

Northumbria Research Link

Citation: Ribeiro, Evaldo, Neave, Nick, Morais, Rosana and Manning, John (2016) Direct Versus Indirect Measurement of Digit Ratio (2D:4D): A Critical Review of the Literature and New Data. *Evolutionary Psychology*, 14 (1). pp. 1-8. ISSN 1474-7049

Published by: SAGE

URL: <http://dx.doi.org/10.1177/1474704916632536>
<<http://dx.doi.org/10.1177/1474704916632536>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/26742/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

www.northumbria.ac.uk/nrl



Direct Versus Indirect Measurement of Digit Ratio (2D:4D): A Critical Review of the Literature and New Data

Evaldo Ribeiro¹, Nick Neave², Rosana N. Morais¹, and John T. Manning³

Abstract

Digit ratio (2D:4D: the relative lengths of the second and fourth digits) is widely used as a correlate of prenatal sex steroids. There are two common methods of measuring 2D:4D, the direct method and the indirect method. The modern interest in 2D:4D began 16 years ago when finger lengths were measured directly, but many studies now report 2D:4D calculated from indirectly measured fingers from photocopies or scans. However, there are concerns about the accuracy of the latter in comparison to the former. The purpose of this article was twofold: to review these concerns and to add new data to the debate. Our review shows that in 2005, directional effects in indirect 2D:4D were reported such that direct 2D:4D > indirect 2D:4D. This finding was challenged by a 2006 report that direct 2D:4D was lower (not higher) than indirect 2D:4D for male right-hand 2D:4D. Two further studies from the same group have claimed that indirect 2D:4D may be lower, higher, or comparable to direct 2D:4D. More recent comparisons of direct 2D:4D versus indirect 2D:4D and a meta-analysis of Chinese studies have replicated the finding of direct 2D:4D > indirect 2D:4D. We considered an additional sample and found significant direct 2D:4D > indirect 2D:4D for three of four ratios. The overall literature is discussed within the context of standards of research (sample size) and publishing (clarity of report). It is concluded that direct 2D:4D does tend to be greater than indirect 2D:4D. Implications for comparative studies and other aspects of research in 2D:4D are discussed.

Keywords

2D:4D, direct versus indirect measurement, statistical power, publishing standards

Date received: January 19, 2016; Accepted: January 21, 2016

Introduction

Digit ratio is the relative lengths of the index finger (2D) and the ring finger (4D), that is, 2D:4D (Manning, Scutt, Wilson, & Lewis-Jones, 1998). Precise measurement of 2D:4D is important because most individuals show differences in the lengths of 2D and 4D of the order of a few millimeters. It has been known for many years that 2D:4D is sexually dimorphic such that male 2D:4D < female 2D:4D (Baker, 1888). However, much of the recent interest in 2D:4D as a marker for the fetal formation of the reproductive system, ejaculate size, and the secretion of prenatal sex steroids dates from the report of Manning, Scutt, Wilson, and Lewis-Jones (1998). Finger length in the Manning et al.'s (1998) paper was measured directly from the fingers, with measurement landmarks consisting of a midpoint in the proximal ventral flexion crease of the finger and the distal tip of the finger. A UK sample of 800 (400 males) participants with an age range of 2 years–25 years gave a mean of 0.98 for males

and 1.00 for females. These means probably represent approximate values for 2D:4D in Caucasians. Further work suggested that ethnic differences existed in mean 2D:4D such that Caucasian 2D:4D > East Asian 2D:4D > Black 2D:4D (Manning et al., 2000; Peters, Mackenzie, & Bryden, 2002) and that there may be between-nation differences in Caucasian mean 2D:4D (Manning, 2002). However, comparative work concerning

¹ Department of Physiology, Federal University of Paraná, Curitiba, Brazil

² Department of Psychology, Northumbria University, Newcastle-upon-Tyne, United Kingdom

³ Applied Sports, Technology, Exercise and Medicine (A-STEM), Swansea University, Swansea, United Kingdom

Corresponding Author:

John T. Manning, Applied Sports, Technology, Exercise and Medicine (A-STEM), Swansea University, Swansea, United Kingdom.

Email: j.manning@swansea.ac.uk



between-population differences in 2D:4D may have been adversely affected by artifacts introduced by the process of indirectly measuring finger lengths.

The Manning et al. (1998) report measured finger length directly on the fingers. This method is costly in terms of participant time and may be affected by such things as movement of the participants' fingers. This led Robinson and Manning (2000) to measure finger length indirectly from photocopies of the hand. It was found that a comparison between direct and indirect 2D:4D gave high intraclass correlation coefficients (ICCs; Robinson & Manning, 2000), and as a consequence, indirectly measured 2D:4D (often from photocopies or scans) has been reported by many subsequent studies. However, there is a problem in the comparison of direct versus indirect 2D:4D using ICCs. Values of ICC are not sensitive to directional effects. Thus, direct 2D:4D may be consistently higher than indirect 2D:4D, and the directionality of the effect would not influence the ICC. If direct 2D:4D > indirect 2D:4D, it is important to understand the properties of indirect 2D:4D in relation to direct 2D:4D.

The purpose of this study was twofold: to review research on the issue of direct versus indirect measurement of digit ratio and to add new data to the debate. Manning, Fink, Neave, and Caswell (2005) have presented evidence that 2D and 4D finger lengths measured indirectly may give mean 2D:4D values that show directional effects such that direct 2D:4D > indirect 2D:4D. If these findings are correct, it is important that direct and indirect mean 2D:4D's should not be included together in between-population studies, and we should understand the possible limitations concerning indirect 2D:4D. Here the attempts to replicate the Manning et al.'s (2005) finding are reviewed. The review considers possible limitations in this literature, which may include issues of power and lack of clarity in report. New data are also presented which consider the directional effect of indirect 2D:4D. The direct 2D:4D > indirect 2D:4D relationship is discussed from the viewpoint of comparative studies and other aspects of research in 2D:4D.

Method

The Review

We report a narrative review of the Manning et al.'s (2005) study and the replications of this work. We searched for reports concerning direct and indirect measurement of 2D:4D using PubMed and the Web of Science. The search terms were digit ratio, 2D:4D, finger lengths, direct measurement, and indirect measurement. Our criteria for inclusion in the review were that mean and *SD* for direct and indirect 2D:4D should be given for the same sample rather than comparisons of subsamples. In addition, it was required that finger length was measured by experimenters rather than self-measured.

Direct Versus Indirect Data

The participants were recruited from students and staff of a UK University. Approval was provided by the local ethics

committee, and informed consent was signed before the study commenced.

Finger lengths were measured directly and from scans with digital callipers, measuring to 0.01 mm. All hands were scanned in pdf format using a Canon Canoscan LiDE 700F scanner whose maximum optical resolution was 4800 dpi, and mechanical resolution was 9600 dpi.

Direct measurements were taken from the fingers. The sequence of measurement was as follows: The second and fourth fingers were measured from the right hand and then the left. This sequence was then repeated such that the second and fourth fingers were measured twice. Measurement was of the ventral surfaces of the second and fourth fingers from a mid-point on the ventral crease proximal to the palm, to the tip of the finger. With regard to the scans, participants were asked to gently place their hands palm down on the glass platen. A scan was taken and checked for clarity at the ventral crease and the tip of the finger. Measurements (indirect) were made using the same landmarks as were used for direct measurement, and fingers were measured twice.

Results

The Review

Manning et al. (2005) suggested that there was a tendency in the 2D:4D literature between 2000 and 2005 for mean direct 2D:4D to be higher than that of mean indirect 2D:4D. Focusing on this possibility, they compared direct and indirect 2D:4D in 50 men and 70 women. They found that mean direct 2D:4D was higher than mean indirect 2D:4D. We have here directional effects such that, in comparison to direct measurements, indirectly measured 4D is relatively longer than 2D. Direct and indirect mean 2D:4D, *SD*s, and the effect sizes (Cohen's *d*) of the difference in 2D:4D between the 2 protocols are given in Table 1 and Figure 1. All four values of *d* were positive (direct 2D:4D > indirect 2D:4D), one was not significant (females right 2D:4D, *d* = +.10), but the remainder were significant with *d* varying from +.20 to +.5.

There have been eight replications of the Manning et al. (2005) UK report. Three of these were from Austria (the Voracek lab) and five from labs in Canada, Saudi Arabia, Spain, South Korea, and the United States. The Voracek lab reported a variety of effects, with direct 2D:4D < indirect 2D:4D; direct 2D:4D \approx indirect 2D:4D; and direct 2D:4D > indirect 2D:4D, and the remaining five labs reported all relationships to be direct 2D:4D > indirect 2D:4D. We first consider the Voracek reports.

The first replication was that of Voracek and Dressler (2006; see Table 1 and Figure 1). The sample comprised 30 Austrian men, and only the right hand was measured. Each hand was measured in four ways: (i) directly on the ventral surface of the fingers; (ii) indirectly from scans of the ventral surface of the fingers; (iii) indirectly from dorsal (palm down) hand outline drawings; and (iv) indirectly from hand-outline drawings (palm up). For our purposes, we must focus on methods (i) and (ii).

Table 1. Mean 2D:4D (Males Right and Left 2D:4D and Females Right and Left 2D:4D) for Direct and Indirectly Measured Fingers and the Difference (Direct–Indirect) and Effect Size of the Difference in Means.

Study	2D:4D Trait	<i>n</i>	Direct Mean (SD)	Indirect Mean (SD)	Direct–Indirect	Effect Size
Manning et al. (2005), Austria & England	Male right	50	.968 (.033)	.956 (.029)	+0.012*	+0.39
	Male left	50	.981 (.034)	.963 (.033)	+0.018*	+0.54
	Female right	50	.982 (.031)	.979 (.028)	+0.003	+0.10
	Female left	50	.987 (.031)	.981 (.030)	+0.006*	+0.20
Voracek and Dressler (2006), Austria	Male right	30	.953 (.031)	.963 (.033)	–0.010*	–0.31
	Male left	60	.969 (.028)	.961 (.028)	+0.008*	+0.29
Voracek and Offenmuller (2007), Austria	Male left	60	.967 (.026)	.955 (.027)	+0.012*	+0.45
	Female right	64	.976 (.033)	.977 (.034)	–0.001	–0.03
	Female left	64	.978 (.028)	.969 (.031)	+0.009*	+0.31
	Male right	75	.956 (.027)	.953 (.024)	+0.003*	+0.12
Dressler and Voracek (2011), Austria	Male left	75	.973 (.03)	.955 (.028)	+0.018*	+0.62
	Female right	75	.961 (.029)	.964 (.031)	–0.003	–0.10
	Female left	75	.980 (.029)	.966 (.029)	+0.014*	+0.48
	Male mean	30	.966 (.03)	.956 (.035)	+0.014*	+0.31
Allaway, Bloski, Pierson, and Lujan (2009), Canada	Female mean	30	.983 (.983)	.973 (.029)	+0.010*	+0.34
	Male right	276	.98 (.04)	.96 (.03)	+0.02*	+0.57
Almasry, El Domiaty, Algaidi, Elbastawisy, and Safwat (2011), Saudi Arabia	Male left	276	.97 (.04)	.96 (.04)	+0.01*	+0.25
	Female right	284	.98 (.04)	.976 (.037)	+0.004*	+0.10
	Female left	284	.99 (.04)	.97 (.04)	+0.02*	+0.50
	Male right	24	.978 (.038)	.971 (.027)	+0.007	+0.21
Costas et al. (2013), Spain	Male left	24	.979 (.038)	.971 (.037)	+0.008	+0.21
	Female right	26	.991 (.980)	.980 (.028)	+0.011	+0.35
	Female left	26	.978 (.033)	.974 (.026)	+0.004	+0.14
	Male right	115	.962 (.03)	.949 (.03)	+0.013*	+0.43
Kim and Cho (2013), Korea	Male left	115	.961 (.03)	.945 (.03)	+0.016*	+0.53
	Female right	157	.972 (.03)	.961 (.03)	+0.011*	+0.37
	Female left	157	.970 (.03)	.962 (.03)	+0.008*	+0.27
	Male right	201	.98 (.03)	.95 (.03)	+0.03	+1.00
Shaw, Kotowski, Bostor, and Levine (2012), United States	Male left	201	.97 (.03)	.95 (.03)	+0.02	+0.67
	Female right	204	.99 (.03)	.97 (.03)	+0.02	+0.67
	Female left	204	.98 (.03)	.97 (.03)	+0.01	+0.67

Note. The effect size is positive when direct 2D:4D > indirect 2D:4D and negative when direct 2D:4D < indirect 2D:4D. The *p* values included are from paired *t* tests reported in the relevant papers.

*significant *p* < .05 or less.

Voracek and Dressler (2006) reported that direct measurement yielded *lower* 2D:4D than that from indirect scans ($d = -0.31$, note the negative effect size). They pointed out that this “was exactly the opposite of what Manning et al. (2005) found.”

The remaining two reports from the Voracek lab consisted of a mix of positive and negative effects (Figure 1). With regard to Voracek and Offenmüller (2007), the effect sizes for direct versus indirect 2D:4D comparisons are difficult to identify as they are reported along with finger ratios that are other than 2D:4D (e.g., 2D:3D, 2D:5D, . . .). However, careful reading of this article indicates of the four possible 2D:4D comparisons, three show positive and significant effect sizes, that is, males right $d = +0.29$; males left $d = +0.45$; and females left $d = +0.31$; all $p < .05$. There was one very small negative effect size, females right $d = -0.03$, not significant (*ns*). This finding for 2D:4D replicates the findings of Manning et al. (2005) by sex and hand and is a nonreplication of Voracek and Dressler (2006).

With regard to Dressler and Voracek (2011), in contrast to Voracek and Offenmüller (2007), the effect sizes for direct versus indirect ratios are clearly set out for 2D:4D as against all other ratios. There are again similar results for 2D:4D to that of Manning et al. (2005), in that there are three positive and significant effect sizes, these are by sex and hand: males right $d = +0.12$ and left $d = +0.62$; females left $d = +0.48$; all $ps < .05$. For female right 2D:4D, there was a negative effect size which was not formally significant ($d = -0.10$, *ns*). Therefore, this is another nonreplication of Voracek and Dressler (2006).

There are five other replications of the direct 2D:4D > indirect 2D:4D hypothesis (Table 1, Figure 2). The reports are from five independent labs, five nations (Canada, Saudi Arabia, Spain, South Korea, and the United States), and two macro-ethnic groups (Caucasians and East Asians). Sample sizes varied from $n = 24$ to $n = 284$. There were 18 effect sizes, and none of these yielded the Voracek and Dressler (2006) effect. Of importance, all effect sizes were positive, that is, they

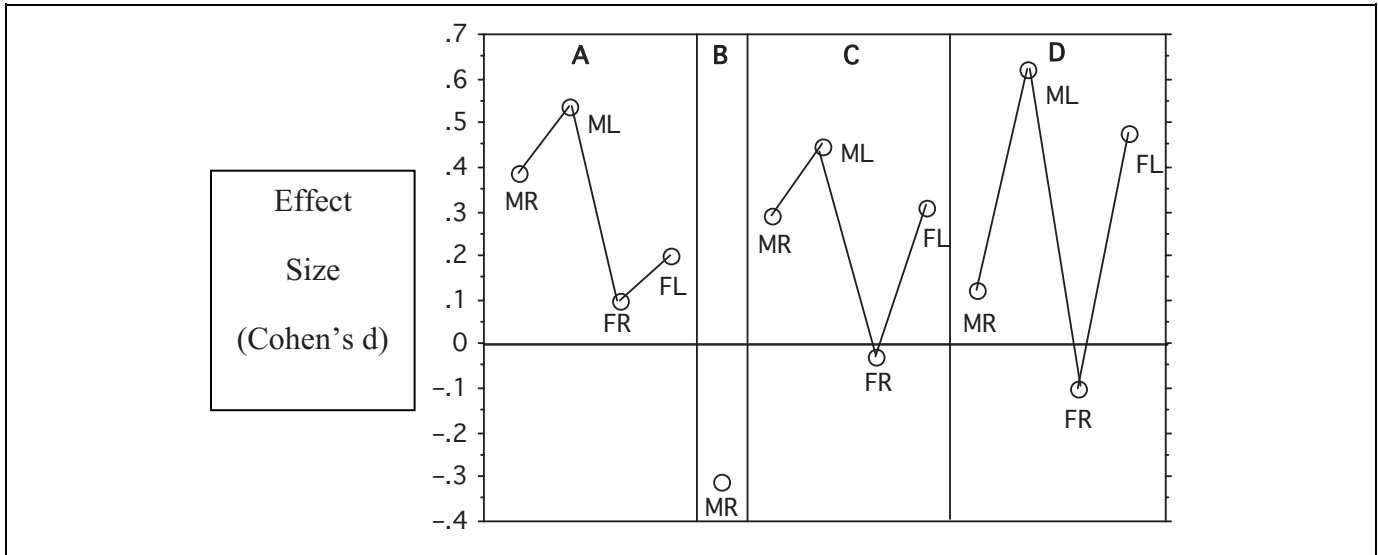


Figure 1. Effect sizes (Cohen's d) for the difference between mean 2D:4D from direct measurement and indirect measurement (direct–indirect). A = Manning et al., 2005; B = Voracek & Dressler, 2006; C = Voracek & Offenmüller, 2007; and D = Dressler & Voracek, 2011). Positive values denote direct mean 2D:4D > indirect mean 2D:4D and negative values denote direct mean 2D:4D < indirect mean 2D:4D. M = male and F = female; R = right hand and L = left hand. The effect sizes for the four 2D:4D ratios are connected with solid lines to facilitate recognition of the similarities in pattern between studies A, C, and D and the discrepant nature of the results of study B.

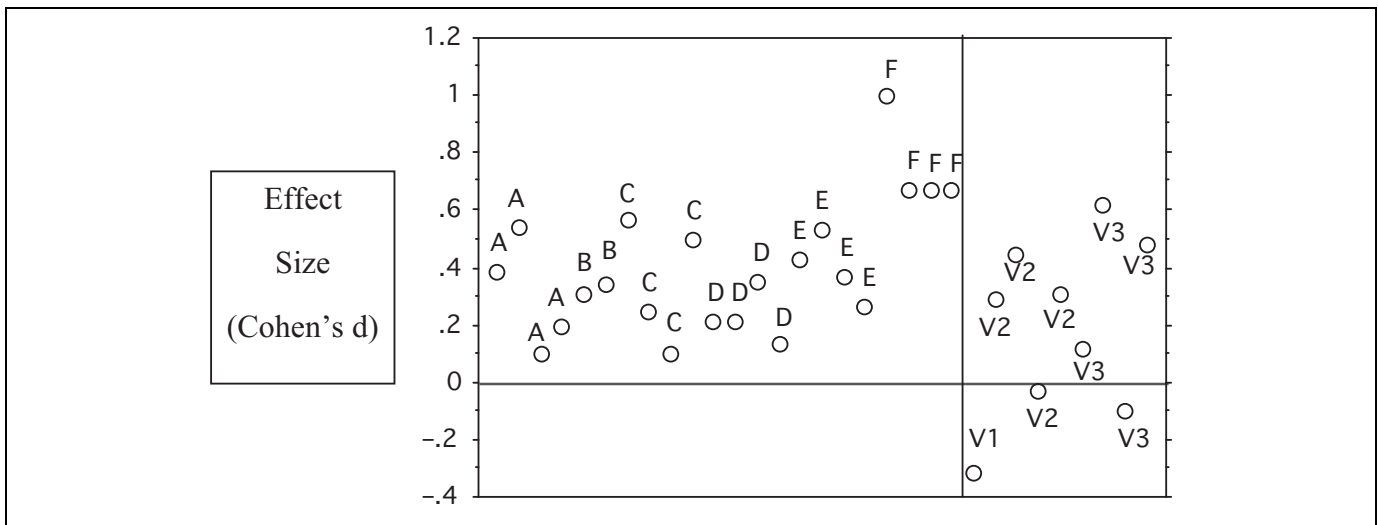


Figure 2. Effect sizes (Cohen's d) for the difference between mean 2D:4D from direct measurement and indirect measurement (direct–indirect). A = Manning et al. (2005); B = Allaway, Bloski, Pierson, and Lujan (2009); C = Almasry et al. (2011); D = Costas et al. (2013); E = Kim and Cho (2013); and F = Shaw, Kotowski, Bostor, and Levine (2012). All these effect sizes ($n = 22$) are positive, that is, they denote direct mean 2D:4D > indirect mean 2D:4D. V denotes studies from the Voracek lab, that is, V1 = Voracek and Dressler (2006); V2 = Voracek and Offenmüller (2007); and V3 = Dressler and Voracek (2011). Of nine effect sizes from this lab, three are negative (direct mean 2D:4D < indirect mean 2D:4D), albeit only one of these is significant.

denoted direct mean 2D:4D > indirect mean 2D:4D, they ranged from $d = +.10$ to $+1.00$.

The effects of direct versus indirect measurements of 2D:4D are best investigated by applying both measurement procedures to the same subjects. However, we can also obtain valuable information by comparing means for direct and indirect 2D:4D across a number of studies. There is one such relevant meta-analysis. Xu and Zheng (2015) considered 2D:4D in

China (mainland and Taiwan), with means for 2D:4D in 28 studies, 112 independent samples and 19,093 participants. As expected, there was a sex difference in 2D:4D with lower mean for males compared to females. However, method of measurement (direct vs. indirect) was a significant moderator variable. With regard to males, for right-hand 2D:4D, there were 14 directly measured ($n = 2,790$) and 15 indirectly measured ($n = 1,698$) samples. Consistent with the findings of Manning

et al. (2005) but counter to that of Voracek and Dressler (2006), the mean for the former was significantly greater than the mean for the latter (direct .953 vs. indirect .942, between study homogeneity statistic $Q_b = 5.0, p < .05$). For left 2D:4D, there were 14 ($n = 2,790$) and 13 ($n = 1,430$) samples for direct and indirect measurement, respectively. Mean direct 2D:4D was .955 compared to .947 for indirect 2D:4D ($Q_b = 2.66, p > .05$). With regard to females, for right hand 10 direct ($n = 2,156$) samples gave a mean of .961 and 15 indirect ($n = 1,958$) samples a mean of .956, $Q_b = .64, p > .05$. For left hand, 10 direct ($n = 2,156$) samples gave a mean of .960 and 15 indirect ($n = 1,792$) samples a mean of .958, $Q_b = .12, p > .05$. Thus, there was an overall tendency for higher mean direct 2D:4D compared to indirect 2D:4D.

Direct Versus Indirect 2D:4D: New Data

There were 114 male and 90 female participants in the sample. With regard to direct measurement, the repeatability (intraclass coefficient; r_1) of the first and second measurements of 2D:4D was high and significant for right 2D:4D (males and females $r_1 = .987, p < .0001$; males $r_1 = .989, p < .0001$; and females $r_1 = .984, p < .0001$) and somewhat lower but still significant for left 2D:4D (males and females $r_1 = .983, p < .0001$; males $r_1 = .989, p < .0001$; and females $r_1 = .977, p < .0001$). In comparison, the first and second indirect measurements gave even higher r_1 values for both right 2D:4D (males and females $r_1 = .998, p < .0001$; males $r_1 = .998, p < .0001$; and females $r_1 = .998, p < .0001$) and left 2D:4D (males and females $r_1 = .998, p < .0001$; males $r_1 = .997, p < .0001$; and females $r_1 = .998, p < .0001$). The high repeatability values for both types of measurement indicated that the between-individual differences in 2D:4D were significantly greater than the measurement error. Therefore, we calculated the mean of the first and second 2D:4D for direct and indirect measurement.

There were significant sex differences in 2D:4D with higher values in females compared to males. The mean values ($\pm SD$) of 2D:4D from direct measurement were: right 2D:4D males = $.969 \pm .030$, females = $.983 \pm .029, t = 3.44, p = .0007, d = .47$; left 2D:4D males = $.966 \pm .029$, females $.978 \pm .033, t = 2.77, p = .006, d = .39$. For indirect measurement, the means were right 2D:4D males = $.961 \pm .031$, females = $.975 \pm .032, t = 3.17, p = .002, d = .44$ and left 2D:4D males = $.961 \pm .029$, females = $.975 \pm .034, t = 3.36, p = .0009, d = .46$. Thus, the effect size (Cohen's d) varied from $d = .39$ (direct) to $d = .47$ (direct).

For both right and left 2D:4D, the male direct measurement means were significantly greater than the indirect means, paired t -test direct (x)–indirect (y): right 2D:4D $x-y = .008, t = 6.89, p < .0001, d = .26$; left 2D:4D $x-y = .005, t = 4.16, p < .0001, d = .21$. For females, the directional effects of indirect measurement were somewhat weaker than those found for males. Direct mean 2D:4D was significantly greater than indirect mean 2D:4D for the right hand ($x-y = .008, t = 4.64, p < .0001, d = .26$). The difference was positive for the left hand but not significant ($x-y = .003, t = 1.42, p = .16, d = .09$).

The effect size for right 2D:4D (male and female) and for left 2D:4D (male only) was about half the effect size for sex differences in 2D:4D.

Discussion

It is difficult to successfully navigate through the world of direct versus indirect 2D:4D measurement. However, the evidence from our review points to the following: (i) Manning et al. (2005) reported differences between mean direct 2D:4D and mean 2D:4D in a Austria/UK study such that the former is greater than the latter. The differences were significant for three of the four possible ratios. (ii) There are three reports from Austrian studies of direct 2D:4D < indirect 2D:4D, they are lab specific and are significant for only one instance of direct right-hand 2D:4D < indirect right-hand 2D:4D for males only. This single study was not replicated in two additional studies by the same lab. In both the latter studies, they found significant direct 2D:4D > indirect 2D:4D, for three of the four possible 2D:4D ratios (i.e., six ratios in total). (iii) Five studies from Canada, Saudi Arabia, Spain, South Korea, and the United States reported 18 positive effect sizes for direct 2D:4D > indirect 2D:4D. (iv) A meta-analysis of Chinese 2D:4D studies reported direct 2D:4D > 2D:4D, and this was significant for right-hand male 2D:4D. With regard to our new data, we have found significant positive effect sizes such that direct 2D:4D > indirect 2D:4D in three of four ratios with d values ranging from .21 to .26. These effect sizes are about half those found for sex differences in 2D:4D. The remaining ratio (female left 2D:4D) showed a positive but nonsignificant effect size ($d = .09$).

Therefore, it seems that there is convincing evidence that direct 2D:4D > indirect 2D:4D. However, there are still concerns with this conclusion. That the overall picture is not as clear as it might be could result from limitations in standards of research (issues of power) and publishing (issues of clarity of report) which may have caused some confusion.

Power Issues

Statistical power calculations appear to support the view that the Voracek and Dressler (2006) study was underpowered. If the true population effect for mean direct 2D:4D—mean indirect 2D:4D was indeed $d = -.31$, then with a sample of $n = 30$, one would have merely 38% power to detect a significant deviation in the mean from 0 (two-tailed test). Note that it is inappropriate to apply a one-tailed test here, as the Voracek and Dressler (2006) effect was opposite to that of Manning et al. (2005). However, even with a one-tailed test, the power to detect a significant deviation from 0 would only be 51%. This is far from a reasonably high power which is often taken as 80%.

Publishing Standards

With such low power, it would have been advisable to include a replication in the Voracek and Dressler (2006) report. The

absence of a replication negatively affects publishing standards. Two further replications were published by Voracek and Offenmüller (2007) and Dressler and Voracek (2011). However, there are issues here that may relate to clarity of report which affect publishing standards and which may have caused confusion in the literature. In these studies, there were six significant positive effect sizes and no significant negative effect sizes for 2D:4D. Despite this, differences between direct and indirect ratios were reported as a mix of positive and negative effect sizes, many of them were significant. Thus, in Voracek and Offenmüller (2007, p. 151), we have the following statement “the current data do not indicate that image-based measurements necessarily yield lower digit ratios (Manning, Fink, Neave, & Caswell, 2005). In point of fact compared with direct finger measurements, they yield lower, comparable, or even higher digit ratios (such as observed in the measurement study of Voracek & Dressler, 2006) for both men and women.” In addition, Dressler and Voracek (2011, p. 75) stated “the hypothesis that image-based digit ratios are uniformly lower than directly measured ones was not supported”. The apparent discrepancy between the actual findings and the report of the findings arises because Voracek and Offenmüller (2007) and Dressler and Voracek (2011) reported direct versus indirect comparisons for 2D:4D and five other finger ratios. It is in these latter ratios, not in 2D:4D, that they found all their significant negative effect sizes. They found no such effects for 2D:4D, and Manning et al. (2005) did not consider finger ratios other than 2D:4D. In addition, there seems to be clarity issues in later reports. Thus, Voracek et al. (2008, p. 508) state “Of note, results from one lab, indicating that image-based methods may uniformly yield lower 2D:4D than direct measurements (Caswell & Manning, 2009; Manning et al., 2005), have not been replicated (Voracek & Dressler, 2006; Voracek & Offenmüller, 2007).” This statement of null replication appears to be incorrect for Manning et al. (2005) versus Voracek and Offenmüller (2007), as the latter replicates the findings of Manning et al. (2005), that is, significant differences in direct 2D:4D > indirect 2D:4D for male right and left hand and female left hand.

In the light of the abovementioned findings, what are we to make of the report of direct 2D:4D < indirect 2D:4D by Voracek and Dressler (2006)? Consequent to this study, five independent groups have failed to replicate this single effect size. There have been two replication attempts by Voracek and colleagues. Putting aside possible issues regarding clarity of report, they failed to find a single significant example of direct 2D:4D < indirect 2D:4D. In addition, after a meta-analysis of Chinese mean 2D:4D's and a consideration of evidence that they were modified by measurement method, Xu and Zheng (2015) concluded that Voracek and Dressler's (2006) report was incorrect. Therefore, it may be best to regard the Voracek and Dressler's (2006) finding as a chance event resulting from power issues with a sample size of $n = 30$. However, there is another, albeit unlikely, possibility. It is well known that there are significant ethnic and national effects on 2D:4D. It may be that there is some aspect of Austrian 2D:4D which lends itself to distortional effects of indirect measurement such that

(in comparison to direct 2D:4D) it shows a mix of positive and negative effects. This should be borne in mind in further work concerning direct versus indirect measurement effects.

If we have arrived at the conclusion that the available evidence indicates direct 2D:4D > indirect 2D:4D, we must address a number of questions. Firstly, what is the mechanism that results in these directional effects of indirect 2D:4D? Secondly, does it matter whether direct or indirect 2D:4D is used in studies? Thirdly, what is the way forward with regard to measurement method in 2D:4D studies?

With regard to the first question, the mechanism that drives indirect 2D:4D down in comparison to direct 2D:4D is unknown. However, it is likely to involve one or both measurement points on the fingers, that is, the tip of the finger and the basal crease. Manning et al. (2005) pointed out that indirect methods involve placing the fingers on the glass plate of the photocopier or scanner. Downward pressure of the hand may distort the fleshy tips of the fingers, the pressure may vary according to sex and even according to personality types. These factors may reduce 2D:4D, particularly if distortion of the fat pad at the tip of the fourth finger was greater than that of the second finger. There has been one test of this suggestion. Voracek et al. (2008) have examined fingertip size in relation to indirect finger measurements. They found large differences in fingertip size but claimed none of these influenced image-based 2D:4D measurements. However, their study did not consider direct measurements. Therefore, the null conclusions of Voracek et al. (2008) should be treated with caution. The second measurement point is to be found at the midpoint of the ventral basal crease of the finger, which is proximal to the palm. This crease is established early in utero, but it is a soft tissue marker and lies about halfway along the proximal phalanx of the digit. One possibility relating to directional effects involves movement of this crease, when the palm is turned downward onto the glass plate of the photocopier or scanner. In short, it is worth considering that nonconcordance of direct and indirect measurement points at the tip and base of the finger may lead to directional effects in indirect 2D:4D.

Considering the second question, if direct 2D:4D shows stronger relationships with target traits than indirect 2D:4D, then measurement protocol does matter in 2D:4D studies. There have been three studies that considered differences in the strength of direct and indirect 2D:4D with target traits (including facial-metrics, Burriss et al., 2007; body size, Almasry, El Domiaty, Algaidi, Elbastawisy, & Safwat, 2011, and dermatoglyphics, Dressler & Voracek, 2010). In comparison to indirect 2D:4D, Burriss et al. (2007) and Almasry, El Domiaty, Algaidi, Elbastawisy, and Safwat reported stronger correlations between direct 2D:4D and facial metrics or body size. Dressler and Voracek (2011) found no significant correlation between 2D:4D (direct or indirect) and dermatoglyphics. Further work is needed in order to clarify whether the strength of associations between 2D:4D and its target trait is dependent on measurement protocol.

With regard to the way forward, we should not include direct and indirect samples together in comparative studies of

2D:4D. This became clear in 2005. Comparative studies before this date should be considered with caution (e.g., see Manning, 2002), and where possible, allowance should be made for differences in measurement method. We should disregard the results of comparative studies after 2005 if they mix direct and indirect samples. Two examples relating to comparative differences in 2D:4D between nations and changes in 2D:4D within nations will suffice to illustrate this problem. An example of the former is afforded by the report of very high mean 2D:4D (>1.00) in Danish men by Bang et al. (2005). This study employed an unusual measurement protocol for 2D:4D which may be the source of their high mean 2D:4D. Voracek and Dressler (2006) considered the effect of the Bang et al. (2005) protocol and concluded that it produced mean 2D:4D which was not significantly different from that of indirect 2D:4D. In view of the power concerns regarding the Voracek and Dressler (2006) report, it would be advisable to reconsider the issue of a high 2D:4D for Danish men. Thus, a comparison of self-reported directly measured 2D:4D for 29 countries showed, for right-hand male 2D:4D, a Danish mean of .982. This was lower than that of 18 of the 29 countries concerned (Manning, Fink, & Trivers, 2014). Therefore, comparative data using direct finger measurement in a large ($n > 250,000$) multinational sample do not support the notion of a very high 2D:4D in Danish men. The magnitude of the mean 2D:4D reported by Bang et al. (2005) may indeed have been related to the method of finger measurement they employed. With regard to comparative changes in 2D:4D within nations, an appropriate example is that of putative change in 2D:4D within Lithuania between samples from the 1880s and modern populations. Digit ratio was measured in the 1880s, and these potentially valuable measures compared to indirect 2D:4D measured in a contemporary sample (Voracek et al., 2007). There were significant differences between the two samples. Time-dependent change in 2D:4D may indicate the modern influence of endocrine disruptors in the environment. However, the sample from the 1880s was measured using a rather unusual direct method, and the modern participants were measured using a standard indirect method. The authors of the study noted the potential for differences in the two measurement protocols to invalidate their comparisons. Their caution was probably correct. It is emphasized that time-dependent and space-dependent comparisons of mean 2D:4D must control for measurement protocol.

What of current studies? What is the “best” method for measuring finger length? Well, this depends on the study. If small numbers of participants are involved and there is adequate time to measure, then direct measurements are indicated. For large samples and/or samples that require fast measurement, indirect measurement may be appropriate. However, for the latter, it may be wise to ensure that participants are instructed to lightly place their hand on the glass plate of the photocopier/scanner.

In conclusion, a comparison of direct and indirect 2D:4D shows the former tends to be larger than the latter (i.e., direct 2D:4D $>$ indirect 2D:4D). Most labs that have considered this issue have concluded that this is so. Reports of direct 2D:4D $<$ indirect 2D:4D are lab specific although we cannot exclude the

possibility that they may be nation-specific. The directional effect of indirect 2D:4D is complex and is little understood, and there are many reports of indirect 2D:4D in the literature. Therefore, we need to understand its effect on 2D:4D. Further work is indicated to elucidate this and should include samples from a number of countries (all nonreplications thus far are reported from Austria, and data from this nation would be particularly valuable) and ethnicities. Suggestions for the appropriate use of direct and indirect methodologies in 2D:4D research are discussed.

Acknowledgments

The authors thank the participants in this study for their time and patience.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Evaldo Ribeiro Jr acknowledges the scholarship from Brazilian Programme “Science Without Borders” (CAPES; no. BEX 6473/14-8), Brazil.

References

- Allaway, H. C., Bloski, T. G., Pierson, R. A., & Lujan, M. E. (2009). Digit ratios (2D:4D) determined by computer-assisted analysis are more reliable than those using physical measurements, photocopies, and printed scans. *American Journal of Human Biology, 21*, 365–370.
- Almasry, S. M., El Domiaty, M. A., Algaidi, S. A., Elbastawisy, Y. M., & Safwat, M. D. (2011). Index to ring digit ratio in Saudi Arabia at Almadinah Almonawarah province: A direct and indirect measurement study. *Journal of Anatomy, 218*, 202–208.
- Baker, F. (1888). Anthropological notes on the human hand. *American Anthropologist, 1*, 51–76.
- Bang, A. K., Carlsen, E., Holm, M., Petersen, J. H., Skakkebaek, N. E., & Jorgensen, N. (2005). A study of finger lengths, semen quality and sex hormones in 360 young men from the general Danish population. *Human Reproduction, 20*, 3109–3113.
- Burriss, R. P., Little, A. C., & Nelson, E. C. (2007). 2D:4D and sexually dimorphic facial characteristics. *Archives of Sexual Behavior, 36*, 377–384.
- Caswell, N., & Manning, J. T. (2009). A comparison of finger 2D: 4D by self-report direct measurement and experimenter measurement from photocopy: methodological issues. *Archives of Sexual Behavior, 38*, 143–148.
- Costas, L., Papadopoulou, E., Perez-Gomez, B., Basagana, X., Alarcán, F., Casabonneai, D., . . . de Sanjosé, S. (2013). Reliability of 2D:4D measurements using a direct method suitable for clinical settings. *Personality and Individual Differences, 55*, 339–342.
- Dressler, S. G., & Voracek, M. (2011). No association between two candidate markers of prenatal sex hormones: Digit ratios (2D:4D and other) and finger-ridge counts. *Developmental Psychobiology, 53*, 69–78.

- Kim, S.-I., & Cho, K.-J. (2013). Difference of second to fourth digit ratio according to the methods of measuring length. *Korean Journal of Physical Anthropology*, *26*, 25–32.
- Manning, J. T. (2002). *Digit ratio: A pointer to fertility, behavior, and health*. New Brunswick, NJ: Rutgers University Press.
- Manning, J. T., Barley, L., Lewis-Jones, I., Walton, J., Trivers, R. L., Thornhill, R., . . . Sved, A. (2000). The 2nd to 4th digit ratio, sexual dimorphism, population differences and reproductive success: Evidence for sexually antagonistic genes. *Evolution and Human Behavior*, *21*, 163–183.
- Manning, J. T., Fink, B., Neave, N., & Caswell, N. (2005). Photocopies yield lower digit ratios (2D:4D) than direct measurements. *Archives of Sexual Behavior*, *34*, 329–333.
- Manning, J. T., Fink, B., & Trivers, R. (2014). Digit ratio (2D:4D) and gender inequalities across nations. *Evolutionary Psychology*, *12*, 757–768.
- Manning, J. T., Scutt, D., Wilson, J., & Lewis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: A predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Human Reproduction*, *13*, 3000–3004.
- Peters, M., Mackenzie, K., & Bryden, P. (2002). Finger length and distal finger extent patterns in humans. *American Journal of Physical Anthropology*, *117*, 209–217.
- Robinson, S. J., & Manning, J. T. (2000). The ratio of 2nd to 4th digit length and male homosexuality. *Evolution and Human Behavior*, *21*, 333–345.
- Shaw, A. Z., Kotowski, M. R., Bostor, F. T., & Levine, T. R. (2012). The effect of prenatal sex hormones on the development of verbal aggression. *Journal of Communication*, *62*, 778–793.
- Voracek, M., Bagdonas, A., & Dressler, S. G. (2007). Digit ratio (2D:4D) in Lithuania once and now: Testing for sex differences, relations with eye and hair color and a possible secular change. *Collegium Antropologicum*, *3*, 863–868.
- Voracek, M., & Dressler, S. G. (2006). High (feminized) digit ratio (2D:4D) in Danish men: A question of measurement method? *Human Reproduction*, *21*, 1329–1331.
- Voracek, M., & Offenmüller, D. (2007). Digit ratios (2D:4D and other) and relative thumb length: A test of developmental stability. *Perceptual and Motor Skills*, *105*, 143–152.
- Voracek, M., Pietschnig, J., & Oeckher, M. (2008). Finger, sex and side differences in finger tip size and lack of association with image-based digit ratio (2D:4D) measurements. *Perceptual and Motor Skills*, *107*, 507–512.
- Xu, Y., & Zheng, Y. (2015). The digit ratio (2D:4D) in China: a meta-analysis. *American Journal of Human Biology*, *27*, 304–9.