

Sex and the Riddle of Variability

Sarah S. Richardson*

Scientific research associations, funders, and publishers have recently introduced sex inclusion mandates requiring the use of male and female specimens in preclinical research designs and the analysis and reporting of data disaggregated by sex. However, it is not a simple matter to incorporate males and females in the same study design while following best practices for rigorous inference in laboratory science using model organisms. This paper analyzes a current scientific debate over differences in variability between male and female laboratory rodents, and specifically over whether potential sources of sex-specific variability such as the estrous cycle constitute components of sex that should be measured. The variability debate surfaces the trade-offs between constructs of sex difference and similarity that face scientific researchers attempting to meet mandates to include both males and females in research design and report sex-specific results. This “riddle of variability” illuminates how laboratory researchers using model organisms make contextual choices (Richardson 2022) at multiple decision points in order to operationalize sex as a biological variable in a particular research design. These judgments are informed by research context and carry consequences for the validity, precision, and generalizability of claims of sex differences derived from preclinical research models.


Keywords

SABV • sex • sex contextualism • sex differences • sex inclusion • model organisms • preclinical research • variability

1. Introduction

Sex inclusion or “sex as a biological variable” (SABV) mandates have recently been instituted by many funders, journals, and pharmaceutical companies, responding to calls for biomedical research to address a history of androcentrism in preclinical science and accelerate the study of health disparities between women and men (White 2021; Woitowich 2020; Haverfield 2021). These policies, developed with the aim of advancing women’s health and increasing rigor and replicability in preclinical science (Clayton and Collins 2014; Klein 2015; Heidari 2016), ask

*Department of the History of Science, Committee on Degrees in Studies of Women, Gender, and Sexuality, Harvard University, Cambridge, MA, USA, srichard@fas.harvard.edu

 <https://orcid.org/0000-0002-9462-3051>

Received 20 January 2024; Revised 05 June 2024; Accepted 30 September 2024
doi:10.3998/ptpbio.5492



preclinical model organism researchers to include “both” sexes in research designs and to report and analyze disaggregated, sex-specific results for males and females.

The US National Institutes of Health (NIH) Office of Research on Women’s Health defines sex in the following way: “Sex is a multidimensional biological construct based on anatomy, physiology, genetics, and hormones. (These components are sometimes referred to together as ‘sex traits.’) All animals (including humans) have a sex. As is common across health research communities, NIH usually categorizes sex as male or female, although variations do occur” (“What are Sex & Gender?” 2024, np). The NIH style guide advises researchers that, “Sex is a biological descriptor based on reproductive, hormonal, anatomical, and genetic characteristics. Typical sex categories include male, female, and intersex. *Sex* is used when describing anatomical, gonadal, chromosomal, hormonal, cellular, and basic biological phenomena” (“NIH Style Guide” 2024, np).

Sex inclusion research policies like that of the NIH seek to render features of male and female biology retrievable across experimental designs and research programs by including and comparing male and female organisms. This operationalization of sex conceptualizes maleness and femaleness as essential kinds about which strong generalizations can be made through the comparison of males and females across contexts, taxa, and levels of biological analysis. Critics (e.g., Richardson et al. 2015; Ritz 2016; Garcia-Sifuentes and Maney 2021; Maney and Rich-Edwards 2023; Massa, Aghi, and Hill 2023; Pape et al. 2024) warn that mandating this binary sex essentialist approach to the study of sex-related variation across research programs ignores the pragmatics of scientific practice, disregards the complexity, context-specificity, and contingency of sex-related variables, and will lead to a proliferation of low quality, uninterpretable, and potentially politically and socially harmful claims about male and female differences.

This essay extends these critiques through an analysis of a current scientific debate over whether researchers need to test for estrous effects when designing, analyzing, and interpreting sex-split studies using rodent models. Analysis of this debate reveals what I term a “riddle of variability” for sex difference research—to search for differences produced by sex, laboratory animal model researchers must partition some forms of variation as non-sex-related and assert strong generalizing claims about homogeneity between males and females as research subjects. These judgments about what counts as sex systematically background many sources of sex-associated variation, resulting in a context-specific construct of sex.

The debate over whether it is necessary to account for estrous-related variation in sex difference research designs is a valuable site for philosophical analysis of scientific operationalizations of sex. Specifically, it supports the contention that sex is a pragmatic, contextual construct (Richardson 2022) in lab-based experimental science and demonstrates the value of a sex contextualist framework for guiding appropriate interpretations of the findings of biological research on sex-related variation.

2. The Debate

2.1. *Sex, Model Organisms, and the Problem of Variability*

Proposals for the routine inclusion of male and female model organisms in sex-split analyses pose many challenges for laboratory model organism research. Laboratory models are not generally designed for modeling variability across populations and among subgroups; rather, they are intended for experimentally isolating particular causal relations against a backdrop of minimal or controllable variability. Rodent models, in particular, are famously prized for their homogeneity, lack of variation, stability, and reliability. After decades of rodent research, there

now exists a wide variety of standardized procedures and traits for the study of behaviors and biological health-related outcomes in rodents. Determining how to rigorously model sex across the extensive range of traits and outcomes, mechanisms, disease models or endophenotypes, and experimental designs in preclinical laboratory research is especially challenging given rising concerns over replicability and the relevance of findings in rodent studies to human health (Kafkafi et al. 2018; Baker 2016; Willyard 2018).

In biomedical research, knowledge about variability is vital when comparing populations or subgroups. In the case of sex, estimates of sex-specific variability are needed to construct experiments that are adequately powered for each sex. If males and females show differences in variability and this is not accounted for in research design and statistical analysis, this violates assumptions of homoscedasticity, or homogeneity in variation. This can lead to a higher Type 1 (false positive) error rate, e.g., overestimating the magnitude of sex differences or asserting sex differences when there are actually none (Zajitschek et al. 2020).

For each trait, variability is assigned by calculating the coefficient of variation (CV). The value of the CV is the ratio of the standard deviation to the mean. It shows the extent of variability in relation to the mean of the population. If the CV value is large, this means that trait values are greatly dispersed in relation to the mean. Since the CV is independent of the unit in which the measurement was taken, it permits comparisons of variability between data with different units or widely different means.

There can be quantitative differences between comparand subgroups, including the sexes, for the trait of variability. For example, when comparing male and female subgroups, male and female means for a trait could be similar but show different variability. Alternatively, means could differ for males and females, but show similar variability. Males and females could have a similar distribution of CVs across many traits but have potentially large differences in variability for particular traits. Males and females could also have a different distribution of CVs across many traits, but this could be explained by a small number of individual traits for which there are large differences in variability. Additionally, variability itself can be variable across timescales (days, estrous cycles) and across contexts (how the rodents are housed or reared, for example; Smarr and Kriegsfeld 2022).

Practically speaking, preclinical laboratory research uses small samples that are usually inadequate to detect differences in variability among subgroups. This is illustrated by a retrospective power analysis by Dayton et al. (2016) of 74 findings of differently-variable phenotypes between male and female rats. The study found that only 16 of these 74 differences in variability would have been detected “when using an experimental cohort large enough to detect a difference in magnitude” (Dayton et al. 2016, 1139). Thus, adequate powering to detect a difference in means between sexes—a principal way that researchers attempt to satisfy the requirements of SABV policies and understand the goals of these policies—is distinct from, and often insufficient for, adequate powering to detect differences in variability between the two groups.

2.2. *Assessing Sex Differences in “Intrinsic Variability”*

Evidence as to whether male or female animals exhibit higher variability is mixed, a state of affairs characterized by one researcher as “a genuine disagreement in the literature” (Pecheva 2023). The debate extends a long history of scientific and cultural contestation around human male and female sex differences in variability (Shields 1982; Richardson 2012) into a new context in which, today, the study of sex differences is considered a central pillar in the agenda to advance women’s health equity. In an often explicitly feminist rebuke to longstanding ideas associating the menstrual cycle with female unreliability, changeability, moodiness, and decep-

tion (Ehrenreich and English 2005; Lunbeck 1994; Young 1990; Fausto-Sterling 1985; Grosz 1994), new voices are highlighting evidence of equivalent variability in males and females and arguing that researchers' continued fixation on the complications of measuring estrous cycle-related variation—without equivalent interest in diurnal cycles and dominance hierarchy variation in testosterone in males—is nothing more than a relic of old sexist biases (Shansky 2019).

Consider a recent study by Levy et al. (2023), published in the general-interest journal *Current Biology*. Levy et al. employed a machine learning tool to observe 16 male and 16 female mice as each individual ran around in a 5-gallon bucket on 20 consecutive days. The tool classified each mouse's motions as a "rear," "dive," or "run," tabulating the frequency for each individual. The results showed high variability among individual mice as well as within each sex category. Males as a group were more variable than females. Furthermore, in females, the phase of each individual mouse's estrous cycle did not correlate with its variability in exploratory behavior. The authors interpreted these findings as showing that estrous cycle is unnecessary to measure in sex difference research in rodents, concluding that, if anything, "females—rather than males—should be the default sex used" (Levy et al. 2023, 1361). "Guess Which Sex Behaves More Erratically (at Least in Mice)?" read the resulting *New York Times* headline: "A new study finds male mice more unpredictable than females, challenging century-old assumptions used to exclude females from research because of their hormones" (Ghorayshi 2023). In the media, lead author Dana Levy positioned the study's findings as advancing rodent researchers' ability to conduct sex comparisons to address issues such as "bias in drugs being produced" and "conditions such as anxiety, depression, and pain ... known to manifest differently in female mice and women" ("Harvard Scientists Make the Case for Female Mice in Neuroscience Research" 2023, np). Janine Clayton, director of the NIH Office of Research on Women's Health, which led the charge for the SABV policy, described the study as endorsing the NIH mandate's premise that "it is in the best interest of the public that sex-specific results are published" (Ghorayshi 2023).

The Levy et al. study is part of a stream of pointed scientific interventions making similar claims in the explicit context of advocating for SABV policies that require modeling sex as a biological variable in laboratory organisms. A prominent example is a 2014 meta-analysis in *Neuroscience and Biobehavioral Reviews* that reported that, "for the average trait of interest to animal researchers, variability does not differ between males and unstaged [measured without regard to estrous phase] females" (Prendergast, Onishi, and Zucker 2014, 3), concluding that, therefore, female mice were henceforth "liberated for inclusion in neuroscience and biomedical research" (1). Today, this study is widely cited—over 740 times according to Google Scholar—as definitively disproving the belief that female rodents are "intrinsically more variable than males" (Prendergast, Onishi, and Zucker 2014, 1).

Studies positing the similarity of male and female intrinsic variability share several assumptions: first, that variability can be represented as an intrinsic propensity of males as a group and females as a group; second, that this property is yielded by comparing overall variability within sex subgroups of a species/strain; and, third, that these generalized claims about variability for a particular sex are appropriate to ground choices in research design. That is, variability is here constructed as an intrinsic propensity of each sex-group, generalizable across many traits. For example, the Dayton et al. (2016) study referenced above found that slightly more than half of traits (74 out of 142) in their sample showed a sex difference in variance. But because there was "no *overall* difference in coefficient of variation between male and female rats" (Dayton et al. 2016, 1139, emphasis added) when averaged across all traits, the study concluded that the results support the interpretation that there is no need to consider sex-specific variability in study design. Similarly, an analysis of sex differences in variability in gene expression in mouse

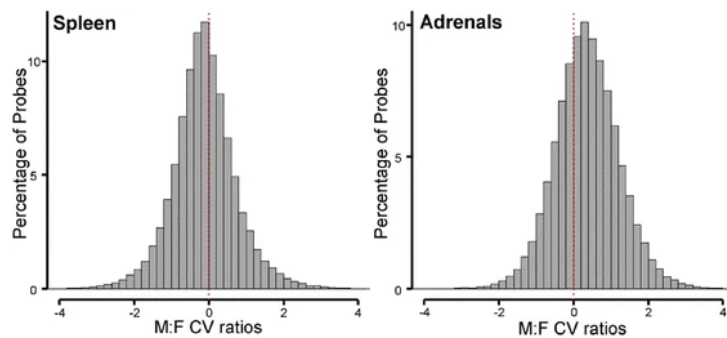


Figure 1: “Tissue-specific bias in the variability of gene expression in a restricted set of microarray studies. In mouse spleen, more probes show higher CV in females than in males. In mouse adrenals, more probes showed higher CV in males than females” (Itoh and Arnold 2015, Figure 2). Reproduced under Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

and human datasets by Itoh and Arnold (2015) found an “overall sexual balance of CV ratios” (2). However, in this study, female mice had higher variation in gene expression in some systems, such as the spleen, whereas males had higher variability in others, such as the adrenals (Figure 1). Nonetheless, Itoh and Arnold conclude that such “tissue-specific sex differences ... cancel each other out when considering *all tissues combined*” (2, emphasis added); hence, they find “no support for the hypothesis that female mice or humans are generally more variable because of their estrous or menstrual cycles or other variables” (4). But for researchers—perhaps the majority—who are not interested in an “average trait of interest” (Prendergast, Onishi, and Zucker 2014, 1) but in a particular trait of focus, claims of overall similarity or difference in variability across many traits and endpoints are of unclear utility for research design.

2.3. Sex-Specific Variability for Particular Traits and Measures

Indeed, ample evidence supports the supposition that there are extensive sex differences in trait-specific variability. A 2020 meta-analysis by Zajitschek et al. of 218 traits observed in 26,900 mice found that across all comparisons, no generalization could be made about which sex was more variable. While “sex biases in variability occur in many mouse traits,” the authors conclude that “the directions of those biases differ between traits” (Zajitschek et al. 2020, 11). To help researchers determine the sample sizes for males and females needed to achieve statistical power given known differences in variability derived from previously published literature, the research team created an app that allows scientists to assess the average difference in variance between females and males for particular mouse strains, traits, procedures, or major trait groups (see https://szaj.shinyapps.io/SexDifference_Shiny/).

More specific to the question of estrous-related variability is a 2022 study by Rocks et al. that directly compared rodent data analyzed by estrous stage to the same data simply presented as a comparison of male and female subgroups. The reanalysis demonstrated that consideration of estrous phase leads to more precision in identification of sex-related biological mechanisms; moreover, this specificity is not predicted by measures of overall differences in variability between males and females. For example, in the case of mouse dendritic spine density, when comparing just two subgroups—males and females—a large sex difference appears, with females showing higher variability than males and only a 38% overlap in male and female distributions (Figure 2a). But when three subgroups are analyzed—males, diestrus rodents (the stage after estrus, when progesterone levels rise), and proestrus rodents (the stage immediately prior to estrus, with characteristic hormonal changes)—males and diestrus rodents are nearly identical, showing 90%

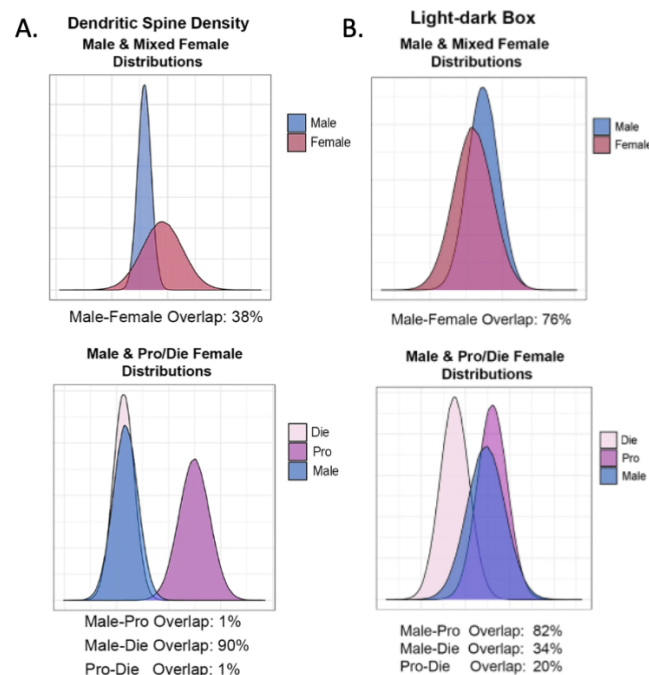


Figure 2: Density distribution plots depict the normal distributions of the (A) dendritic spine density and (B) light-dark box data and for merged female groups compared to males (top, left and right respectively) and males compared to females separated by the estrous cycle stage (bottom, left and right respectively). The degree of overlap between the distributions is given below each plot (Rocks, Cham, and Kundakovic 2022). Reproduced under Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

overlap, while proestrus rodents show only 1% overlap with both males and diestrus rodents. Conversely, in light-dark box experiments used to study anxiety-like behavior in rodents, only a small male-female difference is observed when comparing just males and females. But when three subgroups are analyzed, large variation among females becomes apparent, with only 20% overlap between proestrus and diestrus rodents (Figure 2b). As Rocks et al. conclude, “having the information about the estrous cycle explains where the sex-based variability is coming from and allows for a mechanistic insight” (7). They argue that this “increased resolution ... is because the hormonal status is a sex-specific factor that is more precise than sex” (11). Contrary to the premise of Levy et al. and others that similar levels of variability between the sexes supports the assumption that estrous cycle can be excluded from research analyzing sex-related variation, Rocks et al. (2022) find that “it is possible to not find any difference in the variability between males and females, and still find an effect of the estrous cycle” (12). In other words, “female variability does not equal (and is not predictive of) the estrous cycle effect” (1).

2.4. Pragmatic Considerations

Such findings would seem to affirm longstanding practice. The 4–5-day estrous cycle can produce variations in commonly studied outcomes in rodent laboratory studies. For this reason, standard protocol has historically held that estrous cycle must be assayed when using female rodents. As one researcher summarized this common view in a *Science* article discussing SABV mandates prior to their introduction, “data are uninterpretable” in studies that include both male and female rodents unless researchers take daily vaginal swabs (Wald and Wu 2010).

And yet, there are clear reasons why researchers would rather not have to deal with estrous

hormone-related variation in sex difference studies. Accounting for estrous cycle in a rodent study increases the time and resources needed to conduct studies and introduces new possibilities for error and uncertainty in the findings. As the researcher continued in the article cited above, “scientists may also need to keep as many as four times the number of female animals as male animals to make sure their subjects are cycling in sync. And even with those precautions, the cycle may still lead to less-clear results that are more difficult to publish” (Wald and Wu 2010).

The pragmatic challenges of rigorously measuring estrous phase are important to appreciate. As an illustration, consider a study by Y. Zhao et al. (2018) in *Behavioral Brain Research* that sought to rigorously assess the effect of estrous cycle on the fear potentiated startle response in rats. The study design called for comparing males temporally matched to females at each of four stages of the rat estrous cycle, though in this study, the researchers reliably detected only three stages. To allow estrous cycle-related factors to be adjusted for as covariates in statistical analysis or to be studied directly in relation to a trait of interest, researchers conduct daily vaginal smears of female rodents to determine estrous stage. Because estrous cycle cytology is notoriously difficult to interpret, this study employed two human observers to examine each result to confirm the assessed stage. In this study, the researchers “were unable to histologically differentiate between proestrus and estrus stage” (31). Thus, researchers chose to use three comparands: rodents in the proestrus/estrus stage (comprising one group), rodents in the diestrus stage, and males. As the study went on, however, the rats stopped cycling—with 62% becoming acyclic by the study’s conclusion. Such unknowns, uncertainties, ambiguities, and irregularities are commonplace in attempts to analyze the effect of estrous cycle on studied outcomes, a challenge that multiple research groups are currently attempting to address with AI-informed tools (e.g., Wolcott et al. 2022).

3. What, Then, Is Sex? A Sex Contextualist Analysis

Researchers’ negotiations over whether to test for estrous cycle effects when designing, analyzing, and interpreting sex-split studies reveal how sex is plurally, partially, and pragmatically operationalized in laboratory model organism research design. To meet mandates for sex-inclusive rodent research design, the researcher is faced with the question of whether to define estrous cycle as a component of the variable of sex (which therefore requires measurement in order to account for sex as a biological variable) or as a trait contributing to interindividual variability that is not, in the words of researchers, “sex itself” or “intrinsic sex.”

Levy et al., Prendergast et al., and others (e.g., Mogil and Chanda 2005; Becker, Prendergast, and Liang 2016; Dayton et al. 2016) argue that compliance with sex inclusion mandates can be accomplished *without* requiring research protocols to control for potential sex differences in variability. That is, they conclude from reports of overall similarity in variability in males and females that it is unnecessary to track vaginal cytology at each of the four stages of the rodent estrous cycle when including females in biomedical research. Accepting this premise of male-female homogeneity of “intrinsic variability” (Prendergast, Onishi, and Zucker 2014), however, requires a scaffold of assumptions about sex differences, similarities, and variability, and how to measure them, that involve the contextual judgments of researchers. Specifically, these researchers assert that since reviews show that for many traits, males are just as or more variable than females, no special controls are required for the study of female rodents due to the estrous cycle, nor, for that matter for other potential sources of sex-specific variability, whether male or female (Shansky 2019), when conducting male female-comparisons. For these scientists, studies showing similar within-group variability across males and females support the ease and practicality, without the need for more organisms, time, and resources, of adhering to mandates

to include both males and females in biomedical research and report and analyze sex disaggregated results.

The constitution of sex subgroups (male and female) as representing “sex as a biological variable” and as yielding knowledge about sex, even without measuring sex-related biological variation such as gonadal hormone state, is a stark and striking departure from early arguments for sex inclusion mandates. Evidence that estrous cycle-related hormonal states produce female-specific biological variability, advocates argued at the time, showed the need for thoroughgoing study of females as well as males in all research studies (e.g., Beery and Zucker 2011). A touchstone article arguing for the NIH SABV mandate by neuroscientist and policy architect Larry Cahill, “Why Sex Matters for Neuroscience,” for instance, prominently featured evidence that estrous cycle state produces behavioral effects of large magnitude as an example of “copious sex influences on brain anatomy, chemistry, and function” (Cahill 2006, 477). Similarly, Jill Becker, another SABV policy architect and current editor of the journal *Biology of Sex Differences*, has long argued for the importance of measuring ovarian steroids in study designs that include females (e.g., Becker et al. 2005). Today, however, her position has evolved. Explicitly recognizing that how one conceptualizes the variable of sex may depend on pragmatics and the research question being asked, she now advises that, “for those investigators initiating research on female rats, power calculations based on data from males would likely be sufficient to determine the number of female subjects needed in order to see a sex difference... [I]nclusion of intact females, without regard to estrous cycle, and intact males is a valid approach to learn about females in neuroscience research” (Becker, Prendergast, and Liang 2016, 7).

The operationalization of “sex as a biological variable” in the diverse literature produced in the estrous cycle variability debate yields a concrete example of how sex, in preclinical laboratory practice, is a pragmatic, plural, and contextual construct. A sex contextualist analysis helps answer the question of why researchers deem some forms of sex-specific variation to be non-sex-related. Specifically, a contextualist framework makes clear how pragmatic constraints in the laboratory and in research policy administration inform a particular operationalization of male and female sex in the laboratory. Researchers as well as policy makers supporting sex inclusion research mandates are interested in what is required to allow researchers to easily incorporate the requirement into practice and what is sufficient to merely detect a difference between male and female subgroups. The presence of a policy mandate makes these pragmatic, contextual choices uniquely salient in this case. Fundamentally, advocates of sex inclusion policies in preclinical laboratory science are committed to the importance of the sex subgroup as a category of analysis—or, as Becker et al. put it in the passage quoted above, to research designs adequate “to see a sex difference” (2016, 7). The outcome of the process is the production of simple comparative research designs that represent sex as a discrete, binary biological variable sufficiently measured by comparing means between male and female subgroups.

If we understand operationalizations of sex in preclinical research as always involving contextual judgments (Richardson 2022), both choices—to exclude or include estrous cycle measures in sex-difference research design—are examples of sex contextualism in practice. Conceptualizing the estrous cycle as a sex-related biological variable that must be measured as a part of any sex comparison is a pragmatic contextual judgment that centers gonadal variation as a particularly salient hypothesized sex-related mechanism. In contrast, comparing male and female subgroups without measuring such estrous cycle-related variability is a contextual judgment that favors utilizing sex category as a proxy for sex. Sex contextualism helps us to see that these two constructs of sex are not the same and that plural approaches to operationalizing sex as a biological variable co-exist, even within a single model system and scientific discipline. In this case study, we see this in the radically divergent constructs of sex arrived at by Levy et al., Prendergast et al., and

others, as compared to Zajitschek et al., Rocks et al., and others. Recognizing this pluralism is vital to evaluating inferences, generalizations, and mechanistic hypotheses resulting from any claims of sex difference that emerge from preclinical research.

A sex contextualist approach does not tell us whether one way of constructing sex as a biological variable is more objectively true, sound, or empirically adequate than another, but it does make visible the contextual dimensions of the operationalization of sex in laboratory models, opening them to scrutiny. For example, the Levy et al. study found that 85% of the time, individual mouse identity among females predicted their behavioral repertoire better than any variable including sex; furthermore, Levy et al. recorded high intra- and inter-individual spontaneous exploratory behavior for both males and females (both were slightly higher for males). As coauthor Robert Datta explained to a reporter: “If you give me any random video from our pile, I can tell you which mouse it is. That’s how individualized the pattern of behavior is” (Pecheva 2023, np). The news piece continues: “we find that each female mouse exhibits a characteristic pattern of exploration that uniquely identifies it as an individual,” and the same is true for males; these results are characterized as “surprising” (Pecheva 2023, np). Alternative approaches to the operationalization of sex-related variation that use other approaches than sex category, and offer a different vision of what sex-inclusive science that advances gender justice might look like (Smiley et al. 2024; Massa, Aghi, and Hill 2023), might allow a different interpretation of these data. In contrast to Levy et al.’s strong conclusions about the need for more sex comparisons, perhaps the highly individualized behavioral repertoires identified in this study present a challenge to binary sex as an important and meaningful category to retain for analysis in future work on mouse exploratory behavior.

4. Sex and the Riddle of Variability

The debate over measuring variability in preclinical sex difference science shows that operationalizing “sex as a biological variable” in rodent research requires scientists to make contextual judgements at multiple decision points, each of which becomes a new source of potential incoherence. What results from laboratory-based experimental designs that include male and female model organisms is a riddle—not a simple finding of a difference, but a difference that rests on establishing or assuming other axes of similarity. That is, on the one hand, it is argued that scientists must study female and male biology in preclinical laboratory experimental research in order to capture differences that could be relevant to advancing the health of human women. But operationalized as a binary sex comparison within a single experimental design, studying sex as a biological variable requires, at the same time, the assertion that females are similar enough to males to be included in the same study design.

The paradoxical construction of the estrous cycle as both intrinsic to femaleness and extrinsic to sex itself is present throughout the sex difference literature. This convention of distinguishing sex itself from estrous state is iterated in the title of the Y. Zhao et al. 2018 study previously discussed: “No effect of sex and estrous cycle on the fear potentiated startle response in rats.” Levy et al. (2023) assert that estrous cycle is “a defining feature of the female internal state” (1358), but then proceed to argue that accounting for estrous state is unnecessary for the consideration of sex differences. Similarly, Mogil and Chanda’s (2005) review of nociceptive sensitivity on mouse tail withdrawal tests found that “sex itself accounted for a rather modest percentage of the overall variance” (2), where “sex itself” refers solely to whether the individual was a female or male and explicitly does not include estrous cycle hormonal profile. As seen above, even Rocks et al. slip into this riddle, describing estrous state as “sex-specific factor that is more precise than sex” (11).

This riddle of variability—of which kinds of biological variability constitute sex and should be the focus of experimental laboratory model organism research that includes and compares male and female subgroups, in the face of multiple dimensions of difference, similarity, and variability—reveals that sex does not simply “fall out” of comparisons of males and females as a coherent biological variable that causes differences between individuals or subgroups. It is operationalized through a series of assumptions about what “intrinsic sex” is, materially and statistically realized in a study design and informed by research contexts, including policy, in a sea of uncertainty produced by a dynamic background of many sources of variation, some of which mediate or interact with the thing that is called sex. The issue here for the researcher interested in sex differences is how to operationalize the variable of sex distinct from entanglements and contingencies across varied contexts, in order to get at the “true” effects of sex. Writing in the journal *Biology of Sex Differences*, Smarr and Kriegsfeld (2022) use the metaphors of “white noise in a signal” (2) and “Brownian motion offering precise location measurements for particles” (2) to describe this challenge of locating sex amidst the dynamic complexity and uncertainty of many sources of variability.

The riddle of variability is likely to be persistent, requiring resolution each time a sex-split study design (whether binary or nonbinary) is contemplated. Notably, the estrous cycle debate is just one example of the riddle of variability in sex difference science in laboratory model organisms. Laboratory housing is a large mediator of variability for rodents, representing an interaction between an entirely artifactual laboratory variable and the social-behavioral repertoire of these species. For example, Prendergast et al.’s review of articles that reported housing conditions found that single-housed mice exhibited 37% lower variability when compared to group-housed (Prendergast, Onishi, and Zucker 2014, 1). Similar to estrous cycle, there may be sex-specific dimensions to the structure of variability generated by different housing conditions. It is well known, for instance, that male rodents housed together in a group exhibit much higher levels of fighting behavior than females housed in groups, though such potential sources of variability in males receive far less attention than variables such as the estrous cycle in females.

Like estrous and group housing, average body size often systematically differs between male and female rodents by a ratio as high as 2:1. Body size/weight influences behavioral repertoire and health measures; thus, without adjustments for body size when conducting sex comparisons, a researcher could erroneously attribute to sex a difference that is better explained simply by differences in size. This is well recognized in the literature. For example, the Y. Zhao et al. (2018) study of sex differences in startle response discussed above did not adjust for body weight, but noting that “it cannot be fully excluded that the sex differences in overall startle response was related to differences in body weight between males and females” (31). A 2022 review in the journal *Nature Communications* on “the unresolved issue of whether sex differences in phenotypic traits in mice may be explained by sex differences in body weight” (Wilson et al. 2022, 1) concluded that body size explains a large amount of variation in rodent behavior, either singly or in interaction with other variables. Dayton et al. (2016), looking at sex-specific variability in rat phenotypes, found that “the phenotype of body weight exhibited a significantly different variance between males and females,” which “corresponded to an increase in magnitude of the measured trait, that is, male rats were larger than female” (Dayton et al. 2016, 1141).¹ In short, if male and female rodents are, for example, different sizes, caged differently, and experience different regimes of hormonal cycling, multiple forms of contextual and contingent variation

1. The matter of sex and body size may be vital when translating preclinical, basic biomedical research to human clinical medicine and pharmacology, as illustrated by the case of the much-contested proposal for sex-specific dosing of the sleep drug zolpidem, despite evidence that adjustment for body weight eliminates the statistical significance of sex in zolpidem drug metabolism (H. Zhao et al. 2023).

may need to be controlled for to make valid comparisons between males and females as groups and to reason about which part of any difference between these groups can be attributed to “sex.”

Here, sex contextualism offers some modest normative guidance. In the face of multiply-entangled forms of variability, similarity, and difference, researchers can make visible the constraints and decision points that go into operationalizing the construct of “sex as a biological variable.” Specifically, sex contextualism invites researchers to generate thick characterizations of how sex, in a particular study, is “pragmatically constituted in an observational frame,” rather than presenting sex-related findings as unmediated read-outs of “sex itself” (Richardson 2022, 10). Researchers can make context visible by “contextually defin[ing] sex-related biological variables” (10), “justify[ing] their choices in how they operationalize sex” (10), and appropriately characterizing sex-related variables as producing “variation between and within sex-defined classes in a context-sensitive manner” (Richardson 2022, Table 1, 12). In this way, contextualist frameworks that rigorously specify the partiality, contingency, and artifactuality of particular operationalizations of sex in laboratory model organism research not only better describe research practices, but also offer an ameliorative approach in the face of the risks of sex essentialism, clarifying the strengths, weaknesses, and generalizability of inferences made about sex differences from these studies.

Acknowledgments

Thank you to the journal’s two anonymous reviewers, colleagues Aya Evron, Donna Maney, and Janet Rich-Edwards, audiences at Wesleyan University and the University of Toronto, and members of the Harvard GenderSci Lab for valuable comments that contributed to the development of this essay. This work was supported by the National Science Foundation under Grant Number 2341785.

Literature cited


- Baker, Monya. 2016. “1,500 scientists lift the lid on reproducibility.” *Nature* 533 (7604): 452–454. <https://doi.org/10.1038/533452a>.
- Becker, J. B., A. P. Arnold, K. J. Berkley, J. D. Blaustein, L. A. Eckel, E. Hampson, J. P. Herman, S. Marts, W. Sadee, M. Steiner, J. Taylor, and E. Young. 2005. “Strategies and methods for research on sex differences in brain and behavior.” *Endocrinology* 146 (4): 1650–73. <https://doi.org/10.1210/en.2004-1142>.
- Becker, J. B., B. J. Prendergast, and J. W. Liang. 2016. “Female rats are not more variable than male rats: a meta-analysis of neuroscience studies.” *Biol Sex Differ* 7: 34. <https://doi.org/10.1186/s13293-016-0087-5>.
- Beery, A. K., and I. Zucker. 2011. “Sex bias in neuroscience and biomedical research.” *Neurosci Biobehav Rev* 35 (3): 565–72. <https://doi.org/10.1016/j.neubiorev.2010.07.002>.
- Cahill, L. 2006. “Why sex matters for neuroscience.” *Nat Rev Neurosci* 7 (6): 477–84. <https://doi.org/10.1038/nrn1909>.
- Clayton, Janine A., and Francis S. Collins. 2014. “Policy: NIH to balance sex in cell and animal studies.” *Nature* 509 (7500): 282–283. <https://doi.org/10.1038/509282a>.
- Dayton, A., E. C. Exner, J. D. Bukowy, T. J. Stodola, T. Kurth, M. Skelton, A. S. Greene, and A. W. Cowley, Jr. 2016. “Breaking the Cycle: Estrous Variation Does Not Require Increased Sample Size in the Study of Female Rats.” *Hypertension* 68 (5): 1139–1144. <https://doi.org/10.1161/hypertensionaha.116.08207>.

- Ehrenreich, Barbara, and Deirdre English. 2005. *For her own good: two centuries of the experts' advice to women*. New York: Anchor Books.
- Fausto-Sterling, Anne. 1985. *Myths of gender: biological theories about women and men*. New York: Basic Books.
- Garcia-Sifuentes, Y., and D. L. Maney. 2021. "Reporting and misreporting of sex differences in the biological sciences." *Elife* 10. <https://doi.org/10.7554/eLife.70817>.
- Ghorayshi, Azeen. 2023. "Guess Which Sex Behaves More Erratically (at Least in Mice)." *The New York Times* (Online), 7 March, 2023, Science. <https://www.nytimes.com/2023/03/07/science/female-mice-hormones.html>.
- Grosz, E. A. 1994. *Volatile bodies: Toward a corporeal feminism*. Bloomington: Indiana University Press.
- "Harvard Scientists Make the Case for Female Mice in Neuroscience Research," 8 March, 2023, <https://scitechdaily.com/harvard-scientists-make-the-case-for-female-mice-in-neuroscience-research/>.
- Haverfield, J., & Tannenbaum, C. 2021. "A 10-year longitudinal evaluation of science policy interventions to promote sex and gender in health research." *Health Research Policy and Systems*, 19 (1): 94. <https://doi.org/10.1186/s12961-021-00741-x>.
- Heidari, S., Babor, T. F., De Castro, P., Tort, S., & Curno, M. . 2016. "Sex and Gender Equity in Research: Rationale for the SAGER guidelines and recommended use." *Research Integrity and Peer Review* 1 (2). <https://doi.org/10.1186/s41073-016-0007-6>.
- Itoh, Yuichiro, and Arthur P. Arnold. 2015. "Are females more variable than males in gene expression? Meta-analysis of microarray datasets." *Biology of Sex Differences* 6 (1): 18. <https://doi.org/10.1186/s13293-015-0036-8>.
- Kafkafi, N., J. Agassi, E. J. Chesler, J. C. Crabbe, W. E. Crusio, D. Eilam, R. Gerlai, I. Golani, A. Gomez-Marin, R. Heller, F. Iraqi, I. Jaljuli, N. A. Karp, H. Morgan, G. Nicholson, D. W. Pfaff, S. H. Richter, P. B. Stark, O. Stiedl, V. Stodden, L. M. Tarantino, V. Tucci, W. Valdar, R. W. Williams, H. Würbel, and Y. Benjamini. 2018. "Reproducibility and replicability of rodent phenotyping in preclinical studies." *Neurosci Biobehav Rev* 87: 218-232. <https://doi.org/10.1016/j.neubiorev.2018.01.003>.
- 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. "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>.
- Levy, D. R., N. Hunter, S. Lin, E. M. Robinson, W. Gillis, E. B. Conlin, R. Anyoha, R. M. Shansky, and S. R. Datta. 2023. "Mouse spontaneous behavior reflects individual variation rather than estrous state." *Curr Biol* 33 (7): 1358-1364 e4. <https://doi.org/10.1016/j.cub.2023.02.035>.
- Lunbeck, Elizabeth. 1994. *The psychiatric persuasion: knowledge, gender, and power in modern America*. Princeton, N.J.: Princeton University Press.
- Maney, D. L., and J. W. Rich-Edwards. 2023. "Sex-Inclusive Biomedicine: Are New Policies Increasing Rigor and Reproducibility?" *Womens Health Issues* 33 (5): 461-464. <https://doi.org/10.1016/j.whi.2023.03.004>.
- Massa, M. G., K. Aghi, and M. J. Hill. 2023. "Deconstructing sex: Strategies for undoing binary thinking in neuroendocrinology and behavior." *Horm Behav* 156: 105441. <https://doi.org/10.1016/j.yhbeh.2023.105441>.
- Mogil, J. S., and M. L. Chanda. 2005. "The case for the inclusion of female subjects in basic science studies of pain." *Pain* 117 (1-2): 1-5. <https://doi.org/10.1016/j.pain.2005.06.020>.

- “NIH Style Guide.” 2024. National Institutes of Health (US). Accessed 28 Sept. 2024. <https://www.nih.gov/nih-style-guide>.
- Pape, M., M. Miyagi, S. A. Ritz, M. Boulicault, S. S. Richardson, and D. L. Maney. 2024. “Sex contextualism in laboratory research: Enhancing rigor and precision in the study of sex-related variables.” *Cell* 187 (6): 1316–1326. <https://doi.org/10.1016/j.cell.2024.02.008>.
- Pecheva, Ekaterina, 7 March, 2023, “The Case for Female Mice in Neuroscience Research,” <https://neurosciencenews.com/female-mice-research-neuroscience-22730/>.
- Prendergast, B. J., K. G. Onishi, and I. Zucker. 2014. “Female mice liberated for inclusion in neuroscience and biomedical research.” *Neurosci Biobehav Rev* 40: 1–5. <https://doi.org/10.1016/j.neubiorev.2014.01.001>.
- Richardson, Sarah S. 2012. “Sexing the X: How the X became the ‘Female Chromosome’.” *Signs* 37 (4).
- . 2022. “Sex Contextualism.” *Philosophy, Theory, and Practice in Biology* 14 (2). <https://doi.org/10.3998/ptpbio.2096>.
- Richardson, Sarah S., Meredith Reiches, Heather Shattuck-Heidorn, Michelle Lynne Labonte, and Theresa Consoli. 2015. “Focus on preclinical sex differences will not address women’s and men’s health disparities.” *Proceedings of the National Academy of Sciences* 112 (44): 13419–13420. <http://www.pnas.org/content/112/44/13419.full>.
- Ritz, Stacey A. 2016. “Complexities of Addressing Sex in Cell Culture Research.” *Signs: Journal of Women in Culture and Society* 42 (2): 307–327. <https://doi.org/10.1086/688181>.
- Rocks, Devin, Heining Cham, and Marija Kundakovic. 2022. “Why the estrous cycle matters for neuroscience.” *Biology of Sex Differences* 13 (1): 62. <https://doi.org/10.1186/s13293-022-00466-8>.
- Shansky, Rebecca M. 2019. “Are hormones a ‘female problem’ for animal research?” *Science* 364 (6443): 825–826. <https://doi.org/10.1126/science.aaw7570>.
- Shields, Stephanie A. 1982. “The Variability Hypothesis: The History of a Biological Model of Sex Differences in Intelligence.” *Signs* 7 (4): 769–797.
- Smarr, Benjamin, and Lance J. Kriegsfeld. 2022. “Female mice exhibit less overall variance, with a higher proportion of structured variance, than males at multiple timescales of continuous body temperature and locomotive activity records.” *Biology of Sex Differences* 13 (1): 41. <https://doi.org/10.1186/s13293-022-00451-1>.
- Smiley, K. O., K. M. Munley, K. Aghi, S. E. Lipshutz, T. M. Patton, D. S. Pradhan, T. K. Solomon-Lane, and S. E. D. Sun. 2024. “Sex diversity in the 21st century: Concepts, frameworks, and approaches for the future of neuroendocrinology.” *Horm Behav* 157: 105445. <https://doi.org/10.1016/j.yhbeh.2023.105445>.
- Wald, C., and C. Wu. 2010. “Biomedical research. Of mice and women: the bias in animal models.” *Science* 327 (5973): 1571–2. <https://doi.org/10.1126/science.327.5973.1571>.
- “What are Sex & Gender?”. 2024. Office of Research on Women’s Health (NIH). Accessed 28 Sept. 2024. <https://orwh.od.nih.gov/sex-gender>.
- White, J., Tannenbaum, C., Klinge, I., Schiebinger, L., & Clayton, J. . 2021. “The Integration of Sex and Gender Considerations Into Biomedical Research: Lessons From International Funding Agencies.” *The Journal of Clinical Endocrinology & Metabolism* 106 (10): 3034–3048. <https://doi.org/10.1210/clinem/dgab434>.
- Willyard, C. 2018. “Squeaky clean mice could be ruining research.” *Nature* 556 (7699): 16–18. <https://doi.org/10.1038/d41586-018-03916-9>.

- Wilson, Laura A. B., Susanne R. K. Zajitschek, Malgorzata Lagisz, Jeremy Mason, Hamed Hase-limashhadi, and Shinichi Nakagawa. 2022. "Sex differences in allometry for phenotypic traits in mice indicate that females are not scaled males." *Nature Communications* 13 (1): 7502. <https://doi.org/10.1038/s41467-022-35266-6>
- Woitowich, N. C., Beery, A., & Woodruff, T. 2020. "A 10-year follow-up study of sex inclusion in the biological sciences." *eLife* 9. <https://doi.org/10.7554/eLife.56344>.
- Wolcott, Nora S., Kevin K. Sit, Gianna Raimondi, Travis Hodges, Rebecca M. Shansky, Liisa A. M. Galea, Linnaea E. Ostroff, and Michael J. Goard. 2022. "Automated classification of estrous stage in rodents using deep learning." *Scientific Reports* 12 (1): 17685. <https://doi.org/10.1038/s41598-022-22392-w>.
- Young, Iris Marion. 1990. "Pregnant Embodiment: Subjectivity and Alienation." In *Throwing like a girl and other essays in feminist philosophy and social theory*, 160-174. Bloomington: Indiana University Press.
- Zajitschek, S. R., F. Zajitschek, R. Bonduriansky, R. C. Brooks, W. Cornwell, D. S. Falster, M. Lagisz, J. Mason, A. M. Senior, D. W. Noble, and S. Nakagawa. 2020. "Sexual dimorphism in trait variability and its eco-evolutionary and statistical implications." *Elife* 9. <https://doi.org/10.7554/eLife.63170>.
- Zhao, H., M. DiMarco, K. Ichikawa, M. Boulicault, M. Perret, K. Jillson, A. Fair, K. DeJesus, and S. S. Richardson. 2023. "Making a 'sex-difference fact': Ambien dosing at the interface of policy, regulation, women's health, and biology." *Soc Stud Sci* 53 (4): 475-494. <https://doi.org/10.1177/03063127231168371>.
- Zhao, Y., E. Y. Bijlsma, M. P. Verdouw, and L. Groenink. 2018. "No effect of sex and estrous cycle on the fear potentiated startle response in rats." *Behav Brain Res* 351: 24-33. <https://doi.org/10.1016/j.bbr.2018.05.022>.

© 2025 Author(s)

This is an open-access article, licensed under
Creative Commons Attribution 4.0 International 

This license requires that reusers give credit to the creator(s). It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format.

ISSN 2475-3025