Evolution and Human Behavior 46 (2025) 106781

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.elsevier.com/locate/ens

Updating evidence on facial metrics: A Bayesian perspective on sexual dimorphism in facial width-to-height ratio and bizygomatic width

Alex L. Jones [a,*], Tobias L. Kordsmeyer [b], Robin S.S. Kramer [c], Julia Stern [d], Lars Penke [b]

[a] School of Psychology, Swansea University, Swansea SA2 8PP, United Kingdom
[b] Georg-Elias-Mueller Institute of Psychology & Leibniz ScienceCampus Primate Cognition, University of Göttingen, Goßlerstraße 14, Göttingen 37073, Germany
[c] School of Psychology, University of Lincoln, Lincoln LN6 7TS, UK
[d] Department of Psychology, University of Bremen, Grazer Straße 2c, Bremen 28359, Germany

ARTICLE INFO

Keywords:
Facial-width-to-height
Sexual dimorphism
Bayesian inference
Statistical modelling

ABSTRACT

Facial width-to-height ratio (fWHR) is an extensively studied morphological measure, which was presumably shaped by sexual selection and has been linked to a wide range of perceptual and physiological traits. Underpinning these associations is the premise that fWHR is larger in men, which empirically exhibits a mixed and equivocal pattern in the literature due to variation in measurement, large sample sizes revealing small but significant differences, and a lack of control of body size. In Study 1, in a sample of 1949 faces, we used a Bayesian hierarchical model that incorporates prior information to simultaneously estimate sexual dimorphism in fWHR, adjusted for body size, across five measurement types. While we found larger fWHR in women, comparing this effect to variability in fWHR due to image capture settings revealed no robust evidence of sex differences in fWHR. In Study 2, we investigated sex differences in facial width specifically (also adjusted for body size), again incorporating prior information, and confirmed men have greater face width than women. Advances in this area can be made by shifting focus away from arbitrary ratios like fWHR to direct measures like facial width – as well as carefully considering prior evidence of existing associations.

1. Introduction

The human face conveys a wealth of biological and socially-relevant information, including emotions (Engell et al., 2010; Vuilleumier et al., 2003), health status (Fisher et al., 2014; Jones et al., 2016; Russell et al., 2016), trustworthiness (Bonnefon et al., 2017), attractiveness (Jones & Jaeger, 2019; Kordsmeyer et al., 2024) personality (Jones et al., 2012; Kramer & Ward, 2010), and threat potential (Tipples, 2007). These facial cues and signals carry social consequences, predicting hiring decisions (Zebrowitz & Montepare, 2008), election outcomes (Ballew & Todorov, 2007), and aspects of economic exchange (Jaeger et al., 2019).

Much research has focussed on the facial attributes that communicate this information, covering many aspects of facial shape, colour, and texture (Fisher et al., 2014; Jones, 2018; Mayer et al., 2017). One parameter has received significant amounts of attention - facial width-to-height ratio (fWHR), measured as the distance between the left and right zygions (facial width), divided by a measure of face length. The latter is typically realised as the distance between the midpoint of the brows or the highest point of the eyelids (a proxy for the nasion) and the middle of the lips or highest point of the upper lip (a proxy for the prosthion). Men with larger fWHR (i.e., a relatively wider face) are perceived as more masculine (Mileva et al., 2014) and aggressive (Geniole et al., 2014). Importantly, initial evidence suggested that men with larger fWHR actually possess those qualities – for example, they do engage in more aggressive behaviour across athletic contexts, laboratory tasks, and self-report domains (Carré & McCormick, 2008; Lefevre et al., 2014; Lefevre & Lewis, 2014), exhibit more aggression in committed relationships (Wen & Zheng, 2020), and are more likely to die from interpersonal violence (Stirrat et al., 2012). A key biological assumption that underpins these findings is that fWHR is sexually selected (Hodges-Simeon et al., 2021; Weston et al., 2007), evolved by both intersexual selection (i.e., female mate choice, though evidence is mixed – Geniole et al., 2015; Hodges-Simeon et al., 2021; Valentine et al., 2014) and intrasexual competition. Evidence for the latter is relatively consistent, especially concerning perceptions of aggressiveness – men with relatively wider faces use this characteristic to signal social dominance (Alrajih & Ward, 2014), threat potential (Geniole et al., 2015; Zilioli et al., 2015; but see Dixson et al., 2017), or their position in a hierarchy

* Corresponding author at: School of Psychology, Swansea University, Swansea SA2 8PP, United Kingdom.
E-mail address: alex.l.jones@swansea.ac.uk (A.L. Jones).

https://doi.org/10.1016/j.evolhumbehav.2025.106781
Received 5 February 2025; Received in revised form 11 September 2025; Accepted 30 September 2025
Available online 6 October 2025
1090-5138/

(Kordsmeyer, Freund, et al., 2019) to others.

1.1. Sexual dimorphism in fWHR

Despite early evidence that fWHR is sexually dimorphic in the predicted direction in both facial photographs and skulls (Carré & McCormick, 2008; Haselhuhn & Wong, 2011; Weston et al., 2007), current evidence is mixed. Meta-analyses suggest a small but reliable sex difference when measured from images and skulls (Geniole et al., 2015), as well as from skulls alone, though with varying results for different ethnic populations (Kramer, 2017). A multitude of individual studies, across diverse samples, show no significant differences between sexes across 2D and 3D images (Kramer et al., 2012), and in Caucasian, African or Turkish population images (Lefevre et al., 2012; Özener, 2012). Other studies find supporting evidence for sexual dimorphism when fWHR is measured only in specific ways, such as defining face length as being from the nasion to the chin (Hodges-Simeon et al., 2021). Moreover, some studies suggest that when differences do occur, fWHR is larger in women's faces (Kramer et al., 2012; Lefevre et al., 2012; Rostovtseva et al., 2021; Summersby et al., 2022), casting further doubt on the hypothesis that fWHR is an evolutionarily relevant cue, particularly that it is an intrasexual cue to formidability and threat potential in men.

1.2. Body size and fWHR

One possible explanation for the differences in sexual dimorphism is the role of body size, often termed allometry. A key assumption underpinning the signalling utility of fWHR is that it is unrelated to body size (Carré et al., 2009; Carré & McCormick, 2008; Weston et al., 2007), a claim that has rarely been explicitly tested (Caton & Dixson, 2022). However, there is evidence that fWHR is related to measures of body size, such as height and weight (Burton & Rule, 2013; Deaner et al., 2012; Třebický et al., 2015), as well as muscularity (MacDonell et al., 2018). Further, there is consistent evidence that body mass index (BMI) correlates with fWHR (Coetzee et al., 2010; Eisenbruch et al., 2018; Kramer et al., 2012; Lefevre et al., 2012; Mayew, 2013; Noser et al., 2018; Saribay & Kleisner, 2018), such that individuals with greater BMI have a greater fWHR, and controlling for this variable can remove sex differences in fWHR (Lefevre et al., 2012). However, BMI is not an ideal measure of body size as it removes information on height, instead representing body mass scaled for height.

If fWHR shows positive relationships with body size, however it is measured, then this could lead to a spurious sex difference in fWHR when not controlling for size – sex differences may simply be due to men being larger than women, on average (Puts, 2010). Equally, in instances where there is no observed difference in fWHR, information about body size may mask this difference, given that the ratio is composed of two measures, which may be influenced separately by body size in complex ways (Caton & Dixson, 2022; Lefevre et al., 2012). A lack of consideration of body size generally may also explain the observed behavioural associations seen with fWHR, such as increased aggression – larger males may be more aggressive (Deaner et al., 2012; Sell, Tooby, & Cosmides, 2009), but this may be unrelated to fWHR.

1.3. Variation in fWHR measurement

An additional source of variation in the literature concerns the definition and measurement of fWHR. While facial width is consistently defined as bizygomatic width, facial height is defined in a variety of ways, with each showing varying patterns of sexual dimorphism and associations with behaviour and perceptions (discussed in great detail in Hodges-Simeon et al., 2021). For example, many researchers define facial height as the vertical distance from the upper lip (labiale superius) to the midpoint of the brows, termed fWHR-Brow, where the statistical significance of the sex difference is mixed (Haselhuhn & Wong, 2011; Krenn & Buehler, 2019; Lefevre et al., 2012). A subtle variation to this is to use the same bottom point (upper lip), but to use the highest point of the eyelids as the top, termed fWHR-Eyelid (Kramer et al., 2012; Stirrat & Perrett, 2010). Others define it as the distance between the nasion (fWHR-Nasion), the midline bony depression of the bridge of the nose between the eyes, to the upper lip, finding mixed evidence for a sex difference, or evidence in either direction (Kordsmeyer, Freund, et al., 2019; Kramer, 2017; Özener, 2012; Rostovtseva et al., 2021; Summersby et al., 2022). Another variant is to use the stomion (fWHR-Stomion) – the middle of the lips – as the lower point, and the nasion as the upper point, which shows no significant difference between the sexes (Robertson et al., 2017). Others define height by taking the nasion as the top and the lower chin as the bottom, termed fWHR-Lower, where evidence for dimorphism is also mixed (Hodges-Simeon et al., 2016; Lefevre et al., 2012, 2013; Robertson et al., 2017). Finally, other researchers have defined additional metrics based on available calliper measurements, finding evidence of dimorphism in either direction depending on which is used (Caton & Dixson, 2022).

It is worth stating the practical differences between these definitions. For example, the height component of fWHR-Lower measures from the nasion to the lower chin, while fWHR-Stomion measures from the nasion to the stomion. The difference in height between these metrics in adult faces is on average around 41–45 mm (Young, 1993), which despite being a relatively small distance, may contribute in part to the varying outcomes of fWHR measures, or simply be a source of noise. Still, all of the various measures of fWHR exhibit strong correlations, with some estimates in the region of 0.87 to 0.92 (Hodges-Simeon et al., 2021), which further influences the uncertainty in outcomes. Each variation in measurement produces varying patterns of results, which is perhaps unusual for such tightly correlated variables.

Taken together, these multiple measures suggest that consensus in the definition of fWHR is lacking, and that publication of results in the literature may be biased towards those that produce statistical significance – or at least allow researchers multiple attempts to test their fWHR-related hypotheses. For example, some recommendations call for focussing research efforts on those measures that produce statistically significant outcomes between sexes (Hodges-Simeon et al., 2021), despite considerable issues with *p*-values as a basis for inference, and the likelihood of magnitude errors (that is, a Type-M error) in estimates that can occur (Gelman & Carlin, 2014; McShane et al., 2019).

A further compounding issue across these various measures is that, perhaps in response to the mixed evidence surrounding dimorphism and the relationship of other variables with fWHR, researchers have commendably responded by testing hypotheses in large sample sizes of facial photographs or biometric datasets, with observations in the thousands (e.g., Caton & Dixson, 2022, $n = 6068$; Hodges-Simeon et al., 2021, $n = 2449$; Summersby et al., 2022, $n = 17{,}067$). Despite the increased precision in estimation, the use of *p*-values as a method of inference in large samples can be problematic, as extremely small deviations from the null hypothesis can be highly significant, even if these differences are practically meaningless (Lin et al., 2013). This reflects a key issue with the use of null-hypothesis significance testing as it relates to fWHR research. Differences between sexes are stated in terms of presence or absence, without much consideration about the size of the effect (but see Caton & Dixson, 2022; Geniole et al., 2015), and that any deviation from a point-null is considered worthy evidence of dimorphism, no matter how small (e.g., the 'crud factor', Meehl, 1967; Orben & Lakens, 2020). Each new result with fWHR is also considered independent of the wider literature – outside of meta-analytic approaches (Geniole et al., 2015; Kramer, 2017), significant or non-significant results are stated with only a qualitative consideration of previous research, failing to consider how this evidence mounts in favour of one hypothesis or another.

1.4. A Bayesian perspective

Taken together, there are several issues with research examining sexual dimorphism in fWHR. The inconsistent pattern of sexual dimorphism, a key component of the evolutionary-based signal utility of fWHR, has been examined in many studies with mixed results. This equivocal pattern could be due to the lack of controlling for body size – or using BMI as a control – which may mask or exaggerate differences. The wide range of fWHR measures, which we contend differ very little in practical terms given their correlations (Young, 1993), contributes to this mixed evidence base. Finally, the use of large samples and NHST, as well as the disregarding of uncertainty in the effect in the wider literature, mean that even extremely small, practically insignificant differences can be taken as evidence for a difference. A recent suggestion has been to avoid the use of fWHR altogether and instead focus solely on facial width alone (the consistent numerator in fWHR calculations in the literature). This measure shows a clear, large sex difference which remains after appropriate allometric control, as suggested by the one study available to date (Caton & Dixson, 2022).

We suggest that the principled application of model-based Bayesian inference (Gelman et al., 2020) can address the above issues in the study of fWHR. First, because a Bayesian analysis cannot proceed without a prior distribution, it allows the wealth of existing evidence regarding sex differences in fWHR to be incorporated into a model at the outset. This avoids the usual approach of taking results from NHST analyses in isolation and merging them with the wider literature only qualitatively. Second, a Bayesian analysis delivers the probability of a hypothesis, given the data, in contrast to the probability of the data given the null hypothesis (Kruschke, 2018). This difference allows for greater theoretical understanding - a probability can be placed on the hypothesis that men have greater fWHR than women, in contrast to a rejection of a null hypothesis of no difference, which the *p*-value delivers. Similarly, the Bayesian approach allows for the straightforward acceptance of a null hypothesis - provided one is reasonably defined – which is difficult to do under the NHST framework (Kruschke, 2018). Finally, Bayesian models have the unique advantage of being generative (Haines et al., 2020). This means that the downstream consequences of the model can be directly checked as new data can be created by the model, compared against the observed data, and examined. For example, by holding body size constant, the model can simulate samples of men and women, and the differences in fWHR in this population estimated.

Here, in Study 1, with a large sample of 1949 photographs, we use a Bayesian hierarchical linear model to simultaneously estimate sex differences in fWHR across five different measures, adjusting for height and weight. Taking advantage of the flexibility of these models, we incorporate prior information about sexual dimorphism into our inferences to better represent the current literature. Finally, using posterior predictive simulation techniques (Chambert et al., 2014), we provide probabilistic evidence for the hypothesis of sexual dimorphism in fWHR, from both a point-null perspective as well as through the use of regions of practical equivalence (Kruschke, 2018) which we define from a measurement error perspective. In Study 2, we examine how facial width alone differs between men and women, leveraging the same probabilistic benefits of Bayesian inference, as well as extending previous research by testing dimorphism when controlling for upper body size as opposed to height and weight.

2. Study 1

2.1. Method

2.1.1. Samples of facial photographs

We obtained 1949 facial photographs (818 men, 1131 women) across nine different image sets, which were either open access databases, or collected by the authors for previously published projects. The projects and details of these databases are outlined in Table 1. Faces were drawn from a range of Western societies. Importantly, all images were taken under laboratory conditions, with neutral facial expressions, and facing the camera directly. While image conditions are consistent within a sample, they naturally differ between samples.

Table 1
Dataset summary information.

Sample	Study	Country of origin	Age (M, SD)	Men	Women	Total *N*
1	Scott et al., 2013	UK	21.45 (5.04)	94	130	224
2	Jones et al., 2015	UK	21.64 (5.17)	104	150	254
3	Saribay et al., 2018	Turkey	21.65 (1.89)	91	94	185
4	Jünger, Kordsmeyer, et al., 2018; Jünger, Motta-Mena, et al., 2018; Kordsmeyer et al., 2019	Germany	23.17 (3.39)	0	148	148
5	Stern et al., 2021	Germany	23.2 (3.3)	0	256	256
6	Asendorpf et al., 2011	Germany	32.8 (7.4)	189	165	354
7	Penke & Asendorpf, 2008	Germany	23.7 (2.7)	133	138	271
8	Morrison et al., 2017	Poland	24.12 (3.03)	50	50	100
9	Kordsmeyer & Penke, 2019	Germany	24.23 (3.23)	157	0	157
	Overall			818	1131	1949

For all images, participants' physical measured height (in centimetres) and weight data (in kilograms) were available (except for sample 5, for which only self-reported height and weight were available).

2.1.2. Image landmarking and processing

All faces were automatically landmarked with a set of 68 points using the Python *face recognition* module (https://github.com/ageitgey/face_recognition), built on the Dlib machine learning package (King, 2009). Despite a full set of landmarks being available, we used only those relevant for the derivation of fWHR. Importantly for the current study, the measurement of fWHR from automatic landmarks is tightly correlated with fWHR derived from manually placed landmarks, shows little bias in terms of ethnicity differences, and demonstrates little loss of information (Jones, Schild, & Jones, 2021). Nonetheless, we undertook a validation analysis on a subsample of 528 images (27 % of the total set) for which manually placed landmarks were available, which is described in Table S1 of the supplementary materials. We found no evidence of a differential pattern of fWHR calculation in men and women when using automatic landmarks, mirroring the findings of Jones, Jaeger, & Schild, 2021, Jones, Schild, and Jones (2021). The landmark configuration for the faces is shown in Fig. 1.

2.1.3. Measures of fWHR

We derived five measures of fWHR, common in the literature, from the available landmarks. Facial width was calculated as the Euclidean distance between the left and right zygions, and was divided by five measures of height:

2.1.3.1. fWHR-Brow. Height was defined as the distance between the midpoint of the top of the upper lip or vermilion border (i.e., the 'cupid's bow'), to the midpoint between the brows. The midpoint between the brows was calculated by averaging the landmarks representing the inner points of the brows (e.g., where they begin).

2.1.3.2. fWHR-Eyelid. Height was defined here as the distance between the midpoint of the top of the upper lip (as in fWHR-Brow), to the midpoint of the uppermost point of the eyes. The midpoint was calculated by averaging the two highest points across the two eyelids.

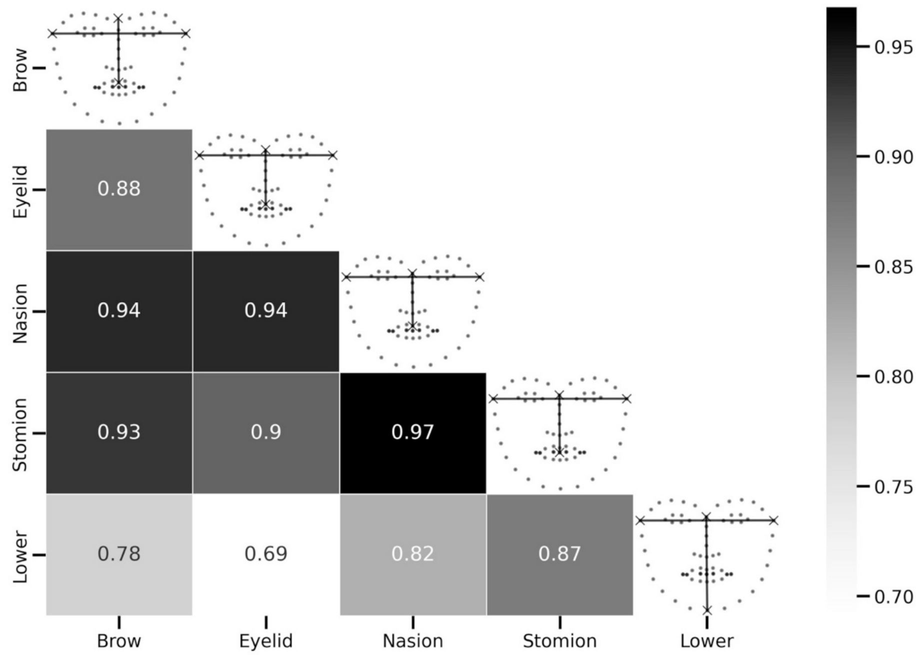

Fig. 1. Facial landmark configurations defining each fWHR measure, along with correlations amongst measures.

2.1.3.3. fWHR-Nasion. This height measure used the same lower point as the previous two measures (the midpoint of the top of the upper lip), but the upper point was defined as the nasion.

2.1.3.4. fWHR-Stomion. A slight alteration of fWHR-Nasion, measuring height from the nasion to the centre of the mouth where the lips meet (the stomion).

2.1.3.5. fWHR-Lower. Here, height was defined as the distance between the nasion and the lowest point of the chin.

These measures, as expected, exhibited strong correlations, ranging from 0.69 at the lowest (fWHR-Eyelid with fWHR-Lower) to almost perfect correlation at 0.97 (fWHR-Stomion with fWHR-Nasion). The full range is given in Fig. 1.

2.2. Analytic strategy – Bayesian model specification

We fitted a single hierarchical linear regression model to the data, the structure and justification of which is described below. The model was estimated using PyMC (Salvatier et al., 2016), and all code and data are available on the OSF (https://osf.io/uzqfh/).

2.2.1. Model structure

The model contained three common (i.e., fixed) effects – sex, with women as the reference category (zero), such that a positive coefficient for this variable indicates larger fWHR in men. Height and weight were the other common effects, which were *z*-score standardised across the entire dataset, so they can be interpreted as the change in fWHR units with a one standard deviation increase in height or weight, with all other predictors held constant. Consequently, the common intercept of the model represented the average fWHR for women's faces. However, we incorporated a careful group (i.e., random) effects structure in the model. Specifically, there was a group-specific intercept for image sample (accounting for variability in fWHR across image sets), a group-specific intercept for each individual image (accounting for within-individual variability in fWHR, across measures), and most importantly, a group-specific intercept for each fWHR measure, that allowed each measure to have its own intercept. In addition, each of the three predictors had group-specific slopes within each fWHR measure. A formal statement of the model is

$$fWHR_{i,j,k} = \beta 0_{i,j,k} + Sex^{*}\beta 1_i + Height^{*}\beta 2_i + Weight^{*}\beta 3_i$$

where i represents each of the five fWHR measures, j indicates which sample the observation is from, with k tracking which image the observation belongs to.

A critical feature of this model structure is that it induces partial pooling (Gelman et al., 2012; Gelman & Pardoe, 2006; Kruschke, 2014; McElreath, 2020). By treating the different measures of fWHR as a group-specific effect, the model can share information across them, making the individual estimates of each measure less noisy via shrinkage (Gelman & Pardoe, 2006). Stated another way, the model estimates a population-level regression model of sex, height, and weight on fWHR, as well as a sub-model for each of the five measures. The estimates for each sub-model are not the same as if each were being estimated in isolation, but are pulled towards the overarching model (Kruschke, 2014). From an inferential perspective, this represents the assumption that the range of fWHR measures are instances of measures of fWHR drawn from a population of candidate measures, and incorporating their variability will shrink their individual estimates towards the population effect. Crucially, this avoids the common problem of testing multiple measure types in isolation and drawing inferences based on the presence or absence of statistical significance, as well as acknowledging the similarity in these measures.

The model used normal prior distributions on all coefficients, informed by sexual dimorphism in fWHR in the literature, and a *t*-distributed likelihood that reduces the influence of any outliers. A full description of the prior evidence and structure can be found in the supplementary materials.

2.3. A region of practical equivalence – what counts as 'different?'

To address the issue of rejecting a point-null hypothesis which can easily occur in large samples, we considered carefully what constitutes a practically null value of a sex difference in fWHR. Estimates should be larger than this region to be deemed noteworthy. By setting this 'region of practical equivalence', or ROPE (Kruschke, 2018), we are able to assess the probability that the hypothesis of a sex difference is in or out of this region, or to what extent it overlaps (Makowski et al., 2019), a considerably more informative conclusion than NHST (Jones, Jaeger, & Schild, 2021). We set our ROPE by conducting a Bayesian re-analysis of Kramer (2016), who examined variation in fWHR solely due to camera focal length, finding that around 0.5 fWHR units can be attributable to variation in camera focal length. As such, differences smaller than this can be considered poor evidence of dimorphism given they could equally be due to unrelated noise (see supplementary materials for a fuller description).

3. Results

3.1. Model estimation

The model was estimated using Hamiltonian Monte Carlo methods (Gelman et al., 2020; McElreath, 2020), and 40,000 samples were drawn from the posterior after a set of 30,000 tuning steps. All parameters in the model converged, with all $\hat{R} \leq 1.01$ (Gelman et al., 2020).

3.2. Common effects

The common effects, indicating the population-level associations between sex, height, and weight, are discussed here. All results are visualised in Fig. 2.

3.2.1. Sex differences

The overall sex difference estimated by the model (adjusted for height and weight) was negative, with an average sex difference of $M = -0.037$, 94 % credible interval $[-0.051, -0.021]$. The posterior probability that this sex difference was negative under the model was 100 % – that is, the model excluded entirely a larger mean fWHR in men. We also conducted a Bayes Factor analysis (Wagenmakers et al., 2010) on this coefficient to describe the change in likelihood of the sex difference being equal to the mean difference set in the prior ($M = 0.008$) after seeing the data. This suggested the current data raised the odds of the alternative model, $BF_{10} = 136$. However, and most importantly, the posterior of the sex difference fell entirely within the ROPE, suggesting the clearly non-zero estimate was far smaller than variation attributable to simple camera focal length.

3.2.2. Height

The population level effect of height was also negative, $M = -0.042$ $[-0.049, -0.034]$, with the entire posterior mass below zero, indicating that an increase of 1 *SD* in height was associated with a decrease in fWHR.

3.2.3. Weight

Conversely, the population level effect of weight was positive, $M = 0.048$ $[0.042, 0.055]$, with the entire posterior mass above zero. *A*1 *SD* unit increase in weight was associated with an increase in fWHR.

3.3. Group-specific effects

The relationships between sex, height, and weight for each individual fWHR measure (estimated with partial pooling) are discussed here. Despite small variations in magnitude, the effects were all in the same direction as the population level effects, indicating a consistent relationship regardless of the chosen fWHR measure. All group-specific

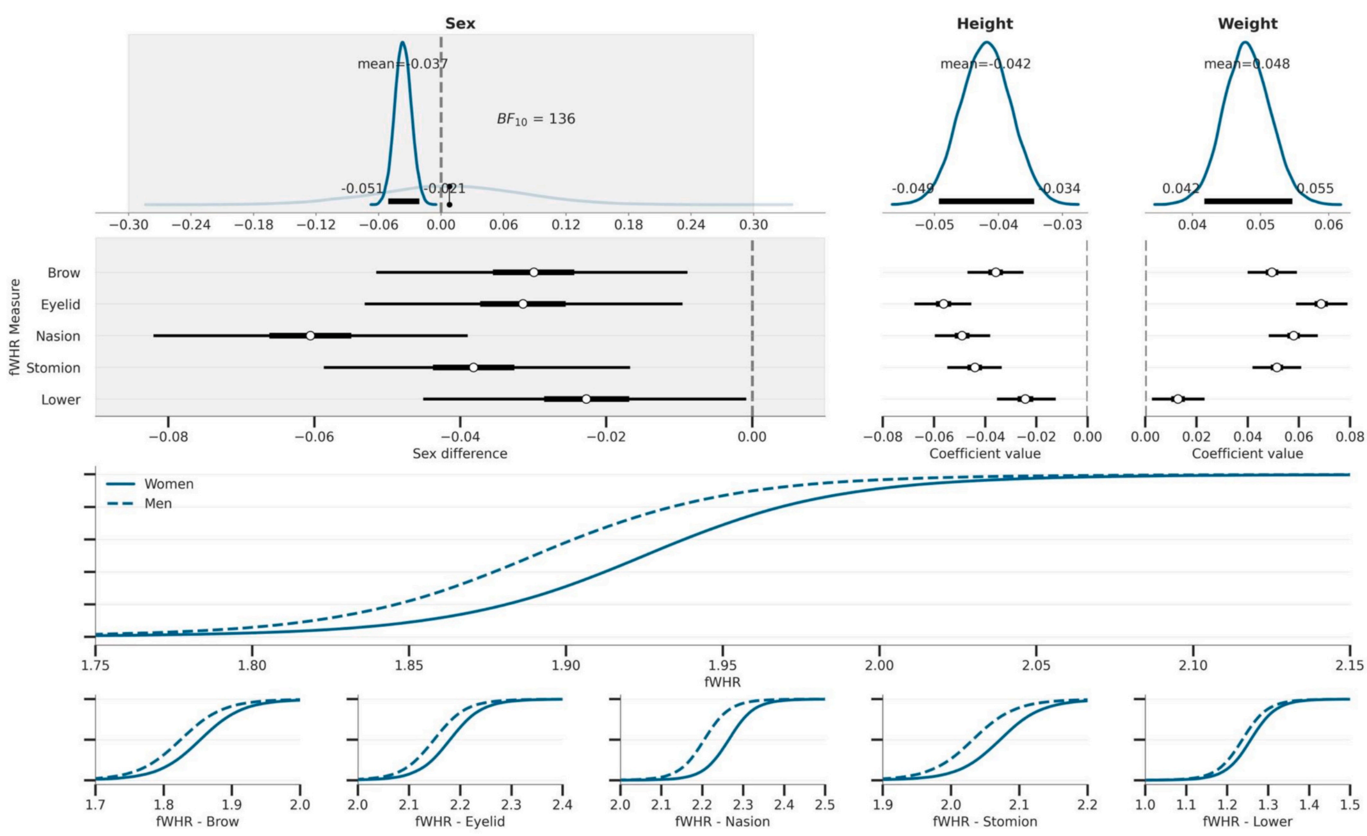

Fig. 2. First row – the prior and posterior distribution of the sex coefficient, with associated interval and Bayes Factor analysis, and the posterior distributions of height and weight. Note the distribution for sex falls entirely in the region of practical equivalence (ROPE). Second row – the posterior distributions of each coefficient under each fWHR measure, illustrating a consistent pattern. Bottom panels – posterior predictive distributions of fWHR for new samples of men and women equated on height and weight, for the population effect (large panel) and each individual measure (bottom row). Lines represent cumulative density functions, illustrating the probabilities of fWHR being equal to or less than a given percentage. Men show consistently smaller fWHR than women.

effects are shown in Fig. 3, and in all but one case (fWHR-Lower), the posterior masses were entirely below (sex, height) or above (weight) zero, implying a 100 % probability of the effect in that direction.

3.3.1. fWHR-Brow

This measure showed a negative effect of sex, $M = -0.030$ [−0.046, −0.014], and height, $M = -0.036$ [−0.044, −0.028], and a positive effect of weight, $M = 0.049$ [0.042, 0.056]

3.3.2. fWHR-Eyelid

This measure showed a negative effect of sex, $M = -0.031$ [−0.047, −0.015], and height, $M = -0.056$ [−0.064, −0.048], and a positive effect of weight, $M = 0.069$ [0.062, 0.076]

3.3.3. fWHR-Nasion

This measure showed a negative effect of sex, $M = -0.061$ [−0.076, −0.045], and height, $M = -0.049$ [−0.057, −0.041], and a positive effect of weight, $M = 0.058$ [0.051, 0.065]

3.3.4. fWHR-Stomion

This measure showed a negative effect of sex, $M = -0.038$ [−0.054, −0.023], and height, $M = -0.044$ [−0.052, −0.036], and a positive effect of weight, $M = 0.051$ [0.045, 0.058]

3.3.5. fWHR-Lower

This measure showed a negative effect of sex, $M = -0.023$ [−0.039, −0.006], and height, $M = -0.024$ [−0.033, −0.016], and a positive effect of weight, $M = 0.013$ [0.006, 0.02]. This measure showed a 0.005 % probability of a male-biased sex difference – that is, just 0.005 % of the posterior of the sex difference was above zero.

3.4. Posterior predictive simulations

One further benefit of Bayesian models is their generative nature,

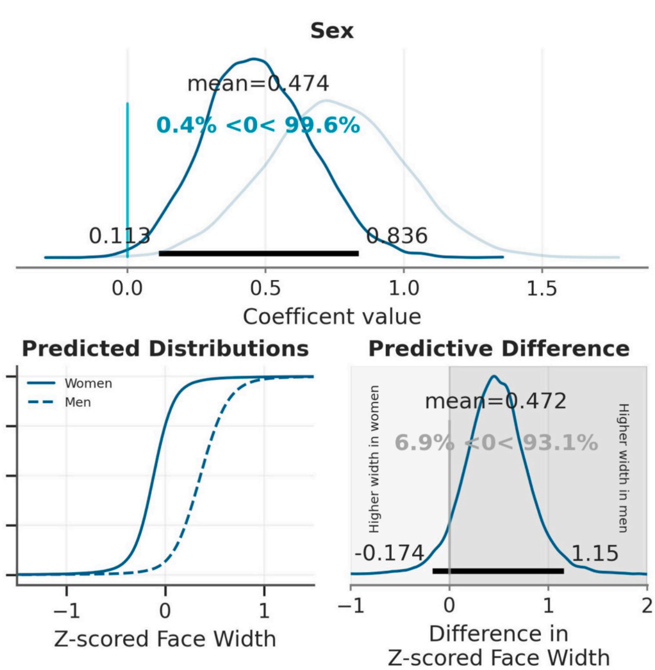

Fig. 3. Results from Analysis One. Top panel – the prior (lighter line) and posterior distributions of the sexual dimorphism in standardised face width. Bottom left – cumulative density functions of predicted distributions of women and men's face width, adjusted for height and weight. Bottom right – the difference between these predicted distributions, indicating the probability that men have higher face width than women. Shaded areas represent direction of effect.

which allows the probing of probabilistic questions that would otherwise be difficult or impossible to address with standard methods (Gelman et al., 2015). Briefly, the posterior distribution of estimated parameters is fed into the likelihood distribution, and a sample is drawn. By doing this across the whole posterior distribution, uncertainty is propagated from parameters (coefficients, variances, etc.) to observed data (Chambert et al., 2014). The model is thus able to generate new datasets based on both the observed data and the model. More simply, a posterior predictive distribution allows for the examination of uncertainty in new observations, rather than uncertainty in parameters (Martin et al., 2021).

Using this approach, we ask what is the probability that a given man has a larger fWHR than a woman, holding height and weight constant? This simple question is deceptively difficult to answer without Bayesian inference. Consider that raw fWHR scores are not adjusted for covariates, and a point estimate alone does not propagate uncertainty in parameters to new data. We answer this question for both the overall population level effects (i.e., integrating across the various measures), as well as for each measure alone, by passing the posterior distributions of the intercepts and slopes to a *t*-distribution, along with the model's error variance, generating a sample of 40,000 women and men. These distributions are illustrated in Fig. 2.

For the overall population effects, men had a 32 % probability of having a larger fWHR than women. Each individual measure showed similar proportions, all being considerably lower than a 50 % probability: fWHR-Brow – 34.73 %, fWHR-Eyelid – 34.17 % fWHR-Nasion – 21.99 %, fWHR-Stomion – 30.84 %, and fWHR-Lower – 38.28 %.

4. Discussion

In a large sample of faces, and across five different measures of fWHR, a hierarchical Bayesian model indicated a very small but consistent sex difference in fWHR in the opposite direction to theoretical predictions, being larger in women. This small effect was entirely within the bounds of differences expected solely from within-person variability resulting from camera focal length (Kramer, 2016).

The finding of sexual dimorphism being larger in women is in line with some estimates (Lefevre et al., 2012; Rostovtseva et al., 2021), and this was consistent across individual measures of fWHR. The use of model-based predictive simulations also suggested that women have roughly a two-thirds greater chance of having larger fWHR than men, a novel estimate of this sex difference. Conducting a Bayes Factor (Wagenmakers et al., 2010) analysis to the sex difference revealed that the current data raised the odds of the alternative hypothesis – that is, the sex difference does not equal that collected from the current literature by a factor of around 137.

However, we stress here that this difference is likely to be unimportant. The hypothesis of zero difference in fWHR between sexes is unrealistic, and small divergences from it are not strong confirmations of theory (Meehl, 1967). Instead, the magnitude of the effect is entirely within the boundaries of differences expected from within-person variation attributable to camera settings (Kramer, 2016). As such, we suggest there is no noteworthy dimorphism in fWHR.

In contrast, we observed consistent effects of height and weight on fWHR, such that taller individuals had lower fWHR and heavier individuals had greater fWHR. The magnitude of these effects was qualitatively similar to the dimorphism difference. For example, a 1 *SD* increase in height was associated with a similar amount of change in fWHR as the average difference between men and women, suggesting that there is nothing particularly salient about fWHR as a secondary sexual characteristic. While separating height and weight is a useful way to measure allometry, these results are broadly consistent with existing literature that uses BMI as a control variable (Hodges-Simeon et al., 2021; Kramer et al., 2012; Lefevre et al., 2012).

5. Study 2

We observed no noteworthy dimorphism in fWHR in Study 1. Here, in Study 2, we examine facial width in isolation as a possible sexually dimorphic facial parameter. In one of the largest studies to date, with calliper-based facial measures – and with comprehensive allometric control – Caton and Dixson (2022) showed a large and robust dimorphism ($d \approx 1$) of facial width (with larger facial width for men, on average), but mixed results of fWHR. Caton and Dixson (2022) also emphasise the statistical issues that arise with ratio measures like fWHR. Here, we undertake two analyses to examine whether facial width exhibits this notable dimorphism. In the first analysis, we again use Bayesian model-based inference to test whether dimorphism in facial width is present across the various samples in our current dataset, adjusting for height and weight. Importantly, we reanalyse the data of Caton and Dixson (2022) to derive a prior for the sex difference in the presence of height and weight, which is fed into our model. In the second analysis, we test whether dimorphism in facial width exists in a smaller subset of our data, controlling for a measure of upper body size, which more closely aligns with the approach of Caton and Dixson (2022).

6. Method

6.1. Analysis one

6.1.1. Samples and facial metrics

The same 1949 facial photographs and landmarks were used for analysis. As in Study 1, we only considered the pair of landmarks necessary to calculate face width. These data were available as part of the calculation of fWHR in Study 1. Participants' height and weight were available for all faces.

6.1.2. Analytic strategy – Bayesian model specification

The model used here was again a single hierarchical linear regression, which had a different formulation to the one used in Study 1, given there was only one measure of face width (as opposed to the multiple measures of face height before).

6.1.3. Model structure

The model again contained three common (i.e., fixed) effects – first, sex, with women as the reference category (zero). That is, a positive coefficient indicates a higher face width in men. As before, the model also contained *z*-scored height and weight as fixed effects. As facial width is measured in Euclidean distance, a unitless metric, we standardised this by *z*-scoring about its grand mean, meaning that a 1 *SD* increase in height or weight indicated a change in facial width *SD*, and the intercept and sex coefficients represented the mean *z*-score of women's faces, and the difference between men and women in *z*-score units, respectively.

We also incorporated a group-specific effect structure in the model. There was an intercept for image sample, accounting for variability in face width across image sets, and similarly to Study 1, each predictor had a group-specific slope within each image set, which again has the property of inducing partial pooling – a separate model is estimated for each dataset, but pulled towards an overarching model. While some groups in the dataset had no women or men in them, partial pooling is still useful as it allows the sex-specific facial width in those groups to flow to those that do have men and women's faces. Formally, the model was:

$$Z - Facial\ Width_{i,j} = \beta 0_j + Sex{*}\beta 1_j + Height{*}\beta 2_j + Weight{*}\beta 3_j$$

where *i* represents each individual face, and *j* indicates which sample the face is from. As before, we used a *t*-distribution as the likelihood, making inferences robust to extreme measures.

We derived a set of informed priors for this model by conducting a re-analysis of the ANSUR II survey (Paquette et al., 2009) used by Caton and Dixson (2022), who demonstrated dimorphism in facial width. This allowed us to derive priors for our data from a rich dataset with physical measurements (see supplementary materials for the details of this analysis).

Finally, we also undertook an additional supplemental analysis that considered the effect of face width together with the multiple face height measures derived from the set of fWHR metrics used in Study 1, to examine whether any dimorphism in face width remained after applying partial pooling towards the effects of face height (see supplementary materials).

6.2. Analysis two

The second analysis focused on examining sexual dimorphism in face width while controlling for upper body size, albeit on a smaller sample taken from the larger set. As a composite measure out of three components of upper body muscularity (shoulder width, chest girth, and mean biceps girth), the highly sexually dimorphic upper body size variable is more closely focused on physical aspects that may relate to aggression and strength (Price et al., 2012; von Borell et al., 2019), and so testing for sexual dimorphism in facial width controlling for this variable may be especially informative.

6.2.1. Samples, upper body, and facial metrics

Participants from samples 4 (Jünger, Kordsmeyer, et al., 2018; Jünger, Motta-Mena, et al., 2018; Kordsmeyer, Freund, et al., 2019) and 9 (Kordsmeyer & Penke, 2019) who had an upper body measure were used for this analysis, comprising a total of $n = 305$ individuals (all 148 women from sample 4, and all 157 men from sample 9). The measure of facial width was used from their facial landmarks. Participants in sample 4 (women) were scanned four times and in sample 9 (men) were scanned three times using the Vitus smart XXL bodyscanner running AnthroScan software (both Human Solutions GmbH, Kaiserslautern, Germany), while wearing standardised tight underwear and standing in a standardised position (for details see Kordsmeyer, Stern, & Penke, 2019). AnthroScan's automatic measures (all according to ISO 20685:2005, German Institute for Standardization, 2006) include the following parameters to measure upper body size (in accordance with Price et al., 2012, see also Kordsmeyer, Stern, & Penke, 2019; Stern et al., 2021): cross shoulder width (AnthroScan #3010, for sample 4's women; for sample 9's men, biacromial width was measured manually, for details, see Kordsmeyer, Stern, & Penke, 2019), bust-chest girth (#4510), and upper arm girth (left: #8520 & right: #8521). These three variables were aggregated, based on *z*-standardised averages of automatic measurements extracted from the three (men) or four (women) body scans, to form the variable upper body size. Reliabilities for the three/four body scans were high for all relevant measures (ICCs [two-way random, single measures] > 0.89, only for women's cross shoulder width the ICC was lower, 0.39).

6.2.2. Model and prior specification

The model here was conceptually simpler, containing just two common effects – sex, with women as the reference category, and *z*-scored upper body size, which was *z*-scored within samples. Facial width was *z*-scored again across the current sample of 305 faces. There were necessarily no group-specific effects in this model. As we had no explicit prior information for this model, we opted to use weakly informative normal priors (mean 0 and an *SD* of 10) for all coefficients, with a *t*-distributed likelihood as before.

7. Results

7.1. Model estimations

The same estimation techniques were used as in Study 1, for both analyses. All parameters in the models converged, with all $\hat{R} = 1$ (Gelman et al., 2020).

7.2. Analysis one – common effects

To orient the reader, analysis one utilised the full sample of 1949 faces. We limit our discussion on the common effects here, given the focus is on the overall dimorphism across samples and not within.

7.2.1. Sex differences

The overall sex difference estimated by the model (adjusted for height and weight) indicated a large and relatively wide posterior, $M = 0.474$ [0.113, 0.836], with a 99.6 % probability of being positive (e.g., men having higher face width than women). Given that face width was standardised, this suggests men have around a half a standard deviation higher face width than women.

7.2.2. Height

The population level effect of height was negative, $M = -0.023$ [−0.034, −0.012], with the entire posterior mass below zero, indicating that an increase of 1 *SD* increase in height was associated with decreasing facial width.

7.2.3. Weight

Conversely, the population level effect of weight was positive, $M = 0.059$ [0.048, 0.069], with the entire posterior mass above zero. *A*1 *SD* unit increase in weight was associated with an increase in face width.

7.2.4. Posterior predictive simulation

As in Study 1, we used a posterior predictive simulation to examine a hypothetical population of women and men's faces measured for face width, equated on height and weight, and assessed the probability that men had a higher face width of any magnitude than women. These distributions are shown in Fig. 3. While the mean difference between these distributions was the same as the estimated sex difference coefficient ($M = 0.472$), the uncertainty was wider, as uncertainty in parameters is propagated to new observations. However, the model indicated that height- and weight-adjusted men had a 93.1 % probability of a higher facial width than women, ranging from −0.174 to 1.15 units.

The supplemental analysis that examined dimorphism in face width in the presence of face height supported these results, confirming the pattern that men have wider faces than women.

7.3. Analysis two – common effects

This analysis used just the faces from samples 4 and 9, comprising women and men, respectively, for a sample size of 305 individuals. As such, the model was conceptually simpler and contained just two common effects, discussed here.

7.3.1. Sex differences

The overall sex difference (adjusted for upper body size) indicated a very large sex difference, $M = 1.896$ [1.814, 1.968], with the model completely excluding zero as a credible hypothesis. Men have around almost a two standard deviation greater face width than women.

7.3.2. Upper body size

The effect of upper body size was positive with a 99.9 % probability, $M = 0.046$ [0.019, 0.074]. *A*1 *SD* unit increase in upper body size was associated with a consistent but relatively small increase in standardised face width, which is very similar in magnitude to the effect of body weight seen in Analysis One.

7.3.3. Posterior predictive simulation

A posterior predictive simulation was again used to generate a population of men and women's faces but equated on upper body size. In this simulation, men had a 99.9 % probability of a higher facial width than women, ranging from 1.2 to 2.6 units.

8. Discussion

In two analyses, we find consistent and large dimorphisms in facial bizygomatic width. In the first analysis, using height and weight as controls as well as prior information from Caton and Dixson (2022), we find that men almost certainly have wider faces (a 97 % probability) than women, in the same large sample for which Study 1 suggested a negligible, female-biased difference in fWHR. This effect was also large, differing by around half a standard deviation. In a second analysis on a smaller subset of the sample, and controlling for upper body size – a variable that shares correlations with behavioural traits that fWHR has also been associated with, such as proneness to anger and aggressiveness (Price et al., 2012; von Borell et al., 2019) – we find that men clearly have much wider faces than women, with a very large effect size.

These findings lend support to those of Caton and Dixson (2022), who first described the consistent dimorphism in bizygomatic width. We extend their findings here using posterior predictive simulations. Beyond a simple average sex difference, the model-implied consequences illustrate that men almost always have a greater face width than women, which points to an evolved, reliable indicator of relevant traits such as formidability (Caton et al., 2022; Sell, Cosmides, et al., 2009). A further implication from both the analyses here is that much like fWHR, bizygomatic width is associated with height (negatively) and weight (positively), and, in our second analysis, associated with upper body size. These results suggest bizygomatic width is not independent of allometry.

9. General discussion

Across two studies and through the lens of Bayesian inference, we examined sexual dimorphism in fWHR and bizygomatic width. In Study 1, by incorporating prior information, as well as controlling for height and weight, we found a small but consistent sexual dimorphism in favour of women (i.e., that women demonstrated larger fWHR). The use of partial pooling allowed us to consider the correlations amongst various measures of fWHR and to not treat them as separate, distinct measures. While this dimorphism was stronger in some measures (e.g., fWHR-Lower, as suggested by Hodges-Simeon et al., 2021), we considered the differences against the backdrop of variability in fWHR that can be attributed to simple camera focal length (Kramer, 2016). Deriving a null region of practical equivalence from this variability, we found that sexual dimorphism fell comfortably within the defined bounds (Kruschke, 2018). Taken together, these results provide robust evidence against the predictions of fWHR being greater in men due to it being a sexually-selected signal of threat and dominance (Geniole et al., 2015).

A second key finding from Study 1 was that fWHR was not independent of body size, being negatively related to height and positively to weight. This suggests that studies linking fWHR to perceptions or behaviour may be confounded. For example, the link between fWHR and aggression may be a byproduct of the association between aggression and weight (Archer & Thanzami, 2007, 2009; Deaner et al., 2012). Controlling for weight indicated height was negatively associated with fWHR, which runs counter to the prediction that fWHR is an evolved signal of threat potential. Previous work that links perceptions of masculinity and dominance to larger fWHR (Merlhiot et al., 2021) may also be confounded by weight, given that body weight is detectable in the face (Coetzee et al., 2010; Mayer et al., 2017; Wolffhechel et al., 2015), and also influences judgments of masculinity and dominance (Holzleitner et al., 2014). Indeed, while Holzleitner et al. (2014) found taller men were rated as having more masculine looking faces, we found that taller men showed smaller, not larger, fWHR, which is again inconsistent with the theoretical underpinnings of fWHR and dimorphism.

Taken together, our findings on fWHR suggest that however it is measured, any sexual dimorphism in this metric is inconsistent with theory and notably is far smaller than variation that might be attributable to variation in camera focal length. This does not preclude differences in fWHR emerging in situations of physical measurement where focal length is not a concern, but these kinds of differences have not often been observed in published literature, and where possible, were incorporated into the prior distribution for the model (Caton & Dixson, 2022; Kramer et al., 2012). Recent demonstrations have also shown that fWHR, when considered amongst other variables, has close to zero influence on social perceptions (Jaeger & Jones, 2022), and other studies have questioned its links with behaviour (Özener, 2012; Zhang et al., 2018) or physiology (Kordsmeyer, Freund, et al., 2019).

In contrast, Study 2, focusing solely on facial bizygomatic width, extended the findings of Caton and Dixson (2022) in a large sample of photographs. Men exhibited a clearly higher facial width than women, on average, when adjusting for body height and weight, and posterior simulations suggested that an individual man has a high probability of having a greater face width than a woman. Interestingly, facial width exhibited the same associations with body height and weight as did fWHR – taller individuals had narrower faces, and heavier individuals had wider faces. Caton and Dixson (2022) measured allometry via a principal component analysis of 93 body measurements (as opposed to height and weight, which we used to derive our prior distributions for our model) and found that facial width was correlated with components mapping onto traits such as bicep, forearm, neck, and chest circumference, concluding that faces evolved as a cue to formidability. Our own analysis with height and weight, informed by the data used for this initial investigation, suggests that at the very least, dimorphism in facial width remains after controlling for these associated measures. As a further step, in a smaller subsample of our data, we used a measure of upper body size as a control variable to assess sexual dimorphism in bizygomatic width, aligning more closely with the study of Caton and Dixson (2022). We found a notably larger dimorphism, and our model implied that almost any given man would have a greater face width than a woman. We also found that facial width shared a similarly sized association with upper body size as it does with weight – a small but positive relationship.

Conducting a sensitivity analysis (Depaoli et al., 2020; Kruschke, 2021) using only weakly informative priors allowed us to examine whether this outcome was sensitive to the prior specifications. Using only weakly informative priors which exert little influence on the data (Lemoine, 2019) showed that, while the effects of height and weight were unaffected, the key sex difference was sensitive to the prior information obtained from the calliper-measured face width data. The posterior probability of this average sex difference being positive dropped from almost 100 % to 82 % and thus captured zero as a credible hypothesis. One explanation for this is that facial width measured from photographs is a noisier, less reliable measure than when taken with calliper-based measures, as it can plausibly vary as a function of head orientation and camera distances (Zhang et al., 2020). Within this analysis, multiple image sets were combined with varying camera settings, and though we incorporated appropriate statistical controls, this may have impacted the precision of estimation. However, it is worth reiterating here the practicality of Bayesian inference and its alternatives – without a prior distribution formulated from a large, gold-standard measurement of facial width, our conclusions would be more uncertain. We contend this conclusion would be unwarranted, given the quality of prior information that can be incorporated. The reliability of measures of face width from facial images is a question easily addressed in existing databases of image-based fWHR measures.

Given the current results, what might be the explanation for greater face width in men? At an ontogenetic level, greater face width may be a consequence of the larger and faster growth of the mandible observed in boys and adolescent males (Matthews et al., 2018). Examining pre-and post-pubertal craniofacial growth trajectories, Matthews et al. (2018) highlighted that a sex difference in the midface and mandible emerges in early childhood and becomes pronounced over time, with the simultaneous flattening of the midface and protrusion of the jaw (along with the brow ridge) occurring more rapidly in males, particularly during puberty. This growth trajectory is in line with the consistent sexual dimorphism observed in the mandible in humans across recent history, which seems unaffected by climactic conditions that would point to natural selection as a driver of sexual dimorphism (Bejdová et al., 2013). As such, the ultimate driver of the facial width difference may be sexual selection. Whether this operates through intra-sexual competition (e.g., the 'protective buttressing' hypothesis, Carrier & Morgan, 2015; Puts, 2016) or through inter-sexual selection by females with an attractiveness preference, similar to non-human primates with conspicuous facial displays (Dixson et al., 2005; Petersen & Higham, 2020), is unclear. Indeed, the very first description of how sexual selection may have shaped fWHR (Weston et al., 2007), pointed out as a secondary finding that male facial width departs from that of females at puberty. While adult male and female face heights are similar, facial width is larger in males. However, Weston et al. (2007) also show that the relationship between skull size and facial width seems similar between males and females, suggesting that larger facial width is simply due to ontogenetic scaling. We have shown here that facial width dimorphism is reliably different, and remains after various controls for body size, arguably more useful than the skull sizes employed in the original work of Weston et al. (2007). While sexual selection is likely implicated in these facial metrics, future work - which will hopefully shift focus from fWHR to facial width – could test how perceptual, social, and psychological variables associated with facial width, and these findings can illuminate the way in which sexual selection shapes the human face.

One limitation to our studies is that, despite using several samples of faces, they are restricted in age range. Some findings suggest dimorphism in fWHR changes throughout life (Summersby et al., 2022), as well as between ethnicities (Kramer, 2015). While this may be the case, it is unlikely that these differences would substantially overturn the estimates presented here. A broader limitation is our use of static images as stimuli. Recent work on fWHR has illustrated the close links between fixed properties of faces, such as bone structure, and the dynamic changes to appearance from facial musculature (Windmann et al., 2023). That is, the role of fWHR in processes such as social perception may be more to do with emotion overgeneralisation (Kramer, 2016; Todorov et al., 2013) as opposed to physiological aspects, and simple photographs cannot discriminate between these sources. Even beyond fWHR, facial properties that have large effect sizes when studied in photographic contexts show mixed predictive utility in live interactions (Zhao et al., 2023). Ultimately, the role faces and their structures play in complex social environments may be largely incompatible with standard research paradigms in the field as we have used here, an approach that has been recently criticised (Satchell, 2019; Satchell et al., 2023).

Here, the use of Bayesian inference allowed us to simultaneously model sexual dimorphism in fWHR across a range of established measures, injecting prior information into the analysis to avoid the pitfalls of NHST that have led to a mixed set of findings for this widely studied facial parameter. We conclude that any sexual dimorphism is in the opposite direction to what is predicted, and that it is certainly smaller than other sources of variability that are irrelevant to evolutionary hypotheses. We extend the findings of others (Caton & Dixson, 2022) who contend that facial width alone carries evolutionarily relevant cues, and that researchers interested in facial morphology and its links with perception and behaviour should shift focus away from fWHR and on to bizygomatic width (Costa et al., 2017). We also suggest that in doing so, researchers consider the prior evidence for these hypotheses and have more stringent standards about what counts as evidence for or against them.

CRediT authorship contribution statement

Alex L. Jones: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Tobias L. Kordsmeyer:** Writing – review & editing, Resources, Data curation. **Robin S.S. Kramer:** Writing – review & editing, Resources, Data curation. **Julia Stern:** Writing – review & editing, Resources, Data curation. **Lars Penke:** Writing – review & editing, Resources, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Thomas Richardson for his insights provided early in the project and advice on the first draft.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2025.106781.

References

Alrajih, S., & Ward, J. (2014). Increased facial width-to-height ratio and perceived dominance in the faces of the UK's leading business leaders. *British Journal of Psychology, 105*(2), 153–161. https://doi.org/10.1111/bjop.12035

Archer, J., & Thanzami, V. (2007). The relation between physical aggression, size and strength, among a sample of young Indian men. *Personality and Individual Differences, 43*(3), 627–633. https://doi.org/10.1016/j.paid.2007.01.005

Archer, J., & Thanzami, V. (2009). The relation between mate value, entitlement, physical aggression, size and strength among a sample of young Indian men. *Evolution and Human Behavior, 30*(5), 315–321. https://doi.org/10.1016/j.evolhumbehav.2009.03.003

Asendorpf, J. B., Penke, L., & Back, M. D. (2011). From dating to mating and relating: Predictors of initial and long–term outcomes of speed–dating in a community sample. *European Journal of Personality, 25*(1), 16–30. https://doi.org/10.1002/per.768

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences of the United States of America, 104*(46), 17948–17953. https://doi.org/10.1073/pnas.0705435104

Bejdová, Š., Krajíček, V., Velemínská, J., Horák, M., & Velemínský, P. (2013). Changes in the sexual dimorphism of the human mandible during the last 1200 years in Central Europe. *HOMO, 64*(6), 437–453. https://doi.org/10.1016/j.jchb.2013.05.003

Bonnefon, J.-F., Hopfensitz, A., & Neys, W. D. (2017). Trustworthiness perception at zero acquaintance: Consensus, accuracy, and prejudice. *The Behavioral and Brain Sciences, 40*, Article e4. https://doi.org/10.1017/S0140525X15002319

Valentine, K. A., Li, N. P., Penke, L., & Perrett, D. I. (2014). Judging a man by the width of his face: The role of facial ratios and dominance in mate choice at speed-dating events. *Psychol. Sci., 25*(3). https://doi.org/10.1177/0956797613511823

von Borell, C. J., Kordsmeyer, T. L., Gerlach, T. M., & Penke, L. (2019). An integrative study of facultative personality calibration. *Evolution and Human Behavior, 40*(2), 235–248. https://doi.org/10.1016/j.evolhumbehav.2019.01.002

Burton, C. M., & Rule, N. O. (2013). Judgments of height from faces are informed by dominance and facial maturity. *Social Cognition, 31*(6), 672–685. https://doi.org/10.1521/soco.2013.31.6.672

Carré, J. M., & McCormick, C. M. (2008). In your face: Facial metrics predict aggressive behaviour in the laboratory and in varsity and professional hockey players. *Proceedings. Biological sciences, 275*(1651), 2651–2656. https://doi.org/10.1098/rspb.2008.0873

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science, 20*(10), 1194–1198. https://doi.org/10.1111/j.1467-9280.2009.02423.x

Carrier, D. R., & Morgan, M. H. (2015). Protective buttressing of the hominin face. *Biological Reviews, 90*(1), 330–346. https://doi.org/10.1111/brv.12112

Caton, N. R., & Dixson, B. J. W. (2022). Beyond facial width-to-height ratios: Bizygomatic width is highly sexually dimorphic when adjusting for allometry. *Biology Letters, 18*(10), 20220211. https://doi.org/10.1098/rsbl.2022.0211

Caton, N. R., Pearson, S. G., & Dixson, B. J. W. (2022). Is facial structure an honest cue to real-world dominance and fighting ability in men? A pre-registered direct replication of Třebický et al. (2013). *Evolution and Human Behavior, 43*(4), 314–324. https://doi.org/10.1016/j.evolhumbehav.2022.04.002

Chambert, T., Rotella, J. J., & Higgs, M. D. (2014). Use of posterior predictive checks as an inferential tool for investigating individual heterogeneity in animal population vital rates. *Ecology and Evolution, 4*(8), 1389–1397. https://doi.org/10.1002/ece3.993

Coetzee, V., Chen, J., Perrett, D. I., & Stephen, I. D. (2010). Deciphering faces: Quantifiable visual cues to weight. *Perception, 39*(1), 51–61. https://doi.org/10.1068/p6560

Costa, M., Lio, G., Gomez, A., & Sirigu, A. (2017). How components of facial width to height ratio differently contribute to the perception of social traits. *PLoS One, 12*(2), Article e0172739. https://doi.org/10.1371/journal.pone.0172739

Deaner, R. O., Goetz, S. M. M., Shattuck, K., & Schnotala, T. (2012). Body weight, not facial width-to-height ratio, predicts aggression in pro hockey players. *Journal of Research in Personality, 46*(2), 235–238. https://doi.org/10.1016/j.jrp.2012.01.005

Depaoli, S., Winter, S. D., & Visser, M. (2020). The importance of prior sensitivity analysis in Bayesian statistics: Demonstrations using an interactive shiny app. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.608045

Dixson, A., Dixson, B., & Anderson, M. (2005). Sexual selection and the evolution of visually conspicuous sexually dimorphic traits in male monkeys, apes, and human beings. *Annual Review of Sex Research, 16*, 1–19.

Dixson, B. J. W., Lee, A. J., Sherlock, J. M., & Talamas, S. N. (2017). Beneath the beard: Do facial morphometrics influence the strength of judgments of men's beardedness? *Evolution and Human Behavior, 38*(2), 164–174. https://doi.org/10.1016/j.evolhumbehav.2016.08.004

Eisenbruch, A. B., Lukaszewski, A. W., Simmons, Z. L., Arai, S., & Roney, J. R. (2018). Why the wide face? Androgen receptor gene polymorphism does not predict men's facial width-to-height ratio. *Adaptive Human Behavior and Physiology, 4*(2), 138–151. https://doi.org/10.1007/s40750-017-0084-x

Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Common neural mechanisms for the evaluation of facial trustworthiness and emotional expressions as revealed by behavioral adaptation. *Perception, 39*(7), 931–941.

Fisher, C. I., Hahn, A. C., DeBruine, L. M., & Jones, B. C. (2014). Integrating shape cues of adiposity and color information when judging facial health and attractiveness. *Perception, 43*(6), 499–508. https://doi.org/10.1068/p7728

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2015). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness, 5*(2), 189–211. https://doi.org/10.1080/19345747.2011.618213

Gelman, A., & Pardoe, I. (2006). Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics, 48*(2), 241–251. https://doi.org/10.1198/004017005000000517

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., … Modrák, M. (2020). *Bayesian workflow* (arXiv:2011.01808). arXiv. https://doi.org/10.48550/arXiv.2011.01808.

Geniole, S. N., Denson, T. F., Dixson, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat. *PLoS One, 10*(7), Article e0132726. https://doi.org/10.1371/journal.pone.0132726

Geniole, S. N., Molnar, D. S., Carré, J. M., & McCormick, C. M. (2014). The facial width-to-height ratio shares stronger links with judgments of aggression than with judgments of trustworthiness. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1526–1541. https://doi.org/10.1037/a0036732

Haines, N., Kvam, P., Irving, L. H., Smith, C., Beauchaine, T. P., Pitt, M., … Turner, B. M. (2020). Theoretically informed generative models can advance the psychological and brain sciences: Lessons from the reliability paradox. *OSF*. https://doi.org/10.31234/osf.io/xr7y3

Haselhuhn, M. P., & Wong, E. M. (2011). Bad to the bone: Facial structure predicts unethical behaviour. *Proceedings of the Royal Society B: Biological Sciences, 279*(1728), 571–576. https://doi.org/10.1098/rspb.2011.1193

Hodges-Simeon, C. R., Albert, G., Richardson, G. B., McHale, T. S., Weinberg, S. M., Gurven, M., & Gaulin, S. J. C. (2021). Was facial width-to-height ratio subject to sexual selection pressures? A life course approach. *PLoS One, 16*(3), Article e0240284. https://doi.org/10.1371/journal.pone.0240284

Hodges-Simeon, C. R., Sobraske, K. N. H., Samore, T., Gurven, M., & Gaulin, S. J. C. (2016). Facial width-to-height ratio (fWHR) is not associated with adolescent testosterone levels. *PLoS One, 11*(4), Article e0153083. https://doi.org/10.1371/journal.pone.0153083

Holzleitner, I. J., Hunter, D. W., Tiddeman, B. P., Seck, A., Re, D. E., & Perrett, D. I. (2014). Men's facial masculinity: When (body) size matters. *Perception, 43*(11), 1191–1202. https://doi.org/10.1068/p7673

Jaeger, B., & Jones, A. L. (2022). Which facial features are central in impression formation? *Social Psychological and Personality Science, 13*(2), 553–561. https://doi.org/10.1177/19485506211034979

Jaeger, B., Sleegers, W. W. A., Evans, A. M., Stel, M., & van Beest, I. (2019). The effects of facial attractiveness and trustworthiness in online peer-to-peer markets. *Journal of Economic Psychology, 75*, Article 102125. https://doi.org/10.1016/j.joep.2018.11.004

Jones, A. L. (2018). The influence of shape and colour cue classes on facial health perception. *Evolution and Human Behavior, 39*(1), 19–29. https://doi.org/10.1016/j.evolhumbehav.2017.09.005

Jones, A. L., & Jaeger, B. (2019). Biological bases of beauty revisited: The effect of symmetry, averageness, and sexual dimorphism on female facial attractiveness. *Symmetry, 11*(2), 279. https://doi.org/10.3390/sym11020279

Jones, A. L., Jaeger, B., & Schild, C. (2021). No credible evidence for links between 2D:4D and COVID-19 outcomes: A probabilistic perspective on digit ratio, ACE variants, and national case fatalities. *Early Human Development, 152*, Article 105272. https://doi.org/10.1016/j.earlhumdev.2020.105272

Jones, A. L., Kramer, R. S. S., & Ward, R. (2012). Signals of personality and health: The contributions of facial shape, skin texture, and viewing angle. *Journal of Experimental Psychology. Human Perception and Performance, 38*(6), 1353–1361. https://doi.org/10.1037/a0027078

Jones, A. L., Porcheron, A., Sweda, J. R., Morizot, F., & Russell, R. (2016). Coloration in different areas of facial skin is a cue to health: The role of cheek redness and periorbital luminance in health perception. *Body Image, 17*, 57–66. https://doi.org/10.1016/j.bodyim.2016.02.001

Jones, A. L., Russell, R., & Ward, R. (2015). Cosmetics alter biologically-based factors of beauty: Evidence from facial contrast. *Evolutionary Psychology, 13*(1), 210–229. https://doi.org/10.1177/147470491501300113

Jones, A. L., Schild, C., & Jones, B. C. (2021). Facial metrics generated from manually and automatically placed image landmarks are highly correlated. *Evolution and Human Behavior, 42*(3), 186–193. https://doi.org/10.1016/j.evolhumbehav.2020.09.002

Jünger, J., Kordsmeyer, T. L., Gerlach, T. M., & Penke, L. (2018). Fertile women evaluate male bodies as more attractive, regardless of masculinity. *Evolution and Human Behavior, 39*(4), 412–423. https://doi.org/10.1016/j.evolhumbehav.2018.03.007

Jünger, J., Motta-Mena, N. V., Cardenas, R., Bailey, D., Rosenfield, K. A., Schild, C., … Puts, D. A. (2018). Do women's preferences for masculine voices shift across the ovulatory cycle? *Hormones and Behavior, 106*, 122–134. https://doi.org/10.1016/j.yhbeh.2018.10.008

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research, 10*, 1755–1758.

Kordsmeyer, T. L., Freund, D., Pita, S. R., Jünger, J., & Penke, L. (2019). Further evidence that facial width-to-height ratio and global facial masculinity are not positively associated with testosterone levels. *Adaptive Human Behavior and Physiology, 5*(2), 117–130. https://doi.org/10.1007/s40750-018-0105-4

Kordsmeyer, T. L., Freund, D., Ueshima, A., Kuroda, K., Kameda, T., & Penke, L. (2024). Halo effect of faces and bodies: Cross-cultural similarities and differences between German and Japanese observers. *Personality Science, 5*. https://doi.org/10.1177/27000710241257814, 27000710241257814.

Kordsmeyer, T. L., & Penke, L. (2019). Effects of male testosterone and its interaction with cortisol on self- and observer-rated personality states in a competitive mating context. *Journal of Research in Personality, 78*, 76–92. https://doi.org/10.1016/j.jrp.2018.11.001

Kordsmeyer, T. L., Stern, J., & Penke, L. (2019). 3D anthropometric assessment and perception of male body morphology in relation to physical strength. *American Journal of Human Biology, 31*(5), Article e23276. https://doi.org/10.1002/ajhb.23276

Kramer, R. S. S. (2015). Facial width-to-height ratio in a large sample of commonwealth games athletes. *Evolutionary Psychology, 13*(1), 197–209. https://doi.org/10.1177/147470491501300112

Kramer, R. S. S. (2016). Within-person variability in men's facial width-to-height ratio. *PeerJ, 4*, Article e1801. https://doi.org/10.7717/peerj.1801

Kramer, R. S. S. (2017). Sexual dimorphism of facial width-to-height ratio in human skulls and faces: A meta-analytical approach. *Evolution and Human Behavior, 38*(3), 414–420. https://doi.org/10.1016/j.evolhumbehav.2016.12.002

Kramer, R. S. S., Jones, A. L., & Ward, R. (2012). A lack of sexual dimorphism in width-to-height ratio in White European faces using 2D photographs, 3D scans, and anthropometry. *PLoS One, 7*(8), Article e42705. https://doi.org/10.1371/journal.pone.0042705

Kramer, R. S. S., & Ward, R. (2010). Internal facial features are signals of personality and health. *The Quarterly Journal of Experimental Psychology, 63*(11), 2273–2287. https://doi.org/10.1080/17470211003770912

Krenn, B., & Buehler, C. (2019). Facial features and unethical behavior – Doped athletes show higher facial width-to-height ratios than non-doping sanctioned athletes. *PLoS One, 14*(10), Article e0224472. https://doi.org/10.1371/journal.pone.0224472

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2 ed.). Academic Press.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science, 1*(2), 270–280. https://doi.org/10.1177/2515245918771304

Kruschke, J. K. (2021). Bayesian analysis reporting guidelines. *Nature Human Behaviour, 5*(10), 1282–1291. https://doi.org/10.1038/s41562-021-01177-7

Lefevre, C. E., Etchells, P. J., Howell, E. C., Clark, A. P., & Penton-Voak, I. S. (2014). Facial width-to-height ratio predicts self-reported dominance and aggression in males and females, but a measure of masculinity does not. *Biology Letters, 10*(10), 20140729. https://doi.org/10.1098/rsbl.2014.0729

Lefevre, C. E., & Lewis, G. J. (2014). Perceiving aggression from facial structure: Further evidence for a positive association with facial width–to–height ratio and masculinity, but not for moderation by self–reported dominance. *European Journal of Personality, 28*(6), 530–537. https://doi.org/10.1002/per.1942

Lefevre, C. E., Lewis, G. J., Bates, T. C., Dzhelyova, M., Coetzee, V., Deary, I. J., & Perrett, D. I. (2012). No evidence for sexual dimorphism of facial width-to-height ratio in four large adult samples. *Evolution and Human Behavior, 33*(6), 623–627. https://doi.org/10.1016/j.evolhumbehav.2012.03.002

Lefevre, C. E., Lewis, G. J., Perrett, D. I., & Penke, L. (2013). Telling facial metrics: Facial width is associated with testosterone levels in men. *Evolution and Human Behavior, 34*(4), 273–279. https://doi.org/10.1016/j.evolhumbehav.2013.03.005

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos, 128*(7), 912–928. https://doi.org/10.1111/oik.05985

Lin, M., Lucas, H. C., & Shmueli, G. (2013). Research commentary—Too big to fail: Large samples and the p-value problem. *Information Systems Research, 24*(4), 906–917. https://doi.org/10.1287/isre.2013.0480

MacDonell, E. T., Geniole, S. N., & McCormick, C. M. (2018). Force versus fury: Sex differences in the relationships among physical and psychological threat potential, the facial width-to-height ratio, and judgements of aggressiveness. *Aggressive Behavior, 44*(5), 512–523. https://doi.org/10.1002/ab.21771

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.02767

Martin, O., Kumar, R., & Lao, J. (2021). *Bayesian modeling and computation in Python*. CRC Press. https://bayesiancomputationbook.com/welcome.html.

Matthews, H. S., Penington, A. J., Hardiman, R., Fan, Y., Clement, J. G., Kilpatrick, N. M., & Claes, P. D. (2018). Modelling 3D craniofacial growth trajectories for population comparison and classification illustrated using sex-differences. *Scientific Reports, 8*(1), 4771. https://doi.org/10.1038/s41598-018-22752-5

Mayer, C., Windhager, S., Schaefer, K., & Mitteroecker, P. (2017). BMI and WHR are reflected in female facial shape and texture: A geometric morphometric image analysis. *PLoS One, 12*(1), Article e0169336. https://doi.org/10.1371/journal.pone.0169336

Mayew, W. J. (2013). Reassessing the association between facial structure and baseball performance: A comment on Tsujimura & Banissy (2013). *Biology Letters, 9*(5), 20130538. https://doi.org/10.1098/rsbl.2013.0538

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and STAN*. CRC Press.

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon Statistical Significance. *The American Statistician, 73*(sup1), 235–245. https://doi.org/10.1080/00031305.2018.1527253

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34*(2), 103–115. https://doi.org/10.1086/288135

Merlhiot, G., Mondillon, L., Méot, A., Dutheil, F., & Mermillod, M. (2021). Facial width-to-height ratio underlies perceived dominance on facial emotional expressions. *Personality and Individual Differences, 172*, Article 110583. https://doi.org/10.1016/j.paid.2020.110583

Mileva, V. R., Cowan, M. L., Cobey, K. D., Knowles, K. K., & Little, A. C. (2014). In the face of dominance: Self-perceived and other-perceived dominance are positively associated with facial-width-to-height ratio in men. *Personality and Individual Differences, 69*, 115–118. https://doi.org/10.1016/j.paid.2014.05.019

Morrison, D., Wang, H., Hahn, A. C., Jones, B. C., & DeBruine, L. M. (2017). Predicting the reward value of faces and bodies from social perception. *PLoS One, 12*(9). https://doi.org/10.1371/journal.pone.0185093

Noser, E., Schoch, J., & Ehlert, U. (2018). The influence of income and testosterone on the validity of facial width-to-height ratio as a biomarker for dominance. *PLoS One, 13*(11), Article e0207333. https://doi.org/10.1371/journal.pone.0207333

Orben, A., & Lakens, D. (2020). Crud (Re)Defined. *Advances in Methods and Practices in Psychological Science, 3*(2), 238–247. https://doi.org/10.1177/2515245920917961

Özener, B. (2012). Facial width-to-height ratio in a Turkish population is not sexually dimorphic and is unrelated to aggressive behavior. *Evolution and Human Behavior, 33*(3), 169–173. https://doi.org/10.1016/j.evolhumbehav.2011.08.001

Paquette, S., Gordon, C., & Bradtmiller, B. (2009). Anthropometric survey (ANSUR) II pilot study: Methods and summary statistics. Technical report NATICK/TR-09/014. https://apps.dtic.mil/sti/pdfs/ADA498172.pdf.

Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: A more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology, 95*(5), 1113–1135. https://doi.org/10.1037/0022-3514.95.5.1113

Petersen, R. M., & Higham, J. P. (2020). The role of sexual selection in the evolution of facial displays in male non-human primates and men. *Adaptive Human Behavior and Physiology, 6*(2), 249–276. https://doi.org/10.1007/s40750-020-00139-z

Price, M. E., Dunn, J., Hopkins, S., & Kang, J. (2012). Anthropometric correlates of human anger. *Evolution and Human Behavior, 33*(3), 174–181. https://doi.org/10.1016/j.evolhumbehav.2011.08.004

Puts, D. (2016). Human sexual selection. *Current Opinion in Psychology, 7*, 28–32. https://doi.org/10.1016/j.copsyc.2015.07.011

Puts, D. A. (2010). Beauty and the beast: Mechanisms of sexual selection in humans. *Evolution and Human Behavior, 31*(3), 157–175. https://doi.org/10.1016/j.evolhumbehav.2010.02.005

Robertson, J. M., Kingsley, B. E., & Ford, G. C. (2017). Sexually dimorphic faciometrics in humans from early adulthood to late middle age: Dynamic, declining, and differentiated. *Evolutionary Psychology, 15*(3), Article 1474704917730640. https://doi.org/10.1177/1474704917730640

Rostovtseva, V. V., Mezentseva, A. A., Windhager, S., & Butovskaya, M. L. (2021). Sexual dimorphism in facial shape of modern Buryats of southern Siberia. *American Journal of Human Biology: The Official Journal of the Human Biology Council, 33*(2), Article e23458. https://doi.org/10.1002/ajhb.23458

Russell, R., Porcheron, A., Sweda, J. R., Jones, A. L., Mauger, E., & Morizot, F. (2016). Facial contrast is a cue for perceiving health from the face. *Journal of Experimental Psychology: Human Perception and Performance, 42*(9), 1354–1362. https://doi.org/10.1037/xhp0000219

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science, 2*, Article e55. https://doi.org/10.7717/peerj-cs.55

Saribay, S. A., Biten, A. F., Meral, E. O., Aldan, P., Třebický, V., & Kleisner, K. (2018). The Bogazici face database: Standardized photographs of Turkish faces with supporting materials. *PLoS One, 13*(2), Article e0192018. https://doi.org/10.1371/journal.pone.0192018

Saribay, S. A., & Kleisner, K. (2018). Are political views and religiosity related to facial morphology? Evidence from a Turkish sample. *Personality and Individual Differences, 124*, 25–30. https://doi.org/10.1016/j.paid.2017.11.035

Satchell, L., Jaeger, B., Jones, A., López, B., & Schild, C. (2023). Beyond reliability in first impressions research: Considering validity and the need to "mix it up with folks.". *Social Psychological Bulletin, 18*, 1–21. https://doi.org/10.32872/spb.10211

Satchell, L. P. (2019). From photograph to face-to-face: Brief interactions change person and personality judgments. *Journal of Experimental Social Psychology, 82*, 266–276. https://doi.org/10.1016/j.jesp.2019.02.010

Scott, N. J., Kramer, R. S. S., Jones, A. L., & Ward, R. (2013). Facial cues to depressive symptoms and their associated personality attributions. *Psychiatry Research, 208*(1), 47–53. https://doi.org/10.1016/j.psychres.2013.02.027

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society B: Biological Sciences, 276*(1656), 575–584. https://doi.org/10.1098/rspb.2008.1177

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences, 106*(35), 15073–15078. https://doi.org/10.1073/pnas.0904312106

Stern, J., Kordsmeyer, T. L., & Penke, L. (2021). A longitudinal evaluation of ovulatory cycle shifts in women's mate attraction and preferences. *Hormones and Behavior, 128*, Article 104916. https://doi.org/10.1016/j.yhbeh.2020.104916

Stirrat, M., & Perrett, D. I. (2010). Valid facial cues to cooperation and trust: Male facial width and trustworthiness. *Psychological Science, 21*(3), 349–354. https://doi.org/10.1177/0956797610362647

Stirrat, M., Stulp, G., & Pollet, T. V. (2012). Male facial width is associated with death by contact violence: Narrow-faced males are more likely to die from contact violence. *Evolution and Human Behavior, 33*(5), 551–556. https://doi.org/10.1016/j.evolhumbehav.2012.02.002

Summersby, S., Harris, B., Denson, T. F., & White, D. (2022). Tracking sexual dimorphism of facial width-to-height ratio across the lifespan: Implications for perceived aggressiveness. *Royal Society Open Science, 9*(5), Article 211500. https://doi.org/10.1098/rsos.211500

Tipples, J. (2007). Wide eyes and an open mouth enhance facial threat. *Cognition and Emotion, 21*(3), 535–557. https://doi.org/10.1080/02699930600780886

Todorov, A., Dotsch, R., Porter, J. M., Oosterhof, N. N., & Falvello, V. B. (2013). Validation of data-driven computational models of social perception of faces. *Emotion, 13*(4), 724–738. https://doi.org/10.1037/a0032335

Třebický, V., Fialová, J., Kleisner, K., Roberts, S. C., Little, A. C., & Havlíček, J. (2015). Further evidence for links between facial width-to-height ratio and fighting success: Commentary on Zilioli et al. (2014). *Aggressive Behavior, 41*(4), 331–334. https://doi.org/10.1002/ab.21559

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience, 6*(6), 624–631. https://doi.org/10.1038/nn1057

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology, 60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Wen, G., & Zheng, L. (2020). Facial width to height ratio predicts physical aggression in committed relationships in men and dominance in women in China. *Personality and Individual Differences, 157*, Article 109832. https://doi.org/10.1016/j.paid.2020.109832

Weston, E. M., Friday, A. E., & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PLoS One, 2*(8), Article e710. https://doi.org/10.1371/journal.pone.0000710

Windmann, S., Steinbrück, L., & Stier, P. (2023). Overgeneralizing emotions: Facial width-to-height revisited. *Emotion, 23*(1), 163–181. https://doi.org/10.1037/emo0001033

Wolffhechel, K., Hahn, A. C., Jarmer, H., Fisher, C. I., Jones, B. C., & DeBruine, L. M. (2015). Testing the utility of a data-driven approach for assessing BMI from face images. *PLoS One, 10*(10), Article e0140347. https://doi.org/10.1371/journal.pone.0140347

Young, J. W. (1993). *Head and face anthropometry of adult U.S. civilians (DOT/FAA/AM-93/10)*. Civil Aeromedical Institute. https://rosap.ntl.bts.gov/view/dot/21363.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*(3), 1497. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Zhang, D., Lin, H., & Perrett, D. I. (2020). Apparent emotional expression explains the effects of head posture on perceived trustworthiness and dominance, but a measure of facial width does not. *Perception, 49*(4), 422–438. https://doi.org/10.1177/0301006620909286

Zhang, W., Hahn, A. C., Cai, Z., Lee, A. J., Holzleitner, I. J., DeBruine, L. M., & Jones, B. C. (2018). No evidence that facial width-to-height ratio (fWHR) is associated with women's sexual desire. *PLoS One, 13*(7), Article e0200308. https://doi.org/10.1371/journal.pone.0200308

Zhao, A. A. Z., Harrison, K., Holland, A., Wainwright, H. M., Ceccato, J.-M., Sidari, M. J., … Zietsch, B. P. (2023). Objectively measured facial traits predict in-person evaluations of facial attractiveness and prosociality in speed-dating partners. *Evolution and Human Behavior, 44*(4), 315–323. https://doi.org/10.1016/j.evolhumbehav.2023.05.001

Zilioli, S., Sell, A. N., Stirrat, M., Jagore, J., Vickerman, W., & Watson, N. V. (2015). Face of a fighter: Bizygomatic width as a cue of formidability. *Aggressive Behavior, 41*(4), 322–330. https://doi.org/10.1002/ab.21544