Evolution and Human Behavior 34 (2013) 273–279

Contents lists available at SciVerse ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

Original Article

Telling facial metrics: facial width is associated with testosterone levels in men

Carmen E. Lefevre [a,*], Gary J. Lewis [b], David I. Perrett [a], Lars Penke [c]

[a] School of Psychology, University of St Andrews, St Mary's Quad, KY16 9JP St Andrews, UK
[b] Sage Center for the Study of the Mind, Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA 93106-9660, USA
[c] Department of Psychology and Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, 7 George Square, EH8 9JZ, Edinburgh, UK

ARTICLE INFO

Article history:
Initial receipt 1 August 2012
Final revision received 25 March 2013

Keywords:
Testosterone
Face-width
Face width to height ratio
Masculinity
Sexual dimorphism
Speed dating
Face structure

ABSTRACT

High facial width-to-height ratio (fWHR) has been associated with a cluster of behavioural traits in men, including aggression and status-striving. This association between face structure and behaviour may be caused by testosterone. Here we investigated the relationship of both baseline and reactive testosterone levels to fWHR. In addition, we investigated the link between testosterone and three well-characterised sexually dimorphic facial metrics. Testosterone was measured in one sample of males ($n = 185$) before and after a speed-dating event. An additional sample provided only baseline testosterone measures ($n = 92$). fWHR was positively associated with testosterone reactions to potential mate exposure and marginally associated with baseline testosterone in Sample 1. We found a positive association with baseline testosterone and fWHR in Sample 2. In addition, face-width-to-lower-height ratio was positively associated with testosterone in both samples, suggesting that, in particular, facial width (scaled by two measures of facial height) is associated with testosterone. Importantly, our results also indicate that there is no association between adult testosterone and the sexual dimorphism of face shape. Thus, while our findings question the status of sexual dimorphism as a proxy measure of testosterone, they do provide evidence that testosterone is linked to fWHR and might underlie the relationship between fWHR and behaviour.

1. Introduction

Recently, the facial width-to-height ratio (fWHR: bizygomatic width divided by upper-face height; see Fig. 1) has been identified as a facial metric with links to a range of behavioural traits in men. fWHR has been shown to predict aggression (Carré & McCormick, 2008; but see Deaner, Goetz, Shattuck, & Schnotala, 2012; Özener, 2012), deception (Haselhuhn & Wong, 2012), and untrustworthiness (Stirrat & Perrett, 2010), but also more positive behaviours such as achievement striving (Lewis, Lefevre, & Bates, 2012) and self-sacrifice towards the in-group (Stirrat & Perrett, 2012). Moreover, several studies have demonstrated that individuals are able to discern other individuals' tendency to aggression (Carré, McCormick, & Mondloch, 2009) and trustworthiness (Stirrat & Perrett, 2010) from fWHR. While this research indicates that the face and, in particular, fWHR act as a cue to behaviour, it is currently unclear what constitutes the physiological mechanisms underlying this anatomy–behaviour association. Here, in two samples, we examine whether levels of baseline and reactive testosterone are associated with fWHR. We also investigate the association of testosterone with three well-characterised sexually dimorphic facial metrics; namely, 1) lower-face to whole-face-height, 2) cheekbone prominence, and 3) face-width to lower-face-height (see Fig. 1 and Lefevre et al., 2012; Penton-Voak et al., 2001), alongside a global morphometric measure of facial masculinity (Pound, Penton-Voak, & Surridge, 2009).

1.1. Face structure and testosterone levels

Recent work reported sexual dimorphism in fWHR, with males showing higher fWHR than females (Carré & McCormick, 2008; Weston, Friday, & Lio, 2007). While this claim of sexual dimorphism in fWHR has since been challenged (Lefevre et al., 2012; Özener, 2012; Stirrat, Stulp, & Pollet, 2012), several of the behaviours linked to fWHR (e.g. aggression, dominance) are also known to be sexually dimorphic. For example, males show consistently higher levels of direct physical aggression than females (Archer, 2006). Testosterone levels, which themselves show large sex differences (e.g. Mazur, Susman, & Edelbrock, 1997), have been proposed as a common underlying factor linking fWHR to behaviour (e.g. Carré & McCormick, 2008).

Consistent with the suggestion that testosterone mediates the association between fWHR and behaviour, a number of behavioural characteristics similar to those associated with fWHR have been linked to both direct and indirect measures of testosterone. For example, basal circulating testosterone levels are associated with levels of dominance in men (e.g. Josephs, Newman, Brown, & Beer,

* Corresponding author. Tel.: +44 01334 463044.
E-mail address: cel37@st-andrews.ac.uk (C.E. Lefevre).

1090-5138/$ – see front matter
http://dx.doi.org/10.1016/j.evolhumbehav.2013.03.005

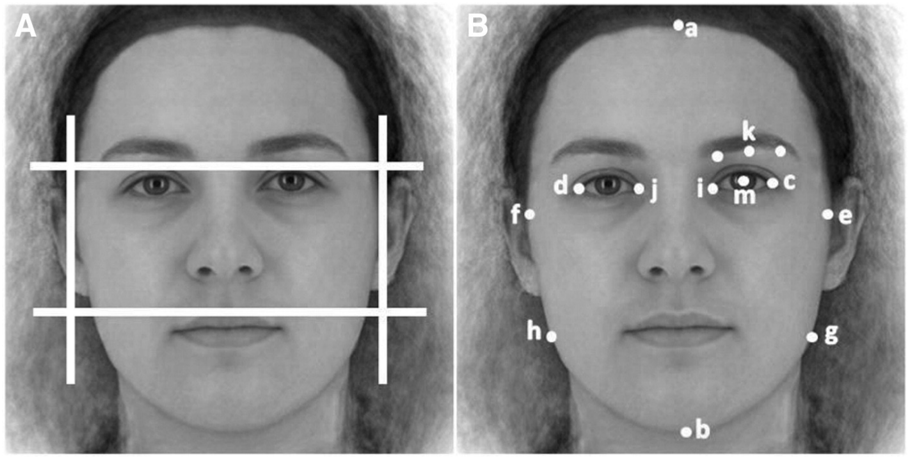

Fig. 1. Examples of measures and measuring points used for morphometric calculations. Panel (A): Measure of fWHR: horizontal lines represent the distance between the upper lip and highest point of the eyelids (upper face height), vertical lines represent the maximum distance between the left and right facial boundary (bizygomatic width). fWHR was calculated as width divided by height. Panel (B): Morphometric calculations: (i) Lower-face/whole-face-height: c–b/a–b, (ii) cheekbone prominence: e–f/g–h, and (iii) face-width/lower-face-height: e–f/c–b. Additional measures used to calculate the masculinity index: (i) eyewidth: c–d/i–j and (ii) mean eyebrow height: mean k–m/a–b.

2003; Mazur & Booth, 1998). Additionally, second-to-forth digit ratio (2D:4D) – a putative proxy measure of in-utero testosterone exposure (Manning, Scutt, Wilson, & Lewis-Jones, 1998, Manning et al., 2000; Williams, Greenhalgh, & Manning, 2003) – has been linked to levels of both self-reported (Bailey & Hurd, 2005) and lab-induced aggression (Millet & Dewitte, 2007).

Work seeking to associate testosterone with facial shape has also provided some support for a potential link with fWHR. In an early study, Verdonck, Gaethofs, Carels, and de Zegher (1999) observed that testosterone administration enhanced craniofacial growth in delayed puberty male adolescents. In particular, these authors report enhanced growth of upper and total face height, and the mandible (jawbone) and ramus (upper part of the jawbone) length. However, these results should be treated with some caution as the sample size was small ($N = 7$) and the boys assessed were all showing delayed puberty, making their development likely not fully comparable to healthy controls.

In addition, several studies have linked testosterone levels to perceived facial masculinity in men. Roney, Hanson, Durante, and Maestripieri (2006) report a moderate correlation ($r = .34$) between ratings of masculinity in natural, unmodified faces and baseline testosterone levels. Similarly, using a forced-choice paradigm, Penton-Voak and Chen (2004) report a weak, but significant, association between testosterone levels and masculine appearance. In pairs of either natural or composite faces the face higher in testosterone was chosen as more masculine 53% and 57% of the time respectively. The authors argue that only men with very high or very low levels of testosterone may be visually distinguishable in terms of their masculinity.

However, other studies find no links between testosterone and masculinity. A study using almost identical methods to Roney et al. (2006), but with a much larger set of men, found no association between perceived facial masculinity and testosterone levels (Peters, Simmons, & Rhodes, 2008). Similarly, Neave, Laing, Fink, and Manning (2003) reported links of perceived facial masculinity with second-to-forth digit ratio (2D:4D), but not with measured baseline testosterone levels; and Ferdenzi, Lemaître, Leongómez, and Roberts (2011) found no association between perceived facial masculinity and 2D:4D ratio.

Fink et al. (2005) incorporated a formal measure of global facial shape, calculating the difference between the average face-shape of men with high and low 2D:4D ratio. This study found that men with a low digit ratio (associated with higher prenatal testosterone) exhibited a wider jaw and zygomatic arch (i.e. face width) as compared to those with a higher digit ratio. However, direct reports of an association between circulating testosterone levels and facial metrics are (to our knowledge) limited to a single study (Pound et al., 2009). In this study, the authors computed a global facial masculinity index using five empirically derived, sexually dimorphic facial metrics (see Penton-Voak et al., 2001), and measured both baseline and reactive levels of testosterone. Reactive testosterone measures were taken after participants won in a manipulated competitive betting task, a scenario previously shown to significantly elevate testosterone levels (e.g. Archer, 2006). While failing to replicate associations between testosterone measures and perceived facial masculinity, Pound et al. (2009) observed that reactive, but not baseline, testosterone was positively associated with a global facial masculinity measure. It is noteworthy that the association between baseline testosterone and facial masculinity also showed a positive trend towards statistical significance in this study, suggesting that the relationship between face shape and baseline testosterone might not be completely absent, but instead may reflect a weaker link than the association with reactive testosterone.

The above findings reported by Pound et al. (2009) are in line with the “challenge hypothesis” (Archer, 2006; Wingfield, Hegner, Dufty, & Ball, 1990), which states that testosterone rises in challenging situations within the mating context, serving as a sexually selected physiological mechanism that calibrates the optimal effort put into intra-sexual competition. Moreover, only the increase in testosterone, rather than baseline testosterone per se, is suggested to associate with aggression and other status-related traits. Accordingly, reactive testosterone may serve as a better predictor of behaviour as compared to baseline levels of testosterone.

1.2. The current study

Determining whether an association exists between facial structure and testosterone is important in order to advance understanding of the underlying basis for links between behaviour and facial characteristics. Particularly, associations between testosterone and fWHR should be examined in light of recent work indicating that fWHR is not sexually dimorphic (Lefevre et al., 2012; Özener, 2012; Stirrat et al., 2012). Moreover, in line with the challenge hypothesis and the work described above, facial metrics associated with aggression and status (e.g. fWHR) may be more closely linked to testosterone reactivity in response to competitive mating opportunities compared to baseline testosterone measures (Ronay & von Hippel, 2010; Roney, Lukaszewski, & Simmons, 2007; Roney, Mahler, & Maestripieri, 2003). Accordingly, here we sought to establish whether fWHR, alongside three sexually dimorphic facial metrics (Lefevre et al., 2012; Penton-Voak et al., 2001) and a global measure of facial masculinity (Pound et al., 2009), predicts baseline testosterone (samples 1 and 2) and testosterone reactivity in response to competitive exposure to potential mates in a speed-dating context in (sample 1).

2. Method

2.1. Participants

2.1.1. Sample 1

Facial metrics and testosterone measures from 188 Caucasian men (mean age = 33.6 years, SD = 7.5, range: 20 to 54 years) who participated in the Berlin Speed Dating Study (Asendorpf, Penke, & Back, 2011) were analysed. All participants were singles drawn from the general population whose motivation to participate was the chance to find a real-life romantic or sexual partner.

2.1.2. Sample 2

This sample consisted of 79 Caucasian male undergraduate students (mean age: 20.50 years, age range: 18–25 years).

2.2. Speed-dating procedure in Sample 1

Participants took part in one of 17 speed-dating events during which each man met between 8 and 14 women (Mean ± SE = 11.4 ± 1.7) of an age similar to their own (within-session age range = ± 4.8 years) for 3 min each. The 'dates' took place in booths equipped with two opposing chairs, cameras, and microphones. Men and women were led to these booths independently to minimise prior contact between the sexes. Women stayed in their allocated booth, while men rotated until each woman had interacted with each man. For details see Asendorpf et al., 2011, and Back et al., 2011b. Body height (m) and weight (kg: dressed, but without jackets or shoes) were measured directly before speed-dating began, from which the body mass index (BMI; kg/m^2) was calculated.

2.3. Facial measures

2.3.1. Sample 1

Participants were recorded with a camcorder on a tripod while standing upright in front of a white background under standardised lighting in order to allow the extraction of various standardised facial photographs. The frame with the most frontal and neutral recording of each participant's face was converted into a picture. Prior to measurement, all pictures were horizontally aligned and scaled to the same inter-pupillary distance (using the Psychomorph software package; Tiddeman, Burt, & Perrett, 2001; http://users.aber.ac.uk/bpt/jpsychomorph). fWHR was measured by calculating the bizygomatic width (maximum horizontal distance from the left facial boundary to the right facial boundary) to upper-face height (vertical distance from the mid-point of the upper-lip to the highest point of the eyelids) ratio from pictures (see Fig. 1A, and Stirrat & Perrett, 2010). We also calculated facial metrics as described by Penton-Voak et al. (2001; see Fig. 1B); namely, (i) lower-face/whole-face-height (vertical distance from mean eye height to gnathion/ vertical distance from trichion to gnathion), (ii) cheekbone prominence (bizygomatic width/horizontal distance between left and right gonion approximation), and (iii) face-width/lower-face-height (bizygomatic width/vertical distance from mean eye height to gnathion). Intercorrelations between facial metrics are reported in Table 1.

Additionally a global facial masculinity index, as previously calculated by Pound et al. (2009), was computed following their exact methodology: Five facial metrics were measured (see Fig. 1), z-transformed and aligned such that a positive value referred to a more masculine score. These measures were then summed.

Table 1
Intercorrelations of facial metrics in Sample 1 (above diagonal) and 2 (below diagonal).

	fWHR	Ch. P.	fW/lFH	lFH/wFH	GM
fWHR	/	−.06	.76**	−.16*	−.32**
Cheekbone prominence (ChP)	−.04	/	.10	−.03	−.52**
Face-width/lower-face-height (fW/lFH)	.65**	.18	/	−.24**	−.61**
Lower-face/whole-face-height (lFH/wFH)	−.14	−.07	−.43**	/	.45**
Global masculinity (GM)	−.25*	−.45**	−.57**	.60**	/

fWHR = facial width-to-height ratio, T = testosterone.
* $p < .05$.
** $p < .01$.

2.3.2. Sample 2

Photographs were taken under standardised lighting conditions during one of the testing sessions. Photographs in this study were taken specifically for morphometric analyses and were thus highly controlled for posture, expression and camera angle. All face metrics, as well as the global facial masculinity, were computed as described for sample 1.

2.4. Testosterone measures

2.4.1. Sample 1

Saliva samples were collected both directly (within 5 min) before and after the speed-dating. The speed-dating events lasted approximately 2 h. The samples were taken by a male research assistant in a waiting room with all male participants of the event present. Male and female participants had been guided to entrances on different sides of a large university building and to separate waiting rooms, so there had been no interaction between sexes at time of the baseline saliva collection. Samples were collected by passive drool using a Salicap tube (www.ibl-international.com) and a straw. In order to control for potential diurnal effects samples were always collected between 3:30 pm and 4:30 pm (baseline) and 5 pm and 7 pm (post-exposure). All samples were visually inspected for blood contamination at time of collection; contaminated samples were excluded from further analysis. To avoid food contamination, participants were asked not to eat and to only drink water during the event.

Immediately following the end of the speed-dating event, specimens were frozen and stored at −20 °C until analysis was performed by the Biopsychological Lab of the Technical University Dresden using IBL luminescence immunoassays (sensitivity 5 pg/ml). After defrosting samples were centrifuged for 10 min and 50 µl of samples was introduced into the respective wells of a microtiter plate. Next 50 µl of enzyme conjugate followed by 50 µl of testosterone antiserum was added to each well. Plates were then incubated for 4 h at room temperature. After discarding incubation solution and washing plates four times using 250 µl of diluted wash buffer, 50 µl Chemiluminescence Reagent AP was added to each well. After 10 min relative luminescence units were measured using a luminometer. Inter- and intra-assay variability was below 12%. For each participant, testosterone levels were analysed in duplicates.

2.4.2. Sample 2

During two testing sessions salivary testosterone was assessed. As in Sample 1, all samples were visually inspected for discoloration indicating blood contamination. Discoloured samples were excluded from analysis. Testosterone levels were calculated as the average of

Table 2
Partial correlations between both baseline testosterone (Pre-T) and post-exposure testosterone (Post-T) with facial structure controlling 1) only age and 2) both age and BMI in Sample 1.

	pre-T (c. Age) df = 178	pre-T (c. Age, BMI) df = 177	post-T (c. Age) df = 178	post-T (c. Age, BMI) df = 177	post-T (c. pre-T & age) df = 177	post-T (c. Pre-T, age, BMI) df = 176
fWHR	.13†	.11	.21**	.19*	.18*	.16*
Cheekbone prominence	.08	.12	.01	.05	−.10	−.03
Face-width/lower-face- height	.14†	.12	.19*	.17*	.18*	.13†
Lower-face/whole-face-height	.04	.04	−.14†	−.14†	−.21**	−.20**
Global masculinity	−.02	−.03	−.11	−.12†	−.12	−.13†

fWHR = facial width-to-height ratio, T = testosterone, c. = control variables.
† $p \leq .10$.
* $p < .05$.
** $p < .01$.

these two samples (for full details see Moore et al., 2011a, 2011b) yielding high reliability for baseline measures.

2.5. Individual differences measures used in Sample 1

In line with Pound et al. (2009) and in order to test for possible moderation effects caused by self-perceived success during the speed dating event (comparable to 'winning' in Pound et al., 2009) three potential moderators were assessed: 1) Self-perceived competitiveness of the speed-dating event as assessed immediately after the event by aggregating two items: "I saw the conversations as competitions to win the women's favour" and "I have experienced the other men as competitors", both on 5-point Likert scales from "strongly disagree" to "strongly agree". The two items were correlated ($r = .43, p < .001$); 2) Self-perceived mating success during the speed dating, measured as number of times each man predicted that he was chosen by a woman directly after each speed-date (perceived mating success). We used predicted female choice instead of actual female choices because the two variables were not correlated ($r = .12$, $p = .12$; see Back et al., 2011a for details) and thus only perceived mating success should be expected to have an effect on testosterone levels; 3) Average flirting behaviour of all female dating partners a man encountered, which was reliably rated for every 30 s interval of the 3-min speed-dates from video recordings showing the woman only by two independent raters (see Back et al., 2011b for details). The average overall-flirting rating for all women each man encountered can be interpreted as a measure of exposure to female courtship behaviour.

3. Results

3.1. Sample 1

3.1.1. Facial metrics and testosterone

Testosterone levels more than three standard deviations from the mean were removed prior to analyses, since outliers of such magnitude are likely caused by imperfect testosterone sampling (e.g. contamination). This reduced the dataset by 3, and 4 participants for baseline and post-exposure testosterone levels, respectively. Furthermore, all testosterone variables were positively skewed and were therefore transformed to normal distribution using square-root transforms.[1] Untransformed means for baseline and post-exposure testosterone were 88.9 (SD = 44.3) and 88.6 (SD = 45.2) pg/ml, respectively.

Consistent with previous work (e.g. Harman, Metter, Tobin, Pearson, & Blackman, 2001), age was negatively associated with baseline ($r = -.31$, $p < .001$) and post-exposure ($r = -.25$, $p < .001$) testosterone levels. Thus, all associations between testosterone and facial measures were age-controlled. Baseline- and post-exposure testosterone levels were significantly correlated ($r = .57, p < .001$). Associations between facial metrics and baseline testosterone are displayed in Table 2. fWHR was associated with testosterone as predicted. Results for lower-face/whole-face-height and face-width/lower-face-height were unexpected because they suggest that more female-typical scores on these metrics were associated with greater testosterone levels.

Although there were no mean differences between baseline- and post-exposure testosterone ($t_{182} = 0.55$, $p = .59$, $d = 0.08$), we further assessed whether facial structure was linked to testosterone change during the speed-dating event. To this end, we examined whether post-exposure testosterone correlated with the assessed facial metrics, controlling for baseline testosterone levels using partial correlations. There was a positive association of testosterone change with fWHR and face-width/lower-face-height, as well as a negative association with lower-face/whole-face-height (see Table 2). As an additional illustration, Fig. 2 details mean baseline and post-exposure testosterone across quartiles for fWHR. Only men with high fWHR show an overall tendency to increase in testosterone during the speed-dating event, although this group difference was non-significant ($p = .66$).

Because BMI has been shown to associate both with face metrics (Coetzee, Chen, Perrett, & Stephen, 2010; Lefevre et al., 2012) and testosterone levels (Osuna, Gómez-Pérez, Arata-Bellabarba, & Villaroel, 2006), we next examined whether testosterone showed independent effects on facial structure controlling for BMI (in addition to age and baseline testosterone). Links between post-exposure testosterone and both fWHR and lower-face/whole-face-height were robust to this additional control (see Table 2). Facial appearance associated with high and low levels of testosterone, respectively, is shown in Fig. 3. As an illustration we created composite faces of 20 individuals with high and low testosterone levels in each sample. In order to visualise shape associations with testosterone over and above those associated with BMI, the two groups of individuals included in the composite faces were matched for average BMI. These images indicate, in line with results, that high testosterone is associated with wider and shorter faces.

We next assessed possible moderation of the relationship between reactive testosterone and fWHR by self-reported competitiveness, self-perceived success (the expected number of follow-up dates) during the speed dating, as well as average rated flirting behaviour of all female dates men encountered. Regression models with post-exposure testosterone as the dependent variable and fWHR, baseline testosterone, the moderator, and the moderator*fWHR interaction as predictors were run. None of these moderations showed a significant effect (all $p > .28$).

3.1.2. Global facial masculinity and testosterone

There was no association between the global facial masculinity index and either baseline ($r = .02$, $p = .77$) or post-exposure

[1] Results including outliers showed significant links between facial metrics and post-exposure testosterone (post T); however, when including outliers associations between fWHR and baseline testosterone with age and BMI as covariates showed non-significant results, albeit in the same direction as reported in the text (rs .11–.12, ps .11–.14).

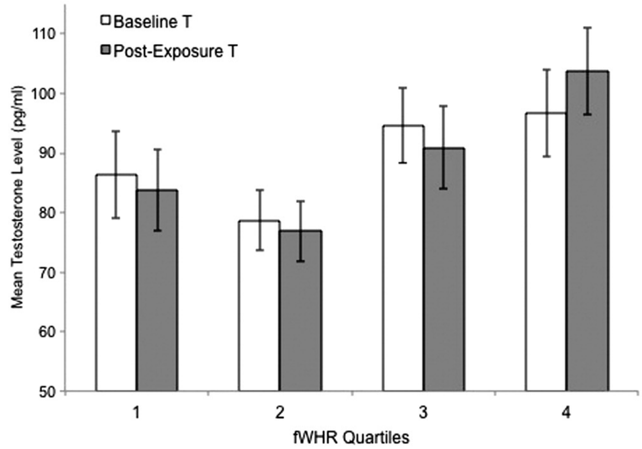

Fig. 2. Mean baseline and post-exposure testosterone levels across fWHR quartiles (1 = lowest quartile, 4 = highest quartile).

testosterone ($r = -.08$, $p = .26$). Furthermore, there was no significant association between global facial masculinity and testosterone change ($r = -.11$, $p = .13$), or testosterone change controlling age and BMI ($r = -.13$, $p = .08$). Additionally, global facial masculinity showed a negative association with age ($r = -.15$, $p = .04$).

Finally, associations between global facial masculinity and facial metrics were assessed: there were a negative correlation with fWHR ($r = -.32$, $p < .001$), face-width/lower-face-height ($r = -.61$, $p < .001$) and cheekbone prominence ($r = -.52$, $p < .001$) and positive correlation with lower-face/whole-face-height ($r = .45$, $p < .001$). Apart from fWHR these metrics form part of the basis for the masculinity index, thus there is an intrinsic interdependence between these metrics and the index.

3.2. Sample 2

Two fWHR data points more than two standard deviations from the mean were removed, leaving 90 cases. All other facial measures

Fig. 3. High and low testosterone averages of each sample. Averages of the twenty men with lowest (left) and highest (right) testosterone levels in Sample 1 (top) and Sample 2 (bottom).

were normally distributed with no outliers. There was no association of testosterone with either age ($r = -.06$, $p = .60$) or BMI ($r = -.05$, $p = .63$). However, since BMI was highly correlated with fWHR ($r = .31$, $p = .005$) and cheekbone prominence ($r = -.26$, $p = .02$), BMI was controlled in subsequent analyses.

Partial correlations revealed positive associations between testosterone and fWHR ($r = .26$, $p = .03$), as well as face-width/lower-face-height ($r = .28$, $p = .02$). In addition, cheekbone prominence was (marginally) negatively correlated with testosterone ($r = -.21$, $p = .07$). There was no association with lower-face/whole-face-height ($r = -.08$, $p = .59$). Finally, there was again no association between testosterone and global facial masculinity ($r = .08$, $p = .49$).

4. Discussion

Evidence from two samples supports a link between circulating testosterone levels and behaviourally-relevant facial structures in adult men; namely, facial width. Of particular interest, facial width-to-height ratio (fWHR) showed positive marginal associations with baseline testosterone in sample 1 and positive significant associations baseline testosterone in sample 2. Moreover, in sample 1 fWHR was positively associated with testosterone reactivity following exposure to potential mates. These latter findings are consistent with predictions arising from the “challenge hypothesis” (Archer, 2006; Wingfield et al., 1990).

Additionally, our results showed a positive association of reactive (Sample 1) and baseline (Sample 1 and 2) testosterone with face-width/lower-face-height, and a negative association of reactive testosterone with lower-face/whole-face-height (Sample 1). These results run contrary to previous assumptions in the literature: More “feminine” scores on these two facial measures were associated with *higher* (rather than lower) levels of testosterone. These results indicate that faces of men with high testosterone levels were relatively shorter and wider than faces of men with lower levels of testosterone. While a limited number of studies have previously associated testosterone with facial elongation (Verdonck et al., 1999) and both measured and perceived masculinity (Penton-Voak & Chen, 2004; Pound et al., 2009), other studies have failed to do so (e.g. Neave et al., 2003). As noted above, we confirmed this association for face-width/lower-face-height across both samples suggesting this finding is robust.

It should be borne in mind that while we found associations between testosterone and face-width/lower-face-height as well as lower-face/whole-face-height (albeit in the “feminine” direction), globally measured facial masculinity was not associated with testosterone in either of our samples. This discrepancy might, in part, be explained by the difference in context between Pound et al.'s (2009) study and the current study. While in Pound's study men's testosterone rose in response to winning in a competitive betting task, here testosterone levels changed in response to meeting potential mates. In addition, we had no means to assess ‘winning’ or perceived ‘winning’ in the participants of the speed-dating study. Being able to discriminate ‘winners’ and ‘losers’ might have yielded stronger associations with post-exposure testosterone than those reported here. However, it is unlikely that these responses would have changed in direction. Taken together, these results suggest that rather than facial sexual dimorphism (i.e. masculinity or femininity) showing links to testosterone, instead only more specific facial metrics are associated with testosterone.

The finding that fWHR, which itself does not appear to be sexually dimorphic (Lefevre et al., 2012; Özener, 2012), is linked to testosterone, is somewhat counterintuitive and thus merits discussion. Moreover, the observation that face-width/lower-face-height is associated with testosterone in the more “feminine” direction also requires further consideration. There are a number of possible explanations for these results. Firstly, while recent findings show no sex difference in fWHR (Lefevre et al., 2012; Özener, 2012), there are known sex differences in facial adiposity: women have larger fat deposits than men, particularly around the cheek-area, even when controlling for BMI (Enlow, 1982). As such, sexual dimorphism in facial bone structure may be “masked” (e.g. fWHR), or even reversed (e.g. face-width/lower-face-height), by this additional facial adiposity in women. However, this explanation does not account for the lack of sexual dimorphism found in fWHR as measured on skulls that was recently reported by Stirrat et al. (2012). Further work, then, establishing the status concerning the sexual dimorphism of these facial metrics, particularly from skull measures, will be valuable.

An alternative, and perhaps more compelling, explanation stems from possible sex differences in the underlying mechanisms influencing facial bone growth. While there is some direct evidence for testosterone affecting facial bone growth in men (Verdonck et al., 1999), other research has highlighted the importance of oestrogen and growth hormone on bone growth in both sexes (Juul, 2001; Ohlsson, Bengtsson, Isaksson, Andreassen, & Slootweg, 1998). Importantly, women, compared to men, have higher average levels of both oestrogen and growth hormone (e.g. Frantz & Rabkin, 1965; Mazur et al., 1997), indicating a stronger influence of these hormones on bone growth in females. In short, because women have more subcutaneous facial fat and higher levels of oestrogen and growth hormone, facial morphology in men and women likely reflects different growth and endocrine mechanisms and is thus not easily comparable.

Furthermore, considering men typically exhibit around five times higher baseline testosterone concentrations compared to women (Mazur et al., 1997), identical effects of testosterone on bone growth in both sexes would likely yield far more striking structural differences between men and women than those that are observed. Following this, differences in testosterone may be directly related to facial bone size (and particularly facial width) within men, while distinct factors may influence these facial metrics in women; that is, factors underlying inter-sex differences in facial metrics may not reflect intra-sex differences. In turn, labelling the poles of facial metrics measured *within*-sex on the basis of *between*-sex differences (i.e. “more masculine” vs. “more feminine”) may be erroneous. This interpretation may account for why more “feminine” face-width/lower-face-height scores associate with higher testosterone: Simply put, influences underpinning intra-sex differences may not be isomorphic with the influences driving inter-sex differences.

As suggested above, differential effects of testosterone on bone growth may be unique to men. Individual differences in facial physiognomy linked to testosterone within males may then reflect variance in total testosterone exposure during (pubertal) development. In line with this argument, our results indicate stronger links between facial structure and reactive testosterone compared to baseline testosterone measures. That is, if facial bone structure is affected by testosterone levels during puberty (see also Verdonck et al., 1999) it is possible that total tissue exposure to the hormone is more closely aligned with appearance outcomes than baseline measures (see Pound et al., 2009 for a similar argument). There is convincing literature indicating a rise in testosterone levels following perceived success in a competitive situation (‘winning’; e.g. Booth, Shelley, Mazur, Tharp, & Kittok, 1989; Mazur et al., 1997) and in exposure to potential mates (e.g. Ronay & von Hippel, 2010; Roney et al., 2003). Two consequences follow: firstly, men who exhibit a stronger hormonal reaction to positive outcomes of competitive or mating situations will be exposed to more testosterone compared to those men who have a lesser reaction. Secondly, irrespective of individual differences in hormone reactivity, those men that experience winning more competitive situations will have higher total exposure.

Finally, fWHR may pose a better proxy measure of testosterone levels compared to other facial metrics or assessments derived from sexual dimorphism: fWHR has been repeatedly associated with dominant behaviours in males and as such reflects individual differences in male behaviour. Measures that are based on sexual dimorphism, such as other measures assessed here and elsewhere, may not capture intra-male variation in testosterone, if indeed testosterone's influence on growth is male specific.

In summary, the current studies provide the first empirical evidence for an association between facial width and testosterone levels. Moreover, the results show that global facial masculinity does not associate with testosterone thus contradicting previous work. These findings provide support for testosterone's role as a physiological link between face structure and behavioural traits that are linked to testosterone and further add to the understanding of mechanisms underlying facial cues to behaviour.

Supplementary Materials

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.evolhumbehav.2013.03.005.

References

Archer, J. (2006). Testosterone and human aggression: An evaluation of the challenge hypothesis. *Neuroscience and Biobehavioral Reviews*, *30*, 319–345.

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*, 16–30.

Back, M. D., Penke, L., Schmukle, S. C., & Asendorpf, J. B. (2011a). Knowing your own mate value: Sex-specific personality effects on the accuracy of expected mate choices. *Psychological Science*, *22*, 984–989.

Back, M. D., Penke, L., Schmukle, S. C., Sachse, K., Borkenau, P., & Asendorpf, J. B. (2011b). Why mate choices are not as reciprocal as we assume: The role of personality, flirting, and physical attractiveness. *European Journal of Personality*, *25*, 120–132.

Bailey, A. A., & Hurd, P. L. (2005). Finger length ratio (2D:4D) correlates with physical aggression in men but not in women. *Biological Psychology*, *68*, 215–222.

Booth, A., Shelley, G., Mazur, A., Tharp, G., & Kittok, R. (1989). Testosterone, and winning and losing in human competition. *Hormones and Behavior*, *23*, 556–571.

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 B*, *275*, 2651–2656.

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. *Psychological Science*, *20*, 1194–1198.

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

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*, 235–238.

Enlow, D. (1982). *Handbook of facial growth.* Philadelphia: Saunders.

Ferdenzi, C., Lemaître, J., Leongómez, J. D., & Roberts, S. C. (2011). Digit ratio (2D:4D) predicts facial, but not voice or body odour, attractiveness in men. *Proceedings of the Royal Society B*, *278*, 3551–3557.

Fink, B., Grammer, K., Mittroecker, P., Gunz, P., Schaefer, K., Bookstein, F. L., & Manning, J. T. (2005). Second to fourth digit ratio and face shape. *Proceedings of the Royal Society B*, *272*, 1995–2001.

Frantz, A. G., & Rabkin, M. T. (1965). Effects of estrogen and sex difference on secretion of human growth hormone. *The Journal of Clinical Endocrinology and Metabolism*, *25*(11), 1470–1480.

Harman, S. M., Metter, E. J., Tobin, J. D., Pearson, J., & Blackman, M. R. (2001). Longitudinal effects of aging on serum total and free testosterone levels in healthy men. *The Journal of Clinical Endocrinology and Metabolism*, *86*, 724–731.

Haselhuhn, M. P., & Wong, E. M. (2012). Bad to the bone: Facial structure predicts unethical behaviour. *Proceedings of the Royal Society B*, *279*, 571–576.

Josephs, R. A., Newman, M. L., Brown, R. P., & Beer, J. M. (2003). Status, testosterone, and human intellectual performance: Stereotype threat as status concern. *Psychological Science*, *14*, 158–163.

Juul, A. (2001). The effects of oestrogens on linear bone growth. *Human Reproduction Update*, *7*, 303–313.

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*, 623–627.

Lewis, G. J., Lefevre, C. E., & Bates, T. C. (2012). Facial width-to-height ratio predicts achievement drive in US presidents. *Personality and Individual Differences*, *52*, 855–857.

Manning, J. T., Barley, L., Walton, J., Lewis-Jones, D. I., Trivers, R. L., Singh, D., Thornhill, R., Rohde, P., Bereczkei, T., P. Henzi, P., Soler, M., & Szwed, A. (2000). The 2nd:4th digit ratio, sexual dimorphism, population differences, and reproductive success: Evidence for sexually antagonistic genes? *Evolution and Human Behavior*, *21*, 163–183.

Manning, J. T., Scutt, D., Wilson, J., & Lewis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: A predictor of sperm numbers and concentrations of testosterone, luteinising hormone and oestrogen. *Human Reproduction*, *13*, 3000–3004.

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *The Behavioral and Brain Sciences*, *21*, 353–363.

Mazur, A., Susman, E. J., & Edelbrock, S. (1997). Sex difference in testosterone response to a video game contest. *Evolution and Human Behavior*, *18*, 317–326.

Millet, K., & Dewitte, S. (2007). Digit ratio (2D:4D) moderates the impact of an aggressive music video on aggression. *Personality and Individual Differences*, *43*, 289–294.

Moore, F. R., Al Dujaili, E. A. S., Cornwell, R. E., Smith, Law, Lawson, J. F., M.J., Sharp, M., & Perrett, D. I. (2011a). Cues to sex- and stress-hormones in the human male face: Functions of glucocorticoids in the immunocompetence handicap hypothesis. *Hormones and Behavior*, *60*, 269–274.

Moore, F. R., Cornwell, R. E., Law Smith, M. J., Al Dujaili, E. A. S., Sharp, M., & Perrett, D. I. (2011b). Evidence for the stress-linked immunocompetence handicap hypothesis in human male faces. *Proceedings of the Royal Society B*, *278*, 774–780.

Neave, N., Laing, S., Fink, B., & Manning, J. T. (2003). Second to fourth digit ratio, testosterone, and perceived male dominance. *Proceedings of the Royal Society B*, *270*, 2167–2172.

Ohlsson, C., Bengtsson, B. A., Isaksson, O. G., Andreassen, T. T., & Slootweg, M. C. (1998). Growth hormone and bone. *Endocrine Reviews*, *19*, 55.

Osuna, J. A., Gómez-Pérez, C. R., Arata-Bellabarba, G., & Villaroel, V. (2006). Relationship between BMI, total testosterone, sex hormone-binding-globulin, leptin, insulin and insulin resistance in obese men. *Systems Biology in Reproductive Medicine*, *52*, 355–361.

Özener, B. (2012). Facial width-to-height ratio in a Turkish population is not sexually dimorphic and is unrelated to aggressive behaviour. *Evolution and Human Behavior*, *33*, 169–173.

Penton-Voak, I. S., & Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution and Human Behavior*, *25*, 229–241.

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 B*, *268*, 1617–1623.

Peters, M., Simmons, L. W., & Rhodes, G. (2008). Testosterone is associated with mating success but not attractiveness or masculinity in human males. *Animal Behaviour*, *76*(2), 297–303.

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 B*, *276*, 153–159.

Ronay, R., & von Hippel, W. (2010). The presence of an attractive woman elevates testosterone and physical risk taking in young men. *Social Psychological and Personality Science*, *1*, 57–64.

Roney, J. R., Hanson, K. H., Durante, K. M., & Maestripieri, D. (2006). Reading men's faces: Women's mate attractiveness judgments track men's testosterone and interest in infants. *Proceedings of the Royal Society B*, *273*, 2169–2175.

Roney, J. R., Lukaszewski, A. W., & Simmons, Z. L. (2007). Rapid endocrine responses of young men to social interactions with young women. *Hormones and Behavior*, *52*, 326–333.

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.

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

Stirrat, M., & Perrett, D. I. (2012). Face structure predicts cooperation: Men with wider faces are more generous to their in-group when out-group competition is salient. *Psychological Science*, *23*, 718–722.

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*, 551–556.

Tiddeman, B., Burt, D. M., & Perrett, D. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications*, *21*, 42–50.

Verdonck, A., Gaethofs, M., Carels, C., & de Zegher, F. (1999). Effect of low-dose testosterone treatment on craniofacial growth in boys with delayed puberty. *European Journal of Orthodontics*, *21*, 137–143.

Weston, E. M., Friday, A. E., & Lio, P. (2007). Biometric evidence that sexual selection has shaped the hominin face. *PLoS One*, *2*, e710.

Williams, J. H. J., Greenhalgh, K. D., & Manning, J. T. (2003). Second to fourth finger ratio and possible precursors of developmental psychopathology in preschool children. *Early Human Development*, *72*, 57–65.

Wingfield, J. C., Hegner, R. E., Dufty, A. M., Jr., & Ball, G. F. (1990). The 'challenge hypothesis': Theoretical implications for patterns of testosterone secretion, mating systems, and breeding strategies. *The American Naturalist*, *136*, 829–846.