



OPEN

## Digit ratio (2D:4D) and transgender identity: new original data and a meta-analysis

Eva-Maria Siegmann<sup>1</sup>✉, Tobias Müller<sup>1,2</sup>, Isabelle Dziadeck<sup>1</sup>, Christiane Mühle<sup>1</sup>, Bernd Lenz<sup>1,3,4</sup> & Johannes Kornhuber<sup>1,4</sup>

Previously reported associations between second-to-fourth digit length ratio (2D:4D), a proxy for prenatal androgen load, and transgender identity have been inconsistent. The objectives of the present study were to provide additional original data and an updated meta-analysis concerning this association. In a study of 464 participants, we compared the 2D:4D of transgender individuals with age- and sex-matched controls. Patients were recruited at a specialized psychiatrist's medical office, whereas controls were hired via flyers, advertisements, and as convenience sample. A random-effects meta-analysis of the literature (17 samples,  $n = 3674$ ) also quantifies the overall magnitude of the difference in 2D:4D between transgender individuals and controls. In our study providing new original data, we found a significantly higher (i.e. feminized) left-hand 2D:4D in the male-to-female transgender (MtF) identity [mean age: 32.3 (18; 61)] than in the male control group [mean age: 34.5 (18; 65)] with a Cohen's  $d = 0.271$ . Concordantly, the meta-analytic results suggest a significant difference in 2D:4D among MtF individuals compared to male controls [ $g = 0.153$ ; 95% CI (0.063; 0.243)], which was even more pronounced when individuals had been diagnosed by a clinician instead of self-identified as transgender [ $g = 0.193$ ; 95% CI (0.086; 0.300)]. In both studies, no significant results were revealed for female-to-male transgender individuals [mean age: 26.1 (18; 53)] versus female controls [mean age: 27.2 (18; 55)]. This original investigation and the updated meta-analysis clarify the association between transgender identity and 2D:4D indicating the influence of prenatal androgen on the development of gender identity in subjects born as males.

Transgender identity or gender dysphoria are defined as experiencing an inconsistency between physical phenotype and one's perceived gender<sup>1</sup>. Evidence suggests that the prevalence has increased over the last decades to up to 5–14 male-to-female transgender (MtF) individuals per 1000 adult males and 2–3 female-to-male transgender (FtM) individuals per 1000 adult females<sup>2</sup>. This growth in prevalence is possibly due to greater social acceptance, de-pathologization, and greater awareness of therapeutic options<sup>2</sup>. Despite the growing amount of research in this area, causal mechanisms are still unclear. Animal studies show an association between perinatal testosterone and the size of the bed nucleus of the stria terminalis (BNST)<sup>3</sup>, a region linked to gender identity in humans<sup>4</sup>. The theory of early organizational testosterone effects on gender identity is further supported by review evidence that transgender individuals' brains show changes away from their natal sex and toward their perceived gender<sup>5</sup>. Additionally, studies examining subjects with congenital adrenal hyperplasia (CAH) or subjects with partial or complete androgen insensitivity syndrome (AIS/CAIS) suggest that (1) levels of gender dysphoria are higher when brain androgenization *mismatches* gender of upbringing and (2) that levels of gender dysphoria are lower when brain androgenization *matches* gender of upbringing<sup>6</sup>. For a more detailed description of gender dysphoria in patients with disorders of sex development and possible confounding variables, see Sadr et al.<sup>6</sup>. In humans, however, it is not possible to directly investigate the effects of prenatal androgens owing to ethical reasons and the long time period between the intrauterine window and the establishment of gender identity. Thus, this research area relies on proxies such as the second-to-fourth digit length ratio (2D:4D), a biomarker determined by the balance of prenatal testosterone to estrogen<sup>7</sup>, and therefore indicating the prenatal androgen load<sup>8</sup>. A lower 2D:4D reflects higher prenatal androgen exposure<sup>9</sup> and hence, the 2D:4D is lower in males than

<sup>1</sup>Department of Psychiatry and Psychotherapy, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Schwabachanlage 6, 91054 Erlangen, Germany. <sup>2</sup>Psychiatric Practice, Treibberg 5, 90403 Nuremberg, Germany. <sup>3</sup>Department of Addictive Behavior and Addiction Medicine, Central Institute of Mental Health (CIMH), Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany. <sup>4</sup>These authors contributed equally: Bernd Lenz and Johannes Kornhuber. ✉email: eva-maria.siegmann@uk-erlangen.de

in females<sup>10,11</sup>. It is frequently used to examine the link between prenatal androgen exposure and postnatal phenotypes including diseases, mental disorders, and behaviors reflecting gender bias or being influenced by first trimester organizational changes induced by testosterone/estrogen ratios<sup>12</sup>, such as alcohol addiction<sup>13</sup>, autism spectrum disorders<sup>14,15</sup>, sexual orientation<sup>16</sup>, the risk for different forms of cancer<sup>17</sup>, and physical prowess<sup>18</sup>. Concerning gender dysphoria and 2D:4D, contradictory evidence exists, which has been summarized in a meta-analytical commentary conducted in 2017<sup>19</sup> where the authors report a small effect size [ $g = 0.190$ ; 95% CI (0.076; 0.304)] indicating feminized 2D:4D of the right hand (but not the left hand) in MtF individuals and no associations at all in FtM individuals. This short systematic review included only few studies with rather small sample sizes or data retrieved from abstracts (which can be influenced by reporting bias), and it did not provide extensive meta-regression and subgroup analyses or distinguish between case-control and cohort studies. In 2020, a more recent meta-analysis<sup>6</sup> reported significantly feminized 2D:4D in MtF individuals ( $d = 0.24$ ). This analysis applied rather strict inclusion criteria (e.g. excluding child samples) and did also not distinguish between patient and cohort samples. The aim of the present research was to re-address this question by adding novel original data from a larger sample size and to subject the entire body of literature to a meta-analysis while taking potentially confounding factors into account. In previous meta-analytical research on 2D:4D and addiction, we found that the effects are more pronounced when cases have been defined via clinical diagnosis instead of via self-disclosure<sup>13</sup>. Therefore, we especially focused on the difference between transgender individuals diagnosed by a specialized clinician according to classification systems versus self-identification as transgender or gender nonconforming. Gender nonconformity extends the term of transgender individuals insofar that it includes other gender-variant identities or non-binary people, as well. For both the original research and the meta-analysis, we tested the hypothesis that 2D:4D is larger, i.e. feminized, in MtF individuals than in male controls. When comparing FtM individuals with female controls, we decided to follow an exploratory approach since previous (meta-analytical) evidence<sup>6,19</sup> is not conclusive. Following the results of our own previous meta-analytical research on 2D:4D<sup>13</sup>, we defined a second hypothesis regarding the meta-analysis, namely, that the difference between transgender individuals and controls is more pronounced when the transgender person has been diagnosed by a specialized clinician.

**Aims of the study.** We wanted to investigate the difference in the second-to-fourth digit length ratio between transgender individuals and control subjects. Therefore, we conducted an original case-control study with large sample size and subjected the entire body of literature to a meta-analysis.

### Study 1: new original data

This study investigated a possible association of 2D:4D and transsexuality in a large sample. It was part of the Finger Length in Psychiatry (FLIP) project<sup>13,20–27</sup>, a line of independent studies aiming to elucidate the role of prenatal sex hormone organization in the etiopathogenesis of psychiatric disorders in childhood, adolescence, and adulthood, such as addictive disorders<sup>22,26,28,29,30</sup>, suicide completion<sup>31</sup>, or behavioral symptoms in children<sup>32</sup>. Another goal was to establish prevention targets and strategies based on this model<sup>33</sup>.

**Materials and methods.** *Sample description.* From 2016 to 2019, we recruited MtF and FtM patients and male and female control groups. In total, datasets of 110 MtF and 151 FtM patients diagnosed with *transsexualism* according to ICD-10<sup>34</sup> and *gender dysphoria* according to DSM-5<sup>35</sup> and 101 male and 102 female age-matched control subjects were analyzed. The patients were enrolled in the study by a psychiatrist specialized in transsexualism (T.M.) at a medical office in Nuremberg, Germany. The control subjects were hired via flyers and advertisements (Facebook advertisements, posts on black boards) or as convenience sample (neighborhood in Erlangen and Bergisch-Gladbach, Germany). They were recruited by a medical student (I.D.) in the Department of Psychiatry and Psychotherapy of the University Hospital Erlangen, Germany.

Inclusion criteria for both groups consisted of providing written informed consent, age  $\geq 18$  years, and body mass index (BMI) of 18.5–35.0 kg/m<sup>2</sup>. Exclusion criteria consisted of severe physical illness, the diagnosis of an androgen insensitivity syndrome or an adrenogenital syndrome, acute suicidality, pregnancy, breastfeeding, and the endocrine disorders Morbus Addison, Morbus Cushing, diabetes type 1 and type 2, and endocrine cancers. We did also not enroll control subjects taking regular medications (except for contraceptives), with any history of psychiatric in-patient treatment, or an ambulatory consultation due to a mental illness during the preceding 10 years.

*Gender identity/Gender Dysphoria Questionnaire and Utrecht Gender Dysphoria Scale.* Independent from each other, three study team members (T.M., C.M., B.L.) translated the 54 items of the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (GIDYQ-AA; 27 for the male and 27 for the female questionnaire)<sup>36</sup> and the 24 items of the Utrecht Gender Dysphoria Scale (UGDS)<sup>37</sup> into German. Afterwards, discrepancies were discussed, and a single translation was selected in agreement. Whereas the original GIDYQ-AA uses a mean with low scores implying gender dysphoria, the original UGDS uses a sum score with high scores implying gender dysphoria. In order to increase comparability, we reversed the GIDYQ scores and converted item scores in the UGDS to mean values. Thus, higher scores of the gender dysphoria scales used in this study represent higher gender dysphoria. The minimum and maximum scores possible were 1 and 5 (similar to Deogracias et al.<sup>36</sup>). High internal consistencies (Cronbach's alpha) were found for the German versions: (1) 0.984 for the male GIDYQ ( $n = 197$ ); (2) 0.985 for the female GIDYQ ( $n = 238$ ); (3) 0.989 for the male UGDS ( $n = 197$ ); and (4) 0.981 for the female UGDS ( $n = 238$ ). Separate Cronbach's alpha values for transgender individuals and controls are provided in Supplementary Table S1 (online supplement). The Pearson correlations between GIDYQ and

UGDS are  $r=0.599$  ( $p<0.001$ ) for the MtF sample ( $n=96$ ),  $r=0.087$  ( $p=0.389$ ) for the male sample ( $n=101$ ),  $r=0.240$  ( $p=0.005$ ) for the FtM sample ( $n=136$ ), and  $r=0.474$  ( $p<0.001$ ) for the female sample ( $n=102$ ).

**Other questionnaires.** We assessed the participants' sexual orientation on a scale from 1 (=attracted only to women) to 7 (=attracted only to men). Following a classification proposed by Lawrence et al.<sup>38</sup> we examined early vs. late onset gender dysphoria in transgender individuals by asking about the onset time point (before vs. after puberty) and about the subjects' discomfort with the assigned gender at birth during their childhood (present vs. absent).

**Second-to-fourth finger length ratio.** Right and left hands of the participants were scanned on common document scanners. The lengths of the index and ring fingers (distance from the middle of the basal crease to the tip of the fingers) were quantified using the GNU Image Manipulation Program (GIMP 2.8; <https://www.gimp.org>). Each finger was measured three times by each of three independent raters (9 times in total). The raters evaluated the scans of patients and controls separately. The files were numbered consecutively (controls) or with a patient's code. Thus, the raters were blind concerning the sex of the control or the patients' status as MtF or FtM individual. 2D:4D values were calculated for the mean of right-hand and left-hand 2D:4D (M2D:4D), right-hand 2D:4D (R2D:4D), left-hand 2D:4D (L2D:4D), and differences between R2D:4D and L2D:4D (2D:4Dr-l) since low 2D:4Dr-l values have been associated with high prenatal testosterone load, as well<sup>8</sup>. We found high inter-rater reliabilities (two-way random interrater correlation coefficient; absolute agreement): (1) M2D:4D:  $n=462$ , 0.990; (2) R2D:4D:  $n=464$ , 0.986; (3) L2D:4D:  $n=462$ , 0.983; and (4) 2D:4Dr-l:  $n=462$ , 0.946.

**Ethical standards.** The study was approved by the Ethics Committee of the Medical Faculty of the Friedrich-Alexander University Erlangen-Nürnberg (FAU; ID 194\_16 B). All participants provided written informed consent.

**Statistical analysis.** For each natal sex individually, we compared M2D:4D, R2D:4D, L2D:4D, and 2D:4Dr-l between the patient and the age-matched control groups using the Student's *t*-test. Levene's test for homogeneity of variance was also used and if necessary, the degrees of freedom were adjusted. The  $\chi^2$  test was used to evaluate differences in the frequency of nominal variables. Pearson's correlations were used to investigate continuous relationships between the 2D:4D measures and GIDYQ and UGDS scores in each group. We present the data as means and standard deviations or relative frequencies. Statistical significance was determined at  $p<0.05$  (2-sided). The data were analyzed using IBM SPSS statistics Version 24 for Windows (SPSS Inc., Chicago, IL, USA).

**Results. Sociodemographic characteristics.** The MtF group did not significantly differ from the male group with regard to age, months of employment during the previous year, BMI, or marital status (Table 1). We found significantly lower body height, body weight, and likeliness to live in a partnership in MtF than in male participants. The FtM group did not differ from the female group in terms of age, months of employment during the previous year, body height, likeliness to live in a partnership, or marital status. The FtM group showed a significantly higher BMI and a higher weight. MtF and FtM groups scored significantly higher on the GIDYQ and UGDS (suggesting higher gender dysphoria) than the male and female control groups.

**Second-to-fourth finger length ratio.** We found higher M2D:4D, R2D:4D, and L2D:4D in MtF than in male participants and lower M2D:4D, R2D:4D, and L2D:4D in FtM than in female participants; however, only the higher L2D:4D in MtF than in male participants reached significance with an effect size of Cohen's  $d=0.271$  (Table 1), a result that partly supported our first hypothesis. The GIDYQ scores correlated significantly negatively with M2D:4D and R2D:4D in the MtF group, positively with L2D:4D, and negatively with 2D:4Dr-l in the male control group. Moreover, the UGDS score correlated negatively with L2D:4D in the male control group (Supplementary Table S2). Our analyses revealed no significant association between the transgender individuals' sexual orientation and M2D:4D, R2D:4D, L2D:4D, and 2D:4Dr-l (Supplementary Table S3). There is no significant difference in M2D:4D, R2D:4D, L2D:4D, and 2D:4Dr-l between early and late onset gender dysphoria in our sample (data not shown). This result needs to be interpreted cautiously since sample sizes in the subgroups were small. We report frequencies of early and late onset gender dysphoria in Supplementary Table S4 (online supplement).

**Discussion.** We found significantly feminized 2D:4D in MtF individuals compared to male controls in the left hand, but not in the right hand and no significant results in FtM individuals versus female controls. However, the nonsignificant differences were in accordance with the prenatal androgen hypothesis. Comprising 464 participants, to our knowledge, this investigation presents the largest sample size among case-control studies of 2D:4D and transgender identity. Nevertheless, a power analysis by Voracek et al.<sup>19</sup> underlying their meta-analytical effect size suggests that a sample size of at least 872 participants would be necessary to be able to detect small effects. Due to small prevalence rates of transgender identity, it is difficult to reach such numbers in clinical settings.

The correlational analyses concerning 2D:4D and the UGDS scores did not reveal any consistent pattern and remain contradictory. When examining the scatterplot of 2D:4D and the GIDYQ score, an inverted U-shaped relationship emerged. Comparable U-shaped associations have been reported in previous studies, such as with respect to 2D:4D and academic performance, altruism, or targeting reaction time (for an overview, see Tektas

	Male-to-female N=110			Male N=101			t, $\chi^2$	P	d	Female-to-male N=151			Female N=102			t, $\chi^2$	P	d
	N	Mean	SD	N	Mean	SD				N	Mean	SD	N	Mean	SD			
Age (years)	110	32.3	12.0	101	34.5	13.8	-1.2	0.235		151	26.1	7.9	102	27.2	8.2	-1.1	0.276	
Months of employment during the previous year	96	7.6	5.0	101	8.1	5.0	-0.7	0.499		138	7.5	5.1	99	7.4	5.0	0.3	0.785	
Body mass index (kg/m <sup>2</sup> )	110	23.9	4.6	101	24.7	3.7	-1.4	0.163		151	26.3	7.2	102	22.7	2.9	5.5	<0.0001	
Height (cm)	110	178	8	101	183	6	-5.1	<0.0001		151	167	7	102	167	7	0.1	0.955	
Weight (kg)	110	75.7	15.5	101	82.7	11.7	-3.7	0.0002		151	73.7	21.0	102	63.4	8.9	5.3	<0.0001	
Living in a partnership (%)	103	42.7		101	64.4		9.6	0.002		140	45.7		101	58.4		3.8	0.052	
Married (%)	103	25.2		100	26.0		<0.1	0.902		140	7.1		102	11.8		1.5	0.217	
GIDYQ mean score <sup>a</sup>	96	3.9	0.3	101	1.1	0.1	77.8	<0.0001		136	4.1	0.2	102	1.1	0.2	119.6	<0.0001	
UGDS mean score	96	4.4	0.5	101	1.1	0.1	61.6	<0.0001		136	4.7	0.3	102	1.5	0.4	72.7	<0.0001	
M2D:4D	109	0.966	0.027	101	0.960	0.027	-1.6	0.114	0.222	150	0.972	0.027	102	0.974	0.029	0.7	0.457	-0.071
R2D:4D	110	0.963	0.030	101	0.959	0.028	-1.0	0.314	0.138	151	0.972	0.028	102	0.975	0.030	0.9	0.392	-0.103
L2D:4D	109	0.969	0.029	101	0.961	0.030	-2.0	0.04973	0.271	150	0.971	0.029	102	0.974	0.032	0.6	0.578	-0.098
2D:4Dr-l	109	-0.007	0.021	101	-0.003	0.021	1.4	0.155	-0.190	150	0.001	0.021	102	0.002	0.024	0.3	0.746	-0.044

**Table 1.** Group comparisons (Student's *t* tests and  $\chi^2$ -tests): male-to-female versus male and female-to-male versus female groups. *GIDYQ* Gender Identity/Gender Dysphoria Questionnaire (absolute range 1–5, higher scores indicate stronger dysphoria), *UGDS* Utrecht Gender Dysphoria Scale (absolute range 1–5, higher scores indicate stronger dysphoria), *2D:4D* second-to-fourth finger length ratio, *M2D:4D* mean of *R2D:4D* and *L2D:4D*, *R2D:4D* right-hand 2D:4D, *R2D* length of the right-hand index finger, *R4D* length of the right-hand ring finger, *L2D:4D* left-hand 2D:4D, *L2D* length of the left-hand index finger, *L4D* length of the left-hand ring finger, *2D:4Dr-l* difference between *R2D:4D* and *L2D:4D*, *d* Cohen's *d*. *p* < 0.05 in bold, <sup>a</sup>reversed.

et al.<sup>20</sup>). Moreover, the low standard deviations (SDs) in both the *GIDYQ* and the *UGDS* could attenuate any association with *2D:4D*, thus entailing non-significant results. The low SDs further bring small Cronbach's alpha values (as can be seen in Supplementary Table S1, online supplement), indicating that the two scales do not differentiate well among members of the same group. Future research will be necessary to clarify the role of *2D:4D* in the severity of gender dysphoria.

We used scanners to evaluate the subjects' finger lengths. This method is said to produce lower *2D:4D* values<sup>39,40</sup>, but to better reflect the sexual dimorphism<sup>41</sup> when compared to direct measurement methods such as caliper. The high interrater correlation coefficients found in our study reflect the reliable accuracy of this method. Manning<sup>12</sup> proposed that results are more precise when measuring *2D:4D* directly than via scans, but a recent meta-analysis did not confirm this assumption<sup>6</sup>.

Our results match the existing, conflicting literature of *2D:4D* and transgender identity where significant results have been observed partly in FtM, but not MtF individuals<sup>42,43</sup>, and partly in MtF, but not FtM individuals<sup>44,45</sup>. This emphasizes the necessity of meta-analytical reviews in this research area.

## Study 2: a meta-analysis

**Materials and methods.** *Search strategy.* A two-step literature search was conducted using Google Scholar and PubMed and including abstracts in English from study inception until December 10, 2018. In the course of the review process we updated this literature search from study inception until February 28, 2020. The following search terms were combined in several ways: *2D:4D*, *2nd to 4th digit ratio*, *second to fourth digit ratio*, *second-to-fourth finger length ratio*, *index and ring digit length ratio*, *transgender*, *transsexual*, *gender-variant identity*, *gender identity disorder*, *gender dysphoria*, *transsexualism*, *Harry Benjamin's syndrome*, *gender divers*, and *gender nonconforming*. The reference lists of retrieved articles were searched manually in the second step. All abstracts were screened applying the selection criteria. The remaining articles were checked for eligibility according to the Preferred Reporting Items for Systemic Reviews and Meta-analyses (PRISMA) statement<sup>46</sup> on the basis of a full-text review.

*Study selection.* We included case-control, cohort, and cross-sectional studies. The eligibility criteria were defined analogously to our meta-analytic review on *2D:4D* and substance and computer use<sup>13</sup> and are detailed in our coding protocol (Supplementary Table S5). The literature search was summarized according to the PRISMA guidelines<sup>47</sup>.

*Data extraction.* All recorded variables can be found in the previously defined coding protocol (Supplementary Table S5).



Data extraction and coding were performed by one investigator (E.-M.S.) and one research assistant (C.R.), independently of one another. Disagreement was resolved by discussion and agreeing upon the extracted values. Regarding quality assessment, the extractors' values were averaged. Following our approach in Siegmann et al.<sup>13</sup>, we assessed the risk of bias with either the Newcastle–Ottawa Scale for case–control studies<sup>48</sup> or an adaptation of the Newcastle–Ottawa Scale for cohort studies<sup>48</sup>, which was specifically designed for cross-sectional studies by Herzog et al.<sup>49</sup> in their systematic review.

**Statistical analysis.** All analyses were conducted and all figures were made using the metafor package<sup>50</sup> within the open-source software environment R, version 3.4.2<sup>51</sup>.

We estimated the standardized mean difference (Hedges' *g*) in 2D:4D among MtF individuals and male controls in addition to among FtM individuals and female controls. In this analysis, a more feminized 2D:4D among MtF individuals compared to male controls is reflected by a positive Hedges' *g*, whereas a more masculinized 2D:4D among FtM individuals compared to female controls is reflected by a negative Hedges' *g*. Furthermore, we tested whether these mean differences were more pronounced among transgender individuals who had been reliably diagnosed by a clinician. For this purpose, we split the sample into transgender patients diagnosed by a clinician using common classification systems (DSM-5, DSM-IV(-TR), ICD-10) versus transgender patients not diagnosed by a clinician. The group of transgender individuals not diagnosed by a clinician consisted of (1) people subthreshold for the diagnosis, (2) people identifying as gender-variant, assessed with a shortened adaptation of the cross gender questionnaire<sup>52</sup>, (3) people identifying as tomboys, and (4) studies reporting correlative data. Correlative data were transformed into Hedges' *g* using common transformation formulas<sup>53</sup>.

Moreover, we estimated Hedges' *g* in the difference of R2D:4D and L2D:4D (2D:4Dr-l) among transgender individuals and controls since low 2D:4Dr-l values have been associated with high prenatal testosterone load<sup>8</sup>. We computed 2D:4Dr-l as the difference between the mean R2D:4D and mean L2D:4D, and related standard deviations were approximated by the pooled standard deviation of the R2D:4D and L2D:4D variances.

We performed univariate random-effects meta-analyses using restricted maximum likelihood estimations in which the point estimate for each study was weighted by the inverse of its variance. Non-independence among effect sizes was accounted for by aggregating, for example, in cases in which studies reported separate data for R2D:4D and L2D:4D. Two studies reported solely one 2D:4D value (right hand only<sup>54</sup>, dominant hand only<sup>42</sup>) entailing that our reported meta-analytic 2D:4D effect size refers to an aggregated 2D:4D rather than to M2D:4D. 2D:4Dr-l did not flow into this outcome measure. Heterogeneity among effect sizes within datasets was assessed using the  $I^2$  statistic.

We ran prespecified meta-regressions for the moderators study quality (i.e. the risk of bias in these studies), mean age of participants, and procedure of measuring 2D:4D (i.e., measurement by multiple independent raters, multiple times by one rater, once by one rater, or by the participants themselves). Thereby, the slope of the meta-regression line ( $\beta$  coefficient) indicated the strength of the association between the moderator and outcome. All meta-regressions were Bonferroni-corrected for multiple testing.

Moreover, we performed prespecified subgroup analyses to investigate the difference in the outcome measures between the right and the left hand, the different methods of measuring 2D:4D (i.e., with or without soft-tissue deformation) and the classification systems [i.e., DSM-5 vs. DSM-IV(-TR) and ICD-10 vs. DSM-IV(-TR)]. A comparison of DSM-5 versus ICD-10 was not possible due to small group sizes. "Measurement without soft-tissue deformation" comprised X-rays and direct measurement from the participants' palm, and "measurement with soft-tissue deformation" comprised photocopies and hand scans. It is a prominent finding in 2D:4D research that associations with target traits are more pronounced in the right than in the left hand e.g.<sup>55,56</sup>. Recent meta-analyses on 2D:4D and gender identity<sup>6,19</sup> report descriptive differences (insofar that the association is stronger with R2D:4D), but did not test for significant differences in these subgroups. We ran one post-hoc subgroup analysis examining differences between studies where the 2D:4D raters have been blinded vs. not blinded to the participants' gender identity.

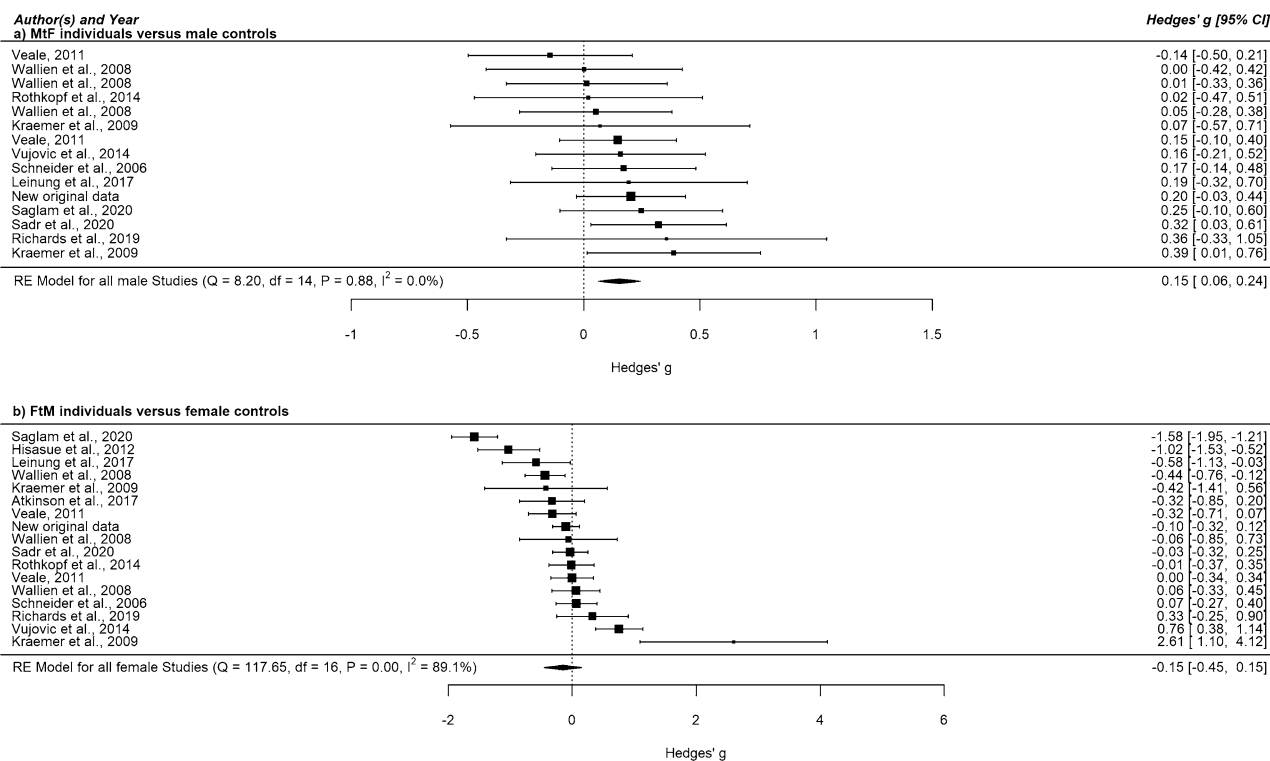
Small study effects were assessed by visual detection of asymmetries in contour-enhanced funnel plots and by the Egger regression test<sup>57</sup>. Following the authors' original proposition, we considered analyses to be biased if the intercept differed from zero at  $p = 0.10$ <sup>57</sup>. We evaluated the sensitivity of our analysis by comparing models with and without effects that we assume to be influential outliers<sup>58</sup>. On the one hand, we defined statistically influential outliers following the approach by Viechtbauer and Cheung<sup>58</sup>: a study may be considered to be influential if at least one of the following is true: (1) the absolute DFFITS value is larger than  $3\sqrt{p/(k-p)}$ , where  $p$  is the number of model coefficients, and  $k$  the number of studies. (2) The lower tail area of a Chi-squared distribution with  $p$  degrees of freedom cut off by the Cook's distance is larger than 50%. (3) The hat value is larger than  $3(p/k)$ . (4) Any DFBETAS value is larger than 1<sup>58</sup>. On the other hand, we considered studies to be influential outliers when the sample was substantially divergent to all other samples included. This is the case for one study with two child samples<sup>59</sup>. The literature is controversial as to whether symptoms of gender identity disorder in children persist into adulthood<sup>45,60,61</sup>, indicating that these samples potentially deviate from the rest of the studies included in our analysis.

$p < 0.05$  (2-sided) was considered statistically significant, except for the Egger test<sup>57</sup> as stated above.

**Results. Eligible studies.** The literature search is summarized in the PRISMA flow chart (Supplementary Figure S1). We identified 12 articles<sup>6,42–45,54,59,62–66</sup> comprising 16 independent samples. Additionally, we conducted one investigation by ourselves described in the first part of this manuscript (Study 1: new original data). This results in 13 eligible articles added up to 17 independent samples. The characteristics of all included studies are detailed in the Supplementary Table S6 (online supplement).

	MtF individuals versus male controls		FtM individuals versus female controls	
No. of subjects analyzed	1815		1859	
Hedges' g (95% CI)	0.153 (0.063; 0.243)		-0.152 (-0.453; 0.149)	
	Diagnosis made by a clinician		Diagnosis made by a clinician	
	Yes	No	Yes	No
No. of subjects analyzed	1206	609	1298	561
Hedges' g (95% CI)	0.193 (0.086; 0.300)	0.057 (-0.111; 0.224)	-0.166 (-0.664; 0.333)	-0.081 (-0.261; 0.098)

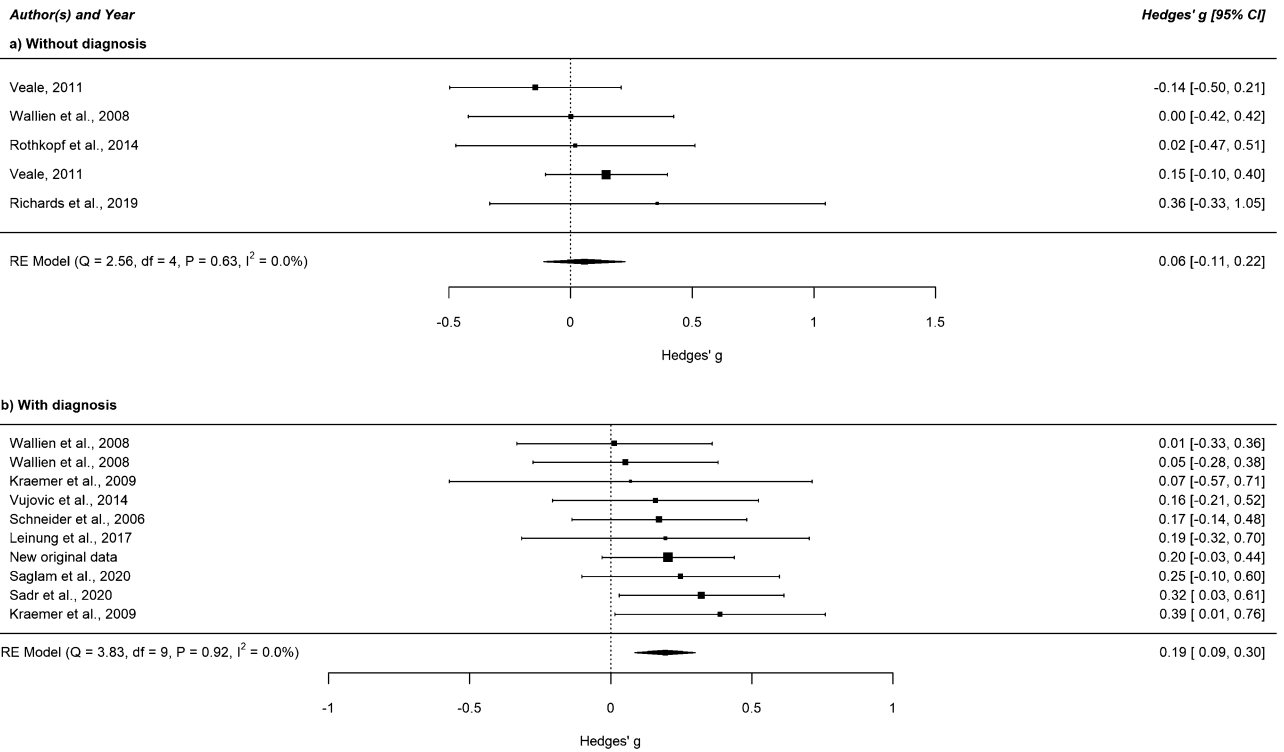
**Table 2.** Meta-analytic results: The standardized mean difference between transgender individuals and controls. *CI* confidence interval, *FtM* female-to-male transgender, *MtF* male-to-female transgender, *No* number.



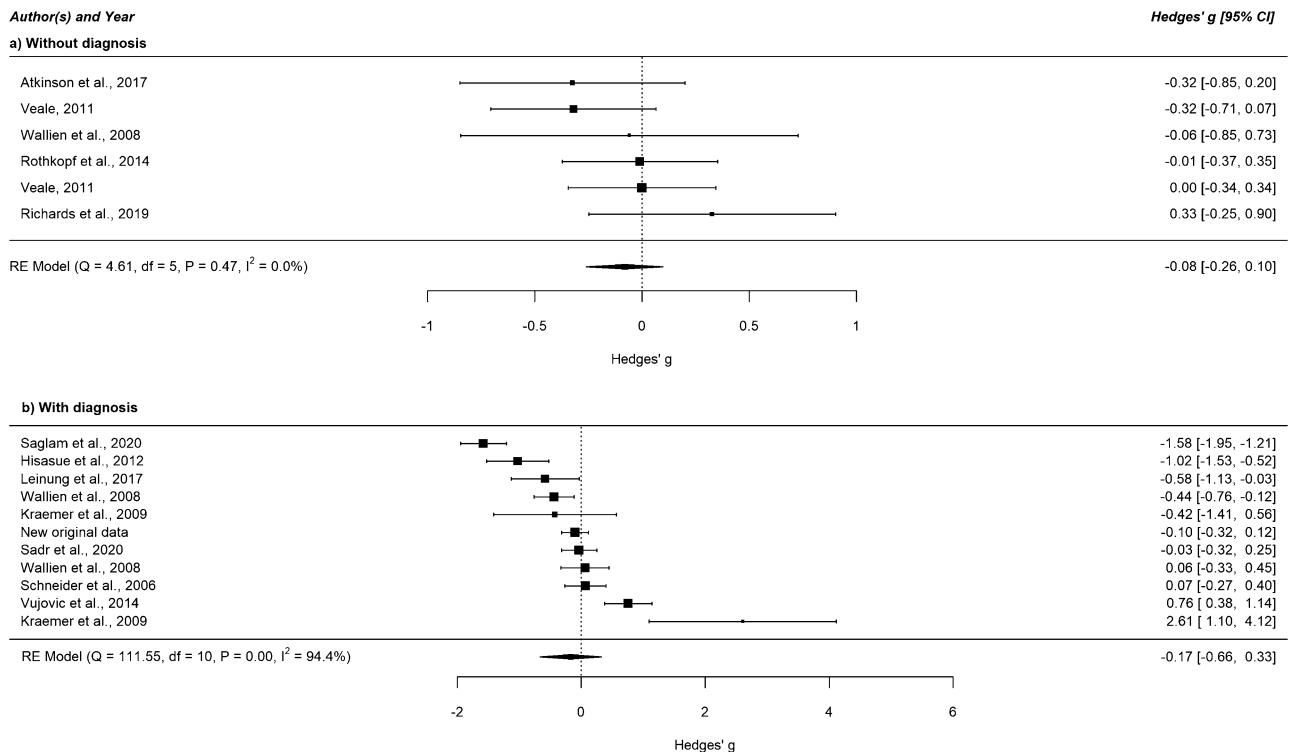
**Figure 1.** Forest plot of the standardized mean difference in 2D:4D among MtF individuals versus male controls and among FtM individuals versus female controls. This figure was created using R software<sup>51</sup> (v3.4.2; <https://www.R-project.org/>). *CI* confidence interval, *FtM* female-to-male transgender, *MtF* male-to-female transgender, *RE Model* restricted maximum likelihood estimation model.

**Meta-analytic results.** The meta-analytic results are shown in Table 2 and Figs. 1, 2 and 3. In line with our first hypothesis, we found that the aggregated 2D:4D was larger in MtF individuals than in male controls. This effect was even more pronounced in MtF individuals diagnosed by a clinician versus male controls, thus supporting our second hypothesis. We did not detect any significant meta-analytical differences between FtM individuals versus female controls, irrespective of a confirmed diagnosis. Concerning 2D:4Dr-l, none of the meta-analyses revealed any significant association with transgender identity (data not shown).

**Meta-regressions and subgroup analyses.** None of the meta-regression analyses concerning study quality, mean age, and the procedure of measuring 2D:4D showed any significant effects with a Bonferroni-corrected threshold of  $p = 0.017$ . In subgroup analyses, effect sizes did not differ between 2D:4D measured on the right or the left hand, 2D:4D measured with or without soft tissue deformation, diagnoses made according to DSM-IV(-TR), ICD-10, or DSM-5, and between studies where 2D:4D raters have been blinded or not to the subjects' case-control status. The exact figures are detailed in the Supplementary Table S7 (online supplement). As mentioned above ("Meta-analytic results" section), the subgroup analyses for our second hypothesis revealed significant differences.



**Figure 2.** Forest plot of the standardized mean difference in 2D:4D among MtF individuals versus male controls separated according to whether diagnoses have been made (or not) by a clinician. This figure was created using R software<sup>51</sup> (v3.4.2; <https://www.R-project.org/>). CI confidence interval, MtF male-to-female transgender, RE Model restricted maximum likelihood estimation model.



**Figure 3.** Forest plot of the standardized mean difference in 2D:4D among FtM individuals versus female controls separated according to whether diagnoses have been made (or not) by a clinician. This figure was created using R software<sup>51</sup> (v3.4.2; <https://www.R-project.org/>). CI confidence interval, FtM female-to-male transgender, RE Model restricted maximum likelihood estimation model.

**Small study effects and sensitivity analyses.** Concerning MtF individuals vs. male controls, the Egger test ( $z = -0.352$ ,  $p = 0.725$ ) and funnel plot (Supplementary Figure S2, online supplement) showed no evidence of small study effects. The funnel plot (Supplementary Figure S3, online supplement) and Egger test ( $z = 1.837$ ,  $p = 0.066$ ) regarding FtM individuals vs. female controls, however, revealed asymmetry in the published literature. Nevertheless, the applied trim and fill method<sup>67</sup> suggested to compensate for this possible bias by adding 0 studies (SE = 2.571). Sensitivity analyses revealed no statistically influential outliers. When excluding one study examining two samples of children<sup>59</sup>, the effect size for FtM individuals versus female controls did not change substantially, whereas the effect size for MtF individuals versus male controls increased [Hedges'  $g = 0.172$ ; 95% CI (0.076; 0.268)].

**Discussion.** In accordance with our hypotheses, we found associations of small effect size between feminized 2D:4D and MtF gender identity that were more pronounced in MtF individuals diagnosed by a specialized clinician. No significant results have been revealed in FtM individuals compared to female controls. Between-study heterogeneity was small for the MtF versus male sample ( $I^2 = 0\%$ ) but high for the FtM versus female comparison ( $I^2 = 89.1\%$ ), possibly resulting in slightly overestimated effect sizes in the latter sample. The high heterogeneity could not be explained by the pre-specified moderators since none of the meta-regression and subgroup analyses revealed any significant results. This contradicts the prominent 2D:4D findings concerning the influence of the measurement with or without soft tissue deformation on the relationship between 2D:4D and target traits<sup>10,11</sup> and of more pronounced effects in the right compared to the left hand due to possible increased sensitivity of the right hand to prenatal testosterone exposure<sup>44,68</sup>. We also did not find a significant association between 2D:4Dr-l and transgender identity, which suggests that 2D:4Dr-l is not as suitable as 2D:4D to indicate relationships in this research area. Sensitivity analysis showed that the influence of single studies on the overall effect should be estimated minor.

### General discussion

Both the original study and the meta-analysis support our hypothesis that MtF individuals have a higher (i.e. feminized) 2D:4D than male controls. The effect was amplified when diagnoses were made according to classification criteria of gender dysphoria or transsexualism instead of self-identifying as transgender. This indicates that in males a lower prenatal androgen load associates with more pronounced gender dysphoria. In terms of practical relevance, the small effect sizes of up to Hedges'  $g = 0.193$  mean that the 2D:4D of 58% of male controls is below the average 2D:4D of MtF individuals<sup>69</sup>. Evaluation of the practical relevance by comparing our results to other meta-analyses in the research field of 2D:4D shows that effect sizes of approximately standardized mean difference = 0.2 are common, e.g. in addictive disorders<sup>13</sup> or physical prowess<sup>18</sup>, but, rarely, higher effects exist as well (e.g. in sexual orientation of women<sup>16</sup> and autism spectrum disorders<sup>14</sup>). However, with a non-overlap of 14.7% between transgender individuals and controls, one has to bear in mind that the practical impact is small, too<sup>69</sup>.

Nevertheless, our analysis implicates that prenatal androgen exposure plays a role in the development of male gender identity. A possible underlying mechanism may be a discrepancy between the two pre-/perinatal testosterone secretion peaks in males: (1) the first between 10 and 20 weeks of gestation entailing bodily changes and (2) the second during the perinatal window entailing organizational effects in the brain<sup>43,70</sup>, which can result in a discrepancy of sex between body and brain<sup>71</sup>. The lack of the second testosterone secretion peak in female infants with rather constant testosterone concentrations during the first year of life<sup>70</sup> might explain why we did not detect any significant associations between 2D:4D and female gender identity. Considering anatomical features of the human brain, the BNST has been linked to gender identity<sup>4</sup>. An animal study examining mice<sup>3</sup> revealed that perinatal metabolites of testosterone masculinize sexually dimorphic cell survival in the BNST, thus suggesting a possible association of perinatal testosterone and gender identity.

A previous meta-analysis<sup>19</sup> reported significant associations of male transgender identity only with R2D:4D but not with L2D:4D, whereas a more recent meta-analysis<sup>6</sup> found significant differences between MtF individuals and male controls in both R2D:4D and L2D:4D. Our original investigation, however, revealed significant associations with only L2D:4D but not with M2D:4D or R2D:4D. Our meta-analysis, in turn, points to an association of male transgender identity with (aggregated) 2D:4D in general with subgroup analyses revealing no significant difference between the meta-analytic effect sizes of R2D:4D and L2D:4D. Following these results, we cannot confirm an increased sensitivity to prenatal testosterone exposure in the right hand as often postulated<sup>44,68</sup>.

In order to draw conclusions about the prenatal androgen load, 2D:4D is used as a proxy<sup>8</sup>, as explained in the "Introduction" section. Recently, criticism has emerged on the use of 2D:4D as a biomarker of prenatal testosterone<sup>72–75</sup>. This is—among other things—due to a lack of X-linked inheritance in classical genetic studies<sup>76,77</sup> or in genome-wide association studies (GWAS)<sup>78</sup>, even though the androgen receptor is located on the X chromosome. Some authors argue that if 2D:4D reflects prenatal testosterone effects and if androgen receptor variants moderate testosterone effects, androgen receptor variants should show systematic relationships with 2D:4D (for a more detailed explanation, see Hönekopp<sup>79</sup>). Additionally, meta-analyses do not confirm an association between 2D:4D and the effectiveness of androgen receptors (genetic CAG and GGC repeat length polymorphism)<sup>72,79</sup>. The largest GWAS on 2D:4D so far only shows a slight positive influence of CAG repeats in women, but not in men<sup>78</sup>. Thus, other mechanisms may also be involved in the development of 2D:4D besides sex hormones and genetics, for example prenatal corticosterone<sup>80</sup> and prenatal stress<sup>23,25,81</sup>.

**Strengths and limitations.** This analysis provides new case–control data from a comparably large sample size and a meta-analysis including a wide array of dimensions of transgender identity, which we incorporated in separate subgroup analyses. We ensure a standardized procedure for our meta-analysis by following the PRISMA statement entirely<sup>46,47</sup> and by assessing the risk of bias of the studies according to the Newcastle–Ottawa Quality



Assessment Scale for case–control and cohort studies<sup>48</sup>. For the sample of MtF individuals versus male controls, heterogeneity was small, indicating low inconsistency of studies' results. This includes variation mainly due to sampling error, i.e. to the variation in study outcomes between studies. Our sensitivity analysis suggests that the effect does not depend on any one single study and therefore, is reasonably robust.

There are some limitations to this study. First, criticism of the concept of 2D:4D as an appropriate proxy for prenatal androgen exposure has emerged as described above in the “General discussion” section, and it is presumable that other mechanisms beside sex hormones and genetics also influence the development of 2D:4D. Second, despite its large sample size, the original investigation is still underpowered and thus, could not detect effects comparable to those described in a previous meta-analysis<sup>19</sup>. This additionally entails that our sample is not fully suited for multiple testing and we, consequently, could not control for alpha inflation; hence impeding the explanatory power of our results. Third, in the meta-analysis of FtM individuals versus female controls, heterogeneity was high, indicating systematic variation not caused by the study-related sampling errors and possibly resulting in overestimated effect sizes for this sample. Fourth, there were some methodological difficulties, such as information and selection bias potentially inflating the case–control difference and impeding the generalizability of results, when applying the study design of case–control studies. This extends to the control self-selection via flyers in the original investigation, an approach in which motivational factors related to personality traits or lifestyle are likely to play a role<sup>82</sup>. Consequently, the control group does not reflect a population-based sample.

**Future research.** Following our meta-analysis, no recommendation can be made as to whether future 2D:4D examinations should be performed analyzing right or left hand 2D:4D or using methods with or without soft tissue deformation (for a definition, see Methods section). We recommend conducting original investigations in this field following a case–control design with reliable diagnoses (preferably made by clinicians) and computing a priori power analyses for determining an adequate sample size.

In summary, our results strengthen the assumption of a multifactorial etiology of gender identity and suggest that prenatal androgen levels are involved. Nevertheless, deeper analyses of other biological and non-biological mechanisms are required. We therefore recommend further studies on the developmental pathways of 2D:4D and transgender identity.

## Data availability

The datasets generated during and/or analyzed during the current study are not publicly available due to data protection limitations but are available from the corresponding author on reasonable request.

Received: 13 November 2019; Accepted: 2 September 2020

Published online: 09 November 2020

## References

- Capetillo-Ventura, N. C., Jalil-Pérez, S. I. & Motilla-Negrete, K. Gender dysphoria: an overview. *Med. Univ.* **17**, 53–58 (2015).
- Zucker, K. J. Epidemiology of gender dysphoria and transgender identity. *Sex. Health* **14**, 404–411 (2017).
- Hisasue, S., Seney, M., Immerman, E. & Forger, N. Control of cell number in the bed nucleus of the stria terminalis of mice: role of testosterone metabolites and estrogen receptor subtypes. *J. Sex. Med.* **7**, 1401–1409 (2010).
- Kruijver, F. P. M. *et al.* Male-to-female transsexuals have female neuron numbers in a limbic nucleus. *J. Clin. Endocrinol. Metab.* **85**, 2034–2041 (2000).
- Guillamon, A., Junque, C. & Gómez-Gil, E. A review of the status of brain structure research in transsexualism. *Arch. Sex. Behav.* **45**, 1615–1648 (2016).
- Sadr, M., Khorashad, B. S., Talaei, A., Fazeli, N. & Hönekopp, J. 2D:4D Suggests a role of prenatal testosterone in gender dysphoria. *Arch. Sex. Behav.* **49**, 421–432 (2020).
- Zheng, Z. & Cohn, M. J. Developmental basis of sexually dimorphic digit ratios. *Proc. Natl. Acad. Sci.* **108**, 16289–16294 (2011).
- Manning, J. T. *Digit Ratio: a Pointer to Fertility, Behaviour and Health* (Rutgers University Press, New Brunswick, 2002).
- Berenbaum, S. A., Bryk, K. K., Nowak, N., Quigley, C. A. & Moffat, S. Fingers as a marker of prenatal androgen exposure. *Endocrinology* **150**, 5119–5124 (2009).
- Hönekopp, J. & Watson, S. Meta-analysis of the relationship between digit-ratio 2D:4D and aggression. *Pers. Individ. Dif.* **51**, 381–386 (2010).
- Xu, Y. & Zheng, Y. The digit ratio (2D:4D) in China: a meta-analysis. *Am. J. Hum. Biol.* **27**, 304–309 (2015).
- Manning, J. T. Prenatal sex steroids and transgender identity: is there a link with digit ratio?. *Endocr. Pract.* **23**, 738–740 (2017).
- Siegmann, E. M., Bouna-Pyrrou, P., Lenz, B. & Kornhuber, J. Digit ratio (2D:4D) in relation to substance and computer use: a meta-analysis. *J. Neural Transm.* **126**, 623–636 (2019).
- Teatero, M. L. & Netley, C. A critical review of the research on the extreme male brain theory and digit ratio (2D:4D). *J. Autism Dev. Disord.* **43**, 2664–2676 (2013).
- Hönekopp, J. Digit Ratio 2D:4D in relation to autism spectrum disorders, empathizing, and systemizing: a quantitative review. *Autism Res.* **5**, 221–230 (2012).
- Grimbos, T., Dawood, K., Burriss, R. P., Zucker, K. J. & Puts, D. A. Sexual orientation and the second to fourth finger length ratio: a meta-analysis in men and women. *Behav. Neurosci.* **124**, 278–287 (2010).
- Bunevicius, A. The association of digit ratio (2D:4D) with cancer: a systematic review and meta-analysis. *Dis. Mark.* **2018**, 7698193 (2018).
- Hönekopp, J. & Schuster, M. A meta-analysis on 2D:4D and athletic prowess: substantial relationships but neither hand out-predicts the other. *Pers. Individ. Differ.* **48**, 4–10 (2010).
- Voracek, M., Kaden, A., Kossmeier, M., Pietschnig, J. & Tran, U. S. Meta-analysis shows associations of digit ratio (2D:4D) and transgender identity are small at best. *Endocr. Pract.* **24**, 386–390 (2018).
- Tektas, O. Y. *et al.* Digit ratio (2D:4D) and academic success as measured by achievement in the academic degree 'Habilitation'. *PLoS ONE* **14**, e0212167 (2019).
- Kornhuber, J. *et al.* Low digit ratio 2D:4D in alcohol dependent patients. *PLoS ONE* **6**, 4–7 (2011).
- Kornhuber, J. *et al.* Low 2D:4D values are associated with video game addiction. *PLoS ONE* **8**, 1–6 (2013).
- Lenz, B., Bouna-Pyrrou, P., Mühle, C. & Kornhuber, J. Low digit ratio (2D:4D) and late pubertal onset indicate prenatal hyperandrogenization in alcohol binge drinking. *Prog. Neuro Psychopharmacol. Biol. Psychiatry* **86**, 370–378 (2018).

24. Lenz, B. & Kornhuber, J. Cross-national gender variations of digit ratio (2D:4D) correlate with life expectancy, suicide rate, and other causes of death. *J. Neural Transm.* **125**, 239–246 (2018).
25. Lenz, B. *et al.* Prenatal and adult androgen activities in alcohol dependence. *Acta Psychiatr. Scand.* **136**, 96–107 (2017).
26. Lenz, B., Mühle, C. & Kornhuber, J. Lower digit ratio (2D:4D) in alcohol dependence: confirmation and exploratory analysis in a population-based study of young men. *Addict. Biol.* **25**, e12815 (2020).
27. Lenz, B. *et al.* Low digit ratio (2D:4D) in male suicide victims. *J. Neural Transm.* **123**, 1499–1503 (2016).
28. Lenz, B., Müller, C. P. & Kornhuber, J. Alcohol dependence in same-sex and opposite-sex twins. *J. Neural Transm.* **119**, 1561–1564 (2012).
29. Lenz, B. *et al.* Sex hormone activity in alcohol addiction: Integrating organizational and activation effects. *Prog. Neurobiol.* **96**, 136–163 (2012).
30. Buchholz, V. N., Mühle, C., Cohort Study on Substance Use Risk Factors, Kornhuber, J., Lenz, B. Lower digit ratio (2D:4D) indicative of excess prenatal androgen is associated with increased sociability and greater social capital. *Front Behav Neurosci.* **13**, 246. <https://doi.org/10.3389/fnbeh.2019.00246> (2019).
31. Lenz, B. *et al.* The androgen model of suicide completion. *Prog. Neurobiol.* **172**, 84–103 (2019).
32. Eichler, A. *et al.* Digit ratio (2D:4D) and behavioral symptoms in primary-school aged boys. *Early Hum. Dev.* **119**, 1–7 (2018).
33. Lenz, B. *et al.* Mindfulness-based stress reduction in pregnancy: an app-based programme to improve the health of mothers and children (MINDFUL/PMI study). *Geburtshilfe Frauenheilkd.* **78**, 1283–1291 (2018).
34. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10)* (1992).
35. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th edition) (DSM-5)* (2013).
36. Deogracias, J. J. *et al.* The gender identity/gender dysphoria questionnaire for adolescents and adults. *J. Sex. Res.* **44**, 370–379 (2007).
37. Schneider, C. *et al.* Measuring gender dysphoria: a multicenter examination and comparison of the Utrecht gender dysphoria scale and the gender identity/gender dysphoria questionnaire for adolescents and adults. *Arch. Sex. Behav.* **45**, 551–558 (2016).
38. Lawrence, A. A. Autogynephilia and the typology of male-to-female transsexualism: concepts and controversies. *Eur. Psychol.* **22**, 39–54 (2017).
39. Ribeiro, E., Neave, N., Morais, R. N. & Manning, J. T. Direct versus indirect measurement of digit ratio (2D:4D): a critical review of the literature and new data. *Evol. Psychol.* **14**, 1–8 (2016).
40. Manning, J. T., Fink, B., Neave, N. & Caswell, N. Photocopies yield lower digit ratios (2D:4D) than direct finger measurements. *Arch. Sex. Behav.* **34**, 329–333 (2005).
41. Caswell, N. & Manning, J. T. A comparison of finger 2D:4D by self-report direct measurement and experimenter measurement from photocopy: methodological issues. *Arch. Sex. Behav.* **38**, 143–148 (2009).
42. Leinung, M. & Wu, C. The biological basis of transgender identity: 2D:4D finger length ratios implicate a role for prenatal androgen activity. *Endocr. Pract.* **23**, 669–671 (2017).
43. Hisasue, S., Sasaki, S., Tsukamoto, T. & Horie, S. The relationship between second-to-fourth digit ratio and female gender identity. *J. Sex. Med.* **9**, 2903–2910 (2012).
44. Schneider, H. J., Pickel, J. & Stalla, G. K. Typical female 2nd–4th finger length (2D:4D) ratios in male-to-female transsexuals: possible implications for prenatal androgen exposure. *Psychoneuroendocrinology* **31**, 265–269 (2006).
45. Kraemer, B. *et al.* Finger length ratio (2D:4D) in adults with gender identity disorder. *Arch. Sex. Behav.* **38**, 359–363 (2009).
46. Liberati, A. *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *BMJ* **339**, b2700 (2009).
47. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **6**, e100097 (2010).
48. Wells, G. *et al.* The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. [https://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp), accessed 3rd November 2018
49. Herzog, R. *et al.* Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? A systematic review. *BMC Public Health* **13**, 1–17 (2013).
50. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
51. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2017). <https://www.R-project.org/>, accessed 4th January 2019
52. Docter, R. F. & Fleming, J. S. Dimensions of transvestism and transsexualism. *J. Psychol. Human Sex.* **5**, 15–38 (1993).
53. Borenstein, M., Hedges, L., Higgins, J. & Rothstein, H. *Introduction to Meta-Analysis* (Wiley, Chichester, 2009).
54. Veale, J. F. Biological and psychosocial correlates of gender-variant and gender-typical identities [dissertation]. Massey University, Albany, New Zealand (2011).
55. Smedley, K. D., McKain, K. J. & McKain, D. N. 2D:4D digit ratio predicts depression severity for females but not for males. *Pers. Individ. Dif.* **70**, 136–139 (2014).
56. Williams, T. J. *et al.* Finger-length ratios and sexual orientation. *Nature* **404**, 455–456 (2000).
57. Egger, M., Smith, G. D., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *Br. Med. J.* **315**, 629–634 (1997).
58. Viechtbauer, W. & Cheung, M. W. Outlier and influence diagnostics for meta-analysis. *Res. Synth. Methods* **1**, 112–125 (2010).
59. Wallien, M. S. C., Zucker, K. J., Steensma, T. D. & Cohen-Kettenis, P. T. 2D:4D finger-length ratios in children and adults with gender identity disorder. *Horm. Behav.* **54**, 450–454 (2008).
60. Janssen, A. & Busa, S. Gender dysphoria in childhood and adolescence. In *David I Driver* (ed. Thomas, S.) 1–10 (Complex Disorders in Pediatric Psychiatry. Elsevier, Amsterdam, 2018); <https://doi.org/10.1016/b978-0-323-51147-6.00001-6>.
61. Cohen-Kettenis, P. T. & Gooren, L. J. G. Transsexualism: a review of etiology, diagnosis and treatment. *J. Psychosom. Res.* **46**, 315–333 (1999).
62. Saglam, T., Bakay, H., Gökler, M. & Turan, S. 2D:4D finger length ratios in individuals with gender dysphoria. *Turk. J. Psychiatry* (2020) ((in press)).
63. Richards, G., Wei, Y. & Hendriks, O. Self-measured digit ratio (2D:4D) and gender variance. *Endocr. Pract.* (2020) ((in press)).
64. Atkinson, B. M., Smulders, T. V. & Wallenberg, J. C. An endocrine basis for tomboy identity: The second-to-fourth digit ratio (2D:4D) in “tomboys”. *Psychoneuroendocrinology* **79**, 9–12 (2017).
65. Rothkopf, I. & Turgeon, S. M. Sex differences in college students' free drawings and their relationship to 2D:4D ratio and recalled childhood play behavior. *Arch. Sex. Behav.* **43**, 311–318 (2014).
66. Vujović, S. *et al.* Finger length ratios in Serbian transsexuals. *Sci. World J.* **2014**, 763563 (2014).
67. Duval, S. & Tweedie, R. Trim and fill: a simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56**, 455–463 (2000).
68. Manning, J. T., Scutt, D., Wilson, J. & Lewis-Jones, D. I. The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Hum. Reprod.* **13**, 3000–3004 (1998).
69. Kim, H. Statistical notes for clinical researchers: effect size Open lecture on statistics. *Restor. Dent. Endod.* **7658**, 328–331 (2015).
70. Forest, M., Sizonenko, P., Cathiard, A. & Bertrand, J. Hypophyso-gonadal function in humans during the first year of life. *J. Clin. Invest.* **53**, 819–828 (1974).

71. Sakuma, Y. Gonadal steroid action and brain sex differentiation in the rat. *J. Neuroendocrinol.* **21**, 410–414 (2009).
72. Voracek, M. No effects of androgen receptor gene CAG and GGC repeat polymorphisms on digit ratio (2D:4D): a comprehensive meta-analysis and critical evaluation of research. *Evol. Hum. Behav.* **35**, 430–437 (2014).
73. van Hemmen, J., Cohen-Kettenis, P. T., Steensma, T. D., Veltman, D. & Bakker, J. Do sex differences in CEOAEs and 2D:4D ratios reflect androgen exposure? A study in women with complete androgen insensitivity syndrome. *Biol. Sex Differ.* **8**, 11 (2017).
74. Yeung, L. & Tse, W. Why does digit ratio research fail to give any implication regarding the organizational effect of prenatal androgen? *J. Individ. Differ.* **38**, 36–45 (2017).
75. Leslie, M. The mismeasure of hands?. *Science* **364**, 923–925 (2019).
76. Ramesh, A. & Murty, J. Variation and inheritance of relative length of index finger in man. *Annu. Hum. Biol.* **4**, 479–484 (1977).
77. Voracek, M. & Dressler, S. G. Brief communication: familial resemblance in digit ratio (2D:4D). *Am. J. Phys. Anthropol.* **140**, 376–380 (2009).
78. Warrington, N. *et al.* Genome-wide association study identifies nine novel loci for 2D:4D finger ratio, a putative retrospective biomarker of testosterone exposure in utero. *Hum. Mol. Genet.* **27**, 2025–2038 (2018).
79. Hönekopp, J. No evidence that 2D:4D is related to the number of CAG repeats in the androgen receptor gene. *Front. Endocrinol. (Lausanne)* **4**, 185 (2013).
80. Lilley, T., Laaksonen, T., Huitu, O. & Helle, S. 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*). *Physiol. Behav.* **99**, 433–437 (2010).
81. Rizwan, S., Manning, J. & Brabin, B. Maternal smoking during pregnancy and possible effects of in utero testosterone: evidence from the 2D:4D finger length ratio. *Early Hum. Dev.* **83**, 87–90 (2007).
82. Jablensky, A. Research methods in psychiatric epidemiology: an overview. *Aust. N. Z. J. Psychiatry* **36**, 297–310 (2002).

## Acknowledgements

We thank the participants for their support of the research project and Sina Therese Kirsten, Terezie Sedlinská, and Colin Rentsch for measuring index and ring finger lengths (Study 1) and for assisting with data extraction (Study 2).

## Author contributions

Conception: B.L., C.M., E.S., J.K. Acquisition: B.L., C.M., I.D., T.M. Analysis: B.L., E.S. Interpretation: B.L., E.S., J.K. Drafting of the manuscript: B.L., E.S., J.K. All authors reviewed the manuscript. The present work was conducted in partial fulfillment of the requirements for obtaining the degree “Dr. med.” (Tobias Müller, Isabelle Dziadeck).

## Funding

Open Access funding provided by Projekt DEAL. This scientific research was promoted by the STAEDTLER Foundation (Staedtler Stiftung) and the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung; IMAC-Mind project: Improving Mental Health and Reducing Addiction in Childhood and Adolescence through Mindfulness: Mechanisms, Prevention and Treatment; 2018–2022; 01GL1745C). C.M. is an associated fellow of the research training group 2162 “Neurodevelopment and Vulnerability of the Central Nervous System” of the Deutsche Forschungsgemeinschaft (DFG; GRK2162/1). The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-72486-6>.

**Correspondence** and requests for materials should be addressed to E.-M.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020