BIOMETRICS 00, 1–20 DOI: 00.0000/a.0000-0000.0000.00000.a
September 2026

# Dimension Reduction for Characterizing Sexual Dimorphism in Biomechanics of the Temporomandibular Joint

**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: Sexual dimorphism is a critical factor in many biological and medical research fields. In biomechanics and bioengineering, understanding sex differences is crucial for studying musculoskeletal conditions such as temporomandibular disorder (TMD). This paper focuses on the association between the craniofacial skeletal morphology and temporomandibular joint (TMJ) related masticatory muscle attachments to discern sex differences. Data were collected from 11 female and 10 male cadaver heads to investigate sex-specific relationships between the skull and muscles. We propose a conditional cross-covariance reduction (CCR) model, designed to examine the dynamic association between two sets of random variables conditioned on a third binary variable (e.g., sex), highlighting the most distinctive sex-related relationships between skull and muscle attachments in the human cadaver data. Under the CCR model, we employ a sparse singular value decomposition algorithm. We further introduce a sequential permutation for selecting sparsity (SPSS) method to identify important variables and determine the optimal number of selected variables.

KEY WORDS: liquid association; multimodal data integration; permutation test; sparse singular value decomposition; variable selection

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

## 1. INTRODUCTION

Sexual dimorphism significantly influences human skull morphology and biomechanics, shaping our understanding of conditions like temporomandibular disorder (TMD). TMD affects $5 - 12\%$ of Americans, with an estimated annual cost of approximately \$4 billion (Stowell et al., 2007). The relationship between temporomandibular joint (TMJ) muscle attachments and skull features is central to TMJ function and TMD development. TMD is multifactorial, with the morphology of the masticatory system contributing significantly to its development. Clinical studies have shown that women are more likely than men to develop TMD, with reported prevalence ratios ranging from 3:1 to 8:1 (Slade et al., 2011). Existing results in TMJ research are often reduced to summary statistics or rely on one-variable-at-a-time approaches (Coombs et al., 2019; She et al., 2021), which lack statistical efficiency and may obscure important associations. By integrating multimodal data from skull and muscle measurements—more accurately, craniofacial skeletal morphology and masticatory muscle attachment measurements—this work aims to provide more precise insights into TMJ biomechanics and craniofacial disorders.

### 1.1 *Motivating Data*

Motivated by a recent study on TMJ muscle attachment morphometry and musculoskeletal characterization (She et al., 2018), we propose an integrative analysis framework that leverages the multivariate structure of multimodal craniofacial data. This approach is designed to identify associations that are highly sensitive to subject-level heterogeneity, explicitly accounting for sex differences. In the study by She et al. (2018), human cadavers and a custom surgical probe were used to quantify three-dimensional muscle attachment morphology. These data were then combined with cone beam computed tomography (CBCT) scans to explore their relationship with musculoskeletal modeling of the TMJ. The resulting dataset represents a valuable resource, as complete muscle shapes and orientations are essential

for accurately characterizing TMJ biomechanics—yet such information cannot be directly obtained through current imaging technologies. The proposed integrative analysis aims to elucidate how different data modalities, such as muscle and skeletal measurements, interact and associate when conditioned on sex.

Data were obtained from 21 cadaver heads (11 females, $73.6 \pm 12.8$ years; 10 males, $75.8 \pm 8.3$ years) without craniofacial abnormalities or TMD, as described in She et al. (2018). CBCT (voxel size $0.2 \times 0.2 \times 0.2\ mm^3$) was used to reconstruct 3D craniofacial models, and dissections identified muscle attachment sites. After scanning with CBCT, solid 3D models of each head were reconstructed. Craniofacial anthropometric dimensions were measured from reconstructed 3D solid models of cadaver heads. TMJ muscle attachment morphometry was quantified using a co-registered CBCT and 3D digitization method (She et al., 2021). A bounding-box approach defined attachment size (length, width, thickness, area), centroid coordinates, and orientation relative to anatomical planes. Measurements were made on eight TMJ muscle attachments. We focus on the temporalis origin (TO), which is critical for load-bearing tasks such as biting and chewing (Hylander and Johnson, 1985). The variables of the TO and skull are summarized in Table S8 of the Supplementary Materials.

### 1.2 *Statistical Problem Formulation and Related Work*

In our TMJ data analysis, let $\mathbf{X} \in \mathbb{R}^{p_1}$ denote the skull characteristic measurements extracted from the CBCT scans and $\mathbf{Y} \in \mathbb{R}^{p_2}$ be the muscle attachment measurements. Then the problem can be statistically formulated as studying the relationship between $\mathbf{X}$ and $\mathbf{Y}$ conditional on sex variable $Z \in \{1, 2\}$. Despite rich statistical literature studying the associations between two sets of multivariate variables, notably the canonical correlation analysis (CCA) and its variants (Li and Gaynanova, 2018; Mai and Zhang, 2019; Shu et al., 2020; Wang et al., 2015; Witten et al., 2009), modeling their interrelationship conditional on a third set of variables is an important new research frontier in modern multivariate

analysis. Here, statistical models and methods are needed for finding features of $\mathbf{X}$ and $\mathbf{Y}$ whose associations differ by sex. Existing CCA-based approaches, in particular, aim to estimate a common pattern across sex groups. The proposed analysis framework therefore addresses this key limitation of CCA-based approaches and enables the discovery of new insights from complex craniofacial data.

In the multivariate analysis literature, dynamic association defined by conditioning on a third variable has been previously proposed. Most relatedly, liquid association is a concept originally proposed by Li (2002) to capture dynamic co-expression in two gene expression profiles given a third gene. Liquid association quantifies the evolving dependence structure between two univariate random variables by incorporating a third variable and measuring a three-way interaction. Extensions along this line of work have been developed over the years (Chen et al., 2011; Ho et al., 2011; Li et al., 2004; Yu, 2018). More recently, Li et al. (2023) introduced the generalized liquid association analysis for high-dimensional settings with three sets of continuous multivariate variables. However, existing methods are designed specifically for continuous conditioning variables within the three-way interaction framework. While treating the binary $Z$ variable as continuous is numerically feasible—for example, applying the penalized tensor decomposition algorithm in Li et al. (2023) with minimal modifications—such an approach leads to ambiguous or questionable interpretations. In particular, both the original liquid association framework (Li, 2002) and its generalized extension (Li et al., 2023) quantify the expected derivative of the conditional association with respect to $Z$, implying smooth trends or continuous modulation. Binary variables, however, encode discrete group membership, and treating them as continuous can obscure group-based interpretations and lead to model misspecification. Our proposed conditional cross-covariance reduction (CCR) model and its estimation method are specifically designed

for binary $Z$ and consequently provide a justified approach and new interpretation within the liquid association literature in contexts such as sex dimorphism in the TMJ.

To investigate sex differences in TMJ mechanics, we propose a CCR model, which provides a simple interpretation and quantification that naturally leads to estimation of sparse linear combinations of TMJ skull and muscle measurements that maximize differences in association by sex. Given the small sample size of the cadaver dataset, traditional cross-validation and penalization methods are unsuitable for selecting the most important variables. To address this, we develop a sequential permutation for selecting sparsity (SPSS) method as a stable data-driven approach for variable selection.

This paper makes several contributions. First, we introduce an interpretable model for characterizing dynamic associations between two sets of variables conditioning on a third binary variable. Second, our estimation method incorporates variable selection through sparse singular value decomposition combined with hard thresholding, which helps reduce potential bias introduced by penalization. Third, the SPSS method enhances robust feature selection, particularly in small-sample settings. Finally, these contributions help provide stable and meaningful insights into sex-specific biomechanical interpretation in TMJ.

The rest of the paper is organized as follows. Section 2 introduces the CCR model and estimation procedures with the SPSS method for variable selection. We numerically show the CCR results and accuracy of the SPSS method in Section 3 and illustrate the TMJ analysis results in Section 4. We conclude with a brief discussion in Section 5. Supplementary Materials include additional numerical results and extensions.

## 2. METHODOLOGY

### 2.1 *Conditional Cross-covariance Reduction Model*

We first introduce the concept of a conditional cross-covariance reduction (CCR) model and tailor it to the case of the binary third variable (e.g., sex). The CCR model accommodates both continuous and discrete third variables, but our development focuses on the discrete case to facilitate our goal of identifying sexual dimorphism.

Let the two sets of random variables be $\mathbf{X} \in \mathbb{R}^{p_1}$ and $\mathbf{Y} \in \mathbb{R}^{p_2}$, and the third random variable $Z \in \mathbb{R}$. Then the conditional cross covariance, which summarizes important aspects of the relationship between $\mathbf{X}$ and $\mathbf{Y}$, can be formulated as $\text{cov}(\mathbf{X}, \mathbf{Y} \mid Z = z) \equiv \mathbf{\Sigma_{XY}}(z) \in \mathbb{R}^{p_1 \times p_2}$. Our CCR model assumes that the matrix $\mathbf{\Sigma_{XY}}(z)$ varies within low-dimensional subspaces for all values of $z$ as follows:

$$\mathbf{\Sigma_{XY}}(z) = \text{cov}(\mathbf{X}, \mathbf{Y} \mid Z = z) = \mathbf{\Gamma}_1 f(z) \mathbf{\Gamma}_2^\top \in \mathbb{R}^{p_1 \times p_2}, \tag{1}$$

for some semi-orthogonal basis matrices $\mathbf{\Gamma}_1 \in \mathbb{R}^{p_1 \times d_1}$ and $\mathbf{\Gamma}_2 \in \mathbb{R}^{p_2 \times d_2}$, and some latent function $f : \mathbb{R} \mapsto \mathbb{R}^{d_1 \times d_2}$. The latent matrix-variate function $f(z) \in \mathbb{R}^{d_1 \times d_2}$, where $d_1 \leqslant p_1$ and $d_2 \leqslant p_2$, is what drives the dynamic covariance between $\mathbf{X}$ and $\mathbf{Y}$. We note that the column spans of $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are identifiable, although the basis matrices themselves are determined only up to right multiplication by an orthogonal matrix. Under the CCR model (1), the linear combinations $\mathbf{\Gamma}_1^\top \mathbf{X}$ and $\mathbf{\Gamma}_2^\top \mathbf{Y}$ capture associations in $\mathbf{X}$ and $\mathbf{Y}$ that vary with $z$. The function $f(\cdot)$ contains the coordinates of the conditional cross-covariance $\mathbf{\Sigma_{XY}}(z)$ relative to $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$. Thus, the important signals in the rows and columns are preserved by $\text{span}(\mathbf{\Gamma}_1)$ and $\text{span}(\mathbf{\Gamma}_2)$, respectively.

When the third variable is binary $Z \in \{1, 2\}$, the CCR model implies that the variation in $\mathbf{\Sigma_{XY}}(z)$ along $z$ is fully characterized by the non-stochastic matrix $\mathbf{\Phi} = \mathbf{\Sigma_{XY}}(1) - \mathbf{\Sigma_{XY}}(2)$. Thus, in the binary-$Z$ setting, "dynamics" correspond to a two-regime contrast, and our framework recovers this contrast as a special case of the general dynamic association model.

Furthermore, the singular value decomposition (SVD) of $\mathbf{\Phi}$ implies that the dimensions of the latent subspaces have to be $d_1 = d_2 = r$ for some integer $r$. Then, we have $\mathbf{\Phi} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ where $\mathbf{U} \in \mathbb{R}^{p_1 \times r}$, $\mathbf{V} \in \mathbb{R}^{p_2 \times r}$ are orthonormal basis matrices and $\mathbf{D} \in \mathbb{R}^{r \times r}$ is a diagonal matrix. The rank $r$ is a pre-specified value. In practice, we may take the rank as 1 or 2 for exploratory analysis and data visualization. The rank selection is still an open question in low-rank matrix approximation, with many ad-hoc approaches proposed in the matrix decomposition literature, and is beyond the scope of this paper.

## 2.2 *Subspace Estimation*

In the CCR model, association patterns in $\mathbf{X}$ and $\mathbf{Y}$ that are affected by $Z$ can be represented through linear combinations of $\mathbf{U}^\top\mathbf{X}$ and $\mathbf{V}^\top\mathbf{Y}$. We estimate the subspace span($\mathbf{U}$) spanned by the columns of $\mathbf{U}$ and the subspace span($\mathbf{V}$) spanned by the columns of $\mathbf{V}$. For $N$ i.i.d. observations $\{\mathbf{x}_i, \mathbf{y}_i, z_i, i = 1, \ldots, N\}$, let the first $n_1$ observations have $z_i = 1$ and the remaining $n_2 = N - n_1$ observations have $z_i = 2$. We center the data within each group because we are interested in the conditional cross-covariance and not the conditional means. Thus, $\sum_{i=1}^{n_1} \mathbf{x}_i = \sum_{i=n_1+1}^{N} \mathbf{x}_i = \mathbf{0}$ and $\sum_{i=1}^{n_1} \mathbf{y}_i = \sum_{i=n_1+1}^{N} \mathbf{y}_i = \mathbf{0}$. We estimate $\mathbf{U}$ and $\mathbf{V}$ by:

$$
\begin{aligned}
(\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}}) &= \arg\max_{\mathbf{U},\mathbf{V}} \left\| \widehat{\mathrm{cov}}(\mathbf{U}^\top\mathbf{X}, \mathbf{V}^\top\mathbf{Y} \mid Z = 1) - \widehat{\mathrm{cov}}(\mathbf{U}^\top\mathbf{X}, \mathbf{V}^\top\mathbf{Y} \mid Z = 2) \right\|_{\mathrm{F}}^2 \\
&= \arg\max_{\mathbf{U},\mathbf{V}} \|\mathbf{U}^\top \widehat{\mathbf{\Sigma}}_{\mathbf{XY}1} \mathbf{V} - \mathbf{U}^\top \widehat{\mathbf{\Sigma}}_{\mathbf{XY}2} \mathbf{V}\|_{\mathrm{F}}^2, \qquad (2)
\end{aligned}
$$

where $\widehat{\mathbf{\Sigma}}_{\mathbf{XY}1} = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{x}_i \mathbf{y}_i^\top$ and $\widehat{\mathbf{\Sigma}}_{\mathbf{XY}2} = \frac{1}{n_2} \sum_{i=n_1+1}^{N} \mathbf{x}_i \mathbf{y}_i^\top$, and $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. Defining $\widetilde{\mathbf{\Phi}} = \widehat{\mathbf{\Sigma}}_{\mathbf{XY}1} - \widehat{\mathbf{\Sigma}}_{\mathbf{XY}2}$, the optimization problem reduces to maximizing $\|\mathbf{U}^\top \widetilde{\mathbf{\Phi}} \mathbf{V}\|_{\mathrm{F}}^2$ under orthogonality constraints. Then the resulting $(\widetilde{\mathbf{U}}, \widetilde{\mathbf{V}})$ are the left and right singular vectors of $\widetilde{\mathbf{\Phi}}$, provided that the leading singular subspaces are identifiable. To enhance interpretability and also to deal with the very small sample size in this paper, we next incorporate variable selection to further reduce the number of parameters.

### 2.3 *Variable Selection and Algorithm*

We consider sparsity on the singular vectors of $\mathbf{\Phi}$, which is achievable by many existing sparse SVD algorithms such as the iterative thresholding algorithm in Yang et al. (2016). In our CCR model, we pre-specify the sparsity levels as $s_1 \leqslant p_1$ and $s_2 \leqslant p_2$ according to the elements in $\mathbf{X}$ and $\mathbf{Y}$ that have dynamic association to each other instead of applying threshold tuning parameters iteratively in the sparse SVD algorithm. Thus, we identify the $s_1$ most important rows in $\mathbf{X}$ and the $s_2$ most important rows in $\mathbf{Y}$ at each iteration. Then we get the singular values having the largest $r$ components and the corresponding singular vectors with $s_1$ and $s_2$ non-zero components. The pre-specified sparsity levels $s_1$ and $s_2$ are enforced via hard thresholding, while the identified loadings are estimated without additional shrinkage. This mitigates the potential bias induced by penalization approaches. Moreover, in small-sample settings, tuning penalty parameters via cross-validation can be unstable. By directly controlling the sparsity levels, our method provides more stable and robust feature selection, consistent with the motivation of the SPSS procedure described in Section 2.5. This sparse estimation performs variable selection for $\mathbf{\Phi}$ since the estimated $\widehat{\mathbf{\Phi}}$ has the $s_1$ and $s_2$ variables most strongly tied to the patterns of association in $\mathbf{X}$ and $\mathbf{Y}$. Thus, we can effectively elucidate associations between modalities with maximal sex-specific differences while identifying a sparse set of driving variables.

We summarize the estimation procedure in Algorithm 1, which yields $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$, with the desired input sparsity levels $s_1$ and $s_2$. The procedure relies on two operators, row-wise truncation $\mathcal{T}_s(\cdot)$ and column-wise orthogonalization $\mathcal{Q}(\cdot)$. For a matrix $\mathbf{A} \in \mathbb{R}^{p\times r}$, let $\mathcal{I}_s(\mathbf{A}) \subseteq \{1,\ldots,p\}$, $s \leqslant p$, be the index set of the $s$ rows of $\mathbf{A}$ with the largest $\ell_2$-norm. We define the row-wise hard thresholding operator $\mathcal{T}_s(\cdot) : \mathbb{R}^{p\times r} \to \mathbb{R}^{p\times r}$ by $\big[\mathcal{T}_s(\mathbf{A})\big]_{ij} = A_{ij} I(i \in \mathcal{I}_s(\mathbf{A}))$. For a matrix $\mathbf{A} \in \mathbb{R}^{p\times r}$, define $\mathcal{Q}(\mathbf{A}) \in \mathbb{R}^{p\times r}$ as any matrix satisfying $\mathrm{span}\big(\mathcal{Q}(\mathbf{A})\big) = \mathrm{span}(\mathbf{A})$ and $\{\mathcal{Q}(\mathbf{A})\}^\top \mathcal{Q}(\mathbf{A}) = \mathbf{I}_r$. This can be obtained via the QR decomposition.

The iteration is initialized with $\widehat{\mathbf{U}}^{(0)}$ and $\widehat{\mathbf{V}}^{(0)}$, the top $r$ left and right singular vectors of $\widetilde{\mathbf{\Phi}}$. At each iteration, the algorithm updates these estimates by first computing the matrix products $\widetilde{\mathbf{\Phi}}\,\widehat{\mathbf{V}}^{(t-1)}$ and $\widetilde{\mathbf{\Phi}}^{\top}\widehat{\mathbf{U}}^{(t)}$, which extract the leading subspace directions. Row-wise hard thresholding is then applied via $\mathcal{T}_{s_1}(\cdot)$ and $\mathcal{T}_{s_2}(\cdot)$ to enforce sparsity, followed by orthonormalization through $\mathcal{Q}(\cdot)$ to obtain updated subspace estimates. Upon convergence, these retained rows represent the variables selected under the sparsity constraint that provide the linear combinations of $\mathbf{X}$ and $\mathbf{Y}$ that are most contrastive for the values of $Z$. Convergence is determined by a tolerance on the maximum subspace difference by $\max\left(\|\widehat{\mathbf{U}}^{(t)}\widehat{\mathbf{U}}^{(t)\top} - \widehat{\mathbf{U}}^{(t-1)}\widehat{\mathbf{U}}^{(t-1)\top}\|_{\mathrm{F}}^2\ , \|\widehat{\mathbf{V}}^{(t)}\widehat{\mathbf{V}}^{(t)\top} - \widehat{\mathbf{V}}^{(t-1)}\widehat{\mathbf{V}}^{(t-1)\top}\|_{\mathrm{F}}^2\right) \leqslant \epsilon$.

**Algorithm 1** CCR via two-way sparse power iteration

**Inputs:**

Sample estimate $\widetilde{\mathbf{\Phi}} \in \mathbb{R}^{p_1 \times p_2}$, rank $r \leqslant \min(p_1, p_2)$, sparsity levels $s_1 \leqslant p_1$, $s_2 \leqslant p_2$.

**Initialize:**

Compute top-$r$ singular vectors $(\widehat{\mathbf{U}}^{(0)}, \widehat{\mathbf{V}}^{(0)})$ of $\widetilde{\mathbf{\Phi}}$.

**repeat**

$\widehat{\mathbf{U}}^{(t)} \leftarrow \mathcal{Q}\Big(\mathcal{T}_{s_1}\Big(\widetilde{\mathbf{\Phi}}\,\widehat{\mathbf{V}}^{(t-1)}\Big)\Big)$

$\widehat{\mathbf{V}}^{(t)} \leftarrow \mathcal{Q}\Big(\mathcal{T}_{s_2}\Big(\widetilde{\mathbf{\Phi}}^{\top}\widehat{\mathbf{U}}^{(t)}\Big)\Big)$

**until** convergence

**Output:**

$\widehat{\mathbf{U}} = \widehat{\mathbf{U}}^{(t)}$, $\widehat{\mathbf{V}} = \widehat{\mathbf{V}}^{(t)}$, $\mathbf{P}_{\widehat{\mathbf{U}}} = \widehat{\mathbf{U}}\widehat{\mathbf{U}}^{\top}$, $\mathbf{P}_{\widehat{\mathbf{V}}} = \widehat{\mathbf{V}}\widehat{\mathbf{V}}^{\top}$, and $\widehat{\mathbf{\Phi}} = \mathbf{P}_{\widehat{\mathbf{U}}}\,\widetilde{\mathbf{\Phi}}\,\mathbf{P}_{\widehat{\mathbf{V}}}$.

### 2.4 *Covariance and Correlation Differences*

We define the *maximal covariance difference* as $\delta_i = \mathbf{U}_i^{\top}\mathbf{\Phi}\mathbf{V}_i$, where $\mathbf{U}_i$ and $\mathbf{V}_i$ are the $i$-th pair of singular vectors of $\mathbf{\Phi}$ for $i = 1, 2, \ldots, r$, and we adopt the sign convention that $\delta_i \geqslant 0$. When $r = 1$, $\delta_1$ is the maximal covariance difference that increases as we increase the sparsity parameters $s_1$ and $s_2$. We define the *associated correlation difference* $\eta_i$ as follows:

$$\eta_i = \mathrm{corr}(\mathbf{U}_i^{\top}\mathbf{X}, \mathbf{V}_i^{\top}\mathbf{Y} \mid Z = 1) - \mathrm{corr}(\mathbf{U}_i^{\top}\mathbf{X}, \mathbf{V}_i^{\top}\mathbf{Y} \mid Z = 2) \in \mathbb{R},\ i = 1, \ldots, r,$$

where by "associated" we mean that the vectors $\mathbf{U}_i$ and $\mathbf{V}_i$ are defined from the maximizing association differences (in covariance scales). It is difficult to simultaneously optimize the subspaces for the difference of two canonical correlation forms in terms of the associated correlation difference $\eta_i$. Thus, the correlation difference is calculated with the subspaces $\mathbf{U}$ and $\mathbf{V}$ that are estimated from the maximization problem in (2). Even if we marginally standardize $\mathbf{X}$ and $\mathbf{Y}$, we still maximize the marginally standardized form of $\delta_i$, not $\eta_i$. We use the associated correlation difference $\eta_i$ to demonstrate that the proposed CCR model avoids the masking of the association between $\mathbf{X}$ and $\mathbf{Y}$ by $Z$.

### 2.5 *Sequential Permutation for Selecting Sparsity*

Algorithm 1 requires the sparsity levels $s_1$, $s_2$. In Supplementary Materials S4, we introduce an information criterion and illustrate its consistency in selecting $s_1$ and $s_2$. However, due to the limited size of the cadaver dataset, we find that using either information criteria or cross-validation to select the sparsity levels can be inaccurate. Instead, we devise a sequential permutation for selecting sparsity (SPSS) approach to select $s_1$ and $s_2$ separately. We employ a leave-two-out (LTO) resampling scheme that iteratively removes one observation from each group $Z \in \{1, 2\}$ and fits the CCR model to the remaining $N-2$ samples. The SPSS approach considers hypotheses related to the increment of the nuclear norm of $\widehat{\mathbf{\Phi}} = \mathbf{P}_{\widehat{\mathbf{U}}}\widetilde{\mathbf{\Phi}}\mathbf{P}_{\widehat{\mathbf{V}}}$ from the output of Algorithm 1. When $r = 1$, the nuclear norm of $\widehat{\mathbf{\Phi}}$ reduces to $\widehat{\delta}_1$.

When $r = 1$, we consider $s_1 = i$ sequentially for $i = 1, 2, \ldots, p_1 - 1$, and stop increasing $s_1$ once $s_1 = i + 1$ does not improve upon $s_1 = i$, in which case we set $s_1 = i$. Specifically, we stop increasing $s_1$ as soon as at least one null hypothesis $H_0^{i,k} : \overline{\delta}_1^{(i+1,k)} - \overline{\delta}_1^{(i,k)} = 0$ fails to be rejected; that is, moving from $i$ to $i + 1$ yields no significant improvement for some $k$, where $\overline{\delta}_1^{(i,k)}$ is the population counterpart of the sample algorithm's output $\widehat{\delta}_1$ at $(s_1, s_2) = (i, k)$.

The hypothesis test $H_0^{i,k}$ vs $H_1^{i,k}$ is performed using a permutation procedure. Let $\widetilde{\delta}_1^{(i,k)}$ be a sample counterpart of $\overline{\delta}_1^{(i,k)}$, and define $D_\ell$, $\ell = 1, \ldots, (n_1 n_2)$, as the differences $\widehat{\delta}_1^{(i+1,k)} -$

$\widehat{\delta}_1^{(i,k)}$ from $(n_1 n_2)$ LTO data splits. Then the observed mean difference is $\widetilde{\delta}_1^{(i+1,k)} - \widetilde{\delta}_1^{(i,k)} = (n_1 n_2)^{-1} \sum_{\ell=1}^{n_1 n_2} D_\ell$. To obtain a reference distribution under the null hypothesis, we use a large number of permutations—for example, 100,000 in our TMJ analysis in Section 4—where, in each permutation, the signs of the $D_\ell$ values are randomly flipped before averaging. The p-value, denoted $p^{(i,k)}$, is the proportion of permuted means at least as large as the observed mean difference $\widetilde{\delta}_1^{(i+1,k)} - \widetilde{\delta}_1^{(i,k)}$, which serves as the test statistic. Thus $s_1$ is set to the smallest $i$ such that the collection of p-values $p^{(i,k)}$, $k = 1, \ldots, p_2$, has at least one value larger than 0.05. An analogous procedure is used to determine $s_2$.

## 3. SIMULATION

### 3.1 *Simulation Setup*

We conduct simulation studies to evaluate the empirical performance of the proposed CCR model. We consider two scenarios for the rank of the cross-covariance, $r = 1$ and $r = 2$. We first set the rank of the cross-covariance of $\mathbf{X}$ and $\mathbf{Y}$ as $r = 1$, $p_1 = 18$, $p_2 = 15$, and true sparsities $s_1^* = s_2^* = 3$, and vary the sample sizes $N = n_1 + n_2$ from 40 to 400 when $n_1 = n_2$. Under the rank-1 scenario ($r = 1$), we get $\mathbf{\Phi}\mathbf{\Phi}^\top = (\rho_1 - \rho_2)^2$ where $\rho_1 - \rho_2$ is a coefficient of the SVD structure of $\mathbf{\Phi}$. That is, maximizing $\mathbf{\Phi}\mathbf{\Phi}^\top$ is equivalent to maximizing $(\rho_1 - \rho_2)^2$. Due to sign indeterminacy, we assume $\rho_1 - \rho_2 \geqslant 0$.

We generate the data in the following way. 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 $i = n_1 + 1, \ldots, N$, we generate $(\mathbf{x}_i, \mathbf{y}_i)$ jointly from a normal distribution with mean zero and covariance $\mathbf{\Sigma}_2$. Here,

$$\mathbf{\Sigma}_z = \begin{pmatrix} \mathbf{\Sigma}_\mathbf{X} & \rho_z \mathbf{U}\mathbf{V}^\top \\ \rho_z \mathbf{V}\mathbf{U}^\top & \mathbf{\Sigma}_\mathbf{Y} \end{pmatrix}, \quad z = 1, 2, \tag{3}$$

where the group index $z$ represents the binary variable $Z \in \{1, 2\}$. To maintain the positive-definiteness of the full covariance matrix and rank-1 condition for the cross-covariance of

$\mathbf{X}$ and $\mathbf{Y}$, we set $\mathbf{U} = \mathbf{\Sigma}_{\mathbf{X}}^{1/2}\mathbf{O}_1$ and $\mathbf{V} = \mathbf{\Sigma}_{\mathbf{Y}}^{1/2}\mathbf{O}_2$ where the $\mathbf{O}_1$ and $\mathbf{O}_2$ are unit length vectors $\mathbf{O}_1 = (1,1,1,0,\ldots,0)^\top/\sqrt{3} \in \mathbb{R}^{18\times 1}$ and $\mathbf{O}_2 = (1,1,1,0,\ldots,0)^\top/\sqrt{3} \in \mathbb{R}^{15\times 1}$. The columns of $\mathbf{U}$ and $\mathbf{V}$ span the subspaces capturing the variation in $\mathbf{X}$ and $\mathbf{Y}$, respectively. The CCR model $\mathbf{\Phi}$ is defined as $\mathbf{\Phi} = (\rho_1 - \rho_2)\mathbf{U}\mathbf{V}^\top$ where $\rho_1 - \rho_2 > 0$. For the covariance matrix $\mathbf{\Sigma}_z$, the marginal covariance matrix $\mathbf{\Sigma}_{\mathbf{X}}$ is set as a block diagonal matrix, $\mathbf{\Sigma}_{\mathbf{X}} = \text{bdiag}(c_1\mathbf{\Sigma}_{\mathbf{X},1}, c_2\mathbf{\Sigma}_{\mathbf{X},2})$, where $\mathbf{\Sigma}_{\mathbf{X},1} \in \mathbb{R}^{s_1^*\times s_1^*}$ corresponds to non-zero elements and takes the form of an autoregressive (AR) structure such that its $(i,j)$th entry equals $\sigma_{ij} = 0.7^{|i-j|}$, $i,j = 1,\ldots,s_1^*$, and $\mathbf{\Sigma}_{\mathbf{X},2} \in \mathbb{R}^{(p_1-s_1^*)\times(p_1-s_1^*)}$ is the identity matrix. The marginal covariance matrix $\mathbf{\Sigma}_{\mathbf{Y}}$ is constructed similarly. Therefore, the true signals in $\mathbf{\Sigma}_{\mathbf{X},1}$ and $\mathbf{\Sigma}_{\mathbf{Y},1}$ will be captured through Algorithm 1. We set $\rho_1 = 0.9$, $\rho_2 = -0.9$, $c_1 = 3$, and $c_2 = 1$. Note that the larger the ratio $c_1/c_2$, the easier it is to detect the true signals.

For the rank-2 simulation scenario ($r = 2$), we modify the cross-covariance in (3) as $\mathbf{U}\mathbf{D}\mathbf{V}^\top$, $z = 1,2$, where $\mathbf{D} = \text{diag}(\rho_{z1}, \rho_{z2})$, $\mathbf{U} = \mathbf{\Sigma}_{\mathbf{X}}^{1/2}\mathbf{O}_1$, and $\mathbf{V} = \mathbf{\Sigma}_{\mathbf{Y}}^{1/2}\mathbf{O}_2$. Here, $\mathbf{O}_1 \in \mathbb{R}^{p_1\times 2}$ and $\mathbf{O}_2 \in \mathbb{R}^{p_2\times 2}$, with the first column being $(1,1,1,0,\ldots,0)/\sqrt{3}$, and the second column being $(0,-1,1,0,\ldots,0)/\sqrt{2}$ under the true sparsity levels $(s_1^*, s_2^*) = (3,3)$. For the rank-2 scenario, we set $p_1 = 18$, $p_2 = 15$, $\rho_{11} = 0.9$, $\rho_{12} = 0.7$, $\rho_{21} = -0.9$, $\rho_{22} = -0.7$, $c_1 = 3$, $c_2 = 1$, and change the sample size $N = n_1 + n_2$ from 40 to 400 when $n_1 = n_2$. There are two contributing linear combinations on each of $\mathbf{X}$ and $\mathbf{Y}$ since $\mathbf{U} \in \mathbb{R}^{p_1\times 2}$ and $\mathbf{V} \in \mathbb{R}^{p_2\times 2}$.

To evaluate the performance of each method in terms of variable selection and subspace estimation accuracy, we used a true positive rate (TPR), a false positive rate (FPR), and a subspace distance. We record the TPR and FPR for each row (variables selected from $\mathbf{X}$) and column (variables selected from $\mathbf{Y}$). Let $\mathcal{I}_{\mathbf{U}} \subseteq \{1,2,\ldots,p_1\}$ the set of true nonzero rows of $\mathbf{U}$. The estimated index set is $\mathcal{I}_{\widehat{\mathbf{U}}} = \{i : \text{there exist non-zero elements on the } i\text{th row of } \widehat{\mathbf{U}}\}$. The TPR is defined as the proportion of correctly selected variables, $\text{TPR}_{\mathbf{X}} = |\mathcal{I}_{\mathbf{U}} \cap \mathcal{I}_{\widehat{\mathbf{U}}}|/s_1^*$, and the FPR as the proportion of falsely selected variables, $\text{FPR}_{\mathbf{X}} = |\mathcal{I}_{\mathbf{U}}^c \cap \mathcal{I}_{\widehat{\mathbf{U}}}|/(p_1 - s_1^*)$.

The definitions for $\text{TPR}_{\mathbf{Y}}$ and $\text{FPR}_{\mathbf{Y}}$ follow analogously by replacing $\mathbf{U}$, $\widehat{\mathbf{U}}$, $p_1$, $s_1^*$ with $\mathbf{V}$, $\widehat{\mathbf{V}}$, $p_2$, $s_2^*$. To assess the $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$, we compute the subspace distance between the true and estimated $\mathbf{U}$ and $\mathbf{V}$, $\text{D}_{\mathbf{U}} = \|\mathbf{P}_{\mathbf{U}} - \mathbf{P}_{\widehat{\mathbf{U}}}\|_{\text{F}}/\sqrt{2r}$ and $\text{D}_{\mathbf{V}} = \|\mathbf{P}_{\mathbf{V}} - \mathbf{P}_{\widehat{\mathbf{V}}}\|_{\text{F}}/\sqrt{2r}$ where $r$ denotes the true rank of the $\mathbf{\Phi}$ and $\|\cdot\|_{\text{F}}$ is the Frobenius norm. We compute the associated correlation differences for the first and second linear combinations as $\widehat{\eta}_1$ and $\widehat{\eta}_2$, respectively.

### 3.2 *Simulation Results*

We first show how the CCR model captures the true signals of the sparsity levels $s_1$ and $s_2$. In Algorithm 1, we set the tolerance level $\epsilon$ as $10^{-11}$ and provide the correct values of $s_1$ and $s_2$ as inputs. The true rank of the underlying model, denoted by $r$, is used for data generation in the simulation. In contrast, our algorithm operates with an estimated rank $\widehat{r}$, which may differ from $r$. For estimation assessments ($\text{D}_{\mathbf{U}}$ and $\text{D}_{\mathbf{V}}$) and variable selection results (TPR and FPR), we use the true scenario rank $r$. To examine the empirical differences of signals at each rank, we compute covariance and correlation differences when $\widehat{r} = 2$, as summarized in Table 1. We do not attempt to estimate the true rank $r$, as its identification remains an open problem in dimension reduction. Table 1 summarizes the simulation results in 100 replications for the two rank scenarios. We change the sample sizes in each group $n_1 = n_2 \in \{20, 30, 50, 100, 200\}$. The TPR and FPR values support that the CCR model accurately selects the variables that produce the most contrastive linear combination by the binary variable. The smaller $\text{D}_{\mathbf{U}}$ and $\text{D}_{\mathbf{V}}$ indicate that the CCR model accurately estimates the subspaces $\mathbf{U}$ and $\mathbf{V}$. Under the rank-1 scenario ($r = 1$), the estimated $\widehat{\delta}_2$ has smaller values (than $\widehat{\delta}_2$ under the rank-2 scenario) and converges to zero with increasing sample size, since the true signals are only in the first canonical directions. Thus, the values of $\widehat{\delta}_2$ under a rank-1 scenario indicate that the second canonical directions are not needed to capture the contrastive difference by $Z$. However, under the rank-2 scenario, $\widehat{\delta}_2$ increases with larger sample sizes and indicates that true signals exist in the second canonical directions under

the rank-2 scenario. In both scenarios, $\widehat{\eta}_1$ is greater than $\widehat{\eta}_2$. The result indicates that the first linear combination captures the most contrastive pattern in $\mathbf{X}$ and $\mathbf{Y}$.

[Table 1 about here.]

We also investigate the empirical values of the maximal covariance difference ($\delta_1$) under the rank-1 scenario. To study $\widehat{\delta}_1$ under the true sparsity levels, we use $(s_1^*, s_2^*) \in \{(3,3),(10,10)\}$ and match the small sample sizes of the real data application by setting $n_1 = 11$ and $n_2 = 10$; we use $p_1 = 18$ and $p_2 = 15$ as illustrative dimensions in the same high-dimensional regime as the application ($p_1 = 16$, $p_2 = 18$ in the real data; see Section 4).

One would expect the maximal covariance differences to increase monotonically as the sparsity levels specified in Algorithm 1 increase because the information available for maximizing $\delta_1$ increases. In Figure 1, this is indeed the case when the true sparsity levels are 10 ($s_1^* = s_2^* = 10$) and the sparsity levels vary from 1 to 10 ($s_1 = s_2 = 1, 2, \ldots, 10$), represented by the solid line and circle dots. In this case, the estimated $\widehat{\delta}_1$ gradually increases. In contrast, when the true sparsity levels are 3 (represented by the dashed and triangular dots), $\widehat{\delta}_1$ rapidly increases as the sparsity levels ($s_1 = s_2$) rise to 3, beyond which no substantial increase is observed. This rapid stabilization suggests that accurately estimating the true sparsity levels is crucial, as it substantially impacts both the sensitivity and robustness of the maximal covariance difference. These patterns demonstrate CCR's ability to detect meaningful associations while avoiding overfitting via SPSS-based sparsity selection.

[Figure 1 about here.]

Overall, the simulation studies suggest that the CCR model performs particularly well when the signal-to-noise ratio is sufficiently large ($c_1/c_2 \geqslant 1$). When the signal is weak ($c_1/c_2 < 1$) and the sample size is limited, estimation variability increases, as expected in low-information settings. In particular, CCR consistently achieves higher true positive rates and lower false positive rates than penalized CCA methods, highlighting its advantage in

variable selection. Detailed numerical comparisons and additional results illustrating the effect of the signal-to-noise ratio $c_1/c_2$—along with further simulation scenarios, including an example with a multi-categorical conditioning variable $Z$—are provided in the Supplementary Materials.

## 4. REAL DATA ANALYSIS

Our goal is to identify the variables associated with sexual dimorphism in the association between features of the TMJ temporalis origin muscle attachment and features of the skull. We set the skull measurement $\mathbf{X} \in \mathbb{R}^{16}$, the temporalis origin (TO) measurements in $\mathbf{Y} \in \mathbb{R}^{18}$, and sex is the binary variable $Z \in \{1, 2\} = \{\text{male}, \text{female}\}$. We center each variable within sex group as discussed in Section 2.2.

First, we investigate the sparsity levels for the CCR using the SPSS method described in Section 2.5. Figure 2 shows boxplots of the p-values $\{p^{(i,k)}\}$ for $i = 1, \ldots, 16$ and $k = 1, \ldots, 18$. The figure indicates that, for both $s_1$ and $s_2$, there is no significant difference between the sparsity levels of 5 and 6. Hence, we take $(s_1, s_2) = (5, 5)$ in our analysis.

[Figure 2 about here.]

Figure 3 displays the linear combinations of skull ($\mathbf{X}$) and TO attachment ($\mathbf{Y}$) measurements identified by the CCR model that exhibit the greatest sex-specific association difference. The maximal covariance difference is $\widehat{\delta}_1 = 119.65$, and the associated correlation difference $\widehat{\eta}_1 = 1.16$ reflects the contrast in correlations between sexes, indicating that the estimated subspaces distinguish sex-specific association patterns.

[Figure 3 about here.]

The selected variables are given in (4) and the corresponding skull images are shown in Figure 4. The selected variables in skull align with previous forensic anthropological results that state that the width of the bicondylar (PlToPr) and the width of the bigonial (GnToGn)

are significant in determining the difference between the sexes (de Oliveira Gamba et al., 2016). The bicondylar width (PlToPr) and the bigonial width (GnToGn) are involved in the dimensions of the medial lateral skull, and the length of the right side of the mandible is related to the size of the anterior and posterior skulls. The two remaining selected skull variables—the anterior facial height (AnH) and the left ramus width (RamusWidth(L))—capture vertical and posterior dimensions of the mandible–cranium complex, which the CCR model also identifies as discriminative between sexes. All variables selected in the TO are related to the muscle attachment size (orange-colored bolded variables in Figure 4).

All data are centered within sex, and the linear combinations $(\widehat{\mathbf{U}}^\top\mathbf{X},\widehat{\mathbf{V}}^\top\mathbf{Y})$ are positively correlated in males (0.5390) and negatively correlated in females (-0.6203). The interpretation of positive and negative signs requires caution, since the estimation problem (2) is defined in the relative contrast between the groups and is invariant to sign changes in $\mathbf{U}$, $\mathbf{V}$, or both. The choice of signs in $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ should be made consistently throughout and needs to be consistent with subsequent interpretation and biomechanical explanation. Within females, smaller mandibular length (MandibleLength(R)) is associated with larger values of $\widehat{\mathbf{U}}^\top\mathbf{X}$, which in turn correspond to smaller values of $\widehat{\mathbf{V}}^\top\mathbf{Y}$. This follows from the negative loading of MandibleLength(R) in $\widehat{\mathbf{U}}$, so that decreasing mandibular length increases $\widehat{\mathbf{U}}^\top\mathbf{X}$, holding other variables fixed. Therefore, a female with a smaller mandibular length (relative to other females), which increases $\widehat{\mathbf{U}}^\top\mathbf{X}$, corresponds to a smaller $\widehat{\mathbf{V}}^\top\mathbf{Y}$. A female with a smaller-than-average mandibular length (among females) may have a larger-than-average TO attachment size (among females). This association is not evident in one-to-one variable relationships, where smaller mandibular length corresponds to smaller attachment areas. The CCR model uncovers associations between selected variables that are masked in simple pairwise comparisons.

[Figure 4 about here.]

$$\begin{aligned}
\text{Skull: } \widehat{\mathbf{U}}^\top \mathbf{X} &= 0.739\ \text{GnToGn} + 0.308\ \text{AnH} + 0.368\ \text{PlToPr} \\
&\quad - 0.323\ \text{RamusWidth(L)} - 0.345\ \text{MandibleLength(R)} \\
\text{TO: } \widehat{\mathbf{V}}^\top \mathbf{Y} &= -0.462\ \text{BoxLength(L)} - 0.482\ \text{BoxLength(R)} \\
&\quad - 0.323\ \text{BoxWidth(L)} - 0.524\ \text{Area(L)} - 0.319\ \text{Area(R)}
\end{aligned} \tag{4}$$

Figure 5 shows the heat maps of the marginal covariances and $\widetilde{\mathbf{\Phi}}$. The selected variables in $\widetilde{\mathbf{\Phi}}$ are shown as gray boxes outlining the corresponding rows and columns. From the covariance heatmaps (left), variation in MandibleLength(R) (the 14th variable) produces very little change in TO for females but substantial change for males. Among males, an increase in MandibleLength(R) is associated with marked increases in TO areas (the 7th and 8th variables). This association is faint when only the most related variables between the skull and TO are selected. Similarly, the 5th and 9th skull variables (PlToPr and RamusWidth(L)) display sex-specific patterns and are also selected by our CCR model with $(s_1, s_2) = (5, 5)$.

[Figure 5 about here.]

For additional biomechanical interpretation, we use the joint reaction force (JRF), calculated as the residual force at the TMJ required to maintain static equilibrium under estimated muscle forces during mandibular motion (She et al., 2021); the JRF magnitude is defined as the length of this vector. A larger JRF indicates greater loading on the TMJ, and for the same level of bite force, JRF increases as 3D mandibular length decreases (Sun et al., 2024). In raw-scale one-to-one comparisons, females with shorter mandibular length and smaller attachment areas exhibit larger JRF magnitudes, resulting in overall higher JRF values than males. Consistently, our result in (4) shows that females tend to have shorter mandibular length (MandibleLength (R)) as the selected size-related variables in TO increase, when other skull variables are held fixed, highlighting a potential high-risk subgroup for TMD

that may not be evident in raw-scale pairwise analyses. These findings suggest that the CCR model offers new insights into the relationship between skull geometry and muscle attachment, pointing to future research directions on abnormal or high-risk subgroups and their biomechanical implications. Additional results are provided in Sections S7.3 and S7.4 of the Supplementary Materials, with the raw-scale plot shown in Figure S5.

## 5. DISCUSSION

This paper proposes the conditional cross-covariance reduction model, which is easy to interpret and is designed to glean new information on the dynamic relationship of two sets of variables conditioning on the third binary variable. Instead of penalizing nonzero components, we apply hard-thresholding to them and introduce a sequential permutation for selecting sparsity method under limited sample sizes, which is practical for the small sample size of the skull and muscle attachment measurements of the TMJ.

The conditional cross-covariance reduction model can be extended to a tensor, or multi-array, formulation to accommodate repeated measures and multiway structured data, thereby capturing both temporal and modality-specific dynamic associations. While the sequential permutation for selecting sparsity method provides a robust, data-driven strategy, it could be further enhanced to quantify uncertainty in the selected features—for instance, by incorporating permutation-based confidence intervals or resampling methods to assess variability. At present, the p-values from this procedure yield a single decision rather than reflecting the full uncertainty range of the estimated effects. Developing such extensions would further strengthen the robustness of the conditional cross-covariance reduction modeling framework for longitudinal and multi-modal dynamic association analysis.

## ACKNOWLEDGEMENTS

We thank the editor, associate editor, and two referees for their valuable comments and constructive suggestions.

## SUPPLEMENTARY MATERIALS

Supplementary materials are available at Biometrics online. Web Appendices, Tables, and Figures referenced in Sections 1.1, 2.5, 3.2, and 4, together with code implementing the CCR and SPSS procedures, are available with this paper at the Biometrics website on Oxford Academic. The implementation code is also available at https://github.com/sparkqkr/CCR.

## FUNDING

The data collection was supported by the National Institute of Dental and Craniofacial Research (NIDCR), National Institutes of Health (NIH), under grant R01DE021134 (Yao), and the analysis was supported under grant R03DE030509 (Zhang and Slate).

## CONFLICT OF INTEREST

None declared.

## DATA AVAILABILITY

The data that support the findings in this paper are available from the Tissue Biomechanics Lab upon request to Dr. Hai Yao (haiyao@clemson.edu).

## References

Chen, J., Xie, J., and Li, H. (2011). A penalized likelihood approach for bivariate conditional normal models for dynamic co-expression analysis. *Biometrics* **67,** 299–308.

Coombs, M. C., She, X., Brown, T. R., Slate, E. H., Lee, J. S., and Yao, H. (2019). Temporomandibular joint condyle–disc morphometric sexual dimorphisms independent of skull scaling. *Journal of Oral and Maxillofacial Surgery* **77,** 2245–2257.

de Oliveira Gamba, T., Alves, M. C., and Haiter-Neto, F. (2016). Mandibular sexual dimorphism analysis in CBCT scans. *Journal of forensic and legal medicine* **38,** 106–110.

Ho, Y.-Y., Parmigiani, G., Louis, T. A., and Cope, L. M. (2011). Modeling liquid association. *Biometrics* **67,** 133–141.

Hylander, W. L. and Johnson, K. R. (1985). Temporalis and masseter muscle function during incision in macaques and humans. *International Journal of Primatology* **6,** 289–322.

Li, G. and Gaynanova, I. (2018). A general framework for association analysis of heterogeneous data. *The Annals of Applied Statistics* **12,** 1700–1726.

Li, K.-C. (2002). Genome-wide coexpression dynamics: theory and application. *Proceedings of the National Academy of Sciences* **99,** 16875–16880.

Li, K.-C., Liu, C.-T., Sun, W., Yuan, S., and Yu, T. (2004). A system for enhancing genome-wide coexpression dynamics study. *Proceedings of the National Academy of Sciences* **101,** 15561–15566.

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.

Mai, Q. and Zhang, X. (2019). An iterative penalized least squares approach to sparse canonical correlation analysis. *Biometrics* **75,** 734–744.

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.

She, X., Wei, F., Damon, B. J., Coombs, M. C., Lee, D. G., Lecholop, M. K., et al. (2018). Three-dimensional temporomandibular joint muscle attachment morphometry and its

impacts on musculoskeletal modeling. *Journal of biomechanics* **79,** 119–128.

Shu, H., Wang, X., and Zhu, H. (2020). D-cca: A decomposition-based canonical correlation analysis for high-dimensional datasets. *Journal of the American Statistical Association* **115,** 292–306.

Slade, G. D., Bair, E., By, K., Mulkey, F., Baraian, C., Rothwell, R., et al. (2011). Study methods, recruitment, sociodemographic findings, and demographic representativeness in the oppera study. *The Journal of Pain* **12,** T12–T26.

Stowell, A. W., Gatchel, R. J., and Wildenstein, L. (2007). Cost-effectiveness of treatments for temporomandibular disorders: biopsychosocial intervention versus treatment as usual. *The Journal of the American Dental Association* **138,** 202–208.

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,** e178578.

Wang, Y. R., Jiang, K., Feldman, L. J., Bickel, P. J., and Huang, H. (2015). Inferring gene–gene interactions and functional modules using sparse canonical correlation analysis. *The Annals of Applied Statistics* **9,** 300–323.

Witten, D. M., Tibshirani, R., and Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10,** 515–534.

Yang, D., Ma, Z., and Buja, A. (2016). Rate-optimal denoising of simultaneously sparse and low rank matrices. *The Journal of Machine Learning Research* **17,** 3163–3189.

Yu, T. (2018). A new dynamic correlation algorithm reveals novel functional aspects in single cell and bulk RNA-seq data. *PLoS computational biology* **14,** e1006391.

*Received September* 2026.

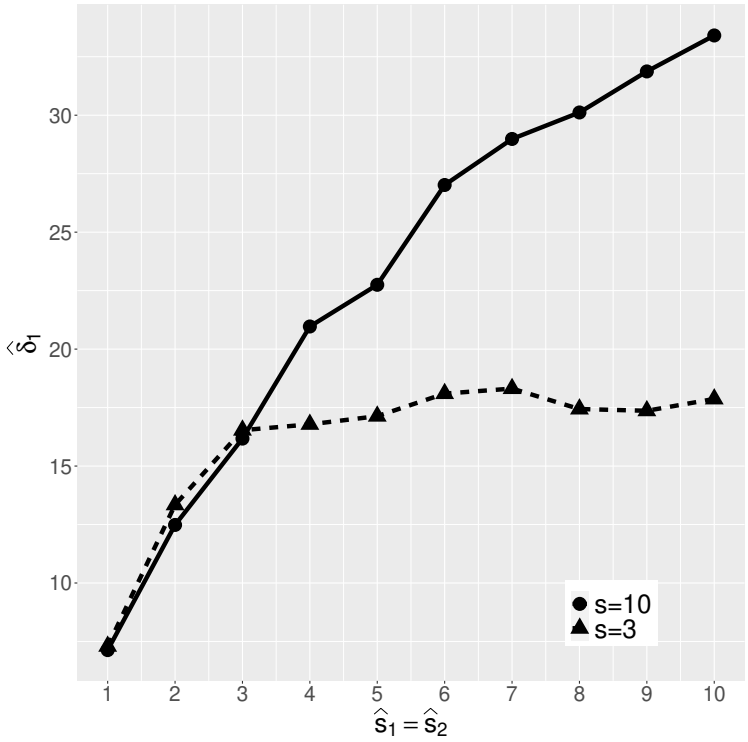
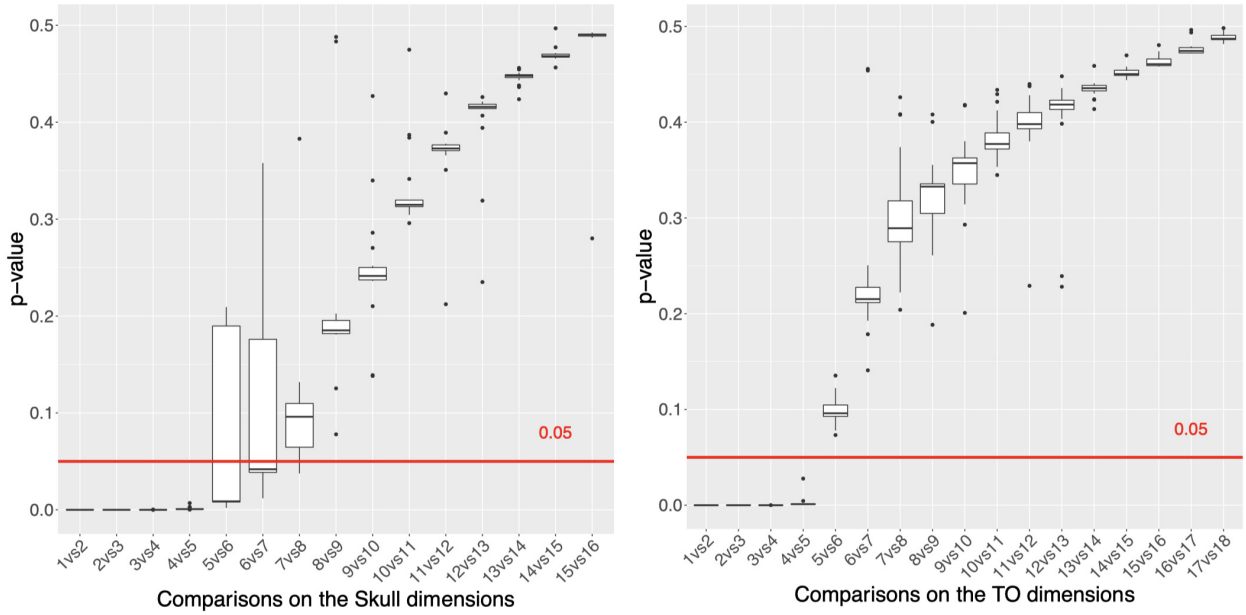



**Figure 1.** Estimated maximal covariance difference $(\widehat{\delta_1})$ where $n_1 = 11$, $n_2 = 10$, $p_1 = 18$, $p_2 = 15$. In solid line with circle dots, the true signals are in the first ten rows and columns $(s_1^* = s_2^* = s = 10)$ and increase $s_1 = s_2 \in \{1, 2, \ldots, 10\}$. In dashed line with triangular dots, the true signals are in the first three rows and columns $(s_1^* = s_2^* = s = 3)$ and increase $s_1 = s_2 \in \{1, 2, \ldots, 10\}$.




(a) Boxplot of p-values on $\mathbf{X}$

(b) Boxplot of p-values on $\mathbf{Y}$

**Figure 2.** Box plots of p-values on the result of the sequential permutation for selecting sparsity (SPSS) method for the maximal covariance difference $\widehat{\delta}_1$ on skull and temporalis origin (TO). From "1vs2", we can select the sparsity level $s_1$ and $s_2$ on panels (a) and (b), respectively. In panel (a), for example, all small p-values on "1vs2" ($< 0.05$) represent that there are significant differences between $s_1 = 1$ and $s_1 = 2$ with fixed $s_2 = k \in \{1, \dots 18\}$. Then, we consecutively move to the next boxplots until there exists any p-value greater than 0.05 (no significant difference). Here, we select the sparsity level $(s_1, s_2) = (5, 5)$.

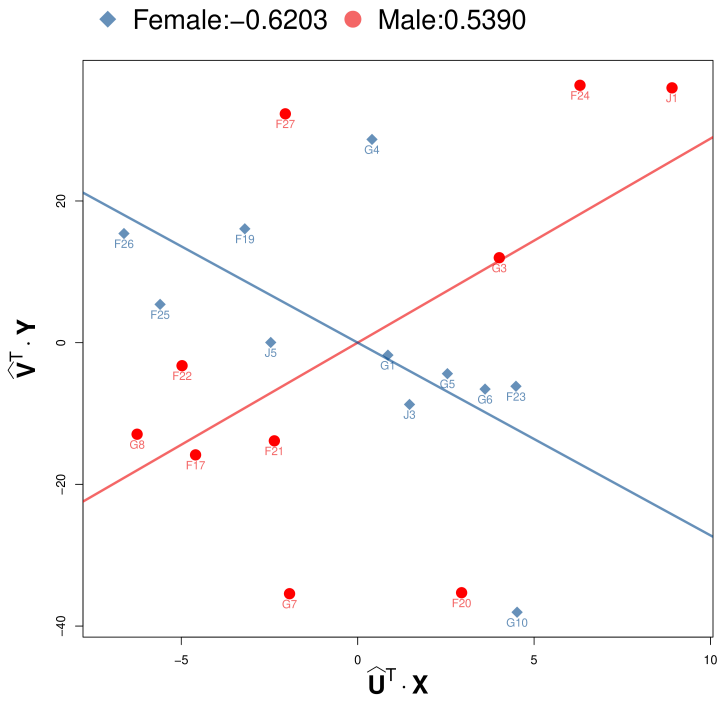



**Figure 3.** Linear combinations of the CCR model on the skull and temporalis origin (TO) measurements. The numbers in the legend represent the correlations between linear combinations by sex. Sparsity levels are set as $s_1 = s_2 = 5$. The $x$-axis ($\widehat{\mathbf{U}}^\top \mathbf{X}$) indicates the linear combination on the skull, and the $y$-axis ($\widehat{\mathbf{V}}^\top \mathbf{Y}$) represents the linear combination on temporalis origin (TO).

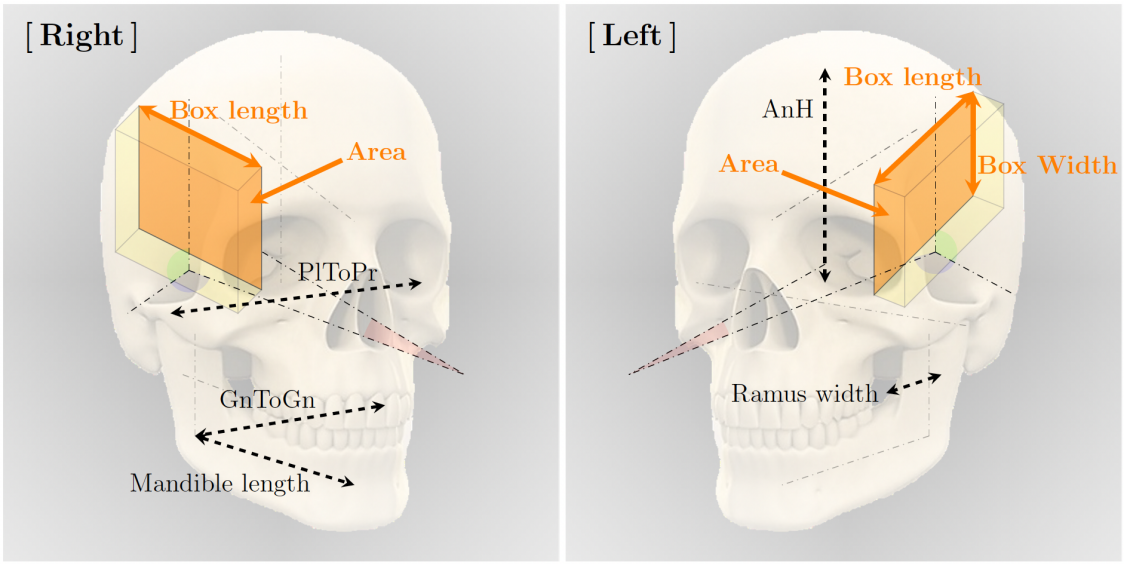



**Figure 4.** Result of the CCR model on the skull and temporalis origin (TO) under $(s_1, s_2) = (5, 5)$ where the 10 selected variables are denoted with variable names. Selected variables on TO are marked as orange-colored bolded letters with solid lines, and selected variables on the skull are distinct as non-bolded letters with dashed lines.

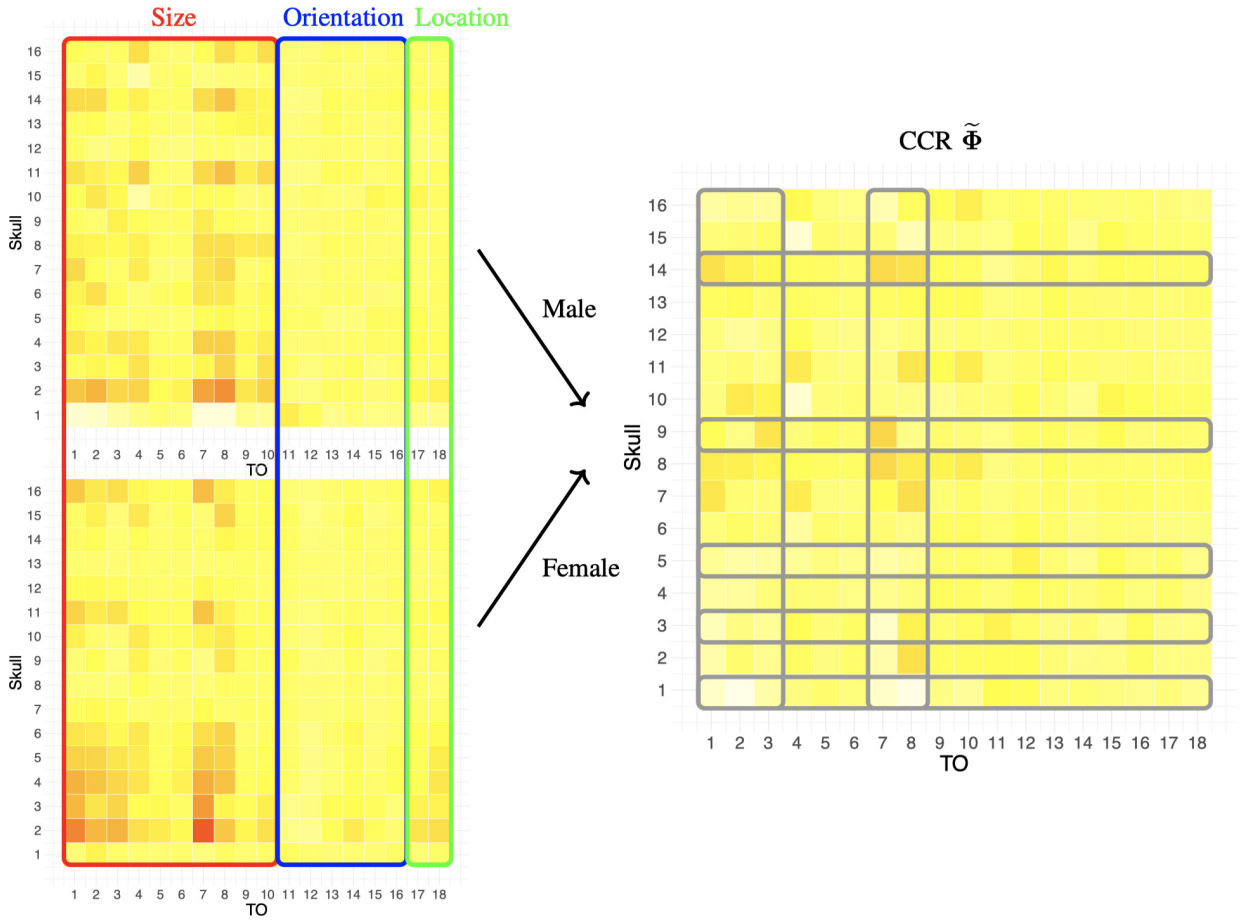



**Figure 5.** Marginal covariance of female and male (left), and sample covariance difference $\widetilde{\mathbf{\Phi}}$ (right). The variables subsetted in size, orientation, and location are displayed in Table S8 in Supplementary Materials. The gray boxes on the right side represent the variables selected under $(s_1, s_2) = (5, 5)$.

**Table 1**

*Numerical evaluations under rank-1 and rank-2 scenarios for the cross-covariance structures based on 100 data replicates. The numbers in parentheses report the standard errors of the subspace distances $D_{\mathbf{U}}$ and $D_{\mathbf{V}}$, the covariance differences $\widehat{\delta}_1$ and $\widehat{\delta}_2$, and the associated correlation differences $\widehat{\eta}_1$ and $\widehat{\eta}_2$. Here, the true sparsity levels ($s_1^* = s_1$, $s_2^* = s_2$) are used for estimation.*

| Rank ($r=1$) | $n_1 = n_2$ 20 | 30 | 50 | 100 | 200 |
|---|---|---|---|---|---|
| $\mathrm{TPR}_{\mathbf{X}}$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\mathrm{TPR}_{\mathbf{Y}}$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\mathrm{FPR}_{\mathbf{X}}$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $\mathrm{FPR}_{\mathbf{Y}}$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $\mathrm{D}_{\mathbf{U}}$ | 0.085 (0.000) | 0.069 (0.000) | 0.060 (0.000) | 0.038 (0.000) | 0.027 (0.000) |
| $\mathrm{D}_{\mathbf{V}}$ | 0.113 (0.001) | 0.084 (0.000) | 0.075 (0.000) | 0.049 (0.000) | 0.037 (0.000) |
| $\widehat{\delta}_1$ | 11.765 (0.254) | 11.825 (0.239) | 12.131 (0.186) | 12.068 (0.141) | 12.035 (0.082) |
| $\widehat{\delta}_2$ | 0.635 (0.028) | 0.514 (0.019) | 0.388 (0.015) | 0.285 (0.012) | 0.184 (0.008) |
| $\widehat{\eta}_1$ | 1.785 (0.006) | 1.778 (0.006) | 1.792 (0.004) | 1.794 (0.003) | 1.793 (0.002) |
| $\widehat{\eta}_2$ | 0.533 (0.019) | 0.424 (0.016) | 0.323 (0.011) | 0.239 (0.009) | 0.150 (0.006) |
| Rank ($r=2$) | $n_1 = n_2$ 20 | 30 | 50 | 100 | 200 |
| $\mathrm{TPR}_{\mathbf{X}}$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\mathrm{TPR}_{\mathbf{Y}}$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\mathrm{FPR}_{\mathbf{X}}$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $\mathrm{FPR}_{\mathbf{Y}}$ | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| $\mathrm{D}_{\mathbf{U}}$ | 0.197 (0.013) | 0.177 (0.011) | 0.112 (0.007) | 0.088 (0.006) | 0.057 (0.004) |
| $\mathrm{D}_{\mathbf{V}}$ | 0.190 (0.014) | 0.141 (0.012) | 0.117 (0.008) | 0.082 (0.005) | 0.058 (0.003) |
| $\widehat{\delta}_1$ | 11.772 (0.248) | 11.861 (0.239) | 12.151 (0.187) | 12.079 (0.142) | 12.055 (0.082) |
| $\widehat{\delta}_2$ | 1.127 (0.078) | 1.224 (0.050) | 1.269 (0.025) | 1.238 (0.017) | 1.253 (0.014) |
| $\widehat{\eta}_1$ | 1.791 (0.006) | 1.787 (0.006) | 1.796 (0.004) | 1.794 (0.003) | 1.796 (0.002) |
| $\widehat{\eta}_2$ | 1.057 (0.060) | 1.174 (0.036) | 1.200 (0.014) | 1.179 (0.009) | 1.189 (0.006) |

*Note:* In the rank-1 scenario, the true value of $\delta_2$ is zero; therefore, we report $\widehat{\delta}_2$ for diagnostic purposes.