No Difference in 2D:4D Ratio between Youth with Elevated Prenatal Androgen Exposure due to Congenital Adrenal Hyperplasia and Controls

Authors: Gideon Nave[1], Christina M. Koppin[2], Dylan Manfredi[1], Gareth Richards[3], Steven J. Watson[4], Mitchell E. Geffner[2,5,6], Jillian E. Yong[2], Robert Kim[2], Heather M. Ross[2], Monica Serrano-Gonzalez[7], Mimi S. Kim[2,5,6]

Author details:

[1] Marketing Department, The Wharton School, University of Pennsylvania

[2] Center for Endocrinology, Diabetes, and Metabolism, Children's Hospital Los Angeles

[3]School of Psychology, Newcastle University

[4]Department of Psychology of Conflict, Risk and Safety, University of Twente

[5]The Saban Research Institute at Children's Hospital Los Angeles

[6]Keck School of Medicine of University of Southern California

[7]Division of Pediatric Endocrinology, Hasbro Children's Hospital, Warren Alpert Medical School of Brown University

Short Title: No difference in 2D:4D between CAH Youth and Controls

Corresponding Author: Gideon Nave

Address: The Wharton School, 3730 Walnut St. JMHH Suite 700, Philadelphia PA 19104

Email: gnave@wharton.upenn.edu

Reprint requests should be addressed to Gideon Nave

Funding Sources: This work was supported by funding support from the Abell Foundation to M.E.G. in the collection of the data and from the National Institutes of Health K23HD084735 (NIH/NICHD) to M.S.K. in the writing of the report.

Disclosure Summary: M.E.G. receives consultant fees from Daiichi Sankyo, Ferring, Millendo Therapeutics, Novo Nordisk, Nutrition & Growth Solutions, Pfizer, Sandoz, and Tolmar; receives royalties from McGraw-Hill and UpToDate; and serves on data safety monitoring boards for Ascendis, Millendo, and Tolmar.

Abstract

The second-to-fourth digit ratio (2D:4D) has been associated with sexual dimorphism, with a lower 2D:4D in males. A large body of research has relied on the 2D:4D as a proxy for prenatal androgen exposure, and includes reports of relationships between 2D:4D and a wide range of human traits. Here, we examine the validity of the 2D:4D proxy by studying the association between 2D:4D and classical Congenital Adrenal Hyperplasia (CAH) due to 21-hydroxylase deficiency, a condition characterized by excessive prenatal exposure to androgens during most of the gestational period. To this end, we retrospectively examine serial radiographs of the left hand obtained clinically in 91 youth with classical CAH (49.5% female) and 70 control youth (44.3% female). Replicating previous reports, we observe associations of the 2D:4D with sex (lower 2D:4D in males), age (increase of 2D:4D through development) and ethnicity (lower 2D:4D in Hispanic individuals). However, we find no evidence for differences in 2D:4D between CAH and controls (full sample: □ = 0.0003 [-0.007, 0.008]; females: □ = -0.004 [-0.015, 0.007]; males: □ = 0.003, [-0.007, 0.013]). Although our findings do not rule out a small association between the 2D:4D and CAH, they cast doubt on the usefulness of the 2D:4D as a biomarker for prenatal androgen exposure in behavioral research.

Keywords: congenital adrenal hyperplasia; 21-hydroxylase deficiency; pediatrics; testosterone; prenatal exposure delayed effects; 2D:4D; digit ratio; androgen excess; prenatal hormones

Highlights

- We conducted a retrospective study of radiographs to quantify the differences in the second-to-fourth digit ratio (2D:4D) between youth with classical Congenital Adrenal Hyperplasia (CAH) and controls.
- We replicated previously reported associations of the 2D:4D with sex (lower 2D:4D in males), age (increase in 2D:4D throughout development) and ethnicity (lower 2D:4D in Hispanic individuals).
- The associations between 2D:4D and CAH diagnosis were small and statistically indistinguishable from zero, both in the full sample and within sex.
- Our findings cast doubt on the reliability of the 2D:4D measure as a biomarker for prenatal androgen exposure in behavioral research.

Introduction

The sex hormone testosterone plays a crucial role in the prenatal sexual differentiation. A growing body of research suggests that testosterone exposure *in-utero* also affects the development of psychological and behavioral traits later in life (Bütikofer et al., 2019; Constantinescu and Hines, 2012; Knickmeyer and Baron-Cohen, 2006; Tapp et al., 2011). Studying the prenatal effects of testosterone is difficult, because measuring hormone levels in-utero requires high-risk procedures such as amniocentesis (Keelan et al., 2012). In an attempt to overcome this challenge, the ratio between the second (index) and fourth (ring) fingers (2D:4D) (Manning et al., 1998) has been proposed to approximate prenatal testosterone exposure. This ration is on average lower in males compared to females (Baker, 1888; Hönekopp and Watson, 2010; Manning et al., 2004).

As of 2019, the number of published studies associating 2D:4D with later-life psychological and behavioral traits exceeds 1,400 (Leslie, 2019). The literature includes reports of associations between 2D:4D and a wide range of outcomes, including aggressive and violent behavior (Bailey and Hurd, 2005; Turanovic et al., 2017, though see Pratt et al., 2016), risk taking (Stenstrom et al., 2011, though see Neyse et al., 2020; van Leeuwen et al., 2020), empathic accuracy (Nitschke and Bartz, 2019, though see Nadler et al., 2019) and even success records of day traders (Coates et al., 2009) and sumo wrestlers (Tamiya et al., 2012).

Despite the growing literature, the validity of 2D:4D as a proxy for prenatal testosterone exposure is not well-established. Animal studies that directly manipulated prenatal androgen levels found conflicting evidence for causal effects of such manipulations on 2D:4D (Huber et al., 2017; Zheng and Cohn, 2011). Human studies of the associations between 2D:4D and hormonal measures from amniotic fluid, umbilical cord blood, or the maternal circulation during pregnancy have also found inconclusive support for the proxy's validity (for review, see Richards, 2017). There is also evidence that 2D:4D is influenced by factors that are unrelated to prenatal hormonal levels. For example, studies of females with complete androgen insensitivity syndrome (CAIS) found that the variance in 2D:4D was similar between CAIS and controls, despite complete insensitivity to androgenic influences in the former population (Berenbaum et al., 2009; van Hemmen et al., 2017). Furthermore, 2D:4D has been found to increase during early-life development, suggesting that the ratio is not prenatally fixed (McIntyre et al., 2005; Richards et al., 2017; Trivers et al., 2006).

Important evidence in support of the 2D:4D proxy's validity comes from studies of patients with Congenital Adrenal Hyperplasia (CAH) due to 21-hydroxylase deficiency. CAH is the most common primary adrenal insufficiency in children. The resulting cortisol deficiency leads to an overproduction of adrenal androgens by the 7th gestational week in classical (severe) CAH, secondary to disrupted steroid biosynthesis (Speiser et al., 2010; White and Speiser, 2002). Such significant changes in the intrauterine environment during early development could adversely impact fetal programming with organizational effects from excess androgens. This is seen in females with classical CAH, who are born with virilized, ambiguous genitalia and can exhibit male-typical play preferences in childhood (Berenbaum and Beltz, 2011; Hines, 2008; Pasterski et al., 2011). Thus, females with classical CAH are born with outward signs of androgen exposure, unlike males with classical CAH, and differential effects can be seen for behavior and fertility (Berenbaum et al., 2012; White and Speiser, 2002).

The first studies that compared 2D:4D between patients with CAH and controls appeared in the early 2000s (Brown et al., 2002; Okten et al., 2002). Both studies were relatively small (less than 20 CAH cases), and both found that measures of 2D:4D from hand-scans were lower (more masculinized) in both males and females with CAH compared to unaffected individuals. The most recent meta-analysis (Hönekopp and Watson, 2010), which included data from two additional studies (Buck et al., 2003; Ciumas et al., 2009), concluded that both males and females with CAH had lower 2D:4D compared to sex-matched controls. The meta-analytic effects were relatively large in size (Cohen's *d* between 0.6 and 0.9), and observed for both hands.

Despite these advances, the set of studies conducted to-date is relatively small, and it is subject to important limitations. First, most studies to-date have relied on 2D:4D measures from photocopies or hand scans. The two exceptions were studies conducted using hand radiographs by Okten et al. (2002)[1] and Buck et al. (2003), and neither of them reported a statistically significant association between CAH status and 2D:4D. This issue is of particular importance, because 2D:4D measures from hand-scans and photocopies were shown to systematically differ from 2D:4D measures obtained directly from the hands. This difference is thought to arise from

[1] Okten et al. (2002) included a small collection of hand radiographs (17 girls and 9 boys), and were unable to find differences between CAH and controls (see Okten et al., 2002, Table 3 p. 51). However, a reanalysis of the data presented in that paper (see Richards et al., 2020, Table 4) revealed a significant difference in radiographic R2D:4D between females with CAH (n=17, M = 0.99, SD = 0.02) and female controls (n=34, M = 1.00, SD = 0.01), t = -2.393, p = 0.021, d = -0.711.

individual differences in the shapes of fat-pads at the fingertips, as well as the degree of pressure applied when the hands are placed on the glass plate of the copier or scanner (Caswell and Manning, 2009; Ribeiro et al., 2016; Wallen, 2009). As noted by Ribeiro et al. (2016), both the fat pads and the degree of pressure might vary by factors such as sex, sexual orientation and personality. As a result, 2D:4D measures from photocopies or hand-scans might generate spurious associations between the 2D:4D proxy and various traits that are unrelated to exposure to prenatal androgens.

A second limitation is that none of the studies comparing 2D:4D between patients with CAH and controls have accounted for the potentially confounding influence of skeletal development. This methodological detail is important because 2D:4D systematically increases during early-life development (McIntyre et al. 2005; Trivers et al. 2006), and patients with CAH can have advancement in skeletal maturation secondary to androgen exposure (Finkielstain et al., 2012). Furthermore, some of the previously reported studies included samples where CAH patients and controls differed considerably in their average age (e.g. Rivas et al., 2014). As a consequence, differences in 2D:4D between patients with CAH and controls might reflect general differences in skeletal development rates or chronological age, rather than specific effects of prenatal androgen exposure.

The current work overcomes the limitations above, in the largest study to-date (N = 161, including 91 CAH cases) comparing the 2D:4D of patients with CAH and controls (both males and females). Our study relies on repeated 2D:4D measures from radiographs, allowing us to directly quantify finger bone structure and to overcome the confounding influences associated with 2D:4D measures from hand-scans (Ribeiro et al., 2016). Our reliance on radiographs further allows us to systematically account for the confounding effects of skeletal development by accurately measuring bone age and controlling for it statistically.

Study Participants and Methods

This was a retrospective study approved by the Children's Hospital Los Angeles (CHLA) Institutional Review Board (CHLA-15-00453, CCI-12-00020). Parents and/or participants gave their written informed consent, with assent from participants between the ages of 7 up to 14 years old. This was in compliance with the Code of Ethical Principles for Medical Research Involving Human Subjects of the World Medical Association (Declaration of Helsinki).

Participants

Ninety-one youth (45 females [49.5%]; 1.1-18.7 years old at the time of imaging) with classical CAH due to 21-hydroxylase deficiency and 70 control youth (31 females [44.3%], 2.6-19.7 years old at the time of imaging) were retrospectively studied. Youth with CAH were patients at the CHLA pediatric endocrinology clinic (CAH Comprehensive Care Center) who were diagnosed with CAH due to 21-hydroxylase deficiency biochemically (elevated basal and/or ACTH-stimulated serum 17-hydroxyprogesterone and androgens) and/or by *CYP21A2* genotype variant. The CAH group consisted of 76% subjects with the salt-wasting (SW) form (*N* = 69, 28 females) and 24% with the simple-virilizing (SV) form (*N* = 22, 17 females) based on clinical phenotype and/or genotype. Controls were selected from a group of healthy participants who had been seen at the CHLA pediatric endocrinology clinic for an assessment of short stature (N = 57), as well as control participants from a different research study (N = 13; Kim et al., 2015). We excluded control participants if (i) they had a pathological etiology for short stature, including a genetic or chromosomal abnormality, and /or (ii) they were medically eligible for, or treated with, growth hormone. Control participants were otherwise healthy and were eligible if they had benign short stature due to familial short stature and/or constitutional delay of growth.

Measures

Hand Radiographs. Serial radiographs of the left hand and wrist were obtained for bone age assessment as part of standard care in youth with CAH, and as part of an evaluation of short stature in otherwise healthy control youth. A smaller proportion of CAH and control participants had x-rays obtained as part of a different research study (Kim et al., 2015). A pediatric endocrinologist (either M.S.K. or M.S-G.) read the radiographs in a blinded fashion to calculate bone age using the Greulich-Pyle method (Gilsanz and Ratib, 2005; Greulich and Pyle, 1953). All participants in the CAH group and 32 participants in the control group (45.7%) had multiple hand radiographs taken at different ages. The total number of CAH hand radiographs was 413 and the total number of control hand radiographs was 119.

Digit Ratios. Two independent raters measured the second and fourth finger lengths from radiographs using digital software (Rasband, 2011; Schneider et al., 2012). First, they drew horizontal lines at the metaphysis of each finger segment (proximal, middle, and distal), and located the midpoint per line. They then measured finger length by drawing a vertical line from

the distal tip through the proximal epiphysis through the midpoints (**Figure 1**). Finally, they calculated the 2D:4D ratio by dividing the average (of the two raters) length measure of the second finger by that of the fourth finger (inter-rater correlation quantified via Pearson's *r* was .97, **Supplemental Figure 1**).

A **B** **C**

Figure 1. Measurements of the 2nd and 4th finger lengths. (A) Radiograph of the left hand. **(B)** Horizontal lines were drawn at the metaphysis of each phalangeal segment and the midpoint located per line (proximal, middle, and distal metaphysis). (**C**) A vertical line was drawn from the distal tip through the proximal epiphysis to give the finger length .

Additional Measures. Our analyses include controls for various factors that may be associated with the 2D:4D ratio based on previous findings. Our primary statistical models include controls for sex (Female=1, Male=0), bone-age (in years) and ethnicity (Hispanic=1; non-Hispanic=0). We also conducted secondary analyses that included additional controls for chronological age (years), height (measured with a stadiometer; cm), age-normalized height (z-score) per United States population data[2], and puberty status (Tanner puberty stage 1-5, assessed by a pediatric endocrinologist). The summary statistics of these variables across the 532 radiographs used in the study are summarized in **Table 1**. The most noticeable difference between the CAH and control patients was in normalized height. While both populations were shorter than average for their age (as implied by the negative mean values), the CAH group was much shorter. This

[2] See https://www.cdc.gov/growthcharts/zscore.htm

difference was expected, and can be attributed to androgen excess, advanced bone age, and glucocorticoid dose in CAH (Finkielstain et al., 2012).

The correlations between the different variables used in the study are summarized in **Table 2**. As expected, there were strong positive correlations between chronological age, bone age, height, and puberty status. This collinearity implies that while it is possible to control for these factors in our analyses, disentangling their individual contributions is not straightforward. Although bone age and chronological age were highly correlated overall, we did observe advanced bone ages in the CAH group (48.4% of patients had at least one radiograph with bone age 2 SD above the age-adjusted mean; see **Figure 2**), a finding which is in line with previous reports (Finkielstain et al., 2012).

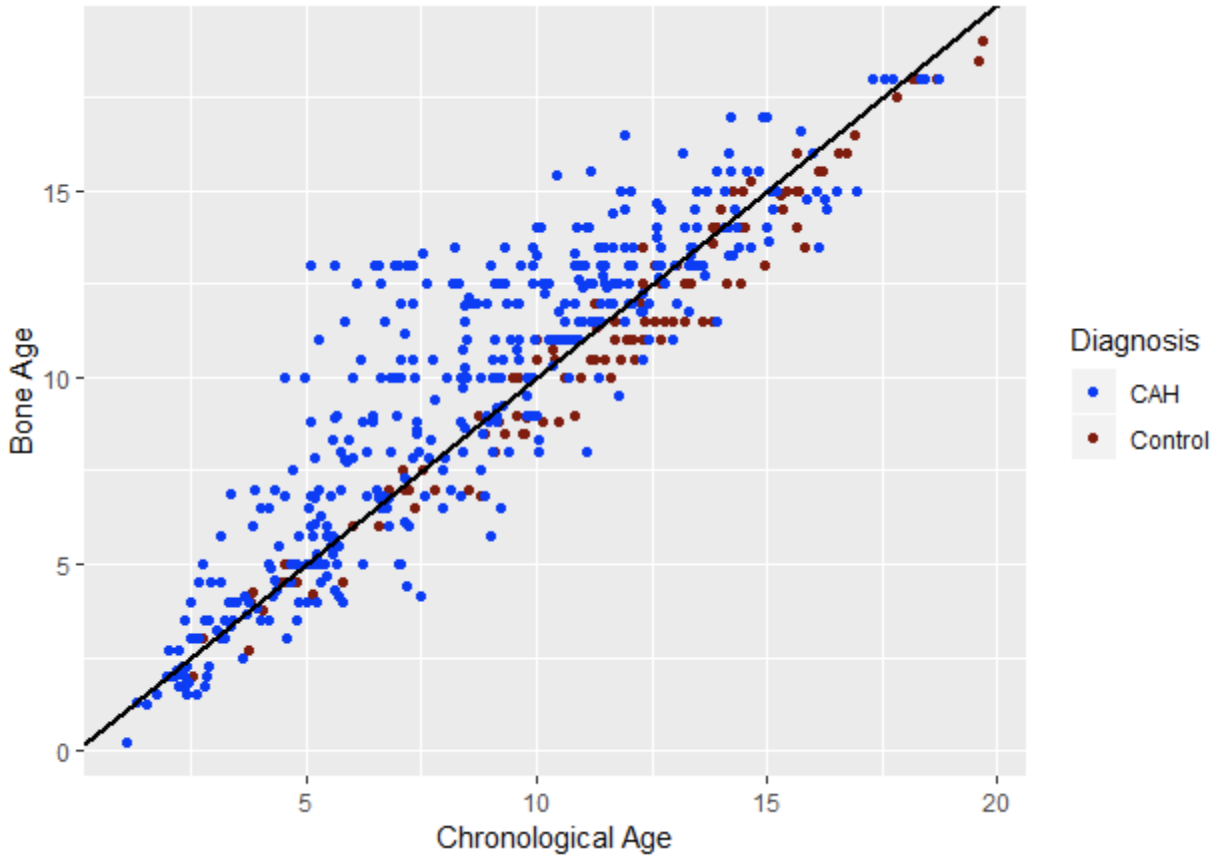
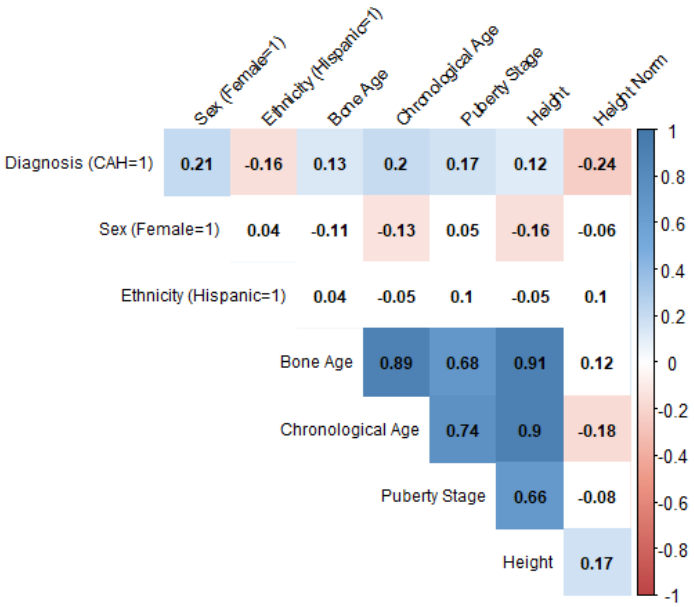


Figure 2. Relationship between chronological age and bone age in the study participants. The black line is a fitted regression line for the full sample.

Table 1: Image characteristics (standard deviations in parenthesis) by CAH diagnosis

Image Characteristic	CAH	Control
Chronological Age	10.13 (3.89)	8.61 (4.01)
Bone Age	10.41 (3.87)	9.47 (4.15)
Height	134.55 (21.91)	129.74 (22.56)
Height Normalized	-0.83 (1.52)	-0.18 (1.16)
% Females	42.11%	62.85%
% White	56.94%	42.72%
% Puberty Stage 1 % Puberty Stage 2 % Puberty Stage 3 % Puberty Stage 4 % Puberty Stage 5	55.61% 12.20% 13.66% 5.37% 13.17%	69.80% 10.07% 9.40% 7.38% 3.36%
Number of Images	413	119
Number of individuals	91	70

Table 2: Pairwise correlations between the study variables across images



Statistical analyses

We estimated mixed-effects linear regression models, implemented using the 'lme4' package in the statistical software 'R' (Bates et al., 2014) to quantify the associations between 2D:4D and the explanatory variables. The outcome variable in all models was the mean left-hand 2D:4D across the two raters. Our main analysis (Model A) included fixed effects for CAH status (CAH=1; Control=0), sex, bone-age, and ethnicity, as well as participant-specific random intercepts. We estimated additional mixed effect regressions (Models B-F) to evaluate the sensitivity of our results to different model specifications. Model B included an interaction effect between CAH status and sex, but was otherwise identical to Model A. Model C included additional controls for chronological age, height, age-normalized height and puberty status, and was otherwise identical to Model A (observations with missing values of any of the additional control variables were omitted). Models D and F were identical to Model A, but were estimated separately for females and males respectively. Models E and G were identical to Model C, but were estimated separately for females and males respectively.

Results

Table 3 summarizes the coefficient estimates of our regression models. We replicated previous findings (Manning et al., 2007) that 2D:4D was greater in females (□ = 0.010, 95% confidence interval (CI) = [(0.003, 0.017)]) and lower in Hispanic (vs. non-Hispanic) participants (□ = -0.009, 95% CI = [(-0.016, -0.002)]). We also replicated the finding of McIntyre et al., (2006) that 2D:4D increases with bone age (□ = 0.002, 95% CI = [(0.001, 0.002)], **Supplemental Figure 2A**). 2D:4D also positively correlated with height (see **Supplemental Figure 2B)** and chronological age, though as these two variables strongly correlated with one another and with bone age, it was not possible to disentangle the differential effects of these three factors.

The regression coefficient for CAH diagnosis was not statistically distinguishable from zero under any of the model specifications in the full sample (□ = 0.0003, 95% CI = [(-0.007, 0.008)]; Model A). This point estimate of this effect's size was opposite in sign than the effect of sex, and its size was only 3% of it. The interaction coefficient between sex and CAH diagnosis was also not statistically distinguishable from zero (□ = -0.006, 95% CI = [-0.021, 0.008]; Model B). Analyses stratified by sex could not identify reliable effects of CAH status in either

females (□ =-0.004, 95% CI = [-0.015, 0.007]; Model D) or males (□ = 0.003, 95% CI = [(-0.007, 0.013)]; Model F). Further analyses with additional controls for chronological age, height (absolute and age normalized) and puberty status yielded similar results (Models C, E and G).

Table 3:
Linear regression of left hand 2D:4D on CAH diagnosis and controls on (95% Confidence intervals in parenthesis)

	Outcome variable:						
	Left 2D:4D						
	All	All	All	Females	Females	Males	Males
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Diagnosis (CAH = 1)	0.0003	0.003	-0.001	-0.004	-0.006	0.003	0.001
	(-0.007, 0.008)	(-0.007, 0.013)	(-0.009, 0.007)	(-0.015, 0.007)	(-0.017, 0.005)	(-0.007, 0.013)	(-0.008, 0.011)
Sex (Female = 1)	0.010**	0.013*	0.009*				
	(0.003, 0.017)	(0.002, 0.024)	(0.002, 0.016)				
Bone Age (Years)	0.002***	0.002***	0.002**	0.001***	0.002	0.002***	0.002*
	(0.001, 0.002)	(0.001, 0.002)	(0.001, 0.003)	(0.001, 0.002)	(-0.00001, 0.003)	(0.001, 0.002)	(0.0003, 0.004)
Ethnicity (Hispanic = 1)	-0.009*	-0.009*	-0.005	-0.011*	-0.010	-0.007	-0.004
	(-0.016, -0.002)	(-0.016, -0.002)	(-0.012, 0.002)	(-0.022, -0.0002)	(-0.021, 0.001)	(-0.017, 0.002)	(-0.013, 0.005)
Diagnosis x Sex		-0.006					
		(-0.021, 0.008)					
Chron. Age (Years)			-0.004***		-0.002		-0.011***
			(-0.007, -0.002)		(-0.004, 0.0004)		(-0.015, -0.007)
Puberty Stage = 2			0.003		0.006*		0.002
			(-0.002, 0.007)		(0.00003, 0.012)		(-0.004, 0.008)
Puberty Stage = 3			-0.003		0.00003		-0.006
			(-0.008, 0.002)		(-0.005, 0.005)		(-0.013, 0.001)
Puberty Stage = 4			-0.003		0.005		-0.006
			(-0.009, 0.004)		(-0.003, 0.013)		(-0.015, 0.003)
Puberty Stage = 5			-0.008*		-0.002		-0.011

			(-0.015, -0.0004)		(-0.011, 0.007)		(-0.021, 0.0003)
Height (cm)			0.001***		0.0003		0.002***
			(0.0005, 0.001)		(-0.00004, 0.001)		(0.001, 0.003)
Height (Normalized)			-0.006***		-0.002		-0.015***
			(-0.009, -0.003)		(-0.005, 0.001)		(-0.019, -0.010)
Intercept	0.900***	0.898***	0.834***	0.916***	0.892***	0.896***	0.737***
	(0.892, 0.907)	(0.890, 0.907)	(0.808, 0.859)	(0.904, 0.927)	(0.862, 0.923)	(0.886, 0.905)	(0.690, 0.785)
Observations	532	532	465	241	206	291	259
Log Likelihood	1,408.720	1,405.082	1,211.243	650.917	541.942	750.327	646.199
Akaike Inf. Crit.	-2,803.441	-2,794.164	-2,394.487	-1,289.834	-1,057.884	-1,488.653	-1,266.398
Bayesian Inf. Crit.	-2,773.504	-2,759.951	-2,336.498	-1,268.925	-1,014.622	-1,466.613	-1,220.160
Note:							*p<0.05 **p<0.01 ***p<0.001

Discussion

We conducted a retrospective study to quantify the relationship between the 2D:4D ratio and various factors, including diagnosis of CAH, which provides a natural human model of excess prenatal androgens from early in fetal development. Our investigation, which is the largest single study of the topic to-date, relied on repeated measures of 2D:4D from finger lengths in hand radiographs. Our results replicate previously reported associations of 2D:4D with regard to sex, age, and ethnicity. However, the associations between 2D:4D and CAH status are not distinguishable from zero.

Our findings are inconsistent with the results of the most recent meta-analysis (Hönekopp and Watson, 2010), which reported relatively large effects of CAH status on 2D:4D (Cohen's *d* between 0.6 and 0.9). Of note, our study had statistical power of over 90% to detect even the smallest effect reported in the meta-analysis ($d = 0.6$).[3] The discrepancy between the results of

[3] Statistical power was calculated using G* Power 3.1 (Faul et al., 2009), for a study without repeated measures in a sample of 91 CAH and 70 Controls (as the combined sample of the current study). The availability of repeated 2D:4D measures in our sample, our reliance on measures obtained from x-rays (which are less noisy than hand-scans (Ribeiro et al., 2016)) as well as our capacity to control for various confounding factors (e.g., bone age and puberty stage), was expected to increase the statistical power of our study even further. Our female sample (N=76) would have 90% power to detect the meta-analytic effect found in (Hönekopp and Watson, 2010) for the left hand of females ($d = 0.75$). Our male sample (N=85) would have 80% power to detect the meta-analytic effect found in (Hönekopp and Watson, 2010) for the left hand of males ($d = 0.63$).

our study and the meta-analysis may arise from the fact that the latter mostly relied on studies that measured 2D:4D on outlined soft tissue from hand-scans and photocopies, rather than bone measures on radiographs. Other potential causes for this discrepancy could be differences in sample characteristics, the existence of publication bias in the meta-analyzed literature (Carter and McCullough, 2014; Lane et al., 2016) or the presence of a statistical artifact due to the small number of meta-analyzed studies (Borenstein et al., 2009). Nonetheless, our results are in line with a previous study that used 2D:4D measures from radiographs (Buck et al., 2003), which did not observe any differences between females with and without CAH. Our studies shared similar approaches, including a similar control group with shorter stature, and measurements of bone tissue from left-hand radiographs.

Our findings do not rule out the possibility of a small association between prenatal androgen elevation due to CAH and 2D:4D. This is especially true for females, for whom the point estimate of the effect of CAH on 2D:4D was negative. Nonetheless, our findings raise questions concerning the validity of behavioral research that has relied on 2D:4D to approximate prenatal testosterone exposure. This is because the relationship between 2D:4D and other phenotypes that are commonly studied in this literature is expected to be even weaker than the association that we observed between 2D:4D and CAH diagnosis.

Indeed, our study examined 2D:4D in relation to a phenotype that is known to have a strong relationship with prenatal androgen exposure, particularly in females, and its phenotypic assessments were based on clinical diagnoses with virtually no measurement error. Conversely, most 2D:4D studies of neurotypical populations report correlations between digit ratios and traits that are complex (e.g., aggressive behavior, risk taking, cognitive empathy). Such traits are measured with less reliability than CAH diagnosis, and are likely associated with many genetic and environmental factors that are not directly related to prenatal testosterone exposure. Furthermore, we measured 2D:4D from bone tissue on radiographs, whereas most previous studies in the literature have relied on 2D:4D measured from photocopies or hand-scans. The latter are noisier measures of digit ratios, due to individual differences in finger fat tissue and the degree of pressure applied when the hands are scanned (Ribeiro et al., 2016). Finally, the much larger effects of age and ethnicity on 2D:4D that we observed demonstrate that many factors

unrelated to prenatal androgen exposure influence the 2D:4D measure, potentially confounding associations between the proxy and behavioral outcomes.

One limitation of our study is that we only used radiographs of the left hand, as a result of largely utilizing x-rays taken for clinical purposes (to determine bone age) that are typically of the left hand (similar to Buck et al., 2003). Although studies using hand-scans documented greater sex differences in the 2D:4D of the right hand (Hönekopp and Watson, 2010), we highlight that our study was able to detect highly significant sex differences in the left hand. Relatedly, a large study of over 3,000 radiographs found that the 2D:4D is symmetrical, with only minor differences between the hands (Robertson et al., 2008).

A second limitation is that our results were based on a sample of youth whose radiographs were taken for clinical purposes, rather than a random sample of the population (similar to Buck et al., 2003). Yet, our analyses systematically controlled for bone age, chronological age, height (absolute and age-normalized), ethnicity, and pubertal status. We have no reason to believe that our results depend on other characteristics of the study participants, materials, and/or context.

In summary, we do not find evidence for a relationship between 2D:4D ratio and CAH status in our sample, yet we do find that the measure is associated with various other factors. This finding suggests that previously reported correlations between 2D:4D and a wide range of behavioral and cognitive traits should be interpreted with caution. We believe, however, that continued investigations of the relationships between prenatal testosterone exposure and behavioral outcomes would be worthwhile. One possibility for conducting such research is to directly study developmentally unique populations such as CAH or complete androgen insensitivity syndrome. Although obtaining large samples of participants from these populations might be difficult, the validity of this approach is based on a more solid empirical ground than relying on the 2D:4D ratio as a proxy from prenatal androgen exposure.

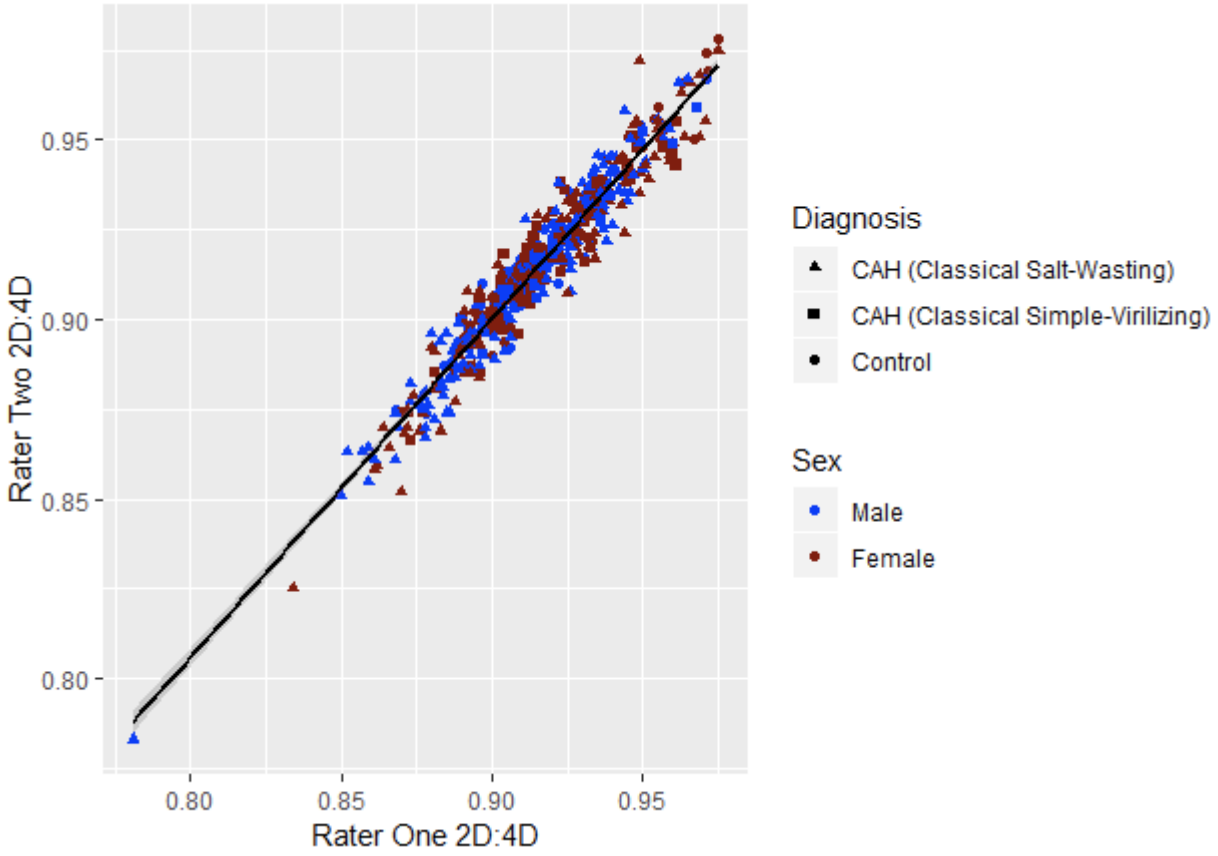
Acknowledgments

We gratefully thank the patients and families who participated in the Natural History Study of CAH from Infancy, and the Testosterone and CAH Study at our center. We thank CARES Foundation for its support of the CHLA CAH Comprehensive Care Center (to M.S.K and M.E.G.), and the Abell Foundation for ongoing support of CAH research at our center (to

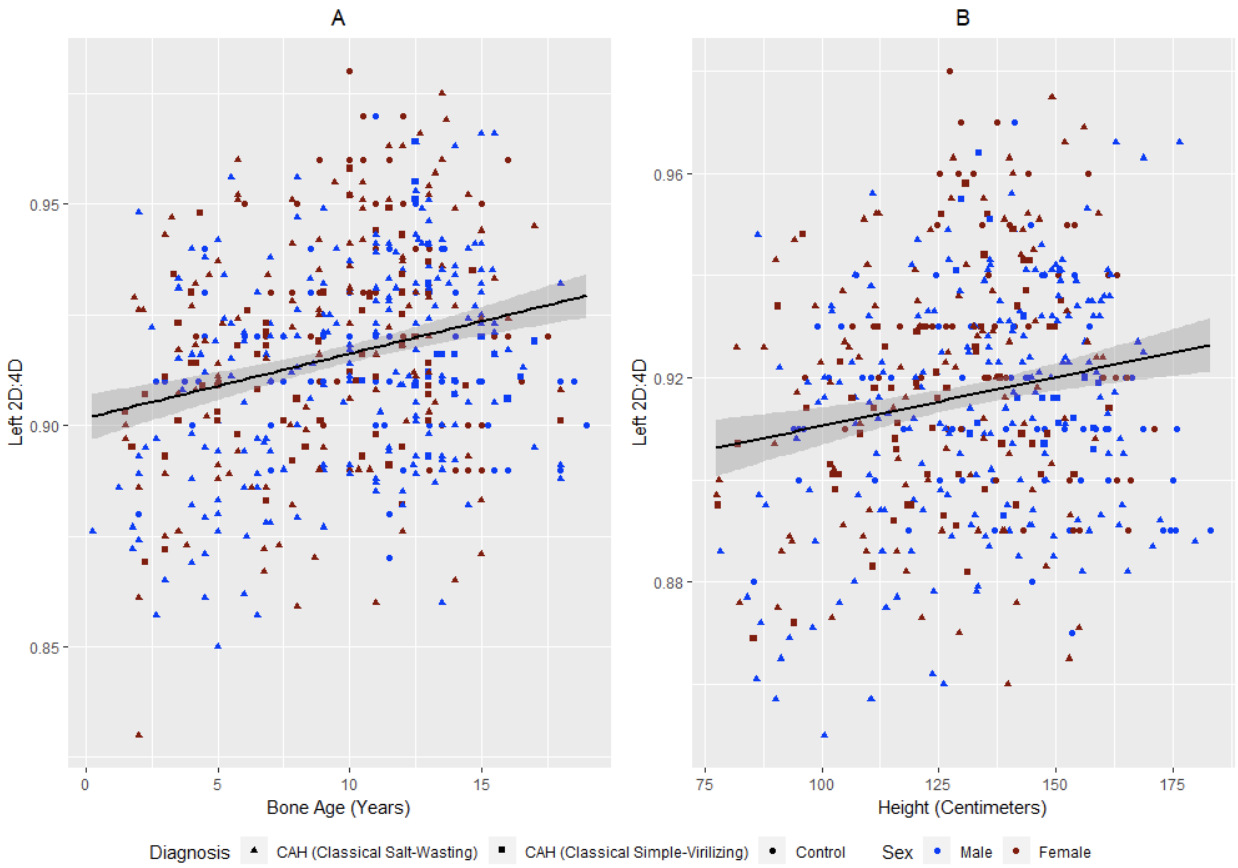
M.E.G.). The writing of this paper was in part funded by K23HD084735-01A1 (to M.S.K.). The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. Gideon Nave thanks the support of Carlos and Rosa de la Cruz and the Wharton Neuroscience Initiative.

Disclosures

Mitchell E. Geffner serves as an advisor to Daiichi Sankyo, Novo Nordisk, Nutrition & Growth Solutions, Pfizer, and Spruce Biosciences; on data safety monitoring boards for Ascendis and Tolmar; and receives royalties from McGraw-Hill and UpToDate.



Supplemental Figure 1. 2D:4D ratio measurements in youth with and without classical CAH due to 21-hydroxylase deficiency. 413 hand radiographs in 91 youth with classical CAH (salt-wasting and simple-virilizing forms) and 119 hand radiographs in 70 controls were measured by two independent raters for 2D:4D.



Supplemental Figure 2. (A) Relationship between bone age (years) and 2D:4D of the left hand (in all participants). **(B)** Relationship between height (cm) and 2D:4D of the left hand (in all participants)**.** The black lines are fitted regression lines.

References

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

Baker, F., 1888. Anthropological Notes on the Human Hand. Am. Anthropol. 1, 51–76.

Bates, D., Maechler, M., Bolker, B., Walker, S., Others, 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1, 1–23.

Berenbaum, S.A., Beltz, A.M., 2011. Sexual differentiation of human behavior: effects of prenatal and pubertal organizational hormones. Front. Neuroendocrinol. 32, 183–200.

Berenbaum, S.A., Bryk, K.K., Nowak, N., Quigley, C.A., Moffat, S., 2009. Fingers as a marker of prenatal androgen exposure. Endocrinology 150, 5119–5124.

Berenbaum, S.A., Bryk, K.L.K., Beltz, A.M., 2012. Early androgen effects on spatial and mechanical abilities: evidence from congenital adrenal hyperplasia. Behav. Neurosci. 126, 86–96.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009. When does it make sense to perform a meta-analysis. Introduction to meta-analysis 357–364.

Brown, W.M., Hines, M., Fane, B.A., Breedlove, S.M., 2002. Masculinized finger length patterns in human males and females with congenital adrenal hyperplasia. Horm. Behav. 42, 380–386.

Buck, J.J., Williams, R.M., Hughes, I.A., Acerini, C.L., 2003. In□utero androgen exposure and 2nd to 4th digit length ratio—comparisons between healthy controls and females with classical congenital adrenal hyperplasia. Hum. Reprod. 18, 976–979.

Bütikofer, A., Figlio, D.N., Karbownik, K., Kuzawa, C.W., Salvanes, K.G., 2019. Evidence that prenatal testosterone transfer from male twins reduces the fertility and socioeconomic success of their female co-twins. Proc. Natl. Acad. Sci. U. S. A. 116, 6749–6753.

Carter, E.C., McCullough, M.E., 2014. Publication bias and the limited strength model of self-control: has the evidence for ego depletion been overestimated? Front. Psychol. 5, 823.

Caswell, N., Manning, J.T., 2009. A Comparison of Finger 2D:4D by Self-Report Direct Measurement and Experimenter Measurement from Photocopy: Methodological Issues. Archives of Sexual Behavior. https://doi.org/10.1007/s10508-007-9202-3

Ciumas, C., Lindén Hirschberg, A., Savic, I., 2009. High fetal testosterone and sexually dimorphic cerebral networks in females. Cereb. Cortex 19, 1167–1174.

Coates, J.M., Gurnell, M., Rustichini, A., 2009. Second-to-fourth digit ratio predicts success among high-frequency financial traders. Proc. Natl. Acad. Sci. U. S. A. 106, 623–628.

Constantinescu, M., Hines, M., 2012. Relating prenatal testosterone exposure to postnatal behavior in typically developing children: methods and findings. Child Dev. Perspect.

Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G., 2009. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav. Res. Methods 41, 1149–1160.

Finkielstain, G.P., Kim, M.S., Sinaii, N., Nishitani, M., Van Ryzin, C., Hill, S.C., Reynolds, J.C., Hanna, R.M., Merke, D.P., 2012. Clinical characteristics of a cohort of 244 patients with congenital adrenal hyperplasia. J. Clin. Endocrinol. Metab. 97, 4429–4438.

Gilsanz, V., Ratib, O., 2005. Hand Bone Age: A Digital Atlas Of Skeletal Maturity. Springer Science & Business Media.

Greulich, W.W., Pyle, S.I., 1953. Radiologic atlas of skeletal development of the hand and wrist (2nd).

Hines, M., 2008. Early androgen influences on human neural and behavioural development.

Early Hum. Dev. 84, 805–807.
Hönekopp, J., Watson, S., 2010. Meta analysis of digit ratio 2D: 4D shows greater sex difference in the right hand. Am. J. Hum. Biol.
Huber, S.E., Lenz, B., Kornhuber, J., Müller, C.P., 2017. Prenatal androgen-receptor activity has organizational morphological effects in mice. PLoS One 12, e0188752.
Keelan, J.A., Mattes, E., Tan, H., Dinan, A., Newnham, J.P., Whitehouse, A.J.O., Jacoby, P., Hickey, M., 2012. Androgen concentrations in umbilical cord blood and their association with maternal, fetal and obstetric factors. PLoS One 7, e42827.
Kim, M.S., Ryabets-Lienhard, A., Dao-Tran, A., Mittelman, S.D., Gilsanz, V., Schrager, S.M., Geffner, M.E., 2015. Increased Abdominal Adiposity in Adolescents and Young Adults With Classical Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency. J. Clin. Endocrinol. Metab. 100, E1153–9.
Knickmeyer, R.C., Baron-Cohen, S., 2006. Topical Review: Fetal Testosterone and Sex Differences in Typical Social Development and in Autism. J. Child Neurol. 21, 825–845.
Lane, A., Luminet, O., Nave, G., Mikolajczak, M., 2016. Is there a Publication Bias in Behavioural Intranasal Oxytocin Research on Humans? Opening the File Drawer of One Laboratory. J. Neuroendocrinol. 28. https://doi.org/10.1111/jne.12384
Leslie, M., 2019. The mismeasure of hands?
Manning, J.T., Churchill, A.J.G., Peters, M., 2007. The effects of sex, ethnicity, and sexual orientation on self-measured digit ratio (2D:4D). Arch. Sex. Behav. 36, 223–233.
Manning, J.T., Scutt, D., Wilson, J., Lewis-Jones, D.I., 1998. The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. Hum. Reprod. 13, 3000–3004.
Manning, J.T., Stewart, A., Bundred, P.E., Trivers, R.L., 2004. Sex and ethnic differences in 2nd to 4th digit ratio of children. Early Hum. Dev. 80, 161–168.
McIntyre, M.H., Cohn, B.A., Ellison, P.T., 2006. Sex dimorphism in digital formulae of children. Am. J. Phys. Anthropol. 129, 143–150.
McIntyre, M.H., Ellison, P.T., Lieberman, D.E., Demerath, E., Towne, B., 2005. The development of sex differences in digital formula from infancy in the Fels Longitudinal Study. Proceedings of the Royal Society B: Biological Sciences 272, 1473–1479.
Nadler, A., Zava, D., Ortiz, T., Watson, N., Carre, J., Camerer, C., Nave, G., 2019. Does testosterone impair mens' cognitive empathy? Evidence from two large-scale randomized controlled trials. bioRxiv 516344.
Neyse, L., Johannesson, M., Dreber, A., 2020. 2D:4D Does Not Predict Economic Preferences: Evidence from a Large, Representative Sample. Representative Sample (April.
Nitschke, J.P., Bartz, J.A., 2019. Lower digit ratio and higher endogenous testosterone are associated with lower empathic accuracy. Horm. Behav. 119, 104648.
Okten, A., Kalyoncu, M., Yariş, N., 2002. The ratio of second- and fourth-digit lengths and congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Early Hum. Dev. 70, 47–54.
Pasterski, V., Geffner, M.E., Brain, C., Hindmarsh, P., Brook, C., Hines, M., 2011. Prenatal hormones and childhood sex segregation: Playmate and play style preferences in girls with congenital adrenal hyperplasia. Horm. Behav. 59, 549–555.
Pratt, T. C., Turanovic, J. J., & Cullen, F. T. (2016). Revisiting the criminological consequences of exposure to fetal testosterone: A meta-analysis of the 2D: 4D digit ratio. *Criminology*, *54*(4), 587-620.

Rasband, W.S., 2011. Imagej, us national institutes of health, bethesda, maryland, usa. http://imagej. nih. gov/ij/.
Ribeiro, E., Neave, N., Morais, R.N., Manning, J.T., 2016. Direct Versus Indirect Measurement of Digit Ratio (2D:4D): A Critical Review of the Literature and New Data. Evol. Psychol. 14, 1474704916632536.
Richards, G., 2017. What is the evidence for a link between digit ratio (2D: 4D) and direct measures of prenatal sex hormones?
Richards, G., Browne, W.V., Aydin E., Constantinescu M., Nave G., Kim M.S., Watson, S.J., 2020. Digit ratio (2D:4D) and Congenital Adrenal Hyperplasia (CAH): Systematic Literature Review and Meta-Analysis. Manuscript under review
Richards, G., Bellin, W., Davies, W., 2017. Familial digit ratio (2D:4D) associations in a general population sample from Wales. Early Human Development. https://doi.org/10.1016/j.earlhumdev.2017.06.006
Rivas, M.P., Moreira, L.M.A., Santo, L.D.E., Marques, A.C.S.S., El-Hani, C.N., Toralles, M.B.P., 2014. New studies of second and fourth digit ratio as a morphogenetic trait in subjects with congenital adrenal hyperplasia. Am. J. Hum. Biol. 26, 559–561.
Robertson, J., Zhang, W., Liu, J.J., Muir, K.R., Maciewicz, R.A., Doherty, M., 2008. Radiographic assessment of the index to ring finger ratio (2D:4D) in adults. J. Anat. 212, 42–48.
Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675.
Speiser, P.W., Azziz, R., Baskin, L.S., Ghizzoni, L., Hensle, T.W., Merke, D.P., Meyer-Bahlburg, H.F.L., Miller, W.L., Montori, V.M., Oberfield, S.E., Ritzen, M., White, P.C., Endocrine Society, 2010. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. J. Clin. Endocrinol. Metab. 95, 4133–4160.
Stenstrom, E., Saad, G., Nepomuceno, M.V., Mendenhall, Z., 2011. Testosterone and domain-specific risk: Digit ratios (2D:4D and rel2) as predictors of recreational, financial, and social risk-taking behaviors. Pers. Individ. Dif. 51, 412–416.
Tamiya, R., Lee, S.Y., Ohtake, F., 2012. Second to fourth digit ratio and the sporting success of sumo wrestlers. Evol. Hum. Behav. 33, 130–136.
Tapp, A.L., Maybery, M.T., Whitehouse, A.J.O., 2011. Evaluating the twin testosterone transfer hypothesis: a review of the empirical evidence. Horm. Behav. 60, 713–722.
Trivers, R., Manning, J., Jacobson, A., 2006. A longitudinal study of digit ratio (2D:4D) and other finger ratios in Jamaican children. Horm. Behav. 49, 150–156.
Turanovic, J.J., Pratt, T.C., Piquero, A.R., 2017. Exposure to fetal testosterone, aggression, and violent behavior: A meta-analysis of the 2D:4D digit ratio. Aggress. Violent Behav. 33, 51–61.
van Hemmen, J., Cohen-Kettenis, P.T., Steensma, T.D., Veltman, D.J., Bakker, J., 2017. Do sex differences in CEOAEs and 2D:4D ratios reflect androgen exposure? A study in women with complete androgen insensitivity syndrome. Biol. Sex Differ. 8, 11.
van Leeuwen, B., Smeets, P., Bovet, J., Nave, G., Stieglitz, J., Whitehouse, A., 2020. No Link Between Sex Hormones Around Birth and Economic Preferences. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3565809
Wallen, K., 2009. Does finger fat produce sex differences in second to fourth digit ratios?

Endocrinology.
White, P.C., Speiser, P.W., 2002. Long-term consequences of childhood-onset congenital adrenal hyperplasia. Best Pract. Res. Clin. Endocrinol. Metab. 16, 273–288.
Zheng, Z., Cohn, M.J., 2011. Developmental basis of sexually dimorphic digit ratios. Proc. Natl. Acad. Sci. U. S. A. 108, 16289–16294.