Article:
Mardia, K.V., Wu, X., Kent, J.T. et al. (Accepted: 2026) Asymmetry Analysis of Bilateral Shapes. Journal of the Royal Statistical Society, Series C. ISSN: 0035-9254 (In Press)

Asymmetry Analysis of Bilateral Shapes

Kanti V. Mardia[1], Xiangyu Wu[1], John T. Kent[1], Colin R. Goodall[1],

Balvinder S. Khambay[2]

[1]Department of Statistics, School of Mathematics, University of Leeds, LS2 9JT, England.

[2]Third Orthodontics Division, Dentistry, School of Health Sciences, College of Medicine and Health, University of Birmingham, England.

Address for correspondence: Kanti V. Mardia, School of Mathematics, University of Leeds, Leeds, LS2 9JT, UK.

Email: k.v.mardia@leeds.ac.uk.

Abstract

Many biological objects possess approximate bilateral symmetry about a midplane. This paper uses landmark-based methods to measure departures from bilateral symmetry. The landmarks for each object come in pairs either side of a midplane or as solos on or near the midplane. After registration, a vector of unsigned elementary features is constructed to describe the asymmetry of each object. Two approaches are proposed to compare the level of asymmetry between two groups of objects. In the first approach the elementary features are combined into a scalar composite asymmetry score for each object; then standard univariate tests are used on the composite asymmetry score to assess asymmetry. In the second approach a univariate test statistic is constructed for each elementary feature; then the maximum of these statistics is used for a union-intersection test. The methodology is illustrated on two data sets collected to assess the success of two types of surgery: (1) cleft lip surgery, where the level of asymmetry is compared between a group of cleft lip subjects and a group of normal

subjects; and (2) orthognathic surgery, where for each patient, the level of asymmetry after surgery is compared to that before surgery.

Keywords: asymmetry scores, cleft lip surgery, orthognathic surgery, registration, shape analysis, union-intersection test.

1 Introduction

Bilateral symmetry, also called left-right symmetry, is a key property in many biological settings. In practice, exact bilateral symmetry seldom holds and it is of interest to study the extent of any asymmetry. Research on the symmetry of objects in the real world has a long history (see, for example, Weyl, 1952). Statistical methods to test for asymmetry were first formalized in Mardia et al. (2000). Subsequently, a wide variety of statistical methods have been developed; see, for example, Kent and Mardia (2001), Bock and Bowman (2006), Ajmera et al. (2022), Ajmera et al. (2023), Patel et al. (2023) and Bains et al. (2025).

The focus of this paper is on landmark-based subjects or objects as in Mardia et al. (2000). A set of K landmarks is identified a priori on each object in a data set so that each object is represented by a $K \times M$ configuration matrix giving the positions of K landmarks in M dimensions. The most important dimensions in practice are $M = 2$ and $M = 3$. Say an object is *bilateral* if its landmarks can be divided into two categories: *pairs* and *solos* according to an $M-1$-dimensional midplane $\mathcal{P}$ which passes through the center of this object. Let there be K_P pairs and K_S solos, so $K = 2K_P + K_S$. The paired landmarks lie on either side of $\mathcal{P}$, while the solos lie on or near $\mathcal{P}$. Say an object is *bilaterally symmetric* about $\mathcal{P}$ if its reflection through $\mathcal{P}$ is exactly the same as the original object up to a relabeling of the landmarks.

For a given set of objects, it is necessary to register them to have the same estimated orientation and location in order that objects can be compared. In particular, for bilateral objects the registered objects should have the same estimated midplane. Further details about registration are given in Section 2.1.

The purpose of the paper is to develop and analyze one-sided tests for vector measures of asymmetry for registered landmark configurations. Two approaches are developed to compare the level of asymmetry between different groups of objects. The two approaches can be described as "combine-then-compare" and "compare-then-combine", leading to a composite test statistic and a union-intersection test statistic, respectively. The two testing approaches and their merits are described in Section 3. Both approaches can be used in a two-group setting or in a paired setting. Earlier work has usually emphasized parametric t-tests; this paper uses nonparametric tests to avoid the assumption of normality.

Figure 1: Four photographs of a cleft patient, taken pre-surgery and at three stages after surgery. From `https://www.nhs.uk/conditions/cleft-lip-and-palate/`.

Two data sets, labeled Data Set A and Data Set B, are used to illustrate the methods of the paper. These data sets are related to two medical questions corresponding to the level of asymmetry in a person's smile. Both sets were collected and pre-processed by one of the authors, B. S. Khambay. These data sets are available at `https://github.com/XW-2025-hub/JRSS-C-Supplementary-Materials` and were earlier analyzed in Patel et al. (2023) and Bains et al. (2025), respectively.

Data Set A is related to cleft lip surgery, which is performed so that form and function can be restored to a normal lip at rest and during movement. Figure 1 illustrates the purpose of cleft lip surgery. The first panel shows a patient before surgery, while the second, third and fourth panels show the patient at various stages after surgery. Data Set A contains two groups of subjects: 13 cleft subjects and 12 control subjects. The purpose of the analysis is to see if the cleft subjects

(after surgery) demonstrate greater asymmetry than the control subjects. The surgery generally takes place in childhood and the comparisons here are between adults.

Data Set B is related to orthognathic surgery, which is performed to correct jaw discrepancies in order to restore the form and function. Although the subjects do not display appreciable asymmetry before surgery, there is concern that the surgery might increase the lip and nose asymmetry as an unintended consequence. Figure 2 shows a face before (left panel) and after (right panel) orthognathic surgery, where the view is from the right side of the face. In the left panel, the jaw protrudes somewhat in comparison to the right. This data set contains paired data on 22 subjects, where each patient is measured before and after surgery. The objective is to see if the post-surgery measurements demonstrate more asymmetry than the pre-surgery measurements.

The collection of information about the smile was the same for both data sets. Each subject was asked to smile from rest (closed lips) to maximum open lip smile. A four-dimensional (three spatial dimensions plus time) movie taken at one-second intervals was made of the face using Di4D motion capture system (`https://di4d.com`). Both data sets were registered in the same way using the ideas of Section 2.1. For the purpose of this paper, three fixed time frames (the beginning, middle and end of the smile) have been used from each movie. Table 1 displays the total number of landmarks (K), landmark pairs (K_P) and solos (K_S) for these two data sets. The two testing methodologies are applied to the two data sets in Sections 4 and 5, respectively. Finally, Section 6 highlights the key contributions of the paper and gives directions for further work.

Table 1: Total number of landmarks, landmark pairs and solos used in Data Sets A and B.

	K	K_P	K_S
Data Set A	24	11	2
Data Set B	16	6	4

Figure 2: View of a face before (left) and after (right) orthognathic surgery.

2 Describing asymmetry

Let $X^*(K \times M)$ denote the landmark configuration for a typical subject or object with K landmarks in M-dimensional space. Denote the elements of X^* by $X^*[k, m]$, $k = 1, \ldots, K,\ m = 1, \ldots, M$. Similarly, denote the rows and columns of X^* by $X^*[k,\ ]$ and $X^*[\ , m]$, respectively. Given a collection of N configurations denote the nth configuration by $X^{(n)*}$. The superscript * on X indicates that it has not yet been registered.

2.1 Registration

The applications in this paper deal with human heads in $M = 3$ dimensions, each represented as a $K \times 3$ configuration matrix of landmark coordinates. Each configuration matrix is only defined up to a rigid body motion and in order to compare the objects it is necessary to register them to a common frame of reference. For this purpose it is helpful to introduce the concept of pose. The *pose* of the head is defined as the location and orientation of the head relative to a chosen coordinate system (Hartley and Zisserman, 2004). Assume that associated with each head there is a set of pose parameters, that is, a location vector $\boldsymbol{\delta}$ and an orthogonal matrix

$$\Gamma = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3) \in \mathbb{R}^{3\times 3}, \tag{2.1}$$

where $\boldsymbol{\delta}$ and Γ specify an element of SE(3), the group of Euclidean transformations in $\mathbb{R}^3$.

The location vector defines the "center" of the head and the three columns of the orthogonal matrix represent the three "biologically meaningful" directions left-to-right, down-to-up and in-to-out, respectively. Recall that a plane P in $\mathbb{R}^3$ can be specified by a normal direction (a unit vector $\boldsymbol{\gamma}$) and a constant c representing the signed distance from the plane to the origin, so that

$$P = \{\boldsymbol{x} \in \mathbb{R}^3 : \boldsymbol{x}^T\boldsymbol{\gamma} = c\}. \tag{2.2}$$

In our application the landmarks come as pairs that are bilaterally positioned about the mid-sagittal plane (the midplane $\mathcal{P}$), and as solos that lie on or near the plane. The position and orientation of the mid-sagittal plane (i.e. the normal vector and signed distance) specify "partial" pose information: the midplane is normal to the left-to-right direction and passes through the center of the head. This partial information leaves unspecified the down-to-up and in-to-out directions, which form a basis for the midplane. Specifying the remaining pose orientation parameters is equivalent to choosing the two orthogonal basis vectors in the midplane. In passing, note that only the left-to-right location parameter is of interest in our analysis later; the location parameters in the other two directions are not needed.

In order to compare configurations, it is necessary to register them, i.e., to translate and rotate them so that they all have the same *canonical pose* for which the center lies at the origin, $\boldsymbol{\delta} = \mathbf{0}$, and the pose directions are given by the three coordinate directions, so that $\Gamma = I_3$. Thus, after registration each head is upright and facing forward.

However, the pose parameters are generally unobserved, and it is necessary to estimate them. In this paper an "external" set of landmarks is used to estimate the pose parameters for the following reason. For all the subjects in both data sets, the top half of the head is nearly bilaterally symmetric; most of the asymmetries appear in the bottom half of the head. Hence, a selection of landmarks in the top half of the head is used to estimate the pose.

More specifically, the pose is estimated here using a clinically derived midplane, as follows. The

two inter-pupillary landmarks are used to determine the left-to-right direction with the center in this direction given by nasion (the anatomical landmark at the intersection of the forehead and nasal bones). The left and right tragus (the anatomical landmark at the front of ear canal) and orbitales (the anatomical landmark at the lowest point of the eye socket) are used to estimate Frankfort plane, which is the horizontal plane. In practice, these features are identified by eye from a three-dimensional image at the start of the smile. Then four landmarks in the forehead are followed through time to preserve the registration. Figure 3 shows the axes after registration; Patel et al. (2023) give more details.

Another way to estimate pose is through Procrustes methods (see, for example, Dryden and Mardia (2016) and Bock and Bowman (2006)) using just the lip landmarks. However, this approach is not suitable for the data sets of this paper because although the lips may be symmetric within themselves, the lips as a whole may be asymmetrically positioned with respect to the upper half of the face. As a result, Procrustes methods can underestimate the amount of asymmetry. This effect has been noted by Patel et al. (2023), Figure 4, and in some experiments performed by us on Data Set A reported at `https://github.com/XW-2025-hub/JRSS-C-Supplementary-Materials`.

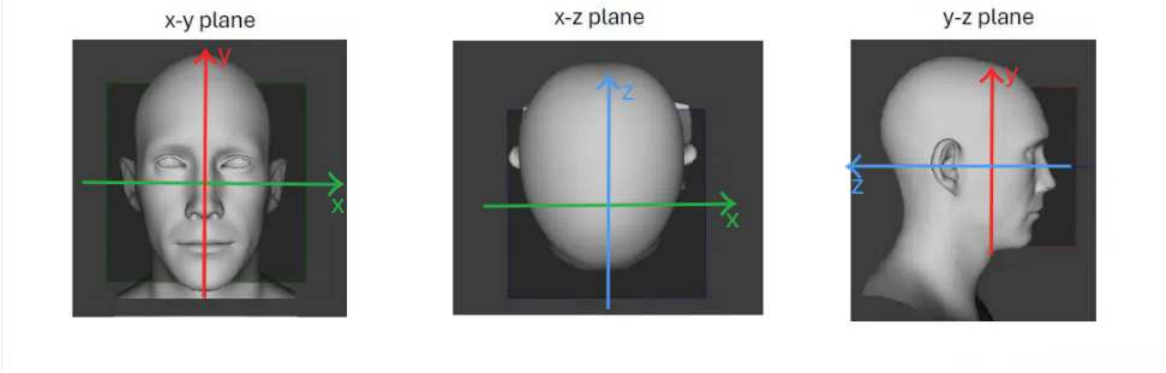

Figure 3: Coordinate system used for registration of human face viewed in the three principal planes. The x-axis is in green, y-axis is in red and z-axis is in blue.

2.2 Elementary features

For the rest of the paper, it is assumed that the bilateral data configurations have been registered as in Section 2.1 so that the estimated pose is canonical. In particular, the estimated midplane $\mathcal{P}$

is orthogonal to the first coordinate axis and passes through the origin. Denote a typical registered configuration by X (without the *).

Let k_L and k_R denote the left and right indices for a typical landmark pair from X and define M *coordinatewise signed elementary features* for this landmark pair,

$$d[(k_L, k_R), 1] = X[k_L, 1] + X[k_R, 1] \tag{2.3}$$

$$d[(k_L, k_R), m] = X[k_L, m] - X[k_R, m], \quad m = 2, \ldots, M. \tag{2.4}$$

There is one feature for each coordinate of each landmark pair, giving a total of MK_P coordinatewise features. At the first coordinate the two landmarks in the pair generally have opposite signs, while at the other coordinates they generally have the same sign. Thus, the quantities given in (2.3) and (2.4) quantify the departure from symmetry. For a bilaterally symmetric subject, the quantities in equations (2.3) and (2.4) are exactly 0.

Similarly, if k_S indexes a typical solo landmark, define a single elementary feature

$$d[(k_S)] = X[k_S, 1], \tag{2.5}$$

so there are in total K_S elementary features for solos. For a bilaterally symmetric subject, the above quantity is 0; no constraints are placed on the other coordinates.

The coordinatewise elementary features can be collected into a *signed elementary feature vector* $\boldsymbol{d} = (d_j)$ of length $J = MK_P + K_S$, as a function of the matrix X. The elements of $\boldsymbol{d}$ can be listed either sequentially with a subscript index, i.e., $d_j,\ j = 1, \ldots, J$, or using parentheses within square brackets, as in (2.3)-(2.5) when it is of interest to emphasize the corresponding landmarks and coordinates.

In many examples, such as the smile data sets analyzed in this paper, the direction of asymmetry is not thought to be scientifically interesting; e.g., whether a cleft lip points to the left or to the right. Hence, the focus in this paper is on the absolute values of the elementary features.

Define the *unsigned elementary feature vector* $\boldsymbol{a}$ with elements

$$a_j = |d_j|, \quad j = 1, \ldots, J. \tag{2.6}$$

The elements of $\boldsymbol{a}$ are indexed in the same way as elements of $\boldsymbol{d}$. In the rest of the paper, each configuration X is reduced to an unsigned elementary feature vector $\boldsymbol{a}$ for further numerical and statistical analysis. For a sample of N configurations, denote the nth feature vectors by $\boldsymbol{d}^{(n)}$ and $\boldsymbol{a}^{(n)}$, $n = 1, \ldots, N$.

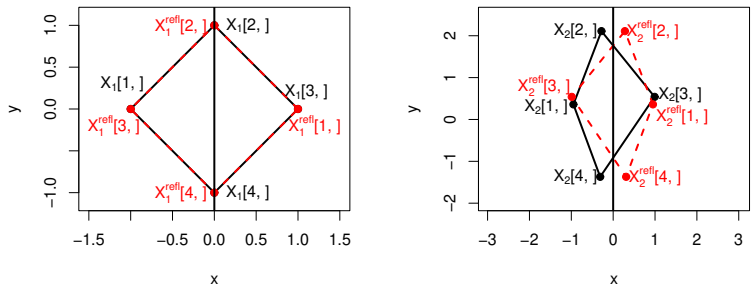

Figure 4: The left and right panels show a square $X^{(1)}$ centered at the origin and an asymmetric quadrilateral $X^{(2)}$, respectively, (in solid black lines) together with their reflections $X^{(1),\text{refl}}$ and $X^{(2),\text{refl}}$ (in red dotted lines) about the vertical midline through the origin.

Illustrative example:

To illustrate the elementary features, consider two registered configurations in $M = 2$ dimensions with $K = 4$ landmarks given by

$$X^{(1)} = \begin{pmatrix} X^{(1)}[1,\ ] \\ X^{(1)}[2,\ ] \\ X^{(1)}[3,\ ] \\ X^{(1)}[4,\ ] \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad X^{(2)} = \begin{pmatrix} X^{(2)}[1,\ ] \\ X^{(2)}[2,\ ] \\ X^{(2)}[3,\ ] \\ X^{(2)}[4,\ ] \end{pmatrix} = \begin{pmatrix} -0.95 & 0.36 \\ -0.28 & 2.11 \\ 0.99 & 0.54 \\ -0.31 & -1.37 \end{pmatrix}. \tag{2.7}$$

There is one landmark pair $(k_L, k_R) = (1, 3)$ and two solos, $k_S = 2, 4$, so $K_P = 1$ and $K_S = 2$. The configurations $X^{(1)}$ and $X^{(2)}$ are shown in the left and right panels of Figure 4, respectively, (solid black lines) together with their corresponding reflections $X^{(1),\text{refl}}$ and $X^{(2),\text{refl}}$ (dotted red

Table 2: Elements of the unsigned and signed elementary feature vectors $\boldsymbol{a} = \boldsymbol{a}^{(2)} \in \mathbb{R}^4$ and $\boldsymbol{d} = \boldsymbol{d}^{(2)}$ for $X^{(2)}$ with $K = 4$ landmarks in $M = 2$ dimensions, and $J = 4$.

Vector index	Landmark indices	Coordinate axis	Features of $\boldsymbol{a}$	Values of $\boldsymbol{a}$	Values of $\boldsymbol{d}$
1	Pair (1,3)	1	$a[(1,3),1] = \lvert X^{(2)}[1,1] + X^{(2)}[3,1]\rvert$	0.04	0.04
2	Pair (1,3)	2	$a[(1,3),2] = \lvert X^{(2)}[1,2] - X^{(2)}[3,2]\rvert$	0.18	−0.18
3	Solo 2	1	$a[(2)] = \lvert X^{(2)}[2,1]\rvert$	0.28	−0.28
4	Solo 4	1	$a[(4)] = \lvert X^{(2)}[4,1]\rvert$	0.31	−0.31

lines) about the midplane $\mathcal{P}$. In $M = 2$ dimensions, the midplane is actually a midline and is given by the vertical line through the origin.

Since $X^{(1)}$ is bilaterally symmetric, its unsigned elementary feature vector vanishes, $\boldsymbol{a}^{(1)} = \boldsymbol{0}$. Table 2 gives the $J = 4$ elements of the unsigned elementary feature vector $\boldsymbol{a}^{(2)}$ and $\boldsymbol{d}^{(2)}$ for $X^{(2)}$. The two elements for the landmark pair (1,3) are listed first, and the two elements for the solos, landmarks 2 and 4, are listed last in the table. It can be seen visually in Figure 4 that for $X^{(2)}$, the solo landmarks 2 and 4 deviate more from the y-axis than landmark 1 differs from the reflected version of landmark 3. Hence the last two elements of $\boldsymbol{a}^{(2)}$ are larger than the first two elements.

2.3 Composite Asymmetry Score

The unsigned elementary feature vector $\boldsymbol{a}$ is a multivariate measure of asymmetry for an object. Each element a_j of $\boldsymbol{a}$ (equation (2.6)) is nonnegative and a larger value of a_j indicates greater asymmetry. Hence it is natural to reduce the J-dimensional vector $\boldsymbol{a}$ to a single number by combining the elements of $\boldsymbol{a}$ into a scalar *composite asymmetry score* $u = \psi(\boldsymbol{a})$, say, where the function $\psi(\boldsymbol{a})$ is monotone increasing in each element a_j of $\boldsymbol{a}$.

One natural choice for the composite asymmetry score is

$$u = u(\boldsymbol{a}) = \sum_{j=1}^{J} a_j^2, \tag{2.8}$$

which was also used by Bock and Bowman (2006); their statistic A is the same as $\frac{2}{K}u$ here. A

more general choice for the composite asymmetry score is

$$u_\rho(\boldsymbol{a}) = \sum_{j=1}^{J} a_j^\rho \tag{2.9}$$

where $\rho > 0$ is a power, e.g. $\rho = 1$ or $\rho = 2$. Thus (2.8) corresponds to the power $\rho = 2$. Patel et al. (2023) and Bains et al. (2025) used a statistic similar to (2.9) with $\rho = 1$,

$$u^* = \sum_{(k_L,k_R)} \left\{b_{(k_L,k_R)}\right\}^{\frac{1}{2}} + \sum_{k_S} a[(k_S)], \quad b_{(k_L,k_R)} = \sum_{m=1}^{3} a^2[(k_L,k_R),m] \tag{2.10}$$

where the first sum is over landmark pairs and the second sum is over solos. At each landmark pair, (2.10) uses $b_{(k_L,k_R)}$, the Euclidean or L_2 norm of a three-dimensional unsigned elementary feature vector because it has a direct physical interpretation. In contrast (2.9) with $\rho = 1$ emphasizes the individual elementary features $a[(k_L,k_R),m]$, so that differences in asymmetry can be explored separately in the three coordinate directions, as in Section 4.3. We note that numerical experiments reported in Mardia et al. (2026) suggest for composite tests there is little difference between the powers $\rho = 1$ and $\rho = 2$ and (2.10) for the examples in this paper.

Another possibility is to use a weighted version of (2.9), $\sum_{j=1}^{J} w_j a_j^\rho$, where the $w_j \geq 0$ are pre-assigned weights. For example, higher weights might be assigned to features for landmarks nearer to the midplane (Mardia et al., 2026). However, for simplicity only the unweighted composite summary statistic (2.8) is used in this paper.

It is possible to carry out a theoretical analysis of the distribution of (2.8) under an isotropic Gaussian assumption on the landmark coordinates (Mardia and Wu, 2025). However, more realistic assumptions require a simulation-based analysis, as given in Section 3.

Consider again the illustrative example in (2.7). Using (2.8) the composite asymmetry scores are $u^{(1)} = 0$ (by construction) and $u^{(2)} = 0.21$ using the $\boldsymbol{a}^{(2)}$ from Table 2.

3 Hypothesis testing

In this section hypothesis tests are developed to compare two groups of objects, where for each elementary feature, the average asymmetry of an individual is thought to be less in one group than the other. A variant of the two-group setting is the paired setting in which each individual object is measured at two different times. For simplicity this section focuses on the two-group setting. The necessary modifications for the paired setting are covered in Section 5.

Let $X^{(1)}, \ldots, X^{(N)}$ denote the registered $K \times M$ matrices of landmark coordinates for N bilateral objects, and let $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(N)}$ denote the corresponding J-dimensional unsigned elementary feature vectors. Suppose the first N_1 configurations come from Group 1 and the remaining N_2 configurations come from Group 2, $N = N_1 + N_2$. Assume the unsigned elementary feature vectors from Group 1 are identically independent distributed (i.i.d.) with a common cumulative distribution function (c.d.f.) F_1. Similarly, suppose the unsigned elementary feature vectors from Group 2 are i.i.d. with a common c.d.f. F_2.

Recall that F_1 is said to be *stochastically less than or equal to* F_2, written $F_1 \leq_{st} F_2$, if

$$F_2(\boldsymbol{a}) \leq F_1(\boldsymbol{a}) \text{ for all } \boldsymbol{a}. \tag{3.1}$$

Similarly, F_1 is said to be *stochastically less than* F_2, $F_1 <_{st} F_2$, if (3.1) holds with with strict inequality for at least one $\boldsymbol{a}$ (e.g., Stoyan, 1983). Let F_{1j} and F_{2j} denote the marginal c.d.f.s of feature a_j, $j = 1, ..., J$. Then $F_1 <_{st} F_2$ implies that $F_{1j} \leq_{st} F_{2j}$ for all $j = 1, \ldots, n$, with strict inequality for at least one j.

Stochastic order provides a natural framework in which to construct a multivariate version of a one-sided hypothesis test. Consider the following hypothesis test:

$$H_0 : F_1 = F_2 \text{ vs. } H_1 : F_1 <_{st} F_2. \tag{3.2}$$

Two general approaches can be used to construct a test statistic. These are: (a) combine-then-

compare, and (b) compare-then-combine. These two approaches are explored in the next subsections.

3.1 Composite hypothesis test: combine-then-compare

Recall from (2.8) that the elements of an unsigned elementary feature vector $\boldsymbol{a}$ can be combined into a composite asymmetry score u, say. Let $G_1(u)$ and $G_2(u)$ denote the c.d.f.s of the scalar u. Since u is a monotone increasing function of $\boldsymbol{a}$, $F_1 <_{st} F_2$ implies $G_1 <_{st} G_2$. Hence, under H_0, $G_1 = G_2$, and under H_1, $G_1 <_{st} G_2$.

Since the composite asymmetry scores are scalars, any standard one-sided univariate test statistic such as the two-sample t-test or the Wilcoxon rank sum test, also known as the Mann-Whitney U test, can be used to compare the two groups. Earlier work has often used the t-test (e.g., Bock and Bowman, 2006; Patel et al., 2023). To allow for long-tailed distributions and outliers, the Wilcoxon rank sum test is used in this paper, Numerical investigations in Mardia et al. (2026) indicate that there is little difference for the examples of this paper.

Let $u^{(1)}, \ldots, u^{(N)}$ denote the composite asymmetry scores for the unsigned elementary feature vectors $\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(N)}$, where the first N_1 observations come from Group 1 and the remaining N_2 observations come from Group 2. Let $R^{(n)},\ n = 1, \ldots, N$ be the ranks of the composite asymmetry scores $u^{(n)},\ n = 1, \ldots, N$, and let R_1 denote the sum of the ranks from Group 1. The Wilcoxon rank sum test statistic is given by

$$T_C = N_1 N_2 + \frac{N_1(N_1+1)}{2} - R_1. \tag{3.3}$$

The distribution of T_C under the null hypothesis is given by the permutation distribution in which the $N!$ possible orders of the $u^{(n)}$ values are randomly assigned. The null distribution is computed exactly or approximately in standard statistical packages. The null hypothesis is rejected for large values of T_C.

The composite approach enables two groups to be compared visually using dot plots, so that

any overlap between the two groups can be easily assessed. Examples are given in Sections 4.2 and 5.2.

3.2 Union-intersection hypothesis test: compare-then-combine

In this approach, a separate test statistic is constructed for each feature in the unsigned elementary feature vector. The most extreme of these separate test statistics is then used as an overall test statistic. For the jth feature, $1 \leq j \leq J$, consider the hypothesis test

$$H_{0j} : F_{1j} = F_{2j} \text{ vs. } H_{1j} : F_{1j} <_{st} F_{2j}. \tag{3.4}$$

where F_{1j} and F_{2j} are defined below (3.1).

Under the overarching assumption that $F_1 \leq_{st} F_2$, the multivariate hypothesis test in (3.2) can be written in a union-intersection formulation,

$$H_0 = \bigcap_{j=1}^{J} H_{0j} \text{ vs. } H_1 = \bigcup_{j=1}^{J} H_{1j}. \tag{3.5}$$

For the jth univariate hypothesis test, a Wilcoxon rank sum test statistic T_{Cj} can be constructed, based on ranks of $a_j^{(n)}$, $n = 1, \ldots, N$. As in Boyett and Shuster (1977), an overall statistic to test H_0 vs H_1 in equation (3.5) can be constructed by taking the maximum of these univariate statistics,

$$T_{UI} = \max_{j=1,\ldots,J} T_{Cj}. \tag{3.6}$$

Since the features T_{Cj} are typically correlated with each other for each subject, it is not possible to derive or approximate the null distribution of T_{UI} theoretically under the null permutation distribution. Instead, simulations must be used. For each of B random permutations of the $N = N_1 + N_2$ subjects, assign the first N_1 subjects to Group 1 and the remaining N_2 subjects to Group 2. The sampling distribution of T_{UI} gives the distribution of T_{UI} under the null hypothesis.

Further, if r is the rank of the observed value of T_{UI}, then its p-value is

$$p = \frac{B + 1 - r}{B + 1} \tag{3.7}$$

For the examples in this paper, $B = 10000$ simulations have been used. Further, the same permutation test has been used in Boyett and Shuster (1977).

An advantage of the union-intersection approach is that when H_0 is rejected, it is possible to identify the features that cause the rejection. Let $T_{UI}^{(\alpha)}$ denote the upper estimated critical value from the permutation simulation at level α (e.g., $\alpha = 0.05$). If the observed UI test statistic T_{UI} satisfies $T_{UI} > T_{UI}^{(\alpha)}$, then the "reason" for the rejection is given by those features j for which $T_{Cj} > T_{UI}^{(\alpha)}$.

4 Analysis of Data Set A: Cleft Lip

4.1 Details of the data

Data Set A was introduced in Section 1. It was collected to investigate the effectiveness of cleft lip surgery. A total of $K = 24$ landmarks have been collected on the lip periphery, with $K_P = 11$ landmark pairs and $K_S = 2$ solos. The landmarks are summarized in Table 3 and illustrated in Figure 5.

This data set contains $N = 25$ adult subjects, of which $N_1 = 12$ are control subjects (matched by age and sex to the cleft subjects from a larger group of control subjects) and $N_2 = 13$ are cleft subjects who had surgery as children. The $M = 3$ dimensional coordinates of these landmarks have been extracted at three time frames: first (closed lip), middle (middle of the smile) and last frame (maximum open lip smile). There are $J = MK_P + K_S = 3 \times 11 + 2 = 35$ elementary features in the signed and unsigned elementary feature vectors $\boldsymbol{d}$ and $\boldsymbol{a}$, respectively.

The rest of this section looks at the composite and union-intersection testing approaches on this data set.

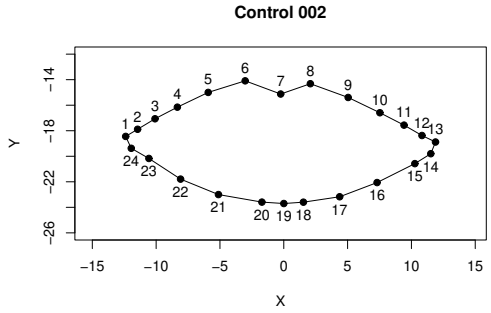

Figure 5: Data Set A. Landmark indices for x-y coordinates on the lip periphery of a control subject at first frame.

Table 3: Data Set A. Indices of landmark pairs and solos in Figure 5 on the lip periphery.

Landmark notation	Indices in Figure 5
pair (k_L, k_R)	(1,13), (2,12), (3, 11), (4,10), (5,9), (6,8), (20,18), (21,17), (22,16), (23,15), (24,14)
solo k_S	7, 19

4.2 The Composite Test

Throughout this section, the unsigned elementary feature vectors are summarized by a composite asymmetry score (2.8). Statistical tests to compare the control (Group 1) and cleft (Group 2) subjects are carried out using the one-sided Wilcoxon rank sum test (3.3). Comparisons between the two groups are made at the initial, middle and final time frames. A large value of the test statistic T_C indicates the cleft subjects are more asymmetric than the control subjects.

Consider the composite asymmetry score u given by (2.8). The corresponding test results are summarized in the first row of Table 4. The p-values are clearly significant at all three time frames, though they become less significant later in the smile. This result suggests that the difference in asymmetry between cleft and control subjects is most pronounced at the initial time frame and weakens over time. The greater asymmetry of the cleft subjects was not clinically surprising. However, there was no prior clinical intuition about whether or how the difference in asymmetry might change in time. Of course, with such a small data set, it is important not to over-interpret the results.

Figure 6 gives dot plots of the composite asymmetry score u to give a visual comparison between the two groups at the three time frames. One striking feature is the substantial overlap

Table 4: Data Set A. Two-sample comparison between cleft and control subjects subjects, with p-values for composite test (row 1) and union-intersection test (row 2) at three time frames.

Test	First frame	Middle frame	Last frame
composite (3.3)	0.0008(***)	0.026(*)	0.020(*)
union-intersection (3.6)	0.003(**)	0.063	0.18

(*) = significant at the 5% significance level. (**) = significant at the 1% significance level. (***) = significant at the 0.1% significance level.

between the two groups at all three times. That is, some cleft subjects are less asymmetric than some control subjects. However, for all three time frames the mean of the cleft patients is larger than the mean of the control subjects, and this difference decreases with time, a result consistent with the p-values. Also note the two outliers at the middle and last time frames, which correspond to the same cleft subject.

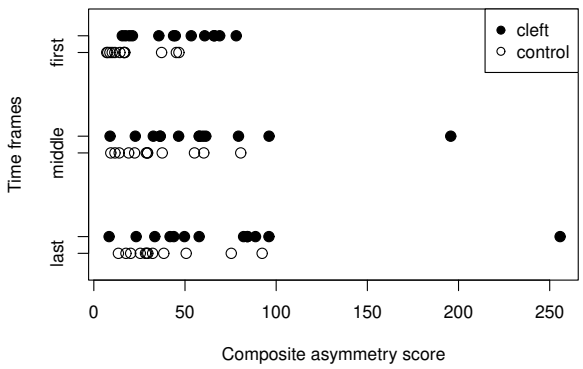

Figure 6: Data Set A. Dot plots of the composite asymmetry score u in (2.8) for both groups at the three time frames. The black dots are cleft subjects, and the open dots are control subjects.

4.3 The Union-Intersection Test

The p-values at the first two time frames for the union-intersection test also suggest that the cleft subjects have a significantly larger asymmetry than the control subjects, though the p-values are less extreme than for the composite tests. Hence for the first two time frames it is possible to ask which elementary features "explain" the rejection of the null hypothesis. At the first time frame, there are several elementary features that are significant at the $\alpha = 0.05$ level: the z-coordinate of landmark pairs (4, 10) and (5, 9), and y, z-coordinates of the pair (6, 8). At the middle time

frame, nothing is significant at the $\alpha = 0.05$ level, but there is a single elementary feature that is significant at the $\alpha = 0.10$ level: the z-coordinate of the landmark pair (6, 8), It is not surprising that the landmark pair (6, 8) stands out, because it lies on the philtrum, which is where cleft palate deformities are located. On the other hand, it is perhaps surprising that it is the z-coordinate that was found to be important in most of the significant features. This result merits further investigation by clinicians.

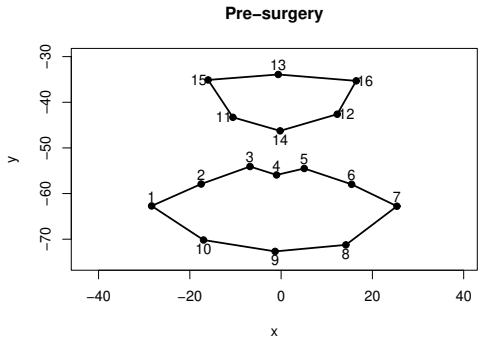

Figure 7: Data Set B. The 16 landmarks on the lip periphery and nasolabial region for a typical subject at the first frame pre-surgery.

Table 5: Data Set B. Indices of landmark pairs and solos in Figure 7 on the lip periphery and nasolabial region.

Landmark notation	Indices in Figure 7
pair (k_L, k_R)	(1,7), (2,6), (3, 5), (8,10), (11,12), (15,16)
k_S	4, 9, 13, 14

Table 6: Data Set B. Indices of these two sets of landmarks in Figure 7 which are used.

	Indices in Figure 7	K_P	K_S	J
Landmark set 1	1 to 7 (upper lip) and 11 to 16 (nasal region)	5	3	18
Landmark set 2	1 to 10 (lip periphery)	4	2	14

5 Analysis of Data Set B: Orthognathic Surgery

5.1 Details of the data and test statistics

Data Set B was introduced in Section 1. The aim of surgery is to reduce the jaw size. However, a possible unintended consequence is to increase the asymmetry around the lip and nasolabial

regions. A total of $K = 16$ landmarks have been collected on the lip periphery and nasolabial region, with $K_P = 6$ landmark pairs and $K_P = 4$ solos. The landmarks are summarized in Table 5 and illustrated in Figure 7. For each subject, there are $J = 3K_P + K_S = 22$ elementary features in $\boldsymbol{a}$. There are $N = 22$ subjects who have undergone orthognathic surgery.

The surgery mainly affects the upper lip and nasolabial region; therefore 13 landmarks at this region were selected, similar to Bains et al. (2025), as shown in Figure 7 and described in Table 6.

Note Data Set B consists of paired data whereas Data Set A consisted of two-group data. After adapting the two-group tests of Section 3 to the paired setting, comparisons between the pre- and post-surgery data for Data Set B are made at the same three time frames that were used for Data Set A.

Let $\boldsymbol{a}_1^{(n)}, \boldsymbol{a}_2^{(n)},\ n = 1, \ldots, N$ denote the unsigned elementary feature vectors for the pre- and post-surgery subjects, with composite asymmetry scores $u_1^{(n)}, u_2^{(n)}$. To carry out a paired comparison between the pre- and post-surgery composite asymmetry scores, the nonparametric Wilcoxon signed-rank test has been used. For the composite approach, the test statistic is given by

$$W_C = \sum_{n=1}^{N} \text{sign}(t^{(n)}) R^{(n)}, \tag{5.1}$$

where $t^{(n)} = u_2^{(n)} - u_1^{(n)}$ is the difference between the post- and pre-surgery composite asymmetry scores for subject n, and the $\{R^{(n)}\}$ are the ranks of the absolute values $\{|t^{(n)}|\}$. As before, a one-sided test is used with rejection if W_C is too large. The null distribution is given by randomizing over the 2^N possible signs of the $t^{(n)}$ values. This distribution is computed or approximated in standard statistics packages.

For the union-intersection test, a separate signed-rank test statistic W_{Cj}, say, is computed for each feature j, with an overall test statistic given by

$$W_{UI} = \max_{j=1,\ldots,J} W_{Cj}, \tag{5.2}$$

where again the null hypothesis is rejected for large values. In this case the null distribution must

be computed by simulation.

Table 7: Data Set B. Paired comparison between pre- and post-surgery subjects, with p-values for composite test (row 1) and union-intersection test (row 2) at three time frames.

Test	First frame	Middle frame	Last frame
composite (5.1)	0.088	0.007(**)	0.001(**)
union-intersection (5.2)	0.39	0.11	0.19

(*) = significant at the 5% significance level. (**) = significant at the 1% significance level. (***) = significant at the 0.1% significance level.

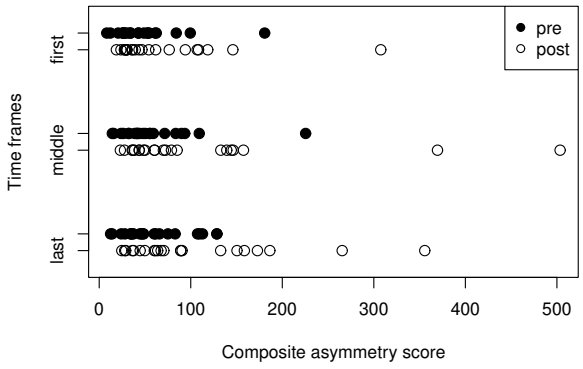

Figure 8: Data Set B. Dot plots of the composite asymmetry score u in (2.8) for the subjects pre-surgery (black dots) and post-surgery (open dots) at the three time frames.

5.2 Test Results for Data Set B

Data Set B was earlier analyzed in Bains et al. (2025), and the analysis here provides further insights. The p-values for the composite test at the three time frames are shown in the first row of Table 7. Notice the p-values are small and decreasing with time, suggesting that surgery leads to more asymmetry, with a greater increase in asymmetry at the later stages of the smile. The increase in asymmetry was also noted in Bains et al. (2025) using (2.10) and a t-test.

Dot plots of the composite asymmetry scores are given in Figure 8 and reinforce the message from the p-values. Although there is substantial overlap between the pre- and post-surgery data, the mean of the post-surgery is larger than the pre-surgery data and the difference increases with time.

The second row of Table 7 gives the p-values from the union-intersection test, with a different conclusion. For this test there is very little evidence that the asymmetry increases post-surgery; none of the test statistics is significant even at the $\alpha = 0.10$ level. Hence, there is no point looking for features to "explain" an increase in asymmetry.

The conflicting conclusions of the two tests may arise from the following explanation. Suppose the effect of surgery is to increase on average the asymmetry of many elementary features by a small amount. Then the cumulative effect of these increases may lead to a noticeable increase in the average composite asymmetry score and hence yield a significant result in the composite test. However, it may be the case that for no elementary feature does the change in asymmetry stand out from the noise. Hence the union-intersection test does not produce a significant result. Since the jaw surgery is not designed to change the asymmetry at any particular landmarks, this explanation may be appropriate here.

6 Discussion

The unsigned elementary feature vector $\boldsymbol{a}$ gives a multivariate description of asymmetry for a bilateral landmark configuration X. This paper gives a systematic construction of one-sided tests for such data. The assumption of stochastic ordering provides a framework to justify the pooling of information in the composite and union-intersection approaches. Each approach has its own strengths. Simple plots of the data are more straightforward for the composite approach, whereas the union-intersection approach can be used to identify important elementary features when significant changes in asymmetry have been detected.

An important application of one-sided multivariate testing is to the investigation of asymmetry in landmark-based medical imaging data. The first step is to construct a vector $\boldsymbol{a}$ of unsigned elementary features to describe the asymmetry of each subject. Then two different approaches (composite and union-intersection) have been used to construct test statistics based on the unsigned elementary feature vectors, both for two-group and paired comparisons.

The introduction of the concept of pose helps to clarify the issues involved in the registration of landmark-based configurations. Further, for the applications in this paper, it is important to use external registration to ensure that the results are medically relevant.

One topic for future work is a more systematic study of power. The null hypothesis fails when one group is more asymmetric than the other under the stochastic ordering assumption. It is difficult to make any general statements about power since the power depends heavily on the nature of the "signal". Two extreme scenarios can be considered. In the first scenario, each unsigned elementary feature is slightly larger on average in one group compared to the other. These small differences may accumulate to have a large effect on the composite test statistic. In the second scenario, there is a large difference on one unsigned elementary feature and no difference on the others. The union-intersection statistic will pick up the difference on this one feature. Hence the composite test might be expected to have higher power in the first scenario, and the union-intersection test might be expected to have higher power in the second scenario. As discussed in Section 5.2, the p-values in Table 7 suggest that for Data Set B we are closer to the first scenario than the second.

Another topic, relevant for the smile application, is to look at dynamic changes in asymmetry. The approach of this paper can be described as three static snapshots. Instead one could look at changing asymmetry through time. The examples of this paper involve the comparison between two sets of observations, either two groups or paired data. A natural extension is to look at multiple comparisons.

Acknowledgment

The authors would like to thank the anonymous referees for their very helpful comments, which have greatly improved the original submission.

References

Ajmera, D. H., Hsung, R. T., Singh, P., Wong, N. S. M., Yeung, A. W. K., Lam, W. Y. H., Khambay, B. S., Leung, Y. Y. and Gu, M. (2022) Three-dimensional assessment of facial asymmetry in class III subjects. Part 1: a retrospective study evaluating postsurgical outcomes. *Clinical Oral Investigations*, **26**, 4947–4966.

Ajmera, D. H., Zhang, C., Ng, J. H. H., Hsung, R. T., Lam, W. Y. H., Wang, W., Leung, Y. Y., Khambay, B. S. and Gu, M. (2023) Three-dimensional assessment of facial asymmetry in class III subjects. Part 2: evaluating asymmetry index and asymmetry scores. *Clinical Oral Investigations*, **27**, 5813–5826.

Bains, A., Bakshi, A., Gu, M. and Khambay, B. (2025) Three-dimensional (3D) dynamic changes in nasolabial symmetry following Le Fort I maxillary advancement in Class III patients. *Advances in Oral and Maxillofacial Surgery*, **20**, 100603. URL: `https://www.sciencedirect.com/science/article/pii/S2667147625000913`.

Bock, M. T. and Bowman, A. W. (2006) On the measurement and analysis of asymmetry with applications to facial modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **55**, 77–91.

Boyett, J. M. and Shuster, J. J. (1977) Nonparametric one-sided tests in multivariate analysis with medical applications. *Journal of the American Statistical Association*, **72**, 665–668.

Dryden, I. L. and Mardia, K. V. (2016) *Statistical Shape Analysis: With Applications in R*. Chichester: Wiley.

Hartley, R. and Zisserman, A. (2004) *Multiple View Geometry in Computer Vision*. Cambridge, UK: Cambridge University Press, 2 edn.

Kent, J. T. and Mardia, K. V. (2001) Shape, Procrustes tangent projections and bilateral symmetry. *Biometrika*, **88**, 469–485.

Mardia, K. V., Bookstein, F. L. and Moreton, I. J. (2000) Statistical assessment of bilateral symmetry of shapes. *Biometrika*, **87**, 285–300.

Mardia, K. V. and Wu, X. (2025) On the distribution of weighted sum of two chi-squares with applications to shape analysis. *Sankhya Series A, a special issue: C. R. Rao, Published Online*. URL: `https://link.springer.com/article/10.1007/s13171-025-00415-8`.

Mardia, K. V., Wu, X., Kent, J. T., Goodall, C. R. and Khambay, B. S. (2026) Asymmetry analysis of bilateral shapes. URL: `https://arxiv.org/abs/2407.17225`.

Patel, Y., Sharp, I., Enocson, L. and Khambay, B. S. (2023) An innovative analysis of nasolabial dynamics of surgically managed adult patients with unilateral cleft lip and palate using 3D facial motion capture. *Journal of Plastic, Reconstructive and Aesthetic surgery*, **85**, 287–298.

Stoyan, D. (1983) *Comparison Methods for Queues and Other Stochastic Models*. Wiley.

Weyl, H. (1952) *Symmetry*. Princeton University Press.