Evolution and Human Behavior xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.ehbonline.org

Original Article

Does the facial width-to-height ratio map onto variability in men's testosterone concentrations?

Brian M. Bird [a], Valeska S. Cid Jofré [b], Shawn N. Geniole [c], Keith M. Welker [d], Samuele Zilioli [e], Dario Maestripieri [f], Steven Arnocky [g], Justin M. Carré [g,*]

[a] Department of Psychology, Laurentian University, Sudbury, ON, Canada
[b] Department of Ecology, University of Chile, Las Palmeras, Ñuñoa, Santiago, Chile
[c] Department of Psychology, Brock University, St. Catharines, ON, Canada
[d] Department of Psychology, University of Massachusetts Boston, Boston, MA, United States
[e] Department of Psychology, Wayne State University, Detroit, MI, United States
[f] Department of Comparative Human Development, University of Chicago, Chicago, IL, United States
[g] Department of Psychology, Nipissing University, North Bay, ON, Canada

ARTICLE INFO

Article history:
Initial receipt 20 July 2015
7 February 2016
Final revision received 16 March 2016
Available online xxxx

Keywords:
Facial width to height ratio
fWHR
Testosterone
Competition
Neuroendocrine function

ABSTRACT

Variation in the facial width-to-height ratio (fWHR) maps onto a number of behavioral and psychological traits among men (e.g., aggression, unethical behavior, negotiation performance). Importantly, observer judgments of many of these traits also correlate strongly with the fWHR, suggesting that it may represent an honest cue to dominance and status. It has been speculated that the relationship between fWHR and these behavioral traits is due to pubertal testosterone concurrently shaping facial structure and traits linked to social dominance. Others, however, have provided some initial, although inconsistent, evidence that circulating testosterone levels in adulthood may underlie associations between the fWHR and behavioral displays. Here, we provide a more powerful test of the second model by examining the relationship between fWHR, baseline testosterone, and competition-induced testosterone reactivity, across seven diverse samples of men (total $N = 780$). We also report a further analysis including data published previously, for a total sample of 1041 men. Analysis of our individual samples, in addition to an internal meta-analysis, demonstrated no significant positive relationship between fWHR and baseline testosterone, or fWHR and three measures of competition-induced testosterone reactivity. We discuss potential reasons for previous discrepancies, and suggest avenues for future research.

1. Introduction

A growing body of evidence indicates that individual differences in facial morphology map onto a diverse range of behavioral and psychological traits, particularly among men (see Geniole, Denson, Dixson, Carre, & McCormick, 2015; Haselhuhn, Ormiston, & Wong, 2015, for meta-analyses). For instance, the facial width-to-height ratio (fWHR)—the distance of the bizogymatic width divided by the distance between the brow and upper lip—is positively correlated with measures of aggressive behavior (Carré & McCormick, 2008; Goetz et al., 2013; Lefevre et al., 2014; Welker, Goetz, Galicia, Liphardt, & Carré, 2014, but see Gómez-Valdés et al., 2013; Özener, 2012), psychopathic traits (Anderl et al., 2016; Geniole, Molnar, Carré, & McCormick, 2014), achievement drive (Lewis, Lefevre, & Bates, 2012), competitive success (baseball study of homeruns: Tsujimura & Banissy, 2013; formidability as a professional combatant: Trebická et al., 2015; Zilioli et al., 2014) unethical behavior (Geniole, Keyes, Carré, & McCormick, 2014; Haselhuhn & Wong, 2012), explicit prejudice (Hehman, Leitner, Deegan, & Gaertner, 2013), and negotiation performance (Haselhuhn, Wong, Ormiston, Inesi, & Galinsky, 2014). Notably, numerous studies also find that perceiver ratings of aggressiveness and dominance are highly correlated with the fWHR (e.g., Carré, McCormick, & Mondloch, 2009; Carré, Morrissey, Mondloch, & McCormick, 2010; Geniole, Molnar, et al., 2014; Short et al., 2012; see Geniole et al., 2015, for meta-analysis), suggesting that the fWHR may serve as a reliable cue to one's propensity for aggressive behavior.

It has been speculated that the link between facial structure and behavioral/psychological traits is due to the common influence of testosterone (T) on craniofacial growth and the expression of sexually dimorphic behaviors and traits (Carré & McCormick, 2008). Indeed, administration of T to males with delayed puberty modulates various indices of craniofacial growth (Verdonck, Gaethofs, Carels, & de Zegher, 1999). Other studies suggest positive associations between adult T concentrations and perceiver ratings of facial masculinity. In one study, composite images of men with high T (versus composites of those with low T) were rated as more masculine by observers (Penton-Voak & Chen, 2004). Similarly, men's T levels are positively correlated ($r =$

* Corresponding author. 100 College Dr., North Bay, Ontario, P1B 8L7.
E-mail addresses: justinca@nipissingu.ca, jcarre14@gmail.com (J.M. Carré).

http://dx.doi.org/10.1016/j.evolhumbehav.2016.03.004
1090-5138/

.34, $n = 38$) with female ratings of their facial masculinity (Roney, Hanson, Durante, & Maestripieri, 2006). However, more recent work has failed to find relationships between subjective ratings of facial masculinity and individual differences in baseline T concentrations (Peters, Simmons, & Rhodes, 2008).

Other research has examined links between objective measures of facial masculinity and individual differences in T concentrations, but results have been mixed (Apicella et al., 2008, 2011; Campbell et al., 2010; Lefevre, Lewis, Perrett, & Penke, 2013; Pound, Penton-Voak, & Surridge, 2009). For instance, Pound et al. found no relationship between facial masculinity (as measured through a global index of facial-masculinity) and baseline T concentrations in a small sample of young men ($r = 0.19$, $n = 47$). Instead, the authors reported a positive correlation between facial masculinity and T concentrations *after* watching a successful competitive interaction. In another study, Lefevre et al. found no relationship between baseline T and an objective measure of facial masculinity. In contrast, the authors found a small positive correlation between fWHR and baseline T concentrations ($r = .13$, $n = 188$) and between fWHR and T responses to a speed dating interaction (Lefevre et al., 2013). Moreover, Lefevre et al. (2013) found that fWHR was positively correlated with acute changes in T concentrations in response to a speed dating paradigm. Collectively, these studies provide mixed evidence for a link between facial structure and baseline T concentrations, and more consistent support for a relationship between facial structure and context-dependent fluctuations in T concentrations.

1.1. The present study

For the present study, we tested the previously reported positive associations between fWHR and T levels, using data from seven independent samples ($N = 780$). Further, we combined our seven samples with data that was previously published and publically available (Lefevre et al., 2013; http://www.sciencedirect.com/science/article/pii/S1090513813000275) to get the most robust test of a potential relationship ($N = 1041$). This study examined three main relationships of interest: (1) Is fWHR associated with baseline T? (2) Is fWHR associated with T responses to competition? (3) Does competition outcome (win vs. loss) moderate the relationship between fWHR and T reactivity? The latter question was motivated by evidence that demonstrated that changes in T concentrations during competition map onto variability in competitive motivation (Mehta & Josephs, 2006) and aggression (Carré, Putnam & McCormick, 2009) in losers, but not winners. To the extent that links between face structure and human behavior are mediated via neuroendocrine function, we wanted to examine whether the relationship between fWHR and T responses to competition would depend on the outcome of the competitive interaction.

2. Method

2.1. Participant samples

2.1.1. Sample 1

Photographs and T samples from 80 male participants between the ages of 18 and 33 ($M_{AGE} = 21.58$, $SD = 3.19$) were used from a previous study investigating testosterone responses to competition (see Norman, Moreau, Welker, & Carré, 2014, study 1, for full details). The majority of participants were Caucasian (86.1%), followed by Asian (5.1%), bi-racial/other (3.8%), First Nations/Aboriginal (2.5%), Black (1.3%), and Latin American (1.3%). Briefly, participants in the original study completed a video game task, which had been pre-programmed at a low level of difficulty, thus allowing participants to experience a string of victories. Saliva samples were collected pre (i.e., baseline) and post video game task.

2.1.2. Sample 2

Photographs and T samples from 114 male participants between the ages of 17 and 56 ($M_{AGE} = 21.78$, $SD = 5.68$) were examined from a previous study investigating testosterone responses to competition (see Carré, Campbell, Lozoya, Goetz, & Welker, 2013). T samples were available for 111 participants, and thus, the final sample size reflects this number. Participant ethnicities were classified as follows: Caucasian (52.6%), bi-racial/other (19.3%), Asian (18.4%), and Black (9.6%). In this study, participants were randomly and evenly assigned to a victory or defeat condition for a video game task. Saliva samples were collected at pre and post video game competition.

2.1.3. Sample 3

Photographs and T samples from 165 male participants between the ages of 18 and 34 ($M_{AGE} = 20.66$, $SD = 2.97$) were used from a larger protocol investigating hormones and competition (Welker & Carré, 2015). Full data for the variables of interest were available for 152 participants, and thus, this sample size was used for the present analysis. Ethnicities were diverse, with Caucasian (37.6%), Black (20%), Asian (18.2%), Middle Eastern (10.3%), bi-racial/other (8.5%), Latin American (4.8%), and First Nations/Aboriginal (0.6%) participants. Briefly, participants played an XBOX-360 video game randomly assigned to high difficulty (i.e., lose condition) or low difficulty (i.e., win condition), and gave a second saliva sample upon completion. Full details for this video game task are available in Carré et al. (2013).

2.1.4. Sample 4

Photographs and T samples available for 159 male participants between the ages of 24 and 35 ($M_{AGE} = 29.09$, $SD = 2.41$) were used from a previous study investigating the relationship between testosterone concentrations, risk aversion, and career choices among business school students (Sapienza, Zingales, & Maestripieri, 2009; see also Maestripieri, Baran, Sapienza, & Zingales, 2010). Participant ethnicities were varied, with Caucasian (43.4%), Native American (22%), Asian (13.8%), bi-racial/other (8.2%), Black (6.3%), and Latin American (6.3%). Participants in the original study had baseline T samples collected, and then engaged in a series of computerized decision-making tasks (Maestripieri et al., 2010; Sapienza et al., 2009). Since this study did not involve an experimental competition paradigm, only the baseline T concentrations were analyzed from this sample.

2.1.5. Sample 5

Photographs and T samples for 95 male participants between the ages of 18 and 30 ($M_{AGE} = 20.36$, $SD = 2.09$) were used from a study examining the relationship between competition, testosterone, and persistence in men (Welker & Carré, 2015). Ethnicities were classified as follows: Caucasian (46.3%), Black (21.1%), bi-racial/other (15.8%), Middle Eastern (6.3%), Asian (5.3%), Latin American (4.2%), and Native American (1.1%). In the original study, participants were randomly assigned to one of three conditions for a competitive number-tracing task: win against a confederate, lose against a confederate, or complete the task alone (control condition). Pre- and post-competition saliva samples were collected for hormonal assay.

2.1.6. Sample 6

Photographs and T samples were collected from 77 male participants aged 18 to 40 ($M_{AGE} = 21.84$, $SD = 3.56$). Participant ethnicities were classified as follows: Caucasian (75.3%), Middle Eastern (9.1%), Asian (9.1%) Black (3.9%), and Latin American. For this study, two saliva samples were collected in the afternoon between 2:00 pm and 4:00 pm across two consecutive days, and were averaged to create a mean T score.

2.1.7. Sample 7

Photographs and T samples were collected from 120 male participants between the ages of 18 and 35 ($M_{AGE} = 25.27$, $SD = 4.98$) that were part of a larger protocol examining the causal role of T on perception, cognition, and decision-making (Carré et al., unpublished). Briefly, participants reported to the lab and completed a battery of self-report

Please cite this article as: Bird, B.M., et al., Does the facial width-to-height ratio map onto variability in men's testosterone concentrations?, *Evolution and Human Behavior* (2016), http://dx.doi.org/10.1016/j.evolhumbehav.2016.03.004

online questionnaires for approximately 1 hour. Following this, participants had 10 mL of blood extracted by a phlebotomist to be used for assessment of basal T levels. Participant ethnicities were categorized as Caucasian (77.5%), First Nations/Aboriginal (13.3%), Asian (4.2%), biracial/other (2.5%), Latin American (1.7%), and Middle Eastern (0.8%).

2.2. Materials

2.2.1. Testosterone

For six participant samples, saliva was collected via passive drool in polystyrene culture tubes. Samples were stored from −20 to −80 °C until assayed using commercially available enzyme linked immunoassay kits (DRG International). For one sample (sample 7), 10 mL blood samples were drawn by a certified phlebotomist. These samples were allowed to clot, and then centrifuged at 3000 rpm. Serum samples were extracted and stored at −60 °C. All samples were assayed in duplicate, and the average of the duplicates was used for analysis. Average intra- and inter assay coefficients of variation (respectively) for each sample were as follows: sample 1 (5.67%, 11.14%), sample 2 (11.72%, 14.88%), sample 3 (6%, 6%), sample 4 (≤10%, ≤15%), sample 5 (9.19%, 16.59%), sample 6 (6%, 7%), sample 7 (7.38%, 16.03%).

2.2.2. Facial measurements

For each sample, fWHR was calculated from emotionally-neutral photographs of each participant using ImageJ (NIH open-source software) as has been reported in previous studies examining fWHR (e.g., Carré & McCormick, 2008; Geniole & McCormick, 2015). fWHR is determined by measuring the bizygomatic width (distance between right and left zygion) and dividing it by the distance between the lip and brow (height of upper face). Two independent researchers measured the corresponding metrics for each face, and consistency was high between raters for each sample (r-values > .90).

3. Results

Based on Lefevre and colleagues' smallest effect found between fWHR and baseline T ($r = .13$), individual samples in the current study were underpowered to detect the presence of such an effect (power < 0.37; G*Power 3, Faul, Erdfelder, Lang, & Buchner, 2007). Therefore, the most powerful test of the relationship between fWHR and baseline T was gleaned from an internal meta-analysis, allowing the combination of effects from each study in order to reach greater precision for estimation (Cumming, 2013). Power analysis (G*Power 3, Faul et al., 2007) indicated that our combined sample of 780 was sufficiently powered to detect the previously reported effect size at a power level of 0.95. Individuals ($n = 20$) with scores greater than 3 standard deviations on any variable had scores Winsorized prior to analysis.

3.1. Internal meta-analysis

An internal meta-analysis was conducted across all seven samples. To achieve this, fWHR and T-measures were first standardized within their respective samples to reduce any potential influence of hormonal assay or facial measurement variation across samples. Bivariate correlations were computed to test the association between the fWHR and baseline T levels. In addition, we also performed partial correlations in which we controlled for participant ethnicity. However, the results from the partial correlations were nearly identical to those obtained using bivariate correlations, and thus, we report the results from the more simple bivariate correlations.

3.1.1. Facial width to height ratio and baseline testosterone

Across our seven samples, fWHR was not significantly correlated with baseline T ($r = -.038$, $p = .284$, 95% CI [−.11, .03]). As the most powerful test of the relationship between fWHR and baseline T, we aggregated data from the seven samples in the present study with those from the two studies presented in Lefevre et al. (2013), which previously showed one marginal and one significant correlation (data freely available online). Using all available samples, there was no significant correlation between fWHR and baseline T ($r = .011$, $p = .725$, 95% CI [−.05, .07]); see Fig. 1 for the fWHR and baseline T relationship for this combined sample of 1041 men.

3.1.2. Facial width to height ratio and competition-induced testosterone dynamics

For those of our seven samples employing either victory, or victory and defeat conditions, Pearson correlations were run to test relationships between the fWHR and competition-induced T-dynamics. For this, we examined three commonly used measures of change in T: absolute change in T (post T minus pre T), percentage change in T (post T minus pre T/pre T), and residualized T change (unstandardized residuals from regressing post T onto pre T). There was no significant association between mean change in T from pre to post competition ($r = -.030$, $p = .542$, 95% CI [−.12, .06]), percentage change in T from pre to post competition ($r = -.029$, $p = .555$, 95% CI [−.12, .07], or post competition T residuals ($r = -.037$, $p = .445$, 95% CI [−.13, .06]). As T levels are known to fluctuate as a function of competition outcome such that winners generally show a rise compared to losers (see Carré & Olmstead, 2015, for review), we further tested whether the relationship between fWHR and T reactivity would be found in either winners or losers. Analyses revealed that these associations remained non-significant when split by competition outcome: change in T from pre to post competition for *winners* ($r = -.05$, $p = .449$, 95% CI [−.18, .08]), change in T from pre to post competition for *losers* ($r = -.001$, $p = .987$, 95% CI [−.16, .15]); percent change in T for *winners* ($r = -.042$, $p = .520$, 95% CI [−.17, .09]), percent change in T for *losers* ($r = -.032$, $p = .685$, 95% CI [−.19, .12]); post T residuals for *winners* ($r = -.08$, $p = .221$, 95% CI [−.21, .05]), or post T residuals for *losers* ($r = -.014$, $p = .861$, 95% CI [−.17, .14]). By the standards of Cohen (1988), all associations observed in this internal meta-analysis failed to meet even the lowest level of effect (i.e., "small" effects: $r = .1$).

3.2. Individual results

While largely underpowered to detect the small effect found in Lefevre et al. (2013), we nevertheless present data from our individual samples for interested readers, as these can provide an indication of how variable samples are in their statistical effects. Descriptive statistics for fWHR and T levels for each sample are provided in Table 1. Correlations between fWHR and T levels at baseline, as well as fWHR and

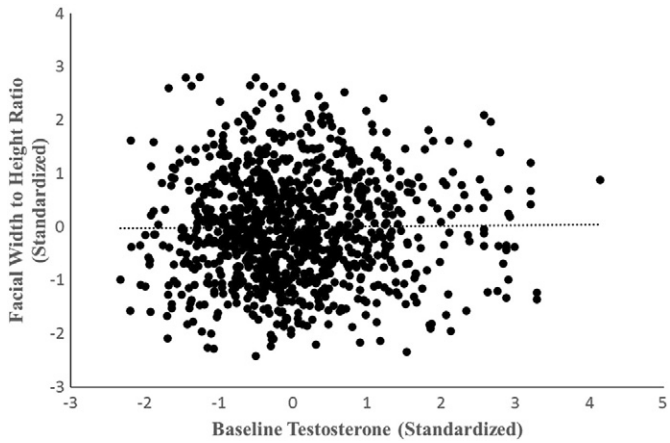

Fig. 1. Scatterplot depicting the null relationship between men's facial width to height ratio and their testosterone levels at baseline, collapsed across seven samples from the present study, plus two samples from Lefevre et al. (2013); Total $N = 1041$. Note: fWHR and baseline testosterone concentrations were standardized within each individual sample.

Please cite this article as: Bird, B.M., et al., Does the facial width-to-height ratio map onto variability in men's testosterone concentrations?, *Evolution and Human Behavior* (2016), http://dx.doi.org/10.1016/j.evolhumbehav.2016.03.004

Table 1
Descriptive statistics for T levels (pg/mL)[a] and fWHR.

Measure	Sample	*n*	Mean	*SD*	Range
T baseline	1	80	82.33	29.4	26.95–190
	2	111	93.83	39.99	22.23–273.59
	3	163	102.48	40.66	26.23–232.86
	4	159	106.21	40.84	22.30–219.50
	5	89	112.74	57.8	18.35–281.88
	6	73	93.00	45.28	34.54–319.96
	7[a]	119	5.34	2.43	1.53–15.09
T post win	1	80	79.45	29.5	26.37–175.92
	2	55	95.21	43.34	23.28–241.56
	3	79	96.77	42.15	13.53–279.9
	5	32	98.06	44.16	22.36–195.35
T post loss	2	56	78.76	31.67	18.34–169.3
	3	84	97.2	39.35	33.25–240.29
	5	28	116.16	49.17	49.44–229.46
fWHR	1	80	1.83	0.13	1.56–2.10
	2	114	1.71	0.12	1.44–2.11
	3	153	1.77	0.15	1.48–2.23
	4	159	1.83	0.14	1.53–2.21
	5	95	1.76	0.16	1.38–2.09
	6	73	1.61	0.13	1.34–1.87
	7	119	1.73	0.15	1.41–2.26

[a] Note that testosterone for sample 7 is assayed from blood serum, and therefore is measured in ng/mL.

change in T following competition (where applicable) are shown in Table 2.

None of the samples showed significant positive correlations between the fWHR and baseline T. Sample 3 revealed a weak, significant negative association ($r = -.161$, $p = .047$), which is directionally opposite to that reported elsewhere (Lefevre et al., 2013). Samples 2, 6, and 7 showed weak, non-significant negative correlations between fWHR and baseline T ($rs = -.15$ to $-.056$, $ps = .115$–$.558$), while samples 1, 4, and 5 revealed weak, non-significant positive correlations ($rs = .053$–$.146$; $ps = .171$–$.509$: see Table 2).

For those samples employing either victory, or victory and defeat conditions, Pearson correlations were run to examine relationships between the fWHR and competition-induced T-dynamics. As shown in Table 2, all comparisons revealed weak, non-significant associations. To test for differential fWHR and T relationships as a function of winning or losing, separate analyses were run in each sample based on the type of competition outcome. With the exception of sample 2—showing significant *negative* correlations between fWHR and T changes following competition among winners only—all samples showed non-significant correlations, regardless of win or loss outcome (see Table 3).

4. Discussion

The rapidly accumulating literature showing positive links between the fWHR and dominance-related phenotypes in men (see Geniole et al., 2015; Haselhuhn et al., 2015, for meta-analyses) has inspired attempts to identify a potential underlying mechanism. Previous work has speculated that the relative influence of T on both craniofacial growth and the expression of sexually dimorphic behaviors or traits may be one such mechanism (Carré & McCormick, 2008). Recent studies have shown some evidence that the fWHR is correlated with T levels in adults, but these associations are weak, and have been inconsistent across the few studies explicitly examining this relationship. The present study tested the association between fWHR and baseline T using seven samples with a combined sample of 780 men; we probed this relationship further by adding two previously published samples from Lefevre et al. (2013), giving a final analysis with a sample size of 1041 men. With our own samples, we also tested the relationship between fWHR and competition-induced T dynamics following a win or a loss, across 4 samples of men ($n = 428$).

Among individual samples, links between facial structure and baseline T levels showed considerable variability. Specifically, fWHR was positively associated with baseline T in three of the seven samples, while negatively in the other four. While under-powered to detect the smallest effect previously found (i.e., $r = .13$), only one of these associations was statistically significant, yet was in the opposite direction to that found previously (Lefevre et al., 2013). Crucially, when all samples were combined for internal meta-analysis to maximize statistical power, the correlation between fWHR and baseline T became effectively non-existent. Adding the two samples from previously published work (i.e., Lefevre et al., 2013) did not significantly affect these findings.

Two of the most recent studies explicitly examining the relationship between T and facial structure also show mixed findings for baseline T, which, when considered in the context of the present study, align with the findings from our individual samples. Pound et al. (2009) found that a global measure of facial masculinity was not significantly associated with baseline T ($r = 0.19$, n.s, $n = 47$), while Lefevre et al. (2013) found a marginally significant relationship between fWHR and baseline T in one sample ($r = .13$, $p \leq .10$, $n = 180$), yet a significant positive relationship in another ($r = .26$, $p = .03$, $n = 79$). Again, the variability across these studies, in combination with those from the present study, suggests that the association between facial structure and T levels at baseline may vary considerably across smaller samples, but appears to be virtually non-existent when using well-powered sample sizes.

We also examined how competition-induced T dynamics mapped onto the fWHR, as a body of research demonstrates that T fluctuates rapidly in the face of competition such that on average, those winning will experience a rise, while those losing will experience a drop (Archer, 2006; Carré & Olmstead, 2015). If these hormonal reactions serve to guide future status-seeking and dominance behavior for winners, while encouraging the avoidance of further loss of status for losers (Mazur, 1985), the predictive power of T might be stronger for reactive, rather than baseline T levels (Lefevre et al., 2013). When considering three commonly employed T change metrics (mean change in T, percentage change in T, or post competition residual T), the present study failed to find any strong or significant positive association for winners, or significant negative association for losers. In fact, the analysis of individual samples showed nearly equal variability in positive or negative associations among winning and losing participants. Given the smaller sample sizes, however, we collapsed across all samples to boost statistical power and estimation precision; this analysis revealed no significant associations between fWHR and mean change in T after competition,

Table 2
Correlations between fWHR and both baseline T, and change in T following competition, stratified by sample.

	Baseline T			Post T–pre T			% change in T			T residuals		
	r	*p*	*n*	*r*	*p*	*n*	*r*	*p*	*n*	*r*	*p*	*n*
S1 fWHR	.093	.413	80	.077	.497	80	.046	.687	80	.092	.415	80
S2 fWHR	−.056	.558	111	−.044	.647	109	.003	.974	109	−.077	.425	109
S3 fWHR	−.161	.047	152	−.008	.918	151	−.027	.743	151	−.071	.385	151
S4 fWHR	.053	.509	159	–	–	–	–	–	–	–	–	–
S5 fWHR	.146	.171	89	−.143	.185	88	−.139	.197	88	−.048	.658	88
S6 fWHR	−.151	.212	70	–	–	–	–	–	–	–	–	–
S7 fWHR	−.145	.115	119	–	–	–	–	–	–	–	–	–

S = sample, fWHR = Facial width to height ratio, T = testosterone (pg/mL). Note: S7 is measured in ng/mL.

Please cite this article as: Bird, B.M., et al., Does the facial width-to-height ratio map onto variability in men's testosterone concentrations?, *Evolution and Human Behavior* (2016), http://dx.doi.org/10.1016/j.evolhumbehav.2016.03.004

Table 3
Correlations between fWHR and change in T, stratified by competition outcome and sample.

Outcome	fWHR	Post T–pre T			% change in T			T residuals		
		r	*p*	*n*	*r*	*p*	*n*	*r*	*p*	*n*
Win	S1 fWHR	.077	.497	80	.046	.687	80	.092	.415	80
	S2 fWHR	−.287	.037	53	−.267	.053	53	−.345	.011	53
	S3 fWHR	.027	.826	71	.028	.816	71	−.057	.637	71
	S5 fWHR	−.241	.191	31	−.247	.181	31	−.226	.221	31
Lose	S2 fWHR	.093	.495	56	.160	.239	56	.108	.430	56
	S3 fWHR	−.048	.674	80	−.076	.505	80	−.091	.423	80
	S5 fWHR	−.146	.478	26	−.296	.141	26	−.104	.614	26

S = sample, fWHR = facial width to height ratio, T = testosterone (pg/mL).

percentage change in T after competition, or residualized post-competition T, regardless of whether it was split by competition outcome (win or lose), or collapsed across winners and losers.

Pound et al. (2009) found that among individuals randomly assigned to a winning condition for a betting task, there was a significant association between a global measure of facial masculinity and post-competition T levels ($r = .36$, $p = .013$); however, this study did not examine fWHR, included only 10 participants in a losing control condition, and was limited in its small sample size for the winning condition ($n = 47$). In a larger sample, Lefevre et al. (2013) found no positive association between the same global masculinity measure used in Pound et al. and men's T levels following a speed-dating event, yet did find that fWHR was positively correlated with post-event T levels. One potential reason for the discrepancy between Lefevre et al.'s positive relationship and the null findings from the present study is that in their speed-dating paradigm, there was no clear way to discriminate winners from losers; rather, individuals were simply hoping to meet future mates, and as such, it may not have provided an effective competition outcome variable. Perhaps more importantly, statisticians strongly recommend that rather than simply employing null hypothesis testing, effect size estimation with confidence intervals (in addition to meta-analyses where possible), should be employed in order to derive scientific conclusions (Cumming, 2013). The inconsistent effects found in these previous studies fall within small to moderate effect size estimates (Cohen, 1988), and in light of these statistical considerations, the null findings in the present study provide a much more robust and convincing test of these relationships.

The non-significant results from the present study suggest that fWHR does not reliably map onto T levels in adulthood at either baseline, or as a function of reactive T in the face of winning or losing a competition. Rather, they suggest that fWHR–neuroendocrine links in men seem to be highly heterogeneous, and effectively non-existent in a large and diverse sample of men, and thus do not appear to be—at least not on their own—a viable explanatory mechanism for linking facial structure to behavioral displays in men.

4.1. Limitations and future directions

Although we find no evidence for a significant association between the fWHR and baseline T, or the fWHR and reactive T following competition, we cannot discount the possibility that fWHR predicts reactive T following an interaction with a potential mating partner, as was noted in Lefevre et al. (2013). Such an effect might be related to the presence of sexual rivals (see Maner, Miller, Coyle, & Kaschak, 2014), rather than a competitive situation, per se. However, the nature of the present study prevented us from testing these relationships, so this should be considered in future investigations. Further, some previous research has noted a correlation between body mass index (BMI) and the fWHR (e.g., Mayew, 2013, see Geniole et al., 2015 for meta-analysis), suggesting that BMI may mask certain effects. BMI data were not available from the samples analyzed in the present study, and therefore future studies attempting replication may want to consider collecting BMI for use as a control variable.

Future research might also consider the modulating influence of the trinucleotide CAG repeat polymorphism in the first exon of the androgen receptor (AR) gene. AR sensitivity is negatively related to the length of CAG repeats, and thus should produce larger phenotypic effects of androgens among those with relatively shorter CAG repeat lengths (Chamberlain, Driver, & Miesfeld, 1994; Choong, Kemppainen, Zhou, & Wilson, 1996). Indeed, recent evidence suggests that basal T concentrations are positively correlated with aggressive and non-aggressive risk-taking behavior, but only among individuals with short CAG repeats (Vermeersch, T'sjoen, Kaufman, Vincke, & Van Houtte, 2010). Furthermore, positive correlations between T and impulsivity is only found among individuals with shorter CAG repeat lengths (Aluja et al., 2015). Thus, a positive relationship between fWHR and T-levels at baseline or following competition may still exist, but it may be reserved for those with relatively shorter CAG repeats.

As has been previously noted, circulating T and facial masculinity both increase in adolescence, and T administration during this same period has been shown to influence the development of facial structure (Verdonck et al., 1999). Studies in both humans (e.g., Weston, Friday, & Liò, 2007) and non-human species (e.g., capuchin monkeys: Lefevre et al., 2014) also find that the fWHR becomes sexually-dimorphic around puberty, suggesting that pubertal T shapes, at least to some degree, variation in the fWHR. Other research shows that accurate identification of male and female faces does not happen until late adolescence, and importantly, accurate identification is strongly predicted by the target's T levels, even after controlling for age (Marečková et al., 2011). Thus, the fWHR's association with certain behavioral dispositions, as well as judgments of such dispositions, may be more closely tied with exposure to T in puberty, rather than to baseline or reactive T levels in adulthood. Other evidence suggests that variation in the fWHR might begin as early as prenatal development. The 2D:4D ratio—a putative negative correlate of prenatal androgen exposure—has been shown to predict more robust and masculine faces among those with lower 2D:4D ratios, and especially so among men (Fink et al., 2005; Meindl, Windhager, Wallner, & Schaefer, 2012; Schaefer, Fink, Mitteroecker, Neave, & Bookstein, 2005). Importantly, the 2D:4D does not appear to map onto adult T levels (see Hönekopp et al., 2007, for meta-analytic review; see also Muller et al., 2011). This evidence, coupled with research showing the organizational and activational effects of T on the prenatal and adolescent brain (reviewed in Berenbaum & Beltz, 2011; Sisk & Zehr, 2005) provides indirect support for the idea that the fWHR and associated behavioral dispositions (e.g., aggression), as well as judgments of such dispositions, may also be linked more closely to T exposure early in development, rather than circulating T concentrations in adulthood.

4.2. Conclusion

In summary, the present study shows that across seven independent samples, the fWHR is neither significantly positively related to men's circulating T-levels, nor to their reactive T-levels following competition. Future research would be better suited towards examining the influence of prenatal and/or pubertal androgens, as well as the potential modulating effect of CAG repeats, on facial structure and behavioral traits.

Supplementary Material

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.evolhumbehav.2016.03.004.

References

Aluja, A., García, L. F., Martí-Guiu, M., Blanco, E., García, O., Fibla, J., & Blanch, À. (2015). Interactions among impulsiveness, testosterone, sex hormone binding globulin and

androgen receptor gene CAG repeat length. *Physiology & Behavior*, *147*, 91–96. http://dx.doi.org/10.1016/j.physbeh.2015.04.022.
Anderl, C., Hahn, T., Schmidt, A., Moldenhauer, H., Notebaert, K., Clément, C. C., & Windmann, S. (2016). Facial width-to-height ratio predicts psychopathic traits in males. *Personality and Individual Differences*, *88*, 99–101. http://dx.doi.org/10.1016/j.paid.2015.08.057.
Apicella, C. L., Dreber, A., Campbell, B., Gray, P. B., Hoffman, M., & Little, A. C. (2008). Testosterone and financial risk preferences. *Evolution and Human Behavior*, *29*(6), 384–390. http://dx.doi.org/10.1016/j.evolhumbehav.2008.07.001.
Apicella, C. L., Dreber, A., Gray, P. B., Hoffman, M., Little, A. C., & Campbell, B. C. (2011). Androgens and competitiveness in men. *Journal of Neuroscience, Psychology, and Economics*, *4*(1), 54–62. http://dx.doi.org/10.1037/a0021979.
Archer, J. (2006). Testosterone and human aggression: An evaluation of the challenge hypothesis. *Neuroscience & Biobehavioral Reviews*, *30*(3), 319–345. http://dx.doi.org/10.1016/j.neubiorev.2004.12.007.
Berenbaum, S. A., & Beltz, A. M. (2011). Sexual differentiation of human behavior: Effects of prenatal and pubertal organizational hormones. *Frontiers in Neuroendocrinology*, *32*(2), 183–200. http://dx.doi.org/10.1016/j.yfrne.2011.03.001.
Campbell, B. C., Dreber, A., Apicella, C. L., Eisenberg, D. A., Gray, P. B., Little, A. C., ... Lum, J. K. (2010). Testosterone exposure, dopaminergic reward, and sensation-seeking in young men. *Physiology & Behavior*, *99*(4), 451–456. http://dx.doi.org/10.1016/j.physbeh.2009.12.011.
Carré, J. M., Campbell, J. A., Lozoya, E., Goetz, S. M., & Welker, K. M. (2013). Changes in testosterone mediate the effect of winning on subsequent aggressive behaviour. *Psychoneuroendocrinology*, *38*(10), 2034–2041. http://dx.doi.org/10.1016/j.psyneuen.2013.03.008.
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. http://dx.doi.org/10.1098/rspb.2008.0873.
Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009a). Facial structure is a reliable cue of aggressive behavior. *Psychological Science*, *20*(10), 1194–1198.
Carré, J. M., Morrissey, M. D., Mondloch, C. J., & McCormick, C. M. (2010). Estimating aggression from emotionally neutral faces: Which facial cues are diagnostic? *Perception*, *39*(3), 356–377. http://dx.doi.org/10.1068/p6543.
Carré, J. M., & Olmstead, N. A. (2015). Social neuroendocrinology of human aggression: Examining the role of competition-induced testosterone dynamics. *Neuroscience*, *286*, 171–186. http://dx.doi.org/10.1016/j.neuroscience.2014.11.029.
Carré, J. M., Putnam, S. K., & McCormick, C. M. (2009b). Testosterone responses to competition predict future aggressive behaviour at a cost to reward in men. *Psychoneuroendocrinology*, *34*, 562–570. http://dx.doi.org/10.1016/j.psyneueu.2008.10.018.
Chamberlain, N. L., Driver, E. D., & Miesfeld, R. L. (1994). The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Research*, *22*(15), 3181–3186. http://dx.doi.org/10.1093/nar/22.15.3181.
Choong, C. S., Kemppainen, J. A., Zhou, Z. X., & Wilson, E. M. (1996). Reduced androgen receptor gene expression with first exon CAG repeat expansion. *Molecular Endocrinology*, *10*(12), 1527–1535. http://dx.doi.org/10.1210/mend.10.12.8961263.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.
Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. NY: Routledge.
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.
Fink, B., Grammer, K., Mitteroecker, 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 of London B: Biological Sciences*, *272*(1576), 1995–2001. http://dx.doi.org/10.1098/rspb.2005.3179.
Geniole, S. N., Denson, T. F., Dixson, B. J., Carre, 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). http://dx.doi.org/10.1371/journal.pone.0132726.
Geniole, S. N., Keyes, A. E., Carré, J. M., & McCormick, C. M. (2014a). Fearless dominance mediates the relationship between the facial width-to-height ratio and willingness to cheat. *Personality and Individual Differences*, *57*, 59–64. http://dx.doi.org/10.1016/j.paid.2013.09.023.
Geniole, S. N., & McCormick, C. M. (2015). Facing our ancestors: Judgements of aggression are consistent and related to the facial width-to-height ratio in men irrespective of beards. *Evolution and Human Behavior*, *36*(4), 279–285. http://dx.doi.org/10.1016/j.evolhumbehav.2014.12.005.
Geniole, S. N., Molnar, D. S., Carré, J. M., & McCormick, C. M. (2014b). 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. http://dx.doi.org/10.1037/a0036732.
Goetz, S. M., Shattuck, K. S., Miller, R. M., Campbell, J. A., Lozoya, E., Weisfeld, G. E., & Carré, J. M. (2013). Social status moderates the relationship between facial structure and aggression. *Psychological Science*, *24*(11), 2329–2334. http://dx.doi.org/10.1177/0956797613493294.
Gómez-Valdés, J., Hünemeier, T., Quinto-Sánchez, M., Paschetta, C., de Azevedo, S., González, M. F., ... González-José, R. (2013). Lack of support for the association between facial shape and aggression: A reappraisal based on a worldwide population genetics perspective. *PloS One*, e52317.
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*(4). http://dx.doi.org/10.1371/journal.pone.0122637.
Haselhuhn, M. P., & Wong, E. M. (2012). Bad to the bone: Facial structure predicts unethical behaviour. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1728), 571–576. http://dx.doi.org/10.1098/rspb.2011.1193.
Haselhuhn, M. P., Wong, E. M., Ormiston, M. E., Inesi, M. E., & Galinsky, A. D. (2014). Negotiating face-to-face: Men's facial structure predicts negotiation performance. *The Leadership Quarterly*, *25*(5), 835–845. http://dx.doi.org/10.1016/j.leaqua.2013.12.003.
Hehman, E., Leitner, J. B., Deegan, M. P., & Gaertner, S. L. (2013). Facial structure is indicative of explicit support for prejudicial beliefs. *Psychological Science*, *24*(3), 289–296. http://dx.doi.org/10.1177/0956797612451467.
Hönekopp, J., Bartholdt, L., Beier, L., & Liebert, A. (2007). Second to fourth digit length ratio (2D:4D) and adult sex hormone levels: New data and a meta-analytic review. *Psychoneuroendocrinology*, *32*(4), 313–321. http://dx.doi.org/10.1016/j.psyneuen.2007.01.007.
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. http://dx.doi.org/10.1016/j.evolhumbehav.2013.03.005.
Lefevre, C. E., Wilson, V. A., Morton, F. B., Brosnan, S. F., Paukner, A., & Bates, T. C. (2014). Facial width-to-height ratio relates to alpha status and assertive personality in capuchin monkeys. *PloS One*, *8*.
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*(7), 855–857. http://dx.doi.org/10.1016/j.paid.2011.12.030.
Maestripieri, D., Baran, N. M., Sapienza, P., & Zingales, L. (2010). Between- and within-sex variation in hormonal responses to psychological stress in a large sample of college students. *Stress: The International Journal on The Biology Of Stress*, *13*(5), 413–424. http://dx.doi.org/10.3109/10253891003681137.
Maner, J., Miller, S., Coyle, J., & Kaschak, M. (2014). Confronting intrasexual rivals: 2D:4D digit ratio predicts behavioral and endocrinological responses to infidelity threat. *Social Psychological and Personality Science*, *5*(1), 119–128. http://dx.doi.org/10.1177/1948550613488948.
Marečková, K., Weinbrand, Z., Chakravarty, M. M., Lawrence, C., Aleong, R., Leonard, G., ... Paus, T. (2011). Testosterone-mediated sex differences in the face shape during adolescence: Subjective impressions and objective features. *Hormones and Behavior*, *60*(5), 681–690. http://dx.doi.org/10.1016/j.yhbeh.2011.09.004.
Mayew, W. J. (2013). Reassessing the association between facial structure and baseball performance: A comment on Tsujimura & Banissy (2013). *Biology Letters*, *9*(5), 20130538.
Mazur, A. (1985). A biosocial model of status in face-to-face primate groups. *Social Forces*, *64*(2), 377–402 (http://www.jstor.org/stable/2578647).
Mehta, P. H., & Josephs, R. A. (2006). Testosterone change after losing predicts the decision to compete again. *Hormones & Behavior*, *50*, 684–692. http://dx.doi.org/10.1016/j.yhbeh.2006.07.001.
Meindl, K., Windhager, S., Wallner, B., & Schaefer, K. (2012). Second-to-fourth digit ratio and facial shape in boys: The lower the digit ratio, the more robust the face. *Proceedings of the Royal Society of London B: Biological Sciences*. http://dx.doi.org/10.1098/rspb.2011.2351 rspb2011.2351.
Muller, D. C., Giles, G. G., Bassett, J., Morris, H. A., Manning, J. T., Hopper, J. L., ... Severi, G. (2011). Second to fourth digit ratio (2D: 4D) and concentrations of circulating sex hormones in adulthood. *Reproductive Biology and Endocrinology*, *9*(57) (http://rbej.com/content/9/1/57).
Norman, R. E., Moreau, B. J. P., Welker, K. M., & Carré, J. M. (2014). Trait anxiety moderates the relationship between testosterone responses to competition and aggressive behavior. *Adaptive Human Behavior and Physiology*, 1–13. http://dx.doi.org/10.1007/s40750-014-0016-y.
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. http://dx.doi.org/10.1016/j.evolhumbehav.2011.08.001.
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*(4), 229–241. http://dx.doi.org/10.1016/j.evolhumbehav.2004.04.003.
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. http://dx.doi.org/10.1016/j.anbehav.2008.02.008.
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*(1654), 153–159. http://dx.doi.org/10.1098/rspb.2008.0990.
Roney, J. R., Hanson, K. N., 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 of London B*, *273*, 2169–2175. http://dx.doi.org/10.1098/rspb.2006.3569.
Sapienza, P., Zingales, L., & Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(36), 15268–15273. http://dx.doi.org/10.1073/pnas.0907352106.
Schaefer, K., Fink, B., Mitteroecker, P., Neave, N., & Bookstein, F. L. (2005). Visualizing facial shape regression upon 2nd to 4th digit ratio and testosterone. *Collegium Antropologicum*, *29*(2), 415–419 (UDC 572.54.087-055.1:612.616.31).
Short, L. A., Mondloch, C. J., McCormick, C. M., Carré, J. M., Ma, R., Fu, G., & Lee, K. (2012). Detection of propensity for aggression based on facial structure irrespective of face race. *Evolution and Human Behavior*, *33*(2), 121–129. http://dx.doi.org/10.1016/j.evolhumbehav.2011.07.002.
Sisk, C. L., & Zehr, J. L. (2005). Pubertal hormones organize the adolescent brain and behavior. *Frontiers in Neuroendocrinology*, *26*(3–4), 163–174. http://dx.doi.org/10.1016/j.yfrne.2005.10.003.
Trebická, 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. http://dx.doi.org/10.1002/ab.21559.

Please cite this article as: Bird, B.M., et al., Does the facial width-to-height ratio map onto variability in men's testosterone concentrations?, *Evolution and Human Behavior* (2016), http://dx.doi.org/10.1016/j.evolhumbehav.2016.03.004

Tsujimura, H., & Banissy, M. J. (2013). Human face structure correlates with professional baseball performance: Insights from professional Japanese baseball players. *Biology Letters*, *9*(3), 20130140. http://dx.doi.org/10.1098/rsbl.2013.0140.

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. http://dx.doi.org/10.1093/ejo/21.2.137.

Vermeersch, H., T'sjoen, G., Kaufman, J. M., Vincke, J., & Van Houtte, M. (2010). Testosterone, androgen receptor gene CAG repeat length, mood and behaviour in adolescent males. *European Journal of Endocrinology*, *163*(2), 319–328. http://dx.doi.org/10.1530/EJE-10-0090.

Welker, K. M., & Carré, J. M. (2015). Individual differences in testosterone predict persistence in men. *European Journal of Personality*, *29*(1), 83–89. http://dx.doi.org/10.1002/per.1958.

Welker, K. M., Goetz, S. M., Galicia, S., Liphardt, J., & Carré, J. M. (2014). An examination of the associations between facial structure, aggressive behavior, and performance in the 2010 World Cup Association Football Players. *Adaptive Human Behavior and Physiology*, *1*(1), 17–29. http://dx.doi.org/10.1007/s40750-014-0003-3.

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

Zilioli, S., Sell, A. N., Stirrat, M., Jagore, J., Vickerman, W., & Watson, N. V. (2014). Face of a fighter: Bizygomatic width as a cue of formidability. *Aggressive Behavior*, *41*(4), 322–330. http://dx.doi.org/10.1002/ab.21544.

Please cite this article as: Bird, B.M., et al., Does the facial width-to-height ratio map onto variability in men's testosterone concentrations?, *Evolution and Human Behavior* (2016), http://dx.doi.org/10.1016/j.evolhumbehav.2016.03.004