Digit ratio (2D:4D) and Congenital Adrenal Hyperplasia (CAH): Systematic Literature Review and Meta-Analysis

Gareth Richards[a,b*], Wendy V. Browne[c], Ezra Aydin[b], Mihaela Constantinescu[d,e], Gideon Nave[f], Mimi S. Kim[g], & Steven J. Watson[a,h]

[a] School of Psychology, Faculty of Medical Sciences, Newcastle University, UK
[b] Autism Research Centre, Department of Psychiatry, University of Cambridge, UK
[c] Faculty of Education, University of Cambridge, UK
[d] Gender Development Research Centre, University of Cambridge, UK
[e] School of Psychology, University of East London, UK
[f] Marketing Department, The Wharton School, University of Pennsylvania, USA
[g] Keck School of Medicine, University of Southern California, USA
[h] Department of Psychology of Conflict, Risk and Safety, University of Twente, Netherlands

* Corresponding author address: School of Psychology, Newcastle University, 2.27 Ridley Building 1, Queen Victoria Road, Newcastle-upon-Tyne, UK; email: gareth.richards@ncl.ac.uk

Declarations of interest

None

Highlights

- Congenital adrenal hyperplasia (CAH) involves elevated prenatal testosterone.
- There is a sex difference in the digit ratio (2D:4D) that may involve prenatal androgen exposure.
- Article presents a systematic review and meta-analysis of CAH and 2D:4D.
- Low (male-typical) 2D:4D is associated with CAH and the effects are small-to-medium in size.
- Effect sizes observed here are ~50% smaller than those of an earlier meta-analysis.

Abstract

The ratio of length between the second and fourth fingers (2D:4D) is commonly used as an indicator of prenatal sex hormone exposure. Several approaches have been used to try to validate the measure, including examining 2D:4D in people with congenital adrenal hyperplasia (CAH), a suite of conditions characterised by elevated adrenal androgen production secondary to defective steroidogenesis. We present a systematic review and meta-analysis that examines the relationship between these two variables. Twelve articles relating to nine CAH cohorts were identified, and 2D:4D comparisons have been made between cases and controls in eight of these cohorts. Altogether, at least one 2D:4D variable has been compared between n=251 females with CAH and n=358 unaffected females, and between n=108 males with CAH and n=204 unaffected males. A previous meta-analysis (Hönekopp & Watson, 2010) reported lower right hand (R2D:4D) and left hand (L2D:4D) digit ratios in patients with CAH relative to sex-matched controls. Our meta-analysis showed the same pattern, with medium effect sizes for R2D:4D and small effect sizes for L2D:4D. Differences of small magnitude were also observed for M2D:4D, and no significant effects were observed for $D_{[R-L]}$. Notably, the only effects that remained statistically significant when stratified by sex were R2D:4D in males and L2D:4D in females, and the average effect size had reduced by 46.70% since the meta-analysis of Hönekopp and Watson (2010). We also found that individual comparisons in this literature were considerably underpowered, and that patterns of sexual dimorphism in 2D:4D were similar in CAH samples as in typically developing populations. Findings are discussed in relation to the prenatal androgen hypothesis as well as alternative explanations.

Keywords: 2D:4D; CAH; Congenital adrenal hyperplasia; Digit ratio; Differences/disorders of sex development; Foetal testosterone; Prenatal sex hormones

Introduction

Digit ratio (2D:4D) is typically lower in males than females, with a slightly larger sex difference present for the right hand (Hönekopp & Watson, 2010). The measure has been suggested to index the level of exposure to foetal testosterone (Brown, Hines, Fane, & Breedlove, 2002; Manning, Scutt, Wilson, & Lewis-Jones, 1998) or the ratio of foetal testosterone to foetal oestradiol (Lutchmaya, Baron-Cohen, Raggatt, Knickmeyer, & Manning, 2004; Manning, 2011; Zheng & Cohn, 2011). Nevertheless, relatively few studies have validated the measure in human populations. Some research has directly manipulated foetal hormones in animal models (Abbott, Colman, Tiefenthaler, Dumesic, & Abbott, 2012; Auger et al., 2013; Huber, Lenz, Kornhuber, & Müller, 2017; Romano, Rubolini, Martinelli, Alquati, & Saino, 2005; Saino, Rubolini, Romano, & Boncoraglio, 2007; Talarovičová, Kršková, & Blažeková, 2009; Zheng & Cohn, 2011), though the effects reported have not always been consistent. For instance, although Zheng and Cohn (2011) and Huber et al. (2017) both examined the effects of prenatal hormone exposure in CD-1 mice, the studies reported effects in opposing directions. Early manipulation of hormones is unethical in human studies, meaning that researchers have had to rely on other methods, such as correlating 2D:4D with hormone concentrations in amniotic fluid (Lutchmaya et al., 2004; Richards, Browne, & Constantinescu, 2020; Richards, Gomes, & Ventura, 2019; Ventura, Gomes, Pita, Neto, & Taylor, 2013), umbilical cord blood (Çetin, Can, & Özcan, 2016; Hickey et al., 2010; Hollier et al., 2015; Mitsui et al., 2016, 2015; Whitehouse et al., 2015), or the maternal circulation (Barona, Kothari, Skuse, & Micali, 2015; Hickey et al., 2010; Richards et al., 2019; Ventura et al., 2013). The results of studies in humans broadly point toward a negative correlation between foetal testosterone exposure and 2D:4D, although statistically significant effects are accompanied by many null findings (Richards, 2017), and publication bias may be an issue.

Another approach for determining the efficacy of 2D:4D has been to examine whether it is associated with medical conditions characterised by atypical androgen activity. Two studies (Berenbaum, Bryk, Nowak, Quigley, & Moffat, 2009; van Hemmen, Cohen-Kettenis, Steensma, Veltman, & Bakker, 2017) have reported evidence of feminised 2D:4D ratios in phenotypically female (46XY) individuals with complete androgen insensitivity syndrome (CAIS), although it should be noted that the variance

for 2D:4D in this population appears to be comparable to that of controls despite the complete lack of androgen sensitivity (Berenbaum et al., 2009; see also commentary by Wallen, 2009). Manning, Kilduff, and Trivers (2013) showed that digit ratios were higher (i.e. more female-typical) in males with Klinefelter syndrome (47XXY) than in their unaffected relatives. However, this effect is difficult to interpret considering that prenatal testosterone levels in males with Klinefelter syndrome do not appear to differ from those of typically developing males (Ratcliffe et al., 1994).

A promising area of research has examined individuals with congenital adrenal hyperplasia (CAH). CAH is a family of autosomal recessive conditions characterised by impairment of one of five enzymes required to synthesise cortisol from cholesterol. This causes an accumulation of adrenocorticotrophic hormone (ACTH) secondary to negative feedback, which results in overstimulation of the adrenal cortex and increased adrenal androgen production (New, 2006). Most cases (90–95%) of CAH are caused by 21–hydroxylase (21–OH) deficiency, with three main phenotypes being distinguishable (for a comparison of symptom profiles, see New, 2006). The most severe form, classical salt-wasting (SW) CAH, involves impairment of aldosterone synthesis, a symptom that is absent overall in classical simple-virilizing (SV) CAH; both SW and SV are characterised by genital ambiguity in female (46XX) patients. Pharmacological treatment for classical CAH due to 21-OH deficiency typically begins soon after birth, and the condition has been found to occur in approximately 1 in 14,000 live births (Pang et al., 1988). Non-classical CAH due to 21-OH deficiency does not present with aldosterone impairment nor typically with genital ambiguity, and can go undetected (Levine et al., 1980) particularly in males. The non-classical or late-onset form is diagnosed when symptoms present later in life (Kisch, Laurian, & Hoerer, 1987; New, Dupont, Pollack, & Levine, 1981), and is more common than classical CAH, with reported prevalence ranging from 1 in 27 to 1 in 300, depending on the ethnic group studied (Hannah-Shmouni et al., 2017; New, 2006).

CAH provides an opportunity for researchers to examine the organisational effects of elevated androgen exposure during gestation. There is some evidence for behavioural masculinisation and defeminisation in CAH, with such issues being particularly pertinent in 46XX female-assigned cases because prenatal androgen concentrations may not only affect external somatic sex structures, but also bipotential areas in the

brain, leading to modification of behavioural/psychological outcomes (see Cohen-Bendahan, van de Beek, & Berenbaum, 2005; Hines, 2004; Hines, Constantinescu, & Spencer, 2015; Jordan-Young, 2012). The early androgenic effects of CAH in males however are less clear, as feedback mechanisms may lead to normalisation of androgen levels via reduced production by the testes (Pang, Levine, Chow, Faiman, & New, 1979; for a discussion, see Mathews et al., 2004). Evidence for this is provided by the observation that amniotic testosterone levels for 46XY CAH foetuses tend not to be clearly distinguishable from those of typically developing 46XY foetuses (Pang et al., 1980; Wudy, Dörr, Solleder, Djalali, & Homoki, 1999), though such observations have relied on very small samples. However, it does appear possible that following an initial elevation, testosterone concentrations may be relatively typical in males with CAH.

Hines et al. (2003) reported that females with CAH outperformed their unaffected female relatives on two tasks assessing targeting performance. The tasks employed included measures of visuomotor spatial ability that have been found to demonstrate a large male advantage in the typically-developing population (Watson, 2001). In a study by Collaer, Brook, Conway, Hindmarsh, and Hines (2009), females with CAH also outperformed unaffected female relatives on motor and visuomotor tasks (grip strength and targeting), which have shown a male advantage in previous research (e.g. Miller, MacDougall, Tarnopolsky, & Sale, 1993), even after controlling for weight and height. The enhanced targeting performance in females with CAH was still present after adjusting for grip strength, leading the researchers to point towards an organisational influence of prenatal androgens on the neural regions dedicated to targeting (Collaer et al., 2009).

Behavioural masculinisation in females with CAH may only occur in traits which show a particularly large sex difference. Alternatively, as studies of CAH populations typically utilise small samples due to the rarity of the condition, they may lack the statistical power required to reliably detect effects of small or medium size. A way to overcome this limitation is to pool the findings of individual studies into a meta-analysis, which provides an indication of the presence (or absence) of an effect as well as its size. Using this technique, Puts et al. (2008) reported that females with CAH display an advantage on spatial tasks, whereas males with CAH display a disadvantage. However, although a more recent meta-analysis (Collaer & Hines, 2020) including a

larger number of samples replicated the finding of reduced overall spatial ability in males with CAH relative to males controls, it did not find any difference between female CAH cases and controls. A possible interpretation of these contradictory findings is that early studies (and hence meta-analyses of those early studies) are more likely to report statistically significant effects in small samples whereas later studies often report smaller (or null) effects when attempting to replicate them in larger samples.

As CAH (at least in females) is associated with elevated prenatal androgen levels, and 2D:4D is hypothesised to indicate individual differences in foetal testosterone exposure, it follows that they should be related. A meta-analysis of early studies of CAH case-control studies (Hönekopp & Watson, 2010) showed that digit ratio for the right hand (R2D:4D) (d = -0.91, $p < 0.001$) and left hand (L2D:4D) (d = -0.75, p = 0.007) were significantly lower (i.e. more male-typical) in females with CAH relative to female controls; R2D:4D (d = -0.94, p = 0.061) and L2D:4D (d = -0.63, p = 0.013) were also lower in males with CAH relative to male controls, albeit the effect for R2D:4D was not statistically significant at the $p < 0.050$ level. This pattern of results is consistent with prenatal androgen exposure being elevated in both males and females with CAH, and so runs contrary to the idea that feedback mechanisms can normalise testosterone levels in males with CAH via downregulation of testicular androgen production.

Although behavioural effects associated with CAH may be explainable by environmental influences (Hines et al., 2015; Jordan-Young, 2012) such as the presence and extent of genital virilisation, alterations in the way that parents, teachers and others interact with children with CAH, it seems unlikely that these could affect a person's digit ratio. However, it should be acknowledged that although CAH research may indicate that elevated prenatal testosterone exposure causes physical differences, such as a masculinised 2D:4D ratio, these findings cannot necessarily be extrapolated to indicate a similar influence on the developing brain.

The current study aims to build on the earlier meta-analysis by Hönekopp and Watson (2010) by updating their analysis to include new studies and incorporating a full systematic review of the relevant literature. Hönekopp and Watson (2010) incorporated their analysis of the relationship between 2D:4D and CAH into an article with a much

broader remit. Therefore, this literature has yet to be comprehensively reviewed. We also extend their analysis in other ways. As it has been suggested that the right-left difference in digit ratio ($D_{[R-L]}$) can provide a further marker of prenatal sex hormone activity, with low R2D:4D relative to L2D:4D hypothesised to indicate high androgen exposure (Manning, 2002; Manning, Kilduff, Cook, Crewther, & Fink, 2014), this variable is also considered in the current study. Furthermore, some studies report on the average 2D:4D across the right and left hands (M2D:4D). Because studies comparing digit ratios between patients with CAH and controls have not so far investigated $D_{[R-L]}$ or M2D:4D, we contacted the authors of relevant papers to request the necessary data. We hypothesised that R2D:4D, L2D:4D, and $D_{[R-L]}$ would each be significantly lower in males and females with CAH relative to male and female controls, respectively. Although not initially considered in our pre-registration (see next section), we also hypothesised that M2D:4D would be significantly lower in males and females with CAH relative to male and female controls, respectively, and, additionally, examined whether 2D:4D variables exhibit similar sexual dimorphism in CAH samples as they do in general population studies.

Material and Methods

We pre-registered our review and analysis plan on the Open Science Framework (osf.io/n2hse) prior to beginning the research. Studies were considered eligible for inclusion where they made at least one comparison of 2D:4D between individuals with a diagnosis of any form of CAH with a control group. We made no limitations on year or language of publication. Studies were excluded where they did not report the statistics necessary to make a comparison between CAH and sex-matched controls, or if they did not report primary data.

We searched (keyword, title, and abstract; no publication date restrictions were imposed) Ovid MEDLINE, Embase, PsychINFO, Web of Science, and Scopus using the following search terms: (Digit ratio OR Digit length ratio OR Digital ratio OR Finger ratio OR Finger length ratio OR 2D:4D OR 2D4D OR Second to fourth OR Second-to-fourth OR Second-fourth OR 2nd to 4th OR 2nd-to-fourth OR 2nd-4th OR Ring to index OR Ring-to-index OR Index to ring OR Index-to-ring) AND (Congenital adrenal hyperplasia OR CAH). We also examined the reference lists of relevant papers,

a bibliographic article of 2D:4D studies published between 1998 and 2008 (Voracek & Loibl, 2009), and an online database of digit ratio research (Fink & Manning, 2018), which (as of 09/12/2018) included 817 references. Additionally, we contacted 70+ researchers within the digit ratio and CAH fields to try to identify any published or unpublished data that we had not already included within our review.

We identified 3,705 articles through literature searches, four from reference lists of relevant papers, and three by contacting authors in the field. Once duplicates had been excluded, this resulted in 3,408 articles. The title for each was read, and the article was excluded from further consideration if it did not appear to relate to either 2D:4D or CAH. The abstracts of 615 potentially relevant articles were then accessed (please note that in cases where the article did not include an abstract, the Introduction, Introduction and Method, or whole article was read). Relevant materials that were not available in English were translated. See **Figure 1** for the PRISMA flow diagram (Moher et al., 2009).

We then used a standard data extraction form created in Microsoft Excel, which included fields for information relating to the paper (e.g. authors, year and place of publication), participants and setting (e.g. country, sample size, sex, age, diagnoses), key study details (e.g. characteristics of participants included in the CAH and control group[s], method[s] used for measuring 2D:4D, descriptive statistics for age and digit ratio variables for each participant group [wherever possible]), and a summary of results. When relevant data were missing or ambiguous, the study authors were contacted for clarification. All data were extracted by GR other than for Nave et al., (2020), which were inputted by SW, and for a Turkish language paper (Kocaman et al., 2017), from which data were extracted by a native Turkish speaker (EA). All data included in the meta-analysis were independently checked by SW, with any disagreements resolved through discussion until a 100% agreement rate was achieved.

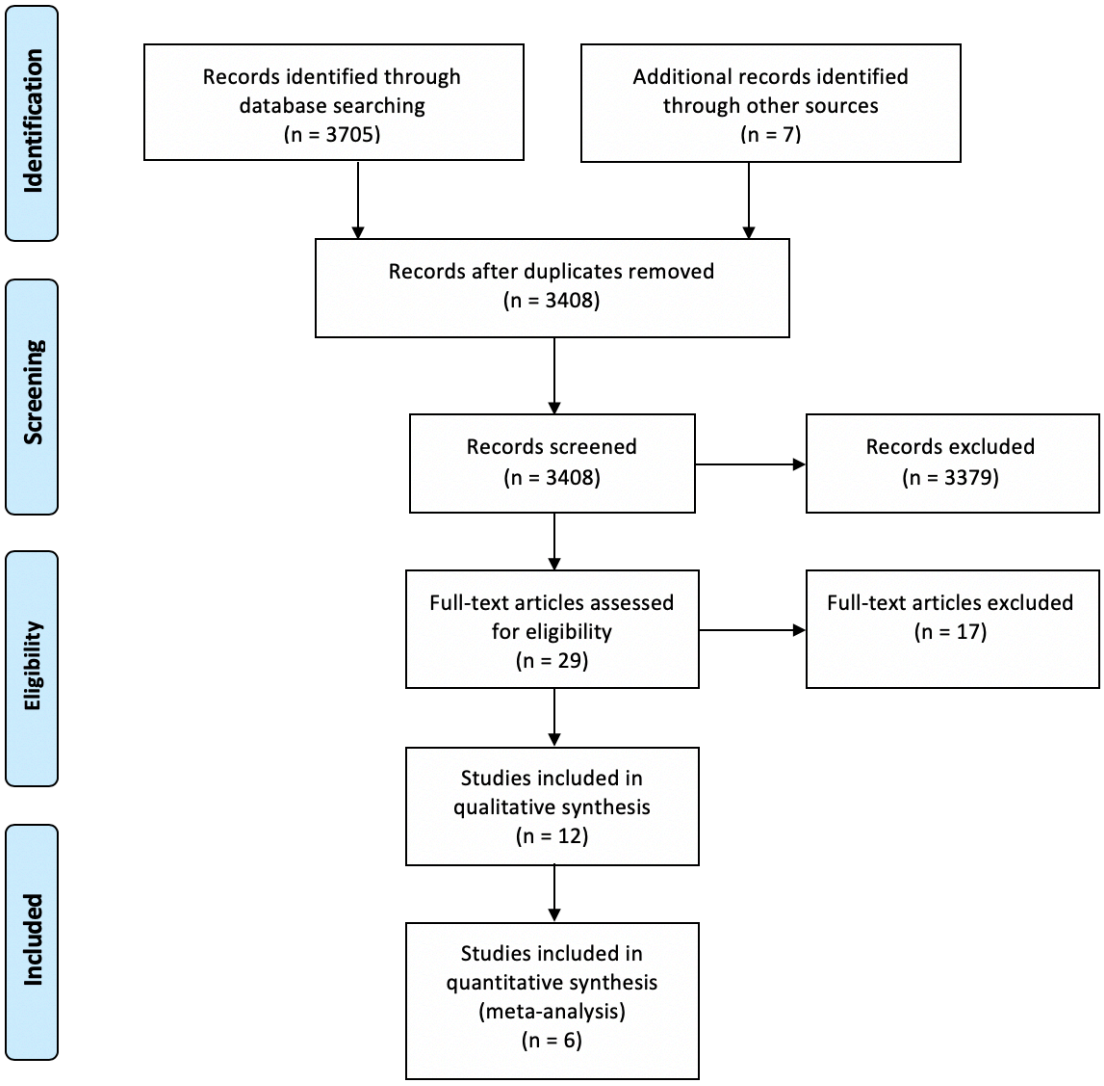

Figure 1. PRISMA flow diagram showing study selection for the systematic literature review and meta-analysis.

Systematic Literature Review and Meta-Analysis

Twelve articles examined 2D:4D in CAH populations and were included in the literature review (**Table 1**). Of these, four (Brown et al., 2002; Buck, Williams, Hughes, & Acerini, 2003; Ciumas, Hirschberg, & Savic, 2009; Ökten, Kalyoncu, & Yariş, 2002) were present in the earlier meta-analysis by Hönekopp and Watson (2010), five (Kim et al., 2017; Kocaman et al., 2016, 2017; Oświęcimska et al., 2012; Rivas et al., 2014) had been published since, one (Nave et al., 2020) was currently under review[1] and two,

[1] The data for Nave et al. were acquired prior to publication of the current article; the current review article was submitted simultaneously with the manuscript that presents the empirical study by Nave et al.

both relating to the same dataset (Constantinescu et al., 2010; Constantinescu, 2009), were unpublished. All were full-length journal articles other than Kim et al. (2017) and Kocaman et al. (2016), which were published abstracts, and subsequently have appeared as full-length journal articles (Kocaman et al., 2017; Nave et al., 2020), Constantinescu (2009), which was an unpublished MPhil thesis, and Constantinescu et al. (2010), which was a conference poster.

The studies included in this review were conducted in six countries (Brazil, Poland, Sweden, Turkey, UK, US), and the type of control group to which patients with CAH were compared differed considerably. Some studies employed healthy adult controls without family history of neuropsychiatric conditions (Ciumas et al., 2009), healthy children who had been seen at an outpatient clinic (Ökten et al., 2002), children who had been screened for autism and psychiatric disorders (Kocaman et al., 2016, 2017), otherwise healthy children who had been assessed at an outpatient endocrine clinic due to concerns over short stature (Buck et al., 2003; Nave et al., 2020), university students (Rivas et al., 2014), and unaffected relatives of patients with CAH (Brown et al., 2002; Constantinescu et al., 2010; Constantinescu, 2009; note that some [but not all] of the control participants in Nave et al. [2020] were relatives of their participants with CAH). Some control groups were matched for chronological age (Buck et al., 2003; Ciumas et al., 2009; Ökten et al., 2002) and handedness (Ciumas et al., 2009; Ökten et al., 2002), and one study (Nave et al., 2020) statistically controlled for individual differences in chronological age, bone age, ethnicity, height, puberty status, and ethnicity. For other studies no such controls were implemented (Rivas et al., 2014) or the details are unclear (Kocaman et al., 2016, 2017). Lack of effective control for age between CAH and comparison groups is a point that has been raised as a possible explanation for the inconsistent nature of findings in this literature (McIntyre, Cohn, & Ellison, 2006, p. 149; Nave et al., 2020). Only one study (Kim et al., 2017) examined whether 2D:4D differs between CAH forms. Although the authors reported no significant difference between patients with classical SW (n=63) or SV (n=20) varieties, further examination is warranted, and particularly so regarding other forms, such as non-classical CAH. Kim et al. (2017) also observed no significant interactions between CAH form, sex, and bone age.

Table 1. Studies of 2D:4D in CAH samples included in the systematic literature review.

Authors	Year	Country	Place of publication	2D:4D measure(s)	Females with CAH		Female controls		Males with CAH		Male controls	
					N	*Age*	*n*	*Age*	*n*	*Age*	*n*	*Age*
Brown et al.	2002	UK	*Hormones and Behavior*	Photocopies	13	Range = 7–44 Average = 15	44	Range = 12–44 Average = 18	16	Range = 5–21 Average = 11	28[1]	Range = 9–34 Average = 15
Ökten et al.[2]	2002	Turkey	*Early Human Development*	Photocopies, X-rays	17	Range = 0–13.3 *M* = 4.6 (*SD* = 4.2)	34[3]	Age-matched	9	Range = 0–13.3 *M* = 4.6 (*SD* = 4.2)	18[3]	Age-matched
Buck et al.[4]	2003	UK	*Human Reproduction*	X-rays	66	Range = 1.1–16.2 Median = 8.5	69	Range = 1.9–17 Median = 9.3	0	-	77	Range = 2.1–20.3 Median = 13.86
Ciumas et al.	2009	Sweden	*Cerebral Cortex*	Direct (reported) Photocopies (not reported)	11	Range = 20–38 *M* = 30 (*SD* = 8)	13	Range = 20–36 *M* = 26 (*SD* = 7)	0	-	13	Range = 21–36 *M* = 28 (*SD* = 6)
Constantinescu[a, 5]	2009	UK	Unpublished MPhil thesis	Direct + photocopies (combined)	40	Range = 4–11.83 *M* = 7.50 (*SD* = 2.20)	17	Range = 4.08–12.42 *M* = 7.13 (*SD* = 2.52)	24	Range = 4–11.25 *M* = 7.43 (*SD* = 1.98)	7	Range = 5.08–10.50 *M* = 7.89 (*SD* = 1.95)
Constantinescu et al.[a, 6]	2010	UK	Unpublished conference poster	Direct + photocopies (combined)	40	Range = 4–11.83 *M* = 7.50 (*SD* = 2.20)	17	Range = 4.08–12.42 *M* = 7.13 (*SD* = 2.52)	24	Range = 4–11.25 *M* = 7.43 (*SD* = 1.98)	7	Range = 5.08–10.50 *M* = 7.89 (*SD* = 1.95)
Oświęcimska et al.	2012	Poland	*Neuroendocrinology Letters*	X-rays	19	Range = 3.7–19 *M* = 13.8 (*SD* = 4.07)	0	–	0	–	0	–
Rivas et al.	2014	Brazil	*American Journal of Human Biology*	Direct	31	*M* = 10.7	100	Range = 16–18	9	*M* = 10.2	100	Range = 16–18
Kocaman et al.[b]	2016	Turkey	*Acta Physiologica*	Direct	30	Range = 3–15	30	Age-matched	0	–	0	–
Kocaman et al.[b, 7]	2017	Turkey	*Anadolu Psikiyatri Dergisi*	Direct	28[8]	*M* = 8.84 (*SD* = 4.06) or *M* = 10.90 (*SD* = 1.46)	49	*M* = 8.84 (*SD* = 4.06) or *M* = 10.90 (*SD* = 1.46)	4[8]	*M* = 8.84 (*SD* = 4.06) or *M* = 10.90 (*SD* = 1.46)	10	*M* = 8.84 (*SD* = 4.06) or *M* = 10.90 (*SD* = 1.46)
Kim et al.[c, 9]	2017	US	*Endocrine Reviews*	X-rays	40	Baseline *M* = 4.6 (*SD* = 2.8) Follow-up *M* = 9.9 (*SD* = 3.3)	0	–	43	Baseline *M* = 4.9 (*SD* = 2.9) Follow-up *M* = 11.7 (*SD* = 3.6)	0	–
Nave et al.[c]	2020	US	Under review (*Hormones and Behavior*)	X-rays	45	Baseline range: 1.1-18.7	31	Baseline range: 2.6-19.7	46	Baseline range: 1.1-18.7	39	Baseline range: 2.6-19.7

Note. [a], [b], and [c] indicate sources reporting on the same cohorts.

[1] Although Brown et al.'s (2002) sample includes data from n=28 male controls, L2D:4D was not recorded for n=1 of these participants; therefore, n=27 for L2D:4D, M2D:4D, and $D_{[R-L]}$.

[2] The age-range reported by Ökten et al. (2002) is based on all (male + female) CAH patients; although control participants were matched for age, the exact age-ranges and Ms and SDs were not reported.

[3] Ökten et al. (2002) collected 2D:4D data from n=52 female controls and n=52 male controls but only compared the CAH samples with n=34 age-matched females and n=18 age-matched males.

[4] Buck et al. (2003) examined X-rays for n=71 females with CAH, n=76 female controls, and n=82 male controls (overall, n=17 were rejected because of poor quality films).

[5] Constantinescu (2009) reported that their sample initially consisted of n=40 females with CAH (age range = 4-11.83, M = 7.49, SD = 2.19), n=25 males with CAH (age-range = 4-11.25, M = 7.29, SD = 2.04), n=18 female controls (age-range = 4.08-12.42, M = 7.33, SD = 2.59), and n=9 male controls (age-range = 5.08-10.50, M = 7.62, SD = 1.77). However, 2D:4D data were unavailable for n=1 male with CAH, n=1 female control, and n=2 male controls; we therefore present here the Ns, age ranges, Ms, and SDs (determined from the original data) based on only those participants for which 2D:4D data were available.

[6] The age range (4-11.8 years) and M age (7.4) reported by Constantinescu et al. (2010) are based on all participants (male + females, with and without CAH); we therefore report here the age ranges, Ms and SDs calculated from the original data for each subgroup.

[7] It is unclear what the age of participants was in Kocaman et al. (2017), as the study reports two separate Ms and SDs; it is unclear whether these relate to subgroups, and so both Ms and SDs are reported here for each group of participants.

[8] Kocaman et al. (2017) reported that they collected data from n=34 children with CAH; n=2 were removed from analysis because they did not provide consent, and n=1 other appears to have been dropped from the analysis (as the overall n=31), though the reason is unclear.

[9] Age-ranges for Kim et al. (2017) are based on all (male + female) CAH patients.

Comparisons of 2D:4D between CAH cases and controls

Findings from studies comparing 2D:4D between CAH samples and control samples are reported in **Table 2**. Significantly lower 2D:4D has been reported in five CAH cohorts (Brown et al., 2002; Ciumas et al., 2009; Kocaman et al., 2016, 2017; Ökten et al., 2002; Rivas et al., 2014). However, three studies (Buck et al., 2003; Constantinescu et al., 2010; Constantinescu, 2009; Nave et al., 2020) reported only null-findings. Notably, these included the largest (Buck et al., 2003: n=66), second largest (Nave et al., 2020; n=45), and third largest (Constantinescu, 2009; n=40) samples of females with CAH, as well as the largest (Nave et al., 2020; n=45) and second largest (Constantinescu, 2009; n=24) samples of males with CAH. However, to interpret these findings accurately, some further consideration of the studies' methodologies is required. Constantinescu and colleagues used an unusual approach for measuring 2D:4D: a combination of both direct (calliper) and indirect (photocopy) measures were collected, with both types of measurements being recorded for a subset of participants. For those from whom only direct measures were available, these were then adjusted so that they resembled photocopy measures. They did this by dividing the overall mean value from the photocopy measurements by the mean for the calliper measurements, then multiplying this by the calliper measurement for each individual. Additionally, although the CAH samples were relatively large (female n=40, male n=24), the comparison samples were not (female n=17, male n=7), meaning that the benefit in terms of statistical power associated with large CAH samples was somewhat undermined by the small control groups. Buck et al. (2003) on the other hand did not examine males with CAH, only recorded L2D:4D (and not R2D:4D), and measured digit ratios from X-rays. Likewise, Nave et al. (2020) compared only L2D:4D (from X-rays) between patients with CAH and controls (although they did examine both males and females).

Brown et al. (2002) provided evidence to suggest the difference in 2D:4D between patients with CAH and controls may be due to environmental influences (e.g. the elevated prenatal testosterone exposure characteristic of CAH) rather than shared genetics, as they observed lower 2D:4D in male patients with CAH than in their unaffected male relatives (R2D:4D, $p = 0.033$, $d = -1.191$; L2D:4D, $p = 0.011$, $d = -1.592$) (Note that these effects are incorrectly reported in the original paper as R2D:4D,

$p = 0.01$, $d = 1.0$; L2D:4D, $p < 0.04$, $d = 0.9$). However, this analysis relied on a very small sample (males with CAH, n=6; unaffected males, n=6), and a larger study (Constantinescu et al., 2010; Constantinescu, 2009) found no such differences in males or females (males with CAH, n=24; unaffected males, n=9; females with CAH, n=40; unaffected females, n=18). Interestingly though, Constantinescu (2009) reported only weak to moderate sized correlations (many of which were not statistically significant) between digit ratios of children and those of their mothers, whereas previous studies (e.g. Hiraishi, Sasaki, Shikishima, & Ando, 2012; Kalichman, Batsevich, & Kobyliansky, 2019; Voracek & Dressler, 2009) suggest that genetic factors contribute substantially to the phenotypic expression of this trait.

The only study to observe a significant effect in the opposite direction than expected was Rivas et al. (2014), who reported L2D:4D in males with CAH to be higher than that of male controls. However, only 9 males with CAH were included in this analysis (in comparison to 100 male controls), and the mean age of the CAH males was 10.2 years whereas the control group consisted of students aged 16-18 years. This sample was also reported to be ethnically diverse, which could be important considering that 2D:4D can vary more by ethnicity than by sex (de Sanctis et al., 2017; Loehlin, McFadden, Medland, & Martin, 2006; Manning, Churchill, & Peters, 2007; Manning, Stewart, Bundred, & Trivers, 2004). In addition, and likely of greater importance, there appear to be errors in the reporting of some of the standard deviations/standard errors (i.e. some were implausibly smaller than others) (see text on p. 560 as well as the error bars on Figure 1 of that paper).

Ökten et al. (2002) reported that 10 girls with CAH who were less than 2 years old had significantly lower R2D:4D and L2D:4D than age-matched female controls. This could suggest that differences in 2D:4D appear early in life, which is consistent with the idea that they relate to prenatal androgen exposure. However, Constantinescu (2009) found only marginally ($p = 0.063$, $d = -0.72$) lower L2D:4D in girls with CAH compared to their unaffected female relatives aged 4–7.99 years, and no difference for R2D:4D; there were also no differences for R2D:4D or L2D:4D in boys of this age. No differences were observed between girls with CAH and unaffected girls or between boys with CAH and unaffected boys aged 8–12.42 years for either R2D:4D or L2D:4D.

No studies reported whether M2D:4D or $D_{[R-L]}$ differed between CAH populations and controls. However, we were able to conduct such analyses from the original data of Brown et al. (2002) and Constantinescu (2009) (see **Table 2**). For Brown et al. (2002), we observed that M2D:4D was significantly lower in females with CAH than female controls. M2D:4D was also lower in males with CAH than male controls, though the effect was just short of statistical significance ($p = 0.051$, $d = -0.633$). A paired-samples *t* test determined that M2D:4D was significantly lower in males with CAH (n=6, $M = 0.911$, $SD = 0.042$) than in their unaffected male relatives (n=6, $M = 0.955$, $SD = 0.033$), $t(5) = -4.043$, $p = 0.01$, $d = -1.164$. However, in Constantinescu's (2009) data, there was no difference in M2D:4D between females with CAH and unaffected females; there was also no difference between males with CAH and unaffected males.

When examining $D_{[R-L]}$ in the data of Brown et al. (2002), we found no significant differences between females with CAH and female controls, or between males with CAH and male controls. A paired-samples *t* test determined that $D_{[R-L]}$ also did not differ between males with CAH (n=6, $M = -0.008$, $SD = 0.032$) and their unaffected male relatives (n=6, $M = 0.004$, $SD = 0.041$), $t(5) = -0.704$, $p = 0.513$, $d = -0.336$. In Constantinescu's (2009) dataset, $D_{[R-L]}$ was marginally lower in males with CAH than unaffected males ($p = 0.082$). However, marginally higher $D_{[R-L]}$ was observed in females with CAH compared to unaffected females ($p = 0.068$).

Digit ratio	Sex	Study	CAH patients			Unaffected controls			Difference				Power[a]
			n	*M*	*SD*	*n*	*M*	*SD*	*t*	*df*	*p*	*d*	
R2D:4D	F	Brown et al. (2002)[1]	13	0.957	0.038	43[2]	0.981	0.032	–2.290	54	**0.026**	**–0.718**	0.450
	F	Ökten et al. (2002) (photocopies)	17	0.96	0.06	34	1.0	0.06	–2.244	49	**0.029**[3]	**–0.667**	0.496
	F	Ökten et al. (2002) (X-rays)	17	0.99	0.02	34	1.00	0.01	–2.393	49	**0.021**[4]	**–0.711**	0.496
	F	Ciumas et al. (2009)[5]	11	0.956	0.024	13	0.985	0.016	–3.533	22	**0.002**	**–1.447**	0.281
	F	Constantinescu (2009)[6]	40	0.960	0.046	17	0.950	0.023	1.097§	53.334	0.278	0.246	0.518
	F	Rivas et al. (2014)	31	0.950	0.0077[7]	100	0.980	0.0026[7]	–33.501	129	**< 0.001**	**–6.887**	0.814
	M	Brown et al. (2002)[1]	16	0.937	0.045	28	0.957	0.038	–1.562	42	0.126	–0.492	0.359
	M	Ökten et al. (2002) (photocopies)	9	0.92	0.04	18	0.97	0.03	–3.653	25	**0.001**	**–1.491**	0.227
	M	Ökten et al. (2002) (X-rays)	9	0.98	0.03	18	0.99	0.02	–1.035	25	0.311[8]	–0.423	0.227
	M	Constantinescu (2009)[6]	24	0.941	0.042	7	0.970	0.038	–1.624	29	0.115	–0.704	0.211
	M	Rivas et al. (2014)	9	0.960	0.0220[7]	100	0.957	0.0031[7]	1.284	107	0.202	0.447	0.309
	F+M	Kocaman et al. (2017)	31	0.96	0.02	59	1.00	0.03	–6.676	88	**< 0.001**	**–1.481**	
L2D:4D	F	Brown et al. (2002)[1]	13	0.952	0.025	43[2]	0.968	0.032[9]	–1.676	54	0.100	–0.523	0.118
	F	Ökten et al. (2002) (photocopies)	17	0.92	0.05	34	0.99	0.06	–4.140	49	**< 0.001**[10]	**–1.230**	0.128
	F	Ökten et al. (2002) (X-rays)	17	0.99	0.04	34	0.99	0.02	0.000	49	1.000[11]	0.000	0.128
	F	Buck et al. (2003)	66	0.925	0.021	69	0.927	0.029	–0.457	133	0.648	–0.079	0.292
	F	Ciumas et al. (2009)[5]	11	0.979	0.027	13	1.005	0.033	–2.086	22	**0.049**	**–0.855**	0.088
	F	Constantinescu (2009)[6]	40	0.944	0.036	17	0.955	0.028	–1.178§	38.663	0.246	–0.325	0.132
	F	Rivas et al. (2014)	31	0.947	0.0114[7]	100	0.977	0.0028[7]	–24.242	129	**< 0.001**	**–4.983**	0.219
	F	Nave et al. (2020)	45	0.917	0.023	31	0.925	0.024	-1.51	74	0.136	-0.352	0.179
	M	Brown et al. (2002)[1]	16	0.931	0.034	27	0.955	0.039[12]	–2.053	41	**0.047**	**–0.644**	0.104
	M	Ökten et al. (2002) (photocopies)	9	0.91	0.06	18	0.94	0.04	–1.553	25	0.133[13]	–0.634	0.081
	M	Ökten et al. (2002) (X-rays)	9	0.98	0.03	18	1.00	0.03	–1.633	25	0.115[14]	–0.667	0.081
	M	Constantinescu (2009)[6]	24	0.959	0.035	7	0.943	0.055	0.906	29	0.372	0.400	0.078
	M	Rivas et al. (2014)	9	0.983	0.0267[7]	100	0.950	0.0028[7]	12.186	107	**< 0.001**[15]	**4.241**	0.095
	M	Nave et al. (2020)	45	0.913	0.023	39	0.913	0.020	-0.155	82	0.877	-0.034	0.166
	F+M	Kocaman et al. (2017)	31	0.96	0.02	59	0.99	0.04	–3.919	88	**< 0.001**	**–0.869**	
M2D:4D	F	Brown et al. (2002)†	13	0.954	0.026	43	0.975	0.030	–2.208	54	**0.032**	**–0.720**	0.175
	F	Constantinescu (2009)†	40	0.952	0.035	17	0.952	0.022	–0.022§	47.658	0.983	0.000	0.201
	M	Brown et al. (2002)†	16	0.934	0.037	27	0.957	0.035	–2.007	41	0.051	–0.643	0.311
	M	Constantinescu (2009)†	24	0.950	0.028	7	0.956	0.031	–0.546	29	0.589	–0.209	0.187
$D_{[R-L]}$	F	Brown et al. (2002)†	13	0.005	0.037	43	0.013	0.024	–0.891	54	0.377	–0.292	
	F	Constantinescu (2009)†	40	0.016	0.041	17	-0.005	0.026	1.863	55	0.068	0.564	

M	Brown et al. (2002)†	16	0.006	0.033	27	0.003	0.028	0.304	41	0.763	0.100
M	Constantinescu (2009)†	24	-0.018	0.053	7	0.027	0.072	–1.804	29	0.082	–0.783

Table 2. Comparisons of digit ratio variables between patients with CAH and unaffected controls.
Note. CAH = congenital adrenal hyperplasia; F = female; M = male; negative *d* values indicate effects in the predicted (i.e. CAH<control) direction; effects in bold are statistically significant ($p < 0.05$).
[a] Estimated statistical power was computed for comparisons of R2D:4D and L2D:4D (i.e. what has actually been explored in the extant literature) using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007); these calculations were based on the effect sizes observed in our meta-analysis (see next section [though note that we substituted *g* for *d* here]), and use of a two-tailed independent samples *t* test with α set at $p < 0.050$.
[1] As we noted errors in the reporting of *SD*s in Brown et al. (2002), we recalculated the *M*s and *SD*s (to three decimal places) and re-ran the statistical tests. We report here the outcome of our re-analysis (and also report the exact *p* values and effect sizes).
[2] In Brown et al. (2002, p. 381) N is listed as 44, though on Figure 1 of that paper, N is listed as 43 (in the dataset we obtained for this study, n=43).
[3] This value is listed as '0.3' in Ökten et al. (Table 1) and as '0.01' in the text (p. 50) of that paper.
[4] This value is listed as '0.07' in Ökten et al. (2002) (Table 3).
[5] Ciumas et al. analysed their data using one-way ANOVA models with females with CAH (n=11), control females (n=13), and control males (n=13) as the three groups; a significant overall effect was reported for R2D:4D ($F = 6.07$, $p = 0.0074$) but not for L2D:4D ($F = 1.9$, $p = 0.178$).
[6] As Constantinescu (2009) did not report the df values for their comparisons, we reran the analyses using the original data (with values rounded to three decimal places), and report their outcomes here.
[7] At least some of the *SD*s reported by Rivas et al. (2014) appear to be erroneous.
[8] This value is listed as '0.7' in Table 3 of Ökten et al.
[9] In the original paper by Brown et al. (2002, p. 383), this value is reported as being '0.005' (the value we report here was calculated from the original dataset).
[10] This value is listed as p = 0.0004 in Table 1 and on p. 50 of Ökten et al.
[11] This value is listed as '0.9' in Table 3 of Ökten et al.
[12] In the original paper by Brown et al. (2002, p. 383), this value was reported as being '0.007' (although is correctly reported as '0.039' elsewhere on p. 383) (the value we report here was calculated from the original dataset).
[13] This value is listed as '0.09' in Ökten et al. (Table 1).
[14] This value is listed as '0.1' in Table 3 or Ökten et al.
[15] This value is listed as '< 0.01' in Rivas et al. (p. 560, and Figure 1).
[§] Equal variances not assumed.
[†] Comparison not reported in the original article (independent *t* tests were conducted using descriptive statistics presented in the original papers, other than for Brown et al. [2002] and Constantinescu [2009], for which comparisons were conducted using the original data).

Meta-analysis of the difference in 2D:4D between CAH cases and controls

To determine quantitatively whether digit ratio variables (R2D:4D, L2D:4D, M2D:4D, and $D_{[R-L]}$) differ significantly between females with CAH and control females, and between males with CAH and control males, we conducted meta-analyses using the R package metafor (Viechtbauer, 2010).

The inclusion criteria for the meta-analysis were that studies had to report primary 2D:4D data for humans with CAH as well as for controls, and that they needed to report effect sizes and/or statistics from which effect sizes could be calculated. We contacted the first/corresponding authors of the relevant studies to request the data necessary to calculate effect sizes if they were not available within the published articles. If we did not hear back, we subsequently contacted other authors for whom contact details could be obtained. The standard deviations reported in Brown et al. (2002) were clearly standard errors (and we checked this using the original dataset), so we were able to calculate the correct values and include this study. Rivas et al. (2014) also reported standard deviations that very likely contain an error given the implausibly large effect sizes generated in this study, and standard deviations far smaller for some subsamples than is typical in 2D:4D literature. However, unlike with Brown et al. (2002), it was not obvious that the reported values definitely referred to standard errors and so we excluded this study from the meta-analysis.

Random-effects models with the restricted maximum-likelihood estimation method were calculated to account for likely heterogeneity in the data. To best compensate for our small samples, we report standardised mean differences in the form of Hedges' *g* (Hedges & Olkin, 1985). We report heterogeneity in terms of I^2. For completeness, we also report Cochran's Q as a formal test for the presence of heterogeneity, though caution that this test is likely underpowered due to the low number of relevant studies identified. Egger's regression was used to formally test the potential for small study effects and publication bias, and we illustrate these using contour enhanced funnel plots.

We present the results of meta-analyses comparing differences in CAH and control samples via Forest Plots in **Figure 2**, and provide contour enhanced funnel plots in

Figure 3. We present summary statistics for all meta-analyses in **Table 3**. First we present analyses which combine male and female samples into a single analysis, then perform separate meta-analyses on male and female subsamples. For the combined samples, R2D:4D, L2D:4D, and M2D:4D are smaller for CAH samples than controls, though no such difference is apparent for $D_{[R-L]}$. This final comparison was the only one for which a significant Egger's test result was observed, suggesting that small study effects are plausible. (z = -2.119, p = 0.034).

Our analyses took into account the risk of bias from taking multiple estimates from the same published (or unpublished) article, as many of our included studies examined both male and female samples. We achieved this by conducting two- and three-level meta-analyses as recommended by Konstantopoulos (2011). That is, we ran a two-level model that controls only for the random effect of the sample from which the estimate is drawn, as is normally the case within random effects meta-analysis. This is the model we present in **Table 3**. We then also ran a three-level model in which the random effect of sample was nested within the article from which the sample was drawn. This adjusts the estimate of effect size to account for any correlation between samples that derives from the samples having a shared source. These results are presented in **Table 4** (please note that we do not recreate the individual sample estimates in **Table 4** as they are identical to those already presented in **Table 3)**. It can be seen from **Table 4** that in most cases nesting by paper as well as sample in the three-level model had no impact on estimates, suggesting there was rarely any variance shared between estimates attributable to male and female samples being drawn from the same article. An exception is for the analysis that compared M2D:4D between CAH and control samples. Here the comparison was statistically significant in the two- but not the three-level model. We also estimated the extent to which sex may moderate any association between CAH and 2D:4D by computing a (two-level) meta-regression model in which sex is included as a categorical moderator (see **Table 4**). This showed that there was no moderation effect of sex for any comparison.

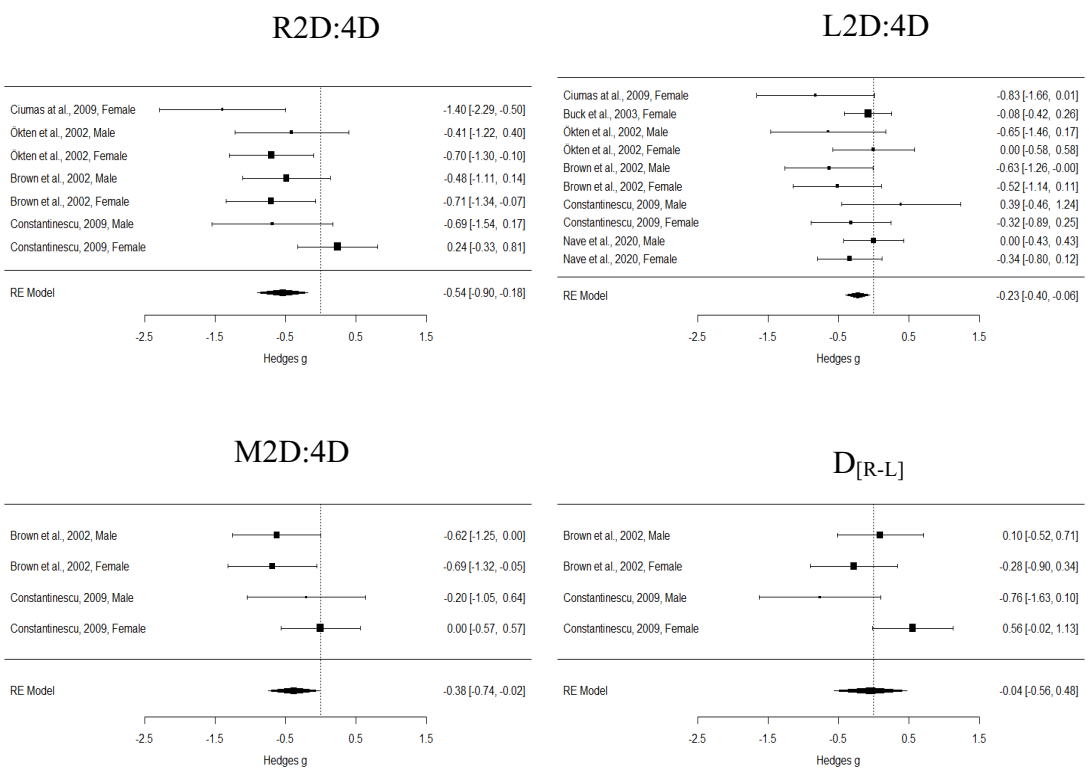

Figure 2. Forest plot summary for each meta-analysis comparing 2D:4D for individuals with CAH to controls for each hand combination.

Note. Negative g values indicate CAH < controls.

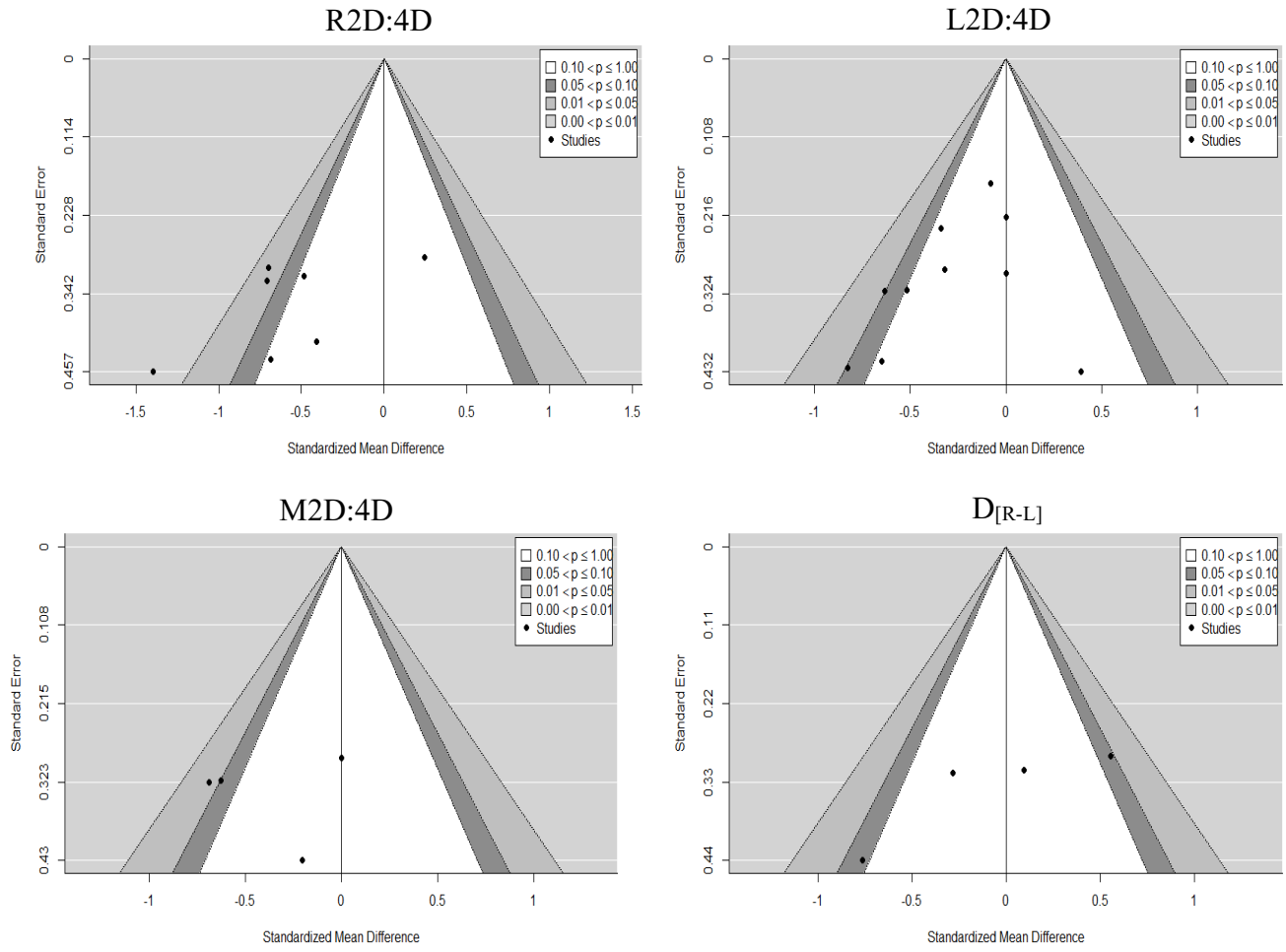

Figure 3. Contour enhanced funnel plots for each meta-analysis comparing mean 2D:4D for individuals with CAH to controls for each hand combination.

Table 3. Summary of meta analyses of the difference between 2D:4D for participants with CAH and controls.

			CAH			Control			Effect Size		Meta-Analyses (95%CI)					Heterogeneity				
Digit Ratio	**Sex**	**Study**	*n*	*M*	*SD*	*n*	*M*	*SD*	*g*	*SE*	*g*	LCI	UCI	*SE*	*p*	Q	df	*p*	τ	I^2
R2D:4D	Male	Ökten et al.	9	0.980	0.030	18	0.990	0.020	-0.410	0.412	-0.513	-0.940	-0.085	0.218	0.019	0.226	2	0.893	0	0
		Brown et al.	16	0.937	0.045	28	0.957	0.038	-0.483	0.318										
		Constantinescu et al.	24	0.941	0.042	7	0.970	0.038	-0.686	0.439										
	Female	Ciumas et al	11	0.956	0.024	13	0.985	0.016	-1.398	0.460	-0.591	-1.233	0.052	0.328	0.072	11.237	3	0.011	0.559	73.74
		Ökten et al.	17	0.990	0.020	34	1.000	0.010	-0.700	0.305										
		Brown et al.	13	0.957	0.038	43	0.981	0.032	-0.708	0.324										
		Constantinescu et al.	40	0.960	0.046	17	0.950	0.023	0.243	0.290										
	Combined										-0.540	-0.901	-0.178	0.184	0.003	11.470	6	0.075	0.333	47.35
L2D:4D	Male	Ökten et al.	9	0.980	0.030	18	1.00	0.030	-0.646	0.418	-0.218	-0.660	0.224	0.226	0.334	5.648	3	0.130	0.303	45.66
		Brown et al.	16	0.931	0.035	28	0.955	0.039	-0.627	0.321										
		Constantinescu et al.	24	0.959	0.035	7	0.943	0.055	0.390	0.433										
		Nave et al.	45	0.913	0.023	39	0.913	0.020	0.000	0.219										
	Female	Ciumas et al	11	0.979	0.027	13	1.005	0.033	-0.825	0.428	-0.245	-0.452	-0.039	0.105	0.020	4.406	5	0.493	0	0
		Buck et al.	66	0.925	0.021	69	0.927	0.029	-0.078	0.172										
		Ökten et al.	17	0.990	0.040	34	0.990	0.020	0	0.297										
		Brown et al.	13	0.952	0.025	43	0.968	0.032	-0.516	0.320										
		Constantinescu et al.	40	0.944	0.036	17	0.955	0.028	-0.320	0.291										
		Nave et al.	45	0.917	0.023	31	0.925	0.024	-0.338	0.235										
	Combined										-0.227	-0.397	-0.056	0.087	0.009	10.153	9	0.338	0	0
M2D:4D	Male	Brown et al.	16	0.934	0.037	28	0.957	0.035	-0.632	0.321	-0.474	-0.978	0.030	0.257	0.065	0.613	1	0.434	0	0
		Constantinescu et al.	24	0.950	0.028	7	0.956	0.031	-0.204	0.430										
	Female	Brown et al.	13	0.954	0.026	44	0.975	0.030	-0.710	0.323	-0.329	-1.001	0.343	0.343	0.338	2.511	1	0.113	0.376	60.17
		Constantinescu et al.	40	0.952	0.035	17	0.952	0.022	0	0.290										
	Combined										-0.379	-0.742	-0.015	0.185	0.041	3.369	3	0.338	0.164	19.38
$\mathbf{D_{[R-L]}}$	Male	Brown et al.	16	0.006	0.033	28	0.003	0.028	0.099	0.314	-0.278	-1.112	0.556	0.425	0.513	2.520	1	0.112	0.471	60.32
		Constantinescu et al.	24	-0.018	0.053	7	0.027	0.072	-0.763	0.441										
	Female	Brown et al.	13	0.005	0.037	44	0.013	0.024	-0.288	0.317	0.146	-0.674	0.966	0.418	0.728	3.748	1	0.053	0.507	73.32
		Constantinescu et al.	40	0.016	0.041	17	-0.005	0.026	0.556	0.294										
	Combined										-0.043	-0.563	0.477	0.265	0.872	7.430	3	0.059	0.409	60.05

Table 4. Summary of meta analyses of the difference between 2D:4D for participants with CAH and controls for males and females combined.

	Meta-analyses (95%CI)						**Sex as a moderator**			
Comparison	**Model**	***g***	**LCI**	**UCI**	**SE**	***p***	**Beta**	**LCI**	**UCI**	**SE**	***p***
R2D:4D	Two-level	-0.540	-0.901	-0.178	0.184	0.003	0.046	-0.760	0.853	0.411	0.910
	Three-level	-0.540	-0.901	0.178	0.184	0.003					
L2D:4D	Two-level	-0.227	-0.397	-0.056	0.087	0.009	0.063	-0.332	0.458	0.201	0.754
	Three-level	-0.227	-0.397	-0.056	0.087	0.009					
M2D:4D	Two-level	-0.379	-0.742	-0.015	0.185	0.041	-0.133	-0.972	0.706	0.428	0.756
	Three-level	-0.364	-0.944	0.215	0.296	0.218					
$D_{[R-L]}$	Two-level	-0.043	-0.563	0.477	0.265	0.872	-0.427	-1.601	0.746	0.599	0.475
	Three-level	-0.043	-0.563	0.477	0.265	0.872					

Note. Positive values for a moderation estimate imply a larger estimate of g for male samples over female samples

We present the results of meta-analyses comparing differences in individual hands (R2D:4D and L2D:4D) via Forest Plots in **Figure 4**, and for aggregated measures (M2D:4D and $D_{[R-L]}$) in **Figure 5**. When breaking the sample down into separate sexes, only two comparisons remained statistically significant: R2D:4D for males, and L2D:4D for females. Egger's test of small study effects did not identify statistically significant effects for any analysis. However, the small number of studies provided low power for this test. For female samples, the R2D:4D analysis was not far from statistical significance (z = -1.877, p = 0.061) while there were no indications of small study effects for L2D:4D (z = -1.552, p = 0.121). Male samples did not show any sign of small study effects for R2D:4D or L2D:4D (z = -0.206, p = 0.837, and z = -0.067, p = 0.947, respectively). We also present contour enhanced funnel plots in **Figure 6**, some of which do suggest that small study effects were plausible for some comparisons. In particular, the plots suggest that some small studies that show effects that are null or in the opposite direction to predicted may have been missing, which, if true, would lead to an artificial inflation of the average weighted effect sizes observed

Notably, the effect sizes reported in these meta-analyses were considerably smaller than those reported by Hönekopp and Watson (2010) (for comparisons, see **Table 7**). To provide an indication of the number of participants that would be required to observe statistically significant effects, we conducted power calculations using G*Power 3.1 (Faul et al., 2007). Based on the effect sizes observed in the present study (note that we substituted *d* for *g* in this case, the difference being negligible), the use of independent samples *t*-tests with equal numbers of cases and controls, α set at $p < 0.05$ (two-tailed), and 80% power, the required sample sizes for R2D:4D (males: n=122; females: n=92) were considerably smaller than those for L2D:4D (males: n=664; females, n=526). Furthermore, the estimated power for the average comparisons reported in this literature was exceptionally low (R2D:4D: females, β = 0.509; males, β = 0.267; L2D:4D: females, β = 0.161, males β = 0.101), indicating considerable propensity for Type 1 errors being reported by small studies.

We additionally conducted a meta-analysis that included only studies that used radiographs to measure 2D:4D. It was only possible to perform meta-analysis for L2D:4D, as only Ökten et al. (2002) used radiographs on the right hand. Ökten et al. (2002) and Nave et al. (2020) used radiographs on the left hand of males. The difference

between male participants with and without CAH remained non-significant (g[95%CI] = -0.225 [-0.829, 0.378], SE = 0.308, p = 0.465, Q(1) = 1.880, p = 0.170, τ = 0.313, I^2 = 46.81). For females, Ökten et al. (2002), Buck et al. (2003) and Nave et al. (2020) used radiographs on the left hand. Here the difference for L2D:4D was no longer statistically significant between control and CAH participants (g[95%CI] = -0.139 [-0.385, 0.108], SE = 0.126, p = 0.270, Q(2) = 1.062, p = 0.588, τ = 0, I^2 = 0).

The data provided by Nave et al. (2020) presented a complication, in that they incorporated multiple cases where participants had 2D:4D measured on more than one occasion. We therefore performed three meta-analyses. The data we presented above take the mean measure for each individual participant across all measures taken for that individual. However, we also performed meta-analyses that took the value from only the first and then only the last measure from each participant. Taking the first value provided results that contrast with the analysis using mean measures as both male and female comparisons identified significant differences in 2D:4D between CAH and non-CAH samples (L2D:4D males: g[95%CI] = -0.363 [-0.668, -0.059], SE = 0.155, p = 0.020, Q(3) = 4.199, p = 0.241, τ < 0.001, I^2 = 0; L2D:4D females: g[95%CI] = -0.302 [-0.535, -0.068], SE = 0.119, p = 0.011, Q(5) = 5.62, p = 0.345, τ = 0.120, I^2 = 16.55). However, taking the final measure provided results that align with using the average measure in that the male L2D:4D difference was not statistically significant (g[95%CI] = -0.162 [-0.682, -0.348], SE = 0.260, p = 0.533, Q(3) = 7.361, p = 0.061, τ = 0.394, I^2 = 58.67), while the female L2D:4D difference remained statistically significant (g[95%CI] = -0.228 [-0.434, -0.022], SE = 0.105, p = 0.031, Q(5) = 4.225, p = 0.518, τ = 0.001, I^2 = 0). These findings may point towards the importance of considering differences in bone maturation between males and females with CAH and male and female controls.

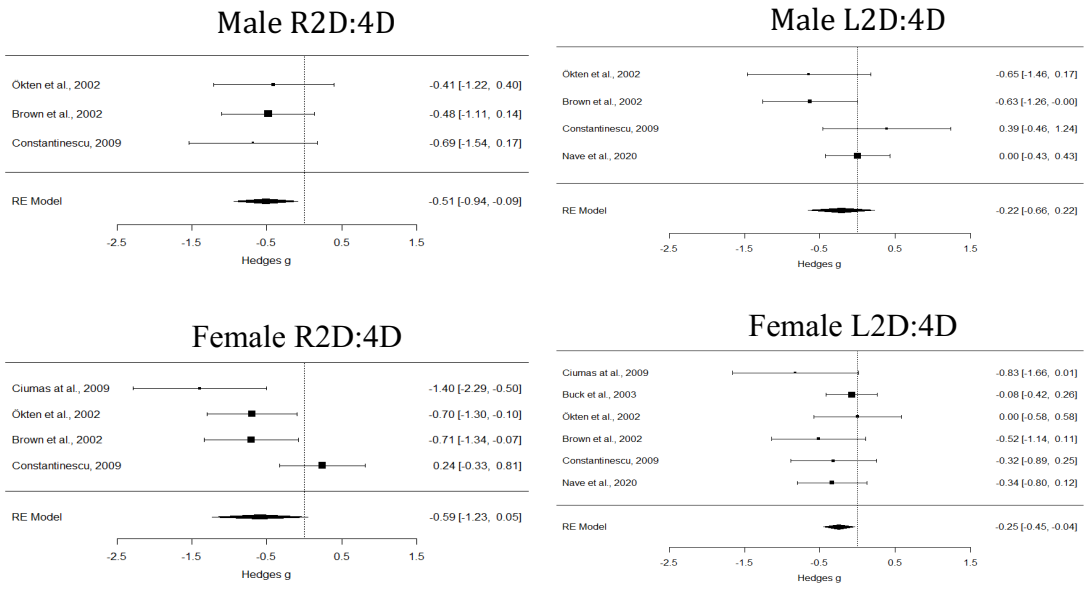

Figure 4. Forest plot summary for each meta-analysis comparing 2D:4D for individuals with CAH to controls for each sex and hand combination.

Note. Negative g values indicate CAH < controls.

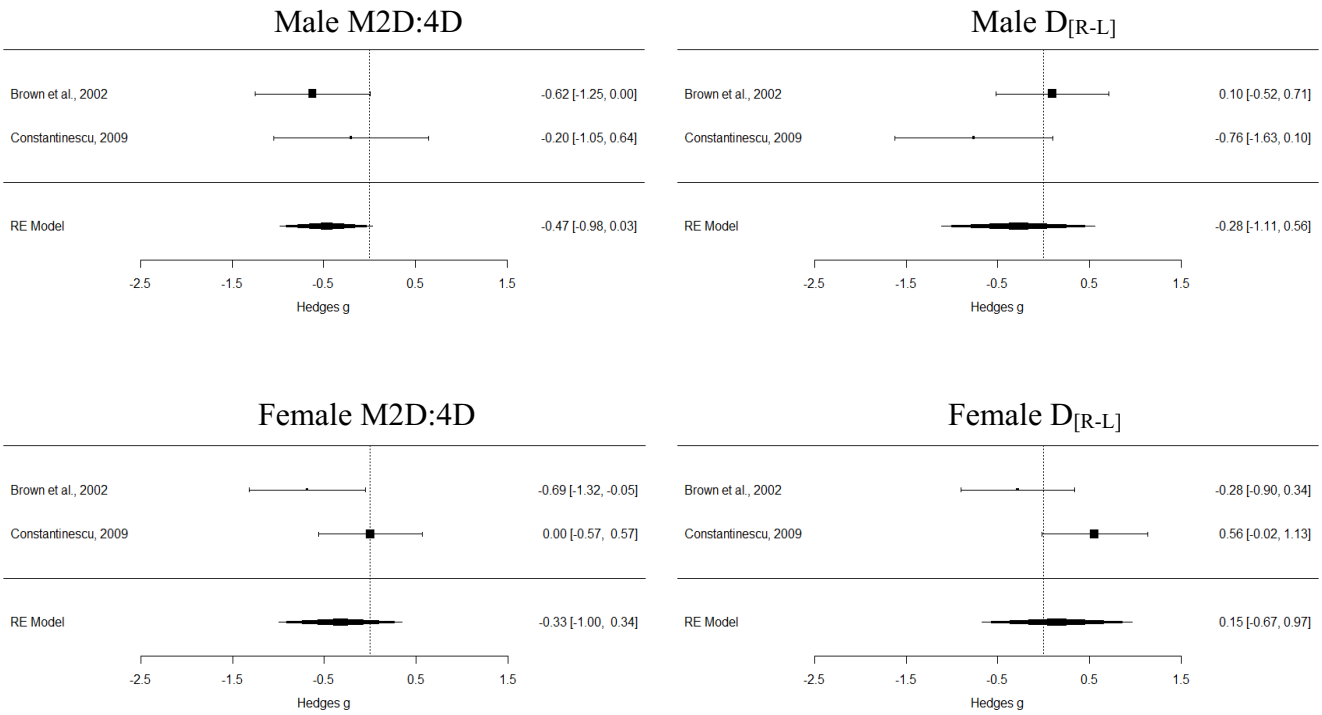

Figure 5. Forest plot summary for each meta-analysis comparing aggregated 2D:4D measures for individuals with CAH to controls for each sex.

Note. Negative g values indicate CAH < controls.

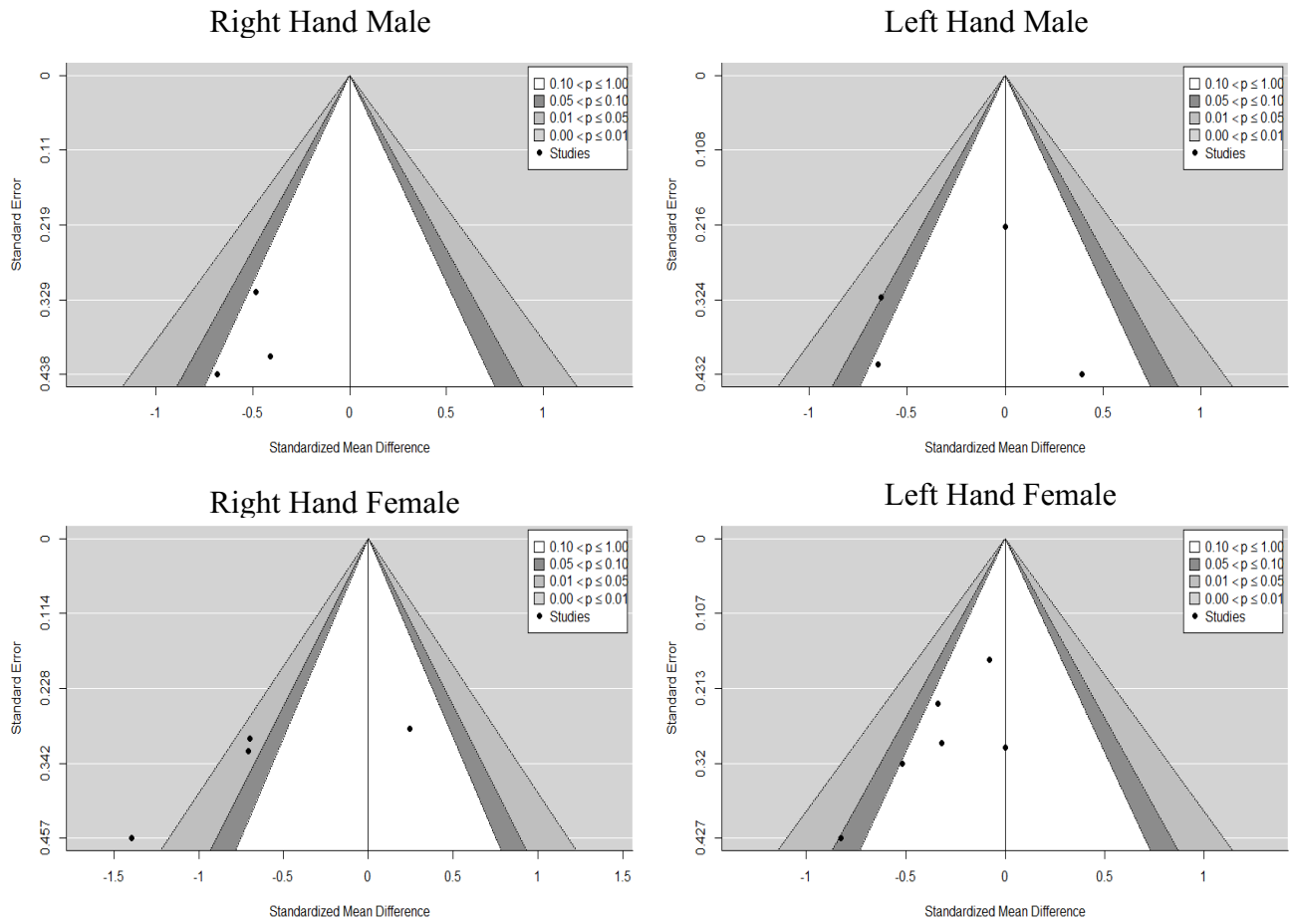

Figure 6. Contour enhanced funnel plots for each meta-analysis comparing mean 2D:4D for individuals with CAH to controls for each sex and hand combination.

Table 7. Comparison of meta-analytic effect sizes observed by Hönekopp and Watson (2010) and by the current study

	Hönekopp & Watson (2010)		**Current study**		**Percentage change**
	p	***d***	***p***	***g***	
Male R2D:4D	0.061	-0.94	0.019	-0.513	-54.57%
Male L2D:4D	0.013	-0.63	0.334	-0.218	-34.60%
Female R2D:4D	< 0.001	-0.91	0.072	-0.591	-64.95%
Female L2D:4D	0.007	-0.75	0.02	-0.245	-32.67%
Mean effect size		-0.81		-0.392	-46.70%

Note. We compare here the original effect size estimates reported by Hönekopp and Watson (2010) *(Cohen's d) with those observed for the current study (Hedge's g); Cohen's d and Hedge's g both indicate the standardised mean difference, and the difference between these two metrics is negligible. Signs for effect size estimates reported by Hönekopp and Watson (2010) have been reversed for ease of comparison with our own.*

Sexual dimorphism in 2D:4D in CAH cases and controls

The magnitude of prenatal androgen elevation associated with CAH might differ by sex due to downregulation of testicular androgen production being possible only in males. Although not specified in our pre-registration plan, we therefore considered it useful to examine whether the pattern of sexual dimorphism observed in typically developing populations (i.e. M<F; R2D:4D, d = -0.457; L2D:4D, d = -0.376; Hönekopp & Watson, 2010) extends to people with CAH, or whether the direction and/or magnitude of such effects differ as a product of diagnostic status.

No individual studies showed statistically significant effects in the expected (i.e. M<F) direction for R2D:4D or L2D:4D (**Table 8**). A longitudinal study (Kim et al., 2017) also reported no significant sex differences for L2D:4D either at baseline or at final follow-up, or in participants who were pre-pubertal or pubertal. The only study for which significant differences were observed was Rivas et al. (2014), which found the opposite pattern of results than expected (R2D:4D and L2D:4D were both higher in males with CAH than in females with CAH). However, the veracity of these results is uncertain. Firstly, it is unclear what measure(s) of dispersal around the mean was/were reported, as some values were approximately tenfold smaller (e.g. 0.0026 for R2D:4D in female controls) than others (e.g. 0.0267 for L2D:4D in males with CAH). Secondly, the effect size for L2D:4D (d = 2.265) was more than twice that of any comparable analysis, and so appeared implausibly large (refer to **Table 8** for direct comparison with other studies).

No studies reported whether M2D:4D or $D_{[R-L]}$ differed between males with CAH and females with CAH. Reanalysis of the datasets presented by Brown et al. (2002) and Constantinescu (2009) yielded no significant sex differences for M2D:4D. Although we also found no sex difference for $D_{[R-L]}$ in the data of Brown et al. (2002), this was not the case for Constantinescu (2009): $D_{[R-L]}$ was significantly lower in males with CAH than in females with CAH. This appeared to be driven by a significant difference in the younger group (< 8 years old), as no such effect was detected in the older group (> 8 years old), and could potentially therefore reflect differences in bone maturation rates.

Table 8. Comparisons of 2D:4D between males and females with CAH.

Study	Digit ratio	Measurement	Males			Females			Difference			
			n	*M*	*SD*	*n*	*M*	*SD*	*t*	*df*	*p*	*d*
Brown et al. (2002)	R2D:4D[a]	Photocopies	16	0.937	0.045	13	0.957	0.038	-1.235	27	0.227	-0.476
	L2D:4D[a]	Photocopies	16	0.931	0.034	13	0.952	0.025	-1.816	27	0.080	-0.692
	M2D:4D[a]	Photocopies	16	0.934	0.037	13	0.954	0.026	-1.664	27	0.108	-0.614
	$D_{[R-L]}$[a]	Photocopies	16	0.006	0.033	13	0.005	0.037	0.099	27	0.922	0.029
Ökten et al. (2002)	R2D:4D[b]	Photocopies	9	0.92	0.04	17	0.96	0.06	-1.792	24	0.086	-0.739
	L2D:4D[b]	Photocopies	9	0.91	0.06	17	0.92	0.05	-0.453	24	0.655	-0.187
	R2D:4D[c]	X-rays	9	0.98	0.03	17	0.99	0.02	-1.019	24	0.318	-0.420
	L2D:4D[c]	X-rays	9	0.98	0.03	17	0.99	0.04	-0.656	24	0.518	-0.271
Constantinescu (2009)	R2D:4D (all subjects)[d,e]	Direct/indirect	24	0.941	0.042	40	0.960	0.046	-1.670	62	0.100	-0.426
	R2D:4D (< 8 years)[d]	Direct/indirect	15	0.933	0.042	23	0.959	0.042	-1.879	36	0.068	-0.619
	R2D:4D (> 8 years)[d,†]	Direct/indirect	9	0.953	0.043	17	0.960	0.051	-0.363	24	0.720	-0.144
	L2D:4D (all subjects)[d,f]	Direct/indirect	24	0.959	0.035	40	0.944	0.036	1.563	62	0.123	0.421
	L2D:4D (< 8 years)[d]	Direct/indirect	15	0.958	0.040	23	0.939	0.029	1.653	36	0.107	0.564
	L2D:4D (> 8 years)[d,†]	Direct/indirect	9	0.960	0.026	17	0.952	0.043	0.646	23.266[§]	0.525	0.210
	M2D:4D (all subjects)[d]	Direct/indirect	24	0.950	0.028	40	0.952	0.035	-0.288	62	0.774	-0.061
	M2D:4D (< 8 years)[d]	Direct/indirect	15	0.945	0.028	23	0.949	0.031	-0.392	36	0.697	-0.134
	M2D:4D (> 8 years)[d]	Direct/indirect	9	0.957	0.030	17	0.956	0.042	0.048	24	0.962	0.026
	$D_{[R-L]}$ (all subjects)[d]	Direct/indirect	24	-0.018	0.053	40	0.016	0.041	-2.811	62	**0.007**	**-0.742**
	$D_{[R-L]}$ (< 8 years)[d]	Direct/indirect	15	-0.025	0.061	23	0.020	0.039	-2.781	36	**0.009**	**-0.923**
	$D_{[R-L]}$ (> 8 years)[d]	Direct/indirect	9	-0.007	0.038	17	0.009	0.045	-0.910	24	0.372	-0.374
Rivas et al. (2014)	R2D:4D[g]	Direct	9	0.960	0.0220	31	0.950	0.0077	2.166	38	**0.037**	**0.820**
	L2D:4D[g]	Direct	9	0.983	0.0267	31	0.947	0.0114	5.981	38	**< 0.001**	**2.265**
Kim et al. (2017)	L2D:4D (baseline)	X-rays	43	0.902	0.035	40	0.911	0.026	-1.260	81	0.211	-0.290
	L2D:4D (follow-up)	X-rays	43	0.918	0.026	40	0.920	0.025	-0.396	81	0.693	-0.078
	L2D:4D (pre-pubertal)	X-rays	16	0.89	0.04	19	0.91	0.02	1.917	33	0.064	-0.650
	L2D:4D (pubertal)	X-rays	27	0.92	0.03	15	0.92	0.03	0.0	40	1.0	0.000
Nave et al. (2020)	L2D:4D (average)	X-rays	45	0.913	0.023	45	0.917	0.023	-0.949	88	0.345	-0.200
	L2D:4D (first scan)	X-rays	45	0.902	0.033	45	0.911	0.027	-1.380	88	0.171	-0.291
	L2D:4D (last scan)	X-rays	45	0.918	0.024	45	0.920	0.023	-0.550	88	0.584	-0.116

Note. Negative d values indicate effects in the predicted direction (i.e. M<F); effects in bold are statistically significant (p <0.05).

[a] *= calculated from original data of Brown et al. (2002)*

[b] *= calculated from data presented by Ökten et al. (2002, Table 1, p. 50)*

[c] = *calculated from data presented by Ökten et al. (2002, Table 3, p. 51)*
[d] = *calculated from the original data of Constantinescu (2009)*
[e] = *we report here the values calculated from the original data due to there being differences in rounding with the values reported by Constantinescu (2009); the values reported by Constantinescu are as follows: males with CAH (n=24, M = 0.9407, SD = 0.04), females with CAH (n=40, M = 0.9598, SD = 0.04), t(62) = -1.67, p = 0.100, d = -0.43*
[f] = *we report here the values calculated from the original data due to there being differences in rounding with the values reported by Constantinescu (2009); the values reported by Constantinescu are as follows: males with CAH (n=24, M = 0.9586, SD = 0.03), females with CAH (n=40, M = 0.9444, SD = 0.03), t(62) = 1.56, p = 0.123, d = 0.40.*
[g] = *calculated from data presented by Rivas et al. (2014, p. 560)*
[†] = *in Constantinescu (2009), it is listed that the > 8 years old group consisted of 19 females and 7 males, whereas in the dataset we analysed, there were 17 females and 9 males.*
[§] = *equal variances not assume*

Meta-analysis of sexual dimorphism in 2D:4D in CAH cases and controls

As with our comparison of 2D:4D variables between CAH and control samples, we used random effects meta-analyses to estimate the difference between male and female 2D:4D. More specifically, we combined CAH and control samples and tested for any moderation effect of population type (CAH or control) on any sex difference. We summarise the values in forest plots (**Figure 7**) and provide greater detail in **Table 8**. For robustness, we again produce two- and three-level meta-analyses that either do (three-level) or do not (two-level) nest estimates of effect within the article from which estimates are drawn. Details of these estimates and analysis of any moderation of population type are presented in **Table 9**.

Only the sex differences for R2D:4D in CAH patients and L2D:4D in control participants were statistically significant (both M<F). Egger's test of small study effects did not identify statistically significant effects except for the comparison of $D_{[R-L]}$, for which CAH and control samples were combined ($z = 2.404$, $p = 0.016$). CAH status did not moderate the sex difference in 2D:4D for any comparison.

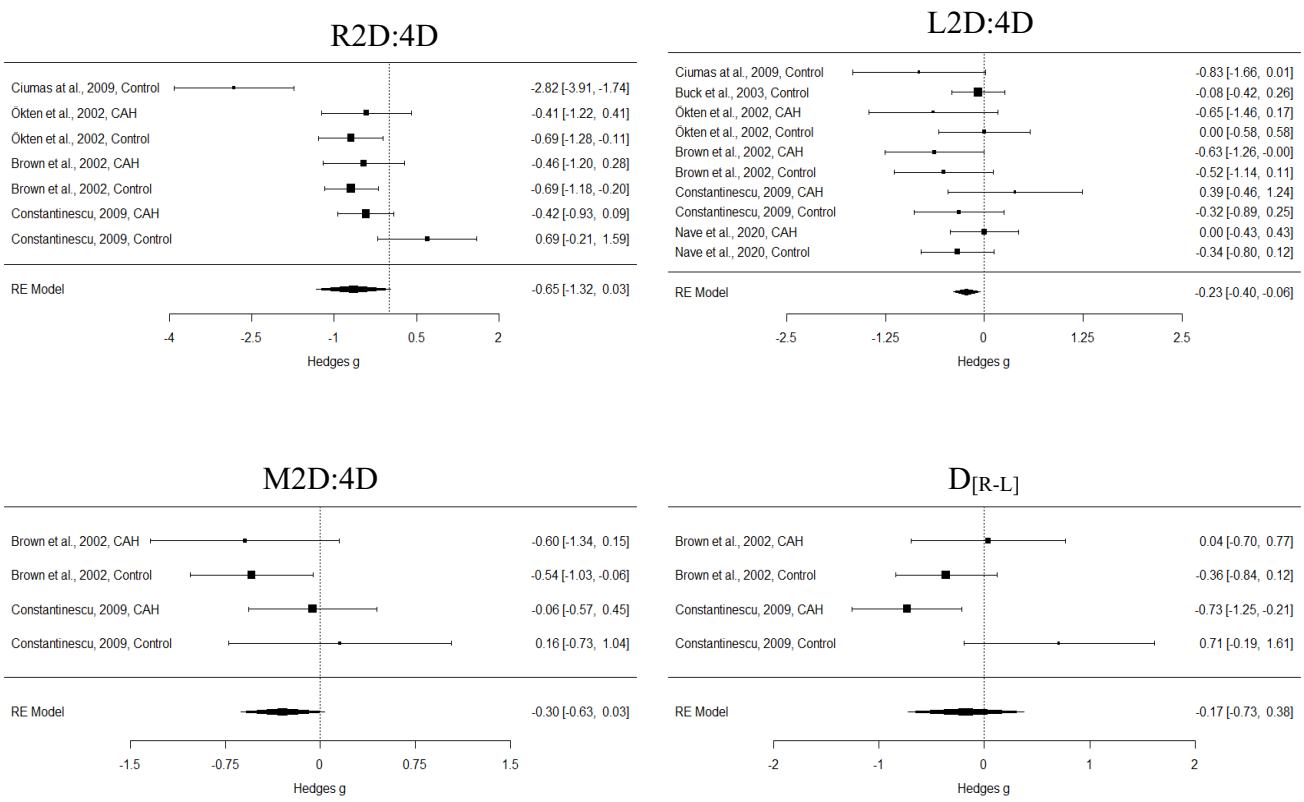

Figure 7. Forest plot summary for each meta-analysis comparing 2D:4D measures between males and females.

Note. Negative g values indicate M<F.

Table 8. Summary of meta analyses of the difference between 2D:4D for male and female participants.

			Male			Female			Effect Size		Meta-Analyses (95%CI)					Heterogeneity				
Digit Ratio	**Population**	**Study**	*n*	*M*	*SD*	*n*	*M*	*SD*	*g*	*SE*	*g*	LCI	UCI	*SE*	*p*	Q	df	*p*	τ	I^2
R2D:4D	CAH	Ökten et al.	9	0.98	0.03	17	0.99	0.02	-0.407	0.416	-0.429	-0.803	-0.055	0.191	0.025	0.012	2	0.994	0	0
		Brown et al.	16	0.937	0.045	13	0.957	0.038	-0.463	0.379										
		Constantinescu et al.	24	0.941	0.042	40	0.960	0.046	-0.421	0.261										
	Control	Ciumas et al.	13	0.945	0.011	13	0.985	0.016	-2.821	0.563	-0.849	-2.196	0.499	0.688	0.217	23.819	3	<0.001	1.315	93.10
		Ökten et al.	18	0.990	0.020	34	1.000	0.010	-0.693	0.300										
		Brown et al.	28	0.957	0.038	43	0.981	0.032	-0.689	0.250										
		Constantinescu et al.	7	0.970	0.038	17	0.950	0.023	0.692	0.461										
	Combined										-0.649	-1.324	0.027	0.345	0.060	24.983	6	<.001	0.830	85.48
L2D:4D	CAH	Ökten et al.	9	0.98	0.03	17	0.99	0.04	-0.262	0.414	-0.118	-0.557	0.320	0.224	0.598	6.387	3	0.094	0.325	54.26
		Brown et al.	16	0.931	0.035	13	0.952	0.025	-0.659	0.384										
		Constantinescu et al.	24	0.959	0.035	40	0.944	0.036	0.416	0.261										
		Nave et al.	45	0.913	0.023	45	0.917	0.023	-0.172	0.211										
	Control	Ciumas et al.	13	0.973	0.047	13	1.005	0.033	-0.763	0.407	-0.303	-0.585	-0.021	0.144	0.035	8.242	5	0.143	0.216	39.35
		Buck et al.	77	0.918	0.026	69	0.927	0.029	-0.326	0.167										
		Ökten et al.	18	1.000	0.030	34	0.990	0.020	0.413	0.294										
		Brown et al.	28	0.955	0.039	43	0.968	0.032	-0.369	0.245										
		Constantinescu et al.	7	0.943	0.055	17	0.955	0.028	-0.310	0.452										
		Nave et al.	39	0.913	0.020	31	0.925	0.024	-0.543	0.245										
	Combined										-0.228	-0.469	0.013	0.123	0.063	16.348	9	0.060	0.258	47.35
M2D:4D	CAH	Brown et al.	16	0.934	0.037	13	0.954	0.026	-0.597	0.382	-0.255	-0.760	0.250	0.258	0.322	1.354	1	0.245	0.194	26.15
		Constantinescu et al.	24	0.950	0.028	40	0.952	0.035	-0.061	0.258										
	Control	Brown et al.	28	0.957	0.035	43	0.975	0.030	-0.556	0.247	-0.295	-0.950	0.361	0.335	0.379	1.856	1	0.173	0.336	46.12
		Constantinescu et al.	7	0.956	0.031	17	0.952	0.022	0.156	0.450										
	Combined										-0.300	-0.631	0.032	0.169	0.076	3.460	3	0.326	0.136	15.56
$D_{[R-L]}$	CAH	Brown et al.	16	0.006	0.033	13	0.005	0.037	0.028	0.373	-0.393	-1.142	0.356	0.382	0.304	2.814	1	0.093	0.437	64.46
		Constantinescu et al.	24	-0.018	0.053	40	0.016	0.041	-0.733	0.266										
	Control	Brown et al.	28	0.003	0.028	43	0.013	0.024	-0.386	0.245	0.102	-0.938	1.141	0.530	0.848	4.212	1	0.040	0.661	76.26
		Constantinescu et al.	7	0.027	0.072	17	-0.005	0.026	0.708	0.461										
	Combined										-0.175	-0.726	0.376	0.281	0.534	8.310	3	0.040	0.454	67.13

Table 9. Summary of meta analyses of the difference between 2D:4D for males and females where controls and participants with CAH are combined.

	Meta-analyses (95%CI)						**CAH vs control as a moderator**				
Comparison	**Model**	***g***	**LCI**	**UCI**	**SE**	***p***	**Beta**	**LCI**	**UCI**	**SE**	***p***
R2D:4D	Two-level	-0.649	-1.324	0.027	0.345	0.060	-0.397	-1.896	1.102	0.765	0.604
	Three-level	-0.924	-2.003	0.155	0.551	0.093					
L2D:4D	Two-level	-0.228	-0.469	0.013	0.123	0.063	-0.195	-0.696	0.306	0.255	0.445
	Three-level	-0.227	-0.476	0.023	0.127	0.075					
M2D:4D	Two-level	-0.300	-0.631	0.032	0.169	0.076	-0.047	-0.861	0.766	0.415	0.909
	Three-level	-0.289	-0.830	0.251	0.276	0.294					
$D_{[R-L]}$	Two-level	-0.175	-0.726	0.376	0.281	0.534	0.461	-0.806	1.727	0.646	0.476
	Three-level	-0.175	-0.726	0.376	0.281	0.534					

Note. Positive values for a moderation estimate imply a larger estimate of g for Control samples over CAH samples

Other reported correlations within CAH samples

Table 10 presents additional findings from studies of 2D:4D in people with CAH. Associations between 2D:4D and age were reported in two samples (Buck et al., 2003; Kim et al., 2017; Nave et al., 2020). Firstly, Buck et al. (2003) reported that 2D:4D correlated positively with age in their cohort, though noted that the effect was not statistically significant. Additionally, the only longitudinal study in the area (Kim et al., 2017; Nave et al., 2020) reported that L2D:4D increased between baseline and final follow-up, and that the effect size (d = 0.46) was small-medium (Cohen, 1988; 0.20 = small, 0.50 = medium, 0.80 = large); further, 2D:4D was lower in pre-pubertal than pubertal participants. Although this could posit a role for pubertal hormones (see Králík, Ingrová, Kozieł, Hupková, & Klíma, 2017), such influence appears unlikely to explain the entirety of the effect because age-related increases in 2D:4D in typically developing samples commence before the onset of puberty (McIntyre, Ellison, Lieberman, Demerath, & Towne, 2005; Trivers, Manning, & Jacobson, 2006).

Although Constantinescu (2009) reported subgroup analyses based on age, they did not report whether 2D:4D correlated with age. We therefore reanalysed these data by using Pearson's correlations to examine this possibility. In males, age did not correlate significantly with R2D:4D (r[22] = 0.178, p = 0.405), L2D:4D (r[22] = -0.275, p = 0.194), M2D:4D (r[22] = -0.036, p = 0.868), or $D_{[R-L]}$ (r[22] = 0.323, p = 0.124). In females, age correlated negatively with $D_{[R-L]}$ (r[38] = -0.320, p = 0.044, though did not correlate significantly with R2D:4D (r[38] = -0.174, p = 0.283), L2D:4D (r[38] = 0.149, p = 0.359), or M2D:4D (r[38] = -0.037, p = 0.819).

Unlike other studies, Kocaman et al. (2016, 2017) presented analyses in which the male and female samples were combined. These authors reported that a Pearson's correlation between 2D:4D and a measure of autistic traits (a Turkish language translation of the Autism Behavior Checklist [ABC]) was significant (direction of effect is unclear). It is also ambiguous whether this effect was observed in CAH participants or controls, and whether it related to R2D:4D or L2D:4D (the effect examined in the analysis in which patients with CAH and controls were combined was not significant). Although the finding is difficult to interpret, it may be relevant in regard to previous studies that have reported correlations between 2D:4D, autism, and autistic traits (Hönekopp, 2012; Manning, Baron-Cohen, Wheelwright, & Sanders, 2001; Myers, van't Westeinde,

Kuja-Halkola, Tammimies, & Bölte, 2018; Schieve et al., 2018; Teatero & Netley, 2013; Voracek, 2008). Kocaman et al. (2017) also reported that 2D:4D did not differ for children who had a difficult birth, a premature birth, or whose mother smoked or had a physical or mental health condition. There was, however, a significant effect of maternal stress within the CAH group, though the direction of this effect is unclear.

Although endocrine status is frequently monitored in patients with CAH, the only study so far to report on circulating hormone levels and 2D:4D in a CAH sample is Oświęcimska et al. (2012). These authors reported that M2D:4D was positively correlated with serum testosterone and dehydroepiandrostenedione sulphate (DHEAS), though there was no association with androstenedione, and they did not report whether there was a correlation with 17-hydroxyprogesterone (17-OHP). (Also note that although both the Abstract and Results sections of this paper report that M2D:4D correlated positively with testosterone and s-DHEA, Figures 1 and 2 reportedly present significant positive correlations between M2D:4D and testosterone and androstenedione, respectively.) These findings are difficult to interpret, as they relate to a small sample, and no other published study has examined such effects in a CAH population. Although some individual studies have reported significant correlations between 2D:4D and circulating testosterone, meta-analyses suggest these variables are not related (Hönekopp, Bartholdt, Beier, & Liebert, 2007; Zhang et al., 2019). It is therefore suggested that unless these effects are replicated, they should be interpreted with caution.

Table 10. Additional findings from studies of 2D:4D in CAH populations.

Authors	Main findings
Buck et al. (2003)	L2D:4D marginally increased with age ($p = 0.08$) (analysis included females with CAH, and male and female controls)
Constantinescu (2009)	No significant correlation between R2D:4D of males with CAH and R2D:4D of their mothers No significant correlation between L2D:4D of males with CAH and L2D:4D of their mothers R2D:4D in females with CAH correlated positively with R2D:4D in their mothers R2D:4D in females with CAH correlated positively with L2D:4D in their mothers No significant correlation between L2D:4D of females with CAH and L2D:4D of their mothers L2D:4D in females with CAH correlated positively with R2D:4D in their mothers †No significant correlation between age and R2D:4D in males †No significant correlation between age and L2D:4D in males †No significant correlation between age and M2D:4D in males †No significant correlation between age and $D_{[R-L]}$ in males †No significant correlation between age and R2D:4D in females †No significant correlation between age and L2D:4D in females †No significant correlation between age and M2D:4D in females †Significant negative correlation between age and $D_{[R-L]}$ in females
Oświęcimska et al. (2012)	Mean 2D:4D in females with CAH correlated positively with serum testosterone Mean 2D:4D in females with CAH correlated positively with serum s-DHEA Mean 2D:4D in females with CAH did not correlate with serum androstenedione
Kocaman et al. (2017)	2D:4D correlated with autistic traits (direction unclear; hand unclear) 2D:4D in children with CAH was related to maternal stress (direction unclear; hand unclear) 2D:4D did not differ in children who had a difficult birth (hand unclear) 2D:4D did not differ in children who had a premature birth (hand unclear) 2D:4D did not differ in children whose mothers smoked (hand unclear) 2D:4D did not differ in children whose mothers had a physical or mental health condition (hand unclear)
Nave et al. (2020)	L2D:4D lower in Hispanic than White participants L2D:4D correlated positively with bone age in males L2D:4D correlated positively with bone age in females L2D:4D correlated negatively with puberty stage L2D:4D correlated positively with height No interaction between sex and CAH

Note. Kocaman et al. (2016) is not included in this Table because it was an earlier publication of the same study presented by Kocaman et al. (2017); Kim et al,. (2017) is also not included because it was an earlier publication of data from Nave et al. (2020)*; Constantinescu et al. (2010) is not included because it is a less complete report of the study by Constantinescu (2009).*
† Effect not reported in the original article (Pearson's correlations were calculated from the original dataset of Constantinescu [2009]).

Discussion

The current study presents a systematic review and meta-analysis of the 2D:4D/CAH literature. We identified 12 articles relating to nine studies, eight of which reported comparisons of 2D:4D between CAH cases and controls. The main findings are that: (1) R2D:4D, L2D:4D, and M2D:4D are all lower in people with CAH compared to typically developing controls, with effect that are small (L2D:4D and M2D:4D) to medium (R2D:4D) in size, (2) when stratified by sex, only the effects for R2D:4D in

males and L2D:4D in females remain statistically significant (i.e. $p < 0.050$), (3) $D_{[R-L]}$ does not differ between CAH cases and controls, and (4) sexual dimorphism in 2D:4D in CAH patients appears to be similar to that observed in typically developing populations (i.e. M<F, small to medium effect sizes); furthermore, we note that: (5) relatively little research in this area has been published since the meta-analysis of Hönekopp and Watson (2010), (6) most studies have examined small samples and lack adequate statistical power, (7) research has been heterogeneous in terms of sample size, country of origin, age-range of participants, type of control group employed, and method used for measuring digit ratio, yet most studies have not controlled for potential confounds such as age and ethnicity, (8) no studies have specifically examined 2D:4D in CAH caused by enzyme deficiencies other than 21-hydroxylase, (9) no studies have specifically examined 2D:4D in non-classical (i.e. late-onset) CAH samples, (10) only one study has examined differences in 2D:4D between patients with salt-wasting and simple virilising forms of classical CAH, and (11) 2D:4D in CAH samples may increase during childhood in a similar manner to that previously reported in typically developing populations. In addition, we note that if Bonferroni adjustment were employed, the effects observed for R2D:4D in males ($g = -0.513$, $p = 0.019$) and L2D:4D in females ($g = -0.245$, $p = 0.020$) would only retain the required α level of $p < 0.013$ if one-tailed tests were used.

Our pattern of results was slightly different from that of the meta-analysis by Hönekopp and Watson (2010), where significant effects were observed for each sex and hand combination other than R2D:4D in males ($d = -0.94$, $p = 0.061$). The addition of new studies has also noticeably reduced the average effect size observed between the previous meta-analysis and the current study, a finding that appears to mirror that of two meta-analyses of CAH and spatial skills conducted over a very similar time-period (Collaer & Hines, 2020; cf. Puts et al., 2008). In Hönekopp and Watson (2010) the standardised mean difference (Cohen's *d*) ranged between -0.63 and -0.94, whereas we report (Hedge's *g*) between -0.218 and -0.591 (average reduction in effect size = 46.70%). Part of this reduction could be explained by our use of Hedge's *g* over Cohen's *d*, which produces less biased estimates when studies have small samples. However the difference between *d* and *g* is negligible, so it does appear that newer studies have produced smaller estimates of difference. Another potential explanation for the disparity in findings between our study and that of Hönekopp and Watson (2010)

is that the latter treated the infant and young toddler sample of Ökten et al. (2002) as independent from their larger sample. As these samples appear unlikely to have been independent (i.e. although not entirely clear within the article, the smaller sample appears to be comprised of participants from the larger sample), they should not have been included in the same meta-analysis. This approach is problematic, as it implies that the same data will be counted twice - artificially lowering the standard error of the estimate, which could potentially account for the significant *p* values.

Although we found some evidence that M2D:4D is lower in males and females with CAH relative to male and female controls when re-examining the original data of Brown et al. (2002), these effects were not replicated when reanalysing data from the larger cohort studied by Constantinescu (2009). Further, although the meta-analysis combining these estimates found a significant difference when males and females were considered together (g = -0.379, p = 0.041), when stratified by sex, the effect was only approaching significance for males (M2D:4D, g = -0.474, p = 0.065) and was not significant for females (g = -0.329, p = 0.338). No reliable differences between CAH cases and controls were observed for $D_{[R-L]}$ (combined: g = -0.043, p = 0.872; males: g = -0.278, p = 0.513; females: g = 0.146, p = 0.728), casting further doubt on the utility of this variable as an indicator of prenatal sex hormone exposure (Richards et al., 2019, 2020).

It appears that digit ratios are typically lower (i.e. more 'male typical') in CAH populations than in sex-matched controls[2]. This provides evidence in favour of the hypothesis that high concentrations of prenatal testosterone lead to the development of low (i.e. 'male-typical') 2D:4D ratios (Manning et al., 1998); however, this should not be overstated because there are also a number of other potential explanations for our findings. For instance, CAH is additionally associated with reduced concentrations of glucocorticoids and mineralocorticoids, both of which play important roles in bone growth. It was therefore interesting to note that all three studies (Buck et al., 2003; Nave et al., 2020; Ökten et al., 2002) that measured 2D:4D from X-rays reported no significant differences between cases and controls. This could suggest that any difference in 2D:4D between patients with CAH and controls relies on soft tissue rather

[2] It should be noted that the meta-analysis determined that $D_{[R-L]}$ was actually slightly (but not significantly) higher in females with CAH relative to female controls.

than bone length, which is consistent with Wallen's (2009) suggestion that the sex difference in 2D:4D may be due to sex differences in the deposition of adipose tissue in the fingers (see also, Trivers, Jacobson, & Manning, 2020). However, although Ökten et al. (2002) claimed to have observed significant effects only when examining 2D:4D measured from photocopies (i.e. not when examining X-rays of the same participants), doubt is cast on this premise. This is because our re-analysis of the data reported in Table 3 of that paper (p. 51) revealed that phalangeal R2D:4D was actually significantly lower in females with CAH than in female controls ($p = 0.021$). Further, meta-analyses of the subset of studies that measured L2D:4D from radiographs showed either significant differences in the expected direction or non-significant differences depending on how the data from Nave et al. (2020) were coded. The ambiguity of these findings suggests that further studies comparing radiographic 2D:4D between patients with CAH and controls will be required for firm conclusions to be drawn.

Another important consideration is that classical CAH is characterised by very low gestational cortisol levels, and that this is typically treated by administration of glucocorticoids and mineralocorticoids starting shortly after birth. As sexual differentiation of digit ratios appears prenatally (Galis, Ten Broek, Van Dongen, & Wijnaendts, 2010; Malas, Dogan, Evcil, & Desdicioglu, 2006) yet 2D:4D remains somewhat labile during early infancy (Knickmeyer, Woolson, Hamer, Konneker, & Gilmore, 2011), it is feasible that either prenatal cortisol deficiency and/or early postnatal hormone replacement could affect its development. Although no published studies have examined prenatal or early postnatal cortisol concentrations in relation to 2D:4D in humans, foetal testosterone and cortisol have been shown to be positively correlated (Gitau, Adams, Fisk, & Glover, 2005; Sarkar, Bergman, Fisk, O'Connor, & Glover, 2007), and an animal study (Lilley, Laaksonen, Huitu, & Helle, 2010) reported an association between maternal corticosterone levels and offspring 2D:4D ratios in field voles. These observations may suggest that further examination of early cortisol exposure is warranted.

It is noteworthy that, unless considering the lower $D_{[R-L]}$ observed in males with CAH in Constantinescu's (2009) data, none of the individual studies for which sex differences in CAH samples (i.e. specific differences between males with CAH and females with CAH) could be examined (Brown et al., 2002; Constantinescu, 2009; Kim et al., 2017;

Nave et al., 2020; Ökten et al., 2002; Rivas et al., 2014) revealed statistically significant effects in the expected direction (i.e. M<F). However, meta-analysis showed that the sex difference for R2D:4D was significant, and the effect size (g = -0.429) is very similar to that reported by Hönekopp and Watson (2010) for typically developing samples (d = -0.457). Although the effect for L2D:4D was not statistically significant (g = -0.118), this is consistent with the smaller effect size associated with this variable. As diagnostic status (i.e. CAH/control) did not moderate the size of the sex difference for any of the digit ratio variables, these findings suggest that sexual dimorphism for 2D:4D in CAH samples is similar to that already widely established in typically developing populations. Although we did not have an *a priori* prediction, it might have been expected that the magnitude of the sexual dimorphism would be different between these groups because the elevation in prenatal androgen exposure experienced by females with CAH may be relatively greater than that experienced by males with CAH (i.e. because excess adrenal androgen production in males could be at least partially compensated for by downregulation of testicular androgen production). No moderating effect of diagnostic status on the sex difference in 2D:4D could therefore suggest that (1) prenatal androgen levels are elevated to a similar degree in males and females with CAH, (2) elevated prenatal androgen exposure does not explain the difference in 2D:4D between CAH cases and controls, or (3) that a moderating effect does exist but that the available data are underpowered to detect it.

When interpreting the current findings, it should be considered that a particular source of variance between (and sometimes within) studies is the comparability of the CAH and control groups. For instance, in some cases participants were genetic relatives, in other cases they were unrelated; some studies recruited patients who had concerns regarding short stature as controls; age sometimes differed considerably between cases and controls. It may be that future studies that use case and control groups that are more closely matched for key variables (e.g. age, ethnicity) could help determine to what extent differences in 2D:4D associated with CAH may be related to prenatal hormones and how much may be attributed to other aspects of the condition. For instance, work comparing genotype in CAH patients with genetic relatives could control for partial penetrance in those relatives who are unaffected carriers.

The studies included in this literature are also diverse in other ways. For instance, they have come from several different countries, and have used photocopies (Brown et al., 2002; Ökten et al., 2002), X-rays (Buck et al., 2003; Nave et al., 2020; Ökten et al., 2002), direct measures (Rivas et al., 2014), and a combination of both direct measures and photocopies (Constantinescu et al., 2010; Constantinescu, 2009) to record 2D:4D. This likely contributed to the relatively high heterogeneity observed for some of these effects when subjected to meta-analysis, which raises some doubt as to the precision of the affected estimates. However, the observation of similar sized sex differences for digit ratios in CAH samples as in typically developing populations may cast doubt on the premise that the ratio is strongly affected by prenatal testosterone exposure. This could be because, although testosterone levels are elevated in females with CAH, the prenatal levels for males with CAH may not differ markedly from those of typically developing males. One might therefore predict an absent or partially attenuated sex difference within CAH samples.

A particularly interesting observation from the current study was that the right-left difference in 2D:4D ($D_{[R-L]}$), low values of which have been hypothesised to indicate high exposure to foetal testosterone (Manning, 2002; Manning et al., 2014), did not differ between male or female cases and controls. Although a recent study (Baxter, Wood, Witczak, Bales, & Higley, 2019) reported that high levels of maternal urinary testosterone and testosterone-estrone ratio measured during the first trimester of pregnancy predicted low $D_{[R-L]}$ in the offspring (14 males, 12 females) of Titi monkeys, these effects did not retain statistical significance once sex had been controlled for as a covariate (see the analyses presented in the online supplementary materials for that paper). Furthermore, the evidence of such a relationship in humans is even less clear. First, if $D_{[R-L]}$ truly does index individual differences in prenatal androgen exposure in humans, it should arguably exhibit marked sexual differentiation. However, from soon after its inception as a proxy for foetal sex hormone levels. Manning (2002, p. 22) reported that "There may indeed be a tendency for low D_{r-l} in males and high D_{r-l} in females, but the dimorphism is an elusive one." Findings from the BBC Internet Study (the largest ever study of digit ratio: male R2D:4D n=126,343; female R2D:4D n=113,725; male L2D:4D n=126,092; female L2D:4D n=113,389) later showed that R2D:4D is only negligibly lower than L2D:4D in males (d = -0.01) and negligibly higher than L2D:4D in females (d = 0.04) (Manning et al., 2007). Even after

considering that the reliability of the self-measured digit ratios used in this study is estimated to be 46% that of expert measurements (Hönekopp & Watson, 2010), and that random measurement errors multiply when ratios are calculated (Voracek, Manning, & Dressler, 2007), the size of any true effect would appear to be very small. When viewed in light of recent studies showing that testosterone measured from amniotic fluid or maternal circulation during the second trimester are uncorrelated with $D_{[R-L]}$ in newborns (Richards et al., 2019), and that the testosterone-to-oestradiol ratio in amniotic fluid does not predict $D_{[R-L]}$ at 4.5 year follow-up (Richards et al., 2020), doubt is cast on the validity of this measure as an indicator of prenatal androgen action in humans.

A possible limitation of the current research is the 'file drawer problem' (Lane, Luminet, Nave, & Mikolajczak, 2016; Rosenthal, 1979), by which studies with small sample sizes and significant results may be more likely to be published than small studies with null findings. This is an issue that has already been posited in specific regard to CAH research (Collaer & Hines, 2020). For instance, Hampson (2016, p. 427) noted that their data were 'an unfortunate example of this phenomenon', as they were 'many years old but were not submitted for publication until now due to the lack of significant group differences'. Although we made extensive efforts to locate unpublished data relating to 2D:4D in CAH samples, we were only able to identify one unpublished dataset (Constantinescu et al., 2010; Constantinescu, 2009). Our meta-analysis did not include enough samples to obtain reliable estimates of publication bias, but visual inspection of the contour enhanced funnels plots (see **Figure 3** and **Figure 6**) indicates a certain degree of asymmetry. Although necessarily speculative, these plots suggest that there may be some small studies observing effects that are null or in the opposite direction than predicted that are missing from the available literature. Of specific relevance to 2D:4D of course, is that more than one predictor variable (e.g. R2D:4D, L2D:4D, M2D:4D, $D_{[R-L]}$) is often used to simultaneously assess the same hypothesis. This makes detection of publication bias more difficult because such bias is likely based on whether *any* statistically significant effect is reported, not for *which* predictor variable the effect is observed. This could partly mitigate the existence of publication bias, because having some non-significant findings would not be a barrier to publication. However, it also means that while it is possible to fail to detect any evidence of publication bias in meta-analysis due to the presence of non-significant

findings, there could be an unknown number of unpublished studies where none of the 2D:4D comparisons were statistically significant.

In addition to CAH, 2D:4D has been examined in a range of conditions associated with atypical androgen activity, such as CAIS (Berenbaum et al., 2009; van Hemmen et al., 2017), cryptorchidism and/or hypospadias (Abbo et al., 2015; Hwang et al., 2014; O'Kelly, DeCotiis, Zu'bi, Farhat, & Koyle, 2020), polycystic ovarian syndrome (PCOS) (Cattrall, Vollenhoven, & Weston, 2005; Lujan, Bloski, Chizen, Lehotay, & Pierson, 2010; Pandit, Setiya, Yadav, & Jehan, 2016; Roy, Kundu, Sengupta, & Hazra, 2016), autism (Hönekopp, 2012; Manning et al., 2001; Schieve et al., 2018; Teatero & Netley, 2013), ADHD (Martel, Gobrogge, Breedlove, & Nigg, 2008), and gender dysphoria/gender identity disorder and gender variance (Richards, Wei, & Hendriks, 2020; Voracek, Kaden, Kossmeier, Pietschnig, & Tran, 2018). Interest has also been expressed in examining digit ratio in populations with sex chromosome aberrations (Voracek & Dressler, 2007, 2009). Although Manning et al. (2013) reported high 2D:4D in males with Klinefelter's syndrome (47XXY) compared to their unaffected relatives, this effect has not yet been replicated, and, as far as we are aware, no research has yet examined 2D:4D in relation to Jacob's syndrome (47XYY), Turner's syndrome (54XO), or triple X syndrome (47XXX). Regarding Turner's syndrome, Necić & Grant (1978, p. 311) noted that 'A short 4th metacarpal is one of the "text-book" signs of Turner's syndrome', which may imply high (feminine) 2D:4D ratios in this patient group. However, it should also be considered that Turner's syndrome is associated with skeletal aberrations, including fusions of bones in the hands (Preger, Steinbach, Moskowitz, Scully, & Goldberg, 1968), which could make it difficult to interpret findings relating to 2D:4D. Further, it has been suggested that there is some symptom overlap (short stature, varying degrees of virilization, amenorrhoea, menstrual irregularities, infertility) between Turner's syndrome and CAH, and that an elevated rate of 21-hydroxylase deficiency occurs within Turner's syndrome populations (Larizza et al., 1994). These observations may represent important confounds that should be taken into account when examining 2D:4D within such patient groups. Other potential avenues would be to examine 2D:4D in 47XYY males and 47XXX females, as well as in relation to enzymatic disorders such as 5α-reductase deficiency, and 17β-hydroxysteroid dehydrogenase 3 deficiency.

Conclusions

Findings from the current study indicate that 2D:4D is lower in patients with CAH than controls, and that this association is not moderated by sex; however, the meta-analytic effects reported here suggest that these effects are substantially smaller than estimated by earlier studies, casting doubt on the usefulness of 2D:4D as proxy. Although a previous meta-analysis (Hönekopp & Watson, 2010) reported larger effect sizes, the difference in our findings should be considered in light of the fact that we included data from a relatively large study (Constantinescu, 2009) that used an atypical method for quantifying 2D:4D. Nevertheless, the average effect sizes observed for this literature (i.e. the strength of the association between 2D:4D and CAH, a clinical phenotype categorically known to be characterised by elevated prenatal androgen exposure) are small enough to suggest that even studies of 2D:4D that incorporate large samples (i.e. in the hundreds) may be underpowered. We also found no compelling evidence to suggest that the right-left difference in 2D:4D ($D_{[R-L]}$) is significantly different between CAH populations and controls. This is consistent with observations that $D_{[R-L]}$ does not show consistent or large sex differences (Manning et al., 2007), and that it is not correlated with mid-trimester amniotic or maternal circulating testosterone concentrations (Richards et al., 2019, 2020).

Declaration of interest

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors would like to thank Prof S. Marc Breedlove for kindly sharing with us the data from the Brown et al. (2002) study, Prof Melissa Hines and Dr Debra Spencer for sharing with us the data from Constantinescu (2009), and each of the authors who

responded to our request to identify relevant published and unpublished datasets in this area.

References

Abbo, O., Ferdynus, C., Kalfa, N., Huiart, L., Sauvat, F., & Harper, L. H. (2015). Male infants with hypospadias and/or cryptorchidism show a lower 2D/4D digit ratio than normal boys. *Archives of Disease in Childhood*, *100*(7), 643–647. https://doi.org/10.1136/archdischild-2014-306454

Abbott, A. D., Colman, R. J., Tiefenthaler, R., Dumesic, D. A., & Abbott, D. H. (2012). Early-to-mid gestation fetal testosterone increases right hand 2D:4D finger length ratio in polycystic ovary syndrome-like monkeys. *PLoS ONE*, *7*(8), e42372. https://doi.org/10.1371/journal.pone.0042372

Auger, J., Le Denmat, D., Berges, R., Doridot, L., Salmon, B., Canivenc-Lavier, M. C., & Eustache, F. (2013). Environmental levels of oestrogenic and antiandrogenic compounds feminize digit ratios in male rats and their unexposed male progeny. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1768). https://doi.org/10.1098/rspb.2013.1532

Barona, M., Kothari, R., Skuse, D., & Micali, N. (2015). Social communication and emotion difficulties and second to fourth digit ratio in a large community-based sample. *Molecular Autism*, *6*(1), 68. https://doi.org/10.1186/s13229-015-0063-7

Baxter, A., Wood, E. K., Witczak, L. R., Bales, K. L., & Higley, J. D. (2019). Sexual dimorphism in titi monkeys' digit (2D:4D) ratio is associated with maternal urinary sex hormones during pregnancy. *Developmental Psychobiology*. https://doi.org/10.1002/dev.21899

Berenbaum, S. A., Bryk, K. K., Nowak, N., Quigley, C. A., & Moffat, S. (2009). Fingers as a marker of prenatal androgen exposure. *Endocrinology*, *150*(11), 5119–5124. https://doi.org/10.1210/en.2009-0774

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. *Hormones and Behavior*, *42*(4), 380–386. https://doi.org/10.1006/hbeh.2002.1830

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. *Human Reproduction*, *18*(5), 976–979. https://doi.org/10.1093/humrep/deg198

Cattrall, F. R., Vollenhoven, B. J., & Weston, G. C. (2005). Anatomical evidence for in utero androgen exposure in women with polycystic ovary syndrome. *Fertility and Sterility*, *84*(6), 1689–1692. https://doi.org/10.1016/j.fertnstert.2005.05.061

Çetin, R., Can, M., & Özcan, E. (2016). The relatıonship between testosterone and oestrogen level of the cord blood And length of fingers of newborns 2D: 4D. *Balıkesır Health Sciences Journal*, *5*(2), 75–82. https://doi.org/10.5505/bsbd.2016.49469

Ciumas, C., Hirschberg, A. L., & Savic, I. (2009). High fetal testosterone and sexually dimorphic cerebral networks in females. *Cerebral Cortex*, *19*(5), 1167–1174.

https://doi.org/10.1093/cercor/bhn160

Cohen-Bendahan, C. C. C., van de Beek, C., & Berenbaum, S. A. (2005). Prenatal sex hormone effects on child and adult sex-typed behavior: Methods and findings. *Neuroscience & Biobehavioral Reviews*, *29*(2), 353–384. https://doi.org/10.1016/j.neubiorev.2004.11.004

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collaer, M. L., Brook, C. G. D., Conway, G. S., Hindmarsh, P. C., & Hines, M. (2009). Motor development in individuals with congenital adrenal hyperplasia: Strength, targeting, and fine motor skill. *Psychoneuroendocrinology*, *34*(2), 249–258.

Collaer, Marcia L., & Hines, M. (2020). No evidence for enhancement of spatial ability with elevated prenatal androgen exposure in congenital adrenal hyperplasia: A meta-analysis. *Archives of Sexual Behavior*, *49*(2), 395–411. https://doi.org/10.1007/s10508-020-01645-7

Constantinescu, M., Spencer, D., Pasterski, V., Neufeld, S., Glover, V., Hindmarsh, P., … Hines, M. (2010). Do sexually dimorphic finger ratios reflect prenatal exposure to androgens? In *Gender Development Research Conference*. San Francisco, CA.

Constantinescu, Mihaela. (2009). *Are finger ratios a useful measure of androgenic influences on sexual differentiation?* University of Cambridge.

de Sanctis, V., Soliman, A. T., Elsedfy, H., Soliman, N., Elalaily, R., & Di Maio, S. (2017). Is the second to fourth digit ratio (2D:4D) a biomarker of sex-steroids activity? *Pediatric Endocrinology Reviews*, *14*(4), 378–386. https://doi.org/10.17458/per.vol14.2017.SSE.SexSteroids

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Galis, F., Ten Broek, C. M. A., Van Dongen, S., & Wijnaendts, L. C. D. (2010). Sexual dimorphism in the prenatal digit ratio (2D:4D). *Archives of Sexual Behavior*, *39*(1), 57–62. https://doi.org/10.1007/s10508-009-9485-7

Gitau, R., Adams, D., Fisk, N. M., & Glover, V. (2005). Fetal plasma testosterone correlates positively with cortisol. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, *90*(2), F166–F169. https://doi.org/10.1136/adc.2004.049320

Hampson, E. (2016). The development of hand preference and dichotic language lateralization in males and females with congenital adrenal hyperplasia. *Laterality*, *21*(4–6), 415–432. https://doi.org/10.1080/1357650X.2015.1102924

Hannah-Shmouni, F., Morissette, R., Sinaii, N., Elman, M., Prezant, T. R., Chen, W., … Merke, D. P. (2017). Revisiting the prevalence of nonclassic congenital adrenal hyperplasia in US Ashkenazi Jews and Caucasians. *Genetics in Medicine*, *19*(11), 1276–1279. https://doi.org/10.1038/gim.2017.46

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hickey, M., Doherty, D. A., Hart, R., Norman, R. J., Mattes, E., Atkinson, H. C., &

Sloboda, D. M. (2010). Maternal and umbilical cord androgen concentrations do not predict digit ratio (2D:4D) in girls: A prospective cohort study. *Psychoneuroendocrinology*, *35*(8), 1235–1244. https://doi.org/10.1016/j.psyneuen.2010.02.013

Hines, M., Fane, B. A., Pasterski, V. L., Mathews, G. A., Conway, G. S., & Brook, C. (2003). Spatial abilities following prenatal androgen abnormality: Targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. *Psychoneuroendocrinology*, *28*(8), 1010–1026.

Hines, Melissa. (2004). *Brain gender*. New York, NY: Oxford University Press.

Hines, Melissa, Constantinescu, M., & Spencer, D. (2015). Early androgen exposure and human gender development. *Biology of Sex Differences*, *6*(3). https://doi.org/10.1186/s13293-015-0022-1

Hiraishi, K., Sasaki, S., Shikishima, C., & Ando, J. (2012). The second to fourth digit ratio (2D:4D) in a Japanese twin sample: Heritability, prenatal hormone transfer, and association with sexual orientation. *Archives of Sexual Behavior*, *41*(3), 711–724. https://doi.org/10.1007/s10508-011-9889-z

Hollier, L. P., Keelan, J. A., Jamnadass, E. S. L., Maybery, M. T., Hickey, M., & Whitehouse, A. J. O. (2015). Adult digit ratio (2D:4D) is not related to umbilical cord androgen or estrogen concentrations, their ratios or net bioactivity. *Early Human Development*, *91*(2), 111–117. https://doi.org/10.1016/j.earlhumdev.2014.12.011

Hönekopp, J. (2012). Digit ratio 2D:4D in relation to autism spectrum disorders, empathizing, and systemizing: A quantitative review. *Autism Research*, *5*(4), 221–230. https://doi.org/10.1002/aur.1230

Hönekopp, J., Bartholdt, L., Beier, L., & Liebert, A. (2007). Second to fourth digit length ratio (2D:4D) and adult sex hormone levels: New data and a meta-analytic review. *Psychoneuroendocrinology*, *32*(4), 313–321.

Hönekopp, J., & Watson, S. (2010). Meta-analysis of digit ratio 2D:4D shows greater sex difference in the right hand. *American Journal of Human Biology*, *22*(5), 619–630. https://doi.org/10.1002/ajhb.21054

Huber, S. E., Lenz, B., Kornhuber, J., & Müller, C. P. (2017). Prenatal androgen-receptor activity has organizational morphological effects in mice. *PLoS ONE*, *12*(11), e0188752. https://doi.org/10.1371/journal.pone.0188752

Hwang, H., Jo, H. W., Yun, B., Kim, T., Yu, D., Park, J., … Woo, S. H. (2014). The second-to-fourth digit ratio in cryptorchidism: A case-control study. *Korean Journal of Urology*, *55*(2), 140–144. https://doi.org/10.4111/kju.2014.55.2.140

Jordan-Young, R. M. (2012). Hormones, context, and "Brain Gender" : A review of evidence from congenital adrenal hyperplasia. *Social Science & Medicine*, *74*(11), 1738–1744. https://doi.org/10.1016/j.socscimed.2011.08.026

Kalichman, L., Batsevich, V., & Kobyliansky, E. (2019). Heritability estimation of 2D:4D finger ratio in a Chuvashian population-based sample. *American Journal of Human Biology*, e23212. https://doi.org/10.1002/ajhb.23212

Kim, M. S., Koppin, C. M., Yong, J. E., Chand, S. S., Serrano-Gonzalez, M., Geffner, M. E., … Nave, G. (2017). 2D:4D increases with age and lacks sexual dimorphism in classical CAH youth. *Endocrine Reviews*, *38*(3 Supplement).

Retrieved from https://www.endocrine.org/meetings/endo-annual-meetings/abstract-details?ID=30357

Kisch, E. S., Laurian, L., & Hoerer, E. (1987). Nonclassical congenital adrenal hyperplasia. *Israel Journal of Medical Sciences*, *23*(4), 287–289.

Knickmeyer, R. C., Woolson, S., Hamer, R. M., Konneker, T., & Gilmore, J. H. (2011). 2D:4D ratios in the first 2 years of life: Stability and relation to testosterone exposure and sensitivity. *Hormones and Behavior*, *60*(3), 256–263. https://doi.org/10.1016/j.yhbeh.2011.05.009

Kocaman, G. M., Özmerdivenli, R., Yektaş, Ç., Arslanoğlu, İ., Özdemir, E., Bolu, S., … Erdoğan, A. (2016). Autistic feature and 2D:4D finger ratio relations children and adolescents with congenital adrenal hyperplasia. *Acta Physiologica*, *218*(S709), 76–77.

Kocaman, G. M., Özmerdivenli, R., Yektaş, Ç., Bolu, S., Haskiliç, Y. E., & Erdoğan, A. (2017). Konjenital adrenal hiperplazisi olan çocuk ve ergenlerde otistik özellikler ve 2D:4D parmak oranlarının ilişkisi. *Anadolu Psikiyatri Dergisi*, *18*(4), 395–403. https://doi.org/10.5455/apd.239006

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, *2*(1), 61–76. https://doi.org/10.1002/jrsm.35

Králík, M., Ingrová, P., Kozieł, S., Hupková, A., & Klíma, O. (2017). Overall trends vs. individual trajectories in the second-to-fourth digit (2D:4D) and metacarpal (2M:4M) ratios during puberty and adolescence. *American Journal of Physical Anthropology*, *162*(4), 641–656. https://doi.org/10.1002/ajpa.23153

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. *Journal of Neuroendocrinology*, *28*(4). https://doi.org/10.1111/jne.12384

Larizza, D., Cuccia, M., Martinetti, M., Maghnie, M., Dondi, E., Salvaneschi, L., & Severi, F. (1994). Adrenocorticotrophin stimulation and HLA polymorphisms suggest a high frequency of heterozygosity for steroid 21-hydroxylase deficiency in patients with Turner's syndrome and their families. *Clinical Endocrinology*, *40*(1), 39–45. https://doi.org/10.1111/j.1365-2265.1994.tb02441.x

Levine, L. S., Dupont, B., Lorenzen, F., Pang, S., Pollack, M., Oberfield, S., … New, M. I. (1980). Cryptic 21-hydroxylase deficiency in families of patients with classical congenital adrenal hyperplasia. *Journal of Clinical Endocrinology & Metabolism*, *51*(6), 1316–1324. https://doi.org/10.1210/jcem-51-6-1316

Lilley, T., Laaksonen, T., Huitu, O., & Helle, S. (2010). Maternal corticosterone but not testosterone level is associated with the ratio of second-to-fourth digit length (2D:4D) in field vole offspring (Microtus agrestis). *Physiology and Behavior*, *99*(4), 433–437. https://doi.org/10.1016/j.physbeh.2009.11.015

Loehlin, J. C., McFadden, D., Medland, S. E., & Martin, N. G. (2006). Population differences in finger-length ratios: Ethnicity or latitude? *Archives of Sexual Behavior*, *35*(6), 739–742. https://doi.org/10.1007/s10508-006-9039-1

Lujan, M. E., Bloski, T. G., Chizen, D. R., Lehotay, D. C., & Pierson, R. A. (2010). Digit ratios do not serve as anatomical evidence of prenatal androgen exposure in clinical phenotypes of polycystic ovary syndrome. *Human Reproduction*,

25(1), 204–211. https://doi.org/10.1093/humrep/dep363

Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R. C., & 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

Malas, M. A., Dogan, S., Evcil, E. H., & Desdicioglu, K. (2006). Fetal development of the hand, digits and digit ratio (2D : 4D). *Early Human Development*, *82*(7), 469–475. https://doi.org/10.1016/j.earlhumdev.2005.12.002

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

Manning, J. T. (2011). Resolving the role of prenatal sex steroids in the development of digit ratio. *Proceedings of the National Academy of Sciences*, *108*(39), 16143–16144. https://doi.org/10.1073/pnas.1113312108

Manning, J. T., Baron-Cohen, S., Wheelwright, S., & Sanders, G. (2001). The 2nd to 4th digit ratio and autism. *Developmental Medicine and Child Neurology*, *43*(3), 160–164. https://doi.org/10.1111/j.1469-8749.2001.tb00181.x

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). *Archives of Sexual Behavior*, *36*(2), 223–233. https://doi.org/10.1007/s10508-007-9171-6

Manning, J. T., & Fink, B. (2020). Digit ratio references. https://doi.org/10.1007/978-3-319-16999-6_3829-1

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., Kilduff, L. P., & Trivers, R. (2013). Digit ratio (2D:4D) in Klinefelter's syndrome. *Andrology*, *1*(1), 94–99. https://doi.org/10.1111/j.2047-2927.2012.00013.x

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. *Human Reproduction*, *13*(11), 3000–3004. https://doi.org/10.1093/humrep/13.11.3000

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 Human Development*, *80*, 161–168. https://doi.org/10.1016/j.earlhumdev.2004.06.004

Martel, M. M., Gobrogge, K. L., Breedlove, S. M., & Nigg, J. T. (2008). Masculinized finger-length ratios of boys, but not girls, are associated with attention-deficit/hyperactivity disorder. *Behavioral Neuroscience*, *122*(2), 273–281. https://doi.org/10.1037/0735-7044.122.2.273

Mathews, G. A., Fane, B. A., Pasterski, V. L., Conway, G. S., Brook, C., & Hines, M. (2004). Androgenic influences on neural asymmetry: Handedness and language lateralization in individuals with congenital adrenal hyperplasia. *Psychoneuroendocrinology*, *29*(6), 810–822. https://doi.org/10.1016/S0306-4530(03)00145-8

McIntyre, M. H., Cohn, B. A., & Ellison, P. T. (2006). Sex dimorphism in digital formulae of children. *American Journal of Physical Anthropology*, *129*(1), 143–

150. https://doi.org/10.1002/ajpa.20240

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*(1571), 1473–1479. https://doi.org/10.1098/rspb.2005.3100

Miller, A. E. J., MacDougall, J. D., Tarnopolsky, M. A., & Sale, D. G. (1993). Gender differences in strength and muscle fiber characteristics. *European Journal of Applied Physiology and Occupational Physiology*, *66*(3), 254–262.

Mitsui, T., Araki, A., Goudarzi, H., Miyashita, C., Ito, S., Sasaki, S., … Nonomura, K. (2016). Effects of adrenal androgens during the prenatal period on the second to fourth digit ratio in school-aged children. *Steroids*, *113*, 46–51. https://doi.org/10.1016/j.steroids.2016.06.009

Mitsui, T., Araki, A., Imai, A., Sato, S., Miyashita, C., Ito, S., … Nonomura, K. (2015). Effects of prenatal Leydig cell function on the ratio of the second to fourth digit lengths in school-aged children. *PLoS ONE*, *10*(3), 1–11. https://doi.org/10.1371/journal.pone.0120636

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., … Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Myers, L., van't Westeinde, A., Kuja-Halkola, R., Tammimies, K., & Bölte, S. (2018). 2D:4D ratio in neurodevelopmental disorders: A twin study. *Journal of Autism and Developmental Disorders*, *48*(9), 3244–3252. https://doi.org/10.1007/s10803-018-3588-8

Nave, G., Koppin, C. M., Manfredi, D., Richards, G., Watson, S. J., Geffner, M. E., … Kim, M. S. (2020). No evidence for difference in 2D:4D between youth with elevated prenatal androgen exposure due to congenital adrenal hyperplasia and controls. *Hormones and Behavior*.

Nečić, S., & Grant, D. B. (1978). Diagnostic value of hand X-rays in Turner's syndrome. *Acta Paediatrica Scandinavica*, *67*(3), 309–312. https://doi.org/10.1111/j.1651-2227.1978.tb16326.x

New, M. I. (2006). Extensive clinical experience: Nonclassical 21-hydroxylase deficiency. *Journal of Clinical Endocrinology & Metabolism*, *91*(11), 4205–4214. https://doi.org/10.1210/jc.2006-1645

New, M. I., Dupont, B., Pollack, M. S., & Levine, L. S. (1981). The biochemical basis for genotyping 21-hydroxylase deficiency. *Human Genetics*, *58*(1), 123–127. https://doi.org/10.1007/bf00284159

O'Kelly, F., DeCotiis, K., Zu'bi, F., Farhat, W. A., & Koyle, M. A. (2020). Increased hand digit length ratio (2D:4D) is associated with increased severity of hypospadias in pre-pubertal boys. *Pediatric Surgery International*, *36*(2), 247–253. https://doi.org/10.1007/s00383-019-04600-3

Ökten, A., Kalyoncu, M., & Yariş, N. (2002). The ratio of second- and fourth-digit lengths and congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Early Human Development*, *70*(1–2), 47–54. https://doi.org/10.1016/S0378-3782(02)00073-7

Oświęcimska, J. M., Książek, A., Sygulla, K., Pyś-Spychała, M., Roczniak, G. R., Roczniak, W., … Ziora, K. (2012). Androgens concentrations and second-to fourth-digit ratio (2D:4D) in girls with congenital adrenal hyperplasia (21-hydroxylase deficiency). *Neuroendocrinology Letters*, *33*(8), 787–791.

Pandit, V. K., Setiya, M., Yadav, S., & Jehan, M. (2016). Digit ratio (2D:4D): A potential anatomical biomarker for predicting the risk of development of polycystic ovarian syndrome. *IOSR Journal of Dental and Medical Sciences*, *15*(8), 58–64. https://doi.org/10.9790/0853-1508105864

Pang, S., Levine, L. S., Chow, D. M., Faiman, C., & New, M. I. (1979). Serum androgen concentrations in neonates and young infants with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Clinical Endocrinology*, *11*(6), 575–584. https://doi.org/10.1111/j.1365-2265.1979.tb03111.x

Pang, S, Levine, L. S., Cederqvist, L. L., Fuentes, M., Riccardi, V. M., Holcombe, J. H., … New, M. I. (1980). Amniotic fluid concentrations of Δ5 and Δ4 steroids in fetuses with congenital adrenal hyperplasia due to 21-hydroxylase deficiency and in anencephalic fetuses. *Journal of Clinical Endocrinology and Metabolism*, *51*(2), 223–229. https://doi.org/10.1210/jcem-51-2-223

Pang, Songya, Wallace, M. A., Hofman, L., Thuline, H. C., Dorche, C., Lyon, I. C. T., … Suwa, S. (1988). Worldwide experience in newborn screening for classical congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Pediatrics*, *81*(6), 866–874.

Preger, L., Steinbach, H. L., Moskowitz, P., Scully, L., & Goldberg, M. B. (1968). Roentgenographic abnormalities in phenotypic females with gonadal dysgenesis: A comparison of chromatin positive patients and chromatin negative patients. *American Journal of Roentgenology*, *104*(4), 899–910. https://doi.org/10.2214/ajr.104.4.899

Puts, D. A., McDaniel, M. A., Jordan, C. L., & Breedlove, S. M. (2008). Spatial ability and prenatal androgens: Meta-analyses of congenital adrenal hyperplasia and digit ratio (2D:4D) studies. *Archives of Sexual Behavior*, *37*(1), 100–111. https://doi.org/10.1007/s10508-007-9271-3.

Ratcliffe, S. G., Read, G., Pan, H., Fear, C., Lindenbaum, R., & Crossley, J. (1994). Prenatal testosterone levels in XXY and XYY males. *Hormone Research*, *42*(3), 106–109. https://doi.org/10.1159/000184157

Richards, G. (2017). What is the evidence for a link between digit ratio (2D:4D) and direct measures of prenatal sex hormones? *Early Human Development*, *113*, 71–72. https://doi.org/10.1016/j.earlhumdev.2017.08.003

Richards, G., Browne, W., & Constantinescu, M. (2020). Digit ratio (2D:4D) and amniotic testosterone and estradiol: An attempted replication of Lutchmaya et al. (2004). https://doi.org/10.1101/2020.07.10.197269

Richards, G., Gomes, M. C., & Ventura, T. (2019). Testosterone measured from amniotic fluid and maternal plasma shows no significant association with directional asymmetry in newborn digit ratio (2D:4D). *Journal of Developmental Origins of Health and Disease*, *10*(3), 362–367. https://doi.org/10.1017/S2040174418000752

Richards, G., Wei, Y., & Hendriks, O. (2020). Self-measure digit ratio (2D:4D) and gender variance. *Endocrine Practice*, *26*(2), 250–251.

https://doi.org/10.4158/1934-2403-26.2.250

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. *American Journal of Human Biology*, *26*(4), 559–561. https://doi.org/10.1002/ajhb.22545

Romano, M., Rubolini, D., Martinelli, R., Alquati, A. B., & Saino, N. (2005). Experimental manipulation of yolk testosterone affects digit length ratios in the ring-necked pheasant (Phasianus colchicus). *Hormones and Behavior*, *48*(3), 342–346. https://doi.org/10.1016/j.yhbeh.2005.03.007

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Roy, R., Kundu, R., Sengupta, M., & Hazra, A. (2016). Assessment of 2nd and 4th digit length ratio as an anatomical marker for predicting the risk of developing polycystic ovarian syndrome. *Journal of the Anatomical Society of India*, *65*(S), S18. https://doi.org/10.1016/j.jasi.2016.08.064

Saino, N., Rubolini, D., Romano, M., & Boncoraglio, G. (2007). Increased egg estradiol concentration feminizes digit ratios of male pheasants (Phasianus colchicus). *Naturwissenschaften*, *94*(3), 207–212. https://doi.org/10.1007/s00114-006-0188-9

Sarkar, P., Bergman, K., Fisk, N. M., O'Connor, T. G., & Glover, V. (2007). Amniotic fluid testosterone: Relationship with cortisol and gestational age. *Clinical Endocrinology*, *67*(5), 743–747. https://doi.org/10.1111/j.1365-2265.2007.02955.x

Schieve, L. A., Tian, L., Dowling, N., Croen, L., Hoover-Fong, J., Alexander, A., & Shapira, S. K. (2018). Associations between the 2nd to 4th digit ratio and autism spectrum disorder in population-based samples of boys and girls: Findings from the Study to Explore Early Development. *Journal of Autism and Developmental Disorders*, *48*(7), 2379–2395. https://doi.org/10.1007/s10803-018-3495-z

Talarovičová, A., Kršková, L., & Blažeková, J. (2009). Testosterone enhancement during pregnancy influences the 2D:4D ratio and open field motor activity of rat siblings in adulthood. *Hormones and Behavior*, *55*(1), 235–239. https://doi.org/10.1016/j.yhbeh.2008.10.010

Teatero, M. L., & Netley, C. (2013). A critical review of the research on the extreme male brain theory and digit ratio (2D:4D). *Journal of Autism and Developmental Disorders*, *43*, 2664–2676. https://doi.org/10.1007/s10803-013-1819-6

Trivers, R. L., Jacobson, A., & Manning, J. T. (2020). Radiographic digit ratios (2D:4D) of Afro-Caribbean children: Comparisons with published data from white children. *Early Human Development*, *146*, 105072. https://doi.org/10.1016/j.earlhumdev.2020.105072

Trivers, R., Manning, J. T., & Jacobson, A. (2006). A longitudinal study of digit ratio (2D:4D) and other finger ratios in Jamaican children. *Hormones and Behavior*, *49*(2), 150–156. https://doi.org/10.1016/j.yhbeh.2005.05.023

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.

Biology of Sex Differences, *8*(1), 11. https://doi.org/10.1186/s13293-017-0132-z

Ventura, T., Gomes, M. C., Pita, A., Neto, M. T., & Taylor, A. (2013). Digit ratio (2D:4D) in newborns: Influences of prenatal testosterone and maternal environment. *Early Human Development*, *89*(2), 107–112. https://doi.org/10.1016/j.earlhumdev.2012.08.009

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Voracek, M., Kaden, A., Kossmeier, M., Pietschnig, J., & Tran, U. S. (2018). Meta-analysis shows associations of digit ratio (2D:4D) and transgender identity are small at best. *Endocrine Practice*, *24*(4), 386–390. https://doi.org/10.4158/EP-2017-0024

Voracek, Martin. (2008). Digit ratio (2D:4D) as a marker for mental disorders: Low (masculinized) 2D:4D in autism-spectrum disorders, high (feminized) 2D:4D in schizophrenic-spectrum disorders. *Behavioral and Brain Sciences*, *31*(3), 283–284. https://doi.org/10.1017/S0140525X08004457

Voracek, Martin, & Dressler, S. G. (2007). Digit ratio (2D:4D) in twins: Heritability estimates and evidence for a masculinized trait expression in women from opposite-sex pairs. *Psychological Reports*, *100*, 115–126. https://doi.org/10.2466/PR0.100.1.115-126

Voracek, Martin, & Dressler, S. G. (2009). Brief communication: Familial resemblance in digit ratio (2D:4D). *American Journal of Physical Anthropology*, *140*(2), 376–380. https://doi.org/10.1002/ajpa.21105

Voracek, Martin, & Loibl, L. M. (2009). Scientometric analysis and bibliography of digit ratio (2D:4D) research, 1998–2008. *Psychological Reports*, *104*(3), 922–956. https://doi.org/10.2466/PR0.104.3.922-956

Voracek, Martin, Manning, J. T., & Dressler, S. G. (2007). Repeatability and interobserver error of digit ratio (2D:4D) measurements made by experts. *American Journal of Human Biology*, *19*(1), 142–146. https://doi.org/10.1002/ajhb.20581

Wallen, K. (2009). Does finger fat produce sex differences in second to fourth digit ratios? *Endocrinology*, *150*(11), 4819–4822. https://doi.org/10.1210/en.2009-0986

Watson, N. V. (2001). Sex differences in throwing: Monkeys having a fling. *Trends in Cognitive Sciences*, *5*(3), 98–99. https://doi.org/10.1016/S1364-6613(00)01595-3

Whitehouse, A. J. O., Gilani, S. Z., Shafait, F., Mian, A., Tan, D. W., Maybery, M. T., … Eastwood, P. (2015). Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1816), 20151351. https://doi.org/10.1098/rspb.2015.1351

Wudy, S. A., Dörr, H. G., Solleder, C., Djalali, M., & Homoki, J. (1999). Profiling steroid hormones in amniotic fluid of midpregnancy by routine stable isotope dilution/gas chromatography-mass spectrometry: Reference values and concentrations in fetuses at risk for 21-hydroxylase deficiency. *Journal of Clinical Endocrinology & Metabolism*, *84*(8), 2724–2728.

https://doi.org/10.1210/jc.84.8.2724

Zhang, K., Yang, X., Zhang, M., Wang, C., Fang, P., Xue, M., … Gong, P. (2019). Revisiting the relationships of 2D:4D with androgen receptor (AR) gene and current testosterone levels: Replication study and meta-analyses: Supplementary materials. *Journal of Neuroscience Research*. https://doi.org/10.1002/jnr.24502

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