ORIGINAL ARTICLE

Relationships of the Ulna-to-fibula Ratio to Baseline and Reactive Steroid Hormone Levels: An Exploratory Study

Martin G. Köllner[1] · **Sinja Braun**[1] · **Hanna Schöttner**[1] · **Gelena Dlugash**[1] · **Marlene Bettac**[1] · **Simon Steib**[2]

Abstract

Objective Organizational hormone effects on the human brain and behavior are often retrospectively assessed via morphological markers of prenatal (e.g., 2D:4D digit ratio) or pubertal (e.g., facial width-to-height ratio, fWHR) hormone exposure. It has been argued that markers should relate to circulating hormones particularly in challenging, dominance/status-relevant situations. However, meta-analytic research indicates that fWHR, a frequently used pubertal marker, is neither reliably sex-dimorphic nor related to steroid hormones. This casts doubt on fWHR's validity for reflecting hormone levels. Ulna-to-fibula ratio (UFR), an alternative, long-bone-length-based pubertal marker, is sex-dimorphic and associated with dominance motivation. However, its hormonal associations were never tested before. We therefore explored UFR's relationships to baseline and reactive hormone levels.

Methods We measured ulna and fibula length as well as shoulder/waist/hip circumference of 81 participants (49 women; after exclusions) via anthropometry. Salivary hormone levels (estradiol, testosterone) at baseline and after a gross-motor one-on-one balancing contest were measured via radioimmunoassay.

Results We replicated UFR's dimorphism, unrelatedness to height, and correlations to other putative markers of organizational hormone effects. On an exploratory basis, we found UFR to be related to overall baseline testosterone and to competition-induced reactive surges in steroid hormones (estradiol, testosterone) overall and in women.

Conclusions Our results hint at UFR's relationship to baseline testosterone and may indicate functional connections between outcomes of pubertal organizational hormone effects and contest-induced steroid reactivity. Pubertal organizational hormone effects may prepare the endocrine system for dominance and status contests. However, the small sample and the exploratory nature of our research demands replication.

Keywords Long bone length · Organizational hormone effects · Puberty · Steroid hormones · Ulna-to-fibula ratio

Extended author information available on the last page of the article

Introduction

Status changes (i.e., gains, losses) and dominance contests are functionally related to steroid hormones (e.g., Mazur, 1985) and the hormones relevant to female (estradiol; Stanton & Schultheiss, 2007) and male dominance (testosterone; Schultheiss et al., 2005) seem to differ. One possible origin of the hormonal mechanisms accompanying dominance contests may be found in outcomes of developmental hormone exposure preparing the endocrine system for challenge situations (Manning et al., 2014). Extending on this idea, this study explores relationships between ulna-to-fibula ratio (UFR), a potential indicator of pubertal steroid exposure (Köllner & Bleck, 2020), with baseline levels and contest-induced surges of estradiol and testosterone.

Organizational Hormone Effects

Hormones not only have transient activational effects, but also *organizational hormone effects*, lasting influences of steroid hormones on brain and behavior during mammalian development (for review, Arnold, 2009). While organizational hormone effects were first found for prenatal development (e.g., Phoenix et al., 1959), recent research made clear that there is at least one additional organizational developmental stage where neuronal plasticity and elevated hormone levels again coincide: puberty (e.g., Schulz & Sisk, 2016, especially fig. 1; cf. standard textbook by Nelson, & Kriegsfeld, 2017). Growing evidence suggests that this is also true for human development (Doll et al., 2016: male phenotypic masculinization; Shirazi et al., 2020b: GnRH deficiency; Shirazi et al., 2020a: visuospatial cognition). In addition, indicators of organizational hormone effects during puberty are meta-analytically related to dominance behavior and aggression (Geniole et al., 2015; Haselhuhn et al., 2015).

Morphometric Markers of Organizational Hormone Effects

In research involving human participants, organizational hormone effects on brain and behavior are mostly retrospectively approximated via morphological *markers* of prenatal (e.g., digit ratio of second and fourth digit, 2D:4D; Manning et al., 2014) or pubertal (e.g., facial width-to-height ratio, fWHR; Carré & McCormick, 2008; for the use of the term "marker", see Hönekopp et al., 2007) steroid exposure (see also bone length research by Martin & Nguyen, 2004).[1] In short, proponents of markers assume simultaneous organizational hormone effects on (a) brain (and thus adult behavior) development and (b) the steroid-driven growth of hormone-sensitive body

---

[1] The marker approach is used by many scientists due to its low costs, non-invasive nature, and the possibility to include it in cross-sectional designs and studies involving healthy adults. Of course there are other methods to assess organizational hormone effects on the human brain, most prominently comparing individuals with endocrine conditions to unaffected individuals (e.g. congenital adrenal hyperplasia, Beltz et al., 2021; complete androgen insensitivity syndrome, Hamann et al., 2014; isolated gonadotropin-releasing hormone deficiency, Swift-Gallant et al., 2021; Shirazi et al., 2021; Shirazi et al., 2022) and also methods involving otoacoustic emissions (McFadden & Pasanen, 1998) or auditory evoked potentials (McFadden & Champlin, 2000).

dimensions during a developmental stage like the prenatal period or puberty. If these body dimensions do not experience further change after the end of a given developmental stage, marker-behavior-relationships in adulthood may reflect organizational hormone effects on brain development (cf. Hönekopp et al., 2007).

The marker approach has sparked an ongoing debate in behavioral endocrinology. Taking 2D:4D as an example, this ratio was criticized for replication problems (Leslie, 2019) and the practice of some researchers to atheoretically correlate it with a multitude of outcome measures (McCormick & Carré, 2020). However, large genome-wide association studies (Warrington et al., 2018) and solid experimental findings from animal research (Zheng & Cohn, 2011) nevertheless support the assumption that 2D:4D reflects prenatal hormone exposure (e.g., Swift-Gallant et al., 2020).

Markers of pubertal organizing hormone effects suffer from even less corroborating research being available (aside from some promising recent findings; Doll et al., 2016; Shirazi et al., 2020a, b). In particular, recent meta-analytic research indicates that fWHR, a frequently used possible pubertal marker, is neither reliably sex-dimorphic (Kramer, 2017; but see Köllner et al., 2018) nor related to baseline steroid hormone levels nor to contest-induced hormone reactivity (Bird et al., 2016), casting doubt on its validity for reflecting hormone levels (but see Welker et al., 2016a, for a low-powered finding possibly hinting at fWHR's relationship to pubertal testosterone in Tsimane men).

Lacking sexual dimorphism is unacceptable for a marker of organizational hormone effects: If a marker does not even reflect the large between-sex differences in developmental hormones (Ober et al., 2008), it is unlikely to reflect more fine-grained within-sex differences between individuals (cf. Köllner et al., 2022b). Above that, fWHR's simultaneously lacking hormonal associations are problematic for two reasons. First, sex-steroids rise markedly during puberty, the time at which fWHR emerges (Weston et al., 2007), and stay high afterwards until at least middle-adulthood (see fig. 1 in Ober et al., 2008, with steroids rising to a several-decades-lasting plateau around the time of puberty; cf. Dimitrakakis & Bondy, 2009, for a similar trajectory of steroid levels across the female lifespan). Thus, based on the overall lifespan trajectory of steroid levels, continued associations between a pubertal marker and circulating baseline hormones in adulthood may potentially be plausible, even though we are not aware of any longitudinal research on the individual stability of high pubertal hormone levels into middle adulthood.[2] Second, and most importantly, organizational hormone effects may prepare the adult endocrine system for adaptive functioning in sexual and competitive, status-related, or dominance-related encounters (cf. Manning et al., 2014). If this should be correct, then we

---

[2] However, it may also be the case that individuals with high numbers of or more sensitive androgen receptors compensatorily produce lower levels of androgens later in life. This in turn would reduce or even change the sign of correlations between adult hormone levels and pubertal markers. While for example the Odense Androgen Study (Nielsen et al., 2010) did not find negative associations between CAG repeat polymorphism and circulating androgen levels in young men, we nevertheless cannot refute this alternative view completely and stress that our study is explorative in nature.

would expect especially clear-cut connections between a pubertal marker and the magnitude of contest-induced hormone reactivity. Indeed, the competitive domain of sports is one of the areas which meta-analytically demonstrated the most substantial relationships to a marker of organizational hormone effects to this date (2D:4D; Hönekopp & Schuster, 2010). Simultaneously, organizational hormone effects seem to be related to steroid level spikes in response to sport competitions (2D:4D asymmetry; Kilduff et al., 2013), which provides a preliminary empirical basis for the above-mentioned expectations.

The Ulna-to-fibula Ratio

Given fWHR's lacking hormonal associations, and consequently, also lacking sexual dimorphism, alternative markers should be identified. An alternative possible pubertal marker using bone lengths is UFR, the ratio of forearm length (measured from the ulna) to lower leg length (measured from the fibula). In general, pentadactyl limbs should provide rich potential sources for markers of developmental sex steroids: Manning (2002) suggests that there is a hox-gene-controlled developmental link between the reproduction-relevant urogenital system and limbs in vertebrate evolution.

UFR is sexually dimorphic, with men displaying longer forearms (cf.Knight, 1991; Tanner, 1990) relative to their lower legs than women (Köllner & Bleck, 2020). UFR was sex-dimorphic (male > female), unrelated to overall body height (and thus no artifact of overall between-sex size differences; cf. Leslie, 2019), related to other potential markers of organizational hormone effects, and associated with dominance motivation in two studies (Köllner & Bleck, 2020; $N = 126$; Köllner et al., 2022b, $N = 250$).

UFR focuses on long bone growth, which is the target of pubertal organizational hormone effects of testosterone and estradiol (e.g. Vanderschueren et al., 2004; cf. Cutler, 1997; Juul, 2001, for details regarding biphasic sex-steroid-effects on pubertal bone growth). The sex-dimorphic ratio of ulna and fibula, two ontogenetically homologous bones, is considered to approximate sex-typical steroid exposure (for details, see Köllner & Bleck, 2020): Köllner and Bleck (2020) found sexual dimorphism of the upper limb (ulna; cf. Purkait, 2001) and used the lower limb as a reference for overall long bone growth. Correspondingly, hypogonadic men affected by Klinefelter syndrome have, besides low androgen-levels, a "feminized" lower UFR, with shorter arms relative to the elongated legs (and also a "feminized" higher 2D:4D; Chang et al., 2015; Manning et al., 2013).

UFR involves large, easily measurable bone structures. It thus goes beyond subtle, potentially more measurement-error prone variations in digit lengths or facial dimensions (further distorted by fat deposits in the cheeks; cf. Lefevre et al., 2013). UFR thus enables low-cost, non-invasive, and reliable body surface assessment of large bones, likely reducing measurement errors compared to subtle fWHR variations. Excellent interrater reliability data for ulna and fibula was reported by Köllner and Bleck (2020) and Köllner et al., (2022b; also compare unsatisfactory reliability for fWHR reported in the same study).

The Present Study

However, UFR's hormonal associations were never tested before, leaving a gap between the postulate of a hormone-sensitive body dimension and assessing actual hormone levels.

The most straightforward test for UFR's hormonal associations would involve testing during the – currently unknown—time in development where UFR's sexual dimorphism emerges. However, as a first exploratory test, adult participants may be acceptable. Thus, we conducted a study including anthropometric measurements and two salivary hormone measurements in adult participants, one before and one after an one-on-one dominance contest in the domain of sport. While refraining from deriving specific formal hypotheses due to the lack of a preregistration for these specific research questions, we expect UFR to be related to baseline hormone levels (which could not be demonstrated for the alternative fWHR dimension; Bird et al., 2016). Further, we agree with Manning et al. (2014) in considering relationships of a marker with challenge-induced reactive hormone changes especially indicative of organizational hormone effects on the endocrine system. Such organizational hormone effects on later hormone reactivity may facilitate challenging encounters (cf. challenge hypothesis; Archer, 2006) and help in maintaining or achieving high status in later adult life (Mazur & Booth, 1998). Thus, we further expect UFR to be related to reactive hormone changes in a contest situation. We also gathered preliminary data on UFR's relationships to hormone changes dependent on contest outcome (i.e., winning and losing), as the biosocial model of status would predict rising steroid levels in winners and falling levels in losers (Mazur, 1985; see also the reciprocal model by Mazur & Booth, 1998). An endocrine system ideally prepared for challenging encounters by higher developmental hormone exposure potentially may display such result-dependent changes in a more pronounced way. However, we have to note that the relationships between hormones and dominance in women are complex and rest on much less available research compared to men. Much of the related work for making our predictions mainly rests on observations in men and/or regarding testosterone (Archer, 2006; Mazur & Booth, 1998) and may not be fully generalizable to women and/or estradiol. Nevertheless, some research indicates that similar relationships in women may exist, like rising or falling estradiol levels in response to winning or losing a dominance contest (Stanton & Schultheiss, 2007; with no corresponding findings for basal or reactive testosterone) or positive relationships of estradiol to intrasexual competition (attention, Fiacco et al., 2021) in premenopausal women. As estradiol seems to fluctuate during female sports competitions, its use for such studies can be recommended (Edwards & Turan, 2020). Thus, to provide a complete descriptive picture, we used a mixed-sex sample as well as simultaneous assessment of estradiol and testosterone for the first exploratory study on UFR's hormonal associations.

Additionally, we aimed at replicating UFR's sexual dimorphism, unrelatedness to overall body height, and convergent validity with waist-to-hip ratio (WHR) and shoulder-to-hip ratio (SHR), other body dimensions developing in a sex-dimorphic way during puberty, with higher scores observed for men than for women, respectively (Köllner & Bleck, 2020; Köllner et al., 2022b).

Method

Sample

We tested 90 and retained 81 healthy adults (49 women; age: $M = 22.33$ years, $SD = 3.10$, age range: 19–35) after exclusions (missing anthropometric and/or salivary measurements, BMI ≤ 17.5, pregnancy, hormonal medication, suspicion regarding fake contest feedback). Participants were mainly recruited via flyers in Friedrich-Alexander University's sport science and sport department as part of a larger experiment on determinants of motor learning in a contest situation (preregistered at https://osf.io/f6q5t). We acknowledge the lack of an a-priori power analysis for our present research questions, as sample size was based on deliberations regarding this larger experiment: The funding grant was limited to 3600€ for participant payments with a participant receiving 20€ per testing day and two testing days. Thus, our sample size was limited to 90 participants.[3]

Design

Our design itself was correlational, but the main experiment featured between-subject-fake-feedback on contest performance (win vs. lose). Sex (female vs. male) was a categorical variable. Hormone concentrations and anthropometric measurements were continuous variables.

Materials

Anthropometric Measurements

BMI (weight in kg/squared height in m) was determined with a bathroom scale (weight) and participants' self-report (height; aided by a measuring tape in case of doubt). A questionnaire assessed age and sex. After initial training, ensuring quality of anthropometric measurements, there was one measuring experimenter per participant. All anthropometric measurements were assessed in cm if not otherwise stated.

---

[3] At the editor's request, we conducted a post-hoc sensitivity power analysis in G*Power 3.1.9.7 (Faul et al., 2009) to determine the minimum detectable effect size of our study, for a 80% test power with a p-level of .05 based on our sample of 81 participants after exclusions. This analysis suggested $\rho(H1) = .31$ with critical r at $\pm.219$ (two-tailed) or $\rho(H1) = .27$ with critical r at .184 (one-tailed if expecting positive associations between UFR and hormones). An analogous analysis was run for the subsample including 49 women (two-tailed: $\rho(H1) = .39$ with critical r at $\pm.282$; one-tailed: $\rho(H1) = .35$ with critical r at $\pm.238$). As this study is the first to examine associations of UFR and circulating hormones, an informed guess regarding what effect sizes would be plausible in the real world is not possible yet. We refrain from further speculation until replication. However, the domain of sports already yielded substantial meta-analytic associations in marker research in the past ($r = -.26$ for 2D:4D and athletic prowess; Hönekopp & Schuster, 2010).

UFR We measured UFR (averaged ulna length across left and right arm/averaged fibula length across left and right leg) via reference points for ulna (caput ulnae, olecranon) and fibula length (caput fibulae, malleolus lateralis) on seated participants according to Köllner and Bleck (2020), using precision calipers (0.02 mm; for details, see Bleck, 2018). Participants were dressed during measurements, aside from removing jackets, watches, jewelry, thick socks, and shoes. During ulna measurement, they were asked to pull pullover sleeves etc. up over the elbow if possible. Given the excellent (Köllner & Bleck, 2020: 0.81 up to 0.99) to nearly perfect (Köllner et al., 2022b: 0.95 to 0.99) interrater reliabilities reported in earlier studies and given the already very resource-intensive testing protocol (two testing days, contest-setup), we did not re-establish interrater reliability separately for this particular study.

Body Circumference Measures We measured WHR (waist/hip) and SHR (shoulder/hip) according to Hughes and Gallup (2003), using a measuring tape to determine circumferences at the smallest width between iliac crest and rib cage, the largest width between thigh and waist, and the greatest width of shoulder blades. Again, as interrater reliability for our lab's established protocol is excellent also for circumference measures (0.83 to 0.91, Köllner et al., 2022b), after initial training one experimenter each took measurements from a given participant.

Salivary Hormone Concentrations

Participants collected the unstimulated saliva samples (5 ml) in sterile polypropylene plastic vials (Greiner Bio-One CELLSTAR™, 50 ml) in separate rooms. The first sample was given while working alone on computer-administered questionnaires unrelated to our research after receiving on-screen instructions, the second after the experimental contest, again in a separate room. Participants sealed the vials directly after collection and experimenters immediately froze the samples when receiving them. Samples were processed and analyzed using radioimmunoassay (ImmunoChem™ Double Antibody Testosterone, MP Biomedicals LLC, Irvin, USA; Ultra-Sensitive Estradiol, Beckman Coulter, Brea, USA) and according to validated protocols (described in Oxford et al., 2017). Saliva samples were assayed in duplicate for more reliable measurements. Lower limits of detection were 0.15 pg/ml (estradiol; $R^2 = 0.97$) and 2.61 pg/ml (testosterone; $R^2 = 0.98$). Median intra-assay CVs were 12.60% (estradiol) and 7.72% (testosterone).

Procedure

For a flow chart of the experiment, please see Fig. 1. Two same-sex (as some dominance-relevant aspects are more pronounced within-sex, Wilson, 1980) participants were scheduled to come to the lab at the same time. However, they were invited to come to different entrances of the building to prevent them from talking to each other prior to the contest. After greeting each other briefly under supervision of
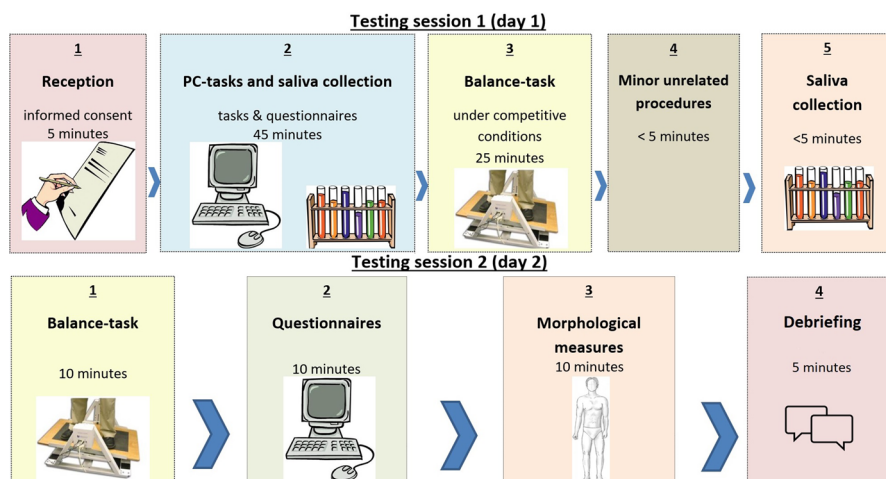
Fig. 1 Procedural flowchart of the study. *Note*. Figure available at https://osf.io/f5hzg/, under a CC-BY4.0 license

the experimenters, being led into separate rooms, and providing informed consent prior to inclusion in the study, participants completed several tests unrelated to our research questions and a demographic questionnaire.

The main task included gross-motor balance (Steib et al., 2018) learning on a stability platform (stabilometer)[4] which can tilt 20° to the left or right, with the goal being to keep it within ± 5° of the horizontal (time in balance within 30 s trials, each followed by 60 s of rest). After one test trial and 3 baseline trials, participants competed in a one-on-one contest spanning 10 training trials against their sex-matched opponent in the adjacent room, trying to score more time in balance than the other participant. The PCs displaying the win/lose-feedback were allegedly linked, but the outcome was predetermined via fake performance feedback: One participant won 8 of 10 trials (win condition) and the other one just 2 of 10 trials (lose condition; see Schultheiss et al., 2005, for a similar contest structure in an implicit learning experiment). This was followed by 3 non-competition trials to assess immediate motor learning after an additional break of three minutes after the contest trials. Finally, 4 non-competition trials (one baseline, three to assess skill retention) were run on the next day.

Saliva samples were obtained 5 min prior to and 5 min after the stabilometer task, and thus approximately 35 min apart (see Fig. 1). The post-contest sample was

---

[4] We aimed to use a competitive task that should be unrelated to gender-differences in competition behavior or stereotypic attitudes (e.g., no mathematical tasks; Dreber et al., 2013). Such "gender-neutral tasks" have been shown to eliminate gender-differences in willingness to compete (Grosse & Riener, 2010). Likewise, the task should be unrelated to pre-existing athletic skills (e.g., no common sport like cycling or soccer) and between-sex interactional dynamics (e.g., no mixed-sex dyads). Balancing on a stabilometer in same-sex dyads is a task that should be novel to most participants, not affected by gender-stereotypes, and unrelated to pre-existing skills.

obtained at least 26 min after the start and at least 11 min after the end of the 10 trials of the win-lose-manipulation. On the subsequent second testing day, along with some unrelated tasks, anthropometric measurements were performed and participants were debriefed. Participants were paid 20€ per testing day (40€ overall).

Statistical Methods

All data files, the study logbook, analysis scripts for reported results, as well as the output files are available at https://doi.org/10.17605/OSF.IO/F5HZG. We used Pearson correlations, *t*-tests, and general linear models in Systat 13. Bayesian values were computed with JASP 0.14.1.0, Spearman's rho and confidence intervals with SPSS 28.

Results

Data Preparation

UFR, height, WHR, and SHR were normally distributed.

Pre- and post-contest hormone levels were not normally distributed (*W*s ranging from 0.55 to 0.86, *p*s < 0.001) and thus logarithmically transformed (ln; cf. Schultheiss et al., 2012). Normality was heavily improved by these transformations for pre- and post-contest testosterone ($W_{pre/post} = 0.96/0.95$, $p = 0.01/0.004$) and estradiol ($W_{pre/post} = 0.93/0.83$, *p*s < 0.001). While perfect normality still could not be achieved by transformation, we side with Rasch and Guiard (2004) regarding the robustness of parametric testing and refer to the problem of bimodal distributions of sex hormones. However, to prevent including invalid cases, as an auxiliary measure we also checked for extreme values that exceed "normal" outliers or natural variations and most likely represent measurement errors. Not a single transformed hormone score exceeded ± 4 *SD*s (99.9% of the population; see Köllner et al., 2021, October 15) of the study mean of the respective sex, thus there were no indications of such extreme outliers. Finally, all significant findings regarding our basic assumptions specified above (UFR's associations to baseline and reactive testosterone/estradiol) also emerged when computing non-parametric Spearman correlations (*rho*; see below).

To obtain a reliable measure of post-contest reactive hormone levels independent from pre-existing baseline differences, post-contest scores were residualized for baseline levels and subsequently *z*-standardized.

Descriptive Statistics

Table 1 provides overall and Table 2 within-sex descriptive statistics (if a *p*-value is not accompanied by a correlation or *t*-value below, the latter can be found in the respective tables). Pronounced sex-differences were observed for body height ($p < 0.0001$) as well as baseline-testosterone ($p < 0.0001$), on average more than five

Table 1 Descriptive statistics and correlations for the overall sample

	M	SD	1	2	3	4	5	6	7	8	9
1. UFR	0.68	0.03	–								
2. Height	172.33	9.65	.08	–							
3. WHR	0.86	0.06	.25*	.38***	–						
4. SHR	1.20	0.10	.24*	.30**	.65***	–					
5. T(bl)	20.94	18.75	.24*	.57***	.54***	.60***	–				
6. T(react)	0.00	1.00	.23*	.29**	.04	–.00	.00	–			
7. E2(bl)	0.59	0.57	.04	.05	–.05	–.03	.13	–.00	–		
8. E2(react)	0.00	1.00	.22*	.04	–.11	–.06	–.01	.35**	.00	–	
9. BMI	22.60	3.08	–.05	.14	.36***	–.17	.03	–.05	.06	.11	–
10. Sex	–	–	1.71#	8.07***	7.05***	7.05***	14.64***	0.88	–0.55	–0.55	0.54

$N = 81$. Descriptive statistics are derived from the raw scores of the respective variables (e.g., ulna/fibula length in cm). Correlations and t-statistics are computed for transformed variables where necessary, which means ln-transformation for T(bl), E2(bl), and BMI; as well as ln-transformation, residualization for ln-transformed baseline levels, and subsequent z-standardization for T(react) and E2(react). UFR = ulna-to-fibula ratio of bone lengths averaged across both body sides; height = body size (cm); WHR = waist-to-hip ratio; SHR = shoulder-to-hip ratio; T(bl) = salivary testosterone level at baseline; T(react) = post-contest testosterone levels residualized for baseline; E2(bl) = salivary estradiol level at baseline (pg/ml); E2(react) = post-contest estradiol levels residualized for baseline; BMI = body mass index; sex = t-statistics for sex differences (t-tests; $df = 79$)

*** $p < .001$. ** $p < .01$. * $p < .05$. # $p < .10$

Table 2 Within-sex descriptive statistics and correlations for women ($n=49$; above diagonal) and men ($n=32$; below diagonal)

	$M_♀$	$SD_♀$	$M_♂$	$SD_♂$	1	2	3	4	5	6	7	8	9
1. UFR	0.68	0.03	0.69	0.02	–	-.02	.21	.15	.15	.33*	.08	.29*	–.10
2. Height	167.12	7.10	180.31	7.33	-.17	–	-.01	-.14	-.02	.26#	.19	.18	.24#
3. WHR	0.83	0.05	0.90	0.05	.10	-.16	–	.39**	.15	-.06	.06	-.06	.36*
4. SHR	1.15	0.07	1.27	0.08	.19	-.29	.49**	–	.27#	-.05	.18	-.10	–.42**
5. T(bl)[a]	7.97	4.41	40.81	14.26	.22	.05	-.41*	-.04	–	-.16	.39**	.07	-.01
6. T(react)	-0.08	1.16	0.12	0.68	-.12	.46**	.06	-.19	-.21	–	.02	.43**	-.05
7. E2(bl)	0.63	0.63	0.53	0.48	-.01	.01	-.18	-.26	.29	-.04	–	.07	.09
8. E2(react)	0.05	1.19	-0.08	0.61	.10	-.09	-.20	.15	.18	-.01	-.28	–	.18
9. BMI	22.49	3.34	22.77	2.68	.04	-.05	.52**	-.02	-.20	-.07	-.02	-.12	–

$N=81$, $n_♀=49$, $n_♂=32$. Descriptive statistics are derived from the raw scores of the respective variables (e.g., ulna/fibula length in cm). Correlations and t-statistics are computed for transformed variables where necessary, which means ln-transformation for T(bl), E2(bl), and BMI; as well as ln-transformation, residualization for ln-transformed baseline levels, and subsequent z-standardization for T(react) and E2(react). UFR = ulna-to-fibula ratio of bone lengths averaged across both body sides; height = body size (cm); WHR = waist-to-hip ratio; SHR = shoulder-to-hip ratio; T(bl) = salivary testosterone level at baseline (pg/ml); T(react) = post-contest testosterone levels residualized for baseline; E2(bl) = salivary estradiol level at baseline (pg/ml); E2(react) = post-contest estradiol levels residualized for baseline; BMI = body mass index

[a] The male/female baseline ratio of testosterone of 5.12 closely replicates the average sex difference found for radioimmunoassays of 5.72 reported by Schultheiss et al., (2019, Table 1), thus indicating our assay's validity

*** $p<.001$. ** $p<.01$. * $p<.05$. # $p<.10$

times higher for men than for women, but not for baseline-estradiol ($p = 0.58$) or residualized post-contest hormone levels ($ps > 0.37$; for t-tests, see Table 1). BMI, used as a basis for determining exclusions in the present study, was positively related to WHR overall and within-sexes, as well as negatively to SHR in women.

Inferential Statistics

Replication of Key Findings Regarding Marker Properties

Testing our expectations (see Tables 1 and 2 for correlations mentioned in the following), we first tried to replicate UFR's performance regarding important "quality criteria" for markers of organizational hormone effects (see Köllner et al., 2022b). As expected, UFR was sex-dimorphic (for t-test see Table 1; $d = 0.39$) but only at non-significant trend-level. UFR was unrelated to body height ($BF_{01} = 5.58$) overall and within-sex, and significantly related to WHR and SHR, other potential markers of pubertal organizational hormone effects.

UFR's Hormonal Associations

Most importantly, UFR was significantly positively associated with baseline-testosterone ($p = 0.03$; $rho = 0.22$, $p < 0.05$; 95% CI [0.03, 0.44]), but not baseline-estradiol ($p = 0.72$; $rho = 0.01$, $p = 0.94$; 95% CI [-0.18, 0.26]). This overall testosterone-related finding was not simply due to general sex-differences in testosterone, as the non-significant within-sex findings for women and men were similar in magnitude and direction (Fig. 2).
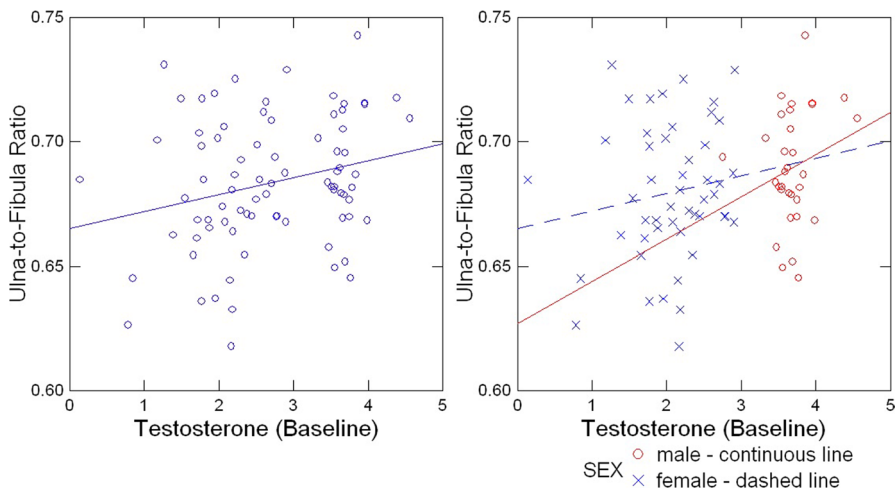

Fig. 2 Relationship between salivary testosterone and ulna-to-fibula ratio (UFR) in the whole sample (left) and dependent on sex (right). *Note*. $N = 81$, $n_{♀} = 49$, $n_{♂} = 32$. Testosterone levels (pg/ml) were ln-transformed. Figure available at https://osf.io/f5hzg/, under a CC-BY4.0 license

UFR also showed significant positive associations to reactive testosterone ($p = 0.04$; $rho = 0.26$, $p = 0.02$; 95% CI [0.02, 0.43]) and estradiol ($p = 0.04$; see Fig. 3; $rho = 0.25$, $p = 0.03$; 95% CI [0.01, 0.42]) overall and in women (testosterone in women: $p = 0.02$; $rho = 0.35$, $p = 0.01$; 95% CI [0.05, 0.56]; estradiol in women: $p < 0.05$; $rho = 0.32$, $p = 0.03$; 95% CI [0.01, 0.53]).

When additionally predicting UFR from feedback condition[5] (win vs. lose) and reactive hormone levels, there were no Condition x Reactive Hormones findings, neither overall ($Fs(1,77) < 0.08$, $ps > 0.78$, $\eta_p^2s < 0.002$) nor for women ($Fs(1, 45) < 0.34$, $ps > 0.56$, $\eta_p^2s < 0.008$), nor for men ($Fs(1, 28) < 1.40$, $ps > 0.24$, $\eta_p^2s < 0.048$). Figure 3 illustrates the absence of an effect of condition for both hormones. Thus, winning or losing seemed to have no influence on the relationship between UFR and contest-induced reactive hormone surges.

Exploratory Analyses

While WHR and SHR were also sex-dimorphic ($p < 0.0001$), interrelated among each other ($p < 0.0001$), and related to baseline-testosterone ($ps < 0.0001$), they were not independent from overall body height ($ps < 0.01$) and did not show any connections to reactive hormone surges in challenge situations ($ps > 0.34$, see Table 1 for t-tests and correlations).

Additional Robustness Checks

Pitfalls of a Mixed-Sex Sample Considering our mixed-sex sample, we applied additional robustness checks for our main overall findings while aiming to use the sample's full power in the analyses. First, we additionally predicted UFR from respective sex hormone while entering sex as a covariate. While the effect was not significant anymore for baseline-testosterone ($F(1, 78) = 2.13$, $p = 0.14$, $\eta_p^2 = 0.027$), both reactive results for testosterone ($F(1, 78) = 4.00$, $p < 0.05$, $\eta_p^2 = 0.049$) and estradiol ($F(1, 78) = 4.78$, $p = 0.03$, $\eta_p^2 = 0.058$) were preserved, respectively. Second, we z-standardized hormone-levels within sex and then used the resulting z-scores in the analyses. Again, the association between UFR and baseline-testosterone was not significant anymore ($r = 0.17$, $p = 0.12$) while both reactive results for testosterone ($r = 0.19$, $p = 0.09$) and estradiol ($r = 0.22$, $p = 0.05$) were preserved, however at non-significant trend-level. Thus, it seems that our reactive findings were more robust than the testosterone-related baseline findings.

---

[5] To check if female and male hormone changes were in the expected directions (e.g., based on Mazur, 1985; Schultheiss et al., 2005; Stanton & Schultheiss, 2007), we looked at the descriptive post-contest hormone scores residualized for baseline. This was done on a descriptive level, given small cell sizes of 15 to 25 participants. Expected increases in winners emerged for testosterone in men ($M = 0.18$, $SD = 0.67$) and estradiol in women ($M = 0.08$, $SD = 1.43$). However, corresponding decreases in losers were not visible, as the residualized scores were positive and near zero for testosterone in men ($M = 0.06$, $SD = 0.70$) and estradiol in women ($M = 0.02$, $SD = 0.92$). While many studies fail to find winner-loser-differences predicted by the biosocial model of status (see review on testosterone by Carré & Olmstead, 2016), in spite of our small cell sizes, we see some descriptive differences between conditions.
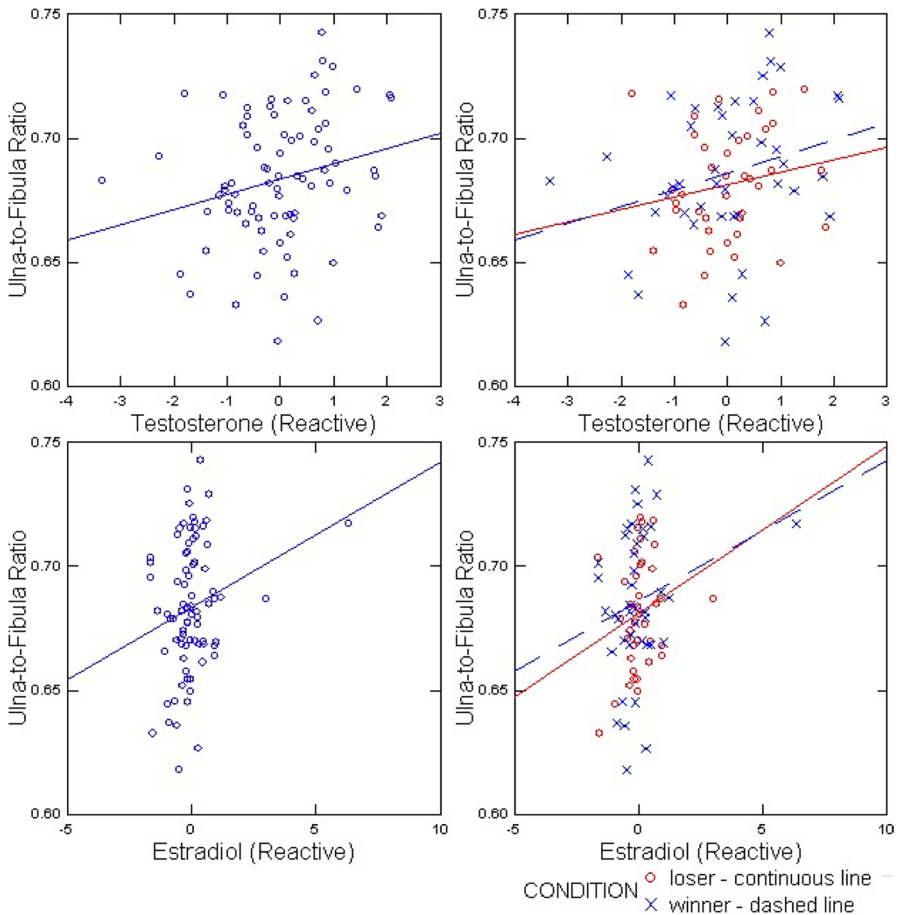
Fig. 3 Relationship between reactive hormone levels and ulna-to-fibula ratio (UFR) in the whole sample (left) and dependent on experimentally manipulated contest outcome (right). *Note*. $N=81$, $n_{loser}=39$, $n_{winner}=42$. Ln-transformed post-contest scores residualized for ln-transformed baseline levels. Please note that the most rightward data point for estradiol is no regression-based outlier, neither overall nor when adding condition as a moderator. Figure available at https://osf.io/f5hzg/, under a CC-BY4.0 license

Applying the above-mentioned approach of adding sex as a covariate also to UFR's marker associations (WHR, SHR), we found the corresponding associations slightly reduced and non-significant (Fs$(1, 78)>2.12$, $ps<0.15$, η_p^2s>0.026).

Outliers To make sure our findings were not driven by outliers, we reran all correlations of UFR and baseline or reactive hormone levels overall and within-sex in general linear models. No regression-based outliers were flagged in any case, thus our results were robust.

However, as an auxiliary check, looking at possible influential data points, we also tested if there were indications of possibly wrong measurements that were

not detected by our previous method of identifying extreme outliers using ± 4 SDs within-sex on baseline and post-contest scores. Applying the same method to residualized post-contest scores, as can be expected from the two bottom-graphs in Fig. 3, there was one data point not fulfilling this criterion, which was also conspicuously atypical in the overall distribution ($+6.33$ SDs). Removing this data point, while leaving the relationship to reactive testosterone intact (overall: $r = 0.27$, $p = 0.02$, women: $r = 0.36$, $p = 0.01$), reduced reactive estradiol's relationship with UFR to slightly below non-significant trend-level (overall: $r = 0.17$, $p = 0.12$, women: $r = 0.21$, $p = 0.15$). Again, there was no Condition x Reactive Hormones interaction, neither overall ($F(1,76) = 0.01$, $p = 0.93$, $\eta_p^2 < 0.001$) nor for women ($F(1, 44) = 0.50$, $p = 0.48$, $\eta_p^2 = 0.011$). Figure 4 shows our findings omitting the data point in question. However, we have to note that this influential case nevertheless was a valid data point regarding the constituting pre- and post-contest scores and also considering our inclusion criteria.

FDR Correction At the editor's request, due to the possible high Type-I-error-rate, we added FDR-corrections. The results of those corrections can be found at https://doi.org/10.17605/OSF.IO/F5HZG in two excel-sheets using the template provided by Pike (2011) and two different bases for making our corrections.

One sheet ("UFR_hormones_FDR") uses the core of our expectations stated in the last paragraph of the introduction of the originally submitted manuscript (changes were made to this paragraph during peer review), which were baseline and reactive associations between UFR and the respective hormones (estradiol, testosterone), UFR's dimorphism, independence from height (which was not expected to
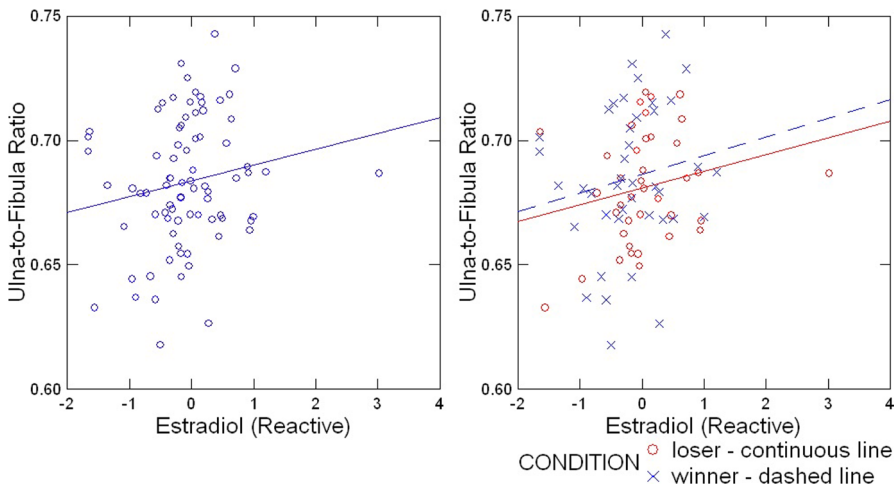

Fig. 4 Relationship between reactive estradiol levels and ulna-to-fibula ratio (UFR) in the whole sample (left) and dependent on experimentally manipulated contest outcome (right) after removing a conspicuously anomalous data point. *Note.* $N = 80$, $n_{loser} = 39$, $n_{winner} = 41$. Ln-transformed post-contest scores residualized for ln-transformed baseline levels. Figure available at https://osf.io/f5hzg/, under a CC-BY4.0 license

be significant anyway), and relationships to WHR and UFR, respectively (8 tests). Based on this, our key results hold up at least according to the graphically sharpened method, aside from UFR's dimorphism and relationship to SHR.

The other sheet uses a more conservative approach ("UFR_hormones_FDR_all") which includes all mentioned relationships not only overall as stated in the original manuscripts introduction, but also within-sex in addition to testing UFR's dimorphism (22 tests). Here, our results would not be significant anymore according to the corrections, independent of applied FDR-correction-method.

Other Checks As estradiol is synthesized from testosterone (Vanderschueren et al., 2004) and assays may feature cross-reactivity, we additionally checked whether the overall associations persisted when controlling for the other hormone, respectively. Both associations were reduced ($\beta_{\text{testosterone/estradiol}} = 0.18/0.16$, $p_{\text{testosterone/estradiol}} = 0.13/0.17$, $\Delta R^2_{\text{testosterone/estradiol}} = 0.028/0.023$), thus seemingly not driven by testosterone or estradiol alone, but by their joint action.

Discussion

Confirming earlier findings, we replicated UFR's dimorphism, unrelatedness to body height, and correlations to WHR and SHR. UFR was associated with baseline-testosterone overall. Within-sex relationships were similar in magnitude, but non-significant, probably due to lacking test power. Relationships to contest-induced reactive steroid hormone surges emerged for testosterone and estradiol (overall and in women). These relationships were not driven by the action of one hormone alone, not moderated by experimentally manipulated contest-outcome (winning or losing), and related regressions contained no flagged outliers. However, removing a data point with extreme estradiol reactivity reduced our estradiol-related findings to slightly below non-significant trend-level.

Interpretation

UFR's Hormonal Associations

UFR's relationships to hormone reactivity were not exclusively attributable to actions of testosterone or estradiol alone and not influenced by contest outcome (winning/losing). Regarding the latter finding, our sample's low overall power precludes premature conclusions regarding (non-)association of pubertal organizational hormone effects with later endocrine reactions to dominance contest outcomes as predicted by the biosocial model of status (Mazur, 1985). Preliminarily, UFR may be interpreted as an indicator of general, non-specific contest-induced hormone reactivity as an outcome of pubertal organizational hormone effects, possibly facilitating challenging encounters (cf. Archer, 2006) and thus apt to maintain or achieve high status in later adult life (Mazur & Booth, 1998). In other words, UFR may

approximate pubertal organizational hormone effects that modulate (or prepare) the endocrine system's functioning in challenging, dominance-relevant situations later in adulthood (see Manning et al., 2014, for a related argument for 2D:4D). Within-sex, our result remained significant only in women, but replication is needed before speculating regarding this additional result.

Overall, given that our findings on reactive estradiol dropped below trend-level when excluding a conspicuous – however valid—data point, it may be tempting to limit our conclusions to reactive testosterone only. This would be in accordance with the original testosterone-centered challenge hypothesis (Archer, 2006), but may be premature, as our sample was comparatively small and as recent research hints at links between estradiol and female competition (contest outcomes, Stanton & Schultheiss, 2007; intrasexual competition, Fiacco et al., 2021; sports, Edwards & Turan, 2020).

Interpreting the association between UFR and baseline-testosterone is a more complex task. For prenatal markers like 2D:4D, observing associations to circulating baseline hormones would allow no differentiation between current and organizational effects and thus threaten the "marker" status for a previous developmental window (Hönekopp et al., 2007). For a potential pubertal marker like UFR, the case is different, as sex steroids rise in puberty (also causing pubertal organizational hormone effects) and remain stable afterwards (see model by Schulz & Sisk, 2016, Fig. 1) until well into middle adulthood, even though we are not aware of longitudinal studies on individual stability of high levels of steroid hormone levels after puberty. Thus, continued associations between UFR and adult hormone levels are plausible for our young-adulthood-sample.

While the findings regarding baseline hormone levels were preserved when using non-parametric testing, they were, unlike the findings regarding reactive hormone surges, not robust to using sex as a covariate or computations with z-scores, standardized within sex. Further, baseline relationships between UFR and hormones may have been diminished or obscured by diurnal variations in hormone levels (e.g. Schultheiss et al., 2012), as restricting time of testing to specific time windows was technically not feasible. The main study required the simultaneous presence of two experimenters and two participants on two consecutive testing days (24 h apart) and thus the schedules of four persons had to be fitted together for each session cycle.

In general, our results should only be regarded as first evidence for UFR's hormonal associations. Larger samples are needed before drawing far-reaching conclusions. Past research has shown that marker research often captures real, but small-in-size effects on development that are difficult to demonstrate given the restricted sample sizes accessible to (psychological) research. For example, digit ratio's association with receptor genes was first denied, with a meta-analysis yielding no evidence for a role of androgen receptor gene efficacy (Voracek, 2014; $N = 2.157$). However, a considerably larger recent meta-analytic genome-wide association study identified genetic loci for digit ratio as well as associations with female androgen receptor sensitivity (Warrington et al., 2018; $N = 15.661$). Applied to our comparatively small sample comprising 81 individuals (for the issues associated with low power, see Button et al., 2013), we currently cannot say with certainty if our observed effects are real. In turn, the absence of findings for example regarding associations between

UFR and baseline estradiol could alternatively be due to the high risk of Type II errors in this study and does not necessarily mean that the picture will not be different in a larger-sample-replication.

UFR's Marker Validity

UFR's only marginal sexual dimorphism likely is due to our small sample as well, as Cohen's *d* is comparable to earlier studies (Köllner & Bleck, 2020; Köllner et al., 2022b). UFR was again unrelated to overall body height, and thus can be considered an indicator of sex-steroid-dependent differential long bone growth, not just representing an artifact of general body-size-dependent bone growth patterns (cf. Köllner et al., 2022b; Leslie, 2019). Convergent validity with other pubertally determined body dimensions was replicated (Köllner et al., 2022b) for the overall sample, adding to the emerging picture that UFR's marker function applies to puberty in particular. However, findings did not persist when using sex as a covariate and not within-sex—likely due to lacking test power, as the non-significant within-sex correlations kept their direction.

Other Potential Markers

While WHR and SHR were also sex-dimorphic, their relationships to salivary hormone levels were, except for baseline-testosterone, absent. This finding also held within-sex, where baseline testosterone showed a non-significant positive trend-level association with SHR in women, and a negative association with WHR in men. SHR and WHR are largely dependent on body fat distribution (Fink et al., 2003), which in turn may be influenced by various post-pubertal factors like nutrition in a stronger way than a more bone-based marker like UFR. In contrast to UFR, SHR and WHR were related to body size and thus possibly partly artifacts of sex-differences in overall size (cf. Leslie, 2019). Overall, we recommend UFR instead of WHR and SHR when estimating pubertal organizational hormone effects.

Limitations and Future Directions

Robustness of Results Our study calls for replication for several reasons. First, our hypotheses were not preregistered, as originally we had no funding available for radioimmunoassays and thus we could not foresee whether it would be possible to analyze the collected saliva samples at all.[6] Second, as a related issue, as mentioned above our sample size was comparatively small, especially for establishing a new discovery with unknown effect size—UFR's hormonal association—for the first time and given the small effect sizes observed here. As stated above, the study was part of a larger one-on-one contest study, involving the necessity for simultaneous

---

[6] See Declarations.

presence of two experimenters and two consecutive testing days, severely limiting the possible observations within our available funding and research resources.

Overall, strong conclusions are not possible based on this study alone. When FDR-correcting for Type I error, our main results were not preserved. It may be premature to discard the results of the present exploratory study as a whole based on this fact alone given that those conservative corrections even failed in replicating some UFR-related findings that clearly emerged in a high-powered study (associations of UFR to other markers; Köllner et al., 2022b) or even repeatedly (UFR's dimorphism; Köllner & Bleck, 2020; Köllner et al., 2022b). But this failure leads to the above-mentioned additional problem of our study, namely that Type II error problems further endangered drawing valid conclusions.

Nevertheless, there were some signs of statistical robustness of our findings which were previously checked for wrong measurements (± 4 *SD*s): Our main results were fully preserved when applying non-parametric testing and the reactive findings were robust to entering sex as a covariate or (at trend-level) even to using within-sex-standardized scores in the analyses. In conclusion, while not discarding our findings as unreliable altogether, we encourage replication in well-powered samples.

Limitations of Salivary Immunoassays Salivary hormone radioimmunoassays like those used in our study were our method of choice due to the availability of funding and a specialized lab at our institution. Immunoassays, including radioimmunoassays, were criticized for yielding results inferior to other methods like serum analysis or tandem mass spectrometry, and seem to have low validity for example for estimating menstrual cycle phase (see empirical comparison of assessment methods by Arslan et al., 2022). Nevertheless, problems with measurement validity have been studied more intensively for enzymatic immunoassays (e.g., Welker et al., 2016b; see Chafkin et al., 2022, for an expansion of this work to chemiluminescent immunoassays, which seem to overestimate testosterone and cortisol and suppress testosterone sex differences when compared to LC–MS/MS). The latter are more susceptible to diverse biochemical parameters than the radioimmunoassays used by us, which use radiation as a physical parameter for analysis (Schultheiss et al., 2019). Thus, while our salivary hormone measurement may not have been ideal and probably also has its limitations, it still likely was a superior choice compared to the widely used enzymatic assays. However, future research projects similar to those for enzymatic and chemiluminescent assays should compare radioimmunoassays to LC–MS/MS as a gold-standard. Where feasible, mass-spectrometry-based (or blood-sample-based, see good performance of serum assessment compared to immunoassays in Arslan et al., 2022) assessment can be used to improve measurement validity.

Other Limitations As another limitation regarding measurement procedures, the practice of relying on self-reported height and only measuring actual height in case of doubt may have introduced some bias in our height variable (e.g., reporting errors, experimenter's judgement). Likewise, we did not consider hormonal fluctuations due to menstrual cycle stage in our analyses, as we had no objective measure available and a simple day counting method is problematic for several reasons, for

example due to many individuals not tracking their cycle appropriately or applying the day count incorrectly (Hampson, 2020).

In addition, while hormonal associations in adulthood add an important piece of evidence to UFR's validation process, developmental studies are indispensable to validate a marker. Future longitudinal research should pinpoint UFR's pubertal origin: Attributability to a specific developmental window is an under-researched criterion for pubertal markers, requiring longitudinal studies (see Lutchmaya et al., 2004) assessing body dimensions, hormonal status, and behavioral and psychological outcomes in prepubertal, peripubertal, and adult participants (cf. Köllner et al., 2021, October 15). Unequivocally proving a marker's pubertal origin requires analyzing the developmental time course of marker development vis-à-vis the developmental increase in steroid hormones and the emergence of marker relationships with behavior.

As soon as UFR's validity has been established, it should be tested whether it is solely a marker in the sense of a mere hormonal by-product of pubertal organizational hormone effects that can be used to estimate them, or if it has additional implications on its own. For example, it is conceivable that it may be related to sports performance regarding upper-body-strength (cf. meta-analytic relationships between 2D:4D and athletic ability; Hönekopp & Schuster, 2010). Also, it may be a signaling cue of threat or dominance to the observer (see ratings of observers of fWHR; meta-analysis by Geniole et al., 2015).

Conclusions

In conclusion, our explorative results for the first time hint at UFR's relationship to baseline-testosterone and also indicate functional connections between outcomes of pubertal organizational hormone effects and contest-induced steroid reactivity overall and in women. Along with UFR's repeatedly demonstrated sexual dimorphism, this may provide an advantage over the frequently-used fWHR marker measure that does not share these properties (Bird et al., 2016), should our observed relationships replicate in larger samples. Especially, if UFR's relationship to circulating hormones should re-emerge in other dominance or status-contest setups, this would support the notion that also pubertal organizational hormone effects prepare the adult endocrine system for dominance and status contests (see Manning et al., 2014, for a related argument for prenatal organizational hormone effects). However, the small sample combined with explorative testing, failing some of the applied robustness checks (e.g., FDR-correction) precludes any strong conclusions until replication.

(equal). Hanna Schöttner: Investigation (equal); writing – review and editing (equal). Simon Steib: Conceptualization (equal); writing – review and editing (equal).

Data availability The data that support the findings (Köllner et al., 2022a) of this study are openly available at the Open Science Framework (OSF) at https://doi.org/10.17605/OSF.IO/F5HZG. Additionally, the logbook, analysis scripts, and output files are also available there.

Declarations

The conducted experiment complies with the current laws of Germany and the Declaration of Helsinki. All participants gave their informed consent prior to their inclusion in the study. The manuscript does not contain clinical studies or patient data. Procedures were approved by the ethics committee of the Friedrich-Alexander University in Erlangen (No. 197_18 B).

Conflict of interest The authors declare that they have no conflict of interest. They have no relevant financial or non-financial interests to disclose.

References

Köllner, M. G., Braun, S., Schöttner, H., Dlugash, G., Bettac, M., & Steib, S. (2022a). *Relationships of the ulna-to-fibula ratio to baseline and reactive steroid hormone levels*. Open Science Framework. https://doi.org/10.17605/OSF.IO/F5HZG

Archer, J. (2006). Testosterone and human aggression: An evaluation of the challenge hypothesis. *Neuroscience and Biobehavioral Reviews, 30*(3), 319–345. https://doi.org/10.1016/j.neubiorev.2004.12.007

Arnold, A. P. (2009). The organizational–activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Hormones and Behavior, 55*, 570–578. https://doi.org/10.1016/j.yhbeh.2009.03.011

Arslan, R. C., Blake, K., Botzet, L., Bürkner, P., DeBruine, L. M., Fiers, T., … Stern, J. (2022). Not within spitting distance: salivary immunoassays of estradiol have subpar validity for cycle phase. https://doi.org/10.31234/osf.io/5r8mg

Beltz, A. M., Demidenko, M. I., Wilson, S. J., & Berenbaum, S. A. (2021). Prenatal androgen influences on the brain: A review, critique, and illustration of research on congenital adrenal hyperplasia. *Journal of Neuroscience Research*. https://doi.org/10.1002/jnr.24900

Bird, B. M., Cid Jofré, V. S., Geniole, S. N., Welker, K. M., Zilioli, S., Maestripieri, D., Arnocky, S., & Carré, J. M. (2016). Does the facial width-to-height ratio map onto variability in men's testosterone concentrations? *Evolution and Human Behavior, 37*, 392–398. https://doi.org/10.1016/j.evolhumbehav.2016.03.004

Bleck, K. (2018). *Measuring lengths of ulna and fibula in living individuals.* In http://www.psych2.phil.uni-erlangen.de/%7Eoschult/humanlab/resources/resources_Ulna_Fibula.htm. Accessed 30 July 2022

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365–376. https://doi.org/10.1038/nrn3475

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. https://doi.org/10.1098/rspb.2008.0873

Carré, J. M., & Olmstead, N. A. (2015). Social Neuroendocrinology of human aggression: Examining the role of competition-induced testosterone dynamics. *Neuroscience, 286*, 171–186. https://doi.org/10.1016/j.neuroscience.2014.11.029

Chafkin, J. E., O'Brien, J. M., Medrano, F. N., Lee, H. Y., Yeager, D. S., & Josephs, R. A. (2022). Chemiluminescent immunoassay overestimates hormone concentrations and obscures testosterone sex differences relative to LC-MS/MS in a field study of diverse adolescents. *Comprehensive Psychoneuroendocrinology, 10*, 100132. https://doi.org/10.1016/j.cpnec.2022.100132

Chang, S., Skakkebæk, A., Trolle, C., Bojesen, A., Hertz, J. M., Cohen, A., Hougaard, D. M., Wallentin, M., Pedersen, A. D., Østergaard, J. R., & Gravholt, C. H. (2015). Anthropometry in Klinefelter syndrome - multifactorial influences due to CAG length, testosterone treatment and possibly intrauterine hypogonadism. *Journal of Clinical Endocrinology and Metabolism, 100*(3), E508–E517. https://doi.org/10.1210/jc.2014-2834

Cutler, G. B. (1997). The role of estrogen in bone growth and maturation during childhood and adolescence. *The Journal of Steroid Biochemistry and Molecular Biology, 61*(3–6), 141–144. https://doi.org/10.1016/S0960-0760(97)80005-2

Dimitrakakis, C., & Bondy, C. (2009). Androgens and the breast. *Breast Cancer Research, 11*(5), 212. https://doi.org/10.1186/bcr2413

Doll, L. M., Cárdenas, R. A., Burriss, R. P., & Puts, D. A. (2016). Sexual Selection and Life History: Earlier Recalled Puberty Predicts Men's Phenotypic Masculinization. *Adaptive Human Behavior and Physiology, 2*, 134–149. https://doi.org/10.1007/s40750-015-0031-7

Dreber, A., von Essen, E., & Ranehill, E. (2014). Gender and competition in adolescence: Task matters. *Experimental Economics, 17*, 154–172. https://doi.org/10.1007/s10683-013-9361-0

Edwards, D. A., & Turan, B. (2020). Within-person coupling of estradiol, testosterone, and cortisol in women athletes. *PeerJ, 8*, e8402. https://doi.org/10.7717/peerj.8402

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fiacco, S., Arpagaus, C., Mernone, L., & Éhlert, U. (2021). Female Intrasexual Competition and Its Link to Menopausal Stage, Sex Hormone Levels, and Personality Characteristics. *Frontiers in Global Women's Health, 2*(740894). https://doi.org/10.3389/fgwh.2021.740894

Fink, B., Neave, N., & Manning, J. T. (2003). Second to fourth digit ratio, body mass index, waist-to-hip ratio, and waist-to-chest ratio: Their relationships in heterosexual men and women. *Annals of Human Biology, 30*(6), 728–738. https://doi.org/10.1080/03014460310001620153

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

Grosse, N. D., & Riener, G. (2010). Explaining Gender Differences in Competitiveness: Gender-Task Stereotypes. *Jena Economic Research Papers, 2010*(017). FriedrichSchiller University Jena and Max Planck Institute of Economics, Jena. Retrieved from: http://hdl.handle.net/10419/32599. Accessed 30 July 2022

Hamann, S., Stevens, J., Hassett Vick, J., Bryk, K., Quigley, C. A., Berenbaum, S. A., & Wallen, K. (2014). Brain responses to sexual images in 46, XY women with complete androgen insensitivity syndrome are female-typical. *Hormones and Behavior, 66*(5), 724–730. https://doi.org/10.1016/j.yhbeh.2014.09.013

Hampson, E. (2020). A brief guide to the menstrual cycle and oral contraceptive use for researchers in behavioral endocrinology. *Hormones and Behavior, 119*, 104655. https://doi.org/10.1016/j.yhbeh.2019.104655

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), e0122637. https://doi.org/10.1371/journal.pone.0122637

Hönekopp, J., & Schuster, M. (2010). A meta-analysis on 2D:4D and athletic prowess: Substantial relationships but neither hand out-predicts the other. *Personality and Individual Differences, 48*(1), 4–10. https://doi.org/10.1016/j.paid.2009.08.009

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*, 313–321. https://doi.org/10.1016/j.psyneuen.2007.01.007

Hughes, S. M., & Gallup, G. G. (2003). Sex differences in morphological predictors of sexual behavior: Shoulder to hip and waist to hip ratios. *Evolution and Human Behavior, 24*(3), 173–178. https://doi.org/10.1016/S1090-5138(02)00149-6

Juul, A. (2001). The effects of oestrogens on linear bone growth. *Human Reproduction Update, 7*(3), 303–313. https://doi.org/10.1111/j.1600-0463.2001.tb05758.x

Kilduff, L., Cook, C. J., Bennett, M., Crewther, B., Bracken, R. M., & Manning, J. (2013). Right–left digit ratio (2D:4D) predicts free testosterone levels associated with a physical challenge. *Journal of Sports Sciences, 31*(6), 677–683. https://doi.org/10.1080/02640414.2012.747690

Knight, B. (1991). *Forensic Pathology*. Edward Arnold.

Köllner, M. G., & Bleck, K. (2020). Exploratory evidence of sex-dimorphic associations of the ulna-to-fibula ratio, a potential marker of pubertal sex steroid exposure, with the implicit need for power. *Adaptive Human Behavior and Physiology, 6*, 93–118. https://doi.org/10.1007/s40750-020-00130-8

Köllner, M. G., Pülschen, L.-S., Stamm, L., & Janson, K. T. (2022b). The ulna-to-fibula ratio as a marker of organizational hormone effects on implicit motive development: A high-powered preregistered replication. *Motivation Science*. Advance online publication. https://doi.org/10.1037/mot0000271

Köllner, M. G., Janson, K. T., & Schultheiss, O. C. (2018). Commentary: Sexual dimorphism of facial width-to-height ratio in human skulls and faces: A meta-analytical approach. *Frontiers in Endocrinology, 9*(227). https://doi.org/10.3389/fendo.2018.00227

Köllner, M. G., Pülschen, L.-S., Stamm, L., & Janson, K. T. (2021, October 15). *The ulna-to-fibula ratio as a marker of organizational hormone effects on the motivational brain: A high-powered preregistered replication*. PsyArXiv. https://doi.org/10.31234/osf.io/hskvy

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

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

Leslie, M. (2019). The mismeasure of hands. *Science, 364*(6444), 923–925. https://doi.org/10.1126/science.364.6444.923

Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R., & Manning, J. T. (2004). 2nd to 4th digit ratios, fetal testosterone and estradiol. *Early Human Development, 77*(1–2), 23–28. https://doi.org/10.1016/j.earlhumdev.2003.12.002

Manning, J. T., Kilduff, L. P., & Trivers, R. (2013). Digit ratio (2D:4D) in Klinefelter's syndrome. *Andrology, 1*, 94–99. https://doi.org/10.1111/j.2047-2927.2012.00013.x

Manning, J. T., Kilduff, L., Cook, C., Crewther, B., & Fink, B. (2014). Digit ratio (2D:4D): A biomarker for prenatal sex steroids and adult sex steroids in challenge situations. *Frontiers in Endocrinology, 5*(9). https://doi.org/10.3389/fendo.2014.00009

Manning, J. T. (2002). *Digit ratio. A pointer to fertility, behavior, and health*. Rutgers University Press.

Martin, J. T., & Nguyen, D. H. (2004). Anthropometric analysis of homosexuals and heterosexuals: Implications for early hormone exposure. *Hormones and Behavior, 45*(1), 31–39. https://doi.org/10.1016/j.yhbeh.2003.07.003

Mazur, A. (1985). A biosocial model of status in face-to-face primate groups. *Social Forces, 64*(2), 377–402. https://doi.org/10.1093/sf/64.2.377

Mazur, A., & Booth, A. (1998). Testosterone and dominance in men. *Behavioral and Brain Sciences, 21*(3), 353–397. https://doi.org/10.1017/S0140525X98001228

McCormick, C. M., & Carré, J. M. (2020). Facing off with the phalangeal phenomenon and editorial policies: A commentary on Swift-Gallant, Johnson, Di Rita and Breedlove (2020). *Hormones and Behavior, 120*(104710). https://doi.org/10.1016/j.yhbeh.2020.104710

McFadden, D., & Champlin, C. A. (2000). Comparison of Auditory Evoked Potentials in Heterosexual, Homosexual, and Bisexual Males and Females. *JARO, 1*, 89–99. https://doi.org/10.1007/s101620010008

McFadden, D., & Pasanen, E. G. (1998). Comparison of the auditory systems of heterosexuals and homosexuals: Click-evoked otoacoustic emissions. *PNAS, 95*(5), 2709–2713. https://doi.org/10.1073/pnas.95.5.2709

Nelson, R. J., & Kriegsfeld, L. J. (2017). *An introduction to behavioral endocrinology* (5th ed.). Sinauer Associates.

Nielsen, T. L., Hagen, C., Wraae, K., Bathum, L., Larsen, R., Brixen, K., & Andersen, M. (2010). The impact of the CAG repeat polymorphism of the androgen receptor gene on muscle and adipose tissues in 20–29-year-old Danish men: Odense Androgen Study. *European Journal of Endocrinology, 162*(4), 795–804. https://doi.org/10.1530/EJE-09-0763

Ober, C., Loisel, D. A., & Gilad, Y. (2008). Sex-specific genetic architecture of human disease. *Nature Reviews Genetics, 9*, 911–922. https://doi.org/10.1038/nrg2415

Oxford, J. K., Tiedtke, J. M., Ossmann, A., Özbe, D., & Schultheiss, O. C. (2017). Endocrine and aggressive responses to competition are moderated by contest outcome, gender, individual versus team competition, and implicit motives. *PLoS ONE, 12*(7), e0181610. https://doi.org/10.1371/journal.pone.0181610

Phoenix, C. H., Goy, R. W., Gerall, A. A., & Young, W. C. (1959). Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. *Endocrinology, 65*(3), 369–382. https://doi.org/10.1210/endo-65-3-369

Pike, N. (2011). Using false discovery rates for multiple comparisons in ecology and evolution. *Methods in Ecology and Evolution, 2*(3), 278–282. https://doi.org/10.1111/j.2041-210X.2010.00061.x

Purkait, R. (2001). Measurements of ulna—A new method for determination of sex. *Journal of Forensic Sciences, 46*(4), 924–927. https://doi.org/10.1520/JFS15071J

Rasch, D., & Guiard, V. (2004). The robustness of parametric statistical methods. *Psychology Science, 46*(2), 175–208.

Schultheiss, O. C., Wirth, M. M., Torges, C. M., Pang, J. S., Villacorta, M. A., & Welsh, K. M. (2005). Effects of implicit power motivation on men's and women's implicit learning and testosterone changes after social victory or defeat. *Journal of Personality and Social Psychology, 88*(1), 174–188. https://doi.org/10.1037/0022-3514.88.1.174

Schultheiss, O. C., Dlugash, G., & Mehta, P. H. (2019). Hormone measurement in social neuroendocrinology: A comparison of immunoassay and mass spectrometry methods. In O. C. Schultheiss & P. H. Mehta (Eds.), *Routledge international handbook of social neuroendocrinology* (pp. 26–41). Routledge.

Schultheiss, O. C., Schiepe, A., & Rawolle, M. (2012). Hormone assays. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *Handbook of Research Methods in Psychology* (Vol. 1: Foundations, planning, measures, and psychometrics, pp. 489–500). American Psychological Association.

Schulz, K. M., & Sisk, C. L. (2016). The organizing actions of adolescent gonadal steroid hormones on brain and behavioral development. *Neuroscience and Biobehavioral Reviews, 70*, 148–158. https://doi.org/10.1016/j.neubiorev.2016.07.036

Schulz, K. M., Molenda-Figueira, H. A., & Sisk, C. L. (2009). Back to the future: The organizational-activational hypothesis adapted to puberty and adolescence. *Hormones and Behavior, 55*(5), 597–604. https://doi.org/10.1016/j.yhbeh.2009.03.010

Shirazi, T. N., Self, H., Cantor, J., Dawood, K., Cárdenas, R., Rosenfield, K., Oritz, T., Carré, J. M., McDaniel, M. A., Blanchard, R., Balasubramanian, R., Delaney, A., Crowley, W., Breedlove, S. M., & Puts, D. A. (2020a). Timing of peripubertal steroid exposure predicts visuospatial cognition in men: Evidence from three samples. *Hormones and Behavior, 121*, 104712. https://doi.org/10.1016/j.yhbeh.2020.104712

Shirazi, T. N., Self, H., Dawood, K., Cárdenas, R., Welling, L. L. M., Rosenfield, K. A., Oritz, T. L., Carré, J. M., Bailey, J. M., Balasubramanian, R., Delaney, A., Crowley, W., Breedlove, S. M., &

Puts, D. A. (2020b). Pubertal timing predicts adult psychosexuality: Evidence from typically developing adults and adults with isolated GnRH deficiency. *Psychoneuroendocrinology, 119*, 104733. https://doi.org/10.1016/j.psyneuen.2020.104733

Shirazi, T. N., Self, H., Dawood, K., Welling, L. L. M., Cárdenas, R., Rosenfield, K. A., Bailey, J. M., Balasubramanian, R., Delaney, A., Breedlove, S. M., & Puts, D. A. (2021). Evidence that perinatal ovarian hormones promote women's sexual attraction to men. *Psychoneuroendocrinology, 134*, 105431. https://doi.org/10.1016/j.psyneuen.2021.105431

Shirazi, T. N., Self, H., Rosenfield, K. A., Dawood, K., Welling, L. L. M., Cárdenas, R., Bailey, J. M., Balasubramanian, R., Delaney, A., Breedlove, S. M., & Puts, D. A. (2022). Low Perinatal Androgens Predict Recalled Childhood Gender Nonconformity in Men. *Psychological Science, 33*(3), 343–353. https://doi.org/10.1177/09567976211036075

Stanton, S. J., & Schultheiss, O. C. (2007). Basal and dynamic relationships between implicit power motivation and estradiol in women. *Hormones and Behavior, 52*(5), 571–580. https://doi.org/10.1016/j.yhbeh.2007.07.002

Steib, S., Wanner, P., Adler, W., Winkler, J., Klucken, J., & Pfeifer, K. (2018). A single bout of aerobic exercise improves motor skill consolidation in Parkinson's disease. *Frontiers in Aging Neuroscience, 10*(328). https://doi.org/10.3389/fnagi.2018.00328

Swift-Gallant, A., Johnson, B. A., Rita, V. D., & Breedlove, S. M. (2020). Through a glass, darkly: Human digit ratios reflect prenatal androgens, imperfectly. *Hormones and Behavior, 120*(104686). https://doi.org/10.1016/j.yhbeh.2020.104686

Swift-Gallant, A., Shirazi, T., Puts, D. A., & Breedlove, S. M. (2021). Evidence for perinatal steroid influence on human sexual orientation and gendered behavior. *Cold Spring Harbor Perspectives in Biology, 14*(a039123). https://doi.org/10.1101/cshperspect.a039123

Tanner, J. M. (1990). *Foetus into Man: Physical Growth from Conception to Maturity, Revised Edition*. Harvard University Press.

Vanderschueren, D., Vandenput, L., Boonen, S., Lindberg, M. K., Bouillon, R., & Ohlsson, C. (2004). Androgens and bone. *Endocrine Reviews, 25*(3), 389–425. https://doi.org/10.1210/er.2003-0003

Voracek, M. (2014). No effects of androgen receptor gene CAG and GGC repeat polymorphisms on digit ratio (2D:4D): A comprehensive meta-analysis and critical evaluation of research. *Evolution and Human Behavior, 35*(5), 430–437. https://doi.org/10.1016/j.evolhumbehav.2014.05.009

Warrington, N. M., Shevroja, E., Hemani, G., Hysi, P. G., Jiang, Y., Auton, A., Boer, C. G., Mangino, M., Wang, C. A., Kemp, J. P., McMahon, G., Medina-Gomez, C., Hickey, M., Trajanoska, K., Wolke, D., Ikram, M. A., Team, T. a. R., Montgomery, G. W., Felix, J. F., (…), & Evans, D. M. (2018). Genome-wide association study identifies nine novel loci for 2D:4D finger ratio, a putative retrospective biomarker of testosterone exposure in utero. *Human Molecular Genetics, 27*(11), 2025–2038. https://doi.org/10.1093/hmg/ddy121https://doi.org/10.1093/hmg/ddy121

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

Welker, K. M., Lassetter, B., Brandes, C. M., Prasad, S., Koop, D. R., & Mehta, P. H. (2016b). A comparison of salivary testosterone measurement using immunoassays and tandem mass spectrometry. *Psychoneuroendocrinology, 71*, 180–188. https://doi.org/10.1016/j.psyneuen.2016.05.022

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

Wilson, E. O. (1980). *Sociobiology: The abridged edition*. Belknap/Harvard.

Zheng, Z., & Cohn, M. J. (2011). Developmental basis of sexually dimorphic digit ratios. *Proceedings of the National Academy of Sciences, 108*(39), 16289–16294. https://doi.org/10.1073/pnas.1108312108

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Martin G. Köllner[1] [iD] **· Sinja Braun[1] · Hanna Schöttner[1] · Gelena Dlugash[1]** [iD] **· Marlene Bettac[1] · Simon Steib[2]**

✉   Martin G. Köllner
     martin.koellner@fau.de

[1]   Human Motivation and Affective Neuroscience Lab, Department of Psychology. Institute
     of Psychology, Friedrich-Alexander University, Nägelsbachstrasse 49b, 91052 Erlangen,
     Germany

[2]   Human Movement, Training and Active Aging Lab, Institute of Sports and Sports Sciences,
     Heidelberg University, Im Neuenheimer Feld 700, 69120 Heidelberg, Germany