

1 **Avoiding misinterpretation of regression lines in allometry:** 2 **is sexual dimorphism in digit ratio spurious?**

3 Wolfgang Forstmeier

4 Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology,
5 82319 Seewiesen, Germany, E-mail: forstmeier@orn.mpg.de

6 **Abstract**

7 The statistical analysis of allometry (size-dependence of traits) is fraught with difficulty that is often
8 underestimated. In light of some recent controversies about statistical methods and the resulting
9 biological conclusions, I here discuss the interpretation of regression lines and show how to avoid
10 spurious effects. General linear models based on ordinary least square (OLS) regression are often
11 used to quantify sexual dimorphism in a trait of interest that is modelled as a function of sex while
12 controlling for size as a covariate. However, an analysis of artificially generated data where males and
13 females differ in size only, but are otherwise built according to the same principles, shows that the
14 OLS method induces a spurious dimorphism where there is none. Hence, OLS-based general linear
15 models should not be regarded as a fail-proof tool that automatically provides the correct answer to
16 whatever question one has in mind. Here I show how to avoid misinterpretation and how to best
17 proceed with answering the recent debate about sexual dimorphism in digit ratio, a trait that is
18 thought to reflect sex-hormone levels during development. The limited data, currently available to
19 me, suggests that the widely accepted sexual dimorphism in digit ratio might well be only a by-
20 product of an allometric shift in shape, urgently calling for a re-examination in larger data sets on
21 humans and other vertebrates.

22 **A controversy about sexual dimorphism as an artefact**

23 The relative length of human index and ring fingers (digit ratio, 2D:4D) has received a lot of research
24 attention, because it is known to be sexually dimorphic (smaller ratios in men than in women [1, 2])
25 and has been suggested to reflect sex-hormone levels experienced during development [2-5]. Now
26 Lolli and coworkers [6] have made a claim that, if true, would have major consequences for this
27 research field. According to their analyses of data on human finger lengths, the apparently lower
28 average digit ratio of men compared to women arises simply as an artefact of allometry: as hands get
29 bigger, there is a general shift in shape such that digit ratio gets smaller. In consequence, there is no

30 particular sexual dimorphism (apart from men being generally larger) that would require an
31 additional explanation invoking specific sex-hormone effects on finger development.

32 This potentially controversial conclusion might however be questioned for methodological reasons.
33 The study by Lolli and coworkers [6] investigates allometry and sexual dimorphism by regressing the
34 length of the second finger 2D (dependent variable) over the length of the fourth finger 4D (predictor
35 variable), yet the implemented method of ordinary least square (OLS) regression fails to account for
36 'biological noise' (aka 'natural variation' [7] or 'biological deviance' [8]) in the predictor variable. An
37 earlier paper [9] already analysed the very same data set using such OLS regression and was
38 subsequently criticized for its inappropriate statistical approach [10]. Indeed, it is easy to show (see
39 further below) that OLS regression systematically induces positive sexual dimorphism (relatively
40 larger trait values in the larger sex) and is therefore not suitable for quantifying sexual dimorphism
41 [10, 11]. Yet, Lolli and coworkers [6] did not address these statistical concerns but rather followed
42 the earlier study [9] (and others, e.g. [12]) in using OLS-based methods. Apparently this was also
43 motivated by a recent methodological review by Kilmer and Rodríguez [13], who incautiously
44 recommend OLS regression for studies of allometry, despite of the inherent problems [7, 10, 14, 15]
45 that have caused a long-lasting debate about regression methods [7, 8, 10]. This may lead to the
46 unfortunate situation where readers are generally confused about which method of regression to
47 use. To bring more clarity into this debate, I here would like to highlight two points:

48 (I) The interpretation of OLS regression lines in the study of allometry and sexual dimorphism is more
49 difficult than is typically recognized [6, 13]. OLS regression lines are suitable for predicting values of y
50 from a given value of x , but they are not designed to automatically reflect the principles that underlie
51 allometric relationships [7]. If we aim at finding out whether males and females differ in size only but
52 are otherwise built according to the same general principles, then OLS regression lines will typically
53 give the wrong answer. For correct interpretation of sexual dimorphism 'after accounting for
54 differences in size', regression slopes need to be adjusted for biological noise in the predictor
55 variable. In contrast, accounting only for technical measurement error in the predictor variable (as
56 suggested by Kilmer and Rodríguez [13]) is not sufficient for removing the spurious dimorphism.

57 (II) I suggest a different way of how to analyse whether digit ratio is sexually dimorphic after
58 accounting for body size. Rather than plotting one measure of size over another (2D over 4D, which
59 calls for major axis regression [7, 10, 15, 16]), I suggest to plot digit ratio directly over mean finger
60 length, and to adjust OLS regression slopes for the biological noise in the predictor. Intriguingly, such
61 analysis of the empirical data made available by Lolli and coworkers [6] indeed suggests that sexual
62 dimorphism in digit ratio may simply emerge from an allometric change in shape with size.

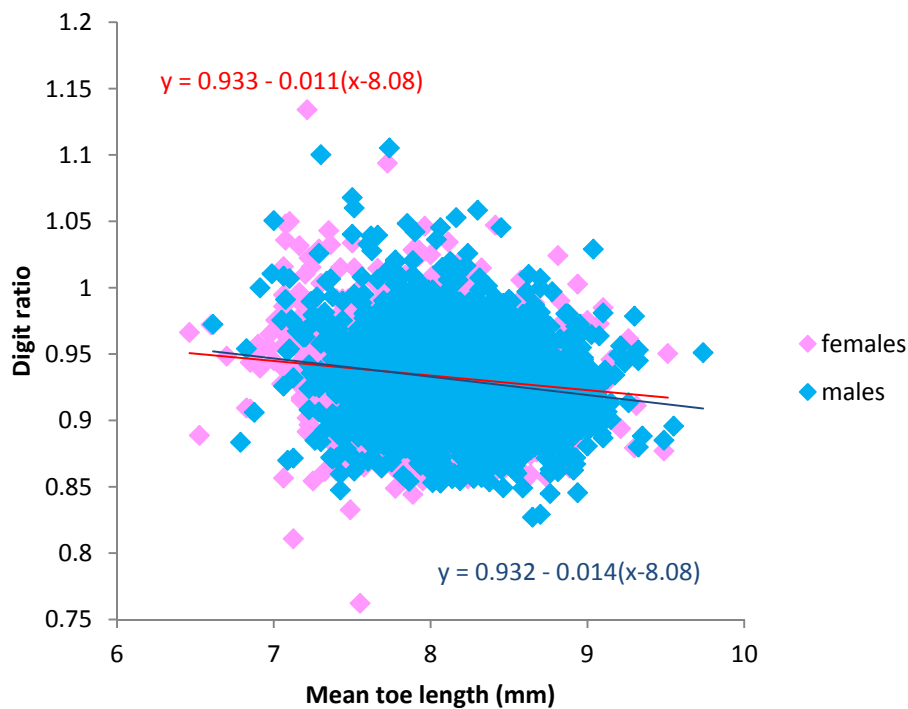
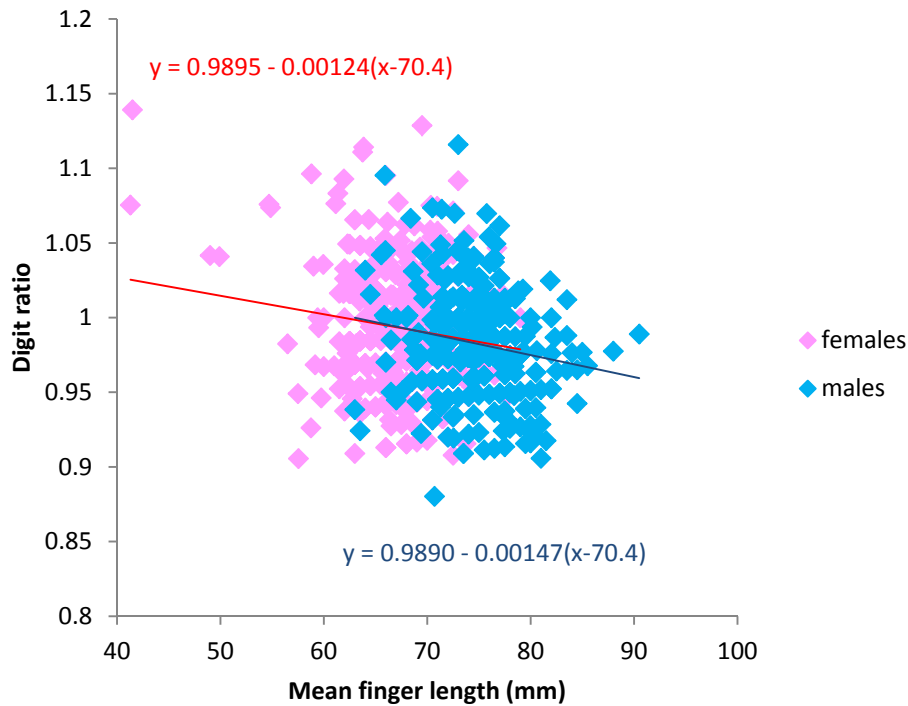
63 Ratios are widely used in many disciplines (e.g. body mass index, waist to hip ratio, body condition
64 index) because many ratios have advantageous properties, which however rarely get appreciated
65 explicitly [15]. More frequently, the inconsiderate use of ratios has been criticized [16-19], typically
66 because ratios often fail to achieve independence of body size. By their very nature, ratios (here
67 2D:4D) are strongly positively correlated with their numerator (here 2D) and strongly negatively
68 correlated with their denominator (here 4D), yet these are mathematical necessities that do not
69 reveal whether a ratio is size dependent. Ratios can in principle be independent of variation in body
70 size (see below), but whether real data on human digit ratios show such independence of size (here
71 average finger length) is a matter of empirical testing.

72 **Does digit ratio shift continuously with size?**

73 Figure 1a shows digit ratios calculated from the data of Lolli and coworkers [6] plotted over the mean
74 length of the two fingers. Intriguingly, within each sex, digit ratio declines as fingers get longer, and
75 the two OLS regression lines for males and females (the appropriateness of which I will address
76 below) practically coincide. This suggests that, as fingers get longer on average, the ratio 2D:4D is
77 shifting continuously, with 4D getting longer disproportionately and 2D lagging behind. Judging from
78 the indicated OLS regression lines, there is apparently no need to invoke any sexual dimorphism in
79 digit ratio beyond the allometric shift. This is because at the population-wide average finger length of
80 70.4 mm, men and women are predicted to have practically equal digit ratios (0.9890 and 0.9895
81 respectively), while without consideration of the allometric shift, the sexes are clearly different in
82 their average digit ratio (men: 0.9828, women: 0.9925, $t_{830} = 3.68$, $p = 0.0003$). Whether the data
83 shown in Figure 1a is truly representative of human allometry will have to be shown in the future by
84 additional analyses of larger data sets, but it seems possible that the phenomenon is more general.

85 The human digit ratio literature has also inspired similar research on numerous other tetrapod
86 species, many of which have also been found to be sexually dimorphic for digit ratio (e.g. [20-27]).
87 Hence, I examined own data on digit ratio in zebra finches (2D:4D, based on toes of the right foot;
88 see [28]). Figure 1b illustrates an interesting similarity with the human case, since within both males
89 and females digit ratio continuously declines as digits get longer and the two OLS regression lines
90 practically coincide. Since the toes of males are only slightly longer than those of females (8.14 mm
91 versus 8.02 mm) and the regression line is relatively shallow, male digit ratios are shifted to only
92 slightly smaller values (mean 0.9310) compared to those of females (mean 0.9336). This sexual
93 dimorphism in digit ratio only barely reaches statistical significance $p = 0.04$ despite a sample size of
94 $n = 3389$, and of course disappears when accounting for the allometric shift. This illustrates further
95 the need to reassess the presence of sexual dimorphism in digit ratio beyond what can be explained
96 by allometry [26], not only in humans but also in other species [20-25, 27].

97



98

99 *Figure 1. A: Digit ratios (2D:4D) of men and women in relation to mean finger length $((2D + 4D) / 2)$*
100 *calculated from the data provided by [6], both left and right hands analysed jointly. OLS regression*
101 *lines are indicated for each sex. B: Digit ratios (2D:4D) from the right foot of male and female zebra*
102 *finches in relation to mean toe length $((2D + 4D) / 2)$. Sex-specific OLS regression lines are shown.*
103 *Intercepts are reported at the population mean by subtracting the population mean from the*
104 *covariate.*

105 Previously [10], I argued that OLS regression can lead to erroneous conclusions about sexual
106 dimorphism because it fails to acknowledge that there is biological noise in the predictor variable
107 (e.g. when regressing 2D over 4D). If our interest lies in exploring sexual dimorphism, we should
108 consider major axis regression (MA) or reduced major axis regression (RMA) when regressing one
109 measure of size over another, because these methods assume equal amounts of statistical noise in
110 the two variables, which appears most sensible for two digits with comparable properties [10]. Now
111 in the above case (Fig. 1), I used OLS regression merely out of convenience, as is often done, but is
112 this really appropriate, given that the x-axis of Figure 1 contains biological noise? Below I will show
113 with simulated data that, in some sense, the regression lines in Figure 1 are indeed somewhat too
114 shallow, being biased downwards by the biological noise in the x-axis. The slopes should in fact be
115 adjusted (see [29]) if, and only if, the aim of line-fitting is to understand the general principles from
116 which males and females are built.

117 **Creating simulated data that include biological noise**

118 Closely adhering to the simulations that I carried out previously [10], I created artificial data on finger
119 lengths following specific predefined rules. In the first scenario, I assume isometry (like in [10]) simply
120 to illustrate how one gets to the situation where ratios are size independent. In the second scenario,
121 I incorporate an allometric shift with size (importantly, the same for the two sexes), and then
122 examine which regression method allows retrieving the correct value for the slope of the underlying
123 line around which the data have been generated. The data generation always consists of three steps:
124 (1) generating between-individual variation in latent size of fingers (representing the sum of genetic
125 and environmental effects that make the fingers of some individuals generally shorter or longer,
126 thereby inducing a strong positive correlation between 2D and 4D across individuals, $r = 0.9$ in [6]),
127 (2) applying underlying rules of isometry or allometric shift, and (3) adding individual biological noise
128 to each finger separately (representing the sum of unique genetic and environmental effects on each
129 finger, which induces the scatter around the regression lines in plots of 2D over 4D, and which is the
130 main source of individual differences in digit ratio).

131 For the first scenario of isometry, latent finger lengths of 5,000 women were drawn from a normal
132 distribution with a mean of 69 mm and a standard deviation (SD) of 5 mm. For 5,000 men I used a
133 mean of 77 mm and also a SD of 5 mm. Next, for women I assumed perfect isometry with a slope of 1
134 ($2D = 4D = \text{latent size}$), corresponding to a digit ratio of 1. For men, I assumed an isometric slope of
135 0.985 ($2D = 0.985 \cdot 4D = 0.985 \cdot \text{latent size}$), corresponding to a digit ratio of 0.985. Finally, for each
136 finger I drew biological noise from a normal distribution with a mean of zero and a SD of 2 mm. Yet
137 to reflect the fact that the amount of noise is typically proportional to size, I multiplied each value of
138 noise with the relative latent size (i.e. with latent size / 73 mm). This means that the noise was on

139 average 10% larger for an individual whose latent size was 10% above the population mean. After
140 adding the created noise to each finger separately, I calculated for each individual its realized mean
141 finger length $((2D + 4D) / 2)$ and its digit ratio $(2D / 4D)$.

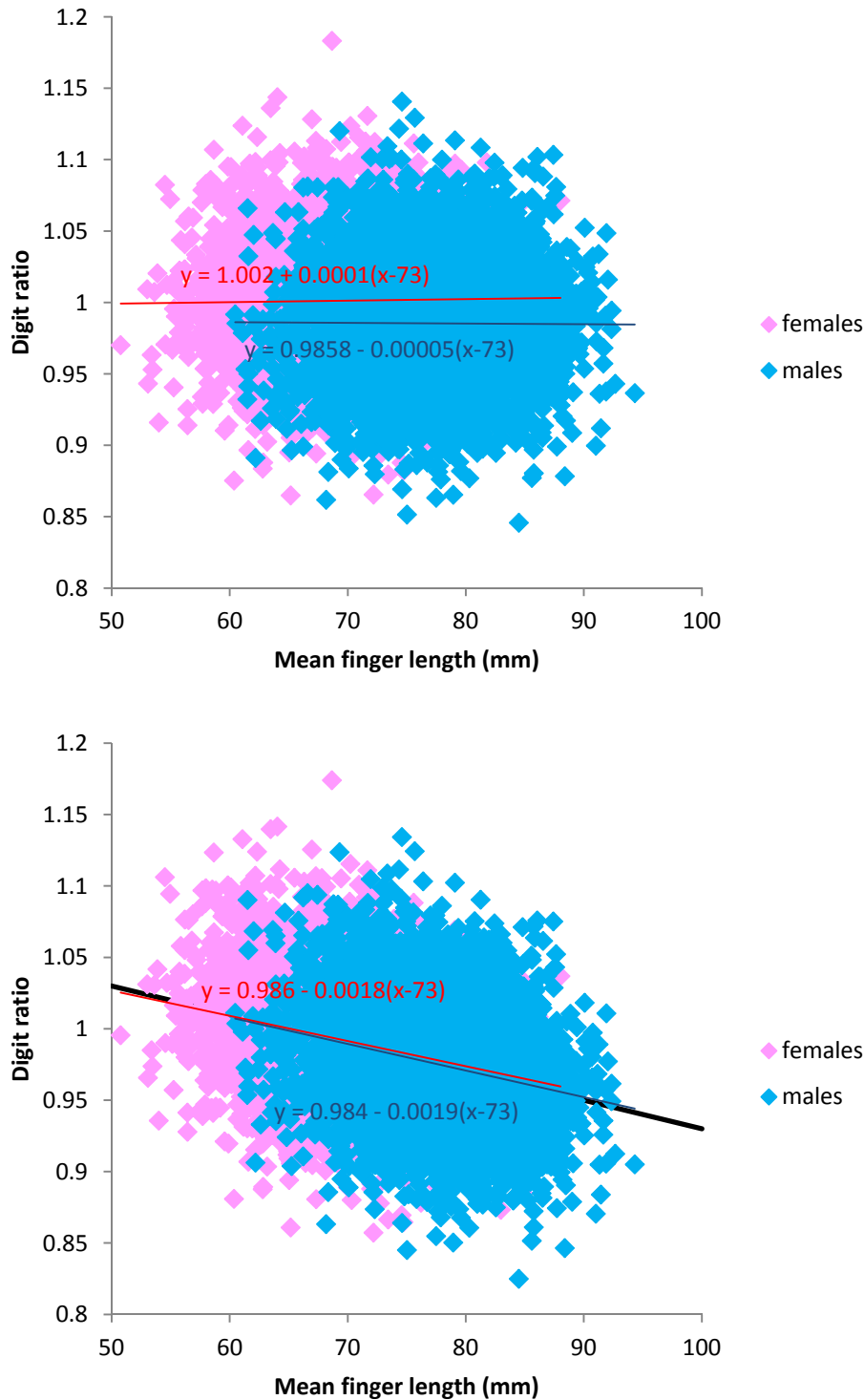
142 In the second scenario, latent sizes were created as above. Yet this time digit ratio was not assumed
143 to be size-independent, but (inspired by Figure 1a) rather to change systematically with latent size,
144 equally in the two sexes (expected digit ratio = $0.984 - 0.002 * (\text{latent size} - 73)$). Using this equation,
145 I calculated expected digit ratios for each individual from its latent size. The expected length of 2D
146 was then determined as $2D = (2 * \text{latent size} * \text{expected digit ratio}) / (1 + \text{expected digit ratio})$, and
147 the expected length of 4D as $4D = 2D / \text{expected digit ratio}$. After this, I again added biological noise
148 to each finger as described above and calculated realized mean finger length and digit ratio.

149 **OLS regression does not automatically retrieve the underlying principles**

150 Figure 2a illustrates the first scenario of isometry, where digit ratios are independent of size, and
151 where sexual dimorphism in digit ratio can be judged from the sex-specific intercepts at the
152 population mean size of 73 mm.

153 Figure 2b illustrates the second scenario of allometric shift where digit ratios are size dependent,
154 together with sex-specific OLS regression lines. The obtained OLS regression slopes for 5,000 males (-
155 0.0019, thin blue line in Fig. 2b) and 5,000 females (-0.0018, red line) were marginally shallower than
156 the slope of the line around which the data were generated (-0.0020, thick black line). Given this
157 marginal difference in slopes which could have arisen by chance from sampling noise, I repeated the
158 simulation based on 100,000 individuals of each sex and this was replicated 20 times to calculate a
159 between-replicate standard error of slope estimates, each obtained from 100,000 individuals.
160 Indeed, the average slope for males (-0.00182 ± 0.00001) and females (-0.00187 ± 0.00001) are
161 clearly shallower than the slope of the underlying digit ratio change with latent size (-0.00200).

162 The reason behind this attenuation in slopes lies in the fact that the x-axis of Figure 2b shows the
163 realized size (after adding biological noise to each finger), while the slope of the underlying
164 relationship (-0.002, black line) refers to the change with latent size (before adding the noise). Which
165 of these slopes is considered as 'correct' is an issue of the purpose of line fitting. The shallower OLS
166 regression lines represent the optimal solution for predicting for each sex an individual's digit ratio
167 from its realized mean finger length. However, these sex-specific lines are too shallow if one wants to
168 retrieve the rules according to which the data had been generated (namely a universal rule that is
169 shared between the sexes rather than two sex-specific rules).



170 Figure 2. Digit ratio in relation to mean finger length calculated from simulated data on 5,000
171 individuals of each sex, together with sex-specific OLS-regression lines. A: Case of isometry (size
172 independence) combined with sexual dimorphism in digit ratio. B: Case of an allometric shift in digit
173 ratio with size (mean finger length) that is shared between the sexes. Note that the thick black line in
174 the background ($y = 0.984 - 0.002(x-73)$) shows the underlying change in digit ratio with latent size
175 rather than with realized mean finger length.

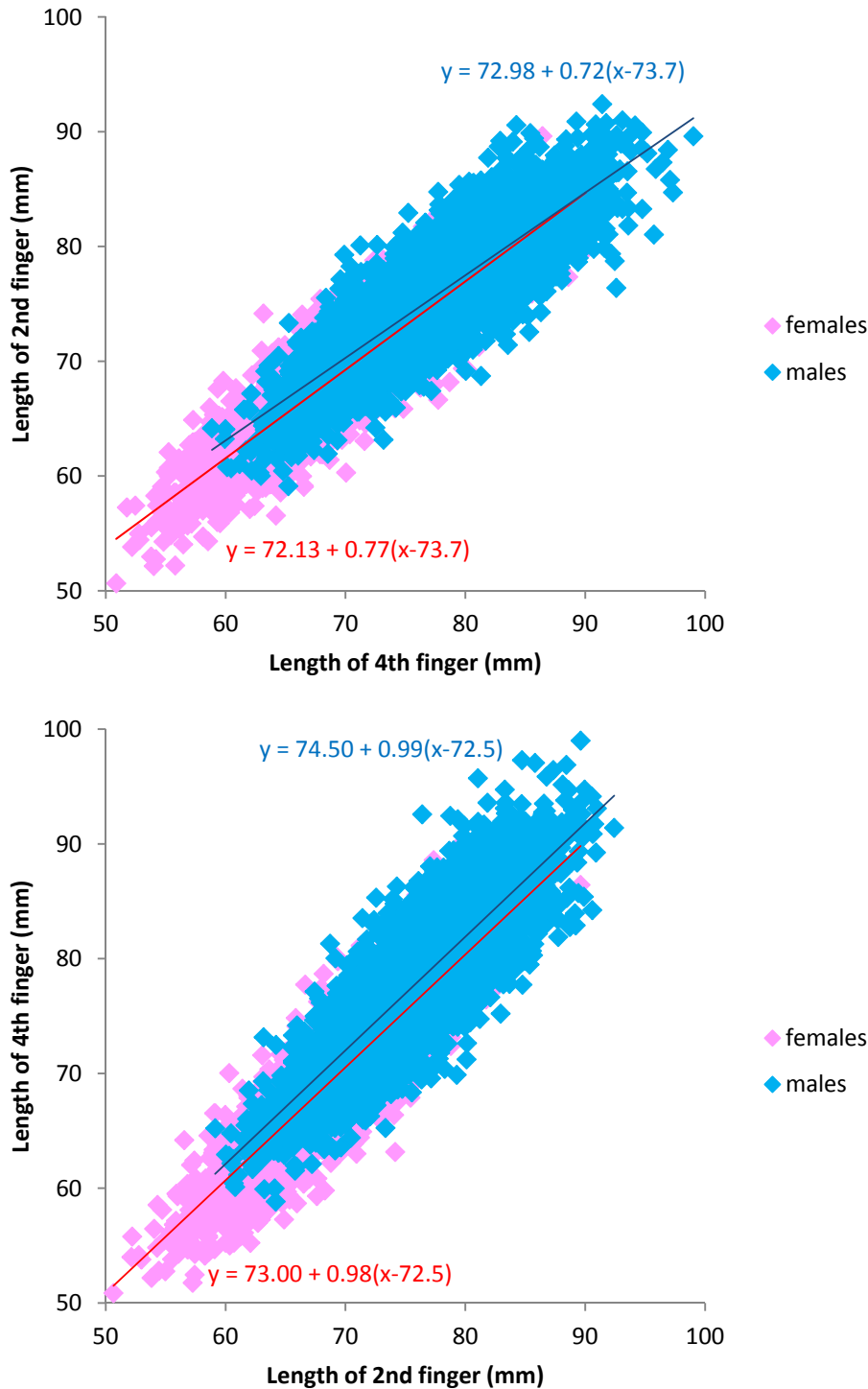
176

177 Importantly, the shallower OLS lines in Figure 2b are not “flawed” since the x-axis of Figure 2b shows
178 the realized mean finger length rather than latent size, but they can be misinterpreted. When
179 researchers say “after accounting for variation in size” they often forget that they have accounted for
180 ‘realized size’ but not for ‘latent size’ (which remains unknown in real data sets) and hence they
181 cannot retrieve the rules from which the sexes were built.

182 The degree of attenuation of slopes caused by the biological noise in the x-axis is relatively
183 straightforward to calculate for such simulated data (see [29]). For each sex, the between-individual
184 variance in latent size is 25 mm^2 (the square of the SD (= 5 mm) of the normal distribution from
185 which latent size was sampled). To each finger we added noise with a variance of 4 mm^2 (SD = 2 mm).
186 Since the realized average finger length is calculated from two fingers ($(D4 + D2) / 2$), the noise in the
187 averages is half of that in each finger ($4 \text{ mm}^2 / 2 = 2 \text{ mm}^2$). Given the additivity of variances, the
188 between-individual variance in realized mean finger length is $25 \text{ mm}^2 + 2 \text{ mm}^2 = 27 \text{ mm}^2$. Yet since
189 the amount of noise was scaled to latent size, there is overall $25 \text{ mm}^2 + 2.2 \text{ mm}^2 = 27.2 \text{ mm}^2$ for
190 males and $25 \text{ mm}^2 + 1.8 \text{ mm}^2 = 26.8 \text{ mm}^2$ for females. The ratio of variance in latent size to variance
191 in realized size ($25/27 = 0.926$) quantifies the amount of attenuation of slopes (0.918 for males and
192 0.933 for females). Dividing the obtained OLS regression slopes from Figure 2b by the attenuation
193 factor, retrieves the slope of the underlying relationship with latent size (-0.002) which was the same
194 by design for both sexes (males: $-0.00182 / 0.918 = -0.002$; females: $-0.00187 / 0.933 = -0.002$).

195 **Careful interpretation of regression lines is needed**

196 The above simulation illustrates that OLS-based general linear models should not be regarded as a
197 fail-proof tool that automatically provides the correct answer to whatever question one has in mind.
198 If we set up a scenario where the sexes only differ in their size but otherwise are constructed
199 according to identical rules, we would ideally like to be able to retrieve that information from our
200 data analysis. Specifically, after correcting for variation in size, we would like to see identical
201 intercepts for the two sexes. This was not the case here, because OLS slopes were systematically
202 biased downwards as a function of the amount of biological noise in the x-axis. In the present case of
203 Figures 1 and 2, the amount of downward bias in slopes is almost negligible (here about 7-8 %), such
204 that the resulting sexual dimorphism (women having slightly higher digit ratios than men at the
205 population average finger length of 73 mm, see Fig. 2b) is even difficult to detect. Here, OLS
206 regression fails to account for the biological noise in the x-axis, but the resulting difference in
207 intercepts is really minimal (though statistically significant, see final paragraph). The resulting bias is
208 minimal because, in this example, the y-axis contains much more biological noise than the x-axis
209 (here making OLS regression much more sensible than RMA regression).



210

211 Figure 3. The data that was generated for Figure 2b, plotted in a way where OLS-regression lines are
212 particularly confusing because in each case the blue line for males lies above the red line for females.
213 A: Plot of 2D over 4D with sex-specific OLS-regression lines suggesting that, at the population mean
214 of 4D of 73.7 mm, males have slightly longer 2D (72.98 mm) than females (72.13 mm) after
215 accounting for the variation in 4D. B: Plot of 4D over 2D with sex-specific OLS-regression lines
216 suggesting that, at the population mean of 2D of 72.5 mm, males have longer 4D (74.50 mm) than
217 females (73.00 mm) after accounting for the variation in 2D.

218 This situation clearly changes when examining apparent sexual dimorphism from OLS regression lines
219 in plots of one measure of size over another. To illustrate that point, I plotted the data from Figure
220 2b in such a way: once arbitrarily plotting 2D over 4D (Fig. 3a) and once arbitrarily plotting 4D over
221 2D (Fig. 3b). The OLS regression slopes show the expected bias, namely an apparent (deceptive)
222 sexual dimorphism with the larger sex (males) showing the 'relatively' larger trait values (irrespective
223 of what the trait is). The lines suggest that males have relatively longer 2D than females (by 0.85 mm
224 $= 72.98 \text{ mm} - 72.13 \text{ mm}$) after accounting for variation in 4D (Fig. 3a), and when plotting the very
225 same data the other way around, the lines suggest that males have relatively longer 4D than females
226 (by $1.50 \text{ mm} = 74.50 \text{ mm} - 73.00 \text{ mm}$) after accounting for variation in 2D (Fig. 3b). The fact that
227 these two interpretations appear paradoxical (it is confusing when both fingers are relatively longer
228 in males after accounting for size), illustrates that it is erroneous to interpret OLS regression lines in
229 such a way (see [7, 10, 11]).

230 Hence, OLS regressions should be interpreted with greatest care, and are not an easy tool in the
231 study of allometry. Prominent claims to the contrary [13] are based on the misunderstanding that
232 biological noise could be ignored when measurement error is accounted for. Measurement error is a
233 minor technical problem that can be effectively reduced by averaging multiple repeated measures of
234 an individual's phenotype. Such measurement error is not the primary reason why individual data
235 points in Figure 3 deviate from the regression lines [7, 8, 29]. The main reason for deviation from the
236 line lies in biological noise. Biological noise, in the present example, is the sum of genetic and
237 environmental effects that let an individual's phenotype deviate from its 'expected value according
238 to its latent size' (e.g. alleles that affect one finger but not the other) [8]. This biological noise is likely
239 of about equal magnitude for 2D and 4D, so it makes little sense to declare one of them to be the
240 dependent variable and the other one the predictor. In such symmetrical problems MA or RMA
241 regression is most helpful [7, 8, 29]. Surely OLS regression can be used in plots of one measure of size
242 over another (e.g. brain mass over body height; [30]) but the obtained regression lines should never
243 be interpreted as proof of sexual dimorphism in y after accounting for variation in x (e.g. the
244 erroneous claim that men have relatively larger brains than women after correcting for body size,
245 [10, 30]). Given that the explicit modelling of biological noise in the x-axis is still far from
246 commonplace, one should always be maximally cautious with formulations like 'after correcting for
247 differences in body size...'.
248

248 **Next steps for digit ratio research**

249 For the interpretation of digit ratio as a marker of sex-hormone effects, it appears essential to know
250 whether digit ratio is sexually dimorphic or not once accounting for any size-related shifts like in
251 Figure 1 (see also [26, 27]). This question should be relatively easy to answer, given that numerous

252 studies have already collected the relevant morphological measurements. Assuming that most
253 researchers are unfamiliar with Bayesian modelling of variation in latent size, what is the most
254 pragmatic way of distinguishing between the two scenarios illustrated in Figure 2a versus Figure 2b?

255 I recommend analysing digit ratio as the dependent variable in a general linear model (effectively an
256 ANCOVA) with sex as a fixed effect and mean finger length as a covariate (main effects without
257 interaction), and interpreting the model output carefully with the attenuation in OLS regression
258 slopes in mind. Such an analysis of the data shown in Figure 2a, yields a highly significant main effect
259 of sex (estimate -0.016 ± 0.001 , $t = -16.3$, $p < 10^{-16}$, note that $n = 10,000$) and a non-significant slope
260 estimate for the covariate 'mean finger length' (estimate 0.00003 ± 0.00007 , $t = 0.4$, $p = 0.7$).

261 Essentially, we want to verify that there is not an overall negative regression slope (like in Figure 2b)
262 that runs through the average coordinates of the male data and the average coordinates of the
263 female data. The slope of such a hypothetical line can be calculated as (average male digit ratio –
264 average female digit ratio) / (average male finger length – average female finger length) = -0.00200 .
265 The lower end of the 95% CI of the model's slope estimate ($0.00003 - 1.96 * 0.00007 = -0.00012$) lies
266 still clearly above that value, even when considering that there is some attenuation of the OLS
267 regression slope.

268 For comparison, the analysis of the data from Figure 2b, yields a small and only barely significant
269 main effect of sex (estimate -0.002 ± 0.001 , $t = -2.3$, $p = 0.02$) and a highly significant negative slope
270 for the covariate 'mean finger length' (estimate -0.0018 ± 0.00007 , $t = -24.4$, $p < 10^{-16}$). Thanks to the
271 very large sample size, the artefacts of spurious dimorphism and attenuation of slopes become
272 visible. Although it seems like there is still a small but significant effect of sex, and also the lower end
273 of the 95% CI of the slope estimate (-0.00196) is only barely above the hypothetical line (-0.00200),
274 these minor deviations should be attributed to the attenuation of the OLS regression slope. How
275 much attenuation should be allowed for? To find out empirically, I modified the parameters of the
276 data simulation shown in Figure 2b until it most closely resembled Figure 1a (when using SD of latent
277 size = 4.2 and SD of noise = 1.85). With these parameters the attenuation of slopes is about 10% (4.2^2
278 / ($4.2^2 + 0.5 * 1.85^2$) = 0.91), so slope estimates should be multiplied by the inverse of that ($1 / 0.91 =$
279 1.097). Generally, the lower the Pearson correlation between 2D and 4D across both sexes (e.g. $r =$
280 0.9 in Figure 3), the higher the noise component and the stronger the attenuation of slopes in our
281 analysis of Figure 2b.

282 Several studies (e.g. [20, 21, 26]) have assessed the size-dependence of digit ratio by plotting it over
283 other measures of size than mean finger length (e.g. body mass), which bears a risk of overlooking a
284 size dependence because the covariate might contain additional biological noise that is not relevant

285 for digit ratio. Hence regressing digit ratio over mean finger length appears most effective in
286 identifying a direct dependency.

287 Sample size is important to consider, if our aim is to reject the idea that there is an allometric shift in
288 digit ratio like shown in Figure 1a. In this analysis, I have treated measurements from both hands as
289 equivalent in order to maximise statistical power. Generally, I would recommend running separate
290 analyses for each hand, only if the patterns differ significantly between hands. In Figure 1a, the
291 Pearson correlation coefficient between digit ratio and mean finger length (across both sexes) is
292 about $r = -0.20$. A total sample size of about 260 individuals is required to achieve a power of 95% for
293 detecting such a correlation (with $\alpha = 0.05$; calculated using G*Power 3.1.7, [31]). If one aims at
294 detecting correlations even within each of the two sexes (here $r = -0.15$ for females and $r = -0.17$ for
295 males), a sample size of about 470 individuals per sex would be required for a statistical power of
296 95%. Studies with a lower statistical power (e.g. [32]) should still report the overall correlation
297 between digit ratio and mean finger length, for a later meta-analytic summary.

298 If the patterns shown in Figure 1 should turn out to be representative of humans and also other
299 animals, it is not easy to see how digit ratio could be a more informative indicator of sex-hormone
300 effects than for instance mean digit length. Ignoring the allometric shift, the sex difference in human
301 digit ratio is only about 0.26 within-sex standard deviations (Cohen's D calculated from Figure 1a),
302 while the sex difference in mean finger length is about six times larger (Cohen's D = 1.56, Figure 1a),
303 hence the latter somehow appears to have greater potential in reflecting levels of sex hormones
304 experienced during development.

305 Proponents of the idea that digit ratio is informative about sex-hormone effects face a two-fold
306 challenge. First, they should demonstrate that sexual dimorphism in digit ratio exists independently
307 of the dimorphism in size (Figure 2a versus Figure 2b), and that the trait is more informative than
308 other sexually dimorphic characters. Second, when testing for the allometric shift, they should
309 demonstrate the objectivity of their analyses in spite of a possible conflict of interest or at least some
310 wishful thinking. To show that one has not e.g. selectively removed outliers that appear influential
311 after inspecting the data (see top left corner of Fig. 1a), I suggest to first deposit complete data sets
312 that have been used in a previous publication (that allow reconstructing the results shown in that
313 publication), and then to analyse the complete data set, following the approach outlined above. This
314 effectively eliminates most 'researcher degrees-of-freedom' [33] and thereby demonstrates a
315 maximum of scientific objectivity [34].

316 Acknowledgements: I thank Stefan van Dongen, Pim Edelaar and Jarrod Hadfield for valuable
317 comments.

318 References

- 319 [1] Honekopp, J. & Watson, S. 2010 Meta-Analysis of Digit Ratio 2D:4D Shows Greater Sex Difference
320 in the Right Hand. *American Journal of Human Biology* **22**, 619-630. (doi:10.1002/ajhb.21054).
- 321 [2] Manning, J.T. 2002 *Digit ratio: a pointer to fertility, behavior, and health*. New Brunswick, NJ,
322 Rutgers University Press.
- 323 [3] Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R. & Manning, J.T. 2004 2nd to 4th digit
324 ratios, fetal testosterone and estradiol. *Early Human Development* **77**, 23-28.
325 (doi:10.1016/j.earlhumdev.2003.12.002).
- 326 [4] Manning, J.T., Scutt, D., Wilson, J. & Lewis-Jones, D.I. 1998 The ratio of 2nd to 4th digit length: a
327 predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen.
328 *Human Reproduction* **13**, 3000-3004.
- 329 [5] Zheng, Z.G. & Cohn, M.J. 2011 Developmental basis of sexually dimorphic digit ratios. *Proceedings*
330 *of the National Academy of Sciences of the United States of America* **108**, 16289-16294.
331 (doi:10.1073/pnas.1108312108).
- 332 [6] Lolli, L., Batterham, A.M., Kratochvíl, L., Flegr, J., Weston, K.L. & Atkinson, G. 2017 A
333 comprehensive allometric analysis of 2nd digit length to 4th digit length in humans. In *Proc. R. Soc. B*
334 (p. 20170356, The Royal Society).
- 335 [7] Smith, R.J. 2009 Use and Misuse of the Reduced Major Axis for Line-Fitting. *American Journal of*
336 *Physical Anthropology* **140**, 476-486. (doi:10.1002/ajpa.21090).
- 337 [8] Shingleton, A.W. 2017 Which line to follow? The utility of different line-fitting methods to capture
338 the mechanism of morphological scaling. *bioRxiv*. (doi:10.1101/216960).
- 339 [9] Kratochvíl, L. & Flegr, J. 2009 Differences in the 2nd to 4th digit length ratio in humans reflect
340 shifts along the common allometric line. *Biology letters* **5**, 643-646.
- 341 [10] Forstmeier, W. 2011 Women have relatively larger brains than men: a comment on the misuse
342 of general linear models in the study of sexual dimorphism. *The Anatomical Record* **294**, 1856-1863.
- 343 [11] Schluter, D. 1992 Brain Size Differences. *Nature* **359**, 181-181.
- 344 [12] Beaty, L.E., Emmering, Q.C. & Bernal, X.E. 2016 Mixed Sex Effects on the Second-to-Fourth Digit
345 Ratio of Túngara Frogs (*Engystomops pustulosus*) and Cane Toads (*Rhinella marina*). *The Anatomical*
346 *Record* **299**, 421-427.
- 347 [13] Kilmer, J.T. & Rodríguez, R.L. 2016 Ordinary least squares (OLS) regression is indicated for studies
348 of allometry. *Journal of evolutionary biology*.
- 349 [14] Green, A.J. 2001 Mass/length residuals: Measures of body condition or generators of spurious
350 results? *Ecology* **82**, 1473-1483.
- 351 [15] Smith, R.J. 2005 Relative size versus controlling for size - Interpretation of ratios in research on
352 sexual dimorphism in the human corpus callosum. *Current Anthropology* **46**, 249-273.
- 353 [16] Ranta, E., Laurila, A. & Elmberg, J. 1994 Reinventing the Wheel - Analysis of Sexual Dimorphism
354 in Body-Size. *Oikos* **70**, 313-321.
- 355 [17] Packard, G.C. & Boardman, T.J. 1988 The Misuse of Ratios, Indexes, and Percentages in
356 Ecophysiological Research. *Physiological Zoology* **61**, 1-9.
- 357 [18] Kratochvíl, L. & Flegr, J. 2017 The mismeasure of (obese) man - body mass index is an inadequate
358 indicator of body condition. *bioRxiv*. (doi:10.1101/149088).
- 359 [19] Nakagawa, S., Kar, F., O'Dea, R.E., Pick, J.L. & Lagisz, M. 2017 Divide and conquer? Size
360 adjustment with allometry and intermediate outcomes. *BMC biology* **15**, 107.
- 361 [20] Burley, N.T. & Foster, V.S. 2004 Digit ratio varies with sex, egg order and strength of mate
362 preference in zebra finches. *Proceedings of the Royal Society of London Series B-Biological Sciences*
363 **271**, 239-244. (doi:10.1098/rspb.2003.2562).
- 364 [21] Kaczmarski, M., Kubicka, A.M., Tryjanowski, P. & Hromada, M. 2015 Females Have Larger Ratio
365 of Second-to-Fourth Digits Than Males in Four Species of Salamandridae, Caudata. *Anatomical*
366 *Record-Advances in Integrative Anatomy and Evolutionary Biology* **298**, 1424-1430.
367 (doi:10.1002/ar.23123).

- 368 [22] Leoni, B., Canova, L. & Saino, N. 2005 Sexual dimorphism in metapodial and phalanges length
369 ratios in the wood mouse. *Anatomical Record Part a-Discoveries in Molecular Cellular and*
370 *Evolutionary Biology* **286A**, 955-961. (doi:10.1002/ar.a.20226).
- 371 [23] Leoni, B., Rubolini, D., Romano, M., di Giancamillo, M. & Saino, N. 2008 Avian hind-limb digit
372 length ratios measured from radiographs are sexually dimorphic. *Journal of Anatomy* **213**, 425-430.
373 (doi:10.1111/j.1469-7580.2008.00964.x).
- 374 [24] Lombardo, M.P. & Thorpe, P.A. 2008 Digit ratios in green anolis lizards (*Anolis carolinensis*).
375 *Anatomical Record-Advances in Integrative Anatomy and Evolutionary Biology* **291**, 433-440.
376 (doi:10.1002/ar.20657).
- 377 [25] Rubolini, D., Pupin, F., Sacchi, R., Gentili, A., Zuffi, M.A.L., Galeotti, P. & Saino, N. 2006 Sexual
378 dimorphism in digit length ratios in two lizard species. *Anatomical Record Part a-Discoveries in*
379 *Molecular Cellular and Evolutionary Biology* **288A**, 491-497.
- 380 [26] Direnzo, G.V. & Stynoski, J.L. 2012 Patterns of Second-to-Fourth Digit Length Ratios (2D: 4D) in
381 Two Species of Frogs and Two Species of Lizards at La Selva, Costa Rica. *The Anatomical Record* **295**,
382 597-603.
- 383 [27] Gomes, C.M. & Kohlsdorf, T. 2011 Evolution of sexual dimorphism in the digit ratio 2D: 4D-
384 relationships with body size and microhabitat use in iguanian lizards. *PloS one* **6**, e28465.
- 385 [28] Forstmeier, W., Mueller, J.C. & Kempenaers, B. 2010 A polymorphism in the oestrogen receptor
386 gene explains covariance between digit ratio and mating behaviour. *Proceedings of the Royal Society*
387 *B-Biological Sciences* **277**, 3353-3361. (doi:10.1098/rspb.2010.1007).
- 388 [29] Warton, D.I., Wright, I.J., Falster, D.S. & Westoby, M. 2006 Bivariate line-fitting methods for
389 allometry. *Biological Reviews* **81**, 259-291.
- 390 [30] Ankney, C.D. 1992 Sex differences in relative brain size: The mismeasure of woman, too?
391 *Intelligence* **16**, 329-336.
- 392 [31] Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. 2009 Statistical power analyses using G* Power
393 3.1: Tests for correlation and regression analyses. *Behavior research methods* **41**, 1149-1160.
- 394 [32] Folland, J.P., Mc Cauley, T.M., Phipers, C., Hanson, B. & Mastana, S.S. 2012 Relationship of 2D:
395 4D finger ratio with muscle strength, testosterone, and androgen receptor CAG repeat genotype.
396 *American journal of physical anthropology* **148**, 81-87.
- 397 [33] Simmons, J.P., Nelson, L.D. & Simonsohn, U. 2011 False-Positive Psychology: Undisclosed
398 Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological*
399 *Science* **22**, 1359-1366. (doi:10.1177/0956797611417632).
- 400 [34] Forstmeier, W., Wagenmakers, E.J. & Parker, T.H. 2017 Detecting and avoiding likely false-
401 positive findings - a practical guide. *Biological Reviews* **92**, 1941-1968. (doi:10.1111/brv.12315).

402