Sex differences in allometry for phenotypic traits indicate that females are not scaled males

Laura A. B. Wilson[1,2*], Susanne R. K. Zajitschek[1,3], Malgorzata Lagisz[1], Jeremy Mason[4,5], Hamed Haselimashhadi[5] Shinichi Nakagawa[1*]

[1]Evolution & Ecology Research Centre, UNSW Data Science Hub, and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, Australia

[2]School of Archaeology and Anthropology, The Australian National University, Canberra, Australia

[3]School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK

[4]Melio Healthcare Ltd., City Tower, 40 Basinghall St., EC2V 5DE, London, UK

[5]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK

*correspondence: laura.wilson@anu.edu.au; s.nakagawa@unsw.edu.au

Abstract

Sex differences in the lifetime risk and expression of disease are well-known. Paradoxically, preclinical research targeted at improving treatment, increasing health span and reducing the financial burden of health care, has mostly been conducted on male animals and cells. Females are assumed to be the same or scaled versions of males, yet sex differences in the allometric relationship between phenotypic traits and body size, needed to evaluate the validity of this assumption, have not been established. We quantify allometry for 297 phenotypic traits in male and female mice, recorded in >2.1 million measurements from the International Mouse Phenotyping Consortium. We find sex differences in allometric parameters (slope, intercept, residual SD) are common. Thus, the allometric relationship varies between the sexes: females are not scaled males. Our results support a complex, trait-specific patterning of sex differences in phenotypic traits, promoting case-specific approaches to therapeutic intervention and drug dosage scaled by body weight.

Keywords: biomarker, sex differences, static allometry, animal model, drug reactions

Introduction

A historic use of male animals in preclinical research and male participants in clinical trials has resulted in a significant bias in healthcare systems around the world (Holdcroft, 2007). The knowledge available on many diseases, their manifestation, time course and the efficacy of treatment options, is highly skewed in favour of males. The need to reach parity of the sexes in biomedical research and to conduct sex-specific analysis of research results has been widely acknowledged (Mogil & Chanda, 2005; Rogers et al., 2008; Kim et al., 2010; Beery & Zucker, 2011; Klein et al., 2015). Efforts to address this issue initially resulted in legislative changes around clinical research, requiring female participants in government-funded clinical trials (e.g., NIH, 1993; Correa-de-Araujo, 2006; Klinge, 2008). Modest improvement to rebalancing representation of the sexes in clinical trials (Zucker & Beery, 2010; Mazure & Jones, 2015; Feldman et al., 2019) has been bolstered by recent revisions to government guidelines in the US for preclinical research, requiring biological sex to be included as a study variable (Clayton & Collins, 2014).

Basing healthcare decisions for women based on research conducted on men (and vice versa, e.g., Wiemann et al., 2007) potentially has profound consequences (Kim et al., 2010; Oh et al., 2015; Tannenbaum et al., 2019). Studies have established that the nature of disease experience and benefits of treatment differ between men and women (Rahore et al., 2002; Gandhi et al., 2004; Canto et al., 2007; Whitley et al., 2009; Wallach et al., 2016; Mauvais-Jarvis et al., 2020). These differences manifest in major pillars of healthcare, impacting cost associated with care and its quality (Wainer et al., 2020). For example, sex differences in pharmacokinetics mean that therapeutic decisions based on studies with male subjects may lead to increased magnitudes of adverse drug reactions in women (Nakagawa & Kajiwara, 2015, Yu et al. 2016). Similarly, the broadly divergent behaviour of male (anti-inflammatory) and female (pro-inflammatory) immune systems translates to antibody response variability, with some vaccines resulting in a stronger immune response in males compared to females (Bouman & Heineman, 2005; Cook, 2008; Klein, 2013; Flanagan, 2014). Moreover, pathophysiological differences between the sexes lead to women being underdiagnosed or undertreated for leading causes of mortality, such as cardiovascular disease and Type 2 diabetes (Mauvais-Jarvis et al., 2020).

With the growing recognition of the importance of sex in biomedicine, a sharper focus on the topic has revealed that some of the initial assumptions and concerns surrounding use of female animals in preclinical research, such as their propensity for greater variation associated with the oestrous cycle (Shansky, 2019), lack empirical support (Mogil & Chanda, 2005; Prendergast et al., 2014; Zajitschek et al. 2020). Nevertheless, questions have been

raised about the value of including female animals in preclinical research, citing a negative impact on the burden of evidence for therapeutic interventions (Fields 2014) and a lack of clarity surrounding the extent to which sex differences may be explained by sex-linked variables, such body mass index or body weight differences between the sexes (Richardson 2015).

Building on empirical studies that have sought to establish the nature of sex differences in biomedicine and to clarify the assumptions surrounding preclinical (Mogil & Chanda, 2005; Becker et al., 2016; Karp et al., 2017; Zajitschek et al., 2020) and clinical (Campesi et al., 2021) research data collected on males and generalized to females, we here tackle the extent to which females can be considered 'small' males in biomedicine. This is a pervasive narrative that impacts research design.

We adopt the framework of static allometry, the measurement of trait covariation among individuals of different size at the same developmental stage, following Huxley (1924, 1932), who proposed an equation to model simple allometry. This equation expresses the growth of two traits, x and y, when regulated by a common growth parameter: $y = ax^b$, or equivalently $\log y = \log(a) + b \log(y)$, where the ratios between the components of the growth rates of y and x correspond to intercept log(a) and a slope b (Pélabon et al., 2013). We quantify the relationship between phenotypic trait and body weight in males and females, statistically evaluating scenarios that describe the magnitude and patterning of sex differences across 297 traits in over 2 million mice from the International Mouse Phenotyping Consortium (IMPC, www.mousephenotype.org; Dickinson et al., 2016).

By providing empirical data on static allometry across phenotypic traits that represent preclinical parameters (e.g., immunology, metabolism, morphology), we aim to clarify if, and the extent to which, trait values for males may be scaled to match those of females. That is, we tackle the assumption that females are small males and identify, for the first time, the trait-specific features of the allometric relationship. We discuss these data considering the discourse on the generalization of male data in preclinical research (Usui et al. 2021), as well as their evolutionary implications, leveraging a large, wildtype dataset to illuminate microevolutionary trends in static allometry. Consideration of the evolutionary context surrounding sex differences may augment understanding of how disease state phenotypes emerge or persist in a population (Morrow & Connallon, 2013; Morrow, 2015). Data on allometric scaling also relate to one of the most salient aspects of sex differences, those concerning adverse drug reactions (ADRs) and the so far unanswered question of whether

weight-adjusted doses would suffice to offset the majority of sex-specific ADRs (Zucker & Prendergast, 2020).

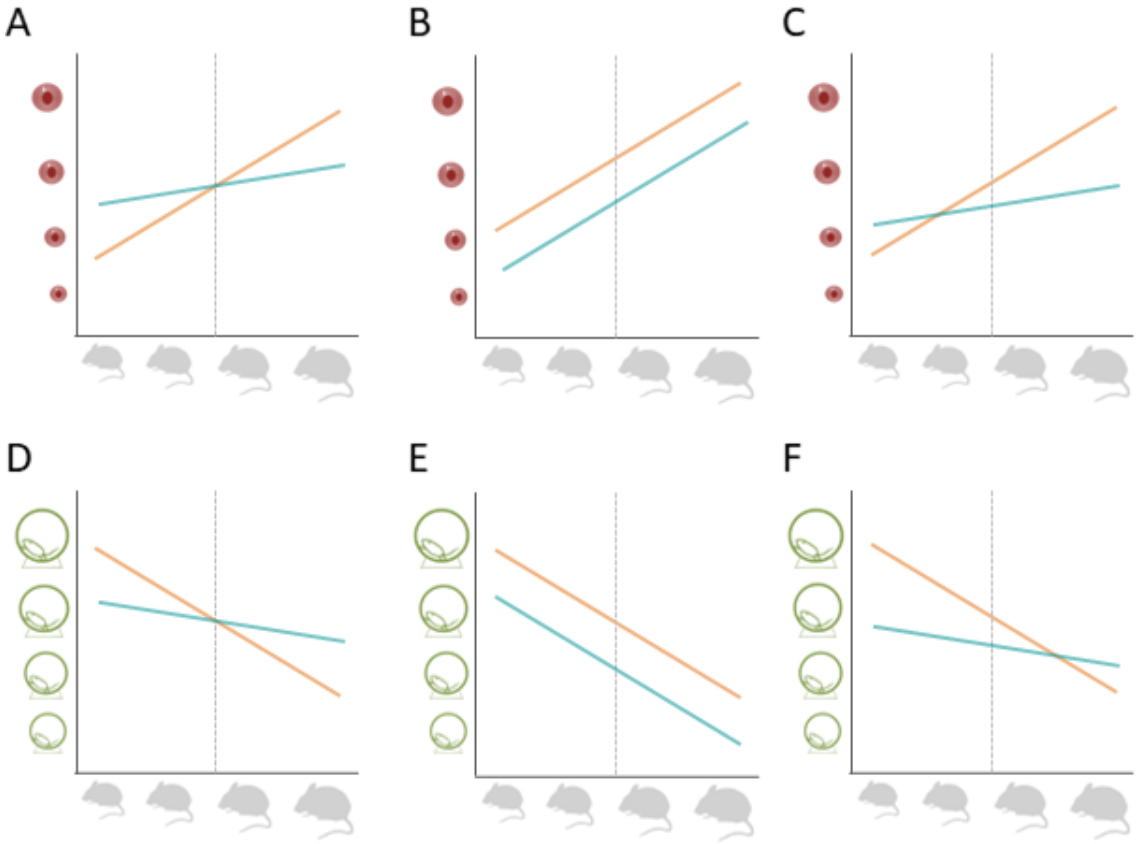

Figure 1. Examples of scenarios of sex differences in a trait of interest ~ weight allometric relationship. Top row shows a hypothetical positive relationship between body weight and eye size and the bottom row negative relationship between body size and activity. Body weights are scaled and centred so that the intercept is at the trait mean represented by a grey dashed line. A) Different positive slopes for the sexes, but same intercepts. B) Same positive slopes for both sexes, but different intercepts. C) Different positive slopes for both sexes, and different intercepts. D) Different negative slopes for the sexes, but the same intercepts. E) Same negative slopes for both sexes, but different intercepts. F) Different negative slopes for both sexes and different intercepts.

Results

Data characteristics

Following initial data cleaning and filtering procedures, the dataset comprised 297 phenotypic traits with a median sample size of 1,585 mice per trait (n = 2,104,527). Representation of males and females was highly similar across most phenotypic traits, with fewer than 10% of traits (29/297) displaying greater than 5% difference in sample size between males and females. The traits were collated into nine functional groupings following Zajitschek et al. (2020) (see Methods): behaviour (57 traits, n = 484,207), eye (27 traits, n = 10,366), hearing (16 traits, n = 201,220), heart (27 traits, n = 196,777), hematology (25 traits,

n = 300,699), immunology (79 traits, n = 89,952), metabolism (9 traits, n = 111,659), morphology (24 traits, n = 364,484), and physiology (33 traits, n = 345,163).

The 297 phenotypic traits were further filtered for non-independence of traits, so that *p* values were merged for traits that were related to one another, resulting in a reduced data set of 181 traits, with a median sample size of 4,044 individuals per trait.

Linear mixed-effects models for static allometry

Our linear mixed-effects models indicated that 8 out of 181 traits (4%) (13 / 297 traits for unmerged p-values) are associated with scenario A (different slope, same intercept, Fig. 1A, 1D); most of these traits belonged to immunology and heart functional groups. Note that the intercept for each sex was set so that we compared mean values for each sex for a given trait. Scenario B (same slope, different intercept, Fig. 1B, 1E) was supported for 70 / 181 (39%) traits (125 / 297 traits for unmerged p-values). For scenario C (different slope, different intercept, Fig. 1C, 1F), 69 / 181 (38%) traits were categorized as consistent (86 / 297 traits for unmerged p-values), and the remaining 34 / 181 (19%) traits showed no significant differences in slope and intercept between males and females. Overall, when a statistically significant difference in allometric pattern was present between the sexes, intercept differences appeared more common than slope differences (39% compared to 4% traits), however both slope and intercept differences were also similarly common (38%). Just under a fifth of traits showed no significant differences between males and females, indicating that, for most traits, sex differences in allometric patterning represent a significant source of variation in trait values.

Taken together, traits in all functional groups showed statistically significant ($\alpha = 0.05$) sex differences. Slope differences between the sexes (scenario A) e most common in immunology and heart groups, while intercept differences (scenario B) were most common for traits in the behaviour and heart functional groups. Traits exhibiting both slope and intercept differences between the sexes (scenario C) were most commonly found in the metabolism and physiology functional groups. Non-significant differences in slope and intercept were most common among traits in the behaviour and morphology functional groups.

Sex bias in allometric parameters

Sex bias in the slope and intercept values, in addition to the magnitude of variance (residual SD), showed considerable variability across functional groups, suggesting trait-specific patterning of sex differences. For scenario A, representing traits with significant differences

in slope, most traits showed greater slope magnitudes for males (n = 6 traits), rather than for females (n = 4 traits) (Fig. 2A). For scenario B, females showed greater intercept magnitudes for morphology, immunology, eye and behaviour functional groups (n = 45 traits), whereas males showed greater intercepts for traits in physiology, metabolism, hematology, heart and hearing functional groups (n = 32 traits) (Fig. 2B). Overall sex bias (65 male traits: 60 female traits, Fig. 2B) was slightly greater for intercept differences, compared to slope differences (7 male traits: 6 female traits, Fig. 2A). Scenario C, which represents significant slope and intercept parameter differences between the sexes, was predominated by mixed bias across five out of nine functional groups (n = 24 traits), indicating that most functional groups contained traits that showed a mixture of directional differences in bias, comprising a combination of male bias in one parameter (slope or intercept) and female bias in the other parameter (slope or intercept) (Figure 2C). Immunology-related traits represent an exception under scenario C, whereby traits with significant differences between the sexes did not show a mixed bias for slope and intercept values. Across functional groups, male bias is slightly more common (5 groups) than female bias (4 groups) for statistically significant sex difference in residual SD, indicating that where traits show differences between the sexes, it is more common for males to be more variable than females, than *vice versa* (Figure 2D) (133 male traits: 71 female traits).

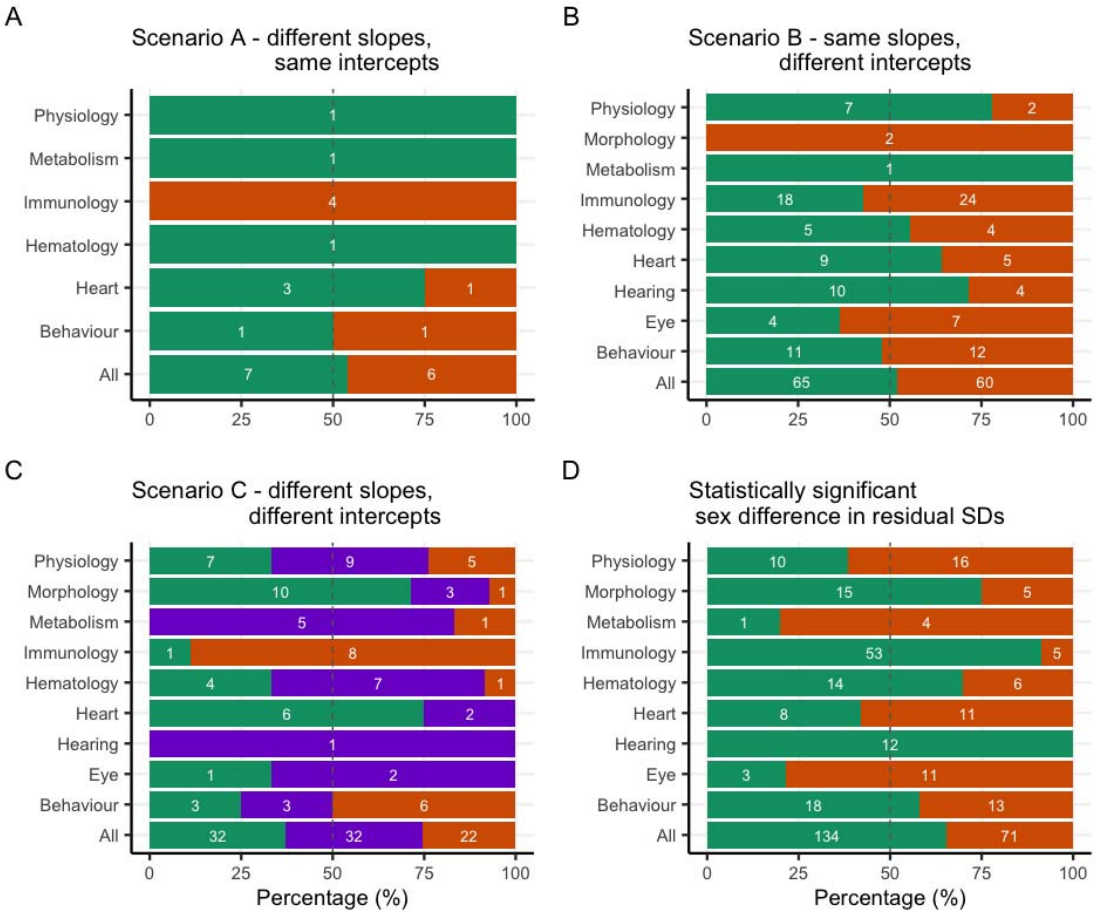

Figure 2. Sex biases for mice phenotypic traits arranged in functional groups. Colours represent significant differences in trait values between the sexes (green – male biased, orange – female biased) for allometric slope (scenario A), intercept (scenario B) or slope and intercept, including traits with mixed (purple) significant differences (i.e. male-biased significant slope and female-biased significant intercept, or female-biased significant slope and male-biased significant intercept) (scenario C), and bias in statistically significant difference in variance (residual SD) between the sexes (D). The number of traits that are either female biased (relative length of orange bars) or male biased (relative length of green bars) are expressed as a percentage of the total number of traits in the corresponding group. Numbers inside the green bars represent the numbers of traits that show female bias within a given group of traits, values inside the orange bars represent the number of male biased traits, and those inside the purple bars represent a combination of female bias (for intercept or slope) and male bias (for intercept or slope).

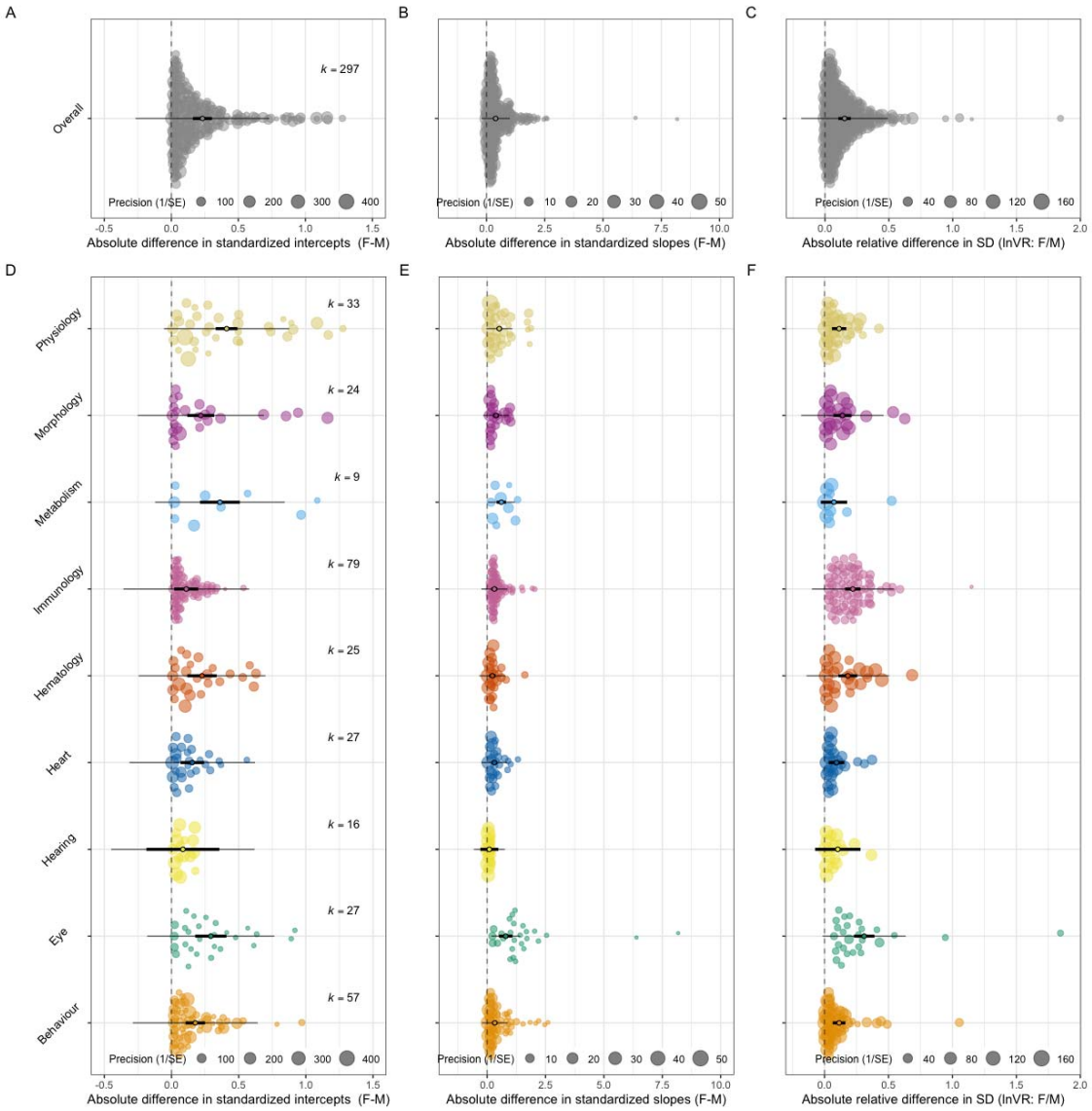

Figure 3. Orchard plots illustrating results of multivariate meta-analysis based on differences between male and female absolute values for allometric intercept (A, D), slope (B, E) and residual variance (SD) (C, F). Plots in greyscale (top row) show overall differences (A – C), and plots below, in colour, show separate results for each functional group (D – F). Orchard plots show model point estimate (black open ellipse) and associated confidence interval (CIs) (thick black horizontal line), 95% prediction intervals (PIs) (thin black horizontal line), and individual effect sizes (filled ellipses), which are scaled by their precision, defined as: precision = 1 / Standard Error (SE) (see Nagakawa et al., 2021).

Meta-analysis and meta-regression of sex differences in slope, intercept and variance

Multi-level meta-analysis of absolute values in allometric slope and intercept, and variance, revealed significant differences between the sexes (Fig. 3A – C), with the greatest effect size evident for intercept value (Fig. 3A). Across functional groups, there was variability in the

magnitude of absolute difference between the sexes, both within parameters (i.e., intercept) and across parameters. For absolute differences in intercept, traits within the physiology functional group showed greatest model point estimate difference between males and females, whereas those within the hearing group showed the smallest magnitude of difference (Fig. 3D). For differences in slope, which showed lower inter-trait variability than differences in intercept, the largest model point estimate difference was observed for eye traits, and the smallest difference for hearing traits (Fig. 3E).

Similarly, for relative difference in residual SD, eye traits showed the largest amount of dimorphism, whereas heart and metabolism traits were most similar in SD values between the sexes (Fig. 3F). Overall, across all parameters (intercept, slope and SD), confidence intervals (CIs) for hearing traits were the only ones to consistently overlap with zero, showing no statistically significant difference between the sexes (Fig. 3D, E, F). For traits within a given functional group, there was considerable variability in the magnitude of difference between the sexes. For sex differences in intercept, inter-trait variability was highest for physiology, morphology and metabolism groups (Fig. 3D), whereas slope differences showed most inter-trait variability for eye and behaviour traits (Fig. 3E), and relative difference in SD was also most variable among traits in eye and behaviour groups (Fig. 3F).

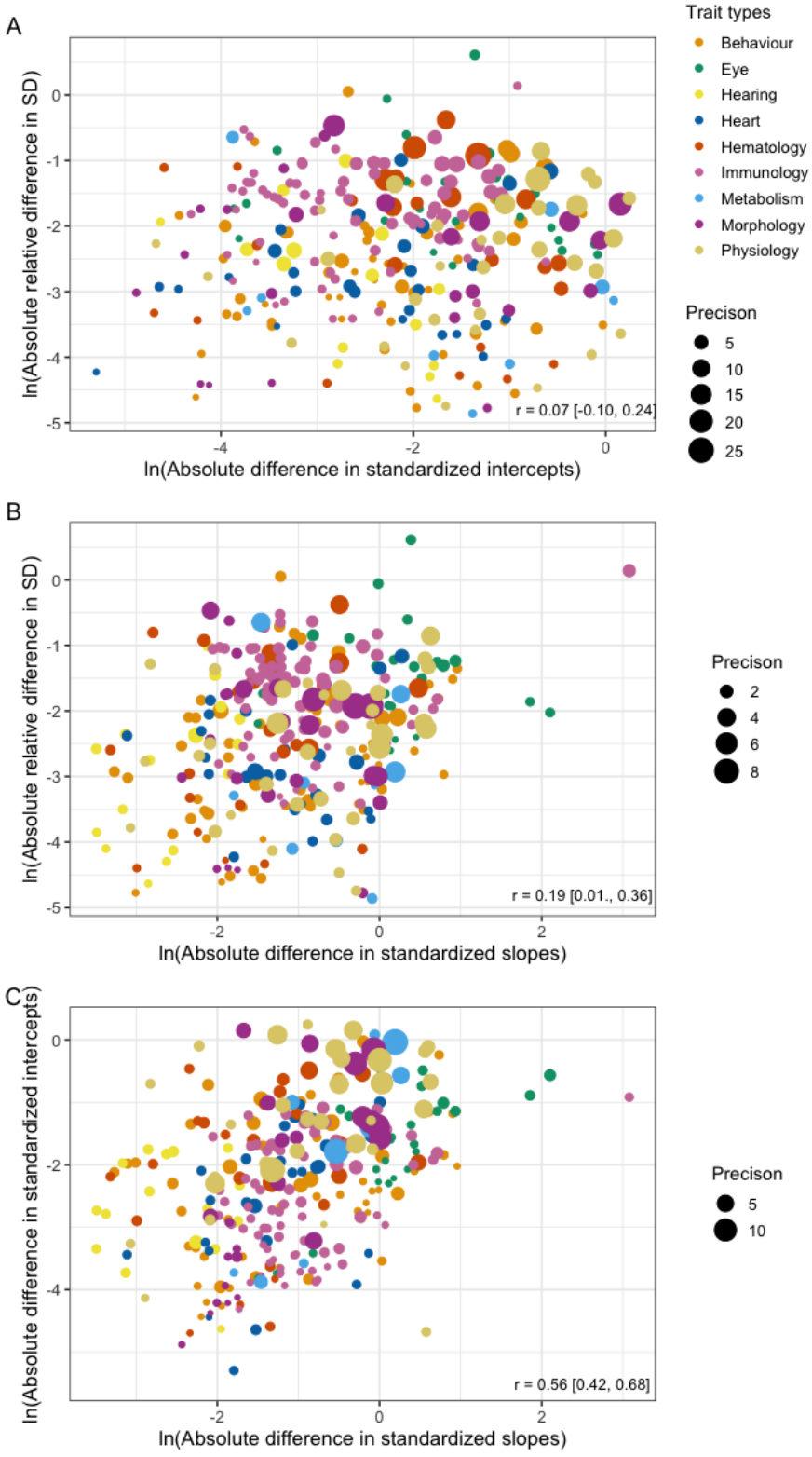

Figure 4. Bivariate ordinations of log absolute difference between males and females for intercept and residual SD (A), slope and residual SD (B), and slope and intercept (C), for biological traits collated into nine functional groups (i.e., Trait types, represented as different circle colours). Individual effect sizes (circles) are scaled by their precision, defined as: precision = 1 / Standard Error.

Relationship between slope/intercept and residual variance

Tri-variate meta-regressions and ordinations of the relationships between slope, intercept and residual variance (Fig. 4) revealed weak correlations between either slope or intercept and residual variance (r = 0.07 – 0.19, Fig. 4A – B), indicating that a greater magnitude of difference between the sexes in either slope or intercept parameter is not strongly associated with greater trait variance. In contrast, absolute differences between the sexes in slope and intercept are strongly correlated (r = 0.56, Fig. 4C), indicating that in cases where there are significant differences in trait values for males and females, should a difference in intercept be present, this is likely accompanied by a difference in allometric slope.

Discussion

Most current medical guidelines are not sex-specific, being informed by preclinical studies that have been conducted only on male animals (Zucker & Beery, 2010; Kim et al., 2010; Zucker et al., 2021) under the assumption that the results are equally applicable to females, or that the female phenotype represents a smaller body size version of the male phenotype (Buch et al., 2019; Campesi et al., 2021). Our study sought to provide comprehensive assessment of this assumption for a large dataset of phenotypic traits in mice. We did not recover strong evidence for the validity of this assumption in a preclinical (mouse) model: we find that females are not 'small' males or, more accurately, not 'scaled' males.

In an era where personalised medicine interventions are within reach and patient-specific solutions represent a realisable frontier in healthcare (e.g., Jackson & Chester, 2014; Javaid & Haleem, 2018; Heath & Pechlivanoglou, 2022), it is now well recognised that sex-based data are much needed to advance care in an equitable and effective manner. The historic neglect of sex as a study variable means that the natural history and trajectory of treatment response in women remains opaque for many chronic diseases. As studies that illuminate the presence and importance of sex differences continue to emerge, many experimental set-ups that use both sexes continue to eschew downstream testing for sex differences, in part due to perceived inflation of sample size required for such analyses (Dayton et al., 2016; Buch et al., 2019; Arnegard et al., 2020; Woitowich et al., 2020).

Explicit male-female comparisons are needed to clarify the nature of sex differences (Garcia-Sifuentes & Maney, 2021; Zucker et al., 2021). Here we address this issue through a novel meta-analytical focus on identifying and characterising allometric scaling relationships for biological traits on a broad scale. We identify slope parameter (b) differences between the sexes as being common (Fig. 2C, Fig. 3E) and where present, often associated with

significant differences in intercept value (Fig. 4C). We therefore demonstrate that the relationship between trait and body mass in mice differs fundamentally in mode (i.e., change in inter-trait covariance) between the sexes and that dimorphism cannot be fully explained by a magnitude shift in intercept value, as would be predicted should female phenotype represent a scaled version of male phenotype. For traits where there are significant differences in both slope and intercept between the sexes (Fig. 3C), it is common for a mixed scenario (male-biased significant slope and female-biased significant intercept, or female-biased significant slope and male-biased significant intercept; note that intercepts represent mean values for each sex) to occur. Therefore, for a given trait, a female value cannot be predicted based on an allometric coefficient extracted from regression data collected on males. Further, we find a male bias in residual SD for traits in morphology, immunology, hematology, hearing, and behaviour functional groups (7 out of 9 functional groups). However, we also find a weak correlation between difference in intercept and residual SD (Fig. 4A), meaning that allometric scaling differences alone do not explain increased residual SD in males compared to females. Or, put another way, among traits that show significant dimorphism in allometric relationships, males do not show greater variance than females just because they have greater body weights than females.

Our results complement recent evidence that supports a complex, trait-specific patterning of sex differences in markers routinely recorded in animal research (Rawlik et al., 2016; Karp et al., 2017; Zajitschek et al., 2020). Specifically, we build on previous studies using phenotypic traits from the International Mouse Phenotyping Consortium that have identified that sexual dimorphism is prevalent among phenotyping parameters (Karp et al., 2017), and moreover that, contrary to long-held assumption, neither females nor males show greater trait variability. We here show that the allometric relationship between trait value and body weight is dimorphic for most traits (75%), and these differences, where present, reflect trait-specific allometric patterns, involving both slope and intercept changes. As such, for slopes greater than zero, some trait values increase faster than body weight (positive allometry; $b > 1$) and some do not increase at the same rate as body weight (negative allometry; $b < 1$).

Sex-based scaling in biomedical studies

Our findings likely have implications for drug therapy, and specifically data surrounding the efficacy of drug dosing scaled by body weight. There exist known sex differences in drug prescription prevalence and usage patterns, as well as response to drug therapy (Watson et al., 2019; Malda et al., 2021). The same therapeutic regimen can elicit different responses due to sex-specific variance in pharmacokinetics and pharmacodynamics profiles (e.g., Yang et al., 2012; Zakiniaeiz et al., 2016), arising from underlying physiologic differences. These

include, for example, significantly dimorphic traits captured among the physiology group in our analysis, such as iron (Jiang et al., 2019) and body temperature (van Hoof, 2015), among the morphology group, such as lean mass and fat mass (Madla et al., 2021), and among the heart functional group, such as QT interval (time between Q wave and T wave) (Regitz-Zagrosek & Kararigas, 2017). Population studies have revealed that there is a higher prevalence of use for most therapeutic drugs in women as compared to men (Fernandez-Liz et al., 2008; Watson et al., 2019). Further, women are 50 – 75% more likely to experience Adverse Drug Reactions (ADRs) (Rademaker, 2001), although these are not fully explained (Koren et al., 2012). Women may be at increased risk of ADRs because they are prescribed more drugs than men, however women are usually prescribed drugs at the same dose as men, meaning that they receive a higher dose relative to body weight in most cases. Scaling of doses on a milligram/kilogram body weight basis has been recommended as a pathway to reducing ADRs (Zucker & Prendergast, 2020), particularly for drugs that exhibit a steep dose-response curve (Chen et al., 2020). Indeed, sex differences in ADRs have been argued to be the result of body weight rather than sex, per se (Richardson et al., 2015). For both assertions to be supported, we would expect to observe a scenario (here, scenario B) whereby most or all phenotypic traits exhibit a scaled relationship between males and females, as a function of body weight. Our results do not provide overwhelming support in favour of scenario B, but rather support a sex- and trait-specific relationship between weight and phenotypic traits. This aligns more closely with evidence that weight-corrected pharmacokinetics are not directly comparable in men and women (Fadiran & Zhang, 2015; Zucker & Prendergast, 2020), and that many sex differences in ADRs persist after body weight correction (Greenblatt et al., 2014, 2019). Nevertheless, the Food and Drug Administration (FDA) has recommended dosage changes for women (e.g., sleep drug zolpidem; Farkas et al., 2013) and weight adjusted dosing of some drugs, such as antifungal drugs and antihypertensive drugs, appear to ameliorate sex differences in pharmacokinetics (Guo et al., 2010; Jarugula et al., 2010). As such, we suggest that where there exists an association between sex and dose, dose-response curves are likely to be sex-specific and clarification of this relationship would be supported (e.g., using meta-analysis, Zhong et al., 2017) rather than using a scaled male-specific dose response curve for females. Since many drugs are withdrawn from the market due to risks of ADRs in women, meta-analytic approaches to illuminating sex-specific dose response curves represents a viable opportunity to reducing the number of ADRs and reaching an important target set by precision medicine (Polasek et al., 2018).

Implications for allometric evolution

The study of allometry has a long history in evolutionary biology, established as a foundational descriptor of morphological variation at ontogenetic, population and evolutionary levels (Cheverud, 1982; Klingenberg, 1998). Allometry may channel phenotypic variation in fixed directions, defining scaling relationships that persist across large evolutionary timescales. For example, craniofacial variation among mammals has been observed to be constrained by allometry, such that small mammals have shorter faces than do larger ones (Cardini & Polly, 2013; Cardini et al., 2015). Conversely, allometry may facilitate morphological diversification, acting as a line of 'evolutionary least resistance', allowing for new morphotypes to originate relatively rapidly among closely related species (Porto et al., 2009; Pélabon et al., 2014). These pathways (allometric constraint vs allometric facilitation) may be a start point for exploring how sex differences in disease phenotypes arise, data that have been cited as a potential unexploited resource relevant for the development of new therapies (Arnold, 2010). Studies of static allometry, as examined herein, have revealed low levels of intraspecific variation in allometric slope, which explains only a small proportion of variation in size (Voje et al., 2014), compared to variation in allometric intercept (Bonduriansky, 2007). Moreover, traits under sexual selection have also revealed low magnitudes of allometric slope change under artificial selection experiments (Egset et al., 2012) and in wild populations (Egset et al., 2011), whereas intercept changes appear clear and heritable. These differences have historically been thought to be due to underlying features of the developmental system acting as an internal constraint (Huxley 1932; Gould, 1966), whereas more recent interpretations suggest that external constraint (selection) more likely acts to maintain slope invariance at the static level (Pélabon et al., 2013), which is consistent with data showing that variation occurs instead at the ontogenetic level, i.e. growth rate and ontogenetic allometric slope are evolvable (e.g., Wilson & Sánchez-Villagra, 2010; Klingenberg, 2010; Wilson, 2013). Broadly consistent with other static allometric studies, we find that where differences in allometry are present, significant intercept shifts alone are more common than are significant slope shifts (Fig. 2A compared to 2B). We focus explicitly on sex differences and observe that many traits show a combination of intercept and slope changes, as well as differences in residual variance. Aside from the evolutionary implications – that allometric slope likely does not have a high evolvability, or capacity to evolve – many of the traits examined here may show a low level of sex difference in slope because the sexes are both experiencing the same selective pressure to maintain functional size relationships across different body sizes.

Our meta-analytic results build a narrative of complexity in sex-based trait interactions and promote a case-specific approach to preclinical research that seeks to inform drug discovery, development and dosage. That females are not 'small' or 'scaled' males in a

388 preclinical mouse model underscores the need to include female data from the earliest
389 experimental stages. Our results evidence the plasticity of allometry at a microevolutionary
390 scale, revealing a pathway for sex variation in phenotypic traits, which may influence study
391 outcomes in biomedicine.
392
393 **Methods**
394 ***Data compilation and filtering***
395 We conducted all data procedures, along with statistical analyses, in the R environment v.
396 4.1.1 (R Development Core Team, 2021). We compiled our data set from the International
397 Mouse Phenotyping Consortium (IMPC) (www.mousephenotype.org, IMPC data release
398 10.1 June 2019), accessed in October 2019. These represent traits recorded in a high-
399 throughput phenotyping setting whereby standard operating procedures (SOPs) are
400 implemented in a pipeline concept. The phenotypic traits represent biomarkers used for the
401 study of disease phenotypes (see Karp et al., 2017), collated into the following nine
402 functional groups: behaviour, eye, hearing, heart, hematology, immunology, metabolism,
403 morphology, and physiology, which are the IMPC's original categorization (also previously
404 used in Zajitschek et al., 2020). These groupings were assigned in relation to the description
405 of the procedure undertaken for data point collection and following the categorisation of
406 pipeline events at adult stage, detailed in the International Mouse Phenotyping Resource of
407 Standardised Screens (IMPReSS, https://www.mousephenotype.org/impress/index).
408
409 For the initial dataset, data points were collated for adult wildtype mice only, filtering to
410 include non-categorical phenotypic trait values for which covariate information on sex and
411 body weight were available. This initial dataset comprised of 2,866,345 data points for 419
412 traits. A series of data cleaning procedures were implemented to remove data points with
413 missing body weight, zero values for a phenotypic trait and duplicated specimen IDs. Data
414 filtering was conducted using the R package dplyr v.1.0.7 (Wickham et al., 2021). The
415 resulting data set comprised 2,104,497 data points for 297 phenotypic traits, all of which had
416 corresponding body weight data, enabling us to estimate an allometric relationship between
417 a trait of interest and body weight. For each phenotypic trait, we had the following variables
418 (covariates): phenotyping center name (location where experimental data were collected),
419 external sample ID (animal ID), metadata group (identifier for experimental conditions in
420 place during the experiment), sex (male / female), weight (body weight in grams), weight
421 days old (day on which weight was recorded), procedure name (description of the
422 experimental procedure as in IMPReSS), parameter name (description of the recorded
423 parameter as in IMPReSS), and data point (phenotypic trait measurement – response
424 variable).

425

426 ***Linear mixed-effects model for static allometry***

427 The static form of allometry, the covariation of a trait with size as measured across a
428 population of adults within a single species (Klingenberg, 1998), was quantified using a
429 linear mixed-effects model approach (Laird & Ware, 1982). Within this framework, the
430 relationship between phenotypic trait value and body weight, accounting for random effects
431 associated with assignment to a metadata group and batch (defined as the date when the
432 measurements are collected), was quantified for each of the 297 traits. Models were
433 constructed using the function *lme* in the R package nlme v. 3.1-153 (Pinheiro et al., 2021)
434 and applied to each phenotypic trait separately. We used the approach described by
435 Nakagawa et al. (2017) that uses within-group centring (wgc) of the continuous predictor
436 (i.e., weight); in this way, the intercepts (x = 0) for each sex represents the population mean
437 for that specific sex. Also, we calculated z-scores (z) from the response (y) so that all
438 regression coefficients are directly comparable across different traits. The applied model
439 was:

440

441 z(log (data point)) ~ wgc(log (weight)) * sex + (1 | batch) + (weight | metadata group)

442

443 The random factor 'batch' labelled a cohort of mice that went through a procedure on the
444 same day (see Karp et al., 2017), whereas 'metadata group' represented occasions when
445 procedural parameters were changed (e.g., different instruments, different observers and
446 different settings). These two random factors along with the 'weight' random slopes would
447 reduce Type I errors due to clustering (Schielzeth & Forstmeier, 2009). Also, to estimate
448 different residual variances between the two sexes, we modelled group-wise
449 heteroscedasticity structure, which was defined using the *lme* function's argument weights =
450 varIdent (form = ~1 | sex).

451

452 For each phenotypic trait, model parameters (regression coefficients and variance
453 components) were extracted, using R package broom.mixed v.0.2.7 (Bolker & Robinson
454 2021), for males and females (slope, intercept, standard error, SE of slope, SE of intercept
455 and residual variance) and corresponding p values for regression coefficients were extracted
456 to assess the significance of sex differences in slope and intercept. Because the *lme*
457 function did not provide statistical significance for differences in residual variances (standard
458 deviations, SDs), we used the method developed by Nakagawa et al (2015) or the logarithm
459 of variability ratio, which compares the difference in SDs between two groups to obtain p
460 values for residual SD differences (see also Senior et al., 2020).

461

We were aware that some of the 297 studied traits were strongly correlated (i.e., non-independent: e.g., traits from left and right eyes and immunological assays with hierarchically clustering and overlapping cell types). Therefore, we collapsed *p* values of these related traits into 181 *p* values, using the procedure (grouping related traits or trait grouping) performed by Zajitschek et al. (2020). We employed Fisher's method with the adjustment proposed by Li and Ji (2005) implemented in the R package, poolr (Cinar & Viechtbauer, 2021), which modelled the correlation between traits; we set this correlation to 0.8.

Static allometry hypotheses and Sex-bias in allometric parameters

Using parameters extracted from the above models, three scenarios were assessed (see Fig 1), describing the form of sex differences in the static allometric relationship between phenotypic trait value and body weight. For a given trait, these were: a) males and females have significantly different slopes but share a similar intercept (Fig. 1A, 1D), b) males and females have significantly different intercepts but share a similar slope (Fig. 1B, 1E), c) males and females have significantly different slopes and intercepts (Fig. 1C, 1F). In addition, we assessed how many traits were significantly different in residuals SDs between the sexes. For these classifications, we used both *p* values from 297 traits and 181 merged trait groups.

For scenarios A – C, which represent significant differences between male and female regression slope and / or intercept parameters and cases where sex differences in SDs were significant, data were collated into functional groupings (as listed above) to assess whether, and to what extent, sex bias in parameter values and variance was present across phenotypic trait values. That is, when males and females differed significantly, we counted which sex displayed the greater parameter value (intercept, slope) and, separately, we also tallied the sex with the higher magnitude of variance. Results were pooled for phenotypic traits within a functional group and visualised using R package ggplot 2 v. 3.3.5 (Wickham 2016) for scenarios A – C, resulting in one set of comparisons for parameter values, and one for variance (SD) values. We should highlight that we only used the data set with 297 traits because the directionality of some trait values became meaningless once traits were merged, although merged *p* values were meaningful as *p* values are not directional (e.g., spending time in light side or dark side).

Meta-analysis of differences in slopes, intercepts and residual SDs

We were aware that our classification approach using *p* values are akin to vote counting, which has limitations (Gurevitch et al., 2018). Therefore, we conduced formal meta-analyses

using the following effect sizes: 1) difference between intercepts (traits mean for males and females), 2) difference between slopes and 3) differences between residuals SDs. We used corresponding SE or, more precisely, the square of SE as sampling variance. We were not able to compare the directionality of effect sizes among traits (e.g., latencies and body sizes), however our main interest in this study was whether males and females were different in intercepts, slopes and residuals SDs irrespective of directionalities. Therefore, we conducted meta-analyses of magnitudes applying the transformation to the mean and sampling variance, which assumes to follow folded normal distribution (Morrissey, 2016, eq. 8), by using the formulas below:

$$\mathrm{ES}_{folded} = \mathrm{SE}\sqrt{\frac{2}{\pi}}\exp\left(\frac{\mathrm{ES}^2}{2\mathrm{SE}^2}\right) + \mathrm{ES}\left[1 - 2\Phi\left(-\frac{\mathrm{ES}}{\mathrm{SE}}\right)\right]$$

$$\mathrm{SE}^2_{folded} = \mathrm{ES}^2 + \mathrm{SE}^2 - \mathrm{ES}^2_{folded}$$

Where Φ is the standard normal cumulative distribution function and ES_{folded} and SE_{folded} are, respectively, transformed effect size (point estimate) and sampling variance, while ES and SE are corresponding point estimate and sampling variance before transformation. Morrissey (2016) has shown that meta-analytic means using such a folding transformation are hardly biased. Therefore, these transformed variables were directly meta-analysed using the *rma.mv* function in the R package, metafor (Viechtbauer, 2010). The intercept models (meta-analytic model) had three random factors: 1) functional group, 2) traits group and 3) effect size identifier (which is equivalent to residuals in a meta-analytic model; see Nakagawa & Santos, 2012), while in the meta-regression models, we fitted functional group as a moderator (see Fig 3). The model structures for all the three effect sizes were identical. We reported parameter estimates and 95% confidence intervals, CI and 95% prediction intervals, PI, which were visualised by the R package, orchaRd (Nakagawa et al., 2021). In a meta-analysis, 95% PI represents the degree of heterogeneity as well as a likely range of an effect size for a future study. We considered the estimate statistically significant when 95% CI did not span zero.

Correlations among differences in slopes, intercepts and residual SDs

We also quantified correlations among the three effect sizes, using a Bayesian tri-variate meta-analytic model, implemented in the R package, brms (Burkner, 2017). We fitted functional grouping as a fixed effect and trait groups as a random effect using the function, *brm*. Notably, we have log transformed ES_{folded} and also transformed SE_{folded} using the delta method (e.g., Nakagawa et al., 2017), accordingly, before fitting effect sizes to the model. We imposed the default priors for all the parameter estimated with the settings of two chains, 1,000 warm-ups and 4,000 iterations. We assessed the convergence of the chains by

Gelman-Rubin statistic (Gelman & Rubin, 1992), which was 1 for all chains (i.e., meaning they were all converged) and we also checked all effective sample sizes for posterior samples (all were over 800). We reported mean estimates (correlations among the three effect sizes) and 95% credible intervals (CI) and if the 95% CI did not overlap with 0, we considered the parameter statistically significantly different from 0.

Data availability

The R code and data generated during this study are freely accessible on GitHub at <to be inserted>. An R Markdown file with the complete workflow for all analyses is provided in the supporting information, available at <to be inserted>.

References

Arnegard, M. E., Whitten, L. A., Hunter, C., & Clayton, J. A. (2020). Sex as a biological variable: a 5-year progress report and call to action. Journal of Women's Health, 29(6), 858-864.

Arnold, A. P. (2010). Promoting the understanding of sex differences to enhance equity and excellence in biomedical science. *Biology of Sex Differences*, *1*(1), 1. https://doi.org/10.1186/2042-6410-1-1

Becker, J. B., Prendergast, B. J., & Liang, J. W. (2016). Female rats are not more variable than male rats: a meta-analysis of neuroscience studies. Biology of Sex Differences, 7(1), 34. https://doi.org/10.1186/s13293-016-0087-5

Beery, A. K., & Zucker, I. (2011). Sex bias in neuroscience and biomedical research. Neuroscience & Biobehavioral Reviews, 35(3), 565-572. https://doi.org/https://doi.org/10.1016/j.neubiorev.2010.07.002

Bolker, B., & Robinson, D. (2021). *broom.mixed: Tidying Methods for Mixed Models*. In (Version 0.2.7.) https://CRAN.R-project.org/package=broom.mixed

Bouman, A., Heineman, M. J., & Faas, M. M. (2005). Sex hormones and the immune response in humans. Human Reproduction Update, 11(4), 411-423.

Bonduriansky, R. (2007). Sexual selection and allometry: a critical reappraisal of the evidence and ideas. Evolution: International Journal of Organic Evolution, 61(4), 838-849.

Buch, T., Moos, K., Ferreira, F. M., Fröhlich, H., Gebhard, C., & Tresch, A. (2019). Benefits of a factorial design focusing on inclusion of female and male animals in one experiment. Journal of Molecular Medicine, 97(6), 871-877. https://doi.org/10.1007/s00109-019-01774-0

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. Journal of statistical software, 80, 1-28.

Campesi, I., Seghieri, G., & Franconi, F. (2021). Type 2 diabetic women are not small type 2 diabetic men: Sex-and-gender differences in antidiabetic drugs. Current Opinion in Pharmacology, 60, 40-45. https://doi.org/https://doi.org/10.1016/j.coph.2021.06.007

Canto, J. G., Goldberg, R. J., Hand, M. M., Bonow, R. O., Sopko, G., Pepine, C. J., & Long, T. (2007). Symptom presentation of women with acute coronary syndromes: myth vs reality. Archives of internal medicine, 167(22), 2405-2413.

Cardini, A., & Polly, P. D. (2013). Larger mammals have longer faces because of size-related constraints on skull form [Article]. Nature Communications, 4, 2458. https://doi.org/10.1038/ncomms3458

Cardini, A., Polly, D., Dawson, R., & Milne, N. (2015). Why the long face? Kangaroos and wallabies follow the same 'rule'of cranial evolutionary allometry (CREA) as placentals. *Evolutionary Biology*, *42*(2), 169-176.

578 Chen, M.-L., Lee, S.-C., Ng, M.-J., Schuirmann, D. J., Lesko, L. J., & Williams, R. L. (2000).
579 Pharmacokinetic analysis of bioequivalence trials: Implications for sex-related issues in
580 clinical pharmacology and biopharmaceutics. Clinical Pharmacology & Therapeutics, 68(5),
581 510-521. https://doi.org/https://doi.org/10.1067/mcp.2000.111184
582 Cheverud, J. M. (1982). Phenotypic, genetic, and environmental morphological integration in the
583 cranium. Evolution, 36(3), 499-516.
584 Cinar, O., & Viechtbauer, W. (2021). *poolr: Methods for Pooling P-Values from (Dependent) Tests.* In
585 (Version 1.0-0.) https://CRAN.R-project.org/package=poolr
586 Clayton, J. A., & Collins, F. S. (2014). Policy: NIH to balance sex in cell and animal studies. Nature
587 News, 509(7500), 282.
588 Cook, I. F. (2008). Sexual dimorphism of humoral immunity with human vaccines. Vaccine, 26(29-30),
589 3551-3555.
590 Correa-de-Araujo, R. (2006). Serious gaps: how the lack of sex/gender-based research impairs health.
591 Journal of Women's Health, 15(10), 1116-1122.
592 Dayton, A., Exner, E. C., Bukowy, J. D., Stodola, T. J., Kurth, T., Skelton, M., Greene, A. S., & Cowley Jr,
593 A. W. (2016). Breaking the cycle: estrous variation does not require increased sample size in
594 the study of female rats. Hypertension, 68(5), 1139-1144.
595 Dickinson, M. E., et al. (2016). High-throughput discovery of novel developmental phenotypes.
596 Nature, 537(7621), 508-514. https://doi.org/10.1038/nature19356
597 Egset, C. K., Bolstad, G. H., Rosenqvist, G., Endler, J. A., & Pélabon, C. (2011). Geographical variation
598 in allometry in the guppy (Poecilia reticulata). Journal of Evolutionary Biology, 24(12), 2631-
599 2638.
600 Egset, C. K., Hansen, T. F., Le Rouzic, A., Bolstad, G. H., Rosenqvist, G., & Pélabon, C. (2012). Artificial
601 selection on allometry: change in elevation but not slope. Journal of Evolutionary Biology,
602 25(5), 938-948. https://doi.org/https://doi.org/10.1111/j.1420-9101.2012.02487.x
603 Fadiran, E. O., & Zhang, L. (2015). Effects of sex differences in the pharmacokinetics of drugs and
604 their impact on the safety of medicines in women. In M. Harrison-Woolrych (Ed.), Medicines
605 for women (pp. 41-68). Springer International Publishing.
606 Farkas, R. H., Unger, E. F., & Temple, R. (2013). Zolpidem and Driving Impairment — Identifying
607 Persons at Risk. New England Journal of Medicine, 369(8), 689-691.
608 https://doi.org/10.1056/NEJMp1307972
609 Feldman, S., Ammar, W., Lo, K., Trepman, E., van Zuylen, M., & Etzioni, O. (2019). Quantifying Sex
610 Bias in Clinical Studies at Scale With Automated Data Extraction. JAMA Network Open, 2(7),
611 e196700-e196700. https://doi.org/10.1001/jamanetworkopen.2019.6700
612 Fernández-Liz, E., Modamio, P., Catalán, A., Lastra, C. F., Rodríguez, T., & Mariño, E. L. (2008).
613 Identifying how age and gender influence prescription drug use in a primary health care
614 environment in Catalonia, Spain. British Journal of Clinical Pharmacology, 65(3), 407-417.
615 Flanagan, K. L. (2014). Sexual dimorphism in biomedical research: a call to analyse by sex.
616 Transactions of The Royal Society of Tropical Medicine and Hygiene, 108(7), 385-387.
617 https://doi.org/10.1093/trstmh/tru079
618 Gandhi, M., Aweeka, F., Greenblatt, R. M., & Blaschke, T. F. (2004). Sex differences in
619 pharmacokinetics and pharmacodynamics. Annual Review of Pharmacology and Toxicology,
620 44, 499-523.
621 Garcia-Sifuentes, Y., & Maney, D. L. (2021). Reporting and misreporting of sex differences in the
622 biological sciences. eLife, 10, e70817. https://doi.org/10.7554/eLife.70817
623 Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences.
624 *Statistical science*, *7*(4), 457-472.
625 Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. Biological Reviews, 41(4), 587-
626 638.
627 Greenblatt, D. J., Harmatz, J. S., Singh, N. N., Steinberg, F., Roth, T., Moline, M. L., Harris, S. C., &
628 Kapil, R. P. (2014). Gender differences in pharmacokinetics and pharmacodynamics of

zolpidem following sublingual administration. The Journal of Clinical Pharmacology, 54(3), 282-290.

Greenblatt, D. J., Harmatz, J. S., & Roth, T. (2019). Zolpidem and gender: Are women really at risk? Journal of clinical psychopharmacology, 39(3), 189-199.

Guo, T., Sun, W. J., Xia, D. Y., & Zhao, L. S. (2010). The pharmacokinetics of fluconazole in healthy Chinese adult volunteers: influence of ethnicity and gender. Journal of Clinical Pharmacy and Therapeutics, 35(2), 231-237.

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. Nature, 555(7695), 175-182.

Heath, A., & Pechlivanoglou, P. Prioritizing Research in an Era of Personalized Medicine: The Potential Value of Unexplained Heterogeneity. Medical Decision Making, 0(0), 0272989X211072858. https://doi.org/10.1177/0272989x211072858

Huxley, J. S. (1924). Constant differential growth-ratios and their significance. Nature, 114, 895-896.

Huxley, J. S. (1932). Problems of Relative Growth. John Hopkins University Press.

Holdcroft, A. (2007). Gender bias in research: how does it affect evidence based medicine? In: SAGE Publications Sage UK: London, England.

Jackson, S. E., & Chester, J. D. (2015). Personalised cancer medicine. International Journal of Cancer, 137(2), 262-266. https://doi.org/https://doi.org/10.1002/ijc.28940

Jarugula, V., Yeh, C. M., Howard, D., Bush, C., Keefe, D. L., & Dole, W. P. (2010). Influence of body weight and gender on the pharmacokinetics, pharmacodynamics, and antihypertensive efficacy of aliskiren. The Journal of Clinical Pharmacology, 50(12), 1358-1366.

Javaid, M., & Haleem, A. (2018). Additive manufacturing applications in orthopaedics: A review. Journal of Clinical Orthopaedics and Trauma, 9(3), 202-206. https://doi.org/https://doi.org/10.1016/j.jcot.2018.04.008

Karp, N. A., Mason, J., Beaudet, A. L., Benjamini, Y., Bower, L., Braun, R. E., Brown, S. D., Chesler, E. J., Dickinson, M. E., & Flenniken, A. M. (2017). Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nature Communications*, *8*(1), 1-12.

Kim, A. M., Tingen, C. M., & Woodruff, T. K. (2010). Sex bias in trials and treatment must end. Nature, 465(7299), 688-689. https://doi.org/10.1038/465688a

Klein, S. L., & Poland, G. A. (2013). Personalized vaccinology: one size and dose might not fit both sexes. Vaccine, 31(23), 2599-2600.

Klein, S. L., Schiebinger, L., Stefanick, M. L., Cahill, L., Danska, J., de Vries, G. J., Kibbe, M. R., McCarthy, M. M., Mogil, J. S., Woodruff, T. K., & Zucker, I. (2015). Opinion: Sex inclusion in basic research drives discovery. Proceedings of the National Academy of Sciences, 112(17), 5257-5258. https://doi.org/10.1073/pnas.1502843112

Klinge, I. (2008). Gender perspectives in European research. Pharmacological Research, 58(3), 183-189. https://doi.org/https://doi.org/10.1016/j.phrs.2008.07.011

Klingenberg, C. P. (1998). Heterochrony and allometry: the analysis of evolutionary change in ontogeny. *Biological Reviews*, *73*(1), 79-123. https://doi.org/10.1017/s000632319800512x

Klingenberg, C. P. (2010). There's something afoot in the evolution of ontogenies. Bmc Evolutionary Biology, 10, Article 221. https://doi.org/10.1186/1471-2148-10-221

Koren, G., Nordeng, H., & MacLeod, S. (2013). Gender differences in drug bioequivalence: time to rethink practices. Clinical Pharmacology & Therapeutics, 93(3), 260-262.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. Biometrics, 963-974.

Li, J., & Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity, 95(3), 221-227.

Madla, C. M., Gavins, F. K. H., Merchant, H. A., Orlu, M., Murdan, S., & Basit, A. W. (2021). Let's talk about sex: Differences in drug therapy in males and females. Advanced Drug Delivery Reviews, 175, 113804. https://doi.org/https://doi.org/10.1016/j.addr.2021.05.014

Mauvais-Jarvis, F., Bairey Merz, N., Barnes, P. J., Brinton, R. D., Carrero, J.-J., DeMeo, D. L., De Vries, G. J., Epperson, C. N., Govindan, R., Klein, S. L., Lonardo, A., Maki, P. M., McCullough, L. D.,

Regitz-Zagrosek, V., Regensteiner, J. G., Rubin, J. B., Sandberg, K., & Suzuki, A. (2020). Sex and gender: modifiers of health, disease, and medicine. The Lancet, 396(10250), 565-582. https://doi.org/https://doi.org/10.1016/S0140-6736(20)31561-0

Mazure, C. M., & Jones, D. P. (2015). Twenty years and still counting: including women as participants and studying sex and gender in biomedical research. BMC Women's Health, 15(1), 94. https://doi.org/10.1186/s12905-015-0251-9

Mogil, J. S., & Chanda, M. L. (2005). The case for the inclusion of female subjects in basic science studies of pain. PAIN, 117(1). https://journals.lww.com/pain/Fulltext/2005/09000/The_case_for_the_inclusion_of_female_subjects_in.1.aspx

Morrissey, M. B. (2016). Meta-analysis of magnitudes, differences and variation in evolutionary parameters. *Journal of Evolutionary Biology*, *29*(10), 1882-1904.

Morrow, E. H. (2015). The evolution of sex differences in disease. Biology of Sex Differences, 6(1), 5. https://doi.org/10.1186/s13293-015-0023-0

Morrow, E. H., & Connallon, T. (2013). Implications of sex-specific selection for the genetic basis of disease. Evolutionary Applications, 6(8), 1208-1217.

Nakagawa, K., & Kajiwara, A. (2015). Female sex as a risk factor for adverse drug reactions. Nihon rinsho. Japanese journal of clinical medicine, 73(4), 581-585.

Nakagawa, S., & Santos, E. S. (2012). Methodological issues and advances in biological meta-analysis. Evolutionary Ecology, 26(5), 1253-1274.

Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M., & Senior, A. M. (2015). Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution*, *6*(2), 143-152.

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface*, *14*(134), 20170213. https://doi.org/doi:10.1098/rsif.2017.0213

Nakagawa, S., Kar, F., O'Dea, R. E., Pick, J. L., & Lagisz, M. (2017). Divide and conquer? Size adjustment with allometry and intermediate outcomes. *BMC biology*, *15*(1), 1-6.

Nakagawa, S., Lagisz, M., O'Dea, R. E., Rutkowska, J., Yang, Y., Noble, D. W. A., & Senior, A. M. (2021). The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond. *Research Synthesis Methods*, *12*(1), 4-12. https://doi.org/https://doi.org/10.1002/jrsm.1424

NIH, 1993. Revitalization Act of 1993, PL 103-43. Available at grants.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm001.htm

Oh, S. S., Galanter, J., Thakur, N., Pino-Yanes, M., Barcelo, N. E., White, M. J., de Bruin, D. M., Greenblatt, R. M., Bibbins-Domingo, K., Wu, A. H. B., Borrell, L. N., Gunter, C., Powe, N. R., & Burchard, E. G. (2015). Diversity in Clinical and Biomedical Research: A Promise Yet to Be Fulfilled. PLOS Medicine, 12(12), e1001918. https://doi.org/10.1371/journal.pmed.1001918

Pélabon, C., Bolstad, G. H., Egset, C. K., Cheverud, J. M., Pavlicev, M., & Rosenqvist, G. (2013). On the relationship between ontogenetic and static allometry. The American Naturalist, 181(2), 195-212.

Prendergast, B. J., Onishi, K. G., & Zucker, I. (2014). Female mice liberated for inclusion in neuroscience and biomedical research. Neuroscience & Biobehavioral Reviews, 40, 1-5. https://doi.org/https://doi.org/10.1016/j.neubiorev.2014.01.001

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2021). nlme: Linear and Nonlinear Mixed Effects Models. In (Version 3.1-153) https://CRAN.R-project.org/package=nlme

Polasek, T. M., Shakib, S., & Rostami-Hodjegan, A. (2018). Precision dosing in clinical medicine: present and future. Expert Review of Clinical Pharmacology, 11(8), 743-746. https://doi.org/10.1080/17512433.2018.1501271

729 Porto, A., de Oliveira, F. B., Shirai, L. T., De Conto, V., & Marroig, G. (2009). The Evolution of
730 Modularity in the Mammalian Skull I: Morphological Integration Patterns and Magnitudes.
731 Evolutionary Biology, 36(1), 118-135. https://doi.org/10.1007/s11692-008-9038-3
732 Rademaker, M. (2001). Do women have more adverse drug reactions? American journal of clinical
733 dermatology, 2(6), 349-351.
734 Rathore, S. S., Wang, Y., & Krumholz, H. M. (2002). Sex-based differences in the effect of digoxin for
735 the treatment of heart failure. New England Journal of Medicine, 347(18), 1403-1411.
736 Rawlik, K., Canela-Xandri, O., & Tenesa, A. (2016). Evidence for sex-specific genetic architectures
737 across a spectrum of human complex traits. Genome biology, 17(1), 1-8.
738 Regitz-Zagrosek, V., & Kararigas, G. (2017). Mechanistic pathways of sex differences in cardiovascular
739 disease. Physiological reviews, 97(1), 1-37.
740 Rogers, W. A., & Ballantyne, A. J. (2008). Exclusion of Women From Clinical Research: Myth or
741 Reality? Mayo Clinic Proceedings, 83(5), 536-542.
742 https://doi.org/https://doi.org/10.4065/83.5.536
743 Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in
744 mixed models. Behavioral ecology, 20(2), 416-420.
745 Senior, A. M., Viechtbauer, W., & Nakagawa, S. (2020). Revisiting and expanding the meta-analysis of
746 variation: The log coefficient of variation ratio. Research Synthesis Methods, 11(4), 553-567.
747 Sex, T., & Group, G. S. R. C. t. A. (2020). Sex and gender in health research: updating policy to reflect
748 evidence. Medical Journal of Australia, 212(2), 57-62.e51.
749 doi:https://doi.org/10.5694/mja2.50426
750 Shansky, R. M. (2019). Are hormones a female problem for animal research? Science, 364(6443),
751 825-826. https://doi.org/doi:10.1126/science.aaw7570
752 Soldin, O. P., & Mattison, D. R. (2009). Sex Differences in Pharmacokinetics and Pharmacodynamics.
753 Clinical Pharmacokinetics, 48(3), 143-157. https://doi.org/10.2165/00003088-200948030-
754 00001
755 Tannenbaum, C., Ellis, R. P., Eyssel, F., Zou, J., & Schiebinger, L. (2019). Sex and gender analysis
756 improves science and engineering. Nature, 575(7781), 137-146.
757 Team, R. C. (2019). R: A language and environment for statistical computing. In *R Foundation for*
758 *Statistical Computing*. http://www.R-project.org
759 Usui, T., Macleod, M. R., McCann, S. K., Senior, A. M., & Nakagawa, S. (2021). Meta-analysis of
760 variation suggests that embracing variability improves both replicability and generalizability
761 in preclinical research. PLoS Biology, 19(5), e3001009.
762 Van Hoof, J. (2015). Female thermal demand. Nature Climate Change, 5(12), 1029-1030.
763 Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of
764 statistical software, 36(3), 1-48.
765 Voje, K. L., Hansen, T. F., Egset, C. K., Bolstad, G. H., & Pélabon, C. (2014). Allometric constraints and
766 the evolution of allometry. Evolution, 68(3), 866-885. https://doi.org/10.1111/evo.12312
767 Wallach, J. D., Sullivan, P. G., Trepanowski, J. F., Steyerberg, E. W., & Ioannidis, J. P. (2016). Sex based
768 subgroup differences in randomized controlled trials: empirical evidence from Cochrane
769 meta-analyses. bmj, 355.
770 Whitley, H. P., & Lindsey, W. (2009). Sex-based differences in drug activity. American Family
771 Physician, 80(11), 1254-1258.
772 Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.
773 https://ggplot2.tidyverse.org
774 Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: A Grammar of Data Manipulation. In
775 (Version 1.0.7) https://CRAN.R-project.org/package=dplyr
776 Wiemann, L. M., Vallarta-Ast, N., Krueger, D., & Binkley, N. (2007). Effect of female database use for
777 T-score derivation in men. Journal of Clinical Densitometry, 10(3), 244-248.

Wilson, L. A. B., & Sánchez-Villagra, M. R. (2010). Diversity trends and their ontogenetic basis: an exploration of allometric disparity in rodents. Proc R Soc Lond B Biol Sci, 277. https://doi.org/10.1098/rspb.2009.1958

Wilson, L. A. (2013). Allometric disparity in rodent evolution. *Ecology and Evolution*, *3*(4), 971-984.

Woitowich, N. C., Beery, A., & Woodruff, T. (2020). A 10-year follow-up study of sex inclusion in the biological sciences. eLife, 9, e56344. https://doi.org/10.7554/eLife.56344

Yang, L., Li, Y., Hong, H., Chang, C.-W., Guo, L.-W., Lyn-Cook, B., Shi, L., & Ning, B. (2012). Sex differences in the expression of drug-metabolizing and transporter genes in human liver. Journal of drug metabolism & toxicology, 3(3).

Yu, Y., Chen, J., Li, D., Wang, L., Wang, W., & Liu, H. (2016). Systematic analysis of adverse event reports for sex differences in adverse drug events. Scientific Reports, 6(1), 1-9.

Zajitschek, S. R. K., Zajitschek, F., Bonduriansky, R., Brooks, R. C., Cornwell, W., Falster, D. S., Lagisz, M., Mason, J., Senior, A. M., Noble, D. W. A., & Nakagawa, S. (2020). Sexual dimorphism in trait variability and its eco-evolutionary and statistical implications. *eLife*, *9*, e63170. https://doi.org/10.7554/eLife.63170

Zucker, I., & Beery, A. K. (2010). Males still dominate animal studies. *Nature*, *465*(7299), 690-690. https://doi.org/10.1038/465690a

Zucker, I., & Prendergast, B. J. (2020). Sex differences in pharmacokinetics predict adverse drug reactions in women. *Biology of Sex Differences*, *11*(1), 32. https://doi.org/10.1186/s13293-020-00308-5

Zucker, I., Prendergast, B. J., & Beery, A. K. (2021). Pervasive Neglect of Sex Differences in Biomedical Research. Cold Spring Harbor Perspectives in Biology, a039156.

Acknowledgements

This research was supported by Australian Research Council grants DP200100361 awarded to SN and ML and FT200100822 awarded to LABW. Research reported in this publication was supported by the European Molecular Biology Laboratory core funding and the National Human Genome Research Institute of the National Institutes of Health under Award Number UM1HG006370. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

LABW and SN designed the research; SN, LABW, SRKZ, ML and HH contributed to the conception and implementation of data analysis; JM contributed to data acquisition; LABW drafted the manuscript with contributions from SN and ML.

Competing interests

The authors declare no competing interests.