Comparing Indirect Methods

of Digit Ratio (2D:4D) Measurement

(Running Head: Methods of Digit Ratio Measurement)

Christoph J. Kemper & Andreas Schwerdtfeger

Johannes Gutenberg-University Mainz, Department of Psychology, Staudingerweg 9,

55099 Mainz, Germany

Corresponding author:

Christoph J. Kemper

Johannes Gutenberg-University Mainz

Department of Psychology

Staudingerweg 9

55099 Mainz

Germany

Phone: ++49 (0)6131 392 25 76

Fax: ++49 (0)6131 392 24 83

E-Mail: kemperc@uni-mainz.de

This is a preprint of an article published in the

American Journal of Human Biology

Published in final edited form as:

***Am J Hum Biol*. 2009, 21(2): 188-191. doi: 10.1002/ajhb.20843**

Abstract

The ratio of the lengths of the second and fourth finger (2D:4D) has been proposed to index prenatal exposure to androgens. Different methods have been utilized to measure digit ratio, however, their measurement precision and economy have not been systematically compared yet. Using different indirect methods (plastic ruler, caliper, computer software), three independent raters measured finger lengths of 60 participants. Generally, measurement precision (ICC, TEM, and relative TEM) was acceptable for each method. However, precision estimates were highest for the computer software, indicating excellent measurement precision. Estimates for the caliper method were somewhat lower followed by ruler which had the lowest precision. On the contrary, the software-based measurements took somewhat longer to complete than the other methods. Nonetheless, we would favor the use of these tools in digit ratio research because of their relative superior reliability which could be crucial when associations with other variables are expected to be low to moderate or sample size is limited. Software offers several promising opportunities that may contribute to an accurate identification of the proximal finger crease (e.g., zooming, adjusting contrast, etc.).

Keywords: digit ratio, finger length, measurement methods, precision, measurement economy

Introduction

The relative length of the second and fourth digit (so-called 2D:4D or digit ratio) represents an individual difference variable putatively related to prenatal gonadal hormone exposure (e.g., Lutchmaya et al., 2004; Manning, 2002; Okten et al., 2002). A lower ratio (smaller 2D than 4D) is indicative of relatively higher prenatal testosterone than oestrogen levels. Correspondingly, males have been found to show lower ratios than females in both animal (Brown et al., 2002) and human studies (e.g., Manning, 2002; Manning et al., 2007). Moreover, digit ratio has been related to various sexually dimorphic variables including cognitive abilities (Bull and Benson, 2006), sexual orientation (Lippa, 2003), aggressive and assertive behavior (Manning and Fink, 2008; Hampson et al., 2008), nociception (e.g., Schwerdtfeger and Heer, in press), and sporting skills (Pokrywka et al., 2005; Manning et al., 2007), although effect sizes seem to be moderate to low and there are several inconsistent findings (e.g., van Anders and Hampson, 2005; Puts et al., 2008).

Inconsistent findings might be attributed to attenuated measurement precision, which may bias statistical tests of significance leading researchers to draw flawed conclusions. For example, random error may inflate error variance associated with a certain measurement possibly obscuring group differences (e.g., sex differences in digit ratio). Furthermore, random error impacts correlation-based univariate and multivariate statistical methods as well. Fleiss and Shrout (1978) cogently argue that relationships among variables (e.g., correlations of digit ratio and traits) may be attenuated and even reversed in direction. Because effect sizes in digit ratio research are usually small to moderate (Manning, 2002), striving for high measurement precision seems to be a prudent objective.

Measurement precision has been under scrutiny by digit ratio researchers for a while. A recent study which relied on caliper and ruler to indirectly measure finger lengths suggests good measurement agreement across different observers (Voracek et al., 2007). These authors report an intraclass correlation coefficient (ICC) of .75 across 17 observers who measured the digit ratio of 50 individuals which represents an acceptable *interrater* agreement. Other studies report good to excellent *intrarater* agreement for the caliper method as well (e.g., Csathó et al., 2003). Although measurement precision of the caliper method is now well-established in digit ratio research, it should be noted that other alternative methods are available that need to be evaluated against this methodology.

****

Insert Table 1 about here

****

In order to get an overall impression of the methods used we performed a literature search using scientific search engines (Web of Science, Pubmed/Medline, Psyndex, British Library Online Contents), thereby scanning for several relevant keywords (digit ratio, finger length, 2D4D, 2D:4D) in the title. We identified 178 studies and 119 of these met our inclusion criteria (human studies with accurate description of the method used; see Table 1). Further analyses revealed that the vast majority of studies used a ruler, caliper, or software to quantify finger lengths. Although many studies report direct measurements of digit lengths, the indirect method relying on photocopies or scanned images of the hands seem to predominate (Table 2), probably because of the superior measurement precision (e.g., Voracek and Dressler, 2006; Voracek and Offenmüller, 2007). Other methods are available that appear promising, but await wider use to assess their comparability with already established methods in future studies (e.g., the tube method; Nicholls et al., 2008; hand outline drawings; Bang et al., 2005; see Voracek and Dressler, 2006 for a comparison of direct and indirect methods; self-report direct measurement; Caswell, & Manning, in press).

Our aim was to directly compare indirect methods of digit ratio measurement (ruler, caliper, and software) with respect to their relative measurement economy (time taken for measurement), and measurement precision (agreement across raters and interrater error).

****

Insert Table 2 about here

****

Method

From a larger sample of undergraduate students at a German university who participated in prior studies (*N* = 193), we randomly selected scanned images of 30 males' and 30 females' hands (age: *M* = 23, *SD* = 4). We included students who provided written informed consent and did not report any injuries of the digits. Hands were scanned with a HP scanner (Scanjet 3500C, resolution 300dpi). Participants were instructed to place their hands on the scanner, slightly spreading their digits. A paper ruler was placed on the scanner surface as well in order to properly scale scanned and printed images afterwards. Scanned images were converted to black and white by an appropriate filter of the Adobe Photoshop® software to eliminate a possible influence due to differences in color between scanned images and photocopies. Converted images were printed in a single printing job on a Dell Laser Printer (model 3100CN) with a resolution of 300dpi.

Digits were measured by three independent raters (VK, TK, AO) who completed a research internship in our department. Raters were thoroughly instructed concerning landmark identification and measurement methods. Identification was trained and results discussed using five problematic sample cases. Thus, raters were trained but had never measured finger length before. Lengths of the second and fourth digit of the right- and left-hand were measured from the ventral proximal crease of the digit to the tip (cf. Voracek et al., 2007). Each rater used four different methods of digit ratio measurement: plastic ruler, vernier caliper, and the computer software Adobe Photoshop® (Adobe Systems Inc., San Jose, USA) and Autometric (DeBruine, 2004), respectively. To control for training effects, the sequence of methods used was balanced across raters for ruler (AO, VK, TK), caliper (VK, TK, AO), and Photoshop (TK, AO, VK). Autometric was used as the last method by all raters because we did not learn about the freeware program Autometric, which was specifically written for the purpose of digit ratio measurement, until the study began. All raters completed the four measurement series within the same two day period (cf. Norton and Olds, 1996). Time spent for measurement was recorded for each method. The sequence of hands to be rated was randomized for each rater-method combination. A semi-transparent plastic ruler and a digital vernier caliper were used to obtain digit measures from the photocopies, accurate to 0.5 mm and 0.01 mm, respectively. To measure digit length in Photoshop, the built-in *measure tool* was used, accurate to 0.1 mm. Instead of using the metric system, Autometric indicates the length of the digits in units of pixels and automatically calculates the digit ratio for each hand after the rater manually sets the landmarks.

For the other methods, digit ratios were calculated by dividing the length of the index finger (2D) by the length of the same-hand ring finger (4D) in SPSS. All statistical analyses were conducted using SPSS 15.0® and Microsoft Excel®. In order to compare measurement methods, three widely used precision estimates were calculated (cf. Weinberg et al., 2005): (a) the intraclass correlation coefficient (ICC), (b) the technical error of measurement (TEM), and (c) the relative technical error of measurement (rTEM). Interrater agreement of digit ratio measurements was estimated with a type of single-score intraclass correlation coefficient, employing a two-way mixed-effects model with absolute agreement definition (ICC Case 3 according to McGraw and Wong, 1996, cf. also Shrout and Fleiss, 1979). TEM was computed as the square root of the squared difference between two corresponding measurements divided by twice the sample size (Dahlberg, 1940). TEM is interpreted as the typical magnitude of error associated with a certain measurement and can be used to estimate intrarater or interrater precision. The TEM can be considered a joint function of the characteristics of the measurement method (e.g., possibilities for accurate landmark identification, gradation of measurement units) as well as the rater who applies the method (e.g., skill). The rTEM is calculated by dividing TEM for a given variable by the grand mean for that variable and multiplying the result by 100. Thus, rTEM represents an estimate of error magnitude as a percentage of object size.

Results

Measurement economy. Raters using Photoshop spent on average 121 minutes (*SD* = 15) on digit measurement of 60 (left and right) hands. Less time was spent using caliper (*M* = 96, *SD* = 33) and ruler (*M* = 79, *SD* = 12), respectively. Least time for digit ratio measurement was spent using Autometric (*M* = 74, *SD* = 10).

Measurement precision.

Estimates of measurement precision are depicted in Table 3. Highest agreement between raters was achieved with computer-based methods of measurement (ICC > .90). Lower ICCs were found for caliper measurements. Agreement was lowest when using plastic ruler (ICC < .80). When taking confidence intervals into account, likely values for true ICCs may reach up to .95 on the upper bound (Photoshop) or down to .67 on the lower bound (ruler). Concerning interrater error, TEM (rTEM) ranged between 0.0088 mm (0.91% of digit ratio) and 0.0157 mm (1.63%). By using software, we obtained the lowest interrater error as estimated by rTEM (approx. 1%). For caliper measurement error was slightly higher (approx. 1.2%). Highest interrater error in our study was observed for measurement with plastic ruler (approx. 1.6%).

****

Insert Table 3 about here

****

Discussion

In most digit ratio studies the question of measurement economy is of minor importance because sample sizes are usually small to moderate. However, large samples were also reported in digit ratio research (e.g., Lippa, 2003, *N* = 2084). For these studies, time spent on digit ratio measurement may be taken into account besides precision. In our study, measurement of digits with Photoshop took the most time. Apparently, additional functions available with software compared to the other methods, e.g. zooming, take their toll. About 25 - 35% less time was spent using caliper and ruler. Least time was spent using Autometric. However, Autometric was used as the last method by all raters, making it difficult to directly compare it with Photoshop, caliper, and ruler, which were counterbalanced across raters. Taking the number of work steps into account, measurement with Autometric is presumably somewhat faster than with Photoshop. About 10 minutes have to be added to the time taken for caliper und ruler measurements due to sorting of prints (if order of measurements is to be randomized) and an unknown amount of time (depending on the hardware and software used) for proper scaling of printouts. When using software and digital images, neither time for sorting nor for scaling/printing is necessary. Thus, in small to moderate samples, differences in economy among the methods are negligible. However, in large samples, time spent for measurement may be worth a closer look.

To compare digit ratio measurement methods with respect to precision, we calculated three widely used precision estimates, the ICC (interrater agreement) and TEM/rTEM (interrater error). By measuring digit ratio with software, we achieved the highest interrater agreement of all methods used. Slightly lower ICCs were found for caliper measurements. Agreement among raters was lowest when they used plastic ruler to measure digit ratios. These results are generally consistent with prior research indicating the upper bound of agreement at .99 for computer-based methods (e.g., Pokrywka et al., 2005), the lower bound at about .73 for plastic ruler (e.g., Lippa, 2003) and ICC for the caliper method somewhere in between (e.g., Fink et al., 2006; Voracek and Dressler, 2006; Voracek and Offenmüller, 2007). Basically, all methods compared in our study had ICCs above .75 suggesting the potential for sufficient interrater agreement (cf. Shrout and Fleiss, 1979). However, taking the confidence intervals of ICCs into account, interrater agreement of measurements by plastic ruler may in some studies be insufficient.

A similar picture emerged for interrater error associated with methods compared. The TEM/rTEM was lowest for the computer-based methods, slightly higher for caliper and highest for ruler measurements suggesting superior precision for Photoshop and Autometric and lowest precision for plastic ruler. Our precision estimates roughly compare to those reported by Voracek and colleagues (2007, TEM: 0.015 - 0.020, rTEM: 1.5% - 2.0%), corroborating the validity of our results. However, precision estimates in the study of Voracek et al. were somewhat lower because these authors put measurement precision to a stronger test by aggregating TEMs/rTEMs across 17 raters. In addition, raters used different methods varying in accuracy (0.01 – 0.5 mm). Basically, all methods in our study had rTEMs below 2%, which is considered an acceptable amount of error for unskilled raters in anthropometry (Norton and Olds, 1996). For computer-based methods, we even obtained rTEMs around 1%, which Norton and Olds report for skilled anthropometrists (rTEM < 1.5%).

As the TEM/rTEM can be considered a joint function of the raters' and methods' characteristics and our raters had no prior experience in anthropometric measurement, we attribute the lower error to beneficial characteristics of computer-based measurement methods, i.e., to a more precise identification of the proximal finger crease. Contrary to the ruler or caliper method – which rely on printed images of the hands – computer software offers the advantage of zooming into the picture or adjusting contrast or color to improve identification of landmarks. Another characteristic of the measurement method is the gradation of the measurement. Interestingly, TEMs/rTEMs were lower for the software methods even though the caliper measures digits with up to ten times higher accuracy (accurate to 0.01 mm) compared to Photoshop (accurate to 0.1 mm) or Autometric (accuracy is dependent on resolution; 0.8 mm with 300 pixels per inch in our study).

Before condensing the results of our study to conclusions, three limitations have to be addressed. Our study was the first to systematically compare different (indirect) methods of digit ratio measurement. Thus, our results are preliminary pending replication. Including other methods in future studies on measurement precision, e.g., direct caliper measurement (e.g., Voracek and Offenmüller, 2007), hand outline drawings (e.g., Bang et al., 2005), or measurement by tube (Nicholls et al., 2008) is highly eligible. Furthermore, we used unskilled raters in our study which is uncommon in digit ratio research. However, the possible impact on precision was most likely marginal because (a) raters were unskilled but not untrained, and (b) interrater error of our raters closely resemble interrater error of skilled raters (experts in the field of digit ratio research; rTEM: M = 1.64, SD = 0.45) in other studies (Voracek et al., 2007). Hence, these findings suggest that raters with no prior experience can sometimes be as precise as skilled raters in measuring digit ratio. Eventually, the question of measurement economy of Autometric remains unclear because of the study design. Thus, our conclusions concerning economy of methods are limited to Photoshop, ruler, and caliper.

Arriving at a conclusion, our study suggests that the compared indirect methods used to quantify digit length have acceptable precision. In the light of typically only small or medium effects sizes in digit ratio research, researchers should think twice before employing a measurement method. When time constraints are crucial, the caliper method seems to offer the best choice due to its acceptable measurement precision. In spite of advantages in measurement economy, the ruler method may in some studies be a suboptimal choice due to relatively low measurement precision. Importantly, if associations with other measures are expected to be low or sample size is rather moderate, we would suggest the method with the highest precision estimates, namely, software. Several commercial and freeware programs are currently available which may be used for this purpose, e.g., Autometric, Gimp, Adobe Photoshop® or Corel Draw®.

Acknowledgments

The authors express their appreciation to Sandra Mai, Volker Keuch, Amara Otte, and Tim Kuhlmann, for their assistance with data collection and analysis.

Literature cited

Bang AK, Carlsen E, Holm M, Petersen JH, Skakkebæk NE, Jørgensen N. 2005. A study of finger lengths, semen quality and sex hormones in 360 young men from the general Danish population. Hum Reprod 20:3109–3113.

Brown WM, Finn CJ, Breedlove, SM. 2002. Sexual dimorphism in digit-length ratios of laboratory mice. Anat Rec 267:231–234.

Bull R, Benson PJ. 2006. Digit ratio (2D:4D) and the spatial representation of magnitude. Horm Behav 50:194–199.

Caswell N, Manning JT. in press. A comparison of finger 2D:4D by self-report direct measurement and experimenter measurement from photocopy: Methodological issues. Arch Sex Behav.

Cohen J. 1992. A power primer. Psychol Bul 112:155–159.

Csathó A, Osváth A, Bicsák E, Karádi K, Manning J, Kállai J. 2003. Sex role identity related to the ratio of second to fourth digit length in women. Biol Psychol 62:147–156.

Dahlberg G. 1940. Statistical methods for medical and biological students. London: G. Allen & Unwin Ltd.

DeBruine LM. 2004. AutoMetric software for measurement of 2D:4D ratios, from www.facelab.org/debruine/Programs/autometric.php

Fink B, Neave N, Laughton K, Manning JT. 2006. Second to fourth digit ratio and sensation seeking. Pers Indiv Differ 41:1253–1262.

Fleiss JL, Shrout PE. 1978. The effects of measurement errors on some multivariate procedures. Am J Public Health 67:1188–1191.

Hampson E, Ellis CL, Tenk CM. 2008. On the relation between 2D:4D and sex-dimorphic personality traits. Arch Sex Behav 37:133–144.

Lippa RA. 2003. Are 2D:4D finger-length ratios related to sexual orientation? Yes for men, no for women. J Pers Soc Psychol 85:179–188.

Lutchmaya S, Baron-Cohen S, Raggatt P, Knickmeyer R, Manning JT. 2004. 2nd to 4th digit ratios, fetal testosterone and estradiol. Early Hum Dev 77:23–28.

Manning JT, Bundred PE, Newton DJ, Flanagan BF. 2003. The second to fourth digit ratio and variation in the androgen receptor gene. Evol Hum Behav 24:399–405.

Manning JT, Fink B. 2008. Digit ratio (2D:4D), dominance, reproductive success, asymmetry, and sociosexuality in the BBC Internet Study. Am J Hum Biol 20:451-461.

Manning JT, Morris L, Caswell N. 2007. Endurance running and digit ratio (2D:4D): implications for fetal testosterone effects on running speed and vascular health. Am J Hum Biol 19:416–421.

Manning JT. 2002. Digit ratio: a pointer to fertility behavior and health. New Brunswick: Rutgers University Press.

McGraw KO, Wong SP. 1996. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1:30–46.

Nicholls MER, Orr CA, Yates MJ, Loftus AM. 2008. A new means of measuring index/ring finger (2D:4D) ratio and its association with gender and hand preference. Laterality 13:71-91.

Norton K, Olds T. 1996. Anthropometrica: a textbook of body measurement for sports and health courses. Sydney: UNSW Press.

Pokrywka L, Rachoń D, Suchecka-Rachoń K, Bitel L. 2005. The second to fourth digit ratio in elite and non-elite female athletes. Am J Hum Biol 17:796–800.

Okten A, Kalyoncu M, Yariş N. 2002. The ratio of second- and fourth-digit lengths and congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Early Hum Dev 70:47–54.

Puts DA, McDaniel MA, Jordan CL, Breedlove SM. 2008. Spatial ability and prenatal androgens: meta-analyses of congenital adrenal hyperplasia and digit ratio (2D:4D) studies. Arch Sex Behav 37:100–111.

Schwerdtfeger A, Heer J. in press. Second to fourth digit ratio (2D:4D) of the right hand is associated with nociception and augmenting-reducing. Pers Indiv Differ.

Shrout PE, Fleiss JL. 1979. Intraclass correlations - uses in assessing rater reliability. Psychol Bul 86:420–428.

van Anders SM, Hampson E. 2005. Testing the prenatal androgen hypothesis: measuring digit ratios, sexual orientation, and spatial abilities in adults. Horm Behav 47:92–98.

Voracek M, Dressler SG. 2006. High (feminized) digit ratio (2D:4D) in Danish men: A question of measurement method? Hum Reprod 21:1329-1331.

Voracek M, Manning JT, Dressler SG. 2007. Repeatability and Interobserver error of digit ratio (2D : 4D) measurements made by experts. Am J Hum Biol 19:142–146.

Voracek M, Offenmüller D. 2007. Digit ratios (2D:4D and other) and relative thumb length: a test of developmental stability. Percept Mot Skills 105:143-152.

Weinberg SM, Scott NM, Neiswanger K, Marazita ML. 2005. Intraobserver error associated with measurements of the hand. Am J Hum Biol 17:368–371.

Weis SE, Firker A, Hennig J. 2007. Associations between the second to fourth digit ratio and career interests. Pers Indiv Differ 43:485–493.

Tables

Tab. 1: Results of literature search.

Literature Search	178
No (new) measurement[1]	18
Non-human data	12
Non-english publication	1
Incomplete description of method[2]	21
Not available[3]	7
Reviewed studies	119

Notes. [1] = Meta-Analysis, Comment, Reanalysis, Review.

[2] = e.g., Abstract. [3] = e.g., Dissertation, Diploma Thesis.

Tab. 2: Frequency of use of different digit ratio measurement methods.

		f(x)	%[2]
Software		20	16.8
Caliper	direct	41	34.5
	indirect[1]	57	47.9
Ruler	direct	1	0.8
	indirect[1]	12	10.1
Self-assessed		8	6.7
Other		2	1.7

Notes. $N = 119$. [1] photocopy, scan/print or x-ray.

[2] percentages do not add up to 100 because several studies (16%) used more than one method.

Tab. 3: Precision estimates for different digit ratio measurement methods.

Method	Hand	ICC	95% CI	TEM	rTEM
Caliper	right	.87	.78-.92	0.0121	1.27%
	left	.89	.82-.93	0.0107	1.12%
Ruler	right	.77	.67-.85	0.0157	1.63%
	left	.79	.70-.86	0.0156	1.61%
Photoshop	right	.92	.87-.95	0.0093	0.97%
	left	.92	.88-.95	0.0087	0.91%
Autometric	right	.90	.81-.94	0.0102	1.06%
	left	.91	.86-.94	0.0088	0.91%

Notes. ICC = intraclass correlation coefficient, CI = confidence interval, TEM = technical error of measurement (averaged across raters) in units of 2D:4D, rTEM = relative technical error of measurement.