Biometrics 00, 1–22 DOI: 00.0000/a.0000-0000.0000.00000.a
September 2026

Supplementary Materials for "Dimension Reduction for Characterizing Sexual Dimorphism in Biomechanics of the Temporomandibular Joint" by Sung Hee Park, Xin Zhang, Elizabeth H. Slate, Shuchun Sun, and Hai Yao

Sung Hee Park[1,2], Xin Zhang[1,*], Elizabeth H. Slate[1,], Shuchun Sun[3], and Hai Yao[3]**

[1]Department of Statistics, Florida State University, Tallahassee, FL, U.S.A.

[2]School of Data Science and Analytics, Kennesaw State University, Marietta, GA, U.S.A.

[3]Clemson-MUSC Bioengineering Program, Department of Bioengineering, Clemson University, SC, U.S.A.

*_email:_ henry@stat.fsu.edu

**_email:_ eslate@stat.fsu.edu

Summary: Section S1 presents additional simulation results under different types of group differences, varying signal strengths of the nonzero components, different covariance structures, and a multi-group setting with a third variable. In Section S2, we provide numerical results comparing our approach with other multivariate association methods. Section S3 includes simulations of the SPSS method used to select predefined sparsity levels. Section S4 introduces an information criterion for determining sparsity levels in cases with larger sample sizes. Section S5 describes two resampling techniques that empirically select the sparsity levels. Section S6 discusses the computational complexity of the CCR model. Section S7 presents further results of the CCR model applied to the marginally standardized skull and temporalis origin (TO).

This paper has been submitted for consideration for publication in *Biometrics*

S1. Additional Simulation Results

S1.1 *Simulation in Different Group Correlation*

From $\mathbf{\Phi}$ in the CCR model, we can control the magnitude of the group difference through ρ_1 and ρ_2 where $\rho_1 - \rho_2 > 0$. As $\rho_1 - \rho_2$ becomes smaller, the group difference decreases, making it more difficult for the CCR model to capture the group-specific association pattern.

In Table S1, we examine the performance of the CCR model under different values of ρ_1 and ρ_2. The worst case is $(\rho_1, \rho_2) = (0.25, -0.25)$, $n_1 = n_2 = 20$ which has the smallest group difference under the smallest sample size. In the worst result, the TPR values are large enough with smaller FPR values. And the subspace distances $\mathrm{D}_{\mathbf{V}}$ and $\mathrm{D}_{\mathbf{U}}$ are reasonably estimated (< 0.5). We can compare the result of the associated correlation difference $\widehat{\eta}_1 = 1.12$ when $(\rho_1, \rho_2) = (0.6, -0.6)$, $n_1 = n_2 = 20$ to the $\widehat{\eta}_1 = 1.16$ in the real data analysis. The $\widehat{\eta}$ values may support our real data analysis result by correctly selecting the variables on the skull and the temporalis origin (TO) muscle to specify the sex dimorphism in TMJ mechanics.

Table S1: Numerical evaluations under rank-1 of cross-covariances in different (ρ_1, ρ_2) over 100 data replicates. The numbers in parentheses report the standard error of the subspace distances $\mathrm{D}_{\mathbf{U}}$, $\mathrm{D}_{\mathbf{V}}$, covariance difference $\widehat{\delta}_1$, and the associated correlation differences $\widehat{\eta}_1$. Here, we used the true sparsity levels ($s_1^* = s_1$, $s_2^* = s_2$) for estimation.

Rank ($r = 1$)	(ρ_1, ρ_2) $(0.25, -0.25)$		$(0.6, -0.6)$		$(0.9, -0.9)$	
$n_1 = n_2$	20	200	20	200	20	200
$\mathrm{TPR}_{\mathbf{X}}$	0.777	0.997	0.983	1.000	1.000	1.000
$\mathrm{TPR}_{\mathbf{Y}}$	0.810	0.990	0.980	1.000	1.000	1.000
$\mathrm{FPR}_{\mathbf{X}}$	0.045	0.001	0.003	0.000	0.000	0.000
$\mathrm{FPR}_{\mathbf{Y}}$	0.048	0.002	0.005	0.000	0.000	0.000
$\mathrm{D}_{\mathbf{U}}$	0.440 (0.034)	0.098 (0.008)	0.177 (0.014)	0.043 (0.002)	0.091 (0.005)	0.028 (0.001)
$\mathrm{D}_{\mathbf{V}}$	0.409 (0.032)	0.109 (0.011)	0.177 (0.015)	0.041 (0.002)	0.094 (0.006)	0.028 (0.001)
$\widehat{\delta}_1$	4.340 (0.144)	3.426 (0.073)	7.857 (0.223)	8.051 (0.077)	11.758 (0.252)	12.099 (0.088)
$\widehat{\eta}_1$	0.841 (0.014)	0.507 (0.010)	1.198 (0.018)	1.191 (0.006)	1.786 (0.006)	1.794 (0.002)

S1.2 *Simulation in Different Signal Strength*

Another parameter we can control is the ratio of c_1/c_2, which we described in Section 3.1. The ratio controls the strength of signals. For example, if we use $c_1/c_2 > 1$, the signals for the non-zero components in $\mathbf{\Sigma}_{\mathbf{X},1}$ and $\mathbf{\Sigma}_{\mathbf{Y},1}$ are larger than the signals for the zero components in $\mathbf{\Sigma}_{\mathbf{X},2}$ and $\mathbf{\Sigma}_{\mathbf{Y},2}$, making it easier to estimate the nonzero components in $\mathbf{X}$ and $\mathbf{Y}$, respectively.

Thus, we simulated four different ratios as $c_1/c_2 \in \{0.5, 1, 3, 5\}$ and the corresponding results are summarized in Table S2. When we have weak signals ($c_1/c_2 = 0.5$) with a small sample size ($n_1 = n_2 = 20$), the CCR model still well captures the true non-zero components ($\text{TPR}_{\mathbf{X}}, \text{TPR}_{\mathbf{Y}} > 0.74$) and the subspace distances are still well estimated ($\text{D}_{\mathbf{V}}, \text{D}_{\mathbf{U}} < 0.4$). We can compare the maximal covariance difference $\widehat{\delta}_1$ to the $\widehat{\delta}_1 = 119.65$ in the real data analysis. That these values are much larger than those in Table S2 supports that the results in Section 4 successfully differentiate the group difference through the CCR model.

Table S2: Numerical evaluations under rank-1 of cross-covariances in different $c_1/c_2 \in \{0.5, 1, 3, 5\}$ over 100 data replicates. The numbers in parentheses report the standard error of the subspace distances $\text{D}_{\mathbf{U}}$, $\text{D}_{\mathbf{V}}$, covariance difference $\widehat{\delta}_1$, and the associated correlation differences $\widehat{\eta}_1$. Here, we used the true sparsity levels ($s_1^* = s_1 = 3$, $s_2^* = s_2 = 3$) for estimation.

Rank ($r = 1$)	c_1/c_2 = 0.5		1		3		5	
$n_1 = n_2$	20	200	20	200	20	200	20	200
$\text{TPR}_{\mathbf{X}}$	0.747	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\text{TPR}_{\mathbf{Y}}$	0.750	1.000	0.997	1.000	1.000	1.000	1.000	1.000
$\text{FPR}_{\mathbf{X}}$	0.051	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\text{FPR}_{\mathbf{Y}}$	0.062	0.000	0.001	0.000	0.000	0.000	0.000	0.000
$\text{D}_{\mathbf{U}}$	0.385 (0.040)	0.027 (0.002)	0.094 (0.005)	0.027 (0.002)	0.091 (0.005)	0.028 (0.001)	0.093 (0.005)	0.026 (0.001)
$\text{D}_{\mathbf{V}}$	0.375 (0.040)	0.031 (0.002)	0.094 (0.008)	0.031 (0.002)	0.094 (0.095)	0.028 (0.001)	0.095 (0.005)	0.028 (0.001)
$\widehat{\delta}_1$	2.145 (0.047)	2.017 (0.015)	3.920 (0.084)	4.033 (0.029)	11.758 (0.252)	12.099 (0.088)	19.588 (0.419)	20.179 (0.144)
$\widehat{\eta}_1$	1.650 (0.026)	1.794 (0.002)	1.783 (0.007)	1.794 (0.002)	1.786 (0.006)	1.794 (0.002)	1.786 (0.006)	1.796 (0.002)

S1.3 *Simulation in Different Covariance Structure*

Here, we added the simulation results in different covariance structures in data generation since data in Section 3 were generated from the autoregressive structure with parameter 0.7.

We considered two covariance structures: compound symmetric (CS) and autoregressive (AR) with parameters in $\{0.3, 0.6, 0.9\}$. The results are displayed in Table S3. There is no substantial difference in the different covariance structures. We could see that the maximal covariance difference $\widehat{\delta}_1$ increases as the parameter increases since the larger parameter generates larger eigenvalues on the corresponding covariance matrices for data generation.

Table S3: Numerical evaluations under rank-1 of cross-covariances in different covariance structures (auto-regressive and compound symmetric structures) with parameter $\{0.3, 0.6, 0.9\}$ over 100 data replicates. The numbers in parentheses report the standard error of the subspace distances $\mathrm{D}_{\mathbf{U}}$, $\mathrm{D}_{\mathbf{V}}$, covariance difference $\widehat{\delta}_1$, and the associated correlation differences $\widehat{\eta}_1$. Here, we used the true sparsity levels ($s_1^* = s_1$, $s_2^* = s_2$) for estimation.

Rank ($r = 1$)	compound symmetric (CS)			auto-regressive (AR)		
Parameter	0.3	0.6	0.9	0.3	0.6	0.9
$\mathrm{TPR}_{\mathbf{X}}$	1.000	1.000	1.000	1.000	1.000	1.000
$\mathrm{TPR}_{\mathbf{Y}}$	1.000	1.000	1.000	1.000	1.000	1.000
$\mathrm{FPR}_{\mathbf{X}}$	0.000	0.000	0.000	0.000	0.000	0.000
$\mathrm{FPR}_{\mathbf{Y}}$	0.000	0.000	0.000	0.000	0.000	0.000
$\mathrm{D}_{\mathbf{U}}$	0.039 (0.002)	0.033 (0.002)	0.028 (0.001)	0.033 (0.002)	0.027 (0.002)	0.023 (0.001)
$\mathrm{D}_{\mathbf{V}}$	0.036 (0.002)	0.033 (0.002)	0.027 (0.001)	0.037 (0.002)	0.030 (0.002)	0.023 (0.001)
$\widehat{\delta}_1$	17.243 (0.127)	19.897 (0.143)	22.559 (0.163)	16.616 (0.124)	19.189 (0.138)	22.295 (0.160)
$\widehat{\eta}_1$	1.798 (0.002)	1.797 (0.002)	1.797 (0.002)	1.793 (0.002)	1.793 (0.002)	1.796 (0.002)

S1.4 *Simulation in Multi-Categorical Case*

We explore a case where the third variable Z has more than two categories, multi-categorical variable. Unlike the binary case presented in the CCR model, we maximize each pairwise covariance difference by stacking the corresponding data matrices for $\mathbf{X}$ and $\mathbf{Y}$, and estimate $\mathbf{U}$ and $\mathbf{V}$ separately for each pairwise comparison.

For simplicity, we assume that Z has three categories. Under the rank-1 scenario, we construct the pairwise covariance differences as follows:

$$\mathbf{\Phi}_{12}\mathbf{\Phi}_{12}^{\top} = (\rho_1 - \rho_2)^2, \;\; \mathbf{\Phi}_{23}\mathbf{\Phi}_{23}^{\top} = (\rho_2 - \rho_3)^2, \;\; \mathbf{\Phi}_{31}\mathbf{\Phi}_{31}^{\top} = (\rho_3 - \rho_1)^2,$$

where ρ_1, ρ_2, and ρ_3 represent group-specific canonical correlations, and each pairwise difference (e.g., $\rho_1 - \rho_2$) corresponds to the singular value in the SVD of the associated pairwise difference matrix. Then, we construct the sample estimates by stacking the pairwise covariance difference matrices as follows:

$$\widetilde{\mathbf{\Phi}}_{\mathbf{X}} = [\widetilde{\mathbf{\Phi}}_{12}, \widetilde{\mathbf{\Phi}}_{23}, \widetilde{\mathbf{\Phi}}_{31}] \in \mathbb{R}^{p_1 \times (3 \cdot p_2)}, \;\; \widetilde{\mathbf{\Phi}}_{\mathbf{Y}} = [\widetilde{\mathbf{\Phi}}_{12}^{\top}, \widetilde{\mathbf{\Phi}}_{23}^{\top}, \widetilde{\mathbf{\Phi}}_{31}^{\top}]^{\top} \in \mathbb{R}^{(3 \cdot p_1) \times p_2},$$

where $\widetilde{\mathbf{\Phi}}_{12}$, $\widetilde{\mathbf{\Phi}}_{23}$, and $\widetilde{\mathbf{\Phi}}_{31}$ denote the pairwise sample covariance difference matrices. We then estimate the sparse matrices $\mathbf{U}$ and $\mathbf{V}$ following the procedure described in Algorithm S1.

Algorithm S1 CCR for multi-categorical Z via two-way sparse iterative thresholding

1: **Inputs:**
$\widetilde{\mathbf{\Phi}}_{\mathbf{X}} \in \mathbb{R}^{p_1 \times (3p_2)}$, $\widetilde{\mathbf{\Phi}}_{\mathbf{Y}} \in \mathbb{R}^{(3p_1) \times p_2}$, rank $r \leqslant \min(p_1, p_2)$, sparsity levels s_1, s_2

2: **Initialize:** Compute top-r singular vectors $(\widehat{\mathbf{U}}^{(0)}, \widehat{\mathbf{V}}^{(0)})$ from $\widetilde{\mathbf{\Phi}}_{\mathbf{X}}$ and $\widetilde{\mathbf{\Phi}}_{\mathbf{Y}}$

3: **Repeat** $t = 1, 2, \ldots$

$$\widehat{\mathbf{U}}^{(t)} \leftarrow Q\Big(T_{s_1}\Big(\widetilde{\mathbf{\Phi}}_{\mathbf{X}}[\widehat{\mathbf{V}}^{(t-1)\top}, \widehat{\mathbf{V}}^{(t-1)\top}, \widehat{\mathbf{V}}^{(t-1)\top}]^{\top}\Big)\Big),$$
$$\widehat{\mathbf{V}}^{(t)} \leftarrow Q\Big(T_{s_2}\Big(\widetilde{\mathbf{\Phi}}_{\mathbf{Y}}^{\top}[\widehat{\mathbf{U}}^{(t)\top}, \widehat{\mathbf{U}}^{(t)\top}, \widehat{\mathbf{U}}^{(t)\top}]^{\top}\Big)\Big)$$

4: **until** convergence

5: **Output:** $\widehat{\mathbf{U}}, \widehat{\mathbf{V}}$

We evaluate the estimation performance of the proposed method through numerical experiments. In the three-group example, the total sample size is $N = n_1 + n_2 + n_3$. For $i = 1, \ldots, n_1$, we generate $(\mathbf{x}_i, \mathbf{y}_i)$ jointly from a normal distribution with mean zero and covariance $\mathbf{\Sigma}_1$. For $i = (n_1 + 1), \ldots, (n_1 + n_2)$, we generate $(\mathbf{x}_i, \mathbf{y}_i)$ jointly from a normal distribution with mean zero and covariance $\mathbf{\Sigma}_2$. Lastly, for $i = (n_1 + n_2 + 1), \ldots, N$, we generate $(\mathbf{x}_i, \mathbf{y}_i)$ jointly from a normal distribution with mean zero and covariance $\mathbf{\Sigma}_3$ where

the covariance matrix for each group is as follows:

$$\mathbf{\Sigma}_z = \begin{pmatrix} \mathbf{\Sigma_X} & \rho_z \mathbf{U}\mathbf{V}^\top \\ \rho_z \mathbf{V}\mathbf{U}^\top & \mathbf{\Sigma_Y} \end{pmatrix}, \quad z = 1, 2, 3,$$

with the group index z represents the categorical variable $Z \in \{1, 2, 3\}$. The $\mathbf{U}$ and $\mathbf{V}$ under rank-1 scenario are the same as in Section 3. We set $p_1 = 18$, $p_2 = 15$, $s_1 = 3$, $s_2 = 3$, $n_1 = n_2 = n_3 \in \{20, 200\}$, $c_1/c_2 \in \{0.5, 3\}$ and the group difference $(\rho_1, \rho_2, \rho_3) \in \{(0.9, 0.1, -0.9), (0.9, -0.4, -0.5)\}$. Table S4 presents the results for the multi-category Z case over 100 replicated datasets.

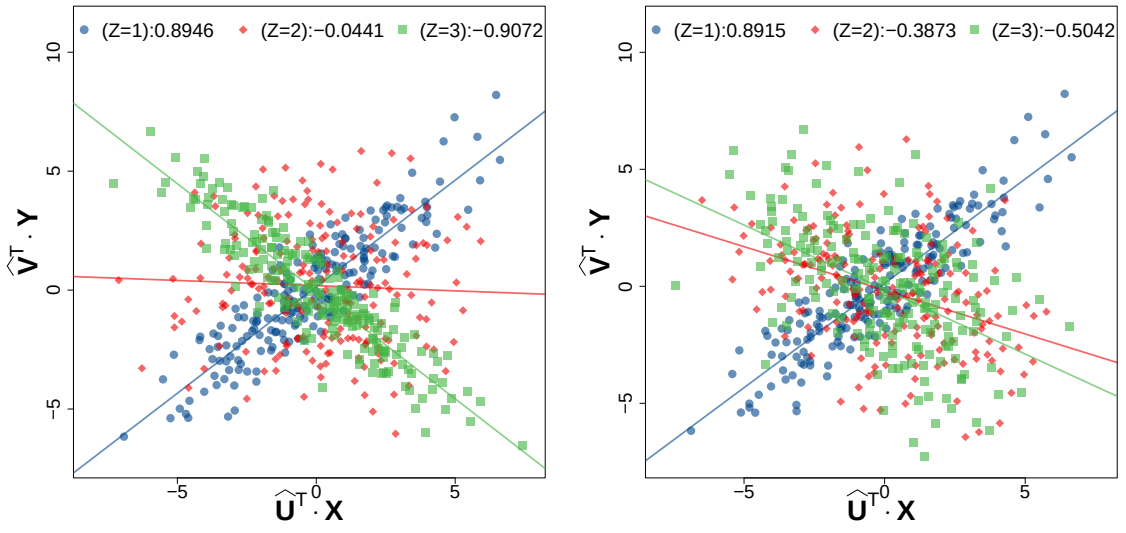
Across both signal regimes (strong and weak), the CCR model demonstrates strong variable selection performance, achieving high TPR and low FPR for both $\mathbf{X}$ and $\mathbf{Y}$. As expected, estimation accuracy improves with increasing sample size and is substantially enhanced under the strong signal setting. In particular, subspace distances $\mathrm{D}_{\mathbf{U}}$ and $\mathrm{D}_{\mathbf{V}}$ decrease rapidly as n increases, with near-oracle recovery observed when $n_1 = n_2 = 200$.

We report the estimated canonical correlations $\widehat{\rho}_k = \mathrm{Corr}(\widehat{\mathbf{U}}^\top \mathbf{X}_k, \widehat{\mathbf{V}}^\top \mathbf{Y}_k)$, $k = 1, 2, 3$, rather than presenting $\widehat{\delta}_1$ and $\widehat{\eta}_1$ explicitly. The accurate recovery of $\widehat{\rho}_1$, $\widehat{\rho}_2$, and $\widehat{\rho}_3$ across both signal regimes indicates that the proposed method effectively captures group-specific dependence structures, with particularly strong performance under moderate to strong signal strength.

We present the linear combinations of $\mathbf{X}$ and $\mathbf{Y}$ for the multi-group case in Figure S1. The plot shows a clear separation among the groups—blue circles, red diamonds, and green squares—indicating that the CCR model effectively extends to multi-group settings by maximizing each pairwise difference.

Table S4: Numerical evaluations under rank-1 cross-covariances in different (ρ_1, ρ_2, ρ_3) over 100 data replicates, comparing strong signal ($c_1/c_2 = 3$) and weak signal ($c_1/c_2 = 0.5$) regimes. The numbers in parentheses report the standard errors of the subspace distances $\mathrm{D}_{\mathbf{U}}$, $\mathrm{D}_{\mathbf{V}}$ and the estimated canonical correlations. True sparsity levels ($s_1^* = s_1$, $s_2^* = s_2$) are used for estimation.

Rank ($r = 1$)	Strong signal ($c_1/c_2 = 3$)				Weak signal ($c_1/c_2 = 0.5$)			
	(ρ_1, ρ_2, ρ_3)				(ρ_1, ρ_2, ρ_3)			
	$(0.9, 0.1, -0.9)$		$(0.9, -0.4, -0.5)$		$(0.9, 0.1, -0.9)$		$(0.9, -0.4, -0.5)$	
$n_1 = n_2$	20	200	20	200	20	200	20	200
$\mathrm{TPR}_{\mathbf{X}}$	0.773	1.000	0.937	1.000	0.497	1.000	0.547	1.000
$\mathrm{TPR}_{\mathbf{Y}}$	0.717	1.000	0.907	1.000	0.530	1.000	0.553	1.000
$\mathrm{FPR}_{\mathbf{X}}$	0.045	0.000	0.013	0.000	0.101	0.000	0.091	0.000
$\mathrm{FPR}_{\mathbf{Y}}$	0.071	0.000	0.023	0.000	0.117	0.000	0.112	0.000
$\mathrm{D}_{\mathbf{U}}$	0.447 (0.034)	0.075 (0.004)	0.216 (0.024)	0.043 (0.002)	0.609 (0.043)	0.037 (0.002)	0.554 (0.044)	0.034 (0.002)
$\mathrm{D}_{\mathbf{V}}$	0.486 (0.036)	0.075 (0.004)	0.256 (0.027)	0.046 (0.003)	0.602 (0.042)	0.035 (0.002)	0.589 (0.042)	0.033 (0.002)
$\widehat{\rho}_1$	0.717 (0.031)	0.896 (0.001)	0.830 (0.021)	0.896 (0.001)	0.711 (0.021)	0.898 (0.001)	0.729 (0.021)	0.898 (0.001)
$\widehat{\rho}_2$	-0.116 (0.021)	0.098 (0.008)	-0.396 (0.019)	-0.403 (0.007)	-0.290 (0.031)	0.104 (0.007)	-0.525 (0.017)	-0.395 (0.006)
$\widehat{\rho}_3$	-0.638 (0.042)	-0.896 (0.001)	-0.445 (0.022)	-0.498 (0.005)	-0.713 (0.022)	-0.896 (0.001)	-0.566 (0.014)	-0.504 (0.006)



(a) $(\rho_1, \rho_2, \rho_3) = (0.9, 0.1, -0.9)$ (b) $(\rho_1, \rho_2, \rho_3) = (0.9, -0.4, -0.5)$

Figure S1: Linear combinations of the CCR model for multi-group case in different group-specific correlations $(\rho_1, \rho_2, \rho_3) \in \{(0.9, 0.1, -0.9), (0.9, -0.4, -0.5)\}$. The numbers in the legend represent the correlation between linear combinations by $Z \in \{1, 2, 3\}$. Sparsity levels are set as $\widehat{s}_1 = \widehat{s}_2 = 3$ and $n_1 = n_2 = n_3 = 200$.

S2. Comparison to Other Multivariate Association Method

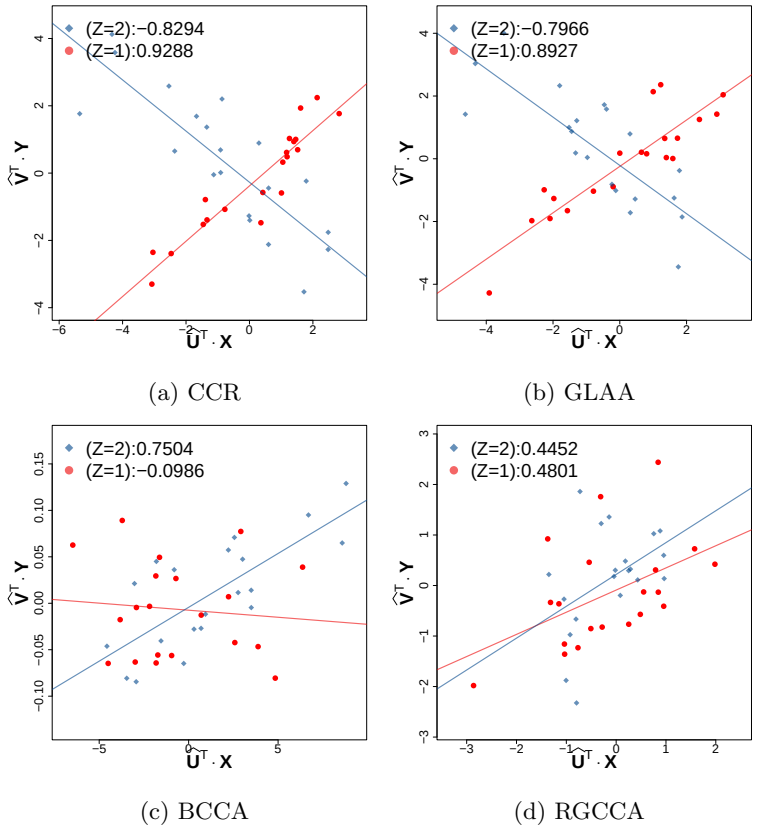
We compared the CCR model with other multivariate association methods to illustrate how the dynamic association between $\mathbf{X}$ and $\mathbf{Y}$ is affected by the binary variable $Z \in \{1, 2\}$. For the other multivariate association methods, we used a generalized liquid association analysis (GLAA; Li et al. 2023), Bayesian canonical correlation analysis (BCCA, Klami et al. 2013), and a regularized generalized canonical correlation analysis (RGCCA; Tenenhaus and Tenenhaus 2011).

Similar to the CCR model, GLAA proposed to specify the dynamic association between $\mathbf{X}$ and $\mathbf{Y}$ by differentiating the continuous $\mathbf{Z} \in \mathbb{R}^{p_3}$. Thus, the GLAA result depends on the smooth differentiation with respect to $\mathbf{Z}$. BCCA extracts the statistical dependencies (correlations) between $\mathbf{X}$ and $\mathbf{Y}$ but also decomposes $\mathbf{X}$ and $\mathbf{Y}$ into shared and data-specific components. By introducing a latent variable, BCCA specifies shared latent structure $\mathbf{C}$ and views specified latent variables $\mathbf{C_X}$ and $\mathbf{C_Y}$. And those latent variables allow flexibility in capturing association patterns that vary across different groups. See Klami et al. (2013) for further details. Lastly, RGCCA proposed to conduct multiblock analyses. For two datasets $\mathbf{X}$ and $\mathbf{Y}$, RGCCA can be applied as sparse canonical correlation analysis, which maximizes the canonical correlation between $\mathbf{X}$ and $\mathbf{Y}$ with sparsity on the canonical vectors.

Similar to the simulation setting in Section 3, we set $p_1 = 18$, $p_2 = 15$, $s_1^* = 3$, $s_2^* = 3$, $N = n_1 + n_2 = 20 + 20 = 40$. For data generation, $\rho_1 = 0.9$, $\rho_2 = -0.9$, $c_1 = 3$, and $c_2 = 1$ for data generation. We treated the binary variable as continuous to apply the GLAA and introduced sparsity only on $\mathbf{X}$ and $\mathbf{Y}$. For BCCA, we set the number of canonical vectors to 1 and others from the default setting in *R package "CCAGFA"*. For RGCCA, we tuned the sparsity parameter in the range between $\min\{p_1, p_2\}$ and one and used 0.33 as the sparsity parameter.

In Figure S2, we plotted the linear combinations of four methods on $\mathbf{X}$ and $\mathbf{Y}$ and marked

observations in different colors by group Z. The first three methods (CCR, GLAA, BCCA) extracted the dynamic association by Z. However, RGCCA cannot distinguish the difference between the group Z. The CCR model specifies the largest correlation difference of 1.76 among other methods. The GLAA has the same estimation metric proportional to $\mathbf{\Phi}$ in the CCR model. Thus, there is no significant difference in correlation between the CCR model and GLAA. However, since the GLAA is a penalized method, the result can be biased when estimating the canonical loadings. The CCR model reduces the bias through hard thresholding in the algorithm and estimates more accurate canonical loading vectors. The subspace distances $\mathrm{D}_{\mathbf{U}}$ and $\mathrm{D}_{\mathbf{V}}$ result in Table S5 also support the less biased result in the CCR model.



(a) CCR (b) GLAA

(c) BCCA (d) RGCCA

Figure S2: Comparison of dynamic association by binary variable $Z \in \{1, 2\}$ with other multivariate association methods. The methods under comparison are: conditional cross-covariance reduction (CCR) model, generalized liquid association analysis (GLAA), Bayesian canonical correlation analysis (BCCA), and regularized generalized canonical correlation analysis (RGCCA).

We replicated the data generations 100 times and compared the results of the other multivariate association methods in Table S5. We used estimated loading vectors as $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ to calculate the subspace distances $\mathrm{D}_{\mathbf{U}}$ and $\mathrm{D}_{\mathbf{V}}$. First, among the other methods, the CCR model has the best result on the variable selection. The CCR model and GLAA result select correct variables on $\mathbf{X}$ and $\mathbf{Y}$ (based on $\mathrm{TPR}_{\mathbf{X}}$, $\mathrm{TPR}_{\mathbf{Y}}$, $\mathrm{FPR}_{\mathbf{X}}$, and $\mathrm{FPR}_{\mathbf{Y}}$). In BCCA, Klami et al. (2013) noted that the elements of inactive components in the loading vectors are not forced exactly to zero but instead shrink toward minimal values under group-wise sparsity assumptions on the loading factors. They also introduced a group-wise spike-and-slab prior for stronger sparsity, but the corresponding implementation is not available. Therefore, we leave the TPRs and FPRs for BCCA blank in Table S5. Similar to Figure S2, the three methods (CCR, GLAA, BCCA) distinguish the group difference ($\widehat{\eta}_1$), with the CCR model achieving the smallest subspace distances.

Table S5: Numerical evaluations of the other multivariate association methods under rank-1 of cross-covariances with $c_1/c_2 = 3$ and $(\rho_1, \rho_2) = (0.9, -0.9)$ over 100 data replicates. The numbers in parentheses report the standard error of the subspace distances $\mathrm{D}_{\mathbf{U}}$, $\mathrm{D}_{\mathbf{V}}$, covariance difference $\widehat{\delta}_1$, and the associated correlation differences $\widehat{\eta}_1$. Here, we used the true sparsity levels ($s_1^* = s_1 = 3$, $s_2^* = s_2 = 3$) for estimation.

Method	n_1, n_2	$\mathrm{TPR}_{\mathbf{X}}$	$\mathrm{TPR}_{\mathbf{Y}}$	$\mathrm{FPR}_{\mathbf{X}}$	$\mathrm{FPR}_{\mathbf{Y}}$	$\mathrm{D}_{\mathbf{U}}$	$\mathrm{D}_{\mathbf{V}}$	$\widehat{\delta}_1$	$\widehat{\eta}_1$
CCR	20	1.000	1.000	0.000	0.000	0.093 (0.005)	0.098 (0.005)	12.580 (0.282)	1.797 (0.006)
	200	1.000	1.000	0.000	0.000	0.029 (0.002)	0.027 (0.002)	12.246 (0.084)	1.796 (0.002)
GLAA	20	1.000	1.000	0.273	0.291	0.322 (0.009)	0.299 (0.010)	12.476 (0.285)	1.796 (0.006)
	200	1.000	1.000	0.272	0.268	0.107 (0.003)	0.098 (0.002)	12.236 (0.084)	1.797 (0.002)
BCCA	20	–	–	–	–	0.496 (0.028)	0.530 (0.031)	11.655 (1.946)	1.260 (0.050)
	200	–	–	–	–	0.402 (0.037)	0.350 (0.036)	7.712 (1.090)	1.357 (0.051)
RGCCA	20	0.240	0.527	0.170	0.072	0.931 (0.014)	0.796 (0.014)	1.480 (0.205)	0.413 (0.043)
	200	0.158	0.491	0.158	0.049	0.936 (0.012)	0.738 (0.012)	1.484 (0.215)	0.467 (0.057)

S3. Simulation on the SPSS Method

We examine the accuracy of the SPSS method for selection of nonzero variables in $\mathbf{X}$ and $\mathbf{Y}$. The simulation setup is the same as in Section 3.1.

From the LTO data splits with $(n_1, n_2) = (10, 11)$, we randomly flip the signs of the differences D_ℓ, $\ell = 1, 2, \ldots, n_1 n_2$ in each permutation under the null hypothesis $\mathrm{H}_0 : \overline{\delta}_1^{(i,k)} - \overline{\delta}_1^{(i+1,k)} = 0$. We permute 1000 times to construct a reference distribution and compute the p-value.

Table S6: Accuracies of the sparsity levels s_1, s_2 from the permutation test at different signal strengths $c_1/c_2 \in \{3, 5, 7, 10\}$.

c_1/c_2	Sparsity level	Accuracy
3	s_1	0.69
	s_2	0.63
5	s_1	1
	s_2	0.88
7	s_1	1
	s_2	1
10	s_1	1
	s_2	1

S4. An Alternative Criterion for Variable Selection

For the algorithm of the CCR model, we need to input the sparsity levels s_1 and s_2. Here, we construct a BIC-type criterion to estimate the sparsity levels. Similar to AIC and BIC, we use minus maximization criterion modified by a penalty derived from the model size. Our proposed information criterion is

$$\mathrm{IC}\big\{s_1, s_2\big\} = -N \cdot \log\big\{\|\mathbf{P}_{\widehat{\mathbf{U}}}\widetilde{\mathbf{\Phi}}\mathbf{P}_{\widehat{\mathbf{V}}}\|_{\mathrm{F}}\big\} + (s_1 + s_2)\log(N).$$

The $\widetilde{\mathbf{\Phi}}$ is the mean the sample covariance difference, (s_1+s_2) is the number of free parameters, $N = n_1 + n_2$ is the number of data point, and $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ are estimated subspaces of the CCR model.

S4.1 *Consistency for the Criterion*

THEOREM 1: *For $\sqrt{n}$-consistent $\widetilde{\mathbf{\Phi}}$, $\widehat{\mathbf{U}}$, and $\widehat{\mathbf{V}}$, $Pr\{(\widehat{s}_1, \widehat{s}_2)_{IC} = (s_1, s_2)\} \to 1$ as $N \to \infty$.*

Proof. To facilitate the proof of Theorem 1, we introduce $\sqrt{n}$-consistency for $\mathbf{\Phi}$ and the subspaces $\mathbf{U}$ and $\mathbf{V}$ as follows.

PROPOSITION 1: By the Central Limit Theorem, $\widetilde{\mathbf{\Phi}}$ is a $\sqrt{n}$-consistent estimator for $\mathbf{\Phi}$. Therefore, the eigenvectors and eigenvalues of $\widetilde{\mathbf{\Phi}}$ are $\sqrt{n}$-consistent for the eigenvectors and eigenvalues of their population counterparts.

We assume the $\sqrt{n}$-consistency for $\widetilde{\mathbf{\Phi}}$, $\widehat{\mathbf{U}}$, $\widehat{\mathbf{V}}$. Let $J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, s_1, s_2) = -\log\left\{\|\mathbf{P}_{\widehat{\mathbf{U}}}\widetilde{\mathbf{\Phi}}\mathbf{P}_{\widehat{\mathbf{V}}}\|_{\mathrm{F}}\right\}$. In the application, we use grid search to determine the $(\widehat{s}_1, \widehat{s}_2)$. To guarantee our information criteria, we need to show that

$$\Pr\Big\{\mathrm{IC}_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{s}_1, \widehat{s}_2) - \mathrm{IC}_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, s_1, s_2) > 0\Big\} \to 1, \text{ as } N \to \infty$$

for the following cases.

$$\begin{cases} \text{(a) } 0 < s_1 \leqslant \widehat{s}_1,\ 0 < s_2 \leqslant \widehat{s}_2 \\ \text{(b) } 0 < \widehat{s}_1 < s_1,\ 0 < s_2 \leqslant \widehat{s}_2 \\ \text{(c) } 0 < s_1 \leqslant \widehat{s}_1,\ 0 < \widehat{s}_2 < s_2 \\ \text{(d) } 0 < \widehat{s}_1 < s_1,\ 0 < \widehat{s}_2 < s_2 \end{cases}$$

By definition of $\mathrm{IC}_N\{s_1, s_2\}$, we have

$$\begin{aligned} &\mathrm{IC}_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{s}_1, \widehat{s}_2) - \mathrm{IC}_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, s_1, s_2) \\ &\qquad = J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{s}_1, \widehat{s}_2) - J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, s_1, s_2) + \big\{(\widehat{s}_1 - s_1) + (\widehat{s}_2 - s_2)\big\}\frac{\log(N)}{N} \end{aligned} \quad (1)$$

First, in (a), by $\sqrt{n}$-consistency,

$$J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{s}_1, \widehat{s}_2) - J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, s_1, s_2) = J(\mathbf{U}, \mathbf{V}, \widehat{s}_1, \widehat{s}_2) - J(\mathbf{U}, \mathbf{V}, s_1, s_2) + O_p(N^{-1/2}).$$

For some $\widehat{s}_1$ and $\widehat{s}_2$, $J(\mathbf{U}, \mathbf{V}, \widehat{s}_1, \widehat{s}_2)$ converges in probability $J(\mathbf{U}, \mathbf{V}, s_1, s_2)$. Then,

$$J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{s}_1, \widehat{s}_2) - J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, s_1, s_2) = J(\mathbf{U}, \mathbf{V}, \widehat{s}_1, \widehat{s}_2) - J(\mathbf{U}, \mathbf{V}, s_1, s_2) + O_p(N^{-1/2}) = O_p(N^{-1/2}).$$

And, it follows from (1) that the dominant term in $\mathrm{IC}_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{s}_1, \widehat{s}_2) - \mathrm{IC}_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, s_1, s_2)$ is

$$\big\{(\widehat{s}_1 - s_1) + (\widehat{s}_2 - s_2)\big\} \cdot \frac{\log(N)}{N}, \text{ which is a positive number.}$$

Next, in (d), since $J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, s_1, s_2) < J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{s}_1, \widehat{s}_2)$, it is suffice to show that

$$J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{s}_1, \widehat{s}_2) - J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, s_1, s_2) = J(\mathbf{U}, \mathbf{V}, \widehat{s}_1, \widehat{s}_2) - J(\mathbf{U}, \mathbf{V}, s_1, s_2) + o_p(1)$$

where $J(\mathbf{U}, \mathbf{V}, s_1, s_2) < J(\mathbf{U}, \mathbf{V}, \widehat{s}_1, \widehat{s}_2) < 0$. We can decompose

$$J_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, i, j) = J(\mathbf{U}, \mathbf{V}, i, j) + o_p(1) \text{ for all } i = 1, \ldots, p_1, j = 1, \ldots, p_2$$

since we have $\sqrt{n}$-consistent $\widetilde{\mathbf{\Phi}}$, $\widehat{\mathbf{U}}$, $\widehat{\mathbf{V}}$ and $\mathbf{U}$,$\mathbf{V}$ affect the $J_N(\mathbf{U}, \mathbf{V})$ and $J(\mathbf{U}, \mathbf{V})$ only through span($\mathbf{U}$) and span($\mathbf{V}$). In addition, it is straightforward that the sum of two terms that both converge to zero at the same rate converges to zero at the same rate ($o_p(1) + o_p(1) = o_p(1)$).

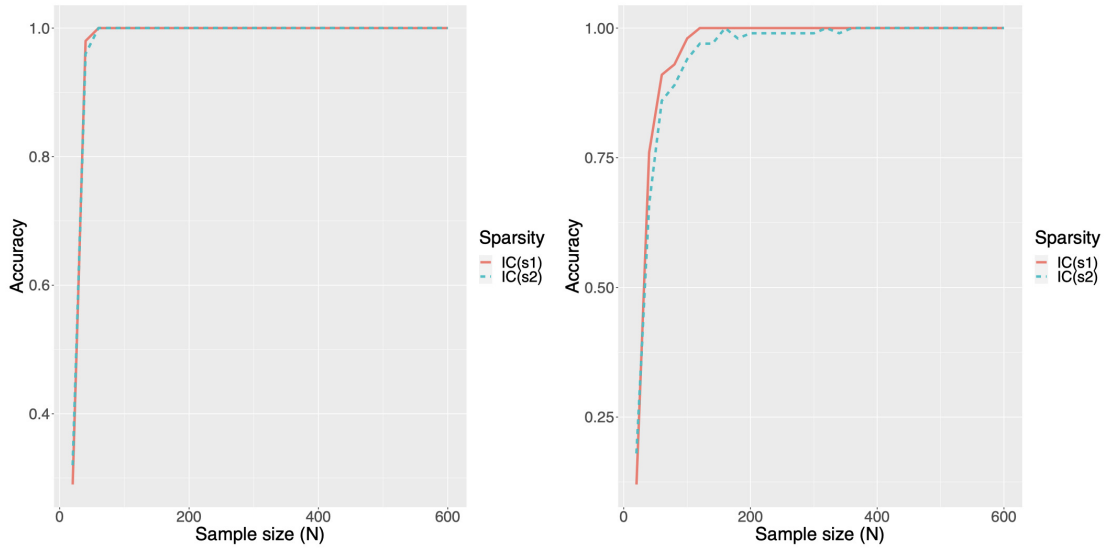
In (b), it is the same as in (d) when $s_1 > \widehat{s}_2$. Also, the steps are the same as in (a) when $s_1 < \widehat{s}_2$. Similarly, in (c), it is the same as in (a) when $\widehat{s}_1 > s_2$. And, the procedures are the same as in (d) when $\widehat{s}_1 < s_2$.

Therefore, $\Pr\Big\{\mathrm{IC}_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, \widehat{s}_1, \widehat{s}_2) - \mathrm{IC}_N(\widehat{\mathbf{U}}, \widehat{\mathbf{V}}, s_1, s_2) > 0\Big\} \to 1$ as $N \to \infty$.

S4.2 *Simulation on the Criterion*

In this section, we perform simulations of the information criterion to select sparsity levels s_1 and s_2 for the CCR model. First, we fix $p_1 = 15$, $p_2 = 15$ and set the true sparsity levels $s_1 = s_2 = 3$. We apply the same covariance structure in the simulation setup in Section 3.1. And we change the sample size $N = n_1 + n_2 \in \{20, 40, 60, \ldots, 600\}$. Thus, we have two multivariate variables $\mathbf{X} \in \mathbb{R}^{15}$, $\mathbf{Y} \in \mathbb{R}^{15}$ and each group has n_1 and n_2 observations and we estimate sparsity levels using the information criteria with 100 replicates. In Figure S3,

we label each sparsity in different colors. For example, IC(s1) denotes accuracy of s_1 in the information criterion. The criterion attains an accuracy of 1 after the sample size is larger than 60 (30 in each subgroup).



(a) Accuracies on (S1) (b) Accuracies on (S2)

Figure S3: Accuracies of the information criterion (IC) on each scenario with 100 replicates when fixing the true sparsity $s_1 = s_2 = 3$ and sample size $N \in \{20, 40, \ldots, 600\}$.

We explore the behavior of the information criterion with different sample sizes in Figure S4. We fix $\widehat{s}_1 = s_1 = 3$ and vary $\widehat{s}_2$ from 1 to 15 in each sample size N from 100 to 600. The information criterion achieves the lowest criterion value on $\widehat{s}_2 = 3$ (which is the true s_2) in all sample sizes. This also supports the criterion is reasonable to determine the number of selected variables in applications when the sample size is large enough ($N \geqslant 60$).

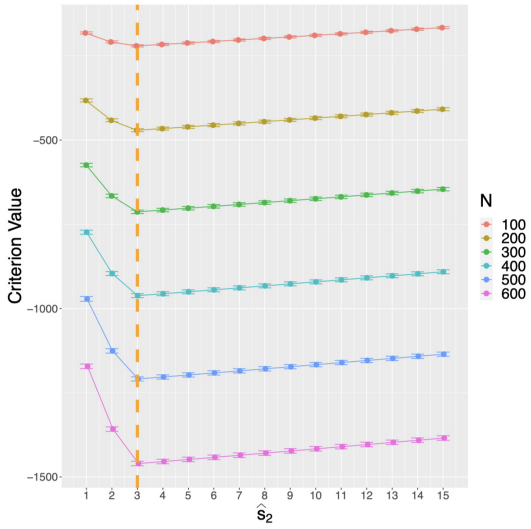


Figure S4: Accuracies of the information criterion (IC) with 100 replicates when fixing the true sparsity $s_1 = s_2 = 3$ and sample size $N \in \{20, 40, \ldots, 600\}$.

S5. Resampling Methods for the Sparsity Levels

We consider two different resampling methods to examine variable selection accuracy under the sparsity assumption. The first is bootstrap sampling. The second is the LTO method, in which one observation from $\mathbf{X}$ and one from $\mathbf{Y}$ are left out in each round, and the CCR model is applied to the remaining $N - 2$ observations.

We simulate data similar to rank-1 scenario in Section 3.1. We set $p_1 = 15$, $p_2 = 18$, $s_1 = 3$, $s_2 = 3$, $n_1 = 20$, and $n_2 = 20$. The ratios in Table S7 are the ratios of each variable selected by each resampling method over 1000 replicates with $(s_1, s_2) = (3, 3)$, that is, the CCR model selects three variables in $\mathbf{X}$ and $\mathbf{Y}$ in each round. The reference ratio is the ratio of variable selection when we randomly select three variables in each of $\mathbf{X}$ and $\mathbf{Y}$. The first three variables in $\mathbf{X}$ and $\mathbf{Y}$ have ratios close to 1, since the selected variables have meaningful signals in $\mathbf{\Sigma}_{\mathbf{X},1}$ and $\mathbf{\Sigma}_{\mathbf{Y},1}$. The result demonstrates that the CCR model successfully denoises the nonsignificant signals.

Table S7: Ratios of selected variables on the CCR model with $\mathbf{X}$ and $\mathbf{Y}$ on the strength of signal is $c_1/c_2 = 3$ where $n_1 = n_2 = 20$. The reference ratios represent ratios when we randomly select variables on sparsity levels s_1, s_2.

$\mathbf{X}$	Bootstrap				LTO	$\mathbf{Y}$	Bootstrap				LTO
	10	50	100	200			10	50	100	200	
1	**0.98**	**1.00**	**1.00**	**1.00**	**1.00**	1	**0.98**	**1.00**	**1.00**	**1.00**	**1.00**
2	**0.99**	**1.00**	**1.00**	**1.00**	**1.00**	2	**0.88**	**1.00**	**1.00**	**1.00**	**1.00**
3	**0.98**	**1.00**	**1.00**	**1.00**	**1.00**	3	**0.99**	**1.00**	**1.00**	**1.00**	**1.00**
4	0	0	0	0	0	4	0.01	0	0	0	0
5	0	0	0	0	0	5	0.02	0	0	0	0
6	0	0	0	0	0	6	0.01	0	0	0	0
7	0	0	0	0	0	7	0.01	0	0	0	0
8	0	0	0	0	0	8	0.01	0	0	0	0
9	0	0	0	0	0	9	0	0	0	0	0
10	0.01	0	0	0	0	10	0	0	0	0	0
11	0	0	0	0	0	11	0.04	0	0	0	0
12	0	0	0	0	0	12	0	0	0	0	0
13	0.01	0	0	0	0	13	0	0	0	0	0
14	0.01	0	0	0	0	14	0	0	0	0	0
15	0.02	0	0	0	0	15	0.01	0	0	0	0
						16	0.01	0	0	0	0
						17	0.02	0	0	0	0
						18	0.01	0	0	0	0
Reference ratio					0.20	Reference ratio					0.17

S6. Computational Complexity

For the computational complexity, let $\mathbf{X} \in \mathbb{R}^{N \times p_1}$ and $\mathbf{Y} \in \mathbb{R}^{N \times p_2}$ with total sample size $N = n_1 + n_2$. First, the computational complexity for calculating the cross-covariance is $\mathbf{\Sigma_{XY}}(z) \in \mathbb{R}^{p_1 \times p_2}$, the computational complexity is $\mathcal{O}(Np_1p_2)$. Then, we need to compute the singular value decomposition (SVD) of $\mathbf{\Phi} = \mathbf{\Sigma_{XY}}(1) - \mathbf{\Sigma_{XY}}(2) \in \mathbb{R}^{p_1 \times p_2}$ and the computational complexity is $\mathcal{O}(\min(p_1^2p_2, p_1p_2^2))$. Lastly, for a sparse SVD for variable selection, the computational complexity is $\mathcal{O}(p_1p_2r)$ when r is a reduced rank. Thus, the total complexity of the CCR model is $\mathcal{O}(Np_1p_2) + \mathcal{O}(\min(p_1^2p_2, p_1p_2^2)) + \mathcal{O}(p_1p_2r)$. For small p_1 and p_2, this simplifies to $\mathcal{O}(Np_1p_2)$, as matrix multiplications and SVD dominate. When p_1 and p_2 are large ($\gg N$), these complexities approximate $\mathcal{O}(p_1^3)$ when $p_1 \approx p_2$.

S7. Real Data Analysis

S7.1 *Variables for Analysis of Skull and Temporalis Origin Muscle*

Table S8: Variables in skull and temporalis origin muscle (TO) with 10 male and 11 female subjects. The skull has 21 subjects with 16 variables and the temporalis origin (TO) has 21 subjects with 18 variables. The asterisks represent the bilaterally measured variables.

	Attribute	Variable	Descriptions and measurement
Skull	Feature points	GnToGn	distance between left & right gonions
		AnL	distance between most anterior & posterior poles
		AnH	distance between most superior & inferior poles
		AnT	distance between most medial & lateral poles
		PlToPr	distance between the most lateral points of the roofs of the left(Pl) & right(Pr) ear canals
		NaToPlPr	vertical distance from nasal bone suture to the line connecting left and right ear canals(Pl & Pr)
		CrToGn*	distance from coronoid process to gonion
	Ramus	Length*	distance from the highest point on the mandibular condyle to the gonion
		Width*	least width perpendicular to the ramus length
	Mandible	Length*	distance from the highest point on the mandibular condyle to the anterior margin of chin
		Angle*	angle formed by tangent between the lower border of the mandible and the posterior border of the ramus from the condyle to gonion
TO	Size	BoxLength*	length of 3D box on the symmetric plane
		BoxWidth*	width of 3D box on the symmetric plane
		BoxThickness*	thickness of 3D box parallel to symmetric plane
		Area*	actual muscle attachment surface
		Volume*	actual volume measured by determination kit
	Spatial orientation	SA*	angle between box plane and sagittal plane
		FA*	angle between box plane and frontal plane
		FHA*	angle between box plane and Frankfurt plane
	Location	Centroid Distance*	distance from origin to centroid of the muscle

S7.2 *Results on Skull and Temporalis Origin*

For additional biomechanical interpretation, we further display mandibular length (MandibleLength (R)) and attachment area (Area(L)) in (4) of the manuscript, which highlight differences in marginal covariance, as shown in Figure 5 of the main manuscript.

Figure S5 shows a scatter plot of mandibular length (MandibleLength (R)) and temporalis

origin (TO) attachment surface areas (Area (L)) with joint reaction force (JRF) magnitude on the raw scale (top), and a boxplot of JRF magnitude by sex (bottom). The JRF vector is calculated as the residual force at the TMJ required to maintain static equilibrium under estimated muscle forces during mandibular motion (She et al., 2021), and its magnitude is defined as the length of this vector. A larger JRF corresponds to greater loading on the TMJ. For the same level of bite force, JRF increases as 3D mandibular length decreases (Sun et al., 2024).

In the raw-scale plot shown in Figure S5, females with shorter mandibular length (MandibleLength (R)) and smaller attachment areas (Area (L)) exhibit larger force magnitudes (top), and the joint reaction force in females is substantially greater than in males (bottom). Notably, our analysis using the CCR model indicates that females tend to have a shorter mandibular length (MandibleLength (R)) as the selected size-related variables in TO increase, while other selected skull variables are held fixed. This finding may help identify a high-risk subgroup of females for TMD, and the association appears faint in the raw-scale plot, which only reflects pairwise variable relationships, as shown in Figure S5.

Thus, the CCR model provides new insights into the association between the linear combinations of the skull and muscle attachment. This suggests a possible future research direction focusing on abnormal or high-risk samples to further investigate the association between bone structure and muscle measurements.

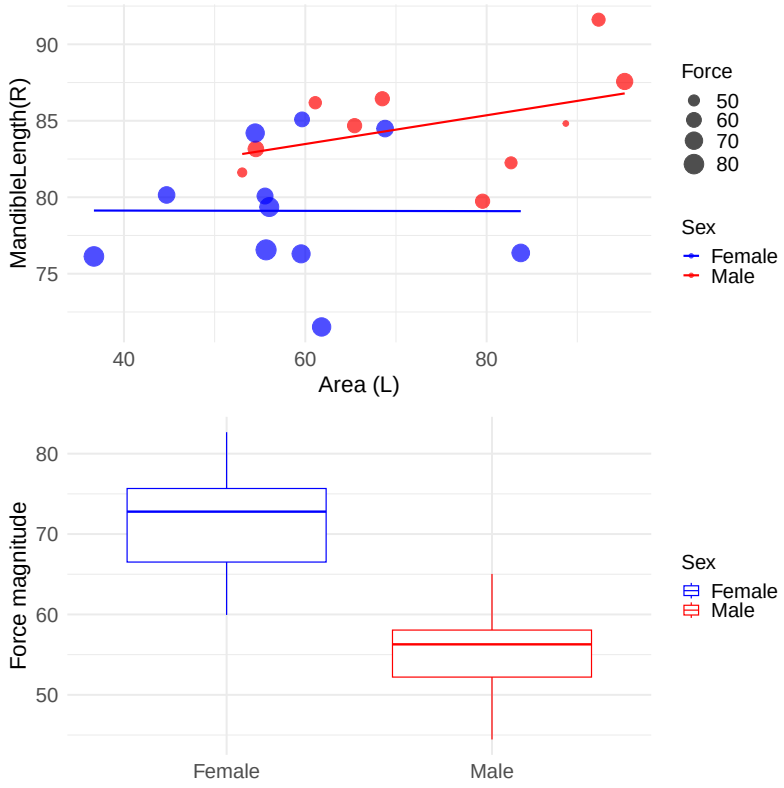


Figure S5: Plots of mandibular length (MandibleLength(R)), and the left temporalis origin surface areas (Area (L)), and joint reaction force magnitude in raw-scale measurements.

S7.3 *Results from the Marginally Standardized Skull and Temporalis Origin*

This section presents CCR model results with marginally standardized skull and temporalis origin (TO) muscle measurements. Marginal standardization refers to standardizing within each group of Z. We use $(s_1, s_2) = (3, 3)$ based on the results of the SPSS method. The standardized vectors and datasets are denoted with the subscript s.

The estimated maximal covariance difference and associated correlation differences are $\widehat{\delta}_s = 1.96$ and $\widehat{\eta}_s = 1.11$, respectively. In (2), we can see that left and right mandible lengths are selected on the standardized skull, and the angle variables (SA, FA) are selected on the standardized TO. The result is different from the non-standardized results in Section 4 since the marginal standardization removed the scale effects. However, since we could not obtain biomechanical evidence supporting these linear combinations in (2), only the non-

standardized results are included in Section 4.

$$\begin{aligned}\text{Skull: } \widehat{\mathbf{U}}_s^\top \mathbf{X}_s &= 0.443\ \ \text{Mandible.Length(L)} + 0.524\ \ \text{Ramus.Width(R)} \\ &\quad + 0.726\ \ \text{Mandible.Length(R)} \\ \text{TO: } \widehat{\mathbf{V}}_s^\top \mathbf{Y}_s &= 0.387\ \ \text{BoxThickness(L)} - 0.602\ \ \text{SA(L)} + 0.697\ \ \text{FA(L)} \end{aligned} \tag{2}$$

Figure S6 displays the result of the CCR model in plots of the linear combinations, and the corresponding graphical example of the standardized skull and TO muscle are displayed in Figure S7.

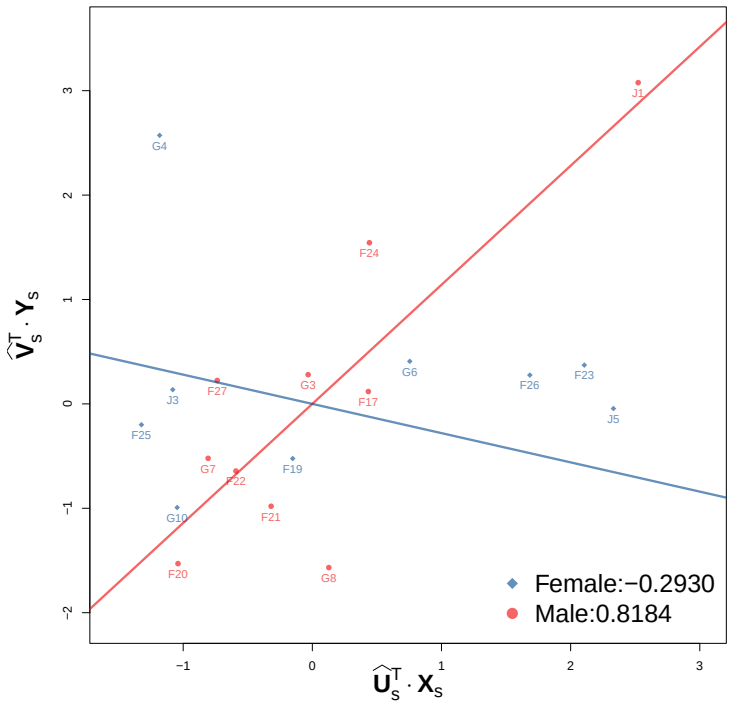


Figure S6: Linear combinations of the CCR model on the marginally standardized skull and temporalis origin (TO) measurements. The numbers in the legend represent the correlations between linear combinations by sex. Sparsity levels are set as $s_1 = s_2 = 3$. The x-axis ($\widehat{\mathbf{U}}_s^\top \mathbf{X}_s$) indicates the linear combination on the standardized skull. The y-axis ($\widehat{\mathbf{V}}_s^\top \mathbf{Y}_s$) represents the linear combination on the standardized temporalis origin (TO).

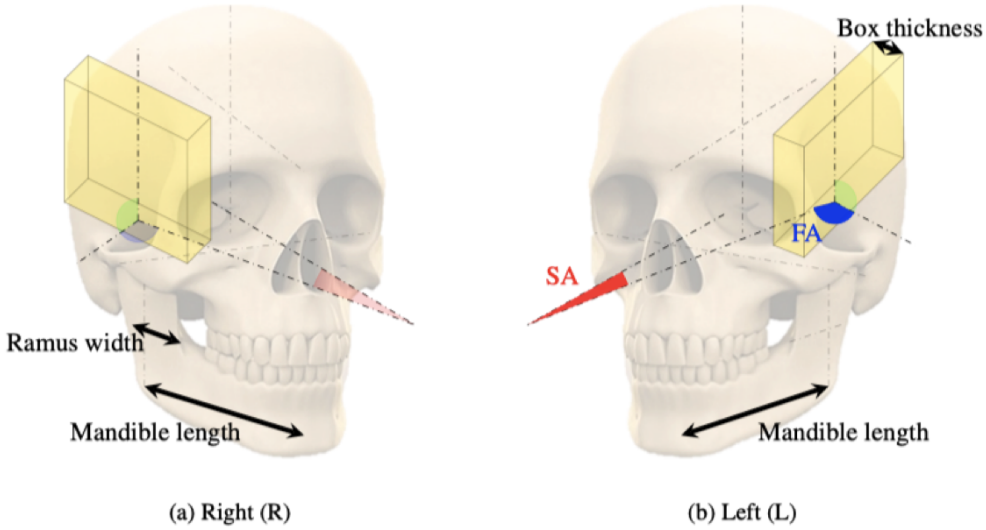


Figure S7: Result of the CCR model on the marginally standardized skull and temporalis origin (TO) under $(s_1, s_2) = (3, 3)$ where selected 6 variables are denoted with variable names.

S7.4 *Results from Other Multivariate Association methods*

Table S9 presents the real-data analysis results from Section 4 (Skull vs. TO muscle), including the estimated correlations, associated correlation difference $(\widehat{\eta}_1)$, and maximal covariance difference $(\widehat{\delta}_1)$ for the proposed CCR, GLAA, and other CCA-based methods (RGCCA and BCCA). Due to the limited sample size in the real-data application (overall $n = 21$), exhaustive cross-validation-based tuning of sparsity parameters may lead to unstable selections across methods. Therefore, we considered a moderate sparsity range that yields comparable sparsity levels across other multivariate association methods (approximately 5–7 selected variables), allowing for a more stable and interpretable comparison of the extracted association structures.

We first note that the maximum covariance difference $(\widehat{\delta}_1)$ under the CCR objective without sparsity constraints is 145.393, indicating that the sparse CCR solution captures a substantial proportion of the covariance contrast represented by the unconstrained CCR formulation. In Table S9, we observe that the estimated correlations and contrast measures vary across

methods, indicating that different methods capture different aspects of the association structure. We also note that the larger estimated $\widehat{\delta}_1$ value for GLAA reflects the larger number of selected variables. In particular, the CCR model selected $(\widehat{s}_1, \widehat{s}_2) = (5, 5)$ variables from $(\mathbf{X}, \mathbf{Y})$, whereas the GLAA selected $(\widehat{s}_1, \widehat{s}_2) = (7, 5)$ variables. Since the covariance contrast is influenced by the scale and dimensionality of the selected representations, methods selecting more variables may naturally yield larger estimated covariance differences.

Among the compared methods, the CCR approach more clearly identifies opposite directions of association between the two groups, which is consistent with the modeling assumption of group-specific cross-covariance structures. This suggests that the CCR method may provide a more interpretable representation of group-specific association patterns in this application, whereas other multivariate association methods do not clearly capture this contrast.

We note that, in real data applications where the true parameters are unknown, a larger $\widehat{\delta}_1$ does not necessarily indicate better performance, as different methods may yield estimates with varying levels of bias or variability. These values should therefore be interpreted with caution.

Table S9: Estimated first two correlations ($\widehat{\rho}_1$, $\widehat{\rho}_2$), associated correlation difference ($\widehat{\eta}_1$), and maximal covariance difference ($\widehat{\delta}_1$) for the skull vs. temporalis origin (TO) analysis across different methods. Since the true parameter values are unknown in real data, these estimates are presented for descriptive comparison rather than as a direct measure of method performance.

	CCR	GLAA	RGCCA	BCCA
$\widehat{\rho}_1$	0.539	0.746	0.877	0.792
$\widehat{\rho}_2$	-0.620	-0.299	0.682	0.577
$\widehat{\eta}_1$	1.159	1.044	0.196	0.215
$\widehat{\delta}_1$	119.649	127.710	21.315	19.417

References

Klami, A., Virtanen, S., and Kaski, S. (2013). Bayesian canonical correlation analysis. *The Journal of Machine Learning Research* **14,** 965–1003.

Li, L., Zeng, J., and Zhang, X. (2023). Generalized liquid association analysis for multimodal data integration. *Journal of the American Statistical Association* **118,** 1984–1996.

She, X., Sun, S., Damon, B. J., Hill, C. N., Coombs, M. C., Wei, F., et al. (2021). Sexual dimorphisms in three-dimensional masticatory muscle attachment morphometry regulates temporomandibular joint mechanics. *Journal of Biomechanics* **126,** 110623.

Sun, S., Xu, P., Buchweitz, N., Hill, C. N., Ahmadi, F., Wilson, M. B., et al. (2024). Explainable deep learning and biomechanical modeling for tmj disorder morphological risk factors. *JCI insight* **9,**.

Tenenhaus, A. and Tenenhaus, M. (2011). Regularized generalized canonical correlation analysis. *Psychometrika* **76,** 257–284.

Received September 2026.