Further evidence that facial width-to-height ratio and global facial masculinity are not positively associated with testosterone levels

Tobias L. Kordsmeyer[*1], Daniel Freund[1], Sara Rodrigues Pita[2], Julia Jünger[1], & Lars Penke[1]

[1]Department of Psychology & Leibniz ScienceCampus Primate Cognition

University of Goettingen

Gosslerstr. 14

37073 Goettingen, Germany

[2]School of Biological Science

University of Edinburgh

West Mains Road

Edinburgh EH9 3JT, Scotland, UK

*Corresponding author.

E-mail address: tob.kor@gmail.com

Abstract

Objectives Facial masculinity, as for example measured by the facial width-to-height ratio (fWHR) or the global facial masculinity index, has been associated with a vast range of behavioural traits, including dominance and aggression. Further, facial masculinity is thought to be influenced by testosterone (T) levels as an underlying mechanism. However, a recent meta-analysis on fWHR and T levels provided non-significant associations in men, which we wanted to examine further in men and additionally in women.

Methods We examined whether fWHR and global facial masculinity are positively associated with salivary baseline T and T reactivity in 140 men (age 18-34 years), as well as with salivary baseline T and hair T concentrations in 151 women (age 18-35 years).

Results No associations of salivary baseline T, T reactivity or hair T levels with fWHR or global facial masculinity were observed. Additional analyses revealed sex differences in sexual dimorphism in fWHR and global facial masculinity: men had generally higher global facial masculinity compared to women, but unexpectedly a lower fWHR.

Conclusions Overall, our results provide further evidence that neither fWHR nor global facial masculinity are related to T levels and question earlier findings on male-biased sexual dimorphism in fWHR.

Keywords: *facial masculinity, facial width to height ratio, salivary testosterone, hair testosterone, sexual dimorphism*

Introduction

Human faces are sexually dimorphic. Structural facial dimorphism is often captured in landmark-based indices, such as the global masculinity index (Penton-Voak et al., 2001). In the last decade, a new index, the facial width-to-height ratio (fWHR; bizygomatic width divided by face height (distance between the upper lip and the brow ridge or nasion); Carré & McCormick, 2008; Weston, Friday, & Liò, 2007), has received considerable attention. fWHR has been associated with a large number of behavioural traits that influence human social interactions. These include traits such as fighting ability (Třebický et al., 2015), dominance and threat behaviour (display of selfish, pejorative or aggressive traits; Geniole, Denson, Dixson, Carré, & McCormick, 2015), perceived formidability (Zilioli et al., 2015), and aggressiveness (Carré & McCormick, 2008; Haselhuhn, Ormiston, & Wong, 2015; Lieberz et al., 2017). An above-average fWHR has also been found in CEOs (chief executive officers) and prosocial leaders of NGOs (non-governmental organizations, *N* = 103, Hahn et al., 2017), hence the association between fWHR and social rank does not only embody antisocial behaviour. Instead, fWHR could be thought of as a predictor of social rank more generally, in both competitive and prosocial settings. A link between observer-rated dominance and fWHR has also been reported in Capuchin monkeys (*N* = 64; Lefevre et al., 2014), thus suggesting the link may be evolutionarily old. Here we study whether intrasexual differences in facial masculinity, as indexed by global facial masculinity and fWHR, are associated with various measures of testosterone in men and women.

Testosterone as an underlying mechanism

One potential underlying mechanism for the association between fWHR and dominance- or aggressiveness-related traits is the steroid hormone testosterone (T). A longitudinal study has shown that exposure to prenatal T is associated with facial masculinity in adult men and women (*r* = .51, *N* = 183; Whitehouse et al., 2015). Thus, interindividual differences in fWHR may be reflected in T concentrations. Across two samples, Lefevre, Lewis, Perrett, and Penke (2013) tested the associations between men's salivary T reactivity (measured as T change from before to after a speed dating event),

baseline T, fWHR and global facial masculinity (a composite measure of five sexually dimorphic components: (i) eye size, (ii) lower face height/face height, (iii) cheekbone prominence, (iv) face width/lower face height, and (v) mean eyebrow height, see Fig. 1 and Penton-Voak et al., 2001). Measures of fWHR showed positive relationships with baseline T (sample 1: $r = .13$, $n = 181$; sample 2: $r = .26$, $n = 79$) and T reactivity levels (sample 1: $r = .21$, $n = 181$). Associations with T reactivity, but not baseline T, remained stable after controlling for age and BMI. However, no association of global facial masculinity emerged with baseline T or with T reactivity (Lefevre et al., 2013). In an earlier study, the link between global facial masculinity, baseline T and T reactivity after a rigged competitive task in men was investigated ($n = 57$; Pound, Penton-Voak, & Surridge, 2009). In contrast to Lefevre and colleagues (2013), Pound and colleagues (2009) findings suggest that facial masculinity is associated with T reactivity ($r = .36$) but not baseline T ($r = .19$), at least in the winning condition. Hodges-Simeon, Sobraske, Somare, Gurven, and Gaulin (2016) investigated the association of pubertal T levels with fWHR in peri-pubertal men and observed no robust relationship between pubertal T and fWHR ($r = .13$; $N = 75$). However, when controlling for age this association was significant ($r = .28$; see Welker, Bird and Arnocky, 2016 for re-analyses of these data also finding a significant positive effect). Eisenbruch, Lukaszewski, Simmons, Arai, and Roney (2017) found no link between fWHR and baseline T and T reactivity ($r = .09$ and $r = .06$, respectively, $N = 133$ men). However, fWHR was significantly correlated with baseline T in an ethnic sub-sample of white men ($r = .23$, $n = 85$). Bird and colleagues (2016) conducted an internal meta-analysis on correlations between fWHR and baseline T and T reactivity levels across seven samples of men (total $N = 793$), which revealed no significant associations (baseline T: $r = -.04$, T reactivity: $r = -.03$). A second meta-analysis encompassing their seven studies and the two samples from Lefevre and colleagues (2013) was conducted, but the link between baseline T and fWHR remained insignificant ($r = .01$, $N = 1041$). Hence, recent evidence seems to show that fWHR may not be linked with T levels (baseline T and T reactivity), whereas findings for global facial masculinity are mixed. This study was conducted to provide further clarity in men, and to additionally explore these relationships in women.

Testosterone measures from hair samples

Previous studies relied primarily on the assessment of T levels from saliva samples. Recently, a new method of measuring hormone concentrations from human hair samples was developed (e.g., Gao et al., 2013; Russel, Koren, Rieder, & Van Uum, 2012), which has been validated showing significant associations between salivary and hair androgen measures (r = .67 for T, Wang et al., 2019). Assessments of hormone levels from hair samples are supposed to reflect mean hormonal concentrations over a longer time compared to salivary baseline measures (approximately one month per one cm of hair; Russel et al., 2012). Given that fWHR and global facial masculinity are stable traits developing over many years in childhood and adolescence, as are the personality traits that have been found to be associated with them, these more long-term measures of T levels from hair represent an intriguing alternative for investigating relationships with facial characteristics (e.g., Grotzinger, Mann, Patterson, Tackett, Tucker-Drob, & Harden, 2018).

Sexual dimorphism in fWHR

fWHR has been postulated to be sexually dimorphic based on biometric analyses of skulls (Weston et al., 2007), which is in line with associations of fWHR with behavioural traits implied in intrasexual selection (a diverse range of aggression- and dominance- related behaviours and traits), mainly in men. Assuming a positive association between fWHR and T concentrations, considerably higher T levels in men, compared to women, may corroborate the idea of sexually dimorphic fWHR. A small degree of male-biased sexual dimorphism in fWHR has indeed been reported in a meta-analysis (k = 32 samples, overall N = 10,853, d = 0.11; Geniole et al., 2015). Larger fWHR in men compared to women was also found in a large sample of Commonwealth Game athletes (N = 3,479, d = 0.31; Kramer, 2015), but differences were non-significant when controlling for body size. Interestingly, ethnicity was reported to moderate these findings, in that significant sex differences in fWHR were only found in White and Black subsamples but not in Asian-Oriental, Indian-Oriental, or other subsamples. In Kramer's (2017) meta-analysis of cranial data, men's skulls had a slightly larger fWHR compared to women's (k = 87

samples, N = 7941, d = 0.09). However, when accounting for ethnic diversity and using facial instead of cranial data, there was little evidence of sexual dimorphism in fWHR (k = 24 samples, N = 4,161; Kramer, 2017). Only weak evidence was found by Lefevre and colleagues (2012), who reported a significant sex difference in fWHR in only one out of four samples (with larger fWHR in women, the effect of which faded when controlling for BMI; overall N = 924, d = 0.03), and a further null-finding was reported by Robertson, Kingsley, and Ford (2017; N = 444, *partial* η^2 = .00). In sum, there is still much controversy whether fWHR is sexually dimorphic or not, and further research concerning this question is warranted.

Study aim

In light of previous null-findings on the link of baseline T levels with fWHR (Bird et al., 2016) and global facial masculinity (e.g., Lefevre et al., 2013), the aim of the present study is to provide further evidence on the associations of these two facial metrics with salivary baseline T levels in men, as well as investigating them in women. To our knowledge, this study is the first to examine salivary T associations in women (but see Whitehouse et al., 2015 on prenatal T and facial masculinity in both sexes). Additionally, competition-induced salivary T reactivity in men (as in Lefevre et al., 2013) and, for the first time, hair T in women are explored as alternative measures. Finally, sexual dimorphism in fWHR and global facial masculinity are examined.

2. Methods

Participants. One hundred and sixty-five men (age: M = 24.2 years, SD = 3.3, range 18-34) and 157 women (age: M = 23.3 years, SD = 3.4, range 18-35) were recruited from a local participant pool (88.6 % students and 98.8% European ethnicity for men, 98.1% students and 93.6% European ethnicity for women). The final sample sizes for baseline T (N = 140, see below) and T reactivity (n = 104) had sufficient power (> .80) to detect effect sizes of Pearson's r > .23 and r > .26, respectively (Cohen, 1992; Faul, Erdfelder, Lang, & Buchner, 2007). Both samples were part of larger studies on hormonal effects

in mating contexts (see Jünger, Kordsmeyer, Gerlach, & Penke, 2018; Kordsmeyer & Penke, *in press* for more information).

3D face measures. Participants' faces were 3D scanned twice for men and thrice for women. For each participant the most suitable scan (in terms of neutral facial expression and standardized head position) was chosen. Due to inadequate facial 3D scans, strong beardedness or the participants not wanting their scans to be used, the samples in this study were reduced to 140 men and 151 women. The remaining scans were aligned horizontally and 27 facial landmarks were placed using Morph Analyser 2.4 (Tiddeman, Duffy, & Rabey, 2000), from which twelve were relevant for this study (Fig. 1). Two independent coders for men's and women's scans each placed the landmarks after receiving extensive training. The landmarks' x- and y-coordinates from the two coders were aggregated and used to calculate fWHR by dividing the bizygomatic width (distance between the left and right *zygion;* Fig. 1) by face height (distance between the *nasion* and the vermillion of the upper lip, *labrale superius*; Stirrat & Perrett, 2010; Weston et al., 2007). Additionally, four established sexually dimorphic facial metrics (eye size, cheekbone prominence, face width/lower face height, and lower face/face height) were computed and aggregated to form a global facial masculinity index (as in Penton-Voak et al., 2001). Intercoder reliabilities were good for all facial metrics (for men: fWHR *ICC* = .96, *p* < .001; global facial masculinity *ICC* = .91, *p* < .001; for women: fWHR *ICC* = .92, *p* < .001, global facial masculinity *ICC* = .85, *p* <.001; except for cheekbone prominence in men *ICC* = .61, *p* = .08, for women: *ICC* = .75, *p* = < .001).
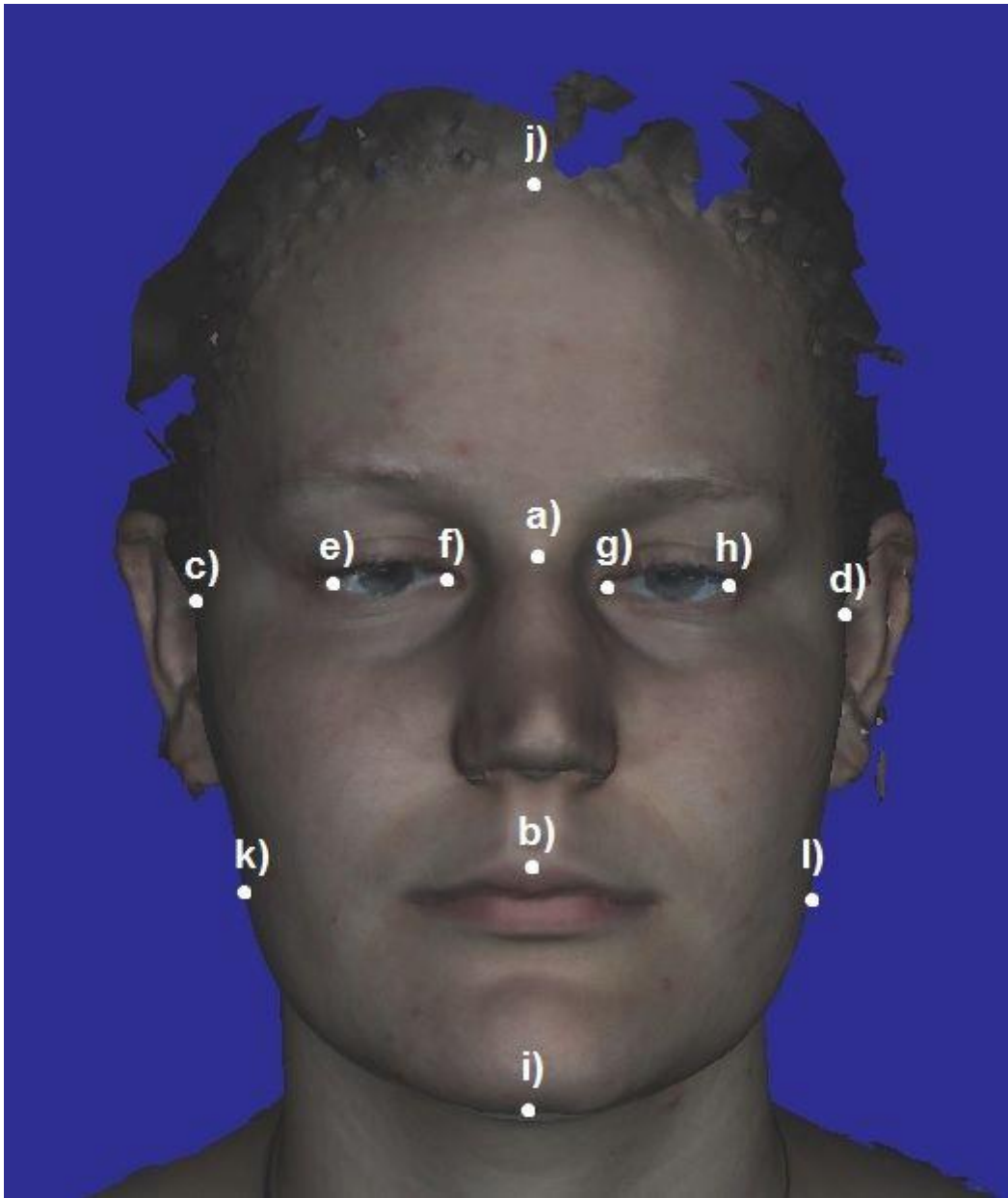
Fig. 1 Landmarks used for morphometric calculations. Measure of fWHR: (a-b)/(c-d); eye size: (e-h)/(f-g); cheekbone prominence: (c-d)/(k-l); face width/lower face height: (c-d)/(e-i); lower face/face height: (e-i)/(j-i).

Hormonal measures. Participants were asked to refrain from drinking alcohol, exercising, taking recreational or non-prescribed clinical drugs on the day of the study, from ingesting caffeine (coffee, tea, coke) or sleeping three hours before the study, and from eating, drinking (except for water), smoking or brushing teeth one hour before their scheduled appointment (Geniole et al., 2013; Lopez, Hay, Conklin, 2009). To check participants' adherence to these instructions and to assess further

potential influences on the saliva samples and hormonal levels, a screening questionnaire was administered at the beginning of the session (Schultheiss & Stanton, 2009). None of the participants indicated to be taking hormonal medication or supplements. All saliva samples were taken in the afternoon to control for circadian variation in hormonal levels (between 2pm and 6pm for men, between 11.30am and 6pm for women; Schultheiss & Stanton, 2009). Participants provided at least 2ml of saliva via unstimulated passive drool through a straw, as recommended by Schultheiss and colleagues (2012). The samples were immediately transported to an ultra-low temperature freezer (-80°C), where salivary testosterone levels (T) are stable for at least 36 months (Granger, Shirtcliff, Booth, Kivlighan, & Schwartz, 2004). At the end of the data collections, saliva samples were shipped on dry ice to the Technical University of Dresden, where men's samples were analysed using chemiluminescence-immuno-assays with high sensitivity (IBL International, Hamburg, Germany) and women's samples via liquid chromatography mass spectrometry (LCMS; Gao, Stalder, & Kirschbaum, 2015). The intra- and inter-assay coefficients (CVs) are below 11%. For men's baseline T levels, saliva samples were collected on two different days (the second sample was collected approx. one week after the first) and then averaged to have a more reliable estimate of baseline T (due to large within-individual variation; Idris, Wan, Zhang, & Punyadeera, 2017). For part of the men's sample (n = 104), T reactivity was measured during the second session after the male participants engaged in a dyadic intrasexual competition (four disciplines: table pinball, snatching game, arm wrestling, turn-taking verbal fluency game; outcome naturally emerging) led by an attractive female confederate (for more detail, see Kordsmeyer & Penke, *in press*). Two post-competition salivary samples were collected, the first immediately after the competition (approx. 15 mins. after onset), the second 15 to 20 minutes later. For women, overall four saliva samples were taken for baseline T measures, two each in the luteal and follicular phases of their menstrual cycle, across two consecutive cycles (i.e., in each cycle one sample 16-18 days before and another 4-11 days before the next menstrual onset, cycle phase was validated with luteinizing hormone tests; for details see Jünger, et al., 2018). Because hormonal levels are known to fluctuate during the menstrual cycle (Tsepelidis et al., 2007), only the two samples

from the luteal phases were used, as were the 3D facial scans from these sessions. Two hair samples were taken from female participants, during their second and fourth visits (for approx. two thirds of the sample in their luteal phases, for the remaining third in their fertile phase; hair T measures are available from n = 128 women as aggregates from two hair samples). Hair at the back of the subject's head was pinned up, and two hair strands (each with a diameter of approximately 3mm, almost full length of participants' hair) were separated. Then, both strands were tightened with a thread, cut as close as possible to the scalp, and packed up in aluminium foil. Finally, the scalp-near ends of the strands were marked on the foil and samples were shipped to the Technical University of Dresden, where only the last grown centimetre of each hair sample was analysed using the LC-MS/MS method (Gao et al., 2013), thus presumably representing the mean testosterone levels within the previous month (Russel et al., 2012). All saliva (baseline, post-competition) and hair T measures were checked for outliers and winsorized to +/- 3 SDs (Mehta, Welker, Zilioli, & Carré, 2015; see supplementary material for details). The Shapiro-Wilk test revealed assumptions of normality were violated for both salivary baseline and post-competition T measures (for men: Ws < .96, ps < .001; for women: Ws < .90, ps < .001); thence, all variables were log10-transformed (after which assumptions of normality were met, Ws > .96, ps > .05, except for the first baseline T measure in men, W = .97, p = .003, assumption met for aggregate of both baseline T measures in men, W = .99, p = .19).

Statistical analysis. Pearson correlations were calculated to examine bivariate associations of (1) fWHR, (2) global facial masculinity, (3) cheekbone prominence, (4) face width/lower face height, (5) lower face/face height and (6) eye size with levels of baseline T in both sexes as well as T reactivity in men. Correlations were run using the means of the two baseline measures taken for both sexes, means of the two hair measures for women, and for the T reactivity aggregate in men. For T reactivity measures, percent changes from baseline levels (using the saliva sample obtained on the day of the second session) to post-competition levels were determined (in accordance with Carré, Iselin, Welker, Hariri, & Dodge, 2014; Roney, Mahler, & Maestripieri, 2003). The difference of baseline and post-competition levels was divided by baseline T. Since both T reactivity values were highly correlated (r =

.65, *p* < .001) we decided to aggregate them, to get one T reactivity value. In line with Lefevre and colleagues (2013), partial correlations controlling for age and BMI were calculated (the latter from participants' measured height and weight; since body size has been suggested to be related to facial metrics and T levels; Kramer, 2017; Třebický et al., 2015). Two-sided *t*-tests were conducted to examine mean differences in fWHR between men and women. To examine sexual dimorphism in global facial masculinity we used raw instead of the usual *z*-standardized scores and excluded eye size, because its absolute scale with considerably larger values compared to the other ratio measures (with values close to 1) would bias mean differences. We aggregated the remaining three components of global facial masculinity as follows: *global facial masculinity = lower face height/face height - cheekbone prominence – face width/lower face height* (thus in slight contrast to Penton-Voak et al., 2001, who formed a similar aggregate based on five components). All analyses were performed using R (R Core Team, 2015).

Data availability. The data and analysis script associated with this research are available at [link].

3. Results

Descriptive statistics and sex differences for all variables are shown in Table 1, correlations between main variables are depicted in Table 2. Baseline T and T reactivity were negatively correlated in men. In women, saliva T and hair T were positively correlated. fWHR correlated negatively with global facial masculinity and was not associated with baseline T in either sex. In men, fWHR was not related to T reactivity, also not after controlling for age and BMI (*r* = -.05, *p* = .62).

Global facial masculinity or any of its four components were not related to baseline T in either sex (men: *r*s < |.17|, *p*s > .06; women: *r*s < |.17|, *p*s > .07) or to T reactivity in men (*r*s < |.14|, *p*s > .19), also not after controlling for age and BMI (*r*s < |.13|, *p*s > .23). Similarly, when analysing the two T reactivity measures in men separately, no significant associations with fWHR, global facial masculinity or any of its four components were found (*r*s < |.14|, *p*s > .17; Table S1 in the online supplementary).

Table 1. *Descriptive statistics and sex differences for all main variables.*

	Men		Women		Sexual dimorphism		
	M	SD	M	SD	t	p	d
Facial width-to-height ratio	2.06	0.15	2.11	0.13	- 3.11	< .01	0.37
Global facial masculinity	-32.72	1.46	-30.09	1.56	4.69	< .001	0.56
Eye size	31.11	1.46	28.40	1.61	15.06	< .001	1.73
Cheekbone prominence	1.13	0.05	1.14	0.05	- 1.05	.29	0.12
Face width/lower face height	1.14	0.06	1.16	0.06	- 3.72	< .001	0.44
Lower face/face height	0.66	0.03	0.64	0.02	5.36	< .001	0.64
Baseline testosterone (1st *pre*-sample)	86.31	32.50	7.97	6.36	30.31	< .001	3.25
Baseline testosterone (2nd *pre*-sample)	77.32	30.78	8.17	8.86	27.53	< .001	2.98
Testosterone reactivity (1st *post*-sample)	85.52	34.06	-	-	-	-	-
Testosterone reactivity (2nd *post*-sample)	91.45	37.44	-	-	-	-	-
Hair testosterone (1st sample)	-	-	0.47	0.32	-	-	-
Hair testosterone (2nd sample)	-	-	0.51	0.50	-	-	-

Note: Men: *n* = 104 – 140; women: *n* = 142 – 151, all testosterone concentrations reported in pg/ml. For sexual dimorphism, positive *t*-values indicate larger values for men.

Table 2. *Bivariate Pearson correlations between all main variables.*

	fWHR	gfm	EZ	CP	fW	lfH	pre T	hair T
Facial width-to-height ratio	-	**-.40**	.11	.14	**.69**	.06	.09	.04
		<.001	.19	.09	<.001	.45	.23	.63
Global facial masculinity	**-.43**	-	**-.46**	**-.61**	**-.63**	**.50**	.09	-.10
	<.001		<.001	<.001	<.001	<.001	.32	.26
Eye size	**.19**	**-.41**	-	.06	.14	.12	-.07	.01
	.02	<.001		.44	.08	.17	.38	.91
Cheekbone prominence	.04	**-.58**	-.05	-	**.19**	-.14	-.07	.13
	.62	<.001	.52		.02	.09	.38	.16
Face width/lower face height	**.72**	**-.66**	.09	**.18**	-	-.07	.07	-.01
	<.001	<.001	.28	.04		.43	.43	.91
Lower face height/face height	.02	**.54**	.15	-.14	-.18	-	.16	-.03
	.85	<.001	.08	.10	.03		.07	.75
Baseline testosterone	.01	.07	.01	-.16	.04	.04	-	**.32**
	.92	.44	.87	.06	.65	.66		<.001
Testosterone reactivity	.01	-.13	.10	.08	.02	-.09	**-.32**	-
	.92	.19	.32	.42	.83	.39	.001	

Note: Correlations for men below, for women above the diagonal; *p*-values are reported below each correlation coefficient. fWHR = facial width-to-height ratio, gfm = global facial masculinity, EZ = eye size, CP = cheekbone prominence, fW = face width/lower face height, lfH = lower face height/face height, pre T = baseline testosterone, hair T = hair testosterone for women.

Examining sexual dimorphism in fWHR, a two-tailed *t*-test showed that fWHR was larger for women than for men (*t* = 3.11, *p* < .01, *d* = 0.37). Global facial masculinity was higher in men compared to women (*t* = 4.69, *p* < .001, *d* = 0.56). Concerning the components of global facial masculinity, men showed a larger eye size than women (*t* = 15.06, *p* < .001, *d* = 1.73), and a larger lower face/face height

(*t* = 5.36, *p* < .001, *d* = 0.64). Face width/lower face height was larger in women (*t* = 3.72, *p* < .001, *d* = 0.44). No significant sex difference emerged for cheekbone prominence (*t* = 1.05, *p* = .29, *d* = 0.12).

4. Discussion

Previous studies suggested associations of fWHR with dominance and aggression. Testosterone (T) has been hypothesized as an underlying mechanism, with some studies reporting positive relationships of baseline or reactive T with fWHR and a related measure, global facial masculinity, though results were inconsistent and a recent meta-analysis revealed non-significant associations for fWHR in men. The aim of the present study was to add further evidence on potential links of fWHR and global facial masculinity with T levels in men and to additionally examine these effects in women, as well as to provide further evidence for a potential sexual dimorphism in fWHR and in global facial masculinity.

Our results suggest that fWHR is neither associated with salivary baseline T in either sex, nor with competition-induced T reactivity (only measured in men) or hair T (only measured in women). This contradicts initial findings, which showed a positive link between fWHR and T levels (Lefevre et al., 2013), but is in line with more recent studies, which revealed no association of fWHR with T levels in adult (meta-analysis, Bird et al., 2016; Eisenbruch et al., 2017) and adolescent men (Hodges-Simeon et al., 2016). Hence, although fWHR has been linked with a vast range of dominance and aggression-related behaviours (Carré & McCormick, 2008; Geniole et al., 2015; Hahn et al., 2017; Lieberz et al., 2017; Třebický et al., 2015; Zilioli et al.; 2015), it does not seem to be robustly related to T levels. Objectively-measured global facial masculinity, an index of sexually dimorphic structural facial characteristics (Lefevre et al., 2013; Penton-Voak et al., 2001), was not related to any of the T measures, contradicting a previous study which indicated an association of facial masculinity with T reactivity (Pound et al., 2009). Nevertheless, our null result corresponds to the findings of Lefevre and colleagues (2013) who found no evidence for such an association. Thus, the relationship of T levels with global facial masculinity is questionable.

As this and earlier studies and meta-analyses have shown that fWHR appears not to be related to baseline or reactive T (e.g., Bird et al., 2016), the question remains which underlying mechanisms

13

can contribute to explaining links of fWHR with focal individuals' behaviour and how they are treated by others (e.g., Geniole et al., 2013; Zilioli et al., 2015). For example, an individual with higher fWHR may be perceived as more dominant or aggressive, without fWHR being calibrated to the individual's traits or behavioural tendencies, instead partly explicable by observers' sensory bias (e.g., Haselhuhn, Wong, & Ormiston, 2013; for similar perceptual biases based on facial characteristics in the context of an economic game and criminal-sentencing decisions, see Eisenbruch, Grillot, Maestripieri, & Roney, 2016, and Wilson & Rule, 2015, respectively). This may result in a self-fulfilling prophecy for the focal individual, in that others treat individuals with higher fWHR as if they were more dominant or aggressive, who in turn habitually tend to behave in such a way. Still, other mechanisms likely have an influence and should be investigated further.

Different measurements of T levels

It is worth mentioning that we did not investigate pubertal or perinatal T levels, which might show stronger links with facial characteristics than baseline or reactive T levels, due to developmental links with traits implicated in craniofacial growth (Hodges-Simeon et al., 2016; Whitehouse et al., 2015). Moreover, due to potentially unreliable measures of salivary T, especially for women's considerably lower T concentrations, the assessment of salivary T is not without pitfalls (Fiers et al., 2014). To tackle this, we aggregated saliva T measurement across two samples in order to get a more stable assessment of both baseline T and T reactivity. Furthermore, we also explored, for the first time, T assessment from women's hair samples as a correlate of intrasexual differences in structural facial dimorphism, which is thought to provide more reliable and aggregated T measures than from saliva (Gao et al., 2013, Russel et al., 2012; while T measures from saliva and hair samples have been shown to be highly correlated, Wang et al., 2019). However, no associations of hair T with any of the facial measures were detected. This contrasts earlier findings on some behaviours and traits which were shown to be related to measures of hair T, but not salivary baseline T (e.g., aggression in $N = 460$ adolescents, partly moderated by hair cortisol; Grotzinger et al., 2018). In general, our results do not support the proposal that T functions as an underlying mechanism for the link between fWHR and dominance.

Sexual dimorphism

In our study, men showed a significantly lower fWHR than women. This is rather unexpected, given that higher fWHR is thought to signal dominance and aggression, and previous findings suggested a higher fWHR for men compared to women (Carré & McCormick, 2008; see Geniole et al., 2015 for a meta-analysis), while other studies found little or no evidence of sexual dimorphism in fWHR (Kramer, Jones, & Ward, 2012; Kramer, 2017; Lefevre et al., 2012; Robertson et al., 2017). One possible explanation would be that fWHR is suited to explain differences in masculinity among men but not between the sexes. Another suggestion comes from research showing that fWHR measures vary across different scanning procedures (e.g., Lefevre et al., 2012 found a lower fWHR in men compared to women in one of four samples, but only for measures from 3D facial scans and not from 2D facial photos). The fact that we used 3D face scans may partly explain this unexpected finding. Furthermore, it has been suggested that sexual dimorphism might be hard to detect on the facial surface due to greater facial adiposity in women, possibly obscuring any differences in the underlying bone structure (Kramer et al., 2012; Lefevre et al., 2012). Since results are mixed and explanations rather speculative, it appears that a sexual dimorphism in fWHR is in need of further investigation. Nevertheless, global facial masculinity was higher for men compared to women, indicating this measure may be better suited to describe between-sex differences in masculine facial characteristics than fWHR, although we only used three out of five components of the original index (Penton-Voak et al., 2001). Despite the broad use in the literature, measurements of facial masculinity based on just a few facial landmarks might not be the most valid method. Instead, approaches like geometric morphometrics might be superior in detecting masculine (sexually dimorphic) face shapes because they reflect biological factors and shape differences more accurately (Windhager, Schaefer, Fink, 2011; Mitteroecker, Windhager, Müller, & Schäfer, 2015).

Conclusion

Overall, our findings provide no evidence for a link of fWHR and global facial masculinity with salivary baseline T in men or women, T reactivity in men, or hair T in women. Hence, hypotheses that T levels

are an underlying mechanism for the development of and/or sexual dimorphism in fWHR or global facial masculinity are not supported. Finally, our study does not support the proposal that fWHR is larger in men. Rather, in our samples, women had a higher fWHR than men. Thus, future studies should take into account that neither fWHR nor the global masculinity index reflect current T levels, rendering it unlikely that T underlies their relationship with dominance-signalling, and that global facial masculinity is better suited than fWHR to capture objective structural sexual dimorphism in the face.

References

Bird, B., Cid Jofré, V., Geniole, S., Welker, K., Zilioli, S., Maestripieri, D., … & Carré, J. (2016). Does the facial

width-to-height ratio map onto variability in men's testosterone concentrations?. *Evolution and*

Human Behavior, *37*, 392–398. https://doi.org/10.1016/j.evolhumbehav.2016.03.004

Carré, J. M., Iselin, A. M. R., Welker, K. M., Hariri, A. R., & Dodge, K. A. (2014). Testosterone reactivity to

provocation mediates the effect of early intervention on aggressive behavior. *Psychological*

Science, 25, 1140-1146. https://doi.org/10.1177/0956797614525642

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 of the Royal Society of*

London B: Biological Sciences, *275*, 2651–2656. https://doi.org/10.1098/rspb.2008.0873

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159. http://doi.org/10.1037/0033-

2909.112.1.155

Eisenbruch, A. B., Grillot, R. L., Maestripieri, D., & Roney, J. R. (2016). Evidence of partner choice heuristics

in a one-shot bargaining game. *Evolution and Human Behavior, 37*, 429-439.

https://doi.org/10.1016/j.evolhumbehav.2016.04.002

Eisenbruch, A. B., Lukaszewski, A. W., Simmons, Z. L., Arai, S., & Roney, J. R. (2017). Why the Wide Face?

Androgen Receptor Gene Polymorphism does not Predict Men's Facial Width-to-Height Ratio.

Adaptive Human Behavior and Physiology. https://doi.org/10.1007/s40750-017-0084-x

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis

program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–

191. https://doi.org/10.3758/BF03193146

Fiers, T., Delanghe, J., T'Sjoen, G., Van Caenegem, E., Wierckx, K., & Kaufman, J. M. (2014). A critical

evaluation of salivary testosterone as a method for the assessment of serum testosterone.

Steroids, *86*, 5–9. https://doi.org/10.1016/j.steroids.2014.04.013

Gao, W., Stalder, T., Foley, P., Rauh, M., Deng, H., & Kirschbaum, C. (2013). Quantitative analysis of steroid hormones in human hair using a column-switching LC–APCI–MS/MS assay. *Journal of Chromatography B*, *928*, 1–8. https://doi.org/10.1016/j.jchromb.2013.03.008

Gao, W., Stalder, T., & Kirschbaum, C. (2015). Quantitative analysis of estradiol and six other steroid hormones in human saliva using a high throughput liquid chromatography–tandem mass spectrometry assay. *Talanta, 143,* 353–358. https://doi.org/10.1016/j.talanta.2015.05.004

Geniole, S. N., Busseri, M. A., & McCormick, C. M. (2013). Testosterone dynamics and psychopathic personality traits independently predict antagonistic behavior towards the perceived loser of a competitive interaction. *Hormones and Behavior, 64,* 790-798. http://doi.org/10.1016/j.yhbeh.2013.09.005

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

Granger, D., Shirtcliff, E. A., Booth, A., Kivlighan, K. T., & Schwartz, E. B. (2004). The "trouble" with salivary testosterone. *Psychoneuroendocrinology*, *29*, 1229–1240. http://doi.org/10.1016/j.psyneuen.2004.02.005

Grotzinger, A. D., Mann, F. D., Patterson, M. W., Tackett, J. L., Tucker-Drob, E. M., & Harden, K. P. (2018). Hair and salivary testosterone, hair cortisol, and externalizing behaviors in adolescents. *Psychological Science, 29*, 688-699. https://doi.org/10.1177/0956797617742981

Hahn, T., Winter, N. R., Anderl, C., Notebaert, K., Wuttke, A. M., Clément, C. C., & Windmann, S. (2017). Facial width-to-height ratio differs by social rank across organizations, countries, and value systems. *PloS One*, *12*, e0187957. https://doi.org/10.1371/journal.pone.0187957

Haselhuhn, M. P., Ormiston, M. E., & Wong, E. M. (2015). Men's facial width-to-height ratio predicts aggression: A meta-analysis. *PLoS One, 10*, e0122637. doi:10.1371/journal.pone.0122637

Haselhuhn, M. P., Wong, E. M., & Ormiston, M. E. (2013). Self-fulfilling prophecies as a link between men's facial width-to-height ratio and behavior. *PloS One, 8*, e72259.

https://doi.org/10.1371/journal.pone.0072259

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

Idris, F. P., Wan, Y., Zhang, X., & Punyadeera, C. (2017). Within-day baseline variation in salivary biomarkers in healthy men. *Omics: A Journal of Integrative Biology, 21*, 74-80. https://doi.org/10.1089/omi.2016.0168Jünger, 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*, 412–423 https://doi.org/10.1016/j.evolhumbehav.2018.03.007

Kordsmeyer, T., & 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

Kramer, R. S. (2015). Facial width-to-height ratio in a large sample of Commonwealth Games athletes. *Evolutionary Psychology*, *13*. https://doi.org/10.1177/147470491501300112

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

Kramer, R. 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*, e42705. https://doi.org/10.1371/journal.pone.0042705

Lakens, D. (2017). Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*, 355-362. https://doi.org/10.1177/1948550617697177

Lefevre, C., Lewis, G., Bates, T., Dzhelyova, M., Coetzee, V., Deary, I., & Perrett, D. (2012). No evidence for sexual dimorphism of facial width-to-height ratio in four large adult samples. *Evolution and Human Behavior, 33*, 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*, 273–279. https://doi.org/10.1016/j.evolhumbehav.2013.03.005

Lefevre, C. E., Wilson, V., Morton, F., Brosnan, S., Paukner, A., & Bates, T. C. (2014). Facial width-to-height ratio relates to alpha status and assertive personality in capuchin monkeys. *PLoS One, 9*, e93369. https://doi.org/10.1371/journal.pone.0093369

Lieberz, K. A., Windmann, S., Geniole, S. N., McCormick, C. M., Mueller-Engelmann, M., Gruener, F., … & Steil, R. (2017). The facial width-to-height ratio determines interpersonal distance preferences in the observer. *Aggressive Behavior*, *43*, 460–470. https://doi.org/10.1002/ab.21704

Lopez, H.H., Hay, A.C. & Conklin, P.H. (2009). Attractive men induce testosterone and cortisol release in women. *Hormones and Behavior, 56,* 84-92. http://doi.org/10.1016/j.yhbeh.2009.03.004

Mehta, P. H., Welker, K. M., Zilioli, S., & Carré, J. M. (2015). Testosterone and cortisol jointly modulate risk-taking. *Psychoneuroendocrinology*, *56*, 88–99. http://doi.org/10.1016/j.psyneuen.2015.02.023

Mitteroecker, P., Windhager, S., Müller, G. B., & Schaefer, K. (2015). The morphometrics of "masculinity" in human faces. *PloS One*, *10*, e0118374. https://doi.org/10.1371/journal.pone.0118374.

Penton-Voak, I. S., Jones, B. C., Little, A. C., Baker, S., Tiddeman, B., Burt, D. M., & Perrett, D. I. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society of London B: Biological Sciences*, *268*, 1617–1623. http://doi.org/10.1098/rspb.2001.1703.x

Pound, N., Penton-Voak, I. S., & Surridge, A. K. (2009). Testosterone responses to competition in men are related to facial masculinity. *Proceedings of the Royal Society of London B: Biological Sciences*, *276*, 153–159. https://doi.org/10.1098/rspb.2008.0990

R Core Team (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org

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*, 1–10. https://doi.org/10.1177/1474704917730640

Roney, J.R., Mahler, S.V., & Maestripieri, D. (2003). Behavioral and hormonal responses of men to brief interactions with women. *Evolution and Human Behavior, 24,* 365-375. http://doi.org/10.1016/S1090-5138(03)00053-9

Russell, E., Koren, G., Rieder, M., & Van Uum, S. (2012). Hair cortisol as a biological marker of chronic stress: current status, future directions and unanswered questions. *Psychoneuroendocrinology*, *37*, 589-601. https://doi.org/10.1016/j.psyneuen.2011.09.009

Schultheiss, O. C., Schiepe, A., & Rawolle, M. (2012). Hormone assays. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf & K. J. Sher (Eds.), *Handbook of Research Methods in Psychology Vol. 1: Foundations, planning, measures, and psychometrics* (pp. 489-500). Washington DC, US: American Psychological Association. http://dx.doi.org/10.1037/13619-000

Schultheiss, O. C. & Stanton, S. J. (2009). The hormonal correlates of implicit power motivation. *Journal of Research in Personality*, *43*, 942–949. http://doi.org/10.1016/j.jrp.2009.04.001

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

Tiddeman, B., Burt, M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE computer graphics and applications, 21*, 42–50. http://dx.doi.org/10.1109/38.946630

Tiddeman, B. P, Duffy, N, & Rabey, G. (2000). Construction and visualisation of three-dimensional facial statistics. *Computer Methods and Programs in Biomedicine, 63*, 9–20. https://doi.org/10.1016/S0169-2607(00)00072-9

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*, 331–334. https://doi.org/10.1002/ab.21559

Tsepelidis, S., Devreker, F., Demeestere, I., Flahaut, A., Gervy, C., & Englert, Y. (2007). Stable serum levels of anti-Mullerian hormone during the menstrual cycle: A prospective study in normo-ovulatory women. *Human Reproduction*, *22*, 1837–1840. http://doi.org/10.1093/humrep/dem101

Wang, W., Moody, S. N., Kiesner, J., Appiani, A. T., Robertson, O. C., & Shirtcliff, E. A. (2019). Assay Validation of Hair Androgens Across the Menstrual Cycle. *Psychoneuroendocrinology, 101*, 175-181. https://doi.org/10.1016/j.psyneuen.2018.10.029

Welker, K. M., Bird, B. M., & Arnocky, S. (2016). Commentary: Facial width-to-height ratio (fWHR) is not associated with adolescent testosterone levels. *Frontiers in Psychology, 7,* 1745. https://doi.org/10.3389/fpsyg.2016.01745

Weston E. M, Friday A. E, & Liò, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PLoS One, 10*, e710. https://doi.org/10.1371/journal.pone.0000710

Whitehouse, A. J., Gilani, S. Z., Shafait, F., Mian, A., Tan, D. W., Maybery, M. T., & Eastwood, P. (2015). Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood. *Proceedings of the Royal Society of London B: Biological Sciences*, 20151351. https://doi.org/10.1098/rspb.2015.1351

Wilson, J. P., & Rule, N. O. (2015). Facial Trustworthiness Predicts Extreme Criminal-Sentencing Outcomes. *Psychological Science, 26*, 1325-1331. https://doi.org/10.1177/0956797615590992

Windhager, S., Schaefer, K., & Fink, B. (2011). Geometric morphometrics of male facial shape in relation to physical strength and perceived attractiveness, dominance, and masculinity. *American Journal of Human Biology*, *23*, 805–814. https://doi.org/10.1002/ajhb.21219

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

Supplementary

Further information on winsorizing: Outliers were winsorized to 3 SDs before examining violation of normal distributions and subsequent log10-transformation. For men, two outliers were winsorized for baseline testosterone and five for reactive testosterone levels. For women, two outliers were winsorized for salivary baseline testosterone and six for hair testosterone levels.

Table S1. *Bivariate Pearson correlations of the two separate testosterone reactivity indices with facial measures.*

	1st testosterone reactivity		2nd testosterone reactivity	
	r	p	r	p
fWHR	-.02	.85	.03	.75
Global facial masculinity	-.10	.33	-.13	.18
Eye size	.07	.51	.09	.33
Cheekbone prominence	.08	.42	.06	.52
Face width/lower face height	.01	.90	.03	.75
Lower face/face height	-.05	.61	-.10	.29

Note: n = 104. fWHR = facial width-to-height ratio.