

Digit Ratios (2D:4D) Determined by Computer-Assisted Analysis are More Reliable than Those Using Physical Measurements, Photocopies, and Printed Scans

HEATHER C. ALLAWAY, TERRI G. BLOSKI, ROGER A. PIERSON, and MARLA E. LUJAN*

Department of Obstetrics, Gynecology and Reproductive Sciences, College of Medicine,
University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W8

Abstract

Prenatal androgens influence the second to fourth digit ratio (2D:4D) of hands with men having lower ratios than women. Numerous methods are used to assess 2D:4D including, physical measurements with calipers, and measurements made from photocopies, scanned images, digital photographs, radiographs, and scaled tubes. Although each method appears relatively reliable, agreement upon a gold standard is necessary to better explore the putative effects of prenatal androgens. Our objective was to assess the level of intra and interobserver reliability when evaluating 2D:4D using four techniques: (1) physical measurements, (2) photocopies, (3) printed scanned images, and (4) computer-assisted image analysis. Physical measurements, photocopies, and printed scanned images were measured with Vernier calipers. Scanned images were also measured with computer-based calipers. Measurements were made in 30 men and 30 women at two different time points, by three experienced observers. Intraclass correlation coefficients were used to assess the level of reliability. Intraobserver reliability was best for computer-assisted (0.957), followed by photocopies (0.939), physical measurements (0.925), and printed scans (0.842; $P = 0.015$). Interobserver reliability was also greatest for computer-assisted (0.892), followed by photocopies (0.858), physical measurements (0.795), and printed scans (0.761; $P = 0.001$). Mean 2D:4D from physical measurements were higher than all other techniques ($P < 0.0001$). Digit ratios determined from computer-assisted, physical measurements, and printed scans were more reliable in men than women ($P = 0.009$, $P = 0.017$, and $P = 0.012$, respectively). In summary, 2D:4D determined from computer-assisted analysis yielded the most accurate and consistent measurements among observers. Investigations of 2D:4D should use computer-assisted measurements over alternate methods whenever possible.

Prenatal androgen exposure influences development of the fingers and leads to distinct differences in hand patterns among men and women. In general, the ratio between the second (2D = index finger) and fourth (4D = ring finger) digits in men is lower than that observed in women (Lutchmaya et al., 2004; Manning et al., 1998). Anatomical evidence of prenatal androgen exposure has gained significance in recent times because there is growing evidence in animal models that adult traits and behaviors including, disease susceptibility may be programmed during fetal development (Abbott et al., 2006; Manikkam et al., 2004;

*Correspondence to: Marla E. Lujan, Obstetrics, Department of Gynecology, and Reproductive Sciences, Royal University Hospital, 103 Hospital Drive, Room 4519, Saskatoon, Saskatchewan, Canada S7N 0W8. lujan@erato.usask.ca.

Manning and Bundred, 2000; Robinson, 2006). Previous studies in humans have shown low digit ratios (2D:4D) to be associated with a variety of traits including, enhanced visual spatial ability (Burton et al., 2005; van Anders and Hampson, 2005), increased athleticism (Honekopp et al., 2006a), female homosexuality (Brown et al., 2002; Hall and Love, 2003; Kraemer et al., 2006; Williams et al., 2000), decreased susceptibility to coronary artery disease (Fink et al., 2003; Manning and Bundred, 2000), and greater fertility potential in men (Manning et al., 1998) but lower fertility potential in women (Cattrall et al., 2005). By contrast, high 2D:4D were shown to be associated with high verbal fluency (Burton et al., 2005), increased emotional behavior (Williams et al., 2003), lower waist-hip ratios (Fink et al., 2003), male homosexuality (Lippa, 2003) and an increased risk of breast cancer (Manning and Bundred, 2000). Unfortunately, these associations with 2D:4D have been dampened by the enormous amount of nonreplication that has been reported among studies investigating similar study populations (e.g. Male and female homosexuality—Lippa, 2003; Robinson and Manning, 2000; Williams et al., 2000; Polycystic ovary syndrome—Bloski et al., 2008; Cattrall et al., 2005; Aggression—Bailey and Hurd, 2005; Hampson et al., 2008).

Studies designed to test the notion that prenatal androgens program certain traits and phenotypes require a standardized, validated method of measuring fingers and determining 2D:4D (Manning et al., 2005). This is particularly important as effect sizes on 2D:4D are generally only moderate to low (Kemper and Schwerdtfeger, in press). At present, several methods have been used to assess 2D:4D including, physical measurements with calipers (Scutt and Manning, 1996), measurements from photocopies (Manning et al., 2005), scanned images (Bailey and Hurd, 2005), digital photographs (Pokrywka et al., 2005), radiographs (Paul et al., 2006; Robertson et al., 2007), inked handprints (Hall and Love, 2003), and scaled tubes (Nicholls et al., 2008). Each of these methods has limitations relating to feasibility and costs but in general, physical measurements, photocopies and scanned images are the most commonly used techniques. Physical measurements show a high degree of repeatability (Scutt and Manning, 1996), but are associated with increased sampling time for participants and measurements can be difficult to obtain because of movement of volunteers' hands (Caswell and Manning, 2007). Printed scanned images and photocopies have the benefit of providing permanent records and shorter sampling times (Caswell and Manning, 2007; Manning et al., 2005); however, lower 2D:4D have been observed with photocopies and the equipment is not generally portable (Burriss et al., 2007; Manning et al., 2005). Digital image technology for the acquisition and measurements of finger lengths may provide a superior alternative to these more traditional techniques because: (1) large permanent databases can be generated requiring little physical space; (2) computer-assisted image enhancement can improve clarity of finger borders and creases which may improve measurement reliability and (3) digital technology may also be considered a more environmentally conscious alternative.

While each of the commonly used methods of acquiring 2D:4D is generally validated within studies to show they yield highly repeatable measurements among a single observer, it must be acknowledged that we know very little about the reliability of 2D:4D measurements made by multiple observers using a single technique (Voracek et al., 2007) or the reliability of measurements made by the same observers using multiple techniques (Kemper and Schwerdtfeger, in press; Manning et al., 2005). The few studies to date that have compared

reliability among measuring techniques have shown notable differences in absolute 2D:4D obtained from multiple techniques (Kemper and Schwerdtfeger, in press; Manning et al., 2005). These findings may be interpreted to mean that 2D:4D are technique dependent and this has serious implications for a field that necessitates a highly precise tool to identify relatively small differences in 2D:4D among study populations (Kemper and Schwerdtfeger, in press). Moreover, it suggests that it may be inappropriate to compare studies using different techniques to acquire finger lengths because even subtle differences in technique (e.g., using a ruler versus a Vernier caliper on a photocopied image) may have the potential to yield divergent findings among studies (Kemper and Schwerdtfeger, in press).

The objective of this study was to assess the level of intra and interobserver reliability associated with evaluating 2D:4D by three experienced observers using three caliper-based techniques (i.e., physical measurements, photocopied and printed scanned images) and one computer-based technique (i.e., computer-assisted image analysis). The effect of the participant's gender and hand being measured on observer agreement was also evaluated. Reliability was defined as the ability of one observer to be consistent in obtaining 2D:4D in a single participant and the ability of multiple observers to obtain the same digit ratio in a single participant. We hypothesized that computer-assisted image analysis would provide more reliable measurements compared to other commonly used techniques given that superior image quality could be generated by judicious manipulation of image settings.

METHODS

Study subjects

Thirty men and thirty women were invited to participate in this study. Participants with a history of injury or illness affecting the hands or fingers were excluded. The purpose of the study was explained and volunteers were given an opportunity to ask questions prior to participating. This study was approved by the Institutional Review Board at the University of Saskatchewan.

Data collection

Volunteers were asked to remove any jewelry or rings that would interfere with obtaining finger length measurements. Lengths of the second and fourth fingers were measured by three experienced observers using four separate techniques. Measurements were made on the ventral surface of the hand by placing the bottom tip of a Vernier caliper (accurate to 0.01 mm) midline of the basal crease of the finger and extending to the tip without exerting pressure as described by Manning et al. (1998). The left index finger was measured first, followed by the left ring, right index, and then right ring finger. Each volunteer's right and left hand were then scanned using a Hewlett Packard ScanJet 3C scanner (100 dpi; Hewlett-Packard Company, Greeley, CO, USA) and stored in a custom-designed database until processing. Participants placed their relaxed hands lightly on the surface of the scanner with the second to fifth fingers held parallel and the tip of the middle finger aligned with the wrist and elbow. Photocopies of both hands were then taken in a similar manner with a Xerox Copycenter C35 (600 dpi; Xerox Corporation, Rochester, New York, NY, USA). Crumpled tinfoil was placed over the hands to enhance image contrast as described by Voracek et al.

(2007). Care was taken to ensure the proximal crease and finger tip borders were well visualized on the photocopies by adjusting brightness. After scanning and photocopying the hands, physical measurements were repeated in the same manner by each observer. The time lapse between physical finger length measurements varied from 10 to 15 min.

Scanned hand images were scaled and printed by a Hewlett Packard LaserJet 5M (600 dpi; Hewlett-Packard Company, Boise, ID, USA) one week later. Each observer measured finger lengths from the printed image by placing the bottom tip of the Vernier calipers on the basal crease of the left index finger and extending to the tip of the finger. Measurements were repeated for the left ring, right index, and right ring finger. A similar method was used to measure finger length on photocopied images. Scanned hand images were also analyzed using computer-assisted image analysis (GNU Image Manipulation Program, GIMP, Version 2.217). Each observer could alter sharpness to aid in visualization of the proximal crease and finger tip border. Because image manipulation runs the risk of distortion and overenhancement of images, observers were advised to only adjust the contrast and sharpness of images in which they felt the basal crease and/or finger tip were difficult to visualize. Mouse controlled calipers were placed on the basal crease of the left index finger and extending to the tip of the finger. Measurements were repeated for the left ring, right index, and right ring finger. As software-based calipers perform linear measurements based on a pixel per inch platform, images were intentionally scanned at 100 dpi to avoid calibrating the software-based calipers to an imaged ruler.

Finger length measurements were repeated in the manner described above from the photocopied, printed scanned images, and computer image analysis the next day.

Statistics

Statistical analyses were performed using Statistical Analysis Software (SAS, Version 9.1, Cary, NC, USA), Statistical Package for the Social Sciences (SPSS, Version 15.0, Chicago, IL, USA), and GraphPad InStat (Version 3, San Diego, CA, USA). Initial data analyses were performed using a SAS mixed model ANOVA with a factorial treatment arrangement. Descriptive statistics (mean \pm standard deviation) were calculated for 2D:4D determined by the four techniques. Differences in mean 2D:4D yielded by the various techniques were determined by Tukey's Multiple Comparison Test on the entire data set and following stratification of the data by sex. Intra and interobserver agreement, for all physical measurements, photocopy, and scan techniques were assessed by pair-wise, two-way random intraclass correlation coefficient (ICC) with absolute agreement. Single measures ICC were reported. Differences in reliability among the four techniques were determined by performing Tukey's Multiple Comparison Tests on the tabulated ICC values. Differences in interobserver reliability on data stratified by judge and technique were assessed by Tukey's Multiple Comparison Tests while *t*-tests were used to determine significant differences in reliability on data stratified by hand measured and gender of the participant. $P < 0.05$ was considered statistically significant.

RESULTS

Mean digit ratios

Mean 2D:4D tabulated by the three observers in all participants using the four techniques were not different (Observer 1 = 0.965 ± 0.032 , Observer 2 = 0.965 ± 0.033 , and Observer 3 = 0.967 ± 0.032 , $P = 0.1413$). Table 1 presents mean 2D:4D derived from physical measurements, photocopies, printed scanned images, and computer assisted image analysis. Mean 2D:4D derived from photocopies, printed scanned images, and computer-assisted image analysis were similar while physical measurements were larger than the other three techniques ($P < 0.0001$). Physical measurements yielded higher mean 2D:4D than photocopies, printed scanned images, and computer-assisted image in both males ($P < 0.0001$) and females ($P < 0.0001$). Overall, mean 2D:4D in women (0.974 ± 0.028) were consistently larger than 2D:4D in men (0.958 ± 0.033 ; $P < 0.0001$) and this difference between sexes was consistent for all measurement techniques. Left hand 2D:4D (0.967 ± 0.032), irrespective of the gender of the participant and measurement technique, were larger than right hand 2D:4D (0.964 ± 0.032 ; $P = 0.0037$).

Intraobserver agreement

The level of intraobserver reliability among 2D:4D determined by physical measurements, photocopies, printed scanned images and computer-assisted image analysis are summarized in Table 2. Overall, each observer demonstrated a similar level of intraobserver reliability when measuring 2D:4D by the four techniques (Mean intraobserver reliability for Observer 1 = 0.885, Observer 2 = 0.921 and Observer 3 = 0.943, $P = 0.3530$). Intraobserver reliability was best for computer-assisted image analysis, followed by photocopies, physical measurements, and printed scanned images ($P = 0.0150$). Intraobserver reliability of 2D:4D obtained by physical measurements were similar to image analysis and photocopied images. However, measurements from image analysis and photocopy technology yielded more reliable 2D:4D than those generated from printed scanned images ($P = 0.0480$).

Interobserver agreement

Table 3 summarizes the level of interobserver reliability when determining 2D:4D by each technique. All pairs of observers demonstrated similar overall reliability when measuring 2D:4D by the four techniques (Mean interobserver reliability for Observers 1 and 2 = 0.840, Observers 1 and 3 = 0.841 and Observers 2 and 3 = 0.799, $P = 0.5510$). Interobserver agreement was best for computer-assisted image analysis, followed by photocopies, physical measurements, and printed scanned images ($P = 0.0014$). Interobserver reliability for 2D:4D obtained from photocopies was similar to that obtained from physical measurements and image analysis. However, both image analysis and photocopies yielded more reliable 2D:4D compared to those generated from printed scanned images ($P = 0.0100$).

The level of interobserver reliability associated with determining 2D:4D in males and females is compared in Table 4. Digit ratios measured in males using physical measurements, printed scanned images, and computer-assisted image analysis were more reliable than those measured in females ($P = 0.0173$, $P = 0.0118$, and $P = 0.0091$, respectively). There was also a tendency toward lower reliability in 2D:4D measured from

photocopies in women but this did not reach statistical significance ($P=0.0662$). No differences in reliability were detected with any of the four techniques when measuring 2D:4D on either the right or left hands (Table 5).

DISCUSSION

The present study was designed to evaluate the level of intra and interobserver reliability associated with determining 2D:4D using three commonly used methods (i.e., physical measurements, photocopies, and printed scanned images) and one emerging technique (i.e., computer-assisted image analysis). In general, 2D:4D obtained from computer-assisted image analysis consistently demonstrated the highest levels of intra and interobserver reliability; whereas, measurements taken from printed scanned images were associated with the lowest levels of reliability. Digit ratios obtained in women were consistently associated with lower levels of reliability using each of the four techniques whereas, measurements made in both the left and right hands showed equal reliability.

Our study supported the hypothesis that computer-assisted image analysis provided the most reliable method of determining 2D:4D. Other investigators have used image analysis to determine 2D:4D from images obtained from flatbed scanners (Bailey and Hurd, 2005; de Bruin et al., 2006; McFadden and Shubel, 2002), radiographs (McIntyre et al., 2005, 2006), and digital photographs (Honekopp et al., 2006a; Pokrywka et al., 2005). Bailey and Hurd (2005) showed good intraobserver reliability when a single observer measured finger lengths in 10 subjects using computer-assisted image analysis on hand scans while McFadden and Shubel (2002) showed good interobserver reliability when three observers measured finger lengths using similar computer-based calipers on scanned images of the hand. Investigations using digital photographs of hands pressed against a glass plate also showed good intra and interobserver reliability when using computer programs to measure finger lengths (Honekopp et al., 2006b; Pokrywka et al., 2005). However, it should be noted that the aforementioned studies only performed these small reliability tests to internally validate their observers' ability to measure finger lengths. This is unlike the current study whose primary objective was to evaluate and compare intra and interobserver reliability when using computer-assisted measurements and other commonly used techniques.

Digit ratios obtained by computer-assisted analysis were similar in reliability to those obtained from photocopies but were more reliable than those obtained from printed scans. Although reliability coefficients were consistently higher when image analysis was used, they were not significantly better than measurements made with Vernier calipers on photocopied images. Comparable levels of intra and interobserver reliability when using computer-guided methods and photocopies, corroborates the observation that photocopy technology generates accurate and clear images from which to measure finger lengths (Manning et al., 2005; Robinson and Manning, 2000; Voracek et al., 2005, 2007). However, despite close reliability, computer-assisted analysis has several advantages over photocopy technology. For example, it was our experience that acquiring hand scans was associated with shorter sampling times which had important implications for recruiting and retaining study subjects. Storage of digital images required less physical storage space and image files, if properly managed, will allow for the development of large databases with longer life

spans. Moreover, improving image quality in digital scans did not involve an excessive and unnecessary use of paper and toner, providing an environmentally sensitive alternative. The ability to scan hands into a computer also raises the issue of exciting future applications of image analysis including, semiautomated and automated determination of finger length measurements.

We elected to print digital images and measure finger lengths using Vernier calipers because others have used this technique assuming this approach to be equivalent to that of photocopies (Voracek et al., 2007; Williams et al., 2003). However, we are unable to locate any study that supported similar levels of interobserver agreement when determining 2D:4D from photocopied and printed hand scans. Although the level of agreement we reported for 2D:4D derived from printed scans was similar to findings made by Voracek et al. (2007), the reliability coefficients were lower than those observed for photocopies and computer-assisted image analysis. In our current study, we believe printed scanned images yielded less reliable 2D:4D because image quality was significantly lost at the time of printing—this occurred despite images being printed at the highest quality available to us (i.e., 600 dpi). Scanned images when projected onto a computer screen were far clearer than the printed images and as such, allowed for easier discernment of digit landmarks.

We consistently reported lower levels of interobserver agreement when measuring 2D:4D in women using each of the four techniques. Differences in reliability when measuring fingers in men and women were unexpected. However, we hypothesize that several factors may have contributed to differences in 2D:4D reliability between genders. Many women enrolled in this study had significantly reduced visibility of basal creases owing to changes in the skin that can occur following long-term ring use. Fingertip borders were also difficult to visualize in hand images of women who had long fingernails. The fact that we noted the highest levels of reliability among female 2D:4D when image settings could be manipulated to better view the hand is consistent with this hypothesis (i.e., photocopies and computer-assisted analysis). Observers also reported more difficulty in measuring index and ring fingers when they perceived there to be negligible differences between the lengths of these two fingers which was often the case among females.

Differences in reliability when measuring 2D:4D in men and women may have implications for discordant findings reported in the literature. For example, prenatal androgen exposure is believed to increase the propensity for aggressive behaviors among species (reviewed in Ryan and Vandenbergh, 2002); yet evidence in humans is sparse and confounding (reviewed in Cohen-Bendahan et al., 2005). Recently, Hampson et al. (2008) reported that lower 2D:4D were associated with increased aggressiveness in men and women. However, these findings were in contrast to a previous report by Bailey and Hurd (2005) that did not conclude a similar relationship when assessing aggressive traits in women using the same standardized questionnaire. That reproducibility in 2D:4D findings tends to be more apparent in males rather than females is also supported by studies in which aspects of personality such as sensation seeking (Hampson et al., 2008; Fink et al., 2006) were shown to correlate with ratios consistently in men but not consistently in women, and by studies that consistently reported no association among 2D:4D with agreeableness and openness in men but showed conflicting associations with 2D:4D in women (Fink et al., 2004; Lippa,

2006). Dewitte and coworker have proposed that context might help to explain this discordance since situational cues were shown to moderate the relationship between 2D:4D and aggression scores, economic decision-making and prosocial “fairness” decision-making in participants subjected to aggressive music videos (Millet and Dewitte, 2007; Millet and Dewitte, 2009), aggressive language tests (Millet and Dewitte, 2009) and sex-related cues (Van den Bergh and Dewitte, 2006). In instances where relationships between personality traits, situational cues, and 2D:4D were readily correlated in men but not in women, our study would suggest that less reliable measurements in women might have also interfered with the accuracy of the measurement. Our observation that reliability was improved when image quality was optimized may be interpreted to mean that future studies involving 2D:4D in women would have a higher likelihood of replication if computer-assisted analysis is used over alternate methods.

Consistent with an enormous body of evidence (reviewed in McIntyre, 2006), male participants had lower mean 2D:4D compared to females. While sexual dimorphism in 2D:4D was detected using all four techniques, it was noted that 2D:4D measured from photocopies, printed scanned images, and computer-assisted measurements were associated with proportionately lower ratios than physical measurements made in both sexes. The observation that physical measurements yield higher mean 2D:4D than photocopies supports the findings of Manning et al. (2005) and Burriss et al. (2007), but contrasts the findings of Voracek and Dressler (2006). This is the first report that printed scans and computer-assisted measurements yield lower 2D:4D than physical measurements, although a trend for lower 2D:4D was noted for computer-assisted measurements by Burriss et al. (2007). Others have previously ascribed lower ratios obtained from photocopies to differences in fat-pads at the fingertips (Manning et al., 2005) because pressing the palm against a hard, flat surface presumably distorts distribution of fat at the finger tips. Observations of lower mean 2D:4D for printed scans and computer-assisted measurements was not unexpected because the palm of the hand was pressed against a hard, flat surface. Manning et al. (2005) have speculated that this deformation of fat at the finger tips alters specular reflection giving rise to an artificial lengthening of the fingers in the two-dimensional image generated. Taking into account an independent measurement of the fingertip curvature may potentially correct for this artificial lengthening (Manning et al., 2005). However, we are unaware of calculations that support this notion.

In conclusion, 2D:4D determined from computer-assisted image analysis and photocopies yielded the most reliable measurements among observers. We recommend investigations reporting 2D:4D to use computer-assisted analysis of hand scans whenever possible given numerous benefits of digital technology over alternate methods and since absolute levels of reliability were highest in women when 2D:4D were obtained using this technology. Enhancement of scanned images must be performed judiciously to avoid the loss of image components and/or the creation of image artifacts at the time of evaluation. Because 2D:4D and reliability in obtaining these values vary among measurement techniques, we recommend that care be taken when extrapolating findings among studies using different methods to measure finger lengths.

Acknowledgments

Contract grant sponsors: Canadian Institutes of Health Research (CIHR), Saskatchewan Health Research Foundation, Strategic Training Initiative in Research in Reproductive Health Sciences (STIRRHs)

LITERATURE CITED

- Abbott DH, Padmanabhan V, Dumesic DA. Contributions of androgen and estrogen to fetal programming of ovarian dysfunction. *Reprod Biol Endocrinol*. 2006; 4:17. [PubMed: 16606451]
- Bailey AA, Hurd PL. Finger length ratio (2D:4D) correlates with physical aggression in men but not in women. *Biol Psychol*. 2005; 68:215–222. [PubMed: 15620791]
- Bloski, TG., Lujan, ME., Chizen, DR., Pierson, RA. Digit ratios as anatomical evidence of prenatal androgen exposure in women with polycystic ovary syndrome. Proceedings of the 54th Annual Meeting of the Canadian Fertility and Andrology Society; Calgary, Alberta, Canada. November 26–29; 2008. p. 65[Abstract]
- Brown WM, Finn CJ, Cooke BM, Breedlove SM. Differences in finger length ratios between self-identified “butch” and “femme” lesbians. *Arch Sex Behav*. 2002; 31:123–127. [PubMed: 11910785]
- Burriss RP, Little AC, Nelson EC. 2D:4D and sexually dimorphic facial characteristics. *Arch Sex Behav*. 2007; 36:377–384. [PubMed: 17203400]
- Burton LA, Henninger D, Hafetz J. Gender differences in relations of mental rotation, verbal fluency, and SAT scores to finger length ratios as hormonal indexes. *Dev Neuropsychol*. 2005; 28:493–505. [PubMed: 15992253]
- Caswell N, Manning JT. A comparison of finger 2D:4D by self-report direct measurement and experimenter measurement from photocopy: methodological issues. *Arch Sex Behav*. 2009; 38:143–148. [PubMed: 17653838]
- Cattrall FR, Vollenhoven BJ, Weston GC. Anatomical evidence for in utero androgen exposure in women with polycystic ovary syndrome. *Fertil Steril*. 2005; 84:1689–1692. [PubMed: 16359966]
- Cohen-Bendahan CC, van de Beek C, Berenbaum SA. Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. *Neurosci Biobehav Rev*. 2005; 29:353–384. [PubMed: 15811504]
- de Bruin EI, Verheij F, Wiegman T, Ferdinand RF. Differences in finger length ratio between males with autism, pervasive developmental disorder-not otherwise specified, ADHD, and anxiety disorders. *Dev Med Child Neurol*. 2006; 48:962–965. [PubMed: 17109783]
- Fink B, Manning J, Neave N. Second to fourth digit ratio and the ‘big five’ personality factors. *Pers Individ Differ*. 2004; 37:495–503.
- Fink B, Neave N, Laughton K, Manning J. Second to fourth digit ratio and sensation seeking. *Pers Individ Differ*. 2006; 41:1253–1262.
- Fink B, Neave N, Manning JT. Second to fourth digit ratio, body mass index, waist-to-hip ratio, and waist-to-chest ratio: their relationships in heterosexual men and women. *Ann Hum Biol*. 2003; 30:728–738. [PubMed: 14675912]
- Hall LS, Love CT. Finger-length ratios in female monozygotic twins discordant for sexual orientation. *Arch Sex Behav*. 2003; 32:23–28. [PubMed: 12597269]
- Hampson E, Ellis CL, Tenk CM. On the relation between 2D:4D and sex-dimorphic personality traits. *Arch Sex Behav*. 2008; 37:133–144. [PubMed: 18075733]
- Honekopp J, Manning JT, Muller C. Digit ratio (2D:4D) and physical fitness in males and females: evidence for effects of prenatal androgens on sexually selected traits. *Horm Behav*. 2006a; 49:545–549. [PubMed: 16403410]
- Honekopp J, Voracek M, Manning JT. 2nd to 4th digit ratio (2D:4D) and number of sex partners: evidence for effects of prenatal testosterone in men. *Psychoneuroendocrinology*. 2006b; 31:30–37. [PubMed: 16005157]
- Kemper CJ, Schwerdtfeger A. Comparing indirect methods of digit ratio (2D:4D) measurement. *Am J Hum Biol*. (in press).
- Kraemer B, Noll T, Delsignore A, Milos G, Schnyder U, Hepp U. Finger length ratio (2D:4D) and dimensions of sexual orientation. *Neuropsychobiology*. 2006; 53:210–214. [PubMed: 16874008]

- Lippa RA. Are 2D:4D finger-length ratios related to sexual orientation? Yes for men, no for women. *J Pers Soc Psychol.* 2003; 85:179–188. [PubMed: 12872893]
- Lippa RA. Finger lengths, 2D:4D ratios, and their relation to gender-related personality traits and the Big Five. *Biol Psychol.* 2006; 71:116–121. [PubMed: 16360883]
- Lutchmaya S, Baron-Cohen S, Raggatt P, Knickmeyer R, Manning JT. 2nd to 4th digit ratios, fetal testosterone and estradiol. *Early Hum Dev.* 2004; 77:23–28. [PubMed: 15113628]
- Manikkam M, Crespi EJ, Doop DD, Herkimer C, Lee JS, Yu S, Brown MB, Foster DL, Padmanabhan V. Fetal programming: Prenatal testosterone excess leads to fetal growth retardation and postnatal catch-up growth in sheep. *Endocrinology.* 2004; 145:790–798. [PubMed: 14576190]
- Manning JT, Bundred PE. The ratio of 2nd to 4th digit length: a new predictor of disease predisposition? *Med Hypotheses.* 2000; 54:855–857. [PubMed: 10859702]
- Manning JT, Fink B, Neave N, Caswell N. Photocopies yield lower digit ratios (2D:4D) than direct finger measurements. *Arch Sex Behav.* 2005; 34:329–333. [PubMed: 15971015]
- Manning JT, Scutt D, Wilson J, Lewis-Jones DI. The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Hum Reprod.* 1998; 13:3000–3004. [PubMed: 9853845]
- McFadden D, Shubel E. Relative lengths of fingers and toes in human males and females. *Horm Behav.* 2002; 42:492–500. [PubMed: 12488115]
- McIntyre MH. The use of digit ratios as markers for perinatal androgen action. *Reprod Biol Endocrinol.* 2006; 4:10. [PubMed: 16504142]
- McIntyre MH, Cohn BA, Ellison PT. Sex dimorphism in digital formulae of children. *Am J Phys Anthropol.* 2006; 129:143–150. [PubMed: 16224778]
- McIntyre MH, Ellison PT, Lieberman DE, Demerath E, Towne B. The development of sex differences in digital formula from infancy in the Fels Longitudinal Study. *Proc Biol Sci.* 2005; 272:1473–1479. [PubMed: 16011922]
- Millet K, Dewitte S. Digit ratio (2D:4D) moderates the impact of an aggressive music video on aggression. *Pers Individ Differ.* 2007; 43:289–294.
- Millet K, Dewitte S. The presence of aggression cues inverts the relation between digit ratio (2D:4D) and prosocial behaviour in a dictator game. *Br J Psychol.* 2009; 100:151–162. [PubMed: 18590604]
- Nicholls ME, Orr CA, Yates MJ, Loftus AM. A new means of measuring index/ring finger (2D:4D) ratio and its association with gender and hand preference. *Laterality.* 2008; 13:71–91. [PubMed: 18050002]
- Paul SN, Kato BS, Hunkin JL, Vivekanandan S, Spector TD. The big finger: the second to fourth digit ratio is a predictor of sporting ability in women. *Br J Sports Med.* 2006; 40:981–983. [PubMed: 17008344]
- Pokrywka L, Rachon D, Sucheka-Rachon K, Bitel L. The second to fourth digit ratio in elite and non-elite female athletes. *Am J Hum Biol.* 2005; 17:796–800. [PubMed: 16254897]
- Robertson J, Zhang W, Liu JJ, Muir KR, Maciewicz RA, Doherty M. Radiographic assessment of the index to ring finger ratio (2D:4D) in adults. *J Anat.* 2008; 212:42–48. [PubMed: 18005121]
- Robinson J. Prenatal programming of the female reproductive neuro-endocrine system by androgens. *Reproduction.* 2006; 132:539–547. [PubMed: 17008465]
- Robinson SJ, Manning JT. The ratio of 2nd to 4th digit length and male homosexuality. *Evol Hum Behav.* 2000; 21:333–345. [PubMed: 11053694]
- Ryan BC, Vandenberg JG. Intrauterine position effects. *Neurosci Biobehav Rev.* 2002; 26:665–678. [PubMed: 12479841]
- Scutt D, Manning JT. Symmetry and ovulation in women. *Hum Reprod.* 1996; 11:2477–2480. [PubMed: 8981138]
- van Anders SM, Hampson E. Testing the prenatal androgen hypothesis: Measuring digit ratios, sexual orientation, and spatial abilities in adults. *Horm Behav.* 2005; 47:92–98. [PubMed: 15579270]
- Van den Bergh B, Dewitte S. Digit ratio (2D:4D) moderates the impact of sexual cues on men's decisions in ultimatum games. *Proc Biol Sci.* 2006; 273:2091–2095. [PubMed: 16846918]

- Voracek M, Dressler SG. High (feminized) digit ratio (2D:4D) in Danish men: a question of measurement method? *Hum Reprod.* 2006; 21:1329–1331. author reply 1331–1332. [PubMed: 16611977]
- Voracek M, Manning JT, Dressler SG. Repeatability and interobserver error of digit ratio (2D:4D) measurements made by experts. *Am J Hum Biol.* 2007; 19:142–146. [PubMed: 17160984]
- Voracek M, Manning JT, Ponocny I. Digit ratio (2D:4D) in homosexual and heterosexual men from Austria. *Arch Sex Behav.* 2005; 34:335–340. [PubMed: 15971016]
- Williams JH, Greenhalgh KD, Manning JT. Second to fourth finger ratio and possible precursors of developmental psychopathology in pre-school children. *Early Hum Dev.* 2003; 72:57–65. [PubMed: 12706312]
- Williams TJ, Pepitone ME, Christensen SE, Cooke BM, Huberman AD, Breedlove NJ, Breedlove TJ, Jordan CL, Breedlove SM. Finger-length ratios and sexual orientation. *Nature.* 2000; 404:455–456. [PubMed: 10761903]

TABLE 1

Mean digit ratios derived from physical measurements, photocopies, printed scanned images, and computer-assisted image analysis

| | Mean 2D:4D (\pm SD) | | | |
|--------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| | Physical | Photocopy | Printed scan | Computer assisted |
| Males ($n= 30$) | 0.966 \pm 0.030 ^a | 0.956 \pm 0.035 ^b | 0.955 \pm 0.032 ^b | 0.953 \pm 0.033 ^b |
| Females ($n= 30$) | 0.983 \pm 0.029 ^a | 0.973 \pm 0.029 ^b | 0.973 \pm 0.028 ^b | 0.968 \pm 0.026 ^b |
| Overall Mean ($n= 60$) | 0.974 \pm 0.031 ^a | 0.964 \pm 0.033 ^b | 0.964 \pm 0.032 ^b | 0.960 \pm 0.031 ^b |

Data for mean 2D:4D is derived from two repeated measurements made in the left and right hands of each participant. Significant differences for within row comparisons are denoted by different letters ($P < 0.05$).

TABLE 2

Level of intraobserver agreement when determining digit ratios from physical measurements, photocopies, printed scanned images, and computer-assisted image analysis

| Observer | Intraclass correlation coefficient | | | |
|----------|------------------------------------|--------------------|--------------------|--------------------|
| | Physical | Photocopy | Printed scan | Computer assisted |
| 1 | 0.873 | 0.928 | 0.822 | 0.916 |
| 2 | 0.934 | 0.925 | 0.861 | 0.962 |
| 3 | 0.969 | 0.964 | 0.843 | 0.994 |
| Mean | 0.925 ^{a,b} | 0.939 ^a | 0.842 ^b | 0.957 ^a |

Significant differences for within row comparisons are denoted by different letters ($P < 0.05$).

TABLE 3

Level of interobserver agreement when determining digit ratios from physical measurements, photocopies, printed scanned images, and computer-assisted image analysis

| Observers | Intraclass correlation coefficient | | | |
|-----------|------------------------------------|----------------------|--------------------|--------------------|
| | Physical | Photocopy | Printed scan | Computer assisted |
| 1,2 | 0.811 | 0.888 | 0.766 | 0.895 |
| 1,3 | 0.815 | 0.855 | 0.789 | 0.905 |
| 2,3 | 0.760 | 0.831 | 0.728 | 0.876 |
| Mean | 0.795 ^{a,b} | 0.858 ^{a,c} | 0.761 ^b | 0.892 ^c |

Significant differences for within row comparisons are denoted by different letters ($P < 0.05$).

TABLE 4

Level of interobserver agreement in males and females when determining digit ratios from physical measurements, photocopies, printed scanned images, and computer-assisted image analysis

| | Intraclass correlation coefficient | | |
|-------------------|------------------------------------|--------------------------|-----------------|
| | Males (<i>n</i> = 30) | Females (<i>n</i> = 30) | <i>P</i> -value |
| Physical | 0.857 ^a | 0.690 ^b | 0.0173 |
| Photocopy | 0.879 ^a | 0.808 ^a | 0.0662 |
| Printed scan | 0.805 ^a | 0.657 ^b | 0.0118 |
| Computer assisted | 0.906 ^a | 0.848 ^b | 0.0091 |

Reliability analysis was performed on data comprised of two repeated measurements made in both hands of each male and female participant. Significant differences for within row comparisons are denoted by different letters ($P < 0.05$).

TABLE 5

Level of interobserver agreement for the left and right hands when determining digit ratios from physical measurements, photocopies, printed scanned images, and computer-assisted image analysis

| | Intraclass correlation coefficient | | |
|-------------------|---|---------------------------|----------------|
| | Left hand (n= 60) | Right hand (n= 60) | P-value |
| Physical | 0.821 | 0.777 | 0.2919 |
| Photocopy | 0.838 | 0.875 | 0.2458 |
| Printed scan | 0.730 | 0.794 | 0.1165 |
| Computer assisted | 0.871 | 0.914 | 0.0850 |

Reliability analysis was performed on data comprised of two repeated measurements made in the left and right hands of each participant.