Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Sex estimation from the three-dimensional nasal septum on computed tomography: an allometric, size-based signal in 1,373 adults

**Michael Sramek**

sramek.michael@mayo.edu

Mayo Clinic in Arizona

**Hannah Gibbs**

Mayo Clinic in Arizona

**India Casaday**

Mayo Clinic in Arizona

**Annica Eells**

Mayo Clinic in Arizona

**Amar Miglani**

Mayo Clinic in Arizona

**Stephen Bansberg**

Mayo Clinic in Arizona

**Tanya Rath**

Mayo Clinic in Arizona

**Michael Marino**

Mayo Clinic in Arizona

**Research Article**

**Keywords:** sex estimation, forensic anthropology, nasal septum, computed tomography, allometry, machine learning

**Posted Date:** November 2nd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11309043/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Abstract

Sex estimation from computed tomography (CT) of the head has relied on manual measurements of the skull, paranasal sinuses, and nasal cavity, with modest accuracy and small samples. The nasal septum, the midline osteocartilaginous partition of the nasal cavity, is present on every head CT but has never been evaluated as a sex estimator, and whether any sex signal in the septum reflects size or shape determines how such an estimator would transfer between populations. We tested two hypotheses in 1,373 adults (394 men, 979 women) with radiologically normal sinus CT: that septal morphometry classifies sex at least as accurately as established cranial methods, and that the discriminating information is allometric size rather than shape. The septum was segmented automatically with a validated convolutional network (held-out Dice 0.983) and 17 morphometric variables were derived from each segmentation. Septal volume was 34% larger in men (16,852 versus 12,564 $mm^3$; Cohen's d 2.33, 95% CI 2.18 to 2.47). Under five-fold cross-validation, sex was classified with an area under the curve (AUC) of 0.95 by random forest and 0.96 by logistic regression on 17 variables; volume alone gave 0.94. After allometric adjustment no shape variable differed by more than d = 0.43, and shape variables alone gave an AUC of 0.58 to 0.63. Both hypotheses were supported: the septum is a strong, fully automatable sex estimator for CT, and its signal is size. Like other size-based skeletal estimators it will require population-specific calibration, which the automated pipeline makes practical.


# Introduction

Sex estimation is a first step in the identification of unknown human remains, and the skull and pelvis remain its principal substrates [1, 2]. Computed tomography (CT) has broadened the method in two ways. Antemortem clinical CT provides large documented-sex samples of living populations from which estimators can be built, and postmortem CT (PMCT), now routine in many forensic institutes, brings those estimators to the deceased before, or instead of, skeletonization [3–6]; clinical CT archives are also increasingly matched against PMCT for identification [7]. In the midface, CT studies of the paranasal sinuses, nasal cavity, and piriform aperture have reported sex differences that support classification, but their accuracy is modest and their samples small. A meta-analysis of 18 studies found that the best single sinus measurements reached sensitivities and specificities of only about 70% [8]; discriminant functions on sinus volumes give 78% to 88% [9]; and the two most recent machine-learning studies, in 195 and 220 adults, reached an area under the curve (AUC) of 0.77 and an accuracy of 84% [10, 11]. All of these methods depend on manual landmarks placed on single slices, which limits sample size and introduces observer error.

The nasal septum has not been evaluated as a sex estimator. It is the midline partition of the nasal cavity, a composite of the quadrangular cartilage anteriorly and the vomer and perpendicular plate of the ethmoid posteriorly, and it is captured on every sinus, maxillofacial, and head CT. It is absent from skeletal standards because the cartilage does not survive and the bony plates are thin and easily damaged, and it has been difficult to measure in three dimensions by hand. The only series to report septal surface area by sex, 100 adults measured manually on CT, found larger values in men but did not

test the difference formally [12]; the perpendicular plate alone has been reported as larger in men in 104 patients [13]; and no study has reported septal volume by sex or evaluated the septum as a sex estimator [14]. Automated segmentation removes the manual bottleneck. We trained and validated a convolutional network that segments the entire osteocartilaginous septum from clinical CT and used it to measure 1,373 adults, a sample an order of magnitude larger than previous midfacial CT studies.

Beyond accuracy, the forensic value of any estimator depends on whether its signal is size or shape. Size-based skeletal estimators are population-specific and track body size, so a threshold derived in one population misclassifies in another [2, 15–17]; shape-based differences are usually smaller and are not population-invariant either, but they appear more uniform across populations than size-and-shape form [18–20]. For the septum, two bodies of evidence make different predictions. The septal cartilage is a growth center of the midface [21–23], and much of craniofacial sexual dimorphism arises from sex differences in growth duration and overall size rather than from sex-specific shape, with allometry accounting for more than half of facial dimorphism in recent multi-population work [24–27], which predicts a septum that differs between the sexes in size with little residual shape difference. Against this, the nasal airway is functionally integrated with ventilatory demand, and men have relatively longer and taller nasal passages than women at a given size [28–30], which predicts residual shape differences in the anterior-posterior and cranial-caudal dimensions of the septum beyond those expected from scaling.

We therefore tested two hypotheses. Hypothesis 1: septal morphometry classifies sex at least as accurately as established cranial estimators, which we operationalized as a cross-validated AUC of 0.90 or higher. Hypothesis 2: the discriminating information is allometric size rather than shape, with three predictions: (a) after adjustment for septal volume no variable differs between the sexes by more than a small effect ($|d| < 0.5$); (b) classifiers built on size-adjusted shape variables alone perform near chance (AUC < 0.70); and (c) septal volume alone classifies sex as accurately as multivariable machine-learning models. The functional alternative predicts the opposite of (a) to (c) for anterior-posterior depth and cranial-caudal height. Measurement reproducibility was a secondary aim, because an estimator intended for forensic casework must be deterministic and auditable.

## Materials and methods

# Study design, population, and ethics

This retrospective cross-sectional study was conducted at a single tertiary academic medical center under institutional review board approval (Mayo Clinic IRB 25-008837) with a waiver of informed consent for use of de-identified imaging and record data; no participants were prospectively enrolled, and the study conforms to the Declaration of Helsinki. Reporting follows the STROBE recommendations for cross-sectional studies and, for the classification models, the TRIPOD + AI statement [31].

Adults who underwent non-contrast sinus CT between 2018 and 2025 with a radiology impression of "normal" or equivalent were identified. Examinations with documented prior nasal or sinus surgery, cleft palate, or septal perforation on the index study were excluded after line-by-line review of the radiology impression. All examinations were acquired under routine clinical protocols (axial slice thickness 1.0 mm or less; multiple scanners and manufacturers) and reconstructed with a bone kernel.

The cohort derivation is given in Online Resource 1, Table S1. Of 1,673 segmented septa, 1,567 belonged to individuals with sex recorded in the institutional record. One examination per individual was retained (23 repeat examinations removed), individuals with a septoplasty or septal repair recorded before the index scan were excluded (n = 27), one segmentation that failed automated quality control was removed, and individuals younger than 18 years at the time of the scan were excluded (n = 31). The primary analytic cohort comprised the 1,373 individuals whose age at scan could be confirmed from the radiology record. The 112 further individuals with recorded sex but no retrievable scan date (so that adult status and age could not be confirmed) were included only in a sensitivity analysis (n = 1,485).

Sex was taken as the sex recorded in the electronic health record, an administrative binary that in this setting corresponds to sex assigned at birth for the overwhelming majority of individuals but was not independently verified; we use "men" and "women" in that sense throughout. Age was computed from date of birth and scan date. Race and ethnicity are reported as the self-identified categories recorded in the health record and are used only to describe the composition of the cohort.

## Segmentation

Septa were segmented with a three-dimensional residual-encoder nnU-Net [32], version 2, in the full-resolution configuration. The model was trained on 125 examinations manually segmented by an otolaryngology resident in 3D Slicer [33] under iterative review by a fellowship-trained rhinologist, with five-fold cross-validation (mean Dice 0.957). On 24 held-out examinations the ensemble achieved a Dice similarity coefficient of 0.983 (SD 0.005), a 95th-percentile Hausdorff distance of 1.82 mm (SD 0.43), and an average symmetric surface distance of 0.31 mm (SD 0.08) [34]. The segmentation protocol captured the entire osteocartilaginous septum from the nasal floor to the skull base and from the membranous septum to the sphenoid rostrum, including the quadrangular cartilage, vomer, and perpendicular plate of the ethmoid (Fig. 1). Training used the framework's default configuration (stochastic gradient descent with Nesterov momentum, initial learning rate 0.01 with polynomial decay, combined Dice and cross-entropy loss, deep supervision, standard nnU-Net augmentation, 1,000 epochs per fold) on a workstation with a 16 GB NVIDIA graphics card. The trained ensemble, with test-time mirroring, was applied to all remaining examinations; output masks were screened by automated rules for volume, extent, and connectivity, and flagged masks were reviewed manually. Trained model weights are available from the corresponding author under the conditions given in the data availability statement.

## Morphometric variables

Seventeen variables were computed deterministically from each binary mask in Python using SimpleITK, scikit-image, trimesh, and NumPy, with the scanner-defined right-anterior-superior axes (Table 2). Volume was the sum of voxel volumes. Surface area was computed on a marching-cubes mesh. The three linear extents were the anterior-posterior depth, cranial-caudal height, and medial-lateral extent of the mask bounding box, and the height-to-width ratio was cranial-caudal height divided by anterior-posterior depth. Parasagittal cross-sectional area (CSA) was sampled on every sagittal slice through the mask at the native in-plane resolution, and the maximum, mean, minimum, and standard deviation across slices were recorded; the slice of maximal CSA corresponds to the plane shown in Fig. 1b. Midline deviation was quantified on coronal slices: the medial-lateral centroid of the septum was computed for each cranial-caudal level, a straight line was fitted to these centroids by least squares, and the perpendicular distances of the centroids from that line were summarized as the maximum, mean, and standard deviation of deviation. The deviation angle is the angle between the fitted line and the cranial-caudal axis; the tortuosity index is the path length of the centroid trace divided by its chord; the medial-lateral range is the range of centroid positions; and the deviation asymmetry index is the normalized imbalance between leftward and rightward deviation. Voxel count and the three centroid coordinates, which depend on scanner resolution and patient position rather than on anatomy, were computed but not used as morphometric variables.

# Statistical analysis

Sex differences in each variable were summarized as means with standard deviations, the percentage difference relative to women, Welch's t-test, and Cohen's d with the pooled standard deviation and its 95% confidence interval (CI). P-values across the 17 variables were adjusted by the Holm method. To test whether dimorphism was independent of age, each variable was also regressed on sex and age, and on sex, age, and their interaction, with the age-adjusted sex coefficient re-expressed as a standardized difference.

Allometric adjustment followed the classical size-and-shape framework [35, 36]. Septal volume was taken as the measure of size. For each of the 12 variables with units of length or area, the natural logarithm of the variable was regressed on the natural logarithm of volume across the pooled sample, and the residual was taken as the size-adjusted (shape) variable; the fitted slope was compared with the isometric expectation (1/3 for lengths, 2/3 for areas). The four dimensionless variables (height-to-width ratio, deviation angle, tortuosity index, and asymmetry index) were used unchanged. Cohen's d was then recomputed for these 16 size-adjusted variables (prediction 2a).

Sex classification used a random-forest classifier [37] with 500 trees, class-balanced sample weighting, and otherwise default settings, and a logistic-regression model with standardized inputs and class-balanced weighting, both implemented in scikit-learn [38]. Models were evaluated by stratified five-fold cross-validation with a fixed random seed, and all performance measures were computed on the pooled out-of-fold predictions. The primary metric was the AUC, with a 95% CI from 1,000 bootstrap resamples of the out-of-fold predictions. Because the cohort was 71% women, we report balanced accuracy,

sensitivity (proportion of men correctly classified), and specificity (proportion of women correctly classified) at the default probability threshold of 0.5 and at the Youden-optimal threshold, in addition to overall accuracy. Variable importance was summarized by mean Gini importance across folds and, because Gini importance is biased toward continuous variables with many split points [39], by the mean decrease in out-of-fold AUC when each variable was permuted.

To partition classification accuracy between size and shape (predictions 2b and 2c), the following nested comparisons were pre-specified: (i) volume alone; (ii) all 17 variables; (iii) the 16 size-adjusted shape variables alone, using isometric scaling (lengths divided by volume$^{1/3}$, areas by volume$^{2/3}$) so that no linear combination of inputs could reconstruct size; and (iv) the shape variables plus volume$^{1/3}$. The difference in AUC between (ii) and (i) was tested by paired bootstrap. Sensitivity analyses repeated the main model with age added, without volume, on the original 22-variable set including the non-anatomical variables, and on the 1,485-person cohort including individuals of unconfirmed age. Analyses used Python 3.10 with scikit-learn 1.7, SciPy, and statsmodels; two-sided alpha was 0.05.

## Use of large language models

A large language model (Claude, Anthropic) was used under the direction and review of the authors to assist with reviewing the author provided python analysis scripts and with editing the manuscript text. The authors designed the study, chose the variables and analyses, interpreted the results, verified the code and every reported value, and take full responsibility for the content.

## Results

# Cohort

The primary cohort comprised 1,373 adults: 394 men (28.7%) and 979 women (71.3%) (Table 1). Mean age was 54.5 years (SD 16.5; range 18 to 90) and did not differ by sex (men 54.4, women 54.5; P = 0.87). By health-record categories, 90.8% of individuals were recorded as White and 91.2% as not Hispanic or Latino. All but one septum in the primary cohort was segmented automatically.

Table 1
Characteristics of the primary cohort and sex-specific percentiles of septal volume

| Characteristic | Men | Women | All |
|---|---|---|---|
| n | 394 | 979 | 1,373 |
| Age at scan, years, mean (SD) | 54.4 (17.1) | 54.5 (16.2) | 54.5 (16.5) |
| Age range, years | 18 to 90 | 18 to 90 | 18 to 90 |
| Recorded as White, n (%) | 349 (88.6) | 897 (91.6) | 1,246 (90.8) |
| Recorded as Black or African American, n (%) | 9 (2.3) | 21 (2.1) | 30 (2.2) |
| Recorded as Asian, n (%) | 5 (1.3) | 10 (1.0) | 15 (1.1) |
| Other, multiracial, or not recorded, n (%) | 31 (7.9) | 51 (5.2) | 82 (6.0) |
| Recorded as Hispanic or Latino, n (%) | 26 (6.6) | 68 (6.9) | 94 (6.8) |
| Charlson comorbidity index, median (IQR) | 1 (0 to 3) | 1 (0 to 2) | 1 (0 to 2) |
| Septal volume, 5th percentile, $mm^3$ | 13,594 | 9,963 | 10,203 |
| Septal volume, 25th percentile, $mm^3$ | 15,372 | 11,362 | 11,810 |
| Septal volume, 50th percentile, $mm^3$ | 16,710 | 12,500 | 13,303 |
| Septal volume, 75th percentile, $mm^3$ | 18,298 | 13,661 | 15,357 |
| Septal volume, 95th percentile, $mm^3$ | 20,455 | 15,329 | 18,748 |

Race and ethnicity are the self-identified categories recorded in the electronic health record and describe the composition of the cohort only. Age did not differ by sex (Welch P = 0.87). SD standard deviation, IQR interquartile range

## Magnitude of sexual dimorphism

Every size-related variable was larger in men (Table 2; Figs. 2 and 4). Mean septal volume was 16,852 $mm^3$ (SD 2,183) in men and 12,564 $mm^3$ (SD 1,686) in women, a 34% difference (d = 2.33, 95% CI 2.18 to 2.47). Surface area was 19% larger (d = 2.19, 95% CI 2.05 to 2.33) and anterior-posterior depth 10% larger (d = 1.79, 95% CI 1.65 to 1.92). Mean parasagittal CSA (d = 1.31), cranial-caudal height (d = 1.02), maximum CSA (d = 0.99), the standard deviation of CSA (d = 0.99), and medial-lateral extent (d = 0.74) showed large to moderate differences. The sex distributions of volume overlapped little: the 5th percentile of male volume (13,594 $mm^3$) was close to the 75th percentile of female volume (13,661 $mm^3$), 2.0% of men fell below the female median, and 1.3% of women exceeded the male median.

Table 2
Sex differences in 17 three-dimensional morphometric variables of the nasal septum (n = 1,373)

| Variable (units) | Men (n = 394), mean (SD) | Women (n = 979), mean (SD) | Difference (%) | Cohen's d (95% CI) | Age-adjusted d | Holm P |
|---|---|---|---|---|---|---|
| Septal volume ($mm^3$) | 16,852 (2,183) | 12,564 (1,686) | 34.1 | 2.33 (2.18 to 2.47) | 2.34 | < 0.001 |
| Surface area ($mm^2$) | 8,781 (682) | 7,409 (603) | 18.5 | 2.19 (2.05 to 2.33) | 2.21 | < 0.001 |
| Anterior-posterior depth (mm) | 92.7 (4.9) | 84.3 (4.7) | 10.0 | 1.79 (1.65 to 1.92) | 1.80 | < 0.001 |
| Cranial-caudal height (mm) | 54.2 (4.6) | 49.9 (4.0) | 8.6 | 1.02 (0.90 to 1.15) | 1.03 | < 0.001 |
| Medial-lateral extent (mm) | 16.5 (2.3) | 14.9 (2.1) | 10.7 | 0.74 (0.62 to 0.86) | 0.75 | < 0.001 |
| Height-to-width ratio | 3.34 (0.49) | 3.41 (0.50) | -2.0 | -0.14 (-0.26 to -0.02) | -0.14 | 0.098 |
| Maximum sagittal CSA ($mm^2$) | 2,661 (477) | 2,231 (413) | 19.3 | 0.99 (0.87 to 1.12) | 0.99 | < 0.001 |
| Mean sagittal CSA ($mm^2$) | 1,015 (155) | 836 (129) | 21.4 | 1.31 (1.18 to 1.43) | 1.31 | < 0.001 |
| Minimum sagittal CSA ($mm^2$) | 4.07 (4.16) | 4.26 (4.69) | -4.5 | -0.04 (-0.16 to 0.07) | -0.04 | 0.919 |
| SD of sagittal CSA ($mm^2$) | 909 (169) | 759 (144) | 19.8 | 0.99 (0.87 to 1.12) | 0.99 | < 0.001 |
| Maximum midline deviation (mm) | 1.55 (0.85) | 1.34 (0.78) | 15.5 | 0.26 (0.14 to 0.38) | 0.26 | < 0.001 |
| Mean midline deviation (mm) | 0.455 (0.214) | 0.394 (0.191) | 15.4 | 0.31 (0.19 to 0.42) | 0.31 | < 0.001 |
| SD of midline deviation (mm) | 0.330 (0.162) | 0.287 (0.144) | 15.3 | 0.29 (0.18 to 0.41) | 0.29 | < 0.001 |
| Medial-lateral range of midline (mm) | 3.61 (1.87) | 3.23 (1.63) | 11.7 | 0.22 (0.11 to 0.34) | 0.22 | 0.003 |
| Deviation angle (°) | 0.626 (3.030) | 0.434 (3.002) | 44.0 | 0.06 (-0.05 to 0.18) | 0.06 | 0.868 |

| Variable (units) | Men (n = 394), mean (SD) | Women (n = 979), mean (SD) | Difference (%) | Cohen's d (95% CI) | Age-adjusted d | Holm P |
|---|---|---|---|---|---|---|
| Tortuosity index | 1.022 (0.018) | 1.020 (0.018) | 0.1 | 0.08 (-0.03 to 0.20) | 0.09 | 0.634 |
| Deviation asymmetry index | 0.115 (0.085) | 0.118 (0.089) | -2.9 | -0.04 (-0.16 to 0.08) | -0.04 | 0.919 |

Difference is the male minus female mean as a percentage of the female mean. Cohen's d uses the pooled standard deviation. Age-adjusted d is the sex coefficient from a linear model with age, divided by the residual standard deviation. P-values are from Welch's t-test with Holm adjustment across the 17 variables. CSA parasagittal cross-sectional area, SD standard deviation, CI confidence interval

Midline deviation variables differed only slightly (maximum, mean, and SD of deviation, d = 0.26 to 0.31; medial-lateral range, d = 0.22), and the dimensionless indices did not differ after adjustment for multiple testing: height-to-width ratio d = -0.14 (Holm-adjusted P = 0.10), minimum CSA d = -0.04, deviation angle d = 0.06, tortuosity index d = 0.09, and asymmetry index d = -0.04.

Age-adjusted standardized differences were indistinguishable from the unadjusted values (for example, volume d = 2.34 after adjustment; Table 2), and no sex-by-age interaction survived adjustment for multiple testing. Volume correlated weakly and positively with age in men (Spearman rho = 0.18) and negligibly in women (rho = 0.04; Online Resource 1, Fig. S1).

# Size versus shape (hypothesis 2a)

Allometric slopes of the length and area variables on volume were below isometry for surface area (0.49 versus 0.67), anterior-posterior depth (0.26 versus 0.33), and cranial-caudal height (0.22 versus 0.33), indicating that larger septa are proportionally thicker and more compact rather than simply scaled-up (Online Resource 1, Table S2). After adjustment for volume, the residual sex differences were small (Fig. 4; Table S2): surface area d = 0.43 (95% CI 0.31 to 0.55), anterior-posterior depth d = 0.38 (0.26 to 0.50), cranial-caudal height d = 0.22 (0.10 to 0.34), and minimum CSA d = -0.16 (-0.27 to -0.04); no other size-adjusted variable differed significantly. At a given volume, therefore, male septa are slightly longer, taller, and larger in surface area, and female septa slightly thicker at their thinnest point, but these shape differences are an order of magnitude smaller than the size difference and all fall below the pre-specified threshold of |d| = 0.5.

# Sex classification (hypothesis 1)

The random-forest classifier discriminated sex from the 17 morphometric variables with an out-of-fold AUC of 0.946 (95% CI 0.932 to 0.960; per-fold range 0.919 to 0.973), overall accuracy of 90.2%, and balanced accuracy of 86.7% (95% CI 84.4 to 88.9) (Table 3; Fig. 3). At the default threshold the model

was conservative in calling men (sensitivity 0.78, specificity 0.95); at the Youden-optimal threshold of 0.33, sensitivity was 0.86 and specificity 0.91 (balanced accuracy 0.88). Logistic regression on the same variables performed at least as well (AUC 0.956, 95% CI 0.944 to 0.968; balanced accuracy 88.9%; sensitivity 0.88, specificity 0.89), and the two models' out-of-fold probabilities were highly correlated (r = 0.96), indicating that the signal is essentially linear.

# Size versus shape in classification (hypotheses 2b and 2c)

Volume alone classified sex with an AUC of 0.942 (95% CI 0.927 to 0.954) and balanced accuracy of 86.7%. Adding the 16 remaining variables to volume raised the logistic-regression AUC by 0.015 (paired bootstrap 95% CI 0.006 to 0.023). Conversely, the size-adjusted shape variables alone gave an AUC of 0.575 (95% CI 0.543 to 0.606) by random forest and 0.630 (0.598 to 0.660) by logistic regression, and restoring a single size variable to the shape set returned the AUC to 0.939. Removing volume from the full set had little effect (AUC 0.944), because surface area and the linear extents carry the same size information.

Variable importance was concentrated in size (Table 3): volume (Gini importance 0.265), surface area (0.211), and anterior-posterior depth (0.145) together accounted for 62% of total importance, and were the only variables whose permutation reduced out-of-fold AUC by more than 0.002. Standardized logistic coefficients gave the same ranking (Online Resource 1, Table S3).

Table 3
Variable importance in the random-forest sex classifier, with the univariate effect size for comparison

| Rank | Variable | Family | Gini importance, mean (SD) | Permutation importance (decrease in AUC) | Cohen's d |
|---|---|---|---|---|---|
| 1 | Septal volume ($mm^3$) | Size | 0.265 (0.011) | 0.048 | 2.33 |
| 2 | Surface area ($mm^2$) | Size | 0.211 (0.011) | 0.019 | 2.19 |
| 3 | Anterior-posterior depth (mm) | Linear | 0.145 (0.006) | 0.020 | 1.79 |
| 4 | Mean sagittal CSA ($mm^2$) | Cross-sectional | 0.061 (0.001) | 0.001 | 1.31 |
| 5 | Medial-lateral extent (mm) | Linear | 0.042 (0.001) | 0.001 | 0.74 |
| 6 | SD of sagittal CSA ($mm^2$) | Cross-sectional | 0.041 (0.001) | 0.002 | 0.99 |
| 7 | Maximum sagittal CSA ($mm^2$) | Cross-sectional | 0.036 (0.003) | 0.000 | 0.99 |
| 8 | Cranial-caudal height (mm) | Linear | 0.034 (0.003) | 0.000 | 1.02 |
| 9 | Mean midline deviation (mm) | Deviation | 0.022 (0.002) | 0.000 | 0.31 |
| 10 | Height-to-width ratio | Proportion | 0.021 (0.001) | 0.000 | -0.14 |
| 11 | Deviation angle (°) | Deviation | 0.019 (0.002) | -0.001 | 0.06 |
| 12 | Minimum sagittal CSA ($mm^2$) | Cross-sectional | 0.018 (0.001) | 0.000 | -0.04 |
| 13 | Deviation asymmetry index | Deviation | 0.018 (0.002) | -0.001 | -0.04 |
| 14 | SD of midline deviation (mm) | Deviation | 0.017 (0.001) | -0.001 | 0.29 |
| 15 | Tortuosity index | Deviation | 0.017 (0.001) | 0.000 | 0.08 |
| 16 | Medial-lateral range of midline (mm) | Deviation | 0.017 (0.001) | -0.001 | 0.22 |
| 17 | Maximum midline deviation (mm) | Deviation | 0.016 (0.001) | -0.001 | 0.26 |

Gini importance is averaged across the five cross-validation folds and sums to 1 across variables. Permutation importance is the mean decrease in out-of-fold AUC when the variable is permuted in the held-out fold (8 repeats per fold). Negative values within sampling error are shown as 0.000. AUC area under the receiver operating characteristic curve, CSA parasagittal cross-sectional area, SD standard deviation

Results were unchanged in sensitivity analyses (Online Resource 1, Table S4): with age added (AUC 0.951), with the original 22-variable set that included voxel count and centroid coordinates (0.951), and in the 1,485-person cohort including individuals of unconfirmed age (0.945, 95% CI 0.931 to 0.958).

## Discussion

Both hypotheses were supported. Septal morphometry classified sex with a cross-validated AUC of 0.95 to 0.96, above the pre-specified benchmark of 0.90 and above the accuracies reported for CT measurements of neighboring midfacial structures. And the signal is size: after allometric adjustment no variable differed between the sexes by more than d = 0.43, shape variables alone classified sex at an AUC of 0.58 to 0.63, and septal volume by itself (AUC 0.94) matched multivariable random-forest and logistic models. The functional alternative received only weak support, in the form of small residual differences in the predicted direction (at a given volume, male septa are slightly longer and taller), which contributed almost nothing to classification. For forensic purposes the male septum is, to a close approximation, a larger female septum.

# Comparison with existing estimators

The accuracy obtained here is at the upper end of what has been reported from the head. Metric and visual cranial methods reach 85% to 92% accuracy when several measurements or traits are combined [2, 16, 40, 41], and the best single cranial dimension, bizygomatic breadth, gives about 88% on 3D CT [42]. Multivariable and machine-learning models on cranial CT measurements reach 94% to 96% [6, 42, 43], geometric morphometrics of the whole cranium up to 98% [44], and deep learning on skull CT or PMCT an AUC of 0.95 to 0.97 [4, 5, 45]. In the midface, paranasal sinus and nasal cavity measurements have given single-measurement sensitivities and specificities near 70% in meta-analysis [8], 78% to 88% from sinus discriminant functions [9], and an AUC of 0.77 to 0.84 in the most recent machine-learning studies [10, 11]. The septum reaches a balanced accuracy of 87% to 89% and an AUC of 0.95 from a single automated measurement, matching the multivariable cranial models and well above every midfacial method, in a sample six times larger than the largest of the midfacial studies. Univariate effect sizes are rarely tabulated for cranial measurements, but the septal volume effect (d = 2.33) exceeds the largest reported in the nasal region, nasal airway volume (d = 1.51) and surface area (d = 1.80) [46].

Three features of the method matter for casework. The measurement is fully automated and deterministic, so inter-observer error, the main limitation of landmark-based midfacial methods, is removed. A single variable carries almost all of the information, so the estimator can be a simple volume

threshold or a one-variable logistic rule that is easy to audit and to recalibrate. And the structure is in the center of the head, protected from the surface damage that compromises thin cranial features, although, as discussed below, its cartilaginous part depends on soft-tissue preservation.

# Why the signal is size, and what that implies

The predominance of size is consistent with the developmental account of craniofacial dimorphism, in which sex differences arise mainly from sex-specific growth trajectories and allometric scaling rather than from sex-specific shape [24–26], and with the role of the septal cartilage as a growth center that scales with the midface [21]. Bastir et al. [30] reported strong dimorphism in the three-dimensional form of the upper airway soft tissues that was likewise dominated by size; our results extend that pattern to the osteocartilaginous septum and quantify it on a much larger sample. The pattern is not universal across the head: in the cranium the most dimorphic regions, the glabella and the mastoid, differ in shape rather than size [18], and in the pharyngeal airway sex differences in three-dimensional shape accompany those in volume [47]. The septum therefore sits at the size-driven end of a spectrum along which craniofacial structures express sexual dimorphism differently. The negative allometry observed here, with surface area and linear extents scaling below isometry, implies that larger septa are relatively thicker; this is what would be expected if septal thickness tracks the mechanical demands of a larger nasal cavity, but that reading was not tested. The small residual differences in depth and height are in the direction predicted by the respiratory account of relatively longer nasal passages in men [29, 30], but their magnitude shows that, in the septum, function adds little to scaling, in the same way that the nasal airway itself differs greatly in volume and surface area between the sexes while its surface-to-volume ratio does not [46]. The midline deviation variables were nearly indistinguishable between the sexes, consistent with earlier CT work finding no sex difference in the magnitude of septal deviation [48], so the common deviated septum is not a sexually dimorphic trait.

The forensic consequence follows directly. Because the dimorphism is allometric, a septal estimator will be sensitive to population differences in body and facial size in exactly the way that other size-based skeletal estimators are [2, 15], and the volume thresholds in Table 1 describe this cohort only. Shape-based information, which might have transferred better between populations, is too weak in the septum to build on. The practical response is population-specific calibration, and the automated pipeline makes that unusually cheap: a documented-sex CT collection from any population can be segmented and a local threshold derived without manual measurement. The appropriate next steps are external validation on clinical CT collections of documented sex from other populations, validation on PMCT, and a head-to-head comparison against established cranial estimators in the same individuals.

# Methodological notes

Two features of the analysis are worth stating explicitly. First, random forests did not outperform logistic regression; the sex signal is linear and dominated by one variable, and we report the random forest for comparability with the recent literature rather than because it is needed. Second, we did not use regression residuals as inputs to the shape-only classifier, because a linear combination of several

allometric residuals with different slopes can partially reconstruct size; the isometric-ratio inputs used here avoid that leakage, and the nested comparison of volume alone against the full variable set is the cleaner statement of how much shape adds.

# Limitations

This is a single-institution study of a clinical population, predominantly recorded as White and not Hispanic or Latino, imaged for sinonasal symptoms; the thresholds reported here should not be applied to other populations without validation, and the cohort is not a probability sample of any population. Sex was taken from the health record and was not independently verified. Body size, stature, and craniofacial dimensions were not available, so septal dimorphism could not be partitioned between a general body-size effect and a local facial effect. The morphometric variables are mask-derived summary measures rather than landmark-based geometric morphometrics, so shape was assessed through ratios, residuals, and deviation metrics rather than through a Procrustes shape space. Cross-validation estimates were obtained within one institution; external validation has not yet been performed. Finally, the segmentation operates on CT of the intact head. In skeletonized remains the cartilaginous septum is absent and the bony plates are often broken; the estimator is therefore suited to antemortem CT and to PMCT of fleshed remains, and a version restricted to the vomer and perpendicular plate would be needed for dry bone.

## Conclusion

The nasal septum is a strongly sexually dimorphic structure on CT, and its dimorphism is almost entirely one of size. Septal volume alone classifies sex with an AUC of 0.94, equaling multivariable machine-learning models, whereas size-adjusted shape carries little sex information. Automated septal volumetry is therefore a candidate size-based sex estimator for antemortem and postmortem CT, subject to the same population-specific calibration that governs other skeletal methods.

## Declarations

# Competing interests.

The authors declare no competing interests.

# Human Ethics and Consent to Participate declarations.

This retrospective analysis of existing clinical imaging was conducted in accordance with the Declaration of Helsinki and its later amendments and with the ethical standards of the Mayo Clinic Institutional Review Board, which approved the study (protocol 25-008837). Because the study used only de-identified imaging and record data collected during routine care, the Institutional Review Board granted a waiver of the requirement for informed consent to participate.

# Consent to publish.

Not applicable; no identifiable individual data or images are published.

# Funding.

No funding was received for this study.

# Author Contribution

Conceptualization: MS, MM. Data curation: MS, HG, IC, AE. Formal analysis: MS. Investigation: MS, HG, IC, AE, AM, SB, TR, MM. Methodology: MS, TR, MM. Software: MS. Supervision: MM. Validation: AM, SB, TR, MM. Visualization: MS. Writing, original draft: MS. Writing, review and editing: all authors. All authors read and approved the final manuscript.

# Acknowledgement

The authors thank the Department of Radiology, the Information Technology department, and the research core staff of Mayo Clinic Arizona for assistance with data access. A preliminary version of the sex-classification result was presented as a poster at the Combined Otolaryngology Spring Meetings, Phoenix, Arizona, April 2026. A companion manuscript reporting sex- and age-specific reference values from the same cohort, without sex classification or allometric analysis, is under review elsewhere.

# Data Availability

The CT examinations cannot be shared because they are identifiable medical images held under institutional policy. The de-identified per-individual dataset analysed here (recorded sex, age at scan, and the 17 morphometric variables), the analysis code, and the trained segmentation weights are held by the corresponding author and can be made available to qualified researchers on reasonable request, subject to Mayo Clinic institutional review and a data use agreement. Aggregate results for every variable are reported in the manuscript and Online Resource 1.

## References

1. Krishan K, Chatterjee PM, Kanchan T, Kaur S, Baryah N, Singh RK (2016) A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. Forensic Sci Int 261. https://doi.org/10.1016/j.forsciint.2016.02.007. :165.e1-165.e8
2. Spradley MK, Jantz RL (2011) Sex estimation in forensic anthropology: skull versus postcranial elements. J Forensic Sci 56:289–296. https://doi.org/10.1111/j.1556-4029.2010.01635.x

3. Dedouit F, Ducloyer M, Elifritz J, Adolphi NL, Yi-Li GW, Decker S et al (2025) The current state of forensic imaging - post mortem imaging. Int J Legal Med 139:1141–1159. https://doi.org/10.1007/s00414-025-03461-x
4. Imaizumi K, Usui S, Nagata T, Hayakawa H et al (2025) Sex-estimation method for three-dimensional shapes of the skull and skull parts using machine learning. Forensic Sci Int 373:112532. https://doi.org/10.1016/j.forsciint.2025.112532
5. Seo T, Yoon Y, Kim Y, Usumoto Y et al (2024) Sex estimation using skull silhouette images from postmortem computed tomography by deep learning. Sci Rep 14:22689. https://doi.org/10.1038/s41598-024-74703-y
6. Bertsatos A, Chovalopoulou ME, Brůžek J, Bejdová Š (2020) Advanced procedures for skull sex estimation using sexually dimorphic morphometric features. Int J Legal Med 134:1927–1937. https://doi.org/10.1007/s00414-020-02334-9
7. Heinrich A, Hubig M, Teichgräber U, Mall G (2025) Automated identification of unknown decedents: matching postmortem CT images with clinical databases. Int J Legal Med 139:2251–2262. https://doi.org/10.1007/s00414-025-03528-9
8. de Mendonça DS, Ribeiro EC, de Barros Silva PG, Rodrigues AA et al (2022) Diagnostic accuracy of paranasal sinus measurements on multislice computed tomography for sex estimation: a systematic review and meta-analysis. J Forensic Sci 67:2151–2164. https://doi.org/10.1111/1556-4029.15117
9. Wanzeler AMV, Alves-Júnior SM, Ayres L, da Costa Prestes MC, Gomes JT, Tuji FM (2019) Sex estimation using paranasal sinus discriminant analysis: a new approach via cone beam computerized tomography volume analysis. Int J Legal Med 133:1977–1984. https://doi.org/10.1007/s00414-019-02100-6
10. Fakher HM, Samir M, Farag AA, Elela NAMA, Abdelshafy SZ, Bayomy HE et al (2026) Machine learning-driven forensic sex prediction using CT-based nasal and maxillary sinus metrics. Int J Legal Med. https://doi.org/10.1007/s00414-026-03951-6
11. de Mendonça DS, Gurgel ML, Ribeiro EC, Rodrigues JVO, de Aguiar ASW, Cevidanes LHS et al (2026) Development of an AI-based model for sex estimation using CT-derived metrics from paranasal sinuses. Int J Legal Med 140:2383–2394. https://doi.org/10.1007/s00414-026-03789-y
12. Daultrey C, Hardman J, Anari S (2018) The Caucasian nasal septum: an in vivo computed tomography study. Aesthet Surg J 38:717–722. https://doi.org/10.1093/asj/sjx249
13. Yang X, Li D, Xue H, Du S, Yang X, Wang J et al (2019) Anatomical characteristics of the perpendicular plate of the ethmoid: an analysis of paranasal sinus computed tomography via three-dimensional reconstruction. J Craniofac Surg 30:604–606. https://doi.org/10.1097/SCS.0000000000005149
14. Maréchal L, Dumoncel J, Santos F, Astudillo Encina W, Evteev A, Prevost A et al (2023) New insights into the variability of upper airway morphology in modern humans. J Anat 242:781–795. https://doi.org/10.1111/joa.13813

15. Garvin HM, Sholts SB, Mosca LA (2014) Sexual dimorphism in human cranial trait scores: effects of population, age, and body size. Am J Phys Anthropol 154:259–269. https://doi.org/10.1002/ajpa.22502
16. Franklin D, Cardini A, Flavel A, Kuliukas A (2013) Estimation of sex from cranial measurements in a Western Australian population. Forensic Sci Int 229. https://doi.org/10.1016/j.forsciint.2013.03.005. :158.e1-158.e8
17. Morandini P, Biehler-Gomez L, Stull K, Cattaneo C (2026) Metric analysis of the postcranial skeleton: a comprehensive approach for biological sex estimation in an Italian population. Int J Legal Med 140:441–461. https://doi.org/10.1007/s00414-025-03599-8
18. Del Bove A, Menéndez L, Manzi G, Moggi-Cecchi J, Lorenzo C, Profico A (2023) Mapping sexual dimorphism signal in the human cranium. Sci Rep 13:16847. https://doi.org/10.1038/s41598-023-43007-y
19. Klingenberg CP (2016) Size, shape, and form: concepts of allometry in geometric morphometrics. Dev Genes Evol 226:113–137. https://doi.org/10.1007/s00427-016-0539-2
20. Hamanová Čechová M, Suchá B, Dupej J, Brůžek J, Villa C, Beňuš R et al (2026) Sex classification using exocranial surfaces in a multi-population sample. Int J Legal Med 140:1033–1046. https://doi.org/10.1007/s00414-025-03694-w
21. Sarnat BG, Wexler MR (1968) Postnatal growth of the nose and face after resection of septal cartilage in the rabbit. Oral Surg Oral Med Oral Pathol 26:712–727. https://doi.org/10.1016/0030-4220(68)90444-1
22. Hall BK, Precious DS (2013) Cleft lip, nose, and palate: the nasal septum as the pacemaker for midfacial growth. Oral Surg Oral Med Oral Pathol Oral Radiol 115:442–447. https://doi.org/10.1016/j.oooo.2012.05.005
23. Goergen MJ, Holton NE, Grünheid T (2017) Morphological interaction between the nasal septum and nasofacial skeleton during human ontogeny. J Anat 230:689–700. https://doi.org/10.1111/joa.12596
24. Rosas A, Bastir M (2002) Thin-plate spline analysis of allometry and sexual dimorphism in the human craniofacial complex. Am J Phys Anthropol 117:236–245. https://doi.org/10.1002/ajpa.10023
25. Bulygina E, Mitteroecker P, Aiello L (2006) Ontogeny of facial dimorphism and patterns of individual development within one human population. Am J Phys Anthropol 131:432–443. https://doi.org/10.1002/ajpa.20317
26. Holton NE, Yokley TR, Froehle AW, Southard TE (2014) Ontogenetic scaling of the human nose in a longitudinal sample: implications for genus Homo facial evolution. Am J Phys Anthropol 153:52–60. https://doi.org/10.1002/ajpa.22402
27. Butovskaya ML, Rostovtseva VV, Dronova DA, Mezentseva AA, Semenova OV, Komarova AV et al (2026) Sexual dimorphism in facial and body shape and their relationship in modern humans of

African, Asian, and European descent. Am J Hum Biol 38:e70282. https://doi.org/10.1002/ajhb.70282

28. Bastir M, Rosas A (2013) Cranial airways and the integration between the inner and outer facial skeleton in humans. Am J Phys Anthropol 152:287–293. https://doi.org/10.1002/ajpa.22359
29. Holton NE, Alsamawi A, Yokley TR, Froehle AW (2016) The ontogeny of nasal shape: an analysis of sexual dimorphism in a longitudinal sample. Am J Phys Anthropol 160:52–61. https://doi.org/10.1002/ajpa.22941
30. Bastir M, Megía I, Torres-Tamayo N, García-Martínez D, Piqueras FJ, Burgos-Flores D et al (2020) Three-dimensional analysis of sexual dimorphism in the soft tissue morphology of the upper airways in a human population. Am J Phys Anthropol 171:65–75. https://doi.org/10.1002/ajpa.23944
31. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B et al (2024) TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 385:e078378. https://doi.org/10.1136/bmj-2023-078378
32. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18:203–211. https://doi.org/10.1038/s41592-020-01008-z
33. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S et al (2012) 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging 30:1323–1341. https://doi.org/10.1016/j.mri.2012.05.001
34. Taha AA, Hanbury A (2015) Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 15:29. https://doi.org/10.1186/s12880-015-0068-x
35. Mosimann JE (1970) Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. J Am Stat Assoc 65:930–945. https://doi.org/10.1080/01621459.1970.10481136
36. Jungers WL, Falsetti AB, Wall CE (1995) Shape, relative size, and size-adjustments in morphometrics. Yearb Phys Anthropol 38:137–161. https://doi.org/10.1002/ajpa.1330380608
37. Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324
38. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
39. Strobl C, Boulesteix AL, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 8:25. https://doi.org/10.1186/1471-2105-8-25
40. Walker PL (2008) Sexing skulls using discriminant function analysis of visually assessed traits. Am J Phys Anthropol 136:39–50. https://doi.org/10.1002/ajpa.20776
41. Wang X, Liu G, Wu Q, Zheng Y, Song F, Li Y (2024) Sex estimation techniques based on skulls in forensic anthropology: A scoping review. PLoS ONE 19:e0311762. https://doi.org/10.1371/journal.pone.0311762

42. Hoshioka Y, Torimitsu S, Makino Y, Yajima D, Chiba F, Yamaguchi R et al (2025) Sex estimation from skull measurements of a contemporary Japanese population using three-dimensional computed tomography images. Int J Legal Med 139:383–391. https://doi.org/10.1007/s00414-024-03319-8
43. Toneva D, Nikolova S, Agre G, Zlatareva D, Hadjidekov V, Lazarov N (2021) Machine learning approaches for sex estimation using cranial measurements. Int J Legal Med 135:951–966. https://doi.org/10.1007/s00414-020-02460-4
44. Gillet C, Costa-Mendes L, Rérolle C, Telmon N, Maret D, Savall F (2020) Sex estimation in the cranium and mandible: a multislice computed tomography (MSCT) study using anthropometric and geometric morphometry methods. Int J Legal Med 134:823–832. https://doi.org/10.1007/s00414-019-02203-0
45. Lye R, Min H, Dowling J, Obertová Z, Estai M, Bachtiar NA et al (2024) Deep learning versus human assessors: forensic sex estimation from three-dimensional computed tomography scans. Sci Rep 14:30136. https://doi.org/10.1038/s41598-024-81718-y
46. Russel SM, Frank-Ito DO (2023) Gender differences in nasal anatomy and function among Caucasians. Facial Plast Surg Aesthet Med 25:145–152. https://doi.org/10.1089/fpsam.2022.0049
47. Tanikawa C, Oka A, Shiraishi Y, Yamashiro T (2023) Sex differences in airway volume and 3-dimensional shape in Japanese adults. Sci Rep 13:13937. https://doi.org/10.1038/s41598-023-41263-6
48. Buyukertan M, Keklikoglu N, Kokten G (2003) A morphometric consideration of nasal septal deviations by people with paranasal complaints; a computed tomography study. Rhinology 41:21–24

## Figures

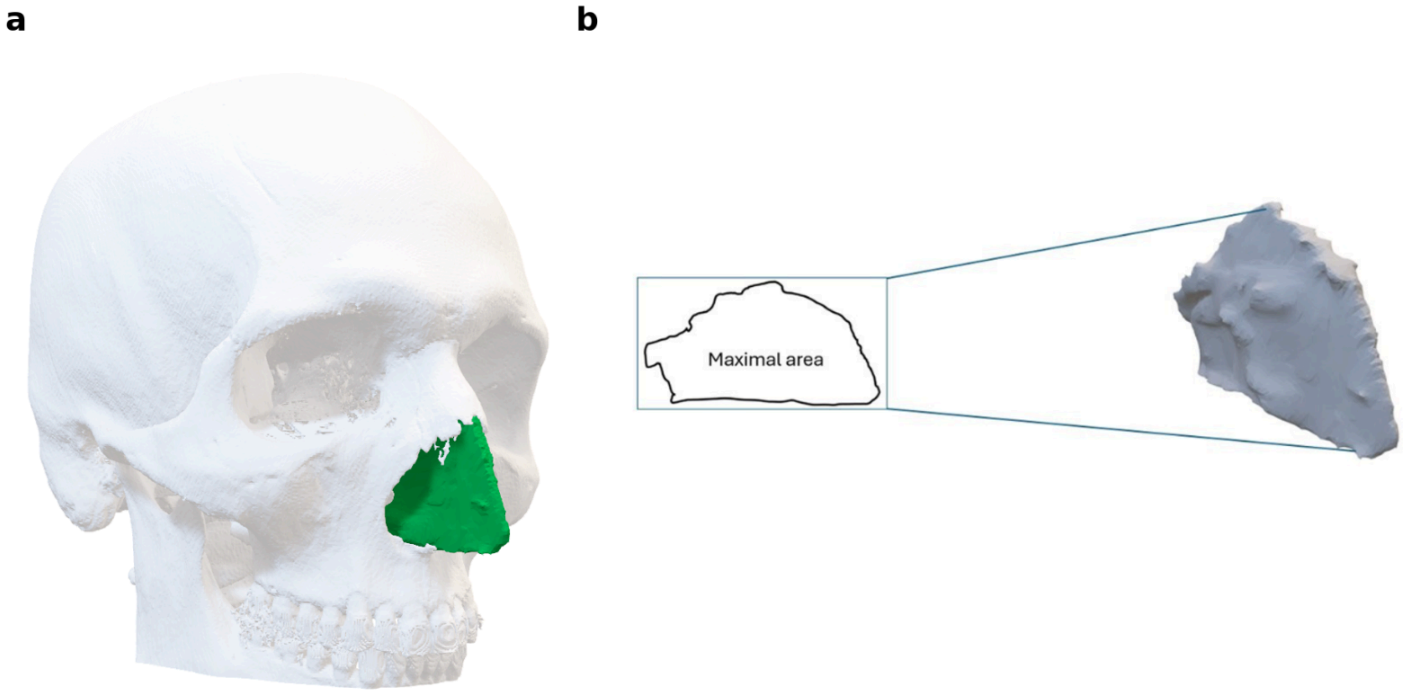


**Figure 1**

The segmented nasal septum. **a** Automated three-dimensional segmentation of the nasal septum (green) rendered against the surrounding skull of the same computed tomography examination, viewed from the left anterior oblique. The segmentation spans the quadrangular cartilage anteriorly and the vomer and perpendicular plate of the ethmoid posteriorly, from the nasal floor to the skull base. **b** Surface rendering of one automated segmentation (right) and the outline of the parasagittal slice of maximal cross-sectional area (left), the plane on which maximum sagittal cross-sectional area is measured

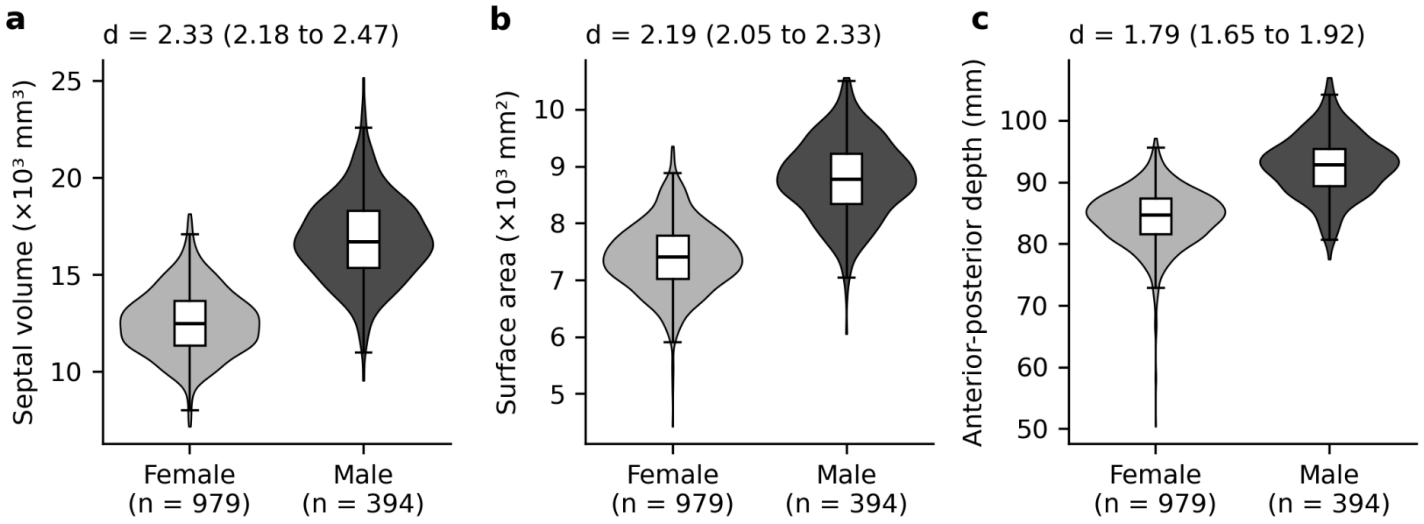



**Figure 2**

Sex differences in the three most dimorphic septal variables. Violin plots with inset box plots (median, interquartile range, and whiskers to 1.5 times the interquartile range) for women (light, n = 979) and men (dark, n = 394): **a** septal volume, **b** surface area, **c**anterior-posterior depth. Cohen's d with its 95% confidence interval is given above each panel. All three differences have Welch P < 0.001

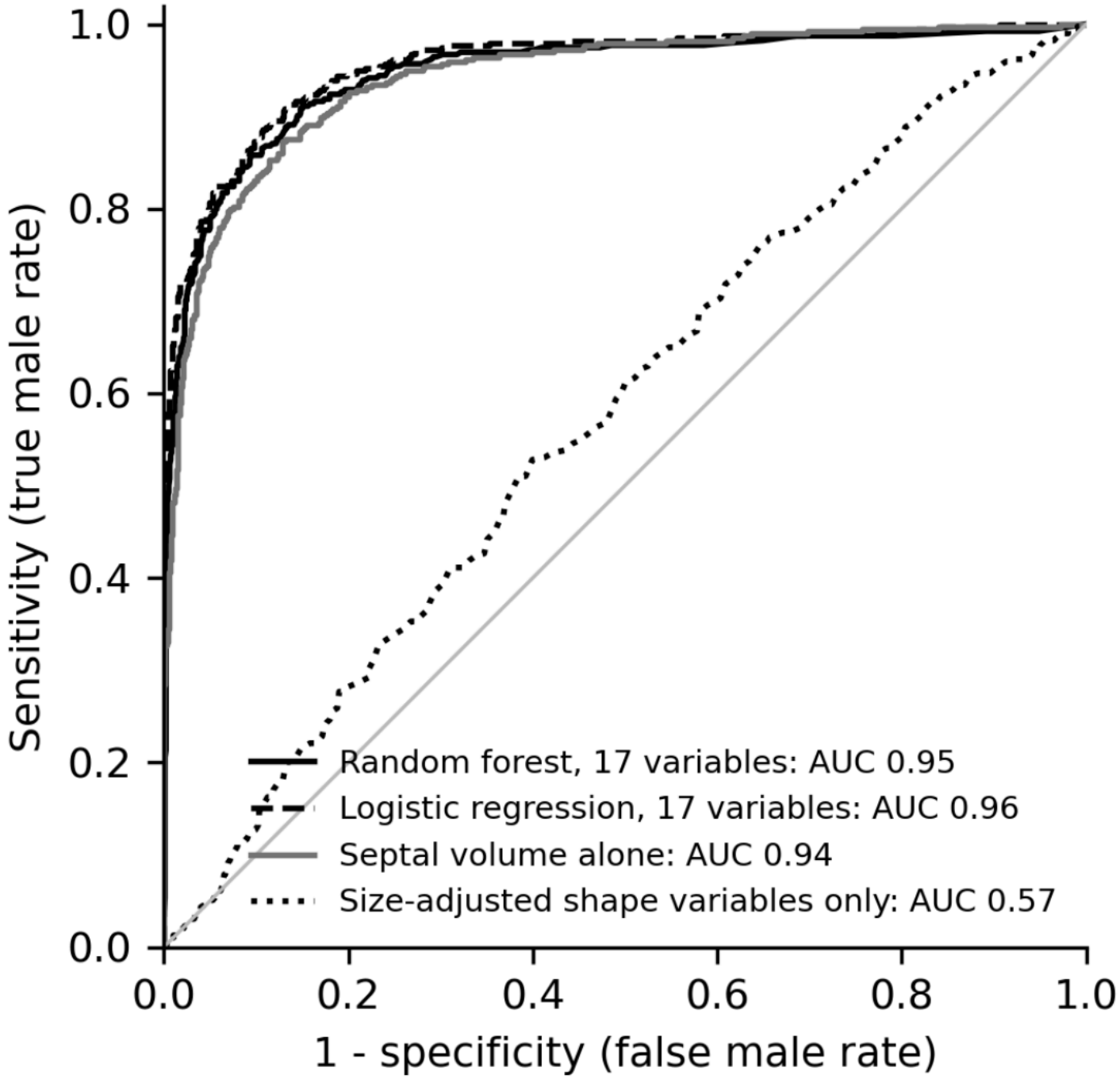



**Figure 3**

Receiver operating characteristic curves for sex classification from septal morphometry, from pooled out-of-fold predictions under stratified five-fold cross-validation (n = 1,373). Solid black, random forest on 17 morphometric variables (AUC 0.95); dashed black, logistic regression on the same variables (AUC 0.96); solid gray, logistic regression on septal volume alone (AUC 0.94); dotted, random forest on the 16 size-adjusted shape variables alone (AUC 0.57). The thin diagonal marks chance performance. AUC area under the curve

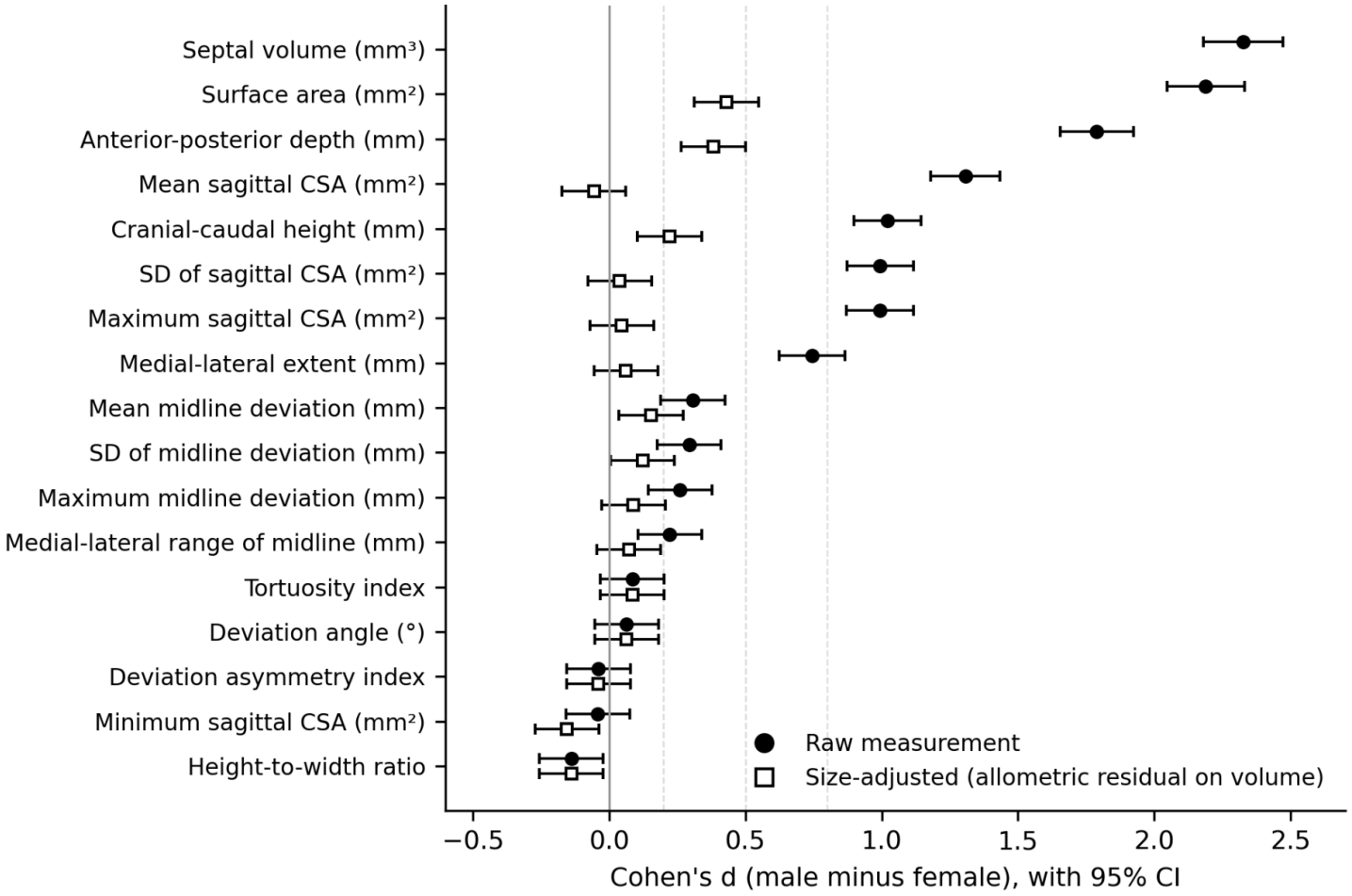



**Figure 4**

Sexual dimorphism before and after size adjustment. For each of the 17 variables, Cohen's d (men minus women, with 95% confidence interval) is shown for the raw measurement (filled circles) and, where applicable, for the size-adjusted variable (open squares), defined as the residual of the log-transformed variable regressed on log septal volume (dimensionless variables are shown once). Vertical dashed lines mark conventional small, medium, and large effects (0.2, 0.5, 0.8). Volume has no size-adjusted counterpart. CSA cross-sectional area, SD standard deviation