



Giving science the finger—is the second-to-fourth digit ratio (2D:4D) a biomarker of good luck? A cross sectional study

James M Smoliga, Lucas K Fogaca, Jessica S Siplon, Abigail A Goldburtt, Franziska Jakobs

Department of Physical Therapy,
High Point University, High Point,
NC, USA

Correspondence to: SJ M Smoliga
jsmoliga@highpoint.edu
(ORCID 0000-0002-1895-5687)

Additional material is published
online only. To view please visit
the journal online.

Cite this as: *BMJ* 2021;375:e067849
<http://dx.doi.org/10.1136/bmj-2021-067849>

Accepted: 30 November 2021

ABSTRACT OBJECTIVES

To explore whether random chance, weak research methodology, or inappropriate reporting can lead to claims of statistically significant (yet, biologically meaningless) biomarker associations, using as a model the relation between a common surrogate of prenatal testosterone exposure, second-to-fourth digit ratio (2D:4D), and a random indicator of good luck.

DESIGN

Cross sectional study.

SETTING

University sports performance laboratory in the United States. Data were collected from May 2015 to February 2017.

PARTICIPANTS

176 adults (74 women, 102 men), including university students, faculty, and staff with no history of injuries, disease, or medical conditions that would affect digit length.

MAIN OUTCOME MEASURES

2D:4D, body composition parameters potentially influenced by androgens (bone mineral content, bone mineral density, body fat percentage), and good luck (using poker hands from randomly selected playing cards as a surrogate).

RESULTS

2D:4D significantly correlated with select body composition parameters (Spearman's r_s range -0.26 to 0.23 ; $P < 0.05$), but the correlations varied by sex, participant hand measured, and the method of measuring 2D:4D (by photocopy or radiography). However, the strongest correlation observed was between right hand 2D:4D in men measured by radiograph and poker hand rank ($r_s = 0.28$, $P = 0.004$).

CONCLUSIONS

Greater prenatal exposure to testosterone, as estimated by a lower 2D:4D, significantly increases good luck in adulthood, and also modulates body composition (albeit to a lesser degree). While these findings are consistent with a wealth of research reporting that 2D:4D is related to many seemingly disparate outcomes, they are not meant to provide confirmatory evidence that 2D:4D is a universal biomarker of nearly everything. Instead, the associations between 2D:4D and good luck are simply due to chance, and provide a “handy” example of the reproducibility crisis within medical and scientific research. Biologically sound hypotheses, pre-registration of trials, strong methodological and statistical analyses, transparent reporting of negative results, and unbiased interpretation of data are all necessary for biomarker studies and other areas of clinical research.

Introduction

The ratio between the length of the second and fourth digit, referred to as the 2D:4D digit ratio (2D:4D) has received considerable attention in the scientific community (fig 1) and mainstream media¹⁻⁵ because of its apparent association with health and behavior. The digit ratio is often claimed to be a surrogate for prenatal androgen exposure, based on cross sectional human studies and experimental animal studies.⁶⁻⁹ These studies generally postulate that a lower 2D:4D reflects greater testosterone exposure (or greater testosterone-to-estrogen ratio), which accounts for men having lower 2D:4D than women.¹⁰ Variations in 2D:4D are also speculated to be rooted in genetic polymorphisms that influence testosterone metabolism and sensitivity.^{11 12} However, little evidence supports this hypothesis in humans,^{13 14} because prospective studies have reported a lack of consistent associations between androgen concentrations in the amniotic fluid or umbilical cord blood and 2D:4D in childhood and adulthood.¹⁵⁻¹⁷

Despite a lack of solid physiological justification for studying 2D:4D, an abundance of studies claim that this anthropometric remnant of the prenatal hormonal environment relates to risk of disease in adulthood (that is, cancers¹⁸ and cardiometabolic disease¹⁹⁻²¹), age of onset, prognosis, and treatment options.²² In 2020, 2D:4D was suggested to help “identify those for whom it would be advisable to exercise social distancing” to avoid contracting covid-19.²³ However, thorough examination of the literature raises questions about the validity and reproducibility of 2D:4D research. It is implausible to think that one biomarker in utero not only predicts risk of myocardial infarction and age of onset,²⁴ but also is associated with the likelihood

WHAT IS ALREADY KNOWN ABOUT THIS TOPIC

Second-to-fourth digit ratios (2D:4D) are commonly used as a surrogate for prenatal exposure to testosterone, although the evidence for this association in humans is weak

Many studies have linked 2D:4D to various aspects of physical and mental health, with proponents of the measurement suggesting that it should be incorporated into clinical practice

WHAT THIS STUDY ADDS

A lower 2D:4D is associated with lower body fat percentage, greater bone mineral content, greater bone mineral density, and greater good luck, especially in men
Spurious associations (that is, false positive findings) are likely to account for statistically significant findings in situations where a weak physiological basis exists for a relation between a predictor and outcome

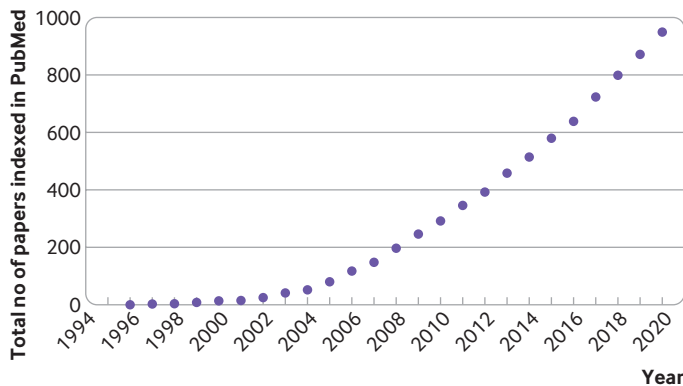


Fig 1 | Cumulative number of PubMed indexed papers about the second-to-fourth digit ratio (2D:4D). The following search was performed to identify papers published each year, with an example of 2020: (“2D:4D” or “digit ratio”) AND (“2020”[date-publication]: “2020”[date-publication])). This search might not capture papers that refer to prenatal testosterone or similar concepts in the abstract but use 2D:4D in the methods

of a person becoming a firefighter,²⁵ having musical ability,^{26 27} showing pro-environmental consumption behavior,²⁸ having a sense of directionality,²⁹ being successful at Sumo wrestling,³⁰ being obsessed with celebrities,³¹ or making a specific choice of Coca-Cola products from a vending machine.³²

Associations between 2D:4D and various outcomes are generally justified through (tenuous) biological explanations, but the possibility of spurious correlations are seldom considered. Digit ratio studies almost always include many comparisons, which increases the likelihood of false positive findings.³³ The digit ratio is also easy to measure, facilitating its inclusion in larger studies, where attempts can be made to correlate it to many other metrics (eg, the BBC internet study³⁴). This practice of including bonus factors without sufficient scientific justification is known to produce spurious associations in other areas of research.³⁵ Even if most studies produce null results, selective reporting and publication bias can create the appearance of consistent positive effect.³⁶ Thus, research involving 2D:4D could provide a prime example of the reproducibility crisis in medicine and science, and of the perpetuation of research based on weak scientific hypotheses,³⁷ non-rigorous methodology, and an over-dependence on confirming hypotheses from weak (potentially spurious) correlations.

With previous research determining that 2D:4D is related to a diversity of outcomes that seem to ultimately shape one’s decisions in life and fate, we aimed to explore the magnitude to which 2D:4D is associated with good luck (using a randomly drawn poker hand selected by each participant as a surrogate measure). To put this in the context of clinically relevant outcomes, we also sought to determine whether 2D:4D was related to body composition parameters, which could plausibly be related to prenatal androgen exposure. We hypothesized that, by random chance alone, 2D:4D would show a statistically significant

relation with good luck, similar in magnitude to body composition parameters. We performed this study to demonstrate that random chance alone could produce seemingly convincing results, rather than validating the use of 2D:4D as a biomarker.

Methods

Study design

An abridged methods section is presented here (full details are provided in appendix 1). This cross sectional study was approved by the institutional review board at High Point University (High Point, NC, USA), and was carried out from May 2015 to February 2017. A priori power analyses are not typically performed to determine sample size in 2D:4D research, so our study sample was based on the time and resources the principal investigator could allot to data collection. A total of 176 individuals gave written informed consent and enrolled in the study. Research participants visited the laboratory in person and underwent body composition testing, had their finger lengths measured using two different procedures, and performed a procedure designed to be a surrogate of good luck.

Study population

Adults aged over 18 years were recruited within a university setting, in a sports performance laboratory (High Point University), including students, faculty, and staff. Individuals with a history of any musculoskeletal or rheumatic diseases, injuries, or surgeries that influenced hands or fingers bilaterally were excluded from the study. Individuals with a history of unilateral hand or finger injury (eg, previous fracture) were allowed to participate, but data from the injured hand or individual fingers were excluded from analysis.

Body composition assessment

We measured bone mineral content, bone mineral density, and body fat percentage using dual energy x ray absorptiometry (DXA) on a Hologic Discovery W scanner (Hologic, Marlborough, MA). Calibration and scan procedures were performed in accordance with manufacturer recommendations (appendix 1).

Digit ratio data collection

Digit ratios were measured by two different procedures in accordance with best practice recommendations (appendix 1). Figure 2 shows example images. Participants were instructed to lightly place their hand on a standard photocopier (MX-3570N, Sharp Electronics, Montvale, NJ), and the researcher then scanned one hand at a time. Digital images were captured and delivered electronically to the research team. To minimize radiation exposure and optimize time efficiency, we also used DXA to obtain images of the phalangeal bones. We used the scanner’s lumbar spine analysis software to capture an image with sufficient detail to identify details of the phalangeal bones.

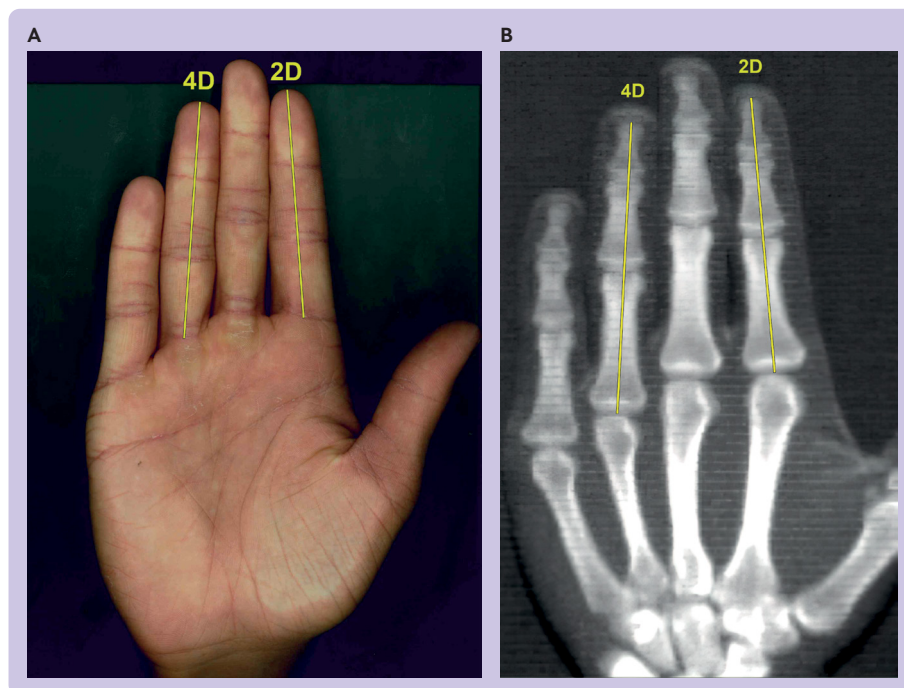


Fig 2 | Example hand images used to measure the second-to-fourth digit ratio (2D:4D). Both images are the right hand of the same participant; yellow lines represent second and fourth digit measurements. (A) Photocopy image: 2D:4D=0.966. (B) Radiographic image: 2D:4D=0.943 (radiographic image has been horizontally flipped to be in the same orientation as the photocopy)

Good luck measurement

Participants were asked to select five cards from a deck of playing cards (United States Playing Card Company, Erlanger, KY). Cards were thoroughly shuffled by an investigator and then fanned out, face down, onto a smooth, flat surface. The participant was then requested to select any five of the face down cards and flip them over. The value and suit of the cards were recorded. This procedure was repeated a second time with a separate, shuffled deck of playing cards, which resulted in each participant having two separate, randomly selected, five card poker hands, drawn from two separate decks of cards.

Poker hands were classified and ranked according to standard poker rules (that is, royal flush as the highest hand, single high card as lowest hand). Each individual's highest ranking hand was then selected, and ranked in relation to all other participants at the completion of the study. For instance, the best poker hand in the study's dataset (a nine high straight) was given the top rank of 1. In the event of a tie, both hands were given the same rank, and the ranking below was given a rank two units below those (eg, if two participants had identical hands, and both were ranked 37th, the next hand below them would be ranked 39th).

Digit ratio measurement

All measurements were performed in Adobe Photoshop using the measure tool. We used mouse guided calipers to measure the second and fourth digits of each hand in accordance with that of previous recommendations.³⁸ For photocopies, the center of the proximal skinfold

nearest the metacarpophalangeal joint was the first point of measurement, and the center of the distal fingertip was the second point. For radiographic (DXA) images, the center of the base of the proximal phalanx was the first point of measurement, and the center of the distal tip of the distal phalanx was the second point.

To minimize risk of rater bias or erroneous measurements, each image was measured by at least two trained raters, each blinded to measurements by the other raters.³⁹ 2D:4D for each hand was computed by the photocopy and radiograph techniques. The percentage difference between raters for each 2D:4D measurement was then computed as follows: $(\text{rater 1} - \text{rater 2}) \div (\text{maximum value of rater 1 or rater 2}) \times 100\%$. If a $\geq 2.0\%$ difference was noted, a third blinded rater repeated the measurement on that image.

Statistical analysis

Statistical analysis was performed by using the average 2D:4D from at least two raters for each hand and each procedure (photocopy and radiography). Significant relations have been reported between 2D:4D and various outcomes when male and female individuals are combined into one group.^{6 40 41} In other instances, the relation between 2D:4D and an outcome is only statistically significant when men and women are analyzed separately.^{27 42} Therefore, we performed all analyses both ways—with sexes combined and separated.

In our (facetious) effort to persuade less statistically savvy readers of the validity of our statistical analyses,

we report as many P values as possible, even where they are not necessary (eg, descriptive statistics). All statistical analyses were performed in SPSS version 27.0 and a priori statistical significance was set at $P \leq 0.05$.

Inter-rater agreement

To determine reliability of the most consistent raters for a given 2D:4D, a one-way random intraclass correlation coefficient was computed for each hand using each technique. Additionally, the mean inter-rater percentage differences were computed for each measurement technique for each hand.

To determine the relation between 2D:4D and sex, hand, and measurement technique, linear mixed effects models were computed by use of a scaled identity as the repeated measures covariance structure. Sex (male *v* female), hand (left *v* right), measurement technique (photocopy *v* radiograph), and all two way interactions between the three factors served as categorical predictors, and 2D:4D served as the dependent variable. A similar linear mixed effects model was also used to compare age, height, and body mass between sexes.

For correlations between 2D:4D and body composition parameters and good luck, preliminary analysis showed that the continuous body composition outcome variables were not normally distributed. Therefore, Spearman's r_s was computed for all correlations.

Patient and public involvement

Conversations with members of the public inspired this study, because many indicated that they had seen 2D:4D mentioned in social media; some believed that scientific research had confirmed that 2D:4D had real life applications, while others expressed doubt that 2D:4D could predict anything. However, patients or the public were not directly involved in this study because of limited resources. A member of the public read this manuscript after submission.

Results

The analysis included 176 individuals (102 men and 74 women). Men were significantly older (age +1.5 years (95% confidence interval 0.2 to 2.9), $P=0.03$), heavier (body mass +16.0 kg (12.9 to 19.1), $P<0.001$), and taller (height +0.16 m (0.10 to 0.22), $P<0.001$) than women. A total of 690 hand images (346 photocopy, 344 radiograph) were included in the analysis. The left hands from one female and one male participant were excluded from analysis, because of a history of broken digits on those hands. Four photocopies (two left hands for men, one left and right for a woman) and six radiographs (three left and three right for men) were not analyzed for technical reasons (that is, missing scan, digits landmarks not clearly visible).

Reliability and rater agreement

All intraclass correlation values were more than 0.90 and the mean percentage difference between raters was less than 1.0% for all 2D:4D measurements, indicating excellent agreement between raters (table 1).

Comparison of 2D:4D by sex, hand, and measurement technique

A summary of 2D:4D data are provided in table 2. Some significant differences were seen: men had a lower 2D:4D than women ($P<0.001$), the left hand had a lower 2D:4D than the right hand for the photocopy technique ($P=0.004$; but a two way interaction showed this association was significant for photocopy only), and the radiograph technique produced lower 2D:4D measurements than the photocopy technique ($P<0.001$). A significant interaction between measurement technique and hand measured ($P=0.006$) was noted, but the interactions between sex and measurement technique ($P=0.70$) and between sex and hand measured ($P=0.92$) were not significant. Despite significant differences in mean 2D:4D, the distribution of 2D:4D between sexes, hands, and techniques showed considerable overlap (fig 3).

Relation between 2D:4D and outcome measures

Correlations between 2D:4D and body composition parameters and good luck ranking are presented in table 3, figure 4, and figure 5. Many statistically significant correlations were seen, with r_s ranging in magnitude from 0.16 to 0.28; the strongest correlation was between the right hand 2D:4D measured by the radiograph technique and poker hand rank in men. However, results were not always consistent between measurement technique. For instance, the correlation between right hand 2D:4D and bone mineral content in women was significant for the photocopy technique ($r_s=0.23$, $P=0.05$), but not for the radiograph technique ($r_s=0.08$, $P=0.5$). In some instances, only men showed a significant correlation (eg, between left and right hand 2D:4D measured using the radiograph technique with bone mineral density), while in others, correlations were only significant when sexes were combined (that is, left and right hand 2D:4D measured using the radiograph technique with body fat percentage).

Discussion

This study intended to explore whether researchers can get lucky in finding statistically significant associations between a biomarker and various outcomes of interest, and whether these relations might reflect random chance rather than biological cause and effect. Failure to recognize these common research pitfalls (eg, scientifically unjustified hypotheses, weak experimental and statistical methodology, and improper reporting; box 1) can allow false positive findings to masquerade as evidence to support unsound theories. We focus on the 2D:4D example, but we urge researchers and clinicians to be especially vigilant when interpreting data from biomarker association studies.

Readers unaware of our study's intent could interpret our results as showing that prenatal testosterone influences body composition in men (maximum $r_s=0.26$), but not as much as it influences good luck ($r_s=0.28$). These results accord with much of what is reported in the 2D:4D literature, including similar

Table 1 | Reliability of rater measurement of the second-to-fourth digit ratio, by measurement technique and participant hand measured. Data are mean (95% confidence interval)

Measurement technique	Left hand		Right hand	
	One-way random intraclass correlation	Difference (%) between raters	One-way random intraclass correlation	Difference (%) between raters
Photocopy	0.987 (0.982 to 0.990)	0.60 (0.52 to 0.67)	0.985 (0.980 to 0.988)	0.65 (0.57 to 0.72)
Radiograph	0.958 (0.944 to 0.968)	0.81 (0.73 to 0.89)	0.966 (0.954 to 0.974)	0.78 (0.70 to 0.86)

magnitude correlations (that is, $r_s=0.15$ to 0.35), lower 2D:4D associated with desirable metrics of performance (that is, better body composition and poker hand), a sex specific effect, and greater association for the right hand than for the left. If study's reported findings are similar to an existing body of research, it might be easy to overlook multiple fallacies and assume that statistically significant findings represent real effects. Thus, our findings could be used (inappropriately) to support theories claiming that prenatal testosterone exposure influences adulthood traits, and they could validate an unfounded hypothesis that 2D:4D might be predictive of future luck.

If we were to interpret these findings seriously, we might suggest that men with a low 2D:4D should participate in activities where good luck is an important contributor to success, while those with high 2D:4D abstain from purchasing lottery tickets. Although we report a lower 2D:4D to be associated with good luck, some studies report that low 2D:4D is also associated with some cancers,¹⁸ and mathematical analysis suggests they might be due to bad luck.^{43 44} Thus, we might postulate that 2D:4D does not have a direct causative influence on one's luck, but rather influences behaviors that modulate luck (that is, carrying a lucky rabbit's foot, frequently interacting with black cats).⁴⁵ This explanation is indeed ridiculous, but scientific "just so" stories are commonly used to explain chance findings and make them fit within an existing paradigm.⁴⁶ We would also caution readers that the good luck described in our study does not necessarily translate to the act of "getting lucky," although previous research indicates that 2D:4D is associated with sexual attractiveness in social situations.⁴⁷

It is problematic when a small subset of positive findings from a larger pool of multiple comparisons are simply assumed to be physiological cause and effect, and spurious correlations are not considered a possibility. Sufficient multiple comparisons allow a reasonable likelihood of finding a difference in group means or a correlation with $P<0.05$, which makes it difficult to disentangle true relations from random chance, especially in the absence of a strong mechanistic hypothesis. Visual examination of figure

3 provides seemingly convincing evidence of a clear relation between 2D:4D and bone mineral density in men (physiologically plausible), but an equally convincing relation is apparent for good luck in figure 4 (clearly spurious). Such chance findings can partly explain why 2D:4D in one hand, but not the other, is significantly associated with the outcome of interest in many studies, and why the more predictive hand is inconsistent between studies. As an example, one meta-analysis relating 2D:4D to athletic performance concludes, "under some circumstances yet to be identified, left hand 2D:4D systematically out-predicts right hand 2D:4D whereas the opposite is true under other circumstances."⁴⁸

Strengths and weaknesses of the study

A strength of this study is that we intentionally performed this study in the same way that hundreds of other 2D:4D studies have been conducted. However, our study attempted to determine whether spurious relations (with poker hand) could easily occur using best practice 2D:4D measurements and how distinct they would be from seemingly physiologically plausible ones (with body composition).

This study had multiple limitations consistent with many 2D:4D studies, as well as medical research attempting to reach conclusions based on correlations or simplistic between-group differences. For example, we did not have a scientific justification for our sample size, which can facilitate misleading results. Although we report significant P values, separate sex analyses were actually underpowered (appendix 2). In underpowered studies, significant findings might be blindly accepted as real, even though they are more likely to be spurious.⁴⁹

This trial was not pre-registered, which is not uncommon for cross sectional studies. Without predefined hypotheses, outcomes, and statistical analyses (including adjustments for covariates and multiple comparisons), readers cannot determine whether significant results were achieved through flexible methodology (eg, various forms of P hacking) and selective reporting. We actually performed different procedures that generated random numbers

Table 2 | Summary of second-to-fourth digit ratios (2D:4D) by sex, measurement technique, and participant hand measured. Data are mean (95% confidence interval). 95% confidence intervals for radiograph ratios for men are equal to three decimal places

	Photocopy		Radiograph	
	Left hand	Right hand	Left hand	Right hand
Women	0.950 (0.944 to 0.956)	0.960 (0.953 to 0.966)	0.926 (0.920 to 0.932)	0.927 (0.920 to 0.933)
Men	0.940 (0.935 to 0.945)	0.952 (0.946 to 0.957)	0.916 (0.910 to 0.921)	0.916 (0.910 to 0.921)

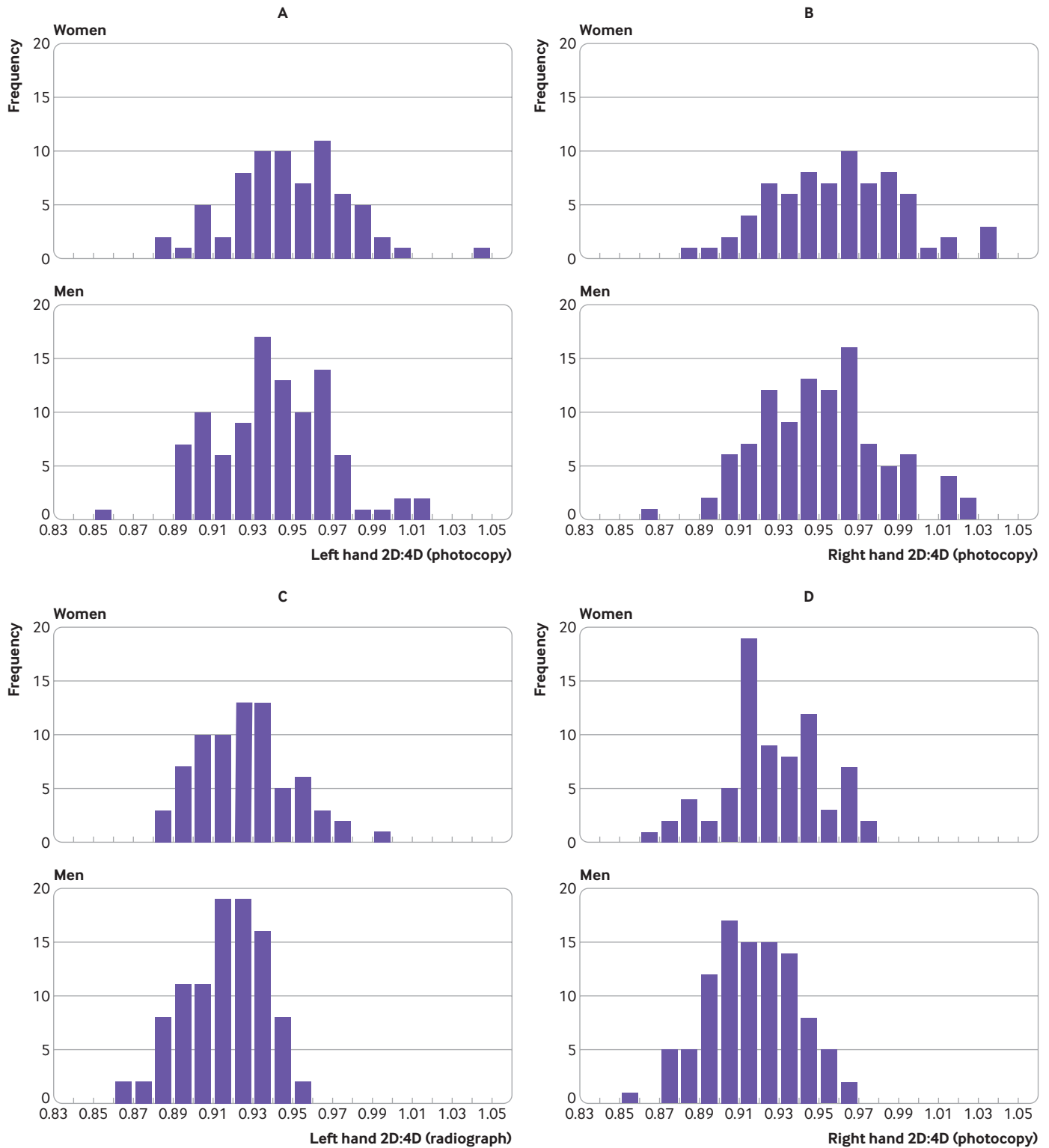


Fig 3 | Histograms of second-to-fourth digit ratios (2D:4D) by participant hand measured, sex, and measurement technique. Despite statistically significant mean differences between men and women for each hand and each technique, the histograms had considerable overlap. (A) Left hand, photocopy technique; (B) right hand, photocopy; (C) left hand, radiograph; (D) right hand, radiograph

(appendix 2), and simply got lucky in that our most fun and interesting procedure (poker hand), had some significant P values (as did all others for at least one comparison). We purposefully omitted these details from the methodology to make our point—even ridiculous hypotheses can be confirmed with sufficient

multiple comparisons, and the apparent validity of these results can be biased through selective reporting. Without pre-registration, reported results might be the endpoint of flexible analyses (see example in fig 6). Our results could have appeared even more convincing if we only reported on 2D:4D measured

Table 3 | Relations between second-to-fourth digit ratios (2D:4D) and age, body composition parameters, and good luck (using poker hand rank as a surrogate)

Sex subgroups and 2D:4D measurement technique	Participant hand measured (No)	Spearman's r_s (P value)			
		Total bone mineral content	Total bone mineral density	Body fat percentage	Best poker hand
Photocopy					
Combined (men + women)	Left (n=171)	-0.12 (0.1)	-0.11 (0.1)	0.13 (0.09)	0.05 (0.6)
	Right (n=175)	-0.13 (0.08)	-0.18 (0.02*)	0.15 (0.05)	0.08 (0.3)
Men	Left (n=99)	-0.07 (0.5)	-0.11 (0.3)	0.03 (0.8)	0.09 (0.4)
	Right (n=102)	-0.22 (0.02*)	-0.26 (0.009*)	0.08 (0.4)	0.11 (0.3)
Women	Left (n=72)	0.12 (0.3)	0.08 (0.5)	0.12 (0.3)	<0.01 (>0.9)
	Right (n=73)	0.23 (0.05*)	0.09 (0.4)	0.15 (0.2)	0.07 (0.6)
Radiograph					
Combined (men + women)	Left (n=171)	-0.20 (0.01*)	-0.19 (0.01*)	0.18 (0.02*)	0.08 (0.3)
	Right (n=173)	-0.24 (0.001*)	-0.25 (<0.001*)	0.20 (0.01*)	0.16 (0.03*)
Men	Left (n=98)	-0.15 (0.1)	-0.20 (0.05*)	-0.02 (0.8)	0.17 (0.1)
	Right (n=99)	-0.24 (0.02*)	-0.26 (0.009*)	-0.01 (0.9)	0.28 (0.004*)
Women	Left (n=73)	0.03 (0.8)	<0.01 (>0.9)	0.14 (0.2)	-0.01 (0.9)
	Right (n=74)	0.08 (0.5)	<0.01 (>0.9)	0.14 (0.2)	0.03 (0.8)

*Some P values round to 0.05 (eg, 0.049 or 0.051), but only those P values that were less than 0.05 are asterisked

on radiograph, which would have seemingly provided fewer multiple comparisons. Even in registered clinical trials, selective reporting and outcome switching are not uncommon.^{50 51}

Conclusions

Our results suggest that a lower 2D:4D, purportedly indicative of greater prenatal testosterone exposure, is associated with favorable body composition parameters and also good luck. When interpreted in the context of the 2D:4D literature, this finding

provides further evidence that 2D:4D might be a universal biomarker of one's fate. In reality, our statistically significant results are actually spurious, and raise the possibility that other claims regarding 2D:4D's association with human health and behavior might also be false positive findings owing to weak experimental and statistical methodology.

The 2D:4D literature provides a valuable example of the necessity for research to have a physiologically sound justification, registered a priori hypotheses with detailed data analysis plans, and publicly

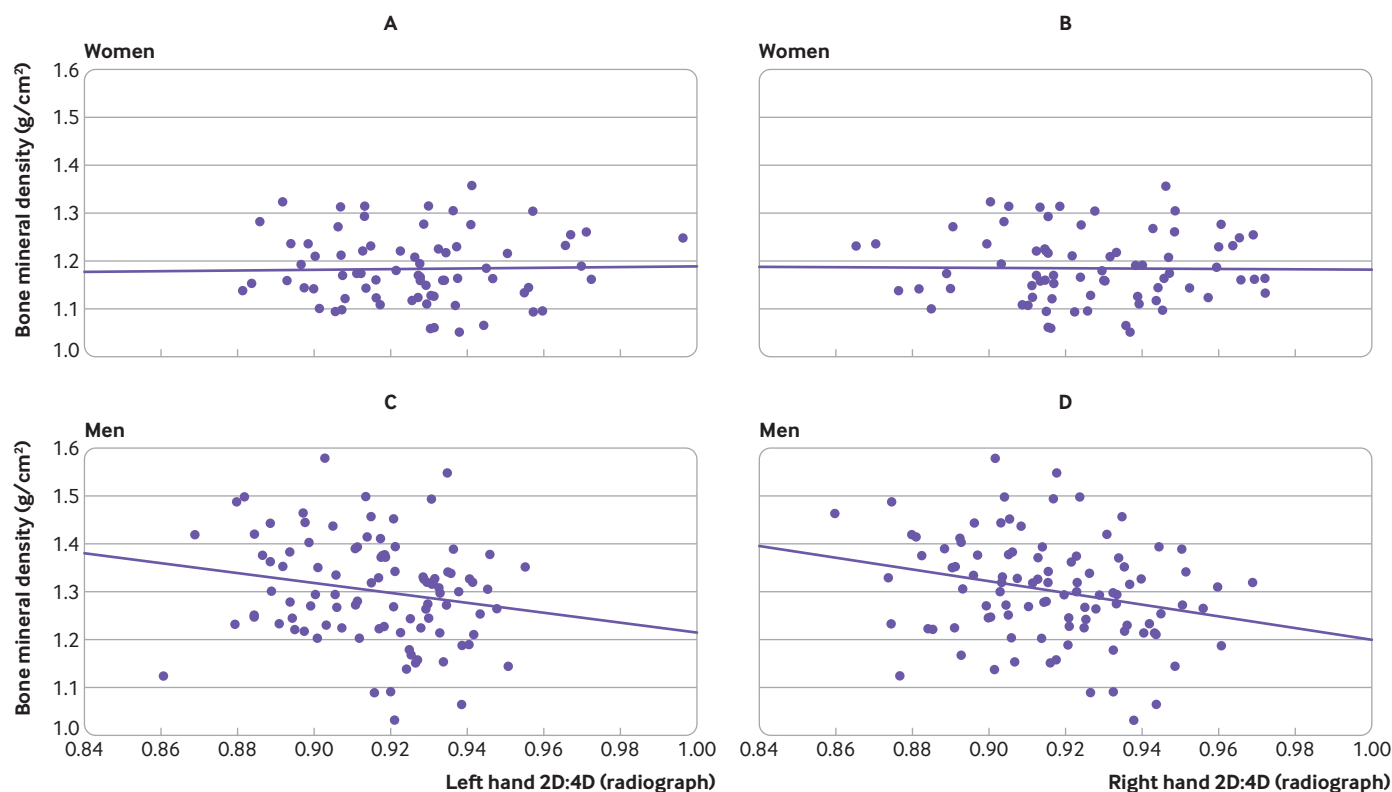


Fig 4 | Scatter plots of second-to-fourth digit ratios (2D:4D) versus bone mineral density by participant sex and hand measured. Only 2D:4D measured by radiographic technique are presented, because the ratios measured by photocopy did not show significant results. (A) left hand, women; (B) right hand, women; (C) left hand, men; (D) right hand, men

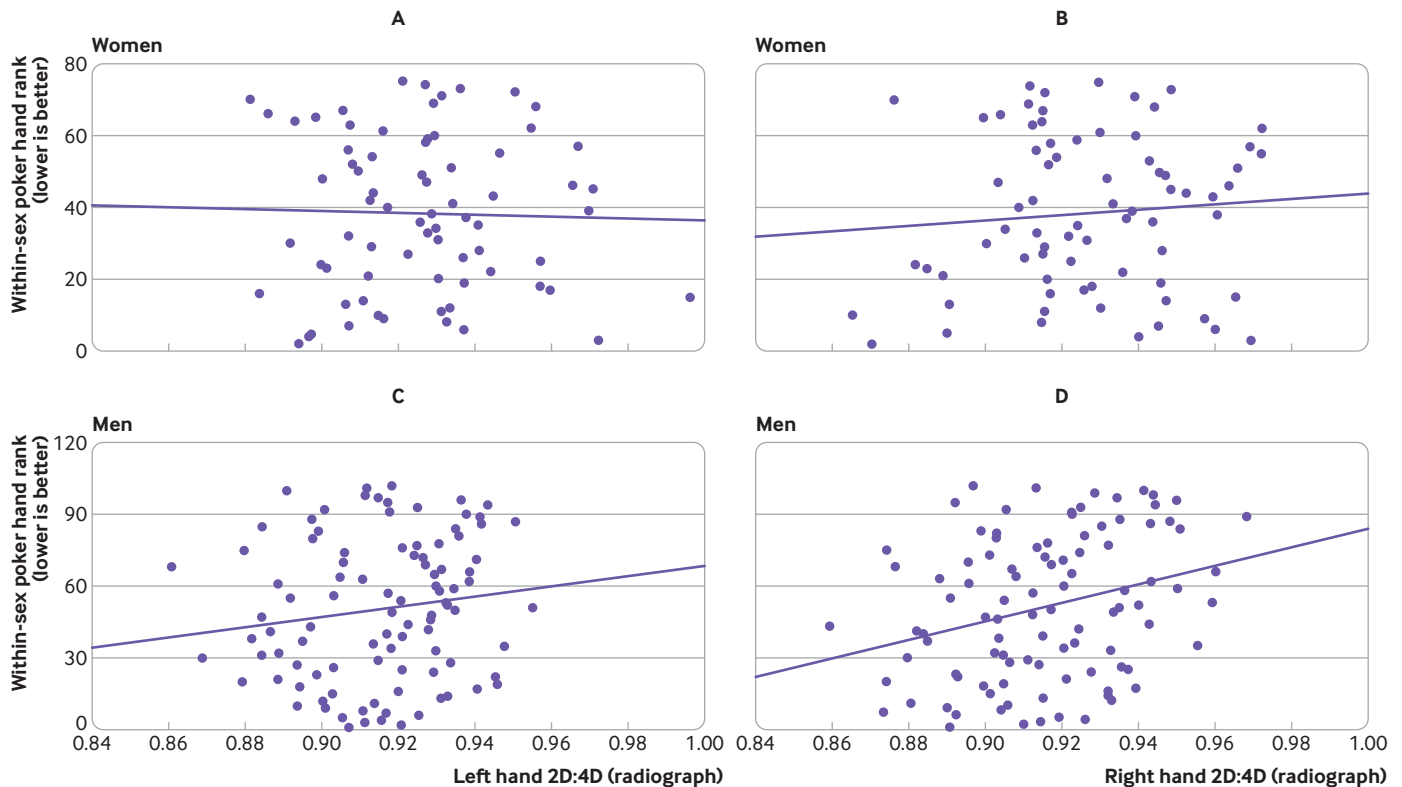


Fig 5 | Scatter plots of second-to-fourth digit ratios (2D:4D) versus poker hand rank by participant hand measured and sex. Only 2D:4D measured by radiographic technique are presented, because the ratios measured by photocopy did not show significant correlations. (A) Left hand, women; (B) right hand, women; (C) left hand, men; (D) right hand, men

available datasets (when feasible) to have a pathway toward clinical relevance. Appropriately powered replication studies and publication of non-significant findings are essential to ensure that poor quality research does not dominate a given field to provide an appearance of a strong body of evidence and spawn biologically unjustified medical recommendations. Before concluding that weak correlations confirm a hypothesis, researchers should consider the possible existence of false positive findings—a dangerous artifact of statistical good luck.

JMS thanks his department for supporting him in investing his time doing this type of research.

Contributors: JMS conceived this idea after reviewing the digit ratio research and becoming suspicious that one biomarker could predict pretty much anything. JMS recruited LKF, JSS, AAG, and FJ to become involved in this project after convincing them that a study about fingers and playing cards was important science. JMS was the principal investigator and led the research design, statistical analysis, data visualization, and manuscript writing and revision. LKF, JSS, and AAG substantially contributed to investigation (data collection and subject recruitment), general project administration, data curation, and contributed to writing aspects of this manuscripts. FJ contributed to data curation and validation. All authors had full access to the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. JMS is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: This study received no specific funding.

Box 1: So-called “pitfalls” to avoid in research on second-to-fourth digit ratios (satirical)

Pre-registration of study protocol

- Substantially reduces flexibility for defining what the primary outcome is, and also creates rigidity in deciding on participant groups and statistical analyses

Performing or reporting a priori power analysis

- Removes flexibility in determining or adjusting sample size and weakens claims of non-significant trends being meaningful

Detailed accounting for multiple comparisons

- Could change statistically significant findings into non-significant ones, which makes for a less interesting (and perhaps less publishable) paper. If multiple comparisons are requested, simply state that they were done and avoid disclosing the denominator used and specifics of how it was determined

Reporting negative findings from other outcomes

- Provides evidence that might contradict evidence otherwise supporting the hypothesis, and might also undermine the appearance of consistent findings within the literature

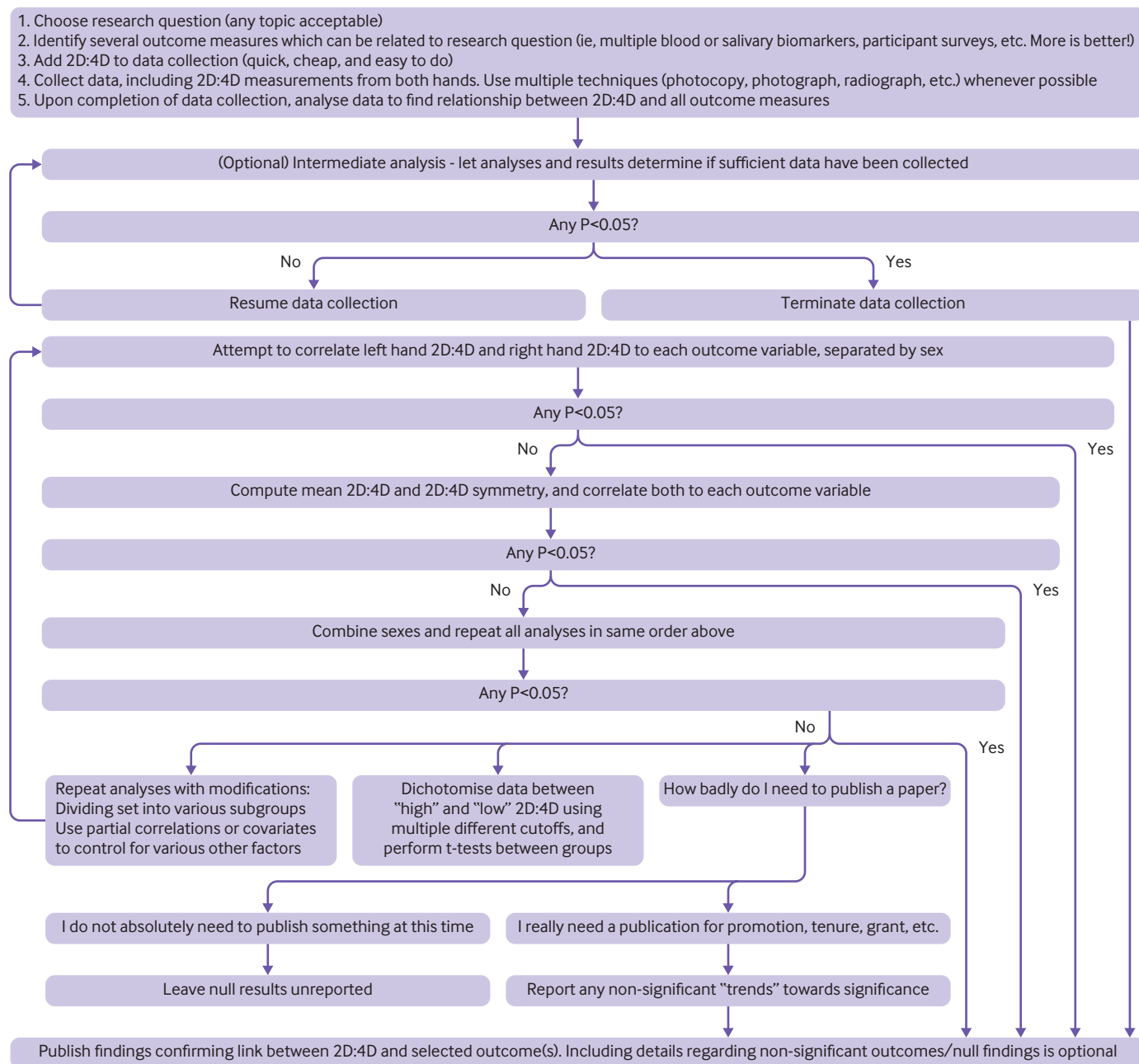


Fig 6 | Hypothetical algorithm for identifying statistically significant relations between second-to-fourth digit ratios (2D:4D) and any outcome measure

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: This study was approved by the High Point University Institutional Review Board (Protocols #201301-153 and #201505-374). All research participants underwent informed written consent.

Data sharing: The full dataset used for this study is available in the supplementary material.

The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities: The principal investigator will directly contact freelance journalists that he has worked with previously, who have written articles about his research for major media outlets; pitch a plain language summary for mainstream media websites where he has written previously (*The Conversation* or *Scientific American* blogs), as well as other similar websites; and provide a Twitter thread on the topic.

Provenance and peer review: Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- 1 Moorhead J. What do your hands say about your health? 2005. <https://www.theguardian.com/society/2005/mar/08/health.lifeandhealth>.
- 2 Walsh F. Index finger length prostate cancer clue. 2010. <https://www.bbc.com/news/health-11880415>.
- 3 Haelle T. What Donald Trump's Hands Might Really Say About Him—According To Science. 2016. <http://www.forbes.com/sites/tarahaelle/2016/03/04/what-donald-trumps-hands-might-really-say-about-him-according-to-science>.
- 4 Gander K. Length of Ring Finger Linked to Men's Risk of Dying from Coronavirus, Study Finds. 2020. <https://www.newsweek.com/length-ring-finger-linked-mens-risk-dying-coronavirus-study-finds-1506692>.
- 5 Narins E. 8 Things Your Ring Finger Length Could Say About Your Personality. 2016. <https://www.cosmopolitan.com/health-fitness/a6573529/things-your-finger-length-could-say-about-your-personality/>.
- 6 Lutchmaya S, Baron-Cohen S, Raggatt P, Knickmeyer R, Manning JT. 2nd to 4th digit ratios, fetal testosterone and estradiol. *Early Hum Dev* 2004;77:23-8. doi:10.1016/j.earlhumdev.2003.12.002
- 7 Manning JT, Scutt D, Wilson J, Lewis-Jones DI. The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Hum Reprod* 1998;13:3000-4. doi:10.1093/humrep/13.11.3000
- 8 Manning JT. Resolving the role of prenatal sex steroids in the development of digit ratio. *Proc Natl Acad Sci U S A* 2011;108:16143-4. doi:10.1073/pnas.1113312108
- 9 Zheng Z, Cohn MJ. Developmental basis of sexually dimorphic digit ratios. *Proc Natl Acad Sci U S A* 2011;108:16289-94. doi:10.1073/pnas.1108312108
- 10 Hönekopp J, Watson S. Meta-analysis of digit ratio 2D:4D shows greater sex difference in the right hand. *Am J Hum Biol* 2010;22:619-30. doi:10.1002/ajhb.21054
- 11 Manning JT, Bundred PE, Newton DJ, et al. The second to fourth digit ratio and variation in the androgen receptor gene. *Evol Hum Behav* 2003;24:399-405. doi:10.1016/S1090-5138(03)00052-7
- 12 Zhang K, Yang X, Zhang M, et al. Revisiting the relationships of 2D:4D with androgen receptor (AR) gene and current testosterone levels: Replication study and meta-analyses. *J Neurosci Res* 2020;98:353-70. doi:10.1002/jnr.24502
- 13 Warrington NM, Shevroja E, Hemani G, et al. 23andMe Research Team. Genome-wide association study identifies nine novel loci for 2D:4D finger ratio, a putative retrospective biomarker of testosterone exposure in utero. *Hum Mol Genet* 2018;27:2025-38. doi:10.1093/hmg/ddy121
- 14 Medland SE, Zayats T, Glaser B, et al. A variant in LIN28B is associated with 2D:4D finger-length ratio, a putative retrospective biomarker of prenatal testosterone exposure. *Am J Hum Genet* 2010;86:519-25. doi:10.1016/j.ajhg.2010.02.017
- 15 Hollier LP, Keelan JA, Jamnadas ESL, Maybery MT, Hickey M, Whitehouse AJ. Adult digit ratio (2D:4D) is not related to umbilical cord androgen or estrogen concentrations, their ratios or net bioactivity. *Early Hum Dev* 2015;91:111-7. doi:10.1016/j.earlhumdev.2014.12.011
- 16 Richards G, Browne WV, Constantinescu M. Digit ratio (2D:4D) and amniotic testosterone and estradiol: an attempted replication of Lutchmaya et al. (2004). *J Dev Orig Health Dis* 2021;1-6. doi:10.1017/s2040174420001294
- 17 Ventura T, Gomes MC, Pita A, Neto MT, Taylor A. Digit ratio (2D:4D) in newborns: influences of prenatal testosterone and maternal environment. *Early Hum Dev* 2013;89:107-12. doi:10.1016/j.earlhumdev.2012.08.009
- 18 Bunevicius A. The Association of Digit Ratio (2D:4D) with Cancer: A Systematic Review and Meta-Analysis. *Dis Markers* 2018;2018:7698193. doi:10.1155/2018/7698193
- 19 Lu H, Ma Z, Zhao J, Huo Z. Second to fourth digit ratio (2D:4D) and coronary heart disease. *Early Hum Dev* 2015;91:417-20. doi:10.1016/j.earlhumdev.2015.04.009
- 20 Manning JT, Bundred PE, Kasielska-Trojan A, Smith-Straney T, Mason L. Digit ratio (2D:4D), myocardial infarction and fibrinogen in men. *Early Hum Dev* 2019;133:18-22. doi:10.1016/j.earlhumdev.2019.04.008
- 21 Oyeyemi BF, Iyiola OA, Oyeyemi AW, Oricha KA, Anifowoshe AT, Alamukii NA. Sexual dimorphism in ratio of second and fourth digits and its relationship with metabolic syndrome indices and cardiovascular risk factors. *J Res Med Sci* 2014;19:234-9.
- 22 Manning JT, Bundred PE. The ratio of 2nd to 4th digit length: a new predictor of disease predisposition? *Med Hypoth* 2000;54:855-7. doi:10.1054/mehy.1999.1150
- 23 Manning JT, Fink B. Understanding COVID-19: Digit ratio (2D:4D) and sex differences in national case fatality rates. *Early Hum Dev* 2020;146:105074. doi:10.1016/j.earlhumdev.2020.105074
- 24 Kyriakidis I, Papaioannidou P, Pantelidou V, Kalles V, Gemitzis K. Digit ratios and relation to myocardial infarction in Greek men and women. *Genet Med* 2010;7:628-36. doi:10.1016/j.genm.2010.11.008
- 25 Voracek M, Pum U, Dressler SG. Investigating digit ratio (2D:4D) in a highly male-dominated occupation: the case of firefighters. *Scand J Psychol* 2010;51:146-56. doi:10.1111/j.1467-9450.2009.00758.x
- 26 Sluming VA, Manning JT. Second to fourth digit ratio in elite musicians: Evidence for musical ability as an honest signal of male fitness. *Evol Hum Behav* 2000;21:1-9. doi:10.1016/S1090-5138(99)00026-4
- 27 Borniger JC, Chaudhry A, Muehlenbein MP. Relationships among musical aptitude, digit ratio and testosterone in men and women. *PLoS One* 2013;8:e57637.
- 28 Hand C. Biology and being green: The effect of prenatal testosterone exposure on pro-environmental consumption behaviour. *J Bus Res* 2020;120:619-26. doi:10.1016/j.jbusres.2019.02.034
- 29 Chai XJ, Jacobs LF. Digit ratio predicts sense of direction in women. *PLoS One* 2012;7:e32816. doi:10.1371/journal.pone.0032816
- 30 Tamiya R, Lee SY, Ohtake F. Second to fourth digit ratio and the sporting success of sumo wrestlers. *Evol Hum Behav* 2012;33:130-6. doi:10.1016/j.evolhumbehav.2011.07.003
- 31 Huh H. Digit ratio and celebrity worship. *Pers Individ Dif* 2012;52:265-8. doi:10.1016/j.paid.2011.10.025
- 32 Aspara J, Van Den Bergh B. Naturally designed for masculinity vs. femininity? Prenatal testosterone predicts male consumers' choices of gender-imaged products. *Int J Res Mark* 2014;31:117-21. doi:10.1016/j.ijresmar.2013.09.001
- 33 Colquhoun D. The reproducibility of research and the misinterpretation of p-values. *R Soc Open Sci* 2017;4:171085. doi:10.1098/rsos.171085
- 34 Reimers S. The BBC internet study: general methodology. *Arch Sex Behav* 2007;36:147-61. doi:10.1007/s10508-006-9143-2
- 35 Brandler WM, Paracchini S. The genetic relationship between handedness and neurodevelopmental disorders. *Trends Mol Med* 2014;20:83-90. doi:10.1016/j.molmed.2013.10.008
- 36 Rosenthal R. The file drawer problem and tolerance for null results. *Psychol Bull* 1979;86:638-41. doi:10.1037/0033-2909.86.3.638
- 37 Betts MG, Hadley AS, Frey DW, et al. When are hypotheses useful in ecology and evolution? *Ecol Evol* 2021;11:5762-76. doi:10.1002/ece3.7365
- 38 Allaway HC, Bloski TG, Pierson RA, Lujan ME. Digit ratios (2D:4D) determined by computer-assisted analysis are more reliable than those using physical measurements, photocopies, and printed scans. *Am J Hum Biol* 2009;21:365-70. doi:10.1002/ajhb.20892
- 39 Voracek M, Manning JT, Dressler SG. Repeatability and interobserver error of digit ratio (2D:4D) measurements made by experts. *Am J Hum Biol* 2007;19:142-6. doi:10.1002/ajhb.20581
- 40 Austin EJ, Manning JT, McInroy K, et al. A preliminary investigation of the associations between personality, cognitive ability and digit ratio. *Pers Individ Dif* 2002;33:1115-24.
- 41 Fink B, Manning JT, Neave N. Second to fourth digit ratio and the 'big five' personality factors. *Pers Individ Dif* 2004;37:495-503.
- 42 Brosnan MJ. Digit ratio as an indicator of numeracy relative to literacy in 7 year old British schoolchildren. *Br J Psychol* 2008;99:75-85.
- 43 Nowak MA, Waclaw B. Genes, environment, and "bad luck". *Science* 2017;355:1266-7. doi:10.1126/science.aam9746
- 44 Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 2015;347:78-81. doi:10.1126/science.1260825
- 45 Rudski J. What does a "superstitious" person believe? Impressions of participants. *J Gen Psychol* 2003;130:431-45. doi:10.1080/00221300309601168
- 46 Nuzzo R. How scientists fool themselves - and how they can stop. *Nature* 2015;526:182-5. doi:10.1038/526182a
- 47 Roney JR, Maestriperi D. Relative digit lengths predict men's behavior and attractiveness during social interactions with women. *Hum Nat* 2004;15:271-82. doi:10.1007/s12110-004-1009-5
- 48 Hönekopp J, Schuster M. A meta-analysis on 2D:4D and athletic prowess: Substantial relationships but neither hand out-predicts the other. *Pers Individ Dif* 2010;48:4-10. doi:10.1016/j.paid.2009.08.009
- 49 Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013;14:365-76. doi:10.1038/nrn3475
- 50 Goldacre B, Drysdale H, Dale A, et al. COMPare: a prospective cohort study correcting and monitoring 58 misreported trials in real time. *Trials* 2019;20:118. doi:10.1186/s13063-019-3173-2
- 51 Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65. doi:10.1001/jama.291.20.2457

Web appendix: Online appendix 1

Web appendix: Online appendix 2

Web appendix: Supporting data used in study