T

The Digit Ratio: Scientific Methodological Challenges and Controversies

Šebánková Blanka and Jaroslav Flegr
Department of Philosophy and History of Science, Faculty of Science, Charles University, Prague, Czech Republic

Synonyms

2D:4D; Digit ratio; Finger ratio; Manning's index; Ring-to-index ratio; Second-to-fourth ratio

Definition

The digit ratio, also known as the 2D:4D ratio or Manning's index, expresses the length ratio between the second (index) and fourth (ring) finger. This ratio is considered an indicator of the concentration of steroid sex hormones to which an individual was exposed during the prenatal period.

Introduction

Men and women differ in a multitude of characteristics. Sex differences in the structure and function of the human body are evident from the fetal age and become more pronounced during puberty. The effects of sex steroid hormones, such as testosterone and estrogen, along with other factors influencing gene regulation in both prenatal and postnatal periods, contribute to the differentiation of sexual phenotypic traits. Sexual dimorphism is manifested, for example, by differences in body size, the ratio of muscle to fat, and bone robustness. Intersex differences can also be seen in the size and shape of the hands. The ratio of the length of the second finger to the fourth finger typically has a lower value in men compared to women. In the European population, this ratio ranges from about 0.96 to 1. Men usually have a longer fourth finger than the second, whereas women tend to have a similar length of fourth and second finger. This characteristic is known to vary considerably between populations. However, the finger length ratio has also been studied, albeit with less intensity, in other animal species (for a review, see Voracek & Loibl, 2009).

Interest in this topic emerged in the second half of the nineteenth century. At the time, scientific debate among anatomists commenced, with the observation that intersexual and interpopulation variations existed in the morphology of the human hand (for a review, see Peters et al., 2002). The focus on the digit ratio was likely also influenced by the parallel publications of the first observations of a shorter index finger in primates. During the post-Darwinian controversy spurred by debates over human origins, the digit ratio served as a tool to support theories of racial differentiation based on evolutionary grounds. In the early twentieth century scientists became

T. K. Shackelford (ed.), *Encyclopedia of Sexual Psychology and Behavior*,
https://doi.org/10.1007/978-3-031-08956-5_780-1

convinced that an elongated second digit was a sign of evolutionary divergence from subhuman primates.

In the latter half of the twentieth century, collective findings in physiology, endocrinology, and neurology demonstrated that prenatal steroid sex hormone concentrations have a profound impact on fetal development. Given the discovery of gene regulation of steroid receptors affecting growth cartilage proliferation in phalanges, the notion that prenatal sex hormone fluctuations influence the length of the fourth digit, and consequently, that the 2D:4D ratio indirectly reflects sex hormone concentrations during fetal development, has gained widespread acceptance (Zheng & Cohn, 2011).

The 2D:4D ratio is sometimes referred to as Manning's index. This publicly known term started to appear in the literature after the first major study by J. Manning in 1998, which demonstrated an association between a higher 2D:4D ratio and poorer sperm quality, reduced levels of serum testosterone, and luteinizing hormone. In recent decades, the 2D:4D ratio has been the subject of intense research. Numerous empirical studies have shown that this ratio correlates with a range of sexually dimorphic traits (see rev. years 1998–2008 in Voracek & Loibl, 2009; see rev. years 2010–2021 in Sorokowski & Kowal, 2023; also see the chapter "Testosterone: Digit Ratio (2D:4D)"). For example, studies have indicated that individuals with a lower 2D:4D ratio perform better in spatial orientation, are more aggressive, have a greater propensity for risk-taking, are more successful in sports, and are more resistant to anxiety. However, individuals with a lower ratio have been shown to be more likely to suffer from certain diseases, such as breast cancer, cardiovascular problems, autism, and infertility. Conversely, a higher 2D:4D ratio is associated with better verbal abilities, empathy, and altruism. A broad area of interest is population differences and also the relationship of this indicator to sexual orientation.

The hypothesis that the digit ratio reflects prenatal sex hormone concentrations remains controversial and has been subject to substantial criticism (Leslie, 2019; Lolli et al., 2017; McCormick & Carré, 2020; Swift-Gallant et al., 2020). A number of published findings have failed to replicate, and many experts have raised fundamental objections to the methodology used. In many fields, though not universally in biomedicine, there is a growing consensus that using ratios as traits is inherently inappropriate and often yields misleading results (Curran-Everett, 2013). Specifically, in the field of 2D:4D ratio research, several researchers have repeatedly pointed out that due to the different allometric relationships between hand size and the lengths of the second and fourth fingers, the digit ratio merely reflects hand size, albeit imperfectly. Consequently, most observed effects attributed to the 2D:4D ratio are actually effects of hand size, which are likely indicative of overall body size effects as well (Lolli et al., 2017).

The aim of this chapter is to provide an overview of the main methodological and conceptual problems that are typical for research in this area and to offer possible solutions.

Validity of Using the 2D:4D Ratio as a Marker of Prenatal Sex Hormone Concentration

Based on the existence of a correlation between the 2D:4D ratio and sexually dimorphic traits, it has been proposed that the digit ratio can be considered an indicator of the degree of prenatal sexual differentiation, specifically the degree of masculinization that occurs during embryonic development under the influence of sex hormones (Manning et al., 1998). If the 2D:4D ratio were to serve this purpose, the validity of this method would first have to be verified. In other words, we need to be reasonably sure that the method (in this case, measuring the ratio of finger lengths) actually measures what we believe it measures (in this case, the prenatal concentration of sex hormones). There is definitely no consensus in the scientific community today on this issue, not even on its sub-questions.

First and foremost, it is not certain whether differences in concentrations of sex hormones are the primary cause of variations in the lengths of individual fingers among individuals. This significant sub-question will be the subject of a separate chapter 7. It remains uncertain the extent to which postnatally measured finger length differences mirror prenatal conditions. It is also unclear how the measured length of fingers relates to the real length of fingers. This is because it depends not only on the length of the respective bones but also on the morphology of the soft tissues, and even on the behavior of the individual throughout their life and during the measurement itself (see chapter 3).

The length of fingers in adulthood depends not only on their development during the prenatal period but also on a range of genetic, epigenetic, and environmental factors that affect an individual after birth. These factors include, for example, physical activity, nutrition, sleep quality, as well as internal factors such as left or right hand preference, both congenital and acquired lateral asymmetry, age, ethnic origin, birth order, and certain infections. If finger growth is indeed influenced by the levels of sex hormone concentrations, there is no reason to assume that this only occurs during the prenatal period. A sudden change in hormone concentrations during puberty could completely erase the existing differences in finger lengths that originated in the prenatal period of an individual's development. Any temporary or permanent fluctuations in sex hormone concentrations, if they occur before the end of growth, can lead to a change in the 2D:4D ratio. It does not matter whether the increase or decrease in sex hormone levels is natural or artificial. In the first case, it may be in relation to a change in sexual behavior —either a decrease or increase in its intensity. Or, it could be consequence, for example, of parasitic infection that increases steroid hormone concentrations to achieve suppression of the host's immune response or to induce behavioral changes that facilitate the transmission of the infection to other uninfected individuals of the host species (Box 1). In the second case, for example, changes in hormone concentrations may be induced intentionally or unintentionally as part of treatment, e.g., immunosuppression or chemotherapy.

The difference in the 2D:4D ratio may be caused by both bone growth and variations in the volume and spatial distribution of soft tissues. Prenatal and postnatal changes in the digit ratio could be due to a higher density of testosterone receptors in the soft tissue of the fourth finger. The volume and distribution of these tissues change over a person's lifetime, and the corresponding values fluctuate depending on the individual's age and weight. Among other factors, these values depend on fat distribution, which is also influenced by the ratio of testosterone to estradiol (Wallen, 2009). In women, the fluctuation of these values depends on the phase of the menstrual cycle (Mayhew et al., 2007). Soft tissue distribution differs between the left and right hands, which might explain why some human characteristics correlate better with the ratio on the left hand, despite the fact that most characteristics correlate with the digit ratio on the right hand (Manning et al., 1998; Flegr et al., 2005). Finger ratios are also likely to depend on how a person uses their hands throughout their life (Richards et al., 2021). For example, the correlation between the 2D:4D ratio and certain traits may depend on which hand is dominant for that person.

The vast majority of studies concerning the relationship between hormones and the 2D:4D ratio have been cross-sectional in nature. The problem with cross-sectional studies is that they do not allow us to determine what is cause and what is effect. It is evident that the finger length in an adult cannot influence hormone concentrations during their prenatal development. However, it cannot be automatically inferred that prenatal hormone concentrations are the cause of changes in finger length and thus in the 2D:4D ratio. It is possible, and quite plausible, that both finger length and hormone concentrations during prenatal development could be influenced by the action of a third, currently unknown, external or internal factor.

For example, studies focusing on the relationship between the 2D:4D ratio and the risk of developing heart disease are usually based on the idea that prenatal exposure to sex hormones affects both the digit ratio and the risk of heart disease. The specific mechanism by which prenatal hormone levels would affect the likelihood of heart disease in later life is unknown. Authors of relevant studies typically speculate on various mechanisms, such as the influence of prenatal hormones on cardiovascular development, but ultimately concede that the relationship may be influenced by a complex interplay of genetic, environmental, and lifestyle factors, which the studies cannot capture.

Therefore, it must be concluded that the question of the validity of estimating prenatal hormone concentrations based on the 2D:4D ratio remains open and definitely deserves more attention. It is clearly not possible to consider the finger ratio as a reliable and valid measure of prenatal hormone concentrations, and any conclusions drawn on the basis of such estimated concentrations should also be supported by other, completely independent data.

Box 1: Association Between 2D:4D Ratio and Infection by *Toxoplasma gondii*

Toxoplasma gondii is a parasite that sexually reproduces only within the intestinal cells of feline species, using virtually all warm-blooded animals as intermediate hosts. Transmission to the definitive host, a feline, occurs through predation. Currently, about one-third of humanity hosts the dormant stages of this parasite for life. For more information, see the chapter "*Toxoplasma* Infection." Infected individuals differ from non-infected individuals in a range of behavioral and morphological differences, many of which deepen over time since infection. One such difference is the 2D:4D ratio. Infected individuals have a lower 2D:4D ratio on the left hand, but not on the right hand, compared to uninfected individuals; however, this difference is significant only in men. The value of the 2D:4D ratio also negatively correlates with the level of specific anti-*Toxoplasma* IgG antibodies, which may reflect the intensity of the infection (Flegr et al., 2005). Studies on toxoplasmosis indirectly confirm that there is a relationship between the 2D:4D ratio and testosterone levels, specifically postnatal testosterone levels. Based on experiments with infections conducted on laboratory animals and observational studies conducted on humans, it has been found that the proximate cause of many, though not all, observed phenotypic changes associated with toxoplasmosis is indeed increased testosterone levels in infected individuals (Tan & Vyas, 2016; Flegr et al. 2008). Although this increase has not been demonstrated in women, it is possible that they have similarly elevated estrogen levels, which correlate with testosterone levels.

Toxoplasma infection is likely a significant confounding factor affecting the results of 2D:4D ratio studies. When statistical models included whether an individual was infected or not, the strength of the association between the 2D:4D ratio and the sex of the individual increased significantly, and the otherwise rarely demonstrated association between digit ratio and age of the individual appeared (Flegr et al., 2008). However, the same study did not demonstrate a correlation between testosterone levels and the 2D:4D ratio.

Research on the model of human toxoplasmosis has brought several insights of broader significance. Perhaps the most important is that the human 2D:4D ratio can change under the influence of environmental factors even in adulthood, particularly affecting the left hand. It also showed that it is appropriate to filter out a number of factors, including latent *Toxoplasma* infection, when studying the influence of the 2D:4D ratio. The prevalence of toxoplasmosis varies significantly between countries, and this may contribute to the between-country variability of the 2D:4D ratio.

Sensitivity and Reproducibility of the Methods of Finger Length Measurement

A frequently discussed issue that can affect the results of studies, and therefore the possibility of their interpretation, is the accuracy of measuring finger lengths. This concerns not only how precisely finger lengths can be measured (method sensitivity) but also how consistent the value obtained are when measured repeatedly over time by the same measurer or when the same measurements are made by different measurers (method reproducibility).

Finger length is measured from the proximal crease between the palm and the finger segment to the fingertip. The methods for measuring fingers vary depending on the type of study and can be divided into direct measurement on the palm of the hand and indirect measurement using an image of the hand. Direct measurement of finger lengths can be done by the individuals themselves, typically using a standard ruler. If the measurement is conducted by a researcher, a more precise caliper is usually used. Measurement by a researcher is suitable for studies with a small number of participants and sufficient time for measurement. Self-measurement is particularly suitable for large samples or hard-to-reach groups. In such cases, the typically lower accuracy of measurement is compensated by including the largest possible number of subjects in the study. As the number of participants increases, the accuracy of estimating the average finger lengths in each group improves. However, the imprecision of measurement still leads to a large variation in measurements, thus there is a risk, despite solid estimates of averages, that observed differences will not appear significant. Where possible, it is always preferable to prioritize measurement over self-measurement.

Indirect measurement of bone lengths using X-ray imaging is less common but possibly the most accurate method. This technique is advantageously used in studies on animals. For human studies, however, it is not generally used purely for research purposes to avoid exposing participants to radioactive radiation.

The most frequently used method for measuring finger lengths involves indirect measurement from photographs or scans of the hand. This is suitable for large samples and offers the main advantage of speeding up data collection. The method also allows the same images to be measured repeatedly, or by two or more people independently, to refine the data. It also facilitates double-blinding of the study. In a single-blind study, the participant does not know which hypothesis is being tested during the study. In a double-blind study, this is also true for the persons conducting the measurements; moreover, they should not be aware of specifics such as the characteristics of the subjects they are measuring, whether it's a man or a woman, infected or uninfected individuals, and so forth. Naturally, it is essential that the order of images measured is randomized. It is not permissible for the measurer to measure male images first and then female, or even for male images to be measured by one person and female by another.

It is definitely not advisable to use multiple methods within a study, as the results obtained can vary significantly. For example, values of the 2D:4D ratio based on indirect measurement are usually lower than those based on direct measurement of fingers (Ribeiro et al., 2016). Even if proper randomization was performed and, for example, an equal number of men and women were measured using both methods, this would increase the variance in the data and thus the risk of a Type II error, i.e., the risk that we do not demonstrate an existing difference between the groups being compared, see below.

Error is not only in combining methods and changing measurers, but also in failing to follow a consistent measurement procedure or in failing to follow instructions on hand positioning by the participant. Another inaccuracy can arise if the participant moves their fingers during measurement and the measurer may not be able to target precisely the individual measurement points. This risk can be minimized by having the person measured sit with their hand resting on a support during the measurement. Measurement inaccuracy and its consequence – stochastic error, can lead to false-negative results of the study, i.e., failure to demonstrate a statistically significant effect of the 2D:4D ratio on the observed trait, even though such an effect actually exists. In the case of 2D:4D ratio studies, this risk is increased by the fact that the effects of the 2D:4D ratio are usually quite small, while the measurement error can be relatively large. Therefore, it is always necessary to use the most accurate measurement method available in the study and sufficiently large samples of participants.

In addition to stochastic error, a much more serious systematic error can also occur. This can be caused by a factor that acts differently on each of the groups compared, or by a factor that influences the person conducting the measurements in a certain way. Sources of systematic errors in large datasets can include indirect effects of very weak confounding variables (Bailey et al., 2016). For example, aggressiveness (readiness and willingness to use violence to resolve conflict) is a trait that typically shows a dependency on gender and reflects levels of androgens. During an experiment, a participant's aggressiveness can manifest as excessive and uneven stretching of the fingers or increased pressure intensity on the copy plate. Thus, measured gender differences in aggressiveness might actually reflect gender differences in behavior, not differences in hand morphology. Similarly, differences between homosexual and heterosexual individuals may be due to differences in behavior during measurement and may depend on whether the direct measurement of fingers or photographing/scanning of hands is performed by a man or a woman (Grimbos et al., 2010).

The risk of systematic error can be reduced by eliminating confounding variables when selecting the test sample—that is, by conducting individual studies on as homogeneous populations as possible. However, in such cases, there is a risk that the results obtained cannot be generalizable to the overall population. If we aim for the possibility of such generalization, it is necessary to conduct several studies, each time on a different homogeneous sample. If we conduct a study on a heterogeneous sample, we have the possibility of blocking the variable, for example by using a paired test. If there are multiple confounding variables, we should make efforts to randomize the assignment of participants or conditions as well as possible in experimental studies. This involves ensuring that individuals with any combination of confounding variable values are equally likely to be assigned to both the experimental and control groups. In observational studies, where individuals sort themselves into experimental and control groups, we must at least record the values of confounding variables for each individual and subsequently attempt to filter out their influence on the target variable using appropriate multivariate statistical methods.

Coping with the Challenges of Multiple Testing: Strategies and Implications in 2D:4D Ratio Research

The artifact of multiple testing is an issue that occurs when multiple hypotheses are tested within a single study or when the same hypothesis is evaluated across several data sets. The outcome of a confirmatory statistical test is a p-value, which indicates the probability of observing data as unevenly or more unevenly distributed as seen in our sample if the null hypothesis is true, i.e., due solely to random chance. For instance, this could be the probability of obtaining as large or even larger a difference in the 2D:4D ratios between men and women under the assumption that the null hypothesis is valid. However, this p-value applies only to the result of a single test. If, for example, we conducted 10 studies and one of them showed a significant difference between men and women at a significance level of 0.02, the probability of achieving such a result by chance is approximately 10 times higher than if the study were conducted only once, i.e., 20%. This probability is relatively high and a corresponding p-value of 0.2 is completely insignificant.

Similarly, if we examine whether the 2D:4D ratio correlates not just with one psychological factor, such as aggressiveness, but sequentially with the five personality factors measured in the Big Five questionnaire in a single study, and the results show a correlation with the factor of neuroticism at a significance level of 0.02, we cannot conclude that the 2D:4D ratio influences psychological traits. Considering the number of tests performed, the p-value would increase to about 0.1, rendering the result insignificant again—it could have occurred purely by chance with approximately a 10% probability.

The artifact of multiple testing in 2D:4D ratio research can also arise from researchers measuring not only the 2D:4D ratio but a variety of anthropometric parameters of the hands. Typically, this includes the ratio of finger lengths on the right hand, left hand, the ratio of corresponding fingers on the right and left hands, or the average of the ratios of the right and left hands. If one of these parameters does not show significance, another might, which increases the risk of the artifact. In large datasets involving multiple comparisons, this is usually coincidental. This artifact manifests externally in that published studies are not consistent, for example, regarding which hand (left or right) is more determinative (see Box 2). Most importantly, the respective results cannot be reproduced with other data.

Several methods are used to address the issue of inflated significance levels arising from multiple comparisons. One such method is Bonferroni correction, a simple approach that adjusts p-values by multiplying them by the number of tests performed. However, this method's simplicity comes at a cost: it can be overly conservative, increasing the risk of Type II errors (failing to detect existing effects). For this reason, it's often recommended as a starting point rather than a definitive approach. Alternatives like stepwise Bonferroni corrections can provide more precise control over Type I errors.

If the study is exploratory, methods like Benjamini-Hochberg or Holm's procedure are often preferred. These methods consider not only the achieved p-values but also the overall distribution of results across all tests. The Benjamini-Hochberg procedure specifically allows researchers to choose an acceptable false discovery rate (FDR), the proportion of false positives they are willing to tolerate among significant findings. For exploratory studies, it is recommended to choose a value for FDR between 0.1 and 0.25. Theoretically, an FDR of 0.20 is probably the most appropriate value. At this level, the probability of a Type I error (proving a non-existent effect) is 20%. This requirement aligns with common standards in statistical testing, which typically demand that the power of the test be at least 0.8, thereby setting the probability of a Type II error (failing to prove an existing effect) also at 20%.

Corrections for multiple testing reduce the probability of a false positive result (Type I error) but at the same time increase the likelihood of a false negative result (Type II error). Moreover, it is quite difficult to perform such a correction properly in the case of 2D:4D ratio studies, because some features, such as the digit ratio on the right and left hands, are not independent. For these reasons, it is advisable to consider the positive results obtained as preliminary, it is better to replicate the study on independent data instead of applying corrections for multiple comparisons. As in other research areas, it is highly desirable to preregister the study, including the hypotheses to be tested and the planned statistical tests, before data collection begins, for example on the Open Science Framework (OSF), a free, open-source web platform. Such preregistration allows for subsequent verification of how many tests were conducted within the study and to assess the level of risk for the occurrence of multiple testing artifacts.

Box 2: Differences in the 2D:4D Ratio Among Populations and Their Relationship to Sexual Orientation

An interesting example of research on the 2D:4D ratio, which may well illustrate the problem of multiple testing, is the study of the relationship between the 2D:4D ratio and sexual orientation by Grimbos et al. (2010). The researchers based their study on the neurohormonal theory of homosexuality, which posits that sexual orientation is influenced by the amount of androgens acting on the developing embryo. Research in this area has yielded inconsistent results to date. Homosexual men in European populations have, on average, lower 2D:4D ratio (thus a more masculine profile). However, studies conducted on the North American male population show the exact opposite, where homosexual men, compared to heterosexuals, have higher or at least the same 2D:4D ratio values on average.

(continued)

Box 2: Differences in the 2D:4D Ratio Among Populations and Their Relationship to Sexual Orientation (continued)

In some studies, homosexuals were divided into two subgroups, those who prefer an active role during sexual acts ("tops") and those who prefer a passive role ("bottoms"). Although the results were in line with the authors' expectations—tops, who are behaviorally and self-perceived as more masculine, showed on average a lower 2D:4D ratio—the division of participants into subgroups doubled the number of tests conducted, significantly increasing the risk of multiple testing artifacts.

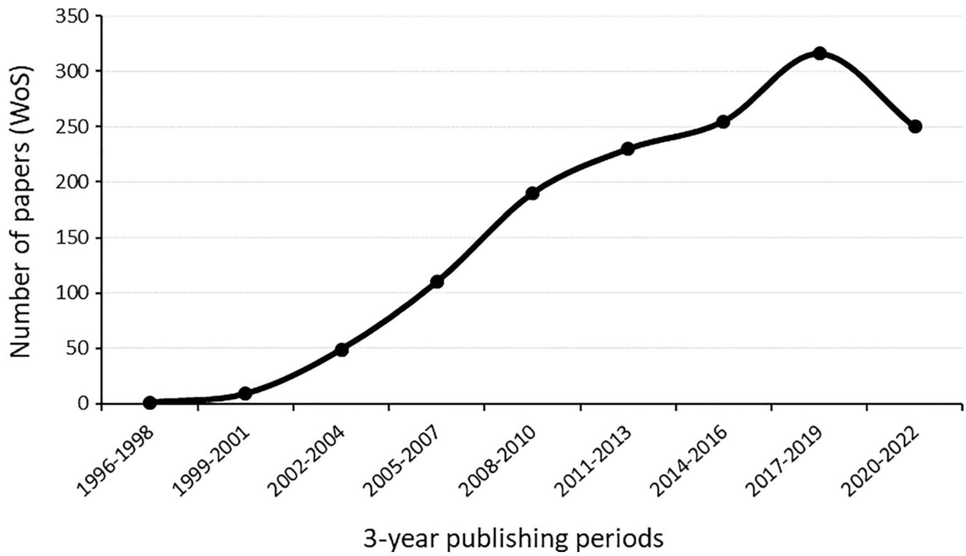
A similar inconsistency was observed in the results concerning female homosexuality. Some studies reported higher (more feminine) 2D:4D ratio in homosexual women, others reported lower (more masculine) value, and some studies found no significant differences between homosexual and heterosexual women. Among subgroups, masculine lesbians showed a lower ratio than feminized lesbians.

Studies examining the association between sexual orientation and the 2D:4D ratio also varied regarding which hand reported statistically significant results. Additionally, the differences in the 2D:4D ratio between homosexuals and heterosexuals are generally smaller than those between the sexes. Both observations suggest the possible influence of artifacts from multiple testing.

Publication Bias in 2D:4D Ratio Studies: The Pitfalls of Selective Reporting

Research on the Manning's index has enjoyed considerable popularity over the last two decades (Fig. 1). Studies addressing this topic are relatively easy, inexpensive, and attractive to a broad scientific and even non-scientific community, leading to a mass production of related studies. Researchers often repeat these studies across different populations and with various methodologies (Voracek & Loibl, 2009). This inevitably leads to a relatively high occurrence of false-positive results due to chance. The threshold for a result to be considered statistically significant is set at 5%, which means that on average, one in every 20 studies might report a false-positive result, even though the factor under examination has no impact on the variable of interest and the observed phenomenon is merely a product of chance. Authors may hesitate to finalize "uninteresting" negative study results into a finished article, especially when authors are also forced to consider the cost of publication fees. Even if the author completes the article and submits it to a scholarly journal, an article reporting negative results or results contrary to the findings of other studies has a significantly lower chance of succeeding in the peer review process than an article reporting a positive, anticipated result. Consequently, the existence of the vast majority of study results indicating the absence of a certain effect, or studies showing an effect opposite to what the author or public expects, may remain hidden from the academic community. Selective publishing of results based on self-censorship and censorship of negative or dissenting results can give readers of scholarly articles a completely false impression that a certain phenomenon undoubtedly exists, even though the opposite may be true.

In the case of 2D:4D ratio studies (inexpensive and easily conducted), it is even more crucial than in other areas not to focus on interpreting individual studies but to concentrate on the results of specific meta-analyses that include measures to minimize publication bias. Such measures include analysis of result heterogeneity, mathematical or visual detection of missing negative results (for example, by publishing a funnel plot). Perhaps the most important, though also the most demanding, is comparing the results of studies that have been published in scholarly journals with those available only in preprint archives, or unpublished and made available by authors upon personal request.



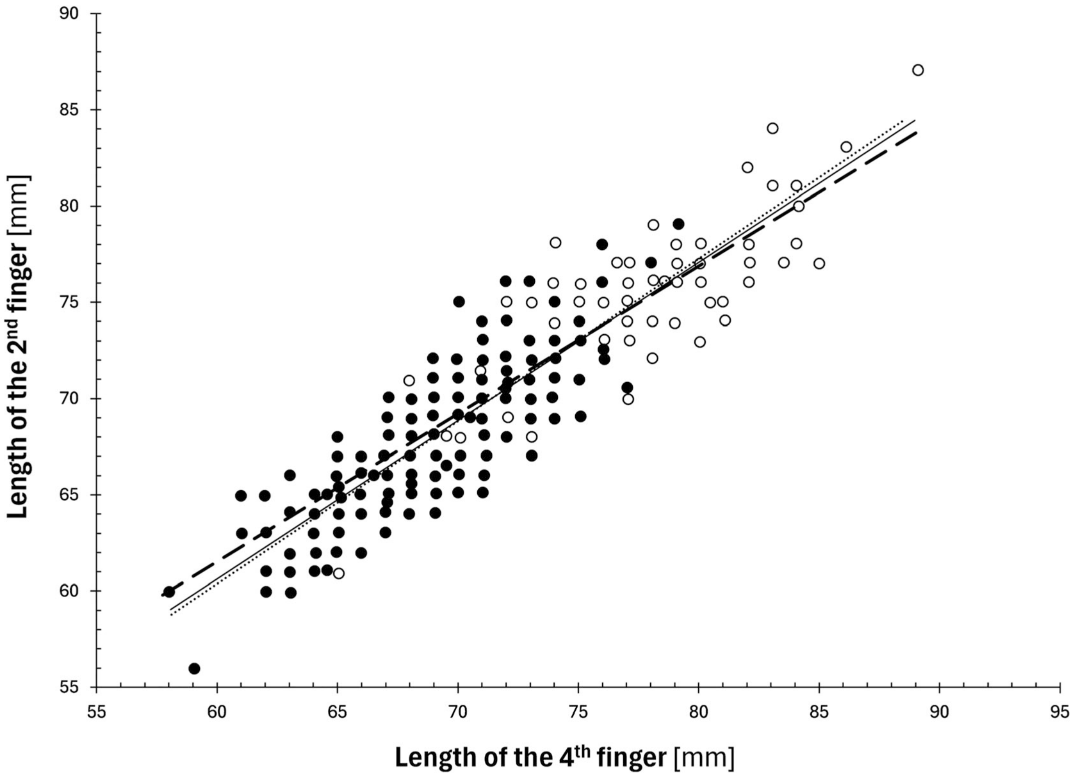
The Digit Ratio: Scientific Methodological Challenges and Controversies, Fig. 1 Number of articles in Web of Science Core Collection discussing the 2D:4D ratio. To identify articles published in each year of 1998–2022, a search was performed using the following terms: (“2D:4D”) OR (“2D4D”) OR (“digit ratio”) OR (“finger ratio”); filtering for “Article” and “Review Article” types. Articles were manually checked for relevance and those not pertinent to the topic were excluded from the count. The curve depicts a significant growth in the topic’s popularity, culminating in the period between 2017 and 2019, when numerous articles emerged that criticized the use of ratios, measurement techniques, statistical methods, and inappropriate assessments of relationships with other variables. The apparent decline in articles from 2020–2022 may reflect a delay in database indexing rather than a true decrease in publication rate

Using Ratio Measures in Research, Particularly in 2D:4D Studies

A fundamental issue is the use of ratio measures, such as 2D:4D, for statistical analysis purposes. Although using ratio measures is a relatively common practice in biomedical research (e.g., BMI, WHR), it has long been known that ratio measures have problematic mathematical properties and their uncritical use can lead to serious biases (Curran-Everett, 2013). This can be attributed, in part, to the increase in measurement error when dividing two imprecisely measured numbers, which leads to a reduction in statistical power and an increased risk of Type II error, i.e., failing to detect an existing effect (Packard, 2009).

In 2D:4D research, an even more serious problem is the fact that the relationship between the lengths of the second and fourth fingers is allometric, not isometric (Kratochvíl & Flegr, 2009). This means that fingers grow at different rates with respect to overall body size and start from different baselines, thus having different constants. In the case of plotting these relationships on an x-y graph, the existence of an allometric relationship is apparent from the fact that the line fitting the data points does not pass through the origin (Fig. 2). This is usually because both compared variables depend on a third variable, here the size of the hand, and each depends differently on this third variable. As hand size (and correspondingly body size) increases, the length of the fourth finger increases more rapidly than that of the second finger. Due to this non-linear dependency, people with larger hands have a relatively longer fourth finger and therefore a lower 2D:4D ratio than people with smaller hands. Since men are on average larger than women, this results in a lower 2D:4D ratio for men than for women (Kratochvíl & Flegr, 2009; Lolli et al., 2017).

From Fig. 2, it is also apparent that the points corresponding to the finger length ratios of men and women are distributed along the same line, indicating that the expected length of the fourth



The Digit Ratio: Scientific Methodological Challenges and Controversies, Fig. 2 Adapted from Kratochvíl and Flegr (2009). Scatterplot of the finger lengths on right hands in 297 biology students. The solid line represents the common line estimated by the ordinary least-square linear regression, with women indicated by solid points and a dotted line, and men by circles and a dashed line. The y-axis intercept of each line is significantly greater than zero ($p < 0.00001$), indicating that the 2D:4D ratio necessarily decreases with increasing finger length. Additionally, there is no significant difference between the sexes in the scaling relationship between the length of the second and fourth fingers

finger for a person with a certain finger length is the same, whether the person is a man or a woman. This conclusion was also confirmed by formal statistical analysis. When the proper statistical method was used to analyze the effect of sex on the ratio of finger lengths, namely, analysis of covariance with the dependent variable the length of the second finger and predictors sex and the length of the fourth finger the effect of sex was entirely non-significant (Kratochvíl & Flegr, 2009; Lolli et al., 2017).

However, if the same analysis of covariance is performed on a sufficiently large sample, the effect of sex may manifest. For instance, Manning (2010) demonstrated a statistically significant influence of sex even when using covariance analysis in a study involving 115,000 people. The effect size subsequently calculated from published data as Cohen's d, however, was only 0.053, which is generally considered negligible according to the established classification of effect sizes. Such a weak effect could be caused by slight bias—respondents measured their own finger lengths, and it is likely that some men and some women adjusted their values to match their expectations. In such a case, men might tend to overestimate the length of the fourth finger, and women the length of the second finger.

A solution to the problems arising from the existence of allometry is the consistent use of individual finger lengths instead of their ratio and statistical control for the influence of the fourth finger length when examining correlations of the second finger length with various characteristics. In case of positive results, it should be considered whether to study the influence of overall body size directly, which might correlate even better with some psychological traits than finger length (Lolli et al., 2017). When publishing any results related to the 2D:4D ratio, primary data should always be disclosed, specifically the measured values of individual finger lengths. The strict requirement for the disclosure of primary data, however, applies to any type of research, and reviewers of scholarly studies and journal editors should insist on it unconditionally.

Does the 2D:4D Ratio Reflect Sex Hormone Concentrations?

Individual differences in the 2D:4D ratio are often explained by exposure of the embryo to different levels of sex hormones, primarily testosterone and estrogen (Manning et al., 1998). These hormones are presumed to influence the development of fingers and other tissues. This assumption is primarily supported by data from animal models, which show the effects of these hormones on limb development (Zheng & Cohn, 2011). However, it is also supported by results from several human studies suggesting a correlation between the 2D:4D ratio and levels of testosterone and estradiol in amniotic fluid (Ventura et al., 2013). Nonetheless, research on the relationship between the 2D:4D ratio and testosterone in umbilical cord blood has not yielded any significant results, and analyzing human fetal blood solely for research purposes is practically impossible.

Attempts to demonstrate a connection between the 2D:4D ratio and polymorphisms of genes associated with androgen and estrogen signaling have produced mixed results (Voracek, 2014). A recent genomic study on a large sample of over 15,000 individuals additionally showed that the heritability of the 2D:4D ratio is relatively low and genetic correlations with sex hormone levels are weak and statistically insignificant (Warrington et al., 2018). Overall, direct evidence of a link between the 2D:4D ratio and prenatal levels of sex hormones in humans is limited and ambiguous.

Findings regarding the relationship between the 2D:4D ratio and levels of sex hormones in adulthood are similarly inconsistent. A meta-analytic study focused on this issue by Hönekopp et al. (2007) revealed only a very weak negative correlation between the 2D:4D ratio and testosterone levels in men and no correlation with estradiol levels. A significant problem with these studies is that steroid hormone levels in the same individual vary significantly over time, both throughout the year and throughout the day. Additionally, results show that especially measured levels of hormones in saliva, which are supposed to reflect the level of free testosterone in the blood, are also influenced by the method of saliva collection. Significantly different levels are measured in saliva collected through free drooling compared to saliva collected by chewing porous materials. Results obtained from a one-time sample collection may not reflect the average hormone concentration in an individual but rather the circumstances of the saliva collection. The weak or nonexistent dependency of the 2D:4D ratio on postnatal testosterone levels in most studies may be due to the difficulty of accurately measuring highly fluctuating concentrations of free testosterone throughout the day.

To verify the hypothesis of a potential postnatal influence of steroid hormones on finger lengths, it would be appropriate to study individuals using male sex hormones for gender transition or bodybuilders using anabolic-androgenic steroids for muscle growth. Studies should also consider the fluctuation of testosterone levels over the years. The impact of high testosterone levels on increasing bone volume is well known, as is the effect on premature cessation of growth due to the closing of growth plates. Additionally, consideration should be given to the fact that testosterone levels significantly change in response to fluctuations in health status affected by diseases. Examples of globally prevalent diseases include polycystic ovary syndrome, where testosterone levels increase, and type 2 diabetes, where they decrease.

These results suggest that the 2D:4D ratio may not be as strongly and specifically affected by prenatal androgens and estrogens as was originally assumed. Nevertheless, most studies rely on the default assumption of a negative correlation between testosterone levels and the 2D:4D ratio value and a positive correlation with estrogen levels.

At present, the relationship of prenatal and postnatal levels of sex hormones with 2D:4D ratio cannot be considered proven. Therefore, when interpreting the results of 2D:4D studies, it is necessary to proceed with maximum caution. It definitely cannot be automatically assumed that measuring finger lengths provides data about prenatal hormone levels, regardless of how frequently such claims appear in the scientific literature.

Conclusion

This chapter has addressed methodological and interpretive challenges concerning the studied relationship between the 2D:4D length ratio, also known as the Manning 's index, and various physical, behavioral, and cognitive characteristics. Despite the allure of the 2D:4D ratio as a tool for understanding sexual differentiation, the field must move beyond uncritical enthusiasm and address the methodological limitations hindering its progress. More rigorous and consistent research is needed to establish the true potential of this measure.

A major issue is that many observed effects likely do not directly relate to the 2D:4D ratio but reflect the influence of overall body size. Given the allometric relationship between finger lengths and hand size, an apparent correlation between the 2D:4D ratio and various traits can arise purely as a result of different body sizes, without any relation to prenatal hormone levels. Using the 2D:4D ratio as a marker for prenatal androgen exposure is further complicated by several methodological problems that can lead to both false positive and false negative results, distorting our understanding of the biological basis of sex differences. Existing evidence on the relationship between the 2D:4D ratio and both prenatal and postnatal hormone levels is weak, inconsistent, and often speculative.

Despite these serious reservations, research in this area is expected to continue using methodology that has been theoretically and practically proven to be suboptimal and potentially misleading. While many observed effects are real, their interpretation is often incorrect. Variability in the 2D:4D ratio probably does not reflect the influence of individual differences in prenatal testosterone concentration, but rather in body size or other variables.

A critical reevaluation of existing evidence, particularly with regard to the likely occurrence of a large number of false positive findings, is a fundamental prerequisite for enhancing the credibility of research in this area. Future research should primarily focus on pre-registered studies that specify in advance which factors will be studied and by which experimental and statistical methods. Given the high risk of false positives and the file drawer effect, it makes little sense to interpret published results based on individual studies; instead, it is always necessary to rely on the results of meta-analyses, particularly those that have tested for the potential existence of publication bias. Caution is also needed in interpreting existing findings and generalizing them to the level of causal developmental mechanisms.

Variability in the 2D:4D ratio is an intriguing phenomenon, and the field of 2D:4D ratio research certainly warrants further exploration. Future research should rely on methodologically robust studies conducted on specific populations with experimentally manipulated or naturally elevated levels of androgens and estrogens. Only through such focused investigations will it be possible to reliably confirm or refute the hypothesis that the 2D:4D ratio serves as a marker of prenatal and postnatal hormone exposure and to elucidate the specific mechanisms influencing the development and variability of this trait.

Cross-References

- ▶ Testosterone: Digit Ratios (2D:4D)
- ▶ *Toxoplasmosa* Infection

References

Bailey, J. M., Vasey, P. L., Diamond, L. M., Breedlove, S. M., Vilain, E., & Epprecht, M. (2016). Sexual orientation, controversy, and science. *Psychological Science in the Public Interest, 17*(2), 45–101. https://doi.org/10.1177/1529100616637616

Curran-Everett, D. (2013). Explorations in statistics: The analysis of ratios and normalized data. *Advances in Physiology Education, 37*(3), 213–219. https://doi.org/10.1152/advan.00053.2013

Flegr, J., Hrušková, M., Hodný, Z., Novotná, M., & Hanušová, J. (2005). Body height, body mass index, waist-hip ratio, fluctuating asymmetry and second to fourth digit ratio in subjects with latent toxoplasmosis. *Parasitology, 130*(6), 621–628. https://doi.org/10.1017/s0031182005007316

Flegr, J., Lindová, J., & Kodym, P. (2008). Sex-dependent toxoplasmosis-associated differences in testosterone concentration in humans. *Parasitology, 135*(4), 427–431. https://doi.org/10.1017/S0031182007004064

Grimbos, T., Dawood, K., Burriss, R. P., Zucker, K. J., & Puts, D. A. (2010). Sexual orientation and the second to fourth finger length ratio: A meta-analysis in men and women. *Behavioral Neuroscience, 124*(2), 278–287. https://doi.org/10.1037/a0018764

Hönekopp, J., Bartholdt, L., Beier, L., & Liebert, A. (2007). Second to fourth digit length ratio (2D:4D) and adult sex hormone levels: New data and a meta-analytic review. *Psychoneuroendocrinology, 32*(4), 313–321. https://doi.org/10.1016/j.psyneuen.2007.01.007

Kratochvíl, L., & Flegr, J. (2009). Differences in the 2nd to 4th digit length ratio in humans reflect shifts along the common allometric line. *Biology Letters, 5*(5), 643–646. https://doi.org/10.1098/rsbl.2009.0346

Leslie, M. (2019). The mismeasure of hands?: Some researchers say a simple ratio of finger lengths can predict personality and health. Others say that's wishful thinking. *Science, 364*(6444), 923–925. https://doi.org/10.1126/SCIENCE.364.6444.923

Lolli, L., Batterham, A. M., Kratochvíl, L., Flegr, J., Weston, K. L., & Atkinson, G. (2017). A comprehensive allometric analysis of 2nd digit length to 4th digit length in humans. *Proceedings of the Biological Sciences, 284*(1857), 20170356. https://doi.org/10.1098/rspb.2017.0356

Manning, J. T. (2010). Digit ratio (2D:4D), sex differences, allometry, and finger length of 12–30-year olds: Evidence from the British Broadcasting Corporation (BBC) internet study. *American Journal of Human Biology, 22*(5), 604–608. https://doi.org/10.1002/ajhb.21051

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

Mayhew, L. M., et al. (2007). Human 2D (index) and 4D (ring) digit lengths: Their variation and relationships during the menstrual cycle. *Journal of Anatomy, 211*, 630–648. https://doi.org/10.1111/j.1469-7580.2007.00801.x

McCormick, C. M., & Carré, J. M. (2020). Facing off with the phalangeal phenomenon and editorial policies: A commentary on Swift-Gallant, Johnson, Di Rita and Breedlove (2020). *Hormones and Behavior, 120*, 104710. https://doi.org/10.1016/j.yhbeh.2020.104710

Packard, G. C. (2009). On the use of logarithmic transformations in allometric analyses. *Journal of Theoretical Biology, 257*(3), 515–518. https://doi.org/10.1016/j.jtbi.2008.10.016

Peters, M., Mackenzie, K., & Bryden, P. (2002). Finger length and distal finger extent patterns in humans. *American Journal of Physical Anthropology, 117*(3), 209–217. https://doi.org/10.1002/ajpa.10029

Ribeiro, E., Neave, N., Morais, R. N., & Manning, J. T. (2016). Direct versus indirect measurement of digit ratio (2D:4D): A critical review of the literature and new data. *Evolutionary Psychology, 14*(1), 1474704916632536. https://doi.org/10.1177/1474704916632536

Richards, G., Medland, S. E., & Beaton, A. A. (2021). Digit ratio (2D:4D) and handedness: A meta-analysis of the available literature. *Laterality, 26*(4), 421–484. https://doi.org/10.1080/1357650X.2020.1862141

Sorokowski, P., & Kowal, M. (2023). Relationship between the 2D:4D and prenatal testosterone, adult level testosterone, and testosterone change: Meta-analysis of 54 studies. *American Journal of Biological Anthropology*, 1–19. https://doi.org/10.1002/ajpa.24852

Swift-Gallant, A., Johnson, B. A., Di Rita, V., & Breedlove, S. M. (2020). Through a glass, darkly: Human digit ratios reflect prenatal androgens, imperfectly. *Hormones and Behavior, 120*, 104686. https://doi.org/10.1016/j.yhbeh.2020.104686

Tan, D., & Vyas, A. (2016). *Toxoplasma gondii* infection and testosterone congruently increase tolerance of male rats for risk of reward forfeiture. *Hormones and Behavior*, *79*, 37–44. https://doi.org/10.1016/j.yhbeh.2016.01.003

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

Voracek, M. (2014). No effects of androgen receptor gene CAG and GGC repeat polymorphisms on digit ratio (2D: 4D): A comprehensive meta-analysis and critical evaluation of research. *Evolution and Human Behavior, 35*(5), 430–437. https://doi.org/10.1016/j.evolhumbehav.2014.05.009

Voracek, M., & Loibl, L. M. (2009). Scientometric analysis and bibliography of digit ratio (2D:4D) research, 1998–2008. *Psychological Reports, 104*(3), 922–956. https://doi.org/10.2466/pr0.104.3.922-956

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

Warrington, N. M., Shevroja, E., Hemani, G., Hysi, P. G., Jiang, Y., Auton, A., Boer, C. G., Mangino, M., Wang, C. A., Kemp, J. P., McMahon, G., Medina-Gomez, C., Hickey, M., Trajanoska, K., Wolke, D., Ikram, M. A., Montgomery, G. W., Felix, J. F., Wright, M. J., Mackey, D. A., Jaddoe, V. W., Martin, N. G., Tung, J. Y., Davey Smith, G., Pennell, C. E., Spector, T. D., van Meurs, J., Rivadeneira, F., Medland, S. E. & Evans, D. M. (2018). Genome-wide association study identifies nine novel loci for 2D:4D finger ratio, a putative retrospective biomarker of testosterone exposure in utero. *Human Molecular Genetics, 27*(11), 2025–2038. https://doi.org/10.1093/hmg/ddy121

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