potential over-fitting with these data.

Table 3 | Results from all-possible (first two columns) and traditional (last column) SPLS for the Healthy Brain Project.

Brain region	% Times chosen with all-possible method	Mean non-zero $\hat{\beta}$ from all-possible method	Traditional SPLS $\hat{\beta}$[c]
[a]Hippocampus	100	0.276	0.262
[a]Parahippocampus	100	0.258	0.180
[b]Amygdala	97.6	0.137	0.065
Anterior cingulate cortex	96.6	0.088	0.091
[a]Entorhinal cortex	96.1	−0.279	−0.159
[b]Medial Parietal cortex	96.1	0.148	0.158
Supplementary motor cortex	96.1	−0.187	−0.310
Thalamus	95.2	−0.104	−0.098
Dorsolateral prefrontal cortex	92.3	0.071	0.020
[b]Lateral parietal superior cortex	91.8	−0.163	−0.101
Pallidum (natural logarithm)	90.3	−0.106	−0.113
[b]Lateral parietal inferior cortex	89.4	0.068	
Post-central gyrus	88.4	0.056	0.079
Putamen	85.0	0.031	
Pre-central gyrus	84.5	−0.024	
Cerebellum	84.1	0.006	0.012
[b]Caudate	83.6	−0.021	
[b]Posterior cingulate cortex	82.6	−0.033	

Regions are ordered by the percentage of time chosen across all models by the all-possible method (first column), with the horizontal lines separating potential groupings of predictors with regard to their strength of association with the outcome.

[a] *Primary hypothesized region.*

[b] *Secondary hypothesized region.*

[c] *Applicable only for those regions chosen by traditional SPLS using optimal tuning parameters.*

SPARSE PARTIAL LEAST SQUARES ANALYSIS

The spls package based on the theory presented by Chun and Keleş (2010) was used for both traditional and all-possible SPLS (**Tables 3**, **4**). Horizontal lines show potential empirically-driven cut-points that indicate varying levels of association between the predictors and outcome.

Within the HBP data set (**Table 3**), all-possible SPLS largely confirmed the proposed hypotheses by choosing two of the primary regions (hippocampus and parahippocampus) 100% of the time and the third (entorhinal cortex) in 96.1% of the models. Additionally, the three largest average non-zero parameter estimates from all-possible (second column) were for the three primary regions: entorhinal cortex (−0.279), hippocampus (0.276) and parahippocampus (0.258). This contrasts traditional SPLS

Table 4 | Results from all-possible (first two columns) and traditional (last column) SPLS for the Cardiovascular Health Study.

Brain region	% Times chosen with all-possible method	Average non-zero $\hat{\beta}$ from all-possible method	Traditional SPLS $\hat{\beta}^{c}$
[a]Parahippocampus	100	0.196	0.111
[a]Hippocampus	100	0.141	0.093
[b]Medial parietal cortex	100	0.228	0.054
[b]Lateral parietal inferior cortex	97.6	0.220	0.060
Pallidum (natural logarithm)	96.6	0.131	0.126
Pre-central gyrus	96.1	0.196	0.021
[b]Lateral parietal superior cortex	94.7	−0.290	
Dorsolateral prefrontal cortex	92.8	0.013	0.026
Supplementary motor cortex	90.8	−0.128	−0.110
[b]Amygdala	89.9	0.053	0.084
[a]Entorhinal cortex	89.4	−0.132	
[b]Posterior cingulate Cortex	88.9	−0.062	
Thalamus	88.9	−0.078	
Post-central gyrus	87.0	−0.070	
Putamen	84.1	0.054	
Anterior cingulate cortex	82.6	0.005	
[b]Caudate	79.7	0.022	
Cerebellum	79.2	0.050	

Regions are ordered by the percentage of time chosen across all models by the all-possible method (first column), with the horizontal lines separating potential groupings of predictors with regard to their strength of association with the outcome.

[a] Primary hypothesized region.

[b] Secondary hypothesized region.

[c] Applicable only for those regions chosen by traditional SPLS using optimal tuning parameters.

in that the region with the largest estimated magnitude (third column) was the supplementary motor cortex (−0.310), yet this was not a hypothesized region. Although chosen a relatively large percentage of the time by all-possible (96.1%), this region was ranked below/tied with all three primary regions and two secondary (amygdala, medial parietal cortex). It also had a smaller average estimate (−0.187) than all three primary regions. Thus, this region was deemed most significant by traditional, but ranked below multiple hypothesized regions by all-possible.

Traditional SPLS also chose post-central gyrus and cerebellum, so that one might conclude these regions are significantly predictive of 3MS, yet cerebellum was the third lowest-ranked region by all-possible (84.1%), and post-central the sixth lowest (88.4%).

Lastly, the additional information gained by all-possible SPLS (ranking according to percent) indicates the lateral parietal inferior cortex is a potentially borderline significant predictor (89.4%), which could not have been known based on the traditional results, as its parameter estimate was set to zero.

Despite being secondary regions, neither the caudate nor posterior cingulate cortex were chosen by either method, so that the results were consistent in this way and may indicate a different relationship in a multivariable setting than has been seen in previous studies involving individual predictors.

The CHS results (**Table 4**) are notably consistent with those from the HBP data. Specifically, two primary regions (parahippocampus, hippocampus) were again chosen in 100% of the models, although the third primary region (entorhinal cortex) was selected less often, at 89.4%. However, this region had a larger average parameter estimate (−0.132) than all other regions selected less than 90% of the time, and some regions selected in greater than 90% of the models (pallidum, dorsolateral prefrontal and supplementary motor cortices, all non-hypothesized). This again shows the utility of all-possible SPLS in that it highlighted a potentially important, borderline predictor that was missed by traditional.

The regions with the largest average magnitudes according to all-possible were the lateral parietal superior (−0.290), medial parietal (0.228) and lateral parietal inferior (0.220) cortices (all secondary), and the parahippocampus (0.196), a primary region, so that the top four largest estimates were associated with hypothesized regions. Alternatively, traditional SPLS assigned the largest parameter estimate to the pallidum (0.126), followed by the parahippocampus (0.111) and the supplementary motor cortex (−0.110), so that two of the three regions with the largest estimates according to traditional SPLS were non-hypothesized. In contrast, all-possible ranked both the pallidum (96.6%) and supplementary motor (90.8%) lower than two primary (parahippocampus, hippocampus) and two secondary (medial parietal, lateral parietal inferior cortices) regions (and also lower than lateral parietal superior in the case of the supplementary motor cortex).

Lastly, the posterior cingulate cortex and caudate, despite being secondary regions, were not chosen by either method. This finding for the caudate is consistent with that from the HBP.

DISCUSSION

The purpose of this study was to illustrate that all-possible SPLS provides additional, useful information not attainable by traditional SPLS: relative rankings and parameter estimates for non-selected predictors. Simulation verified that predictors not associated with the outcome are selected less often as sparsity increases, while strong, and in most cases weak, associations remain robust. Additionally, conducting all-possible SPLS a large number of times showed that, on average, the percentage of time chosen and mean non-zero standardized estimates were consistent with the structure of the simulated data. A real data example indicated all-possible SPLS was more successful at highlighting hypothesized relationships than traditional SPLS, and also gave useful information about borderline variables that could not otherwise have been known.

Given the CHS and HBP data sets differed with respect to neuroimaging protocols and demographics, it is notable that all-possible SPLS detected hypothesized associations across these cohorts, suggesting robustness in the method. Specifically, MR scanners had different field strengths: the CHS MRIs were obtained with a 1.5 Tesla, the HBP with a 3.0. Additionally, protocols with different spatial resolutions were used across groups: the CHS applied a 5.0 mm slice, whereas the HBP applied a 1.5. Lastly, the cohorts were significantly different with regard to race, obesity and age (although these factors were controlled for in all models). Despite these differences between data sets, the method yielded consistent results overall, indicating its utility as variable selection technique.

A weakness of all-possible SPLS is its relative nature (i.e., ranking by percentage) with no strict cut-off value due to a lack of distributional properties. For example, in the simulation in Section Percentage of Time Chosen and Average Non-Zero Standardized Estimates (**Table 1**), the average percentage defined three distinct groups, but with no insight into significance (or lack thereof). However, viewing the predictors in this way allows one to see more detail than the dichotomous results of traditional SPLS, and to apply a cut-off if desired, where the value would be based on empirical experience, rather than guided by theory.

By utilizing simulation and a well-studied predictor-outcome relationship across two independent studies, the current findings validate this variation of SPLS as a useful technique for selecting variables in situations where other approaches (namely, OLS) fail. The results of this study suggest all-possible SPLS could be used for hypothesis generation without having to restrict the set of predictors due to multicollinearity or a comparatively small sample size, which geneticists, neuroscientists, economists and social scientists often encounter. The additional information given by all-possible SPLS is especially useful in exploratory analyses, as it allows for a more thorough understanding of the data than can be provided by the binary results of traditional SPLS.

ACKNOWLEDGMENTS

This work was supported by the National Institute on Aging (NIA) AG-023629 and the National Science Foundation Graduate Research Fellowship DGE-0940903. Health ABC was supported by NIA contracts N01-AG-6-2103, N01-AG-6-2106 and N01-AG-6-2101; NIA grant R01-AG028050 and NINR grant R01-NR012459. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging. CHS was supported by contract numbers N01-HC-85239, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01-HC-15103, N01-HC-55222, N01-HC-75150, N01-HC-45133, P30-AG-024827-01 and grant number HL080295 from the National Heart, Lung, and Blood Institute (NHLBI), with added contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG-15928, AG-20098, AG-027058 and AG-023629 from the NIA, R01-HL-075366 from the NHLBI and the University of Pittsburgh Claude D. Pepper Older Americans Independence Center P30-AG-024827-07. A full list of principal CHS investigators and institutions can be found at https://chs-nhlbi.org/pi.

REFERENCES

Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS regression). *Wiley Interdiscip. Rev. Comput. Stat.* 2, 97–106. doi: 10.1002/wics.51

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression Diagnostics.* New York, NY: John Wiley & Sons. doi: 10.1002/0471725153

Brickman, A. M., Schupf, N., Manly, J. J., Luchsinger, J. A., Andrews, H., Tang, M. X., et al. (2008). Brain morphology in older African Americans, Carribean Hispanics, and Whites from northern Manhattan. *Arch. Neurol.* 64, 1053–1061. doi: 10.1001/archneur.65.8.1053

Bryan, R. N., Wells, S. W., Miller, T. J., Elster, A. D., Jungreis, C. A., Poirier, V. C., et al. (1997). Infarctlike lesions in the brain: prevalence and anatomic characteristics at MR imaging of the elderly-data from the cardiovascular health study. *Radiology* 202, 47–54.

Chun, H., and Keleş, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *J. R. Statist. Soc. B.* 72, 3–25. doi: 10.1111/j.1467-9868.2009.00723.x

Chung, D., Chun, H., and Keleş, S. (2009). *An Introduction to the 'Spls' Package, Version 1.0.* Available online at: http://cran.r-project.org/web/packages/spls/

Dickerson, B. C., Goncharova, I., Sullivan, M. P., Forchetti, C., Wilson, R. S., Bennett, D. A., et al. (2001). MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. *Neurobiol. Aging.* 22, 747–754. doi: 10.1016/S0197-4580(01)00271-8

Farrar, D. E., and Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *Rev. Econ. Stat.* 49, 92–107. doi: 10.2307/1937887

Fried, L. P., and Borhani, N. (1991). The cardiovascular health study: design and rationale. *Ann. Epidemiol.* 1, 263–276. doi: 10.1016/1047-2797(91)90005-W

Garthwaite, P. (1994). An interpretation of partial least squares. *J. Am. Stat. Assoc.* 89, 122–127. doi: 10.1080/01621459.1994.10476452

Harrell, F. E. Jr. (2001). *Regression Modeling Strategies With Applications to Linear Models, Logistic Regression and Survival Analysis.* New York, NY: Springer.

Hoerl, A. E., and Kennard, R. W. (2000). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 42, 80–86. doi: 10.1080/00401706.2000.10485983

Koivunen, J., Scheinin, N., Virta, J. R., Aalto, S., Vahlberg, T., Någren, K., et al. (2011). Amyloid PET imaging in patients with mild cognitive impairment: a 2-year follow-up study. *Neurology* 76, 1085–1090. doi: 10.1212/WNL.0b013e318212015e

Krishnan, A., Williams, L. J., McIntosh, A. R., and Abdi, H. (2011). Partial least squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* 56, 455–475. doi: 10.1016/j.neuroimage.2010.07.034

Lopez, O. L., Jagust, W. J., Dulberg, C., Becker, J. T., DeKosky, S. T., Fitzpatrick, A., et al. (2003). Risk factors for mild cognitive impairment in the cardiovascular health study cognition study: part 2. *Arch. Neurol.* 60, 1394–1399. doi: 10.1001/archneur.60.10.1394

McDonald, G. C. (2009). Ridge regression. *Wiley Interdiscip. Rev. Comput. Stat.* 1, 93–100. doi: 10.1002/wics.14

Packard, M. G., and Knowlton, B. (2002). Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.* 25, 563–593. doi: 10.1146/annurev.neuro.25.112701.142937

Raji, C. A., Ho, A. J., Parikshak, N. N., Becker, J. T., Lopez, O. L., Kuller, L. H., et al. (2010). Brain structure and obesity. *Hum. Brain Mapp.* 31, 353–364. doi: 10.1002/hbm.20870

Rosano, C., Aizenstein, H., Brach, J., Longenberger, A., Studenski, S., and Newman, A. B. (2008). Special article: gait measures indicate underlying focal gray matter atrophy in the brain of older adults. *J. Gerontol. A Biol. Sci. Med. Sci.* 63, 1380–1388. doi: 10.1093/gerona/63.12.1380

Rosano, C., Aizenstein, H. J., Newman, A. B., Venkatraman, V., Harris, T., Ding, J., et al. (2012a). Neuroimaging differences between older adults with maintained versus declining cognition over a 10-year period. *Neuroimage* 62, 307–313. doi: 10.1016/j.neuroimage.2012.04.033

Rosano, C., Aizenstein, H. J., Studenski, S., and Newman, A. B. (2007a). A regions-of-interest volumetric analysis of mobility limitations in community-dwelling older adults. *J. Gerontol. A Biol. Sci. Med. Sci.* 62, 1048–1055. doi: 10.1093/gerona/62.9.1048

Rosano, C., Becker, J., Lopez, O., Lopez-Garcia, P., Carter, C. S., Newman, A. B., et al. (2005). Morphometric analysis of gray matter volume in demented older adults: exploratory analysis of the cardiovascular health study brain MRI database. *Neuroepidemiology* 24, 221–229. doi: 10.1159/000085140

Rosano, C., Bennett, D. A., Newman, A. B., Venkatraman, V., Yaffe, K., Harris, T., et al. (2012b). Patterns of focal gray matter atrophy are associated with bradykinesia and gait disturbances in older adults. *J. Gerontol. A Biol. Sci. Med. Sci.* 67, 957–962. doi: 10.1093/gerona/glr262

Rosano, C., Brach, J., Longstreth, W. T. Jr., and Newman, A. B. (2006). Quantitative measures of gait characteristics indicate prevalence of underlying subclinical structural brain abnormalities in high-functioning older adults. *Neuroepidemiology* 26, 52–60. doi: 10.1159/000089240

Rosano, C., Brach, J., Studenski, S., Longstreth, W. T. Jr., and Newman, A. B. (2007b). Gait variability is associated with subclinical brain vascular abnormalities in high-functioning older adults. *Neuroepidemiology* 29, 193–200. doi: 10.1159/000111582

SAS Institute Inc. (2008). *SAS/STAT 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

Shackman, A. J., Sarinopoulos, I., Maxwell, J. S., Pizzagalli, D. A., Lavric, A., and Davidson, R. J. (2006). Anxiety selectivity disrupts visuospatial working memory. *Emotion* 6, 40–61. doi: 10.1037/1528-3542.6.1.40

Simonsick, E. M., Newman, A. B., Nevitt, M. C., Kritchevsky, S. B., Ferrucci, L., Guralnik, J. M., et al. (2001). Measuring higher level physical function in well-functioning older adults: expanding familiar approaches in the Health ABC study. *J. Gerontol. A Biol. Sci. Med. Sci.* 56, M644–M649. doi: 10.1093/gerona/56.10.M644

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. doi: 10.1002/hbm.10062

Squire, L. R., and Wixted, J. T. (2011). The cognitive neuroscience of human memory since H.M. *Annu. Rev. Neurosci.* 34, 259–288. doi: 10.1146/annurev-neuro-061010-113720

Teng, E. L., and Chui, H. C. (1987). The Modified Mini-Mental State (3MS) examination. *J. Clin. Psychiatry* 48, 314–318.

Thirion, J. P. (1998). Image matching as a diffusion process: an analogy with Maxwell's demons. *Med. Image Anal.* 2, 243–260. doi: 10.1016/S1361-8415(98)80022-4

Tobias, R. D. (1997). *An Introduction to Partial Least Squares Regression*. Cary, NC: SAS Institute Inc. Available online at: http://ftp.sas.com/techsup/download/technote/ts509.pdf

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Venkatraman, V. K., Aizenstein, H. J., Newman, A. B., Yaffe, K., Harris, T., Kritchevsky, S., et al. (2011). Lower digit symbol substitution score in the oldest old is related to magnetization transfer and diffusion tensor imaging of the white matter. *Front. Aging Neurosci.* 3:11. doi: 10.3389/fnagi.2011.00011

Wold, S., Ruhe, A., Wold, H., and Dunn, W. J. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* 5, 735–743. doi: 10.1137/0905052

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58, 109–130. doi: 10.1016/S0169-7439(01)00155-1

Wu, M., Carmichael, O., Lopez-Garcia, P., Carter, C. S., and Aizenstein, H. J. (2006). Quantitative comparison of AIR, SPM, and the fully deformable model for atlas-based segmentation of functional and structural MR images. *Hum. Brain Mapp.* 27, 747–754. doi: 10.1002/hbm.20216

Yue, N. C., Arnold, A. M., Longstreth, W. T. Jr., Elster, A. D., Jungreis, C. A., O'Leary, D. H., et al. (1997). Sulcal, ventricular, and white matter changes at MR imaging in the aging brain: data from the cardiovascular health study. *Radiology* 202, 33–39.

Zhang, Y., Brady, M., and Smith, S. M. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20, 45–57. doi: 10.1109/42.906424

Zola-Morgan, S., and Squire, L. R. (1993). Neuroanatomy of memory. *Annu. Rev. Neurosci.* 16, 547–563. doi: 10.1146/annurev.ne.16.030193.002555

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 April 2013; accepted: 11 February 2014; published online: 03 March 2014. Citation: Olson Hunt MJ, Weissfeld L, Boudreau RM, Aizenstein H, Newman AB, Simonsick EM, Van Domelen DR, Thomas F, Yaffe K and Rosano C (2014) A variant of sparse partial least squares for variable selection and data exploration. Front. Neuroinform. ***8****:18. doi: 10.3389/fninf.2014.00018*

This article was submitted to the journal Frontiers in Neuroinformatics.