Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Early Human Development 152 (2021) 105272

Contents lists available at ScienceDirect

Early Human Development

journal homepage: www.elsevier.com/locate/earlhumdev

No credible evidence for links between 2D:4D and COVID-19 outcomes: A probabilistic perspective on digit ratio, ACE variants, and national case fatalities

Alex L. Jones [a,*], Bastian Jaeger [b], Christoph Schild [c]

[a] Swansea University, United Kingdom
[b] Tilburg University, the Netherlands
[c] University of Siegen, Germany

ARTICLE INFO

Keywords:
COVID-19
Digit ratio
2D:4D
ACE
Bayesian inference

ABSTRACT

Research into COVID-19 susceptibility and outcomes are critical, but claims must be carefully evaluated to inform policy decisions. In a recent series of articles, Manning and Fink [1–3] use national-level data to describe associations between case-fatality ratios and male and female finger ratios (2D:4D), a suggested measure of prenatal androgen exposure, as well as angiotensin-converting enzyme (ACE) allele and genotype frequencies. The authors suggest that 2D:4D is linked with ACE variant prevalence, and that higher male 2D:4D is associated with higher case fatality ratios, and point to 2D:4D as a useful prognostic measure for COVID-19 susceptibility. A critical review and robust Bayesian analysis of the hypothesis is described here, finding no conclusive evidence of COVID-19 mortality and 2D:4D, nor associations between 2D:4D and ACE1 allele or ACE2 genotype frequency. This absence of evidence is present for data taken from the second wave of COVID-19 in October 2020. Problematic theoretical grounding, individual-level conclusions drawn from national-level data, and issues with statistical inference in the original articles are discussed. Taken together, the current data offer no clear utility of 2D:4D in determining COVID-19 outcomes.

1. Introduction

COVID-19 is a public health emergency, with researchers across broad disciplines contributing reports that progress understanding of disease susceptibility and progression. A clinical feature of the infection is that males experience more severe symptoms, and have higher mortality than females [4], consistent with other acute respiratory infections [5]. While there is currently little supporting data, it has been suggested that androgen sensitivity facilitates more severe infection through greater levels of transmembrane protease serine 2 (TMPRSS2; [6]), and accordingly, males with more severe COVID-19 infections present with phenotypes of high levels of androgen sensitivity, such as male pattern baldness [7]. Related, angiotensin converting enzyme 2 (ACE2) has greater activity in males [8], and in other coronaviruses, has been shown to provide an entry point to the cell through cleaving with TMPRSS2 [9]. ACE2 also has a fundamental role in COVID-19 severity, acting as the main binding of the virus to cell surfaces [10], and its depletion leaves angiotensin II to damage cells [11].

In a recent series of articles, Manning and Fink report the association of national-level index to ring finger ratio (2D:4D) with national case fatality rates [1–3], and test the associations of national-level ACE1 allele frequency and the ACE2 genotype on the same outcome [3]. 2D:4D has long been thought to correlate negatively with prenatal androgen exposure, such that shorter index fingers relative to ring fingers indicate more prenatal androgen exposure [12]. However, interestingly Manning and Fink report that countries with higher male 2D:4D ratios actually have *higher* case fatality ratios, suggesting that higher levels of prenatal testosterone may confer protection from COVID-19 symptoms [2,3]. Manning and Fink suggest that this may be due to the frequency of the ACE2 genotype within a nation, higher levels of which reduce COVID-19 severity, while also demonstrating negative correlations between national level ACE1 allele frequency and ACE2 genotype frequency and national level 2D:4D [3], which are both thought to relate to androgen exposure [6,12].

Taken together, the current evidence is used to suggest that 2D:4D could be used as a prognostic device for the severity of COVID-19 [1], or

* Corresponding author.
E-mail address: alexjonesphd@gmail.com (A.L. Jones).

https://doi.org/10.1016/j.earlhumdev.2020.105272
Received 23 September 2020; Received in revised form 4 November 2020; Accepted 10 November 2020
Available online 17 November 2020

inform whom should engage in more stringent forms of social distancing [1]. While this is an intriguing hypothesis, we outline below reasons why the existing evidence for it is weak.

1.1. 2D:4D as a marker of prenatal testosterone

The hypothesis that 2D:4D is associated with COVID-19 outcomes rests on the assumption that 2D:4D is a valid measure of prenatal androgen exposure. In recent years, accumulating evidence has suggested that this relationship may not be robust. Both Hollier et al. [13] and Hickey et al. [14] demonstrate no relationships between adult 2D:4D ratio and androgen concentrations in umbilical cord blood, providing a critical test of the prenatal androgen exposure hypothesis. There is also a lack of association between testosterone levels in amniotic fluid or mother's plasma and 2D:4D, at least in infants [15]. In addition, the association between human androgen receptor genes and 2D:4D are almost exactly null [16], as well as the association between circulating testosterone and continuous 2D:4D measures [17,18], and changes in testosterone and physical exercise [19], all of which suggests the basic biological mechanisms underpinning 2D:4D as a measure of any kind of testosterone exposure are weak to non-existent. In response to a similar criticism by Jones et al., [20] Manning and Fink [2] point out that 2D:4D shows associations with prenatal androgen exposure in discussions of sexual orientation [21] and gender dysphoria [22], but the direct evidence that higher levels of prenatal androgen impact later outcomes is missing. However, others demonstrate that 2D:4D is associated with amniotic fluid testosterone, but only in limited cases, such as for the left hands of females only [23], which makes its association with more severe COVID-19 for males difficult to reconcile.

However, the relationships between 2D:4D and physical endurance and performance offer some theoretical grounding for the relationship between COVID-19 outcomes and digit ratio, in that lung capacity and efficiency may be linked to testosterone exposure. For example, male rowers with lower 2D:4D show faster rowing times [24], and those with lower digit ratios record faster running times in endurance running competitions [25]. These relationships are generally strong, but vary in their estimated strength in males [26,27], and are either present [27,28] or absent in females [25,29]. 2D:4D of either hand is not associated with VO_2 max, a measure of maximal oxygen uptake but the differences between the two hands may be [30], though this difference is not investigated frequently elsewhere. While equivocal, these findings do suggest a stronger link with 2D:4D and aspects of cardiovascular functioning.

1.2. The ecological fallacy

A concerning aspect of the hypothesis suggested by Manning and Fink is that it is based solely on national-level data. Manning and Fink use aggregated country-level 2D:4D measures taken from the BBC internet study, which comprises of thousands of self-reported digit measurements from around the world [31]. These are then related to country-level case fatality ratios. While there is nothing inherently incorrect with analysing aggregate-level data (sometimes referred to as ecological correlations) [32], and in some cases they yield the same conclusions as individual-level data, this analytical approach becomes problematic when inferences are extended to the individual level, and assuming the relationships are the same at both levels is misleading [33]. In this case, Manning and Fink infer many individual-level conclusions from their data, such as suggesting that an individual's ACE2 genotype will influence 2D:4D [3], and recommendations of who should social distance based on digit ratios [1]. In both recent articles, findings are framed heavily in terms of the relationships between hormonal and genetic pathways, which occur within individuals, which the data has no bearing on. While trait relationships with 2D:4D have been shown to exist at national and individual levels, such as in the case of alcohol and cigarette consumption [34], the default position should not assume this duality is consistent, particularly in the context of a public health crisis, especially when the absence of such a duality is a significant problem in inference [33].

1.3. Statistical inference

All claims rest on the robustness of the statistical inferences made, and there are several issues with the analyses presented in the recent series of articles suggesting links between COVID-19 and digit ratio [1–3]. The first is that the outcome measure of case fatality ratio (the number of deaths divided by the number of cases) is likely a biased measure, given that each nation my record cases differently – for example, either by recording all confirmed cases, or by reporting only cases where death or recovery has occurred, which means a ratio of deaths to cases will be biased in some way [35]. In addition, there may be time-lags in how cases are reported within a nation, which is especially true during the early stages of the pandemic, upon which the initial findings were reported [1], which means the measure is likely noisy. It is also worth noting that the World Health Organisation reports give cumulative death rates and cumulative cases, which are used to define the case fatality rate, which are more closely aligned to a measure of a proportion of a group that die over a specific time, rather than a case fatality rate [35].

Second, there is an overreliance on simple bivariate correlations for the various combinations of male, female, left and right hands, and ACE1 and ACE2 frequencies, and case fatality ratios as defined in the articles. This approach greatly increases the likelihood of finding a statistically significant result when there is no relationship, due to false positive rates [36,37], which there are no multiple comparisons for. This is particularly concerning for right and left 2D:4D, as even though there are arguments to be made that these two variables are not simply a bilateral version of the same trait, but are affected differently by androgen exposure [38], there are no specific hypothesis tests regarding left or right hands in any of the existing articles [1,3] and conclusions are drawn directly about 2D:4D and its relationship with COVID-19 outcomes at a general level. Thus, corrections for multiple comparisons should be made for at least this aspect of the data. To compound this, the sample size is relatively small, though this is understandable given the nature of the dataset. Nonetheless, correlations are unstable in small samples where any outliers can sufficiently skew the estimated relationship higher to find a significant effect [39,40].

Most concerning is that there is a dichotomous approach to inference [41] - effects are described as either present or absent, with an exclusive focus on *p*-values of regression or correlation coefficients. This is often misleading and can obscure what the models actually imply about the relationship between 2D:4D, ACE and COVID-19 outcomes [42]. Interpretation and dissection of statistical models are essential to understand the credibility and plausibility of results, at a time when replication and reliability are low [43] and global pandemics demand reliable results [44]. Crucially, the regression models described by Manning and Fink imply unrealistically large effects between digit ratio and expected mortality from COVID-19, which questions the credibility of the results. Sahin [45] offers an improved parameterisation of the models that more closely approximates the question, but the focus is once more on statistical significance, and includes unjustified covariates to improve model fits.

As an example, Manning and Fink [1] report a multiple regression model predicting log case fatality ratio from national right-hand 2D:4Ds of males and females. The male coefficient is significant; $B = 51.84$, $p = .025$, while for females it is not; $B = -11.26$, $p = .52$. Examining the data reveals the authors use a log-transformation with base-10 of the case fatality ratio, and so the coefficients must be interpreted on this scale – that is, for a one-unit increase in national-level right-hand 2D:4D, log case fatality ratio increases by 51.84. It is more useful to cast this in terms of the actual case fatality ratio by applying an inverse log-10 transformation, which yields the interpretation that a one-unit increase in national male right hand 2D:4D has a multiplicative effect on

the case fatality ratio of $10^{51.84}$, which is an implausibly large effect. For example, the case fatality ratio of the United Kingdom in the data used by Manning and Fink [1] was 11.148, and the right hand 2D:4D ratio is 0.985. Applying this to a hypothetical individual (while ignoring the ecological fallacy) would suggest that a male with a right hand 2D:4D of half a standard deviation above average (an increase of 0.002 units, from 0.985 to 0.987), would have a mortality risk of $11.144 \times (10^{0.002 * 51.84}) = 14.151\%$, or a relative increase in risk of 26.97%. This is similar to the differences in risk of cardiovascular disease when comparing men who have high blood pressure and high cholesterol or healthy levels of both [46]. In short, according to the model of Manning and Fink [1], even a very small change in digit ratio should produce an implausibly large increase in mortality, which questions the credibility of their reported results. That these relationships hold over time unfortunately does not aid in their interpretation [2,3].

1.4. A Bayesian interpretation

Taken together, the above points suggest that the hypothesis relating 2D:4D to COVID-19 outcomes has a weakening theoretical basis, makes individual-level conclusions from national-level data, and has problematic statistical inferences in terms of measures, corrections, and interpretability. Here, using Bayesian methods, we quantify the evidence for this hypothesis using data from Manning and Fink [3], re-examining their analysis of WHO Situation Report 165 [47]. Given that, as of October 2020, COVID-19 is entering a 'second wave' globally [48], we also analyse data from the WHO Situation Report of 18th October [49] to examine whether the relationship exists.

The Bayesian approach has a number of advantages for the current question. First, it answers the question policy makers and researchers are interested in, specifically what the probability of 2D:4D being associated with COVID-19 outcomes is given the available evidence. Current analytic approaches, based on frequentist statistics, only provide the reverse – the probability of observing the data given the hypothesis [50]. Another advantage over frequentist estimation in this context is that Bayesian approaches are the only methods available for testing epistemic probability [51]. Frequentist methods force researchers to assume infinite resamples of their data under identical conditions, which works well in experimental settings, but not for the current question – pandemics of the same disease are vanishingly rare, and the exact conditions under which they occur (global travel, healthcare systems, government responses, etc) vary dramatically. Bayesian methods are not constrained by this assumption, and thus can answer the question directly. Moreover, using a Bayesian approach also allows for the incorporation of prior belief into the analysis, allowing for the state of existing 2D:4D literature and theory to be taken into account. Given the need for robust evidence during the COVID-19 pandemic and the suboptimal quality of COVID-19 research in other areas [44], a Bayesian interpretation can offer interested researchers or policy makers confidence that investing further in 2D:4D as a prognostic tool is a worthwhile endeavour, or whether the evidence is inconclusive.

2. Study one – a reanalysis of case report 165

2.1. Method

Manning and Fink [3] report the analysis of WHO situation report 165 (3rd July 2020). As COVID-19 mortality has increased over time, the effects of the suggested relationships should be clearer with more deaths. The dataset contains 41 countries of which the number of cases and number of deaths are known, as well as measures of 2D:4D for both males and females and left and right hands [31]. The national frequency of the ACE1 allele is known for 37 of those countries, and the frequency of the ACE2 genotype is known for 39. For each ACE variable, the countries with missing data were omitted when entered into an analysis with other variables.

Manning and Fink [3] run several sets of analyses, but most prominent to their conclusions are two subsets: first, the correlations between national 2D:4D of female and male left and right hands, and ACE1/ACE2 frequency, and second, multiple regression models predicting log case fatality ratio from 2D:4D of female and male right and left hands, and the two ACE frequencies. The latter analyses are broken into separate models, such that log case fatality ratio is predicted from right hand male and female 2D:4D and ACE1, right hand male and female 2D:4D and ACE2, and so on. The versions of these analyses carried out using Bayesian inference are described below.

2.1.1. Correlational analyses

The same set of correlations (eight in total; correlating each variant of male and female, left and right digit ratio with both ACE polymorphism frequencies) were estimated using Bayesian methods. However, while frequentist approaches assume a uniform prior of correlation coefficients between −1 and 1, (that is, the probability of obtaining a correlation of −1 is identical to that of 0), a normal prior with mean zero and standard deviation of 0.25 was used here for the correlation coefficient. This reflects the belief that correlations between 2D: and ACE1 or ACE2 are more probable in the region of ±0.25 (with 68% probability), and unlikely to be higher than ±0.50. This is a crucial decision, that reflects that reflect the small-to-medium effect sizes observed in meta-analytic studies of the association of 2D:4D with hormonal and androgen-linked traits, such as spatial navigation, aggression, and gender roles [16,52–54], while also incorporating the effect sizes seen between correlations of 2D:4D and cardiovascular and lung function [25,26,30]. The flexibility of Bayesian analyses allows for these existing effect size estimates to be considered in the analysis of the current data.

In addition, a *t*-distributed likelihood was used, rather than a normal distribution. This provides a robust correlation – one that is not sensitive to outliers in the data, of which Pearson correlations are sensitive to, especially with small datasets [39,40]. The *t*-distribution has heavier tails, which means that it is less affected by outliers in the data. Thus, this method allows for a fairer test of the hypothesis that national level 2D:4D is linked with ACE alleles and genotypes. It is worth noting explicitly that a *t*-distributed likelihood does not remove or ignore outliers in anyway; it simply considers as less extreme than does the standard assumption of normality. This property was the driving force behind the formulation of the distribution [55].

2.1.2. Binomial regression models

Manning and Fink [1–3] use multiple regression models where log case fatality rate (the number of deaths divided by the number of cases) is predicted from various combinations of male, female, right and left 2D:4D, and ACE1 alleles and ACE2 genotype frequencies. As discussed above, this model formulation is difficult to interpret, and taking the ratio of number of reported cases and reported deaths is likely a biased estimate of the fatality risk of COVID-19 [35].

The approach here is to build a binomial regression model that uses the data as it is provided, with no obfuscating transforms. Each nation has a given number of deaths, and a number of cases. The observed death count in a given nation can be thought of as a sample drawn from a binomial distribution, unique to that nation. A binomial distribution has two parameters – the number of events n and the probability of an event (i.e., a death from COVID-19) occurring, p. A linear model was built that directly predicted p using the logistic function on the predictors. Each nation was treated as its own separate binomial distribution by being assigned its own random intercept, allowing for generalisation to more nations than those observed. The fixed effect of male, female, and left and right 2D:4D, as well as ACE1 and ACE2 was estimated across countries. The coefficients of this model can be interpreted in exactly the same way as those from a logistic regression, by exponentiating the coefficients. These can then be interpreted more in line with the WHO report, indicating the odds of being killed by COVID-19 on the sampled day, as the predictor increases by one standard deviation. Following

Manning and Fink [3], four models were estimated, predicting *p* from male and female right 2D:4D and ACE1 prevalence, male and female right 2D:4D and ACE2 prevalence, and so on.

Each model thus had three predictors – male and female 2D:4D, and ACE variant. For each coefficient, a normal prior was set with mean zero and a standard deviation of 0.30. These priors represent the belief about the magnitude of the coefficients in log-odds, and translate directly to the odds of being dead from COVID-19 at the time of sampling most likely between 0.74 and 1.35, with the probability of effects greater than that being small. That is, coefficients are estimated with prior knowledge that any effect associated with 2D:4D is likely to be small in magnitude [14,16], but also allowing for the larger effect sizes seen in studies testing relationships between 2D:4D and cardiovascular function [25,30]. Priors for the intercept term (one per country) were normal with mean -2 and standard deviation 1, allowing a weakly-informative prior on the baseline odds of *p*, consistent with the fatality rate of COVID-19.

2.1.2.1. Prior predictive checks. A common criticism of Bayesian analysis is that the choice of priors is subjective, or that the results of such an analysis are not an objective measure of probability, though there are clear rebuttals, including the main issue that likelihoods, and not priors, are typically the problem [56]. In order to assuage these criticisms, we performed a prior predictive check (PrPC) on our models [57]. A PrPC uses the prior specifications of the model, before it has seen the outcome data (that is, the number of deaths per country), to make a set of repeated predictions. If the model prior specification is sensible, then the model should be capable of producing datasets that approximate the observed data. We checked our prior specifications using a simple statistic – the total observed number of deaths. For case report 165, the total global deaths in the observed data was 451,373. By taking 1000 sets of predictions from our prior models and computing the sum, we demonstrate that our prior specifications will consider outcomes much larger than what is observed, as shown in Fig. 1. For each model, the prior specification entertains global death counts as high as four million with relatively common frequency, though considers smaller death counts as more probable. Thus, our prior specification will not obfuscate or distort effects.

2.1.3. Bayesian decision-making

As opposed to a single number yielded by a frequentist regression or correlation, a Bayesian analysis will return a posterior distribution of coefficients. This distribution reflects, given the data, the most probable values of the statistic of interest. These posterior distributions can be summarised in various ways to make decisions about hypotheses. The first approach taken here is to calculate the mean of the posterior, in addition to calculating the 95% highest density interval (HDI), or credible interval – the area of the posterior distribution that contains

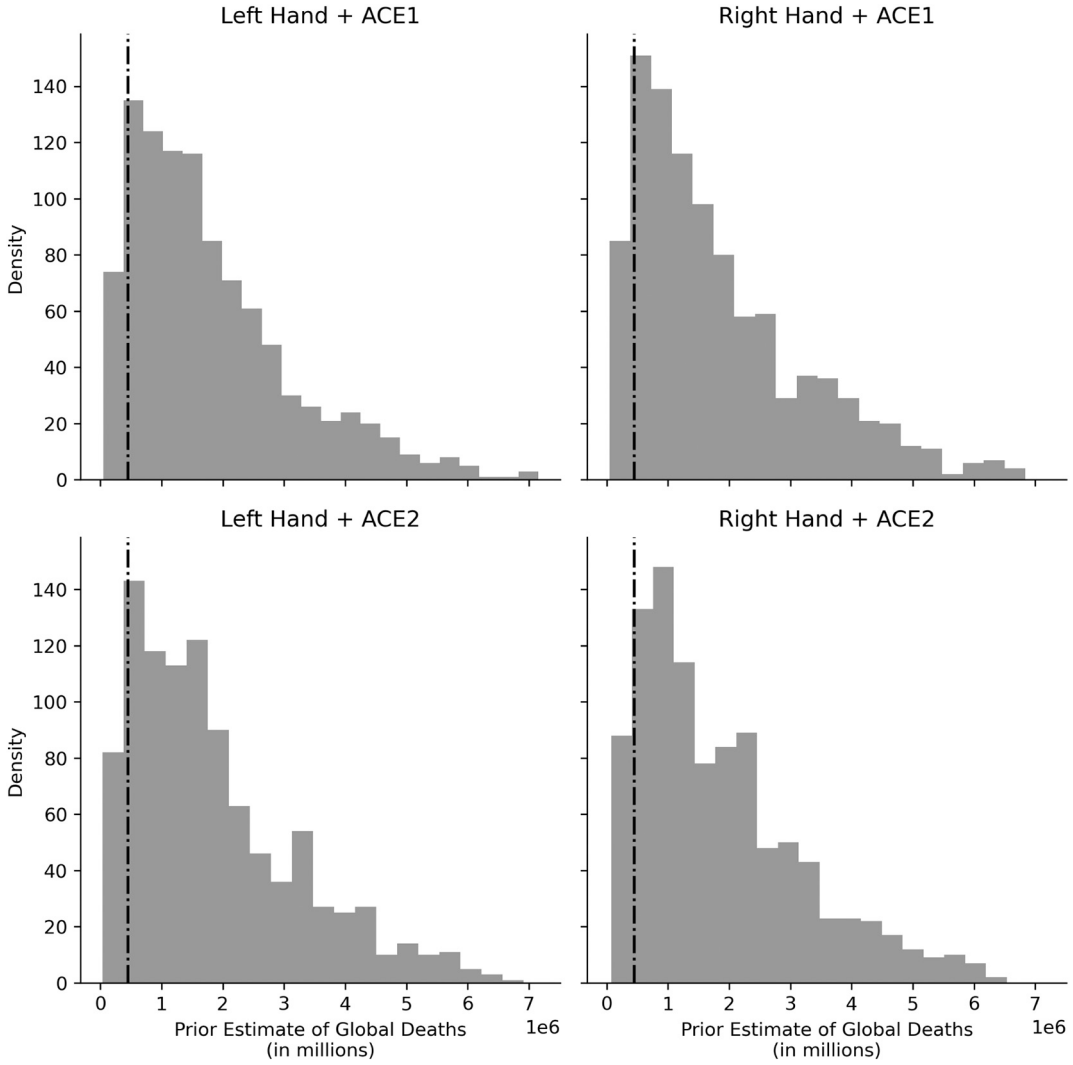

Fig. 1. Prior predictive checks for each model, illustrating the sum of deaths in each set of predictions. The actual death count observed in case report 165 is shown by the dashed line, which the prior specification captures well. Much larger effects are also considered possible, and thus our priors are not restrictive.

95% of its values. The lower and upper bounds of the interval represent the values with which it is directly possible to assert, with 95% certainty, where the true value of the coefficient lies, based on the data [50].

The HDI approach allows for an expression of (un)certainty around a statistic, which directly informs belief about the strength of evidence. A second approach used is that of the region of practical equivalence (ROPE; [58,59]). A ROPE is a set of posterior values that can be considered to be null, in much the same way a point estimate of zero is used in null-hypothesis significance testing. For example, a region of −0.05 to 0.05 for a regression coefficient might be considered a ROPE, as it contains zero and small effects that are of little practical or theoretical relevance. By defining a ROPE, it is possible to calculate the proportion of a posterior that falls within the ROPE and use that to inform decisions about a hypothesis. If a posterior distribution falls entirely within a ROPE, then the effect is practically null and can be discounted; if it falls entirely outside, then the null values can be fully rejected [58]. If the posterior overlaps with the ROPE, the effect is inconclusive. Usefully, in that case, the percentage of the posterior in the ROPE can be used as a continuous measure of evidence of the lack of conclusiveness, as well as the direction [59]. For example, if 90% of the posterior is in the ROPE then the evidence is inconclusive, but there is greater certainty that the effect is null. Conversely, if 10% of the posterior is in the ROPE, then the evidence is still inconclusive, but there is greater certainty the null values could be rejected with more data. This approach offers much more information than *p*-values.

For the correlational analyses, a ROPE of between −0.10 and 0.10 was used. This ROPE considers values surrounding zero, from −0.10 to 0.10 as practically null, which is the threshold of a small effect size definition for correlations of ±0.10 [60]. For the binomial regression models, log-odds coefficients were exponentiated to the odds-ratio scale, and a ROPE of 0.95 and 1.05 was used that corresponds to a region capturing a small reduction or increase in the odds of being killed by COVID-19 on the day of measurement. This ROPE corresponds to about half the magnitude of a small effect size for regression coefficients [50,58,60], and is thus a liberal ROPE (larger ROPEs being more conservative and difficult to overcome). Readers are welcome to define their own ROPE's on the posterior distributions and can accept or reject parameter values they deem worthwhile.

All analyses were estimated using Markov Chain Monte Carlo methods in PyMC3 [61] and data and code are available on the Open Science Framework (osf.io/6v5u7).

2.2. Results

2.2.1. Correlation analysis

The robust Bayesian correlations revealed weaker relationships between all national 2D:4D measures and ACE alleles and genotypes than reported by Manning and Fink [3], though the negative direction of all of the correlations was preserved. While the full posterior distributions, reflecting the belief in the strength of the correlation coefficient, can be averaged to provide a single point estimate (which is consistently lower than those reported by Manning and Fink), the examination of the 95% credible intervals are illuminating. For each correlation, the credible interval contained zero and small positive values, and the ROPE fell directly within the interval, indicating an inconclusive result. The amount of the credible interval falling within the ROPE was also calculated, providing an estimate of how many values of the coefficient was practically null, ranging from approximately nine to 24%. Full posterior distributions are shown in Fig. 2.

2.2.2. Binomial regression models

The probability of death in each country was predicted using a binomial regression model, using the different combinations of national male, female, left and right 2D:4D, as well as ACE1 allele and ACE2 genotype prevalence, deriving a set of posterior distributions for all coefficients. All these were estimated on the log-odds scale, and thus were exponentiated to express them as odds ratios. Odds above one indicated a higher likelihood of being killed from COVID-19 in the sampled data. The full posterior distributions are shown in Fig. 3.

These posteriors represent the belief of how much, given the data, COVID-19 deaths in each country will change given a one-standard-deviation change in the predictor. Across all models and coefficients, the credible intervals were generally wide, with estimates of odds being above and below one, indicating that whether deaths will go up or down with the predictor cannot be expressed with at least 95% confidence. In all coefficients, the posterior overlapped with the ROPE, and in some cases substantially so. In two cases, for male left hand 2D:4D, did the proportions fall below 5%. While the means of the distributions indicate the central tendency of the direction of the effect here (i.e., that it is a positive effect), the posterior overlaps with the ROPE, and the credible interval begins with a null effect (i.e., an odds of one).

The models can be used to make predictions that incorporate uncertainty to demonstrate the usefulness of the inferences that can be made with the data. As an example, consider the left hand 2D:4D model with ACE2 genotype frequency (which has the largest coefficient estimates), and the predictions for the United Kingdom. The observed number of deaths for the UK was 43,995 out of 283,761 cases. Holding

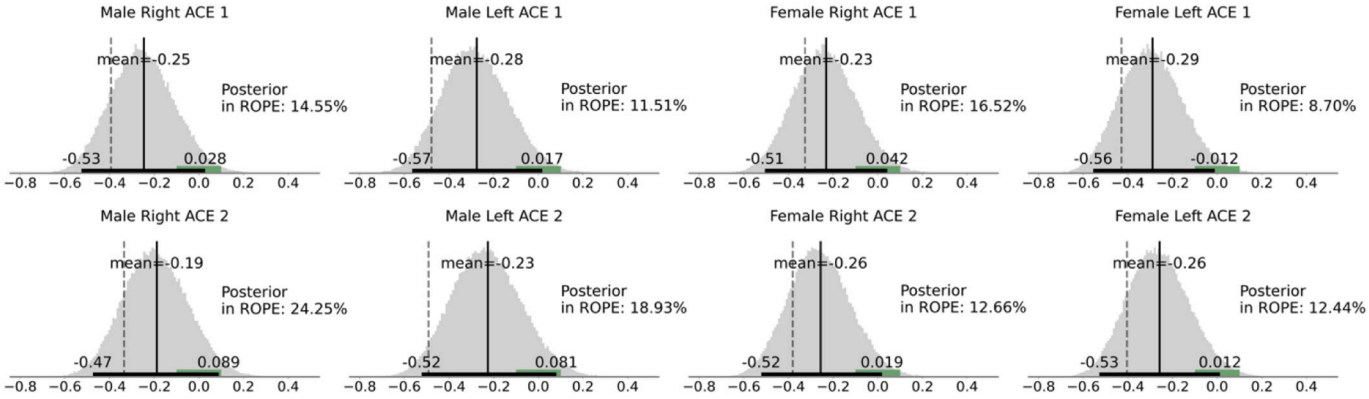

Fig. 2. The posterior distributions of each correlation coefficient, between male, female, left, and right 2D:4D, and ACE I or II polymorphism prevalence. The solid vertical black line is the mean of the distribution (labelled), and the dashed grey vertical line is the frequentist correlation coefficient. The horizontal black line denotes the 95% credible interval, or the range of values that can be stated the true correlation lies in with 95% certainty. The small green bar represents the ROPE of −0.10 to 0.10, which falls partially inside the posterior for all correlations, the proportion of which is displayed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

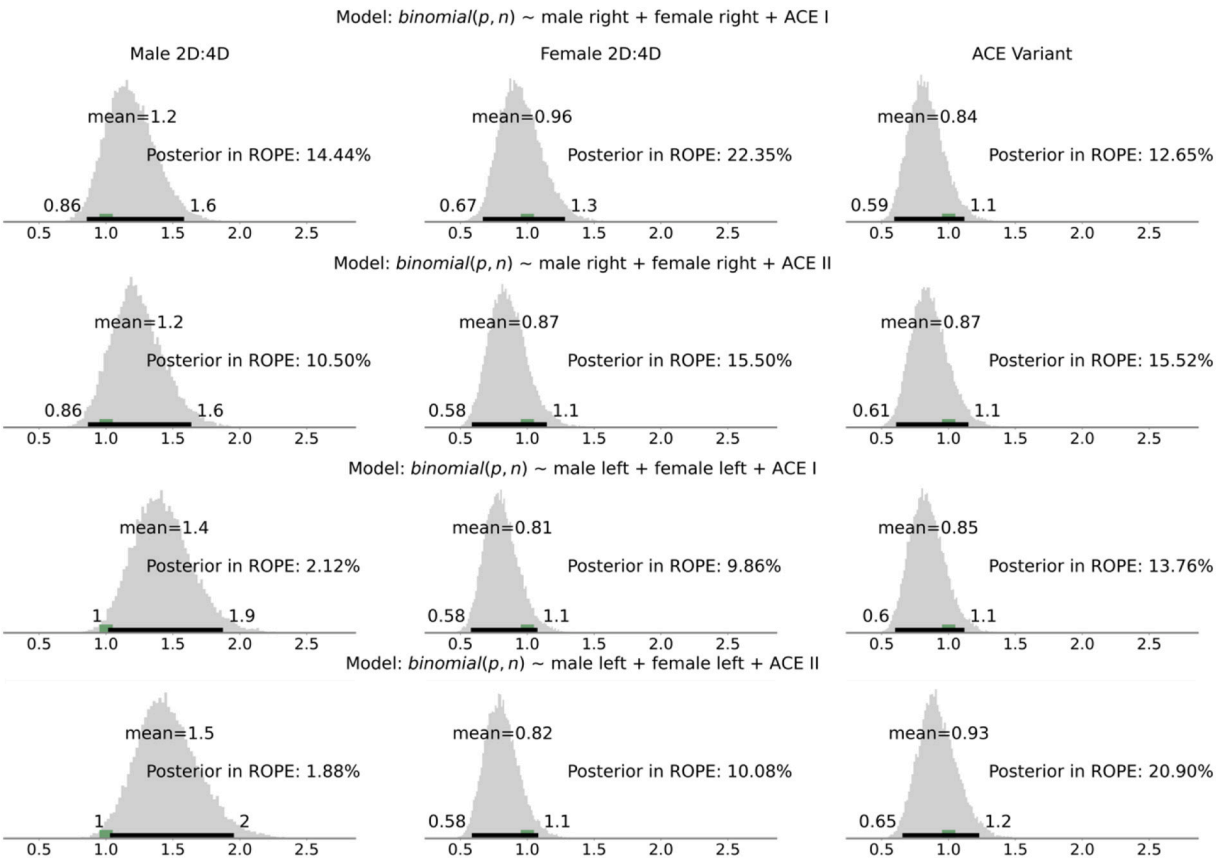

Fig. 3. Coefficients of binomial regressions for the different combinations of model specifications. Each posterior represents the odds ratio of being killed by COVID-19 on the date sampled, from a one standard-deviation increase in the predictor. Means represent the average of the posterior, and the solid black lines the 95% credible intervals or highest posterior density. The small green bar represents the ROPE of 0.95 to 1.05. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

constant female 2D:4D and ACE2 genotype frequency, and assuming an increase in the UK's male 2D:4D of 0.02 units (equivalent to half a standard deviation in global male left 2D:4D, from 0.986 to 0.988), leads to a predicted death count of 43,553, but with a lower 95% credible interval of 33,145, and an upper interval of 57,175. That is, predicted deaths might decrease by 25%, or increase by 30%. This illustrates the poor informativeness of 2D:4–D as a predictor of COVID-19 mortality – these are very wide margins when measured in human lives.

3. Study 2 – case report as of 18th October 2020

The above reanalysis of the data presented by Manning and Fink [3] suggests that the effects of 2D:4D and ACE on COVID-19 outcomes are generally inconclusive, aside from somewhat uncertain effects of male left-hand 2D:4D that seem to be positive, though one cannot exclude no effect. However, Manning and Fink [2,3] show statistically significant effects at multiple time points, though these come with substantial caveats as described above. As COVID-19 is now entering a second wave [48], it is possible that nations with the hypothesized relationships suggested by Manning and Fink [3] – i.e., positive relationships between male 2D:4D and COVID-19 outcomes, and negative relationships between 2D:4D and ACE1 and ACE2 frequencies. Here, we analyse the case report given by the WHO on October 18th, repeating the same binomial regressions as above with the updated case and death data.

3.1. Results

The posteriors for the October data report are shown in Fig. 4. In this more recent data, during the second wave, there are again no conclusive effects. For all coefficients, the 95% credible interval contained one, or no effect, and the lowest overlap with the ROPE was 5.59%, and the effects of left hand male 2D:4D are weaker.

4. General discussion

The hypothesis that COVID-19 outcomes are associated with national level 2D:4D, as well as ACE1 allele and ACE2 genotype frequency has been tested critically here using updated data and Bayesian methods, which offer a series of advantages. Manning and Fink [1–3] suggest that nations with higher male 2D:4D have higher case fatality ratios, and infer this result is due to a host of genetic and hormonal processes influence the relationships between 2D:4D and bodily reactions to COVID-19 infection.

We first tested the relationship between national 2D:4D and ACE frequencies using robust correlations that are not sensitive to outliers [62]. The mean of the posterior distributions of these tests, which can be taken as a point-estimate of the correlation in the same way as a frequentist estimate, are all negative, and in some cases around half the size of the original estimate reported by Manning and Fink [3]. This indicates that the original estimates are affected by outliers. Second, in all but one correlation, the 95% credible interval contains positive values. That is, given this data, asserting 95% confidence in the correlation estimate implies being forced to consider a positive effect, counter to the original hypothesis of Manning and Fink. To be certain of only a negative effect would mean reducing the credible interval, which is a difficult stance to reconcile with evidence that can aid a global pandemic. Finally, the amount of posterior in the ROPE varied from moderate to relatively small; suggesting that, for some effects, a decision about the presence of

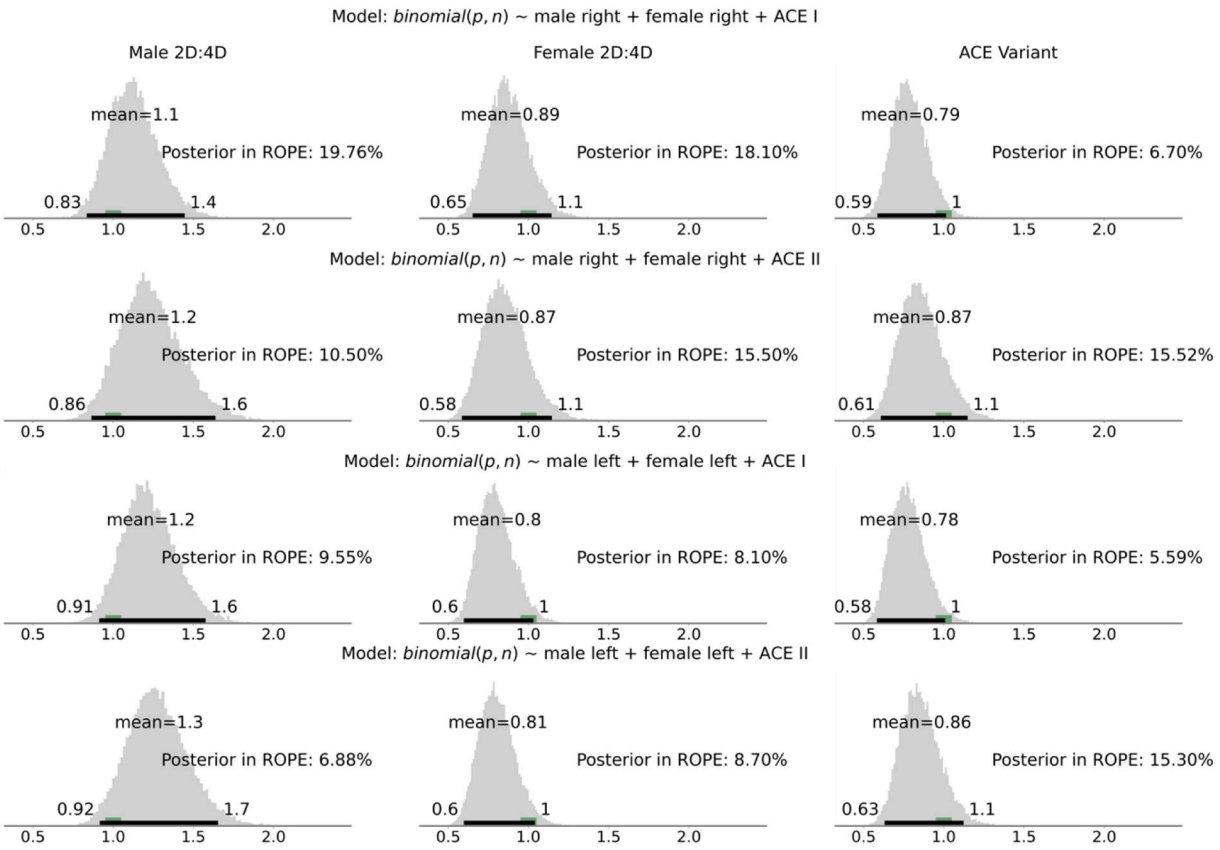

Fig. 4. Coefficients of binomial regressions for the different combinations of model specifications, for October 2020 data. Each posterior represents the odds ratio of being killed by COVID-19 on the date sampled, from a one standard-deviation increase in the predictor. Means represent the average of the posterior, and the solid black lines the 95% credible intervals or highest posterior density. The small green bar represents the ROPE of 0.95 to 1.05. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

an effect was close. However, the most conservative conclusion is that the results are inconclusive for national-level associations between 2D:4D and ACE1 allele frequency or ACE2 genotype frequency.

The binomial regressions on the case report 165 data showed similar outcomes, providing no conclusive evidence of an effect of any of the predictors, under any combination of hand, sex, or ACE frequencies. Manning and Fink and Sahin [1–3,45] suggest that higher male 2D:4D has a positive relationship with COVID-19 fatality, as well as a negative association for females. The posterior distributions of the regression coefficients, expressed as odds, showed no conclusive evidence of these effects, though the general direction of the effects were preserved. However, the credible intervals included null (that is, an odds of one) or possible values in the opposite direction. For some coefficients, overlap in the posterior with the ROPE was small, indicating less uncertainty the absence of an effect. However, there is no conclusive evidence of an association between 2D:4D and COVID-19 mortality at that time point.

As the WHO provides continual updates of the COVID-19 pandemic, any relationships that may exist between 2D:4D and COVID-19 outcomes should become stronger over time as the outcome variable is estimated with greater precision. This is particularly true as COVID-19 enters its second wave – nations with greater male 2D:4D and lower ACE frequencies should see more deaths out of their reported cases, given the hypothesis. Repeating our analysis for the most recent data at the time of writing yielded the same conclusions as case report 165, with the posterior distributions of all coefficients overlapping with negative or null effects. Thus, the posterior distributions convey how much belief we should allocate to the hypothesis, which is that it is inconclusive given the existing data.

The role of scientific research in understanding and treating COVID-19 is paramount. The hypotheses proposed by Manning and Fink [1–3] combine existing evidence around testosterone exposure (as indexed by 2D:4D) and COVID-19 outcomes, which offers a potentially useful clinical measure of susceptibility to the disease. Though this hypothesis is well grounded in biological theory, the current data do not speak to it clearly, particularly because national-level datasets do not necessarily offer insight into individual level outcomes [33], and the current analyses are not easily interpretable. By using Bayesian methods, the current analyses are robust, allow for the incorporation of prior information, and provide clean interpretations of the relationships that are currently obfuscated, and further add to criticism of the hypothesis [20]. Our simple claim here is that, conditional on the existing data, there is no clear evidence of the hypothesis being supported. While Manning and Fink rightly urge for individual level data to corroborate these findings [1–3], these are likely to be costly and would come with serious ethical implications, and with the existing national level data, we see no clear evidence to pursue these research questions.

CRediT authorship contribution statement

Alex Jones: Data curation, Formal analysis, Methodology, Investigation, Validation, Writing – original draft.

Bastian Jaeger: Conceptualization, Investigation, Methodology, Project administration, Validation, Writing – review and editing.

Christoph Schild: Conceptualization, Investigation, Methodology, Validation, Writing – review and editing.

Acknowledgements

The authors would like to thank Chelsea Parlett-Pelleriti for helpful discussion around statistical models.

References

[1] J.T. Manning, B. Fink, Understanding COVID-19: digit ratio (2D:4D) and sex differences in national case fatality rates, Early Hum. Dev. 146 (2020 Jul 1) 105074.

[2] J.T. Manning, B. Fink, Evidence for (mis-)understanding or obfuscation in the COVID-19 and digit ratio relationship? A reply to Jones et al. Early Hum Dev. 148 (2020 Sep 1) 105100.

[3] J.T. Manning, B. Fink, Understanding COVID-19: a hypothesis regarding digit ratio (2D:4D), ACE I/D polymorphism, oxygen metabolism and national case fatality rates, Early Hum. Dev. 151 (2020 Dec 1) 105161.

[4] J.-M. Jin, P. Bai, W. He, F. Wu, X.-F. Liu, D.-M. Han, et al., Gender differences in patients with covid-19: focus on severity and mortality, Front Public Health [Internet] 8 (2020) 152 [cited 2020 Sep 13];8. Available from: https://www.frontiersin.org/articles/10.3389/fpubh.2020.00152/full.

[5] R. Channappanavar, C. Fett, M. Mack, P.P. Ten Eyck, D.K. Meyerholz, S. Perlman, Sex-based differences in susceptibility to severe acute respiratory syndrome coronavirus infection, J Immunol Baltim Md 1950 198 (10) (2017) 4046–4053, 15.

[6] C.G. Wambier, A. Goren, S. Vaño-Galván, P.M. Ramos, A. Ossimetha, G. Nau, et al., Androgen sensitivity gateway to COVID-19 disease severity, Drug Dev Res [Internet] 81 (2020 May) 771–776, 15 [cited 2020 Sep 14]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7273095/.

[7] A. Goren, J. McCoy, C.G. Wambier, S. Vano-Galvan, J. Shapiro, R. Dhurat, et al., What does androgenetic alopecia have to do with COVID-19? An insight into a potential new therapy, Dermatol Ther [Internet] 33 (4) (2020 Apr), e13365, 8 [cited 2020 Sep 14]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7228378/.

[8] G.M. Bwire, Coronavirus: why men are more vulnerable to covid-19 than women? Sn Compr Clin Med. 4 (2020 Jun) 1–3.

[9] A. Heurich, H. Hofmann-Winkler, S. Gierer, T. Liepold, O. Jahn, S. Pöhlmann, TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein, J. Virol. 88 (2) (2014 Jan) 1293–1307.

[10] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, Science. 367 (6485) (2020 Mar 27) 1444.

[11] J.R. Delanghe, M.M. Speeckaert, M.L. De Buyzere, The host's angiotensin-converting enzyme polymorphism may explain epidemiological findings in COVID-19 infections, Clin Chim Acta Int J Clin Chem. 505 (2020 Jun) 192–193.

[12] J.T. Manning, Resolving the role of prenatal sex steroids in the development of digit ratio, Proc. Natl. Acad. Sci. 108 (39) (2011 Sep 27) 16143–16144.

[13] L.P. Hollier, J.A. Keelan, E.S.L. Jamnadass, M.T. Maybery, M. Hickey, A.J. O. Whitehouse, Adult digit ratio (2D:4D) is not related to umbilical cord androgen or estrogen concentrations, their ratios or net bioactivity, Early Hum. Dev. 91 (2) (2015 Feb 1) 111–117.

[14] M. Hickey, D.A. Doherty, R. Hart, R.J. Norman, E. Mattes, H.C. Atkinson, et al., Maternal and umbilical cord androgen concentrations do not predict digit ratio (2D:4D) in girls: a prospective cohort study, Psychoneuroendocrinology. 35 (8) (2010 Sep 1) 1235–1244.

[15] G. Richards, M. Gomes, T. Ventura, Testosterone measured from amniotic fluid and maternal plasma shows no significant association with directional asymmetry in newborn digit ratio (2D:4D), J. Dev. Orig. Health Dis. 10 (3) (2019 Jun) 362–367.

[16] M. Voracek, 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, Evol. Hum. Behav. 35 (5) (2014 Sep 1) 430–437.

[17] A.A. Beaton, N. Rudling, C. Kissling, R. Taurines, J. Thome, Digit ratio (2D:4D), salivary testosterone, and handedness, Laterality. 16 (2) (2011 Mar 1) 136–155.

[18] D.C. Muller, G.G. Giles, J. Bassett, H.A. Morris, J.T. Manning, J.L. Hopper, et al., Second to fourth digit ratio (2D:4D) and concentrations of circulating sex hormones in adulthood, Reprod. Biol. Endocrinol. 9 (1) (2011 Apr 27) 57.

[19] M. Kowal, P. Sorokowski, A. Żelaźniewicz, J. Nowak, S. Orzechowski, G. Żurek, et al., No relationship between the digit ratios (2D:4D) and salivary testosterone change: study on men under an acute exercise, Sci. Rep. 10 (1) (2020 Jun 22) 10068.

[20] A.L. Jones, L.P. Satchell, B. Jaeger, C. Schild, (Mis-)understanding COVID-19 and digit ratio: methodological and statistical issues in Manning and Fink (2020), Early Hum Dev. 148 (2020 Sep 1) 105095.

[21] A. Swift-Gallant, B.A. Johnson, V. Di Rita, S.M. Breedlove, Through a glass, darkly: human digit ratios reflect prenatal androgens, imperfectly, Horm. Behav. 120 (2020 Apr 1) 104686.

[22] M. Sadr, B.S. Khorashad, A. Talaei, N. Fazeli, J. Hönekopp, 2D:4D suggests a role of prenatal testosterone in gender dysphoria, Arch. Sex. Behav. 49 (2) (2020 Feb 1) 421–432.

[23] T. Ventura, M.C. Gomes, A. Pita, M.T. Neto, A. Taylor, Digit ratio (2D:4D) in newborns: influences of prenatal testosterone and maternal environment, Early Hum. Dev. 89 (2) (2013 Feb 1) 107–112.

[24] D. Longman, J.T. Stock, J.C.K. Wells, Digit ratio (2D:4D) and rowing ergometer performance in males and females, Am. J. Phys. Anthropol. 144 (3) (2011) 337–341.

[25] J.T. Manning, L. Morris, N. Caswell, Endurance running and digit ratio (2D:4D): implications for fetal testosterone effects on running speed and vascular health, Am J Hum Biol Off J Hum Biol Counc. 19 (3) (2007 Jun) 416–421.

[26] J.T. Manning, M.R. Hill, Digit ratio (2D:4D) and sprinting speed in boys, Am J Hum Biol Off J Hum Biol Counc. 21 (2) (2009 Apr) 210–213.

[27] M. Voracek, B. Reimer, S.G. Dressler, Digit ratio (2D:4D) predicts sporting success among female fencers independent from physical, experience, and personality factors, Scand. J. Med. Sci. Sports 20 (6) (2010) 853–860.

[28] S.N. Paul, B.S. Kato, J.L. Hunkin, S. Vivekanandan, T.D. Spector, The Big Finger: the second to fourth digit ratio is a predictor of sporting ability in women, Br. J. Sports Med. 40 (12) (2006 Dec 1) 981–983.

[29] M.W. Peeters, A.L. Claessens, The left hand second to fourth digit ratio (2D:4D) does not discriminate world-class female gymnasts from age matched sedentary girls, PLoS One 7 (6) (2012 Jun 29), e40270.

[30] R. Hill, B. Simpson, J. Manning, L. Kilduff, Right–left digit ratio (2D:4D) and maximal oxygen uptake, J. Sports Sci. 30 (2) (2012 Jan 1) 129–134.

[31] J.T. Manning, B. Fink, Digit ratio (2D:4D), dominance, reproductive success, asymmetry, and sociosexuality in the BBC Internet Study, Am J Hum Biol Off J Hum Biol Counc. 20 (4) (2008 Aug) 451–461.

[32] W.S. Robinson, Ecological correlations and the behavior of individuals, Am. Sociol. Rev. 15 (3) (1950) 351–357.

[33] R. Kievit, W.E. Frankenhuis, L. Waldorp, D. Borsboom, Simpson's paradox in psychological science: a practical guide, Front Psychol [Internet] 4 (2013) 513 [cited 2020 May 21];4. Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00513/full.

[34] J.T. Manning, B. Fink, Digit ratio, nicotine and alcohol intake and national rates of smoking and alcohol consumption, Personal. Individ. Differ. 50 (3) (2011 Feb 1) 344–348.

[35] P. Spychalski, A. Błażyńska-Spychalska, J. Kobiela, Estimating case fatality rates of COVID-19, Lancet Infect. Dis. 20 (7) (2020 Jul 1) 774–775.

[36] P. Ranganathan, C.S. Pramesh, M. Buyse, Common pitfalls in statistical analysis: the perils of multiple testing, Perspect Clin Res. 7 (2) (2016) 106–107.

[37] J.P. Simmons, L.D. Nelson, U. Simonsohn, False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant, Psychol Sci [Internet] 22 (11) (2011 Oct) 1359–1366, 17 [cited 2020 Sep 22]; Available from: https://journals.sagepub.com/doi/10.1177/0956797611417632.

[38] J. Hönekopp, S. Watson, Meta-analysis of digit ratio 2D:4D shows greater sex difference in the right hand, Am. J. Hum. Biol. 22 (5) (2010) 619–630.

[39] S.J. Devlin, R. Gnanadesikan, J.R. Kettenring, Robust estimation and outlier detection with correlation coefficients, Biometrika. 62 (3) (1975) 531–545.

[40] Y. Kim, T.-H. Kim, T. Ergün, The instability of the Pearson correlation coefficient in the presence of coincidental outliers, Finance Res Lett. 13 (2015 May 1) 243–257.

[41] G. Cumming, Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis, Routledge, New York, 2011, 536 p.

[42] B.B. McShane, D. Gal, A. Gelman, C. Robert, J.L. Tackett, Abandon statistical significance, Am Stat 73 (sup1) (2019 Mar 29) 235–245.

[43] Open Science Collaboration, Estimating the reproducibility of psychological science, Science 349 (6251) (2015 Aug 28), aac4716.

[44] P.E. Alexander, V.B. Debono, M.J. Mammen, A. Iorio, K. Aryal, D. Deng, et al., COVID-19 coronavirus research has overall low methodological quality thus far: case in point for chloroquine/hydroxychloroquine, J. Clin. Epidemiol. 123 (2020) 120–126.

[45] H. Sahin, A further analysis of Manning and Fink (2020), Early Hum. Dev. 2 (2020 Jul) 105121.

[46] J.D. Berry, A. Dyer, X. Cai, D.B. Garside, H. Ning, A. Thomas, et al., Lifetime risks of cardiovascular disease, N. Engl. J. Med. 366 (4) (2012 Jan 26) 321–329.

[47] World Health Organisation, Coronavirus Disease (COVID-19) [Internet]. 2020 Jul. Report No.: 165, Available from, https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200703-covid-19-sitrep-165.pdf?sfvrsn=b27a772e_6.

[48] S. Xu, Y. Li, Beware of the second wave of COVID-19, Lancet 395 (10233) (2020 Apr 25) 1321–1322.

[49] World Health Organisation, Weekly Epidemiological Update [Internet], Available from, https://www.who.int/publications/m/item/weekly-epidemiological-update—20-october-2020, 2020 Oct.

[50] J.K. Kruschke, T.M. Liddell, The Bayesian new statistics: hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective, Psychon. Bull. Rev. 25 (1) (2018 Feb 1) 178–206.

[51] A. Etz, J. Vandekerckhove, Introduction to Bayesian inference for psychology, Psychon. Bull. Rev. 25 (1) (2018 Feb 1) 5–34.

[52] T.C. Pratt, J.J. Turanovic, F.T. Cullen, Revisiting the criminological consequences of exposure to fetal testosterone: a meta-analysis of the 2d:4d digit ratio*, Criminology. 54 (4) (2016) 587–620.

[53] D.A. Puts, M.A. McDaniel, C.L. Jordan, S.M. Breedlove, Spatial ability and prenatal androgens: meta-analyses of congenital adrenal hyperplasia and digit ratio (2D:4D) studies, Arch. Sex. Behav. 37 (1) (2007 Dec 12) 100.

[54] M. Voracek, J. Pietschnig, I.W. Nader, S. Stieger, Digit ratio (2D:4D) and sex-role orientation: further evidence and meta-analysis, Personal. Individ. Differ. 51 (4) (2011 Sep 1) 417–422.

[55] S.E. Fienberg, N. Lazar, William Sealy Gosset, in: C.C. Heyde, E. Seneta, P. Crépel, S.E. Fienberg, J. Gani (Eds.), Statisticians of the Centuries [Internet], Springer, New York, NY, 2001, pp. 312–317 [cited 2020 Oct 28]. Available from, https://doi.org/10.1007/978-1-4613-0179-0_67.

[56] A. Gelman, C.P. Robert, "Not only defended but also applied": the perceived absurdity of Bayesian inference, Am. Stat. 67 (1) (2013) 1–5.

[57] J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, A. Gelman, Visualization in Bayesian workflow, J. R. Stat. Soc. Ser. A 182 (2) (2019) 389–402.

[58] J.K. Kruschke, Rejecting or accepting parameter values in Bayesian estimation, Adv. Methods Pract. Psychol. Sci. 1 (2) (2018 Jun 1) 270–280.

[59] D. Makowski, M.S. Ben-Shachar, S.H.A. Chen, D. Lüdecke, Indices of effect existence and significance in the bayesian framework, Front Psychol [Internet] 10

(2019) 2767 [cited 2020 Sep 16];10. Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02767/full.

[60] J. Cohen, Statistical Power Analysis for the Behavioral Sciences [Internet], Routledge, 1969 [cited 2019 Jan 10]. Available from: https://www.taylorfrancis.com/books/9781134742707.

[61] J. Salvatier, T.V. Wiecki, C. Fonnesbeck, Probabilistic programming in Python using PyMC3, PeerJ Comput Sci. 2 (2016 Apr 6), e55.

[62] G.A. Rousselet, C.R. Pernet, Improving standards in brain-behavior correlation analyses, Front Hum Neurosci [Internet] 6 (2012) 119 [cited 2020 Sep 17];6. Available from: https://www.frontiersin.org/articles/10.3389/fnhum.2012.00119/full#h2.