CellPress

Cell
Leading Edge

Perspective

Sex contextualism in laboratory research: Enhancing rigor and precision in the study of sex-related variables

Madeleine Pape,[1,*] Miriam Miyagi,[2] Stacey A. Ritz,[3] Marion Boulicault,[4] Sarah S. Richardson,[5,6] and Donna L. Maney[7,8]
[1]Institute of Social Sciences, University of Lausanne, Lausanne, Switzerland
[2]Center for Computational Molecular Biology, Brown University, Providence, RI, USA
[3]Department of Pathology & Molecular Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[4]Department of Philosophy, University of Edinburgh, Edinburgh, Scotland
[5]Department of the History of Science, Harvard University, Cambridge, MA, USA
[6]Committee on Degrees in Studies of Women, Gender, and Sexuality, Harvard University, Cambridge, MA, USA
[7]Department of Psychology, Emory University, Atlanta, GA, USA
[8]Harvard-Radcliffe Institute, Harvard University, Cambridge, MA, USA
*Correspondence: madeleine.pape@unil.ch
https://doi.org/10.1016/j.cell.2024.02.008

SUMMARY

Understanding sex-related variation in health and illness requires rigorous and precise approaches to revealing underlying mechanisms. A first step is to recognize that sex is not in and of itself a causal mechanism; rather, it is a classification system comprising a set of categories, usually assigned according to a range of varying traits. Moving beyond sex as a system of classification to working with concrete and measurable sex-related variables is necessary for precision. Whether and how these sex-related variables matter—and what patterns of difference they contribute to—will vary in context-specific ways. Second, when researchers incorporate these sex-related variables into research designs, rigorous analytical methods are needed to allow strongly supported conclusions. Third, the interpretation and reporting of sex-related variation require care to ensure that basic and preclinical research advance health equity for all.

INTRODUCTION

Over the past 30 years, scientists and policymakers in the United States (US) and elsewhere have called for more attention to sex in biomedical research. The key arguments for this increased attention are two-fold: first, that consideration of sex is necessary to advance the health of women,[1] and second, that factoring sex into research designs is required to ensure rigor, reproducibility, precision, and transparency in both clinical and preclinical research.[2,3] It is often argued that without considering and understanding sex-related variation at the basic and preclinical level, flawed assumptions could be carried into clinical trials and ultimately clinical practice, which, coupled with the tendency to over-generalize from data derived from male models and datasets, could have particularly detrimental consequences for women.[4] In other words, without adequate consideration of sex, the fear is that biomedical research could fall short of standards of rigor and precision, compromising its potential to advance health for all.

As a result of these calls to action, several funding bodies have introduced policies requiring the consideration of sex and/or gender throughout the research spectrum.[5] One example is the Sex as a Biological Variable (SABV) policy of the National Institutes of Health (NIH) in the US. This mandate requires that NIH-funded research include female and male animals/participants (unless a single-sex approach is clearly justified) and that findings be disaggregated by sex.[6] Such policies are not limited to funding agencies; for example, many journals have adopted the Sex and Gender Equity in Research (SAGER) guidelines,[7] which similarly recommend disaggregation of data by sex.

Since these policies were implemented, consideration of sex and inclusion of female and male animals/participants in biomedical studies has become more common.[8] A study of research proposals in Canada showed, for example, that between 2011 and 2019, the integration of sex into research designs increased from 22% to 83%.[9] This finding suggests that the policies have likely increased the inclusivity of biomedical research. This increase has, however, occurred without concomitant emphasis on scientific rigor with respect to precise operationalization of variables, appropriate choices of analytical approaches, and accurate reporting.[10–12] For example, a recent analysis showed that, in a sample of papers published in 2019, 70% of claims of sex-specific effects were not supported by appropriate statistical evidence.[11] Given that mandates like the SABV policy are situated within broader efforts to enhance rigor and reproducibility,[13] work remains to ensure that the study of sex-related variation achieves these aims.

Here, we discuss three main ways in which researchers can improve the precision and rigor of research involving sex categories. First, researchers can operationalize "sex" in ways that

focus on concrete, measurable variables rather than relying on the proxy categories of female and male. Being specific about how these variables (for example, chromosome complement, hormone levels, or gendered social and environmental exposures) contribute to research observations can support causal hypotheses about the role of sex-related variation in health outcomes. This approach builds on the concept of sex contextualism,[14] at the core of which is the observation that the relationship between sex-related factors and experimental outcomes will vary in context-specific ways across different research settings. Our argument also draws from the foundational work of others who have pointed out that sex is plurally defined and comprises multiple variables that can vary dynamically.[15–19] Second, we call for enhanced rigor in research design and analytical methods in order to improve the accuracy of claims about the potential influences of sex-related factors, and to expand beyond "sex differences" alone as a means of describing and explaining variation. Third, we urge transparency and care in the reporting of findings and the generalization of results. Transparency relies on accurate accounts of the actual distributions of data and appropriate interpretation of any statistically significant difference between category means; in addition, care should be taken when using findings about sex-related variation in laboratory models to make predictions about diverse and socially complex human populations.

Increasing rigor and precision in the consideration of sex goes hand-in-hand with the policy goals of addressing gender/sex disparities in health outcomes. Contextualist approaches to sex-related variation can enhance the accuracy and efficacy of clinical interventions in diverse human populations, while also minimizing the potential harms of such research. Studies of sex-related variation occur in a broader cultural milieu in which women and men are regularly constructed as opposites, gender inequities continue in both the public and private spheres, and sex and gender minorities face existential threats—precisely because they challenge deeply held convictions about the nature of sex categories.[20] Scientific findings about sex-related variation filter into the public consciousness and can shape gender/sex stereotypes[21] and attitudes toward minoritized groups.[22] There is thus much at stake—for both human health and gender equity more broadly—in striving to uphold the highest scientific standards in the study of sex-related variation in basic science.

OPERATIONALIZING SEX

Sex is not a causal mechanism

In the years since funding agencies and journals began to introduce policies for the consideration of sex, it has become commonplace to refer to sex as a "biological variable." As a starting point for situating sex-related factors in their context, we emphasize that sex is not a variable that is in and of itself a biological mechanism; rather, sex is better understood as a system of classification. Once assigned, sex categories are frequently treated as causal mechanisms, which is evident in claims that biological phenomena are "driven," "influenced," or "impacted" by sex. However, sex requires careful operationalization in order to move beyond a set of assigned categories, such as female and male, and toward the actual mechanisms of interest. By acknowledging that it is not sex itself that generates effects in experimental settings, but rather any one (or more) of the mechanisms that assigned sex categories are taken to represent, we can be more precise about the relationships between biological variables and observed outcomes.[17,23,24] Importantly, this approach can also allow for more accurate descriptions and explanations of overlap and similarity between sex-classified groups and the diversity and heterogeneity within them (Table 1).

As an example, the use of the category "women" as a proxy variable—instead of the more precise variable of uterine presence/absence—underestimates incidence of uterine cancer in human populations by 23%–53% and racial disparities in cervical cancer by 44%; it also obscures the incidence of these cancers in intersex and transgender people.[25,26] In this case, adequately operationalizing sex beyond an identity or assigned sex category dramatically enhances the value of the data. The presence of a uterus is an example of how the use of sex-related variables relevant to the question at hand can result in a more precise operationalization of sex, including in basic research settings.

Operationalization is particularly important when considering variables that correspond imperfectly with assigned sex categories. Hormone levels, for example, may correlate with assigned sex categories but can also vary independently of them. In most cisgender women early in the menstrual cycle or in post-menopause, estradiol levels overlap with those of cisgender men.[38] In all people, regardless of sex assigned at birth, estradiol levels can vary in response to social context and behaviors as well as hormone therapy[39]; this variation reflects the intersection of social factors (e.g., access to hormonal contraception, post-menopausal hormone replacement therapies, or gender-affirming care) with a sex-related variable.

A conceptual shift from sex category toward plausible mechanisms associated with sex-related variables can support more accurate and clinically relevant interpretations of data. This framing may help to identify specific sex-related independent variables with effects that can be investigated or controlled for as part of a study and consequently determine the types of data that need to be collected to facilitate consideration of these variables. Careful consideration of how sex is operationalized can also help researchers more accurately consider and explain patterns of variation that are not well-classified in terms of strictly binary female/male categories.

Sex depends on context

Whether and how any sex-related variable matters—and what patterns of difference and variation it contributes to—cannot be assumed *a priori*, as these relationships will vary across research questions, model species, and lab settings.[14] Consequently, researchers should consider which sex-related variables or covariates could be most relevant to the experimental setting at hand. When designing a study with the explicit aim of exploring sex-related factors, researchers should be aware that contextual factors such as age, environment, ovarian cycle phase, time of day, time of year, cell type, etc., each affect levels of hormones, gene expression, and other correlates of sex. The dynamic nature of hormone levels and gene expression dictates

Table 1. Practicing rigor and precision in the consideration of sex-related variables

Common issues in the use of sex-related variables	Solutions	Examples/further reading
sex categories are treated as a biological variable without further operationalization	operationalize sex categories in terms of sex-related variables that are context- and model-specific, appropriate to the research question, concrete, and measurable; recognize the limited explanatory value of assigned sex categories used alone	instead of using the category "women" as a proxy variable, the more precise variable of uterine presence/absence increases accuracy in estimating incidence of uterine cancer in human populations and racial disparities in cervical cancer;[25,26] Yang et al. investigated how androgen-mediated signaling contributes to sex-related variation in the development of immune phenotypes;[27] Bongen et al. acknowledged in their study of sex-linked gene expression and immunity: "we need to improve our understanding of the biological factors that underlie sex differences so that we do not rely on the crude labels of 'male' and 'female' when predicting disease risks"[28]
process for classifying research subjects by sex category is not reported	include an explicit description of how sex categories were assigned (in the Methods section or in supplemental information)	Massa et al. explained their operationalization in the Methods section: "female (defined as having small anogenital distance at weaning and presence of ovaries at time of death) and male (defined as having large anogenital distance at weaning and presence of testes postmortem)"[29]
environmental variables outside of the hypothesized sources of variation are not identified or considered; this issue often appears as an attribution of all sex-related variation to "hormones"[30] rather than more thoughtful consideration of a broader range of possible mechanisms	carefully consider factors in the lab environment (such as housing density or exposure to different types of conspecifics) and/or gendered exposures in the interpretation of sex-related variation	sex-related differences in functioning of the hypothalamo-pituitary-adrenal axis can be at least partly attributed to gender-related differences in social buffering;[31] Klein and Nelson found sex-related variation in immune responses in meadow voles when they were housed in pairs, but not when housed individually[32]
dividing a sample into two categories (F/M) with no *a priori* hypothesis about how or why they may differ	rather than comparisons of means, consider statistical approaches that can reveal more complex relationships	Smiley et al. outline a variety of alternative statistical approaches for analyzing sex variability, such as data reduction or clustering[33]
claiming a statistically significant difference between the treatment responses of sex-classified groups without testing for one (the DISS error)	test for a sex-by-treatment/exposure interaction in all cases, even when underpowered to detect small interactions	see Figure 1 and further elaboration by Rich-Edwards and Maney[34]
inaccurate use of terminology such as "dimorphism" and "sex specific"	describe sex-related variation in ways that accurately reflect the actual distributions of the data and what can and cannot be concluded from null hypothesis significance testing	Yang et al.[27] and Naqvi et al.[35] described immunological factors and gene expression, respectively, as "sex-biased" rather than sex-specific or dimorphic; see further reading for recommendations on how to graph categorical data in ways that depict overlap[23,24,34]
interpreting a finding of a statistically significant difference between female- and male-classified groups of non-human animals as pointing to a clinically significant difference warranting development of sex-category-specific treatments in humans	recognize the specificities of the chosen model and describe potential limitations when generalizing to other models or to a clinical setting; in many cases, due to divergent natural histories, gene expression, and other factors,[35] sex-related variation in a given non-human model cannot be generalized to humans; in some cases, findings cannot be generalized even across strain within species	Fischer and Riddle reviewed evidence of sex-related variation in aging processes in fruit flies, nematodes, mice, and humans, and reported that although sex differences exist, they are often not conserved among species[36]; sex-related variation in gut microbiota does not generalize from one strain of laboratory mouse to another[37]

careful consideration of the tissues and time points that are sampled, in the context of the goals of a given study.[29,30,40] Similarly, other correlates of sex categories should be approached as highly context-specific.

In considering context, researchers can first and foremost address the particularities of the model itself; that is, the characteristics unique to a model species or strain and developmental stage (Table 1). The species-specificity of findings about sex-related variation is well known to researchers working in fields such as behavioral neuroendocrinology and neuroethology. Decades of research on rodents, songbirds, and fish, for example, has shown that female-male differences in mating behaviors, neuroanatomy, and neurochemistry can vary dramatically even across closely related species.[41–43] In a large study of gene expression in 12 tissues in five mammalian species, including humans and mice, Naqvi and colleagues showed that samples tended to cluster by species, not sex, suggesting that although the genes themselves are conserved across species, sex-related differences in their expression are not.[35] They argued that sex bias in the expression of most genes has evolved recently and, as such, is not shared among most mammals. Even within species, sex-related variation may not generalize to other animals of the same species; in rodents in particular, strain, parent-offspring relations, and even litter effects could be highly relevant to the size and direction of any observed sex-related variation.

Cross-species differences have been leveraged to understand the mechanisms that underlie biological variation and its evolution. But modeling human health in non-human animals relies on cross-species similarity, not variation; such variation can thus be a barrier to generalizability. *Drosophila* and rodents, for example, have very different sex-related biologies from each other and from humans, making it difficult to predict whether or how a sex-related difference in one of these models will manifest in humans. Although all three taxa share basic biology that includes homologous protein-protein interactions, DNA sequences, and epigenetic modifications, the extent to which sex-related factors interact with these features, and how they do so, varies widely. Modeling human sex-related variation in depression, anxiety, addiction, or post-traumatic stress disorder can be challenging because the associated behaviors in rodents do not always mirror the direction of female-male differences in humans.[44–47] For example, male rats exhibit greater anxiety than their female counterparts across various models of anxiety, including the open field test and elevated plus maze, whereas in human populations, women tend to be diagnosed with anxiety more often than men.[44] This lack of generalizability is not necessarily related to lack of homology of the mechanisms underlying such conditions; rather, it may be related to the profound diversity of sex-related, social, and environmental factors that interact to produce observed outcomes.

These caveats are particularly important when modeling human diseases and disorders that seem to be more prevalent in one sex. It is commonly argued, for example, that for disorders that are more prevalent in one gender, e.g., men, the underlying biological mechanisms must be modeled in non-human animals of the associated sex category, e.g., males.[48,49] But contextualism teaches us that it is an empirical question whether a female or male non-human model at a particular stage of development or gonadal status is most apt for a given scientific inquiry.

A similar principle applies in research involving dissociated cells grown in culture. It is often said, when advocating for the consideration of sex in basic science, that "every cell has a sex."[50] *In vitro* research presents several conceptual challenges, however, to such an endeavor. First, even when the sex category of the cell donor is known, only certain aspects of sex can be represented *ex vivo*.[51] Typically, sex is operationalized in these cases by the chromosome complement (or the presumed complement based on the sex category of the donor); however, the environment in the culture dish, including the hormonal milieu, differs profoundly from that in the intact animal (or human). Cell cultures, therefore, cannot be understood to be representing "women" or "men." Further, although it is rarely acknowledged, almost all primary cells in culture (particularly human ones) originate from a gendered body with gendered experiences prior to cell harvesting. Thus, although the culture dish may seem to offer a highly controlled, context-free environment, there are many factors to consider. A contextualist approach shifts the focus from the sex category of the cell donor to the sex-related processes that can actually be modeled in the *in vitro* environment.

As a second step, researchers can identify features of the environment that may covary with sex-related variables (Table 1).[52–54] For example, at many institutions, male mice are frequently housed singly to prevent fighting whereas female mice are housed together to save on housing costs. In this case, a researcher could consider whether findings that appear to be linked to sex categories are in fact related to housing density. Even when male and female animals are housed under identical conditions, however, the natural history of a species and its sex-related behaviors could result in apparent sex-related findings; for example, female- and male-classified animals may vary in their stress responses to social isolation, which could affect the variable of interest.[55] Overcoming such confounds can be challenging,[56] making it all the more important to acknowledge and consider them when interpreting findings.

Third, researchers could consider that—whether in the laboratory or the clinic—sex difference research occurs in a gendered social world, meaning that social gender beliefs, practices, and assumptions are part of the context in which, and could possibly affect, the research being undertaken. The potential influence of these factors has been well established in studies of the history of the sciences of sex, including in the cases of gendered interpretations of gametes, steroidal hormones, sex determination processes, and the X and Y chromosomes.[57–59] These biases are not limited to studies on humans or human-derived systems. One study showed that, when primed, observers' expectations about sex-related behavior in red-backed salamanders affected the accuracy of behavioral observations.[60] In another study, this research group showed that even without priming, observers can bring preconceived expectations about sex differences into the lab: unprimed women and men observers both believed that the male salamanders would be more aggressive and the female ones better foragers.[61] These examples illustrate that gendered assumptions can influence scientific observation. The degree to which such biases can affect outcomes, particularly in biomedical research in non-human animal models, is not well understood.

In biomedical research, sex is often presented as referencing physiological features such as chromosomes, hormones, and secondary sex characteristics, whereas "gender" is defined as a human-specific phenomenon, relating to one's identity, behaviors, and exposure to socially conditioned and unequal norms, relationships, and environments, which include laws, policies, and opportunity structures.[62] A common practice in biomedical research is to attempt to separate causal factors linked to sex from those linked to gender, in order to support a focus on the former, particularly when using non-human models.[3] For example, although it is widely recognized that clinician bias and other gendered pathways may explain substantial portions of gender disparities in human populations with conditions such as depression,[63] fibromyalgia,[64] and autism,[65] animal models of these conditions often proceed as if the disparities can be explained solely by physiological factors such as hormones, divorced from gender. The study of sex-related biology in animal models in isolation from the gendered contexts that shape human systems is a crucial limitation of preclinical research. Rather than attempting to model human experiences of gender, which most basic and preclinical research cannot do, a contextualist approach supports researchers to think beyond a sex/gender distinction and consider how sex-related variation in any model will always be context-specific.

Vitally, a sex contextualist approach to sex does not deny the possibility of material differences between sexed/gendered bodies. Nor does it deny the importance of non-human models to advancing scientific knowledge about human health and illness. Instead, it shifts the consideration of sex from the disaggregation of data by binary female/male categories to the question of: "how should we operationalize the concept of sex in the particular context of our research?" In other words, it moves each researcher toward a conceptual definition of sex that can be more useful for achieving their research aims and increases clarity about the limitations of any given operationalization of sex. As we discuss in the following section, a contextualist approach can also inform data analysis, which presents opportunities to carefully consider how sex categories are utilized in the study of variation.

TREATMENT OF SEX-RELATED FACTORS IN RESEARCH DESIGN

A priori hypotheses about sex

One common practice in the analysis of sex-related variation is to separate groups of samples, animals, or participants according to an assigned (e.g., based on measures of chromosome complement or anogenital distance) or self-reported sex category. Often, researchers separate female and male groups even when they are not looking to test *a priori* hypotheses about sex. Researchers should be aware, however, that splitting a sample into groups creates a perceptual bias: members of different groups are subsequently assumed to be more different than they actually are, and members of the same group more similar.[22,66] When data for female and male-categorized groups are presented separately in every study, as encouraged by various policies for the consideration of sex, the perception of large female-male differences for biological measures may be magnified. The *a priori* classification of samples by sex without clear justification or hypotheses can in turn lead to "HARKing," or hypothesizing after the results are known.[34,67] When complying with requirements to consider sex, researchers should ensure that all variables included in the analysis are well-chosen, relevant to the research question, and connected to evidence-based hypotheses.

Statistical comparisons of sex-classified samples

Just as dividing a sample into categories without *a priori* hypotheses can present challenges to rigor and reproducibility, so can the seemingly simple act of comparing across those categories. A diversity of analytical approaches can be used in the study of sex-related variation to support precise understandings of mechanisms and distributions without comparing explicitly across sex categories (Table 1). But, despite the fact that most policies for the consideration of sex do not require statistical comparisons between female- and male-classified samples, comparing them remains a common practice.[68] Further, analytical strategies for comparing across sex-classified subgroups are often inappropriately deployed. Several recent studies have shown that a large majority of claims of sex-specific effects, for example, of a drug or exposure, were not supported by valid statistical evidence.[11,68,69]

Here, we offer a hypothetical example to first illustrate a valid approach to testing for sex differences in responses to a treatment. Then, we use this same hypothetical to explain one of the most common errors in testing for these effects. Imagine a study testing for an effect of drug X on tumor volume in a mouse model of cancer. A decade ago, researchers in the US may have had only two groups of mice in such a study; 16 males receiving the drug and 16 males receiving a placebo control. After 2016, to comply with the SABV and SAGER guidelines, they now have four groups: male and female mice treated with drug X, and male and female mice treated with placebo (Figure 1). The design of their study is thus considerably more complex than a simple two-group comparison and requires a suitable analytical approach. To allow proper estimates of whether and how the effects of the treatment depend on sex-related variable(s), appropriate analytical strategies will incorporate sex (precisely operationalized) into a model that includes all animals/participants. Such a model does not ignore sex-related variation, nor does it aggregate data from females and males; instead, it allows formal, quantitative tests of whether the effect of interest depends on sex as operationalized in the context of the study.[12,34,70,71] A common strategy is to test for a statistical interaction between sex-related variables and the effect of a treatment or exposure in an analysis of variance (ANOVA) or a regression. In Figure 1, this interaction is not significant, at $p = 0.89$, meaning there is no statistical evidence to indicate that female and male mice responded differently to the drug.

Unfortunately, researchers test for interactions only rarely.[11,68,70–73] More commonly, they skip this important step and conduct two separate tests for the effect of drug X: one in the female group and one in the male group. In our example, these two tests produce p values of 0.10 and 0.02, respectively. That is, the effect of drug X on tumor volume was not statistically significant in the female group, but it was in the male group (Figure 1). Many researchers interpret such a result to mean that the drug had "different" effects in the female and male groups when in fact the effect of the drug was not compared between the two groups at all (Table 1).[11,68]

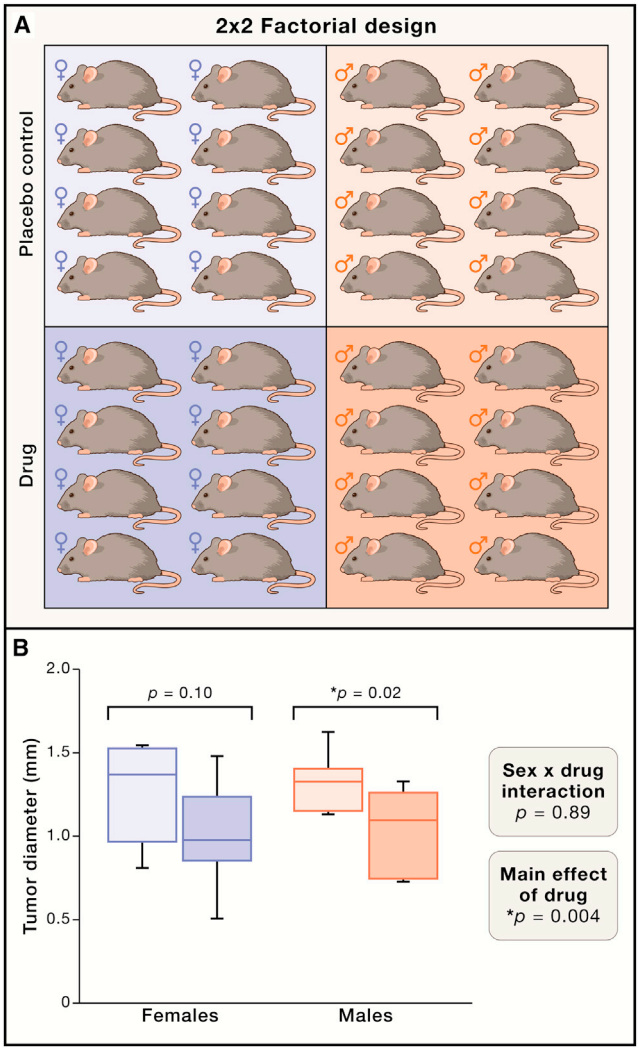

Figure 1. Incorporating a sex-related variable into a simple research design

In this hypothetical study, researchers are testing for an effect of drug X on tumor volume in a mouse model of cancer.

(A) The study has a 2X2 factorial design, where one factor is drug treatment and the other is sex (here, operationalized as a categorical variable such as presence of ovaries or Y chromosome).

(B) Given this design and the mandate to present disaggregated results, the researchers are likely to conduct two separate tests for an effect of drug X, one in female mice and one in male mice, without comparing females and males statistically.[11] Finding a statistically significant effect in male but not in female mice, the researchers then commit a "DISS" error (see text): they conclude a sex-specific effect of the drug despite having not compared the effect of the drug between the sex-classified groups.[12,34] There is thus a risk that this spurious finding of difference would appear in their research report.[11] An appropriate analytical strategy would have shown no statistical evidence that female and male mice responded differently (sex-by-treatment interaction, p = 0.89) and would have shown a highly significant effect of drug X that was independent of sex (main effect of treatment, p = 0.004). Both results, which are relevant to human health, would have been missed using the DISS approach. * statistically significant. Box-and-whisker plots depict hypothetical data analyzed in a two-way univariate ANOVA using SPSS v.29. Data are available from the authors.

This fallacy, dubbed the "difference in sex-specific significance (DISS)" error,[12] was behind a large majority of claims of sex-specific effects in a sample of the 2019 biomedical literature.[11] It continues to be a popular method of "testing" for sex differences[68] and is explicitly endorsed in publicly available online training materials on how to consider sex in biomedical research.[49,74,75] One online course states, for example, that "presenting analyses separately by sex provides the clearest picture of where exposures might differ for men and women."[74] The course goes on to argue that interaction terms are generally not preferred because they are less "intuitive" and more "difficult to calculate and interpret" than the results of separate analyses.

Note that in our hypothetical example (Figure 1), the drug had a statistically significant effect (main effect, p = 0.004) when including all the animals in the study. Thus, provided this result is considered together with the size of the effect, the conclusion from this approach is that the findings could have clinical significance regardless of sex. The main effect of drug X would have been masked by a DISS approach—that is, testing for effects separately within sex subgroups. In the face of the missed generalizable finding, researchers may decide to frame their entire research report around an unsupported "sex-specific effect." The hypothetical authors from Figure 1 could, based on an incomplete analysis, call for future development of drug X *only in men*, an unsupported argument that could undermine women's (and men's) health instead of advancing it. The potential consequences of missing a difference, that is, of a false negative finding, have been much discussed[1,4,76–78]; the potential costs of false *positive* findings, however, have not been as strongly considered. False positive findings of female-male differences, which are made likely by a DISS approach,[34] potentially represent a threat to reproducibility and health equity as significant as that posed by false negatives.

Moving beyond binary comparisons

In our example above, sex is operationalized using a categorical, binary trait, and its associations analyzed using a relatively simple test (ANOVA). As more sophisticated analytical methodologies become more common and accessible, opportunities are arising to move the study of sex-related variation beyond the comparison of categorical variables to better account for dynamic, complex phenotypes.[29,33,79] These strategies include working with bimodal models of continuous sex-related variables or multivariate models of sex as a collection of traits, as well as more exploratory approaches, such as discriminant analysis and probabilistic modeling, which integrate multiple sex-related variables into complex phenotypic endpoints. Such approaches, which could help researchers move away from *a priori* splitting of a sample into two groups, could allow for more precise descriptions and explanations of variation while also reducing tendencies to over-binarize sex-related data.[30] Such a strategy is critical when findings of sex-related variation are used in the development of precise medical interventions that can be tailored to individual needs.[80]

INTERPRETING SEX-RELATED VARIATION AND TRANSLATING IT INTO PRACTICE

When a difference is not really a difference

Even when appropriate analytical approaches are used, interpretations of sex-related variation are not always commensurate

with the story the data actually tell.[23] For example, statistical significance alone is typically used to declare females and males to be different. When a p value falls below 0.05, be it for a simple comparison of means between female and male groups or for a more complex interaction between another variable and sex (however operationalized), the result is often interpreted as "females and males differ." Sometimes this difference is even described as "profound" or "fundamental."[23] But, as is the case for any biological question, p values reveal very little about the size of any categorical difference or its clinical relevance. A p value tells us only that sex-related factors likely explain part of the variation.

Findings of statistically significant female-male differences have led some researchers and policy advocates to imagine a future of sex-specific medical treatments, which they believe would represent a definitive advance for precision, reproducibility, and gender equity in medicine. But a two-sizes-fit-all approach based on binary sex categories may be only marginally better than a generalized model of a biological pathway, particularly when the causal mechanisms underlying a sex-related effect overlap substantially between the categories or when the causal mechanisms remain obscure or incompletely understood, as is often the case even in preclinical research. The two-sizes approach could introduce inaccuracies and harms at the clinical level—not only for sex and gender minorities but also for cisgender women and men who do not approximate their category mean (Figure 2).[81]

For example, despite calls to "recognize sex-specific symptoms of heart disease,"[82] such as chest pain for men and shortness of breath for women, research shows that an individual's sex category is not strongly predictive of symptoms of cardiovascular disease. Common symptoms of heart attack—chest pain, shortness of breath, nausea—are experienced by women and men alike,[83,84] with average differences too small to be clinically informative. One study of chest pain characteristics (CPCs) found that 31 of 34 CPCs (91%) had similar likelihood ratios for the diagnosis of acute myocardial infarction in women and men, with the three remaining CPCs (8.8%) deemed clinically unhelpful.[83] This example highlights the importance of developing more precise ways to address variation (and overlap) across women and men, which cannot be achieved with either a "one-size" or purely sex category-based "two-sizes" approach (Figure 2).

Some of the potential pitfalls of binary, sex-specific approaches in the clinic are illustrated by the sleep aid zolpidem. In 2013, the US Food and Drug Administration (FDA) lowered the recommended initial dosage for women by 50%, claiming that women are more susceptible to next-day drowsiness. No published studies were cited in their decision, however, and no clinical trials demonstrated particular risks to women regarding drug clearance, driving impairment, or adverse reactions.[85] In practice, the FDA's recommendation could mean that many women are being deprived of adequate insomnia treatment without appropriate scientific evidence to support the treatment recommendations.[80,85] Nonetheless, zolpidem is widely cited as an example of a pressing need for more female-male comparisons in preclinical research.[4,76]

Moving toward underlying mechanisms, combined with closer engagement with complex distributions in sex-related data, facilitates precise and equitable care for all patients regardless of how far they fall from an accepted category mean, including but not limited to members of underrepresented groups already marginalized in healthcare (Figure 2). By explicitly operationalizing sex-related variables from the outset, we are better positioned to evaluate findings of statistical significance and identify the potential clinical implications of sex-related associations with greater precision, leading to interventions that better serve diverse populations.

Precision in description

How researchers choose to describe findings can have important implications for how sex-related variation is understood within the biomedical knowledge base. Notably, the language used to describe sex-related associations often paints a stark and sometimes misleading picture of differences between female- and male-categorized subjects in a given study. For example, a difference is commonly described as a "dimorphism" regardless of the extent of overlap in the distributions for the female and male groups. Strictly speaking, dimorphism means that a trait occurs in two distinct forms, but it is common to see the expression "sex dimorphism" used when data from female and male groups overlap substantially or when the authors have not tested statistically for a sex difference.[40,86]

Like the phrase sex dimorphism, the term "sex specific" is often used misleadingly. The term strongly suggests that an effect is occurring *exclusively* in one sex category and not another, when what is more typically observed is a statistical difference in the *degree* to which the two groups exhibit the effect. The term sex-specific is commonly used even without evidence of a difference in degree; for instance, when authors have chosen a DISS approach (see above) and test for effects within each group separately without comparing them statistically with each other. For most sex comparisons, neither "dimorphism" nor "sex-specific" are likely to accurately reflect the real distribution of the data or what can be concluded from null hypothesis significance testing. Nonetheless, the terms are used pervasively in the literature in ways that can present statistical differences between categories as absolute.[11,87]

A more precise approach when interpreting and translating findings of sex-related associations is to describe the distributions for female and male groups, specify the possible contributing mechanism(s), and acknowledge limitations (Table 1). For example, in a recent study of the relationship between sex-linked gene expression and immunity, Bongen and colleagues carefully described how patterns in the mechanisms contributing to mean differences between women and men vary across developmental stages, with older women and men featuring a greater degree of similarity in the expression of some genes.[28] Incorporating attention to mechanisms and using judicious language, the authors wrote, "[w]e need to improve our understanding of the biological factors that underlie sex differences so that we do not rely on the crude labels of 'male' and 'female' when predicting disease risks."

When studies are not designed to test whether the effects of an exposure or treatment depended on sex-related variable(s), researchers should avoid extrapolating to the clinical context, acknowledge the exploratory nature of the study, and call for replication of any incidental findings that appear to be related

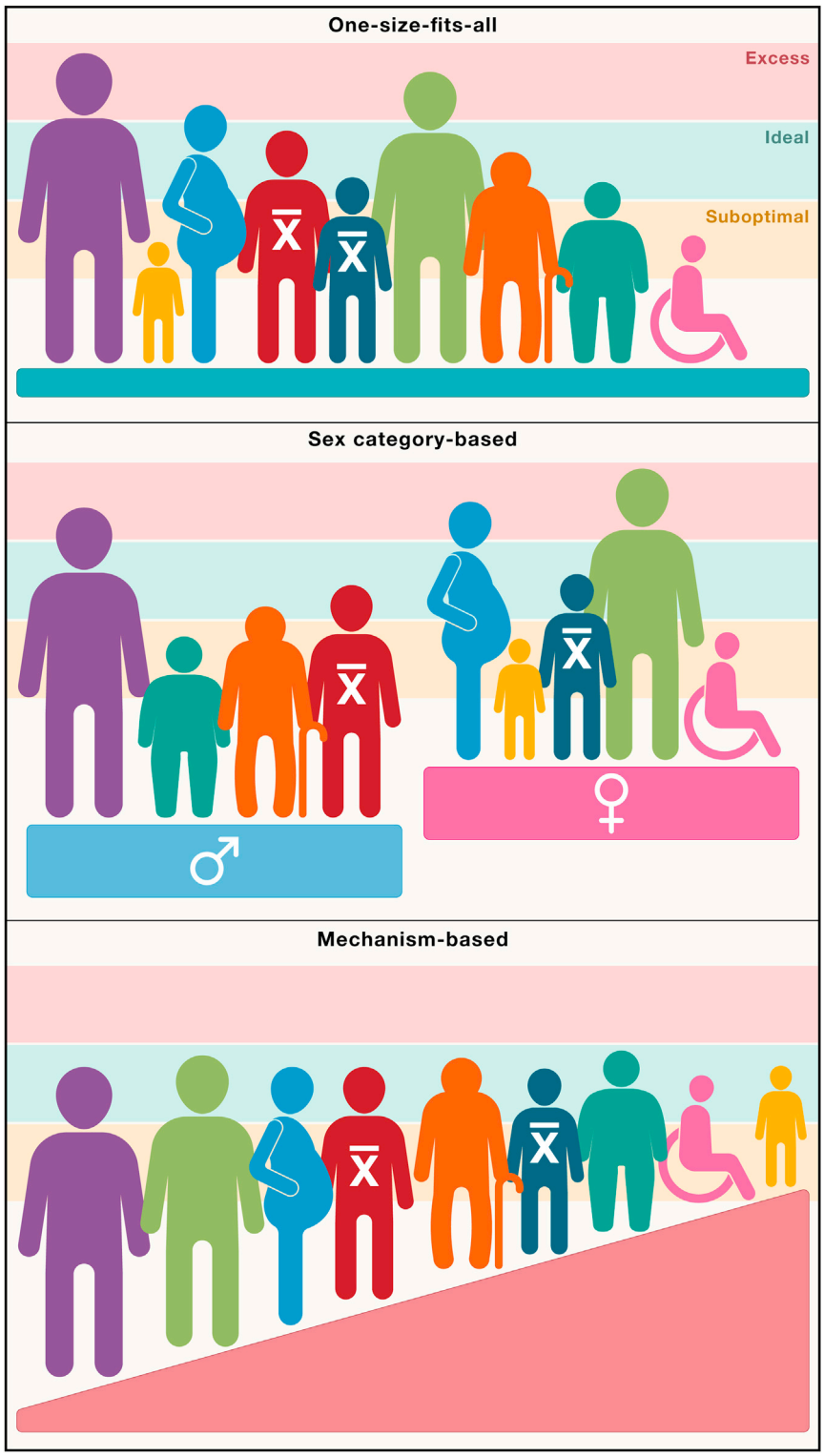

Figure 2. Research aimed at understanding the influence of sex and gender on health must go past a "two-sizes-fit-all" approach

This figure depicts strategies for treating a diverse population; patients whose heads are in the "ideal" zone represent those receiving optimal treatment. The patients marked by "x" represent the average woman and man. In the "one-size-fits-all" approach (top), the average man receives optimal treatment, but the average woman is underserved. With an approach based on binary sex category assignment (middle), individuals who approximate their accepted category mean receive an optimal dose. Many others, however, remain underserved. In practice, those who could be underserved or even harmed by a female-male sex category-based approach include, but are not limited to, those in marginalized groups: disabled, elderly, non-binary, trans, intersex, and other individuals. Importantly, even among those people who would consider themselves to belong to a binary gender category (e.g., cisgender men), there will always be a certain proportion of individuals whose experiences fall outside of a two-sizes-fit-all approach.[81] Rigorous approaches that move past sex categories as proxies and toward understanding mechanisms and addressing variation will serve the most people (bottom), advancing both precision and equity. Figure created with Biorender.com.

to those variables. For example, in a study of periodontitis, Sayad and colleagues observed patterns suggestive of possible sex-related influences on the regulation of long-coding RNAs but acknowledged that the mechanism remains unclear.[88] They emphasized that their findings "add another layer of complexity in the mechanisms of regulation of immune responses among males and females" and called for further research to improve understanding of the mechanisms at play. Researchers should assiduously recognize complexity in the consideration of sex-related variation, and be encouraged to do so by reviewers, editors, and funders.

Thoughtful consideration of sex-related variables requires accurate reporting of the process of sex classification itself, which is seldom included in manuscripts, illustrating the extent to which such categories are considered to be self-evident. Non-human animals are typically categorized at weaning on the basis of morphology (e.g., approximation of anogenital distance); sometimes genotyping or other approaches are used. In research on humans, sex is typically determined by self-report of participants; best practices for how to collect this information are without consensus and rapidly changing.[89] In some fields of study, human data are filtered to exclude all individuals for whom self-reported sex category does not match that assigned by the researcher (such as by genotype in human genomics studies), systematically removing appreciable diversity of configurations of sex from the study.[90] In both human and non-human animal research, transparency and reproducibility

could be improved by normalizing the reporting of the method of sex classification and the implications of any exclusion criteria (Table 1).

Many basic and preclinical researchers are motivated by the potential contributions of their work in non-human animal models and cells to advancing human health, and this motivation often shapes the discussion portion of manuscripts. In the current climate of intense interest in precision medicine, researchers should be careful not to overstate the relevance of findings to human experiences of illness and disease—particularly in the case of sex categories.[80] Building on the attention to context described above, the goals of rigor and reproducibility are strengthened by discussion of the extent to which findings about sex-related variables may or may not be generalizable beyond the model and the laboratory. Research on sex- and gender-related variation in health and illness is by nature interdisciplinary and is therefore ideally informed by the social sciences and gender studies literature, particularly that which considers the social dimensions of sex-related variation in health and biological outcomes.[91–94]

CONCLUSIONS

Policies aimed at redressing a historical tendency to utilize exclusively male models in many areas of preclinical research, such as mandates to consider sex, have contributed to shifting ways of thinking about the relevance of sex to animal models, data analysis, and interpretation of findings. Biomedical research is now more likely to include women and female non-human animals,[7] which represents a step toward inclusivity. Such policies have been less successful, however, at promoting rigor and reproducibility.[11,12] Indeed, in some respects, policies for the consideration of sex may have proliferated and codified imprecise operationalizations and substandard approaches to data analysis, which may impede the larger goal of health equity. Thus, there is now not only an opportunity but also a clear need to develop tools to ensure that such policies do not inadvertently compromise rigor and precision and, in turn, negatively impact efforts to advance health and equity.

A more contextual use of sex categories is vital to ensure that any sex-related findings are accurate, reproducible, and clinically relevant. This practice begins, first, with recognition of the key distinction between assigned categories and biologically relevant variables. Identifying context-specific, sex-related factors as part of this approach brings further precision by concretizing what could otherwise be an imprecise and abstract categorical variable. Identifying and parsing potential confounds, such as housing and researcher expectations in the case of research with laboratory animals, is critical for the interpretation of findings of apparent sex-related variation. Second, consideration of sex-related variables during the analysis stage of a study must adhere to the principles of hypothesis-driven experimental research. Female-male comparisons alone, even when conducted using appropriate statistical tests, may be inadequate for capturing variability and may misdirect research through an over-reliance on sex categories. Third, researchers should be precise in the presentation of data, rigorous in the interpretation of sex-related findings, and cautious, acknowledging the limitations of their approach. There is considerable benefit to bringing a more critical, thoughtful lens to how assumptions about sex-related variation manifest themselves in scientific research and the language used to describe it. We encourage researchers and policymakers alike to strive for new standards of rigor and precision in the consideration of sex in biomedical research and to embrace the opportunities for innovation and discovery that this endeavor will most certainly bring.

ACKNOWLEDGMENTS

We thank the Radcliffe Institute for Advanced Study at Harvard University for supporting the workshop that made this paper possible. The Exploratory Seminar titled "Sex in Context," which took place from the 30th to the 31st of March 2023, was co-organized by Sarah Richardson and Madeleine Pape; participants included Madina Agénor, Elizabeth Barr, Jill Becker, Marion Boulicault, Cato Clune-Taylor, Marina DiMarco, Kelsey Ichikawa, Rebecca Jordan-Young, Nancy Krieger, Donna Maney, Miriam Miyagi, Robin Nelson, Stacey Ritz, Tamara Rushovich, and Marcia Stefanick. We also thank the Flax Foundation for supporting one of our early career contributors (M.P.) to attend the workshop. The work for this article was supported in part by grants from the National Institutes of Health, United States (1R01GM152543-01 to D.L.M.), National Science Foundation, United States (PRFB-2305910 to M.M.), Robert Wood Johnson Foundation (79892 to S.R.), and the Swiss National Science Foundation, Switzerland (PZ00P1_208890/1 to M.P.).

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Stachenfeld, N.S., and Mazure, C.M. (2022). Precision medicine requires understanding how both sex and gender influence health. Cell *185*, 1619–1622.
2. Clayton, J.A., and Collins, F.S. (2014). Policy: NIH to balance sex in cell and animal studies. Nature *509*, 282–283.
3. Clayton, J.A. (2016). Studying both sexes: a guiding principle for biomedicine. FASEB J. *30*, 519–524.
4. 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., et al. (2015). Opinion: Sex inclusion in basic research drives discovery. Proc. Natl. Acad. Sci. USA *112*, 5257–5258.
5. White, J., Tannenbaum, C., Klinge, I., Schiebinger, L., and Clayton, J. (2021). The Integration of Sex and Gender Considerations Into Biomedical Research: Lessons From International Funding Agencies. J. Clin. Endocrinol. Metab. *106*, 3034–3048.
6. National Institutes of Health (NIH) (2014). Consideration of sex as a biological variable in NIH-funded research. NOT-OD-15-102 Guidance. https://orwh.od.nih.gov/sites/orwh/files/docs/NOT-OD-15-102_Guidance.pdf.
7. Heidari, S., Babor, T.F., De Castro, P., Tort, S., and Curno, M. (2016). Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. Res. Integr. Peer Rev. *1*, 2.
8. Woitowich, N.C., Beery, A., and Woodruff, T. (2020). A 10-year follow-up study of sex inclusion in the biological sciences. eLife *9*, e56344.
9. Haverfield, J., and Tannenbaum, C. (2021). A 10-year longitudinal evaluation of science policy interventions to promote sex and gender in health research. Health Res. Policy Syst. *19*, 94.
10. Rich-Edwards, J.W., Kaiser, U.B., Chen, G.L., Manson, J.E., and Goldstein, J.M. (2018). Sex and Gender Differences Research Design for Basic, Clinical, and Population Studies: Essentials for Investigators. Endocr. Rev. *39*, 424–439.

11. Garcia-Sifuentes, Y., and Maney, D.L. (2021). Reporting and misreporting of sex differences in the biological sciences. eLife *10*, e70817.
12. Maney, D.L., and Rich-Edwards, J.W. (2023). Sex-Inclusive Biomedicine: Are New Policies Increasing Rigor and Reproducibility? Womens Health Issues *33*, 461–464.
13. Collins, F.S., and Tabak, L.A. (2014). Policy: NIH plans to enhance reproducibility. Nature *505*, 612–613.
14. Richardson, S.S. (2022). Sex contextualism. Philos. Theor. Pract. Biol. *14*, 2.
15. Fausto-Sterling, A. (1993). The five sexes. The Sciences *33*, 20–24.
16. Mol, A. (2002). The Body Multiple: Ontology in Medical Practice (Duke University Press).
17. Springer, K.W., Mager Stellman, J., and Jordan-Young, R.M. (2012). Beyond a catalogue of differences: A theoretical frame and good practice guidelines for researching sex/gender in human health. Soc. Sci. Med. *74*, 1817–1824.
18. Hyde, J.S., Bigler, R.S., Joel, D., Tate, C.C., and Van Anders, S.M. (2019). The Future of Sex and Gender in Psychology: Five Challenges to the Gender Binary. Am. Psychol. *74*, 171–193.
19. van Anders, S.M., Steiger, J., and Goldey, K.L. (2015). Effects of gendered behavior on testosterone in women and men. Proc. Natl. Acad. Sci. USA *112*, 13805–13810.
20. Sudai, M., Borsa, A., Ichikawa, K., Shattuck-Heidorn, H., Zhao, H., and Richardson, S.S. (2022). Law, policy, biology, and sex: Critical issues for researchers. Science *376*, 802–804.
21. Maney, D.L. (2015). Just Like a Circus: The Public Consumption of Sex Differences. In Ethical Issues in Behavioral Neuroscience, G. Lee, J. Illes, and F. Ohl, eds., pp. 279–296.
22. Saguy, T., Reifen-Tagar, M., and Joel, D. (2021). The gender-binary cycle: the perpetual relations between a biological-essentialist view of gender, gender ideology, and gender-labelling and sorting. Philos. Trans. R. Soc. Lond. B Biol. Sci. *376*, 20200141.
23. Maney, D.L. (2016). Perils and pitfalls of reporting sex differences. Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20150119.
24. Ritz, S.A., and Greaves, L. (2022). Transcending the Male–Female Binary in Biomedical Research: Constellations, Heterogeneity, and Mechanism When Considering Sex and Gender. Int. J. Environ. Res. Public Health *19*, 4083.
25. Beavis, A.L., Gravitt, P.E., and Rositch, A.F. (2017). Hysterectomy-corrected cervical cancer mortality rates reveal a larger racial disparity in the United States. Cancer *123*, 1044–1050.
26. Bauer, G.R. (2023). Sex and Gender Multidimensionality in Epidemiologic Research. Am. J. Epidemiol. *192*, 122–132.
27. Yang, C., Jin, J., Yang, Y., Sun, H., Wu, L., Shen, M., Hong, X., Li, W., Lu, L., Cao, D., et al. (2022). Androgen receptor-mediated CD8+ T cell stemness programs drive sex differences in antitumor immunity. Immunity *55*, 1268–1283.e9.
28. Bongen, E., Lucian, H., Khatri, A., Fragiadakis, G.K., Bjornson, Z.B., Nolan, G.P., Utz, P.J., and Khatri, P. (2019). Sex Differences in the Blood Transcriptome Identify Robust Changes in Immune Cell Proportions with Aging and Influenza Infection. Cell Rep. *29*, 1961–1973.e4.
29. Massa, M.G., Aghi, K., and Hill, M.J. (2023). Deconstructing sex: Strategies for undoing binary thinking in neuroendocrinology and behavior. Horm. Behav. *156*, 105441.
30. Williams, J.S., Fattori, M.R., Honeyborne, I.R., and Ritz, S.A. (2023). Considering hormones as sex- and gender-related factors in biomedical research: challenging false dichotomies and embracing complexity. Horm. Behav. *156*, 105442.
31. Hostinar, C.E., Sullivan, R.M., and Gunnar, M.R. (2014). Psychobiological mechanisms underlying the social buffering of the hypothalamic-pituitary-adrenocortical axis: a review of animal models and human studies across development. Psychol. Bull. *140*, 256–282.
32. Klein, S.L., and Nelson, R.J. (1999). Social interactions unmask sex differences in humoral immunity in voles. Anim. Behav. *57*, 603–610.
33. Smiley, K.O., Munley, K.M., Aghi, K., Lipshutz, S.E., Patton, T.M., Pradhan, D.S., Solomon-Lane, T.K., and Sun, S.E.D. (2024). Sex diversity in the 21st century: Concepts, frameworks, and approaches for the future of neuroendocrinology. Horm. Behav. *157*, 105445.
34. Rich-Edwards, J.W., and Maney, D.L. (2023). Best practices to promote rigor and reproducibility in the era of sex-inclusive research. eLife *12*, e90623.
35. Naqvi, S., Godfrey, A.K., Hughes, J.F., Goodheart, M.L., Mitchell, R.N., and Page, D.C. (2019). Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. Science *365*, eaaw7317.
36. Fischer, K.E., and Riddle, N.C. (2018). Sex Differences in Aging: Genomic Instability. J. Gerontol. A Biol. Sci. Med. Sci. *73*, 166–174.
37. Elderman, M., Hugenholtz, F., Belzer, C., Boekschoten, M., van Beek, A., de Haan, B., Savelkoul, H., de Vos, P., and Faas, M. (2018). Sex and strain dependent differences in mucosal immunology and microbiota composition in mice. Biol. Sex Differ. *9*, 26.
38. Frederiksen, H., Johannsen, T.H., Andersen, S.E., Albrethsen, J., Landersoe, S.K., Petersen, J.H., Andersen, A.N., Vestergaard, E.T., Schorring, M.E., Linneberg, A., et al. (2020). Sex-specific Estrogen Levels and Reference Intervals from Infancy to Late Adulthood Determined by LC-MS/MS. J. Clin. Endocrinol. Metab. *105*, 754–768.
39. Stanton, S.J., and Schultheiss, O.C. (2007). Basal and dynamic relationships between implicit power motivation and estradiol in women. Horm. Behav. *52*, 571–580.
40. Eliot, L., Ahmed, A., Khan, H., and Patel, J. (2021). Dump the "dimorphism": Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. Neurosci. Biobehav. Rev. *125*, 667–697.
41. Smith, G.T. (2013). Evolution and hormonal regulation of sex differences in the electrocommunication behavior of ghost knifefishes (Apteronotidae). J. Exp. Biol. *216*, 2421–2433.
42. Ball, G.F. (2016). Species variation in the degree of sex differences in brain and behaviour related to birdsong: adaptations and constraints. Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20150117.
43. Dumais, K.M., and Veenema, A.H. (2016). Vasopressin and oxytocin receptor systems in the brain: Sex differences and sex-specific regulation of social behavior. Front. Neuroendocrinol. *40*, 1–23.
44. Kokras, N., and Dalla, C. (2014). Sex differences in animal models of psychiatric disorders. Br. J. Pharmacol. *171*, 4595–4619.
45. Becker, J.B., McClellan, M.L., and Reed, B.G. (2017). Sex differences, gender and addiction. J. Neurosci. Res. *95*, 136–147.
46. Ma, L., Xu, Y., Wang, G., and Li, R. (2019). What do we know about sex differences in depression: A review of animal models and potential mechanisms. Prog. Neuropsychopharmacol. Biol. Psychiatry *89*, 48–56.
47. Shansky, R.M. (2015). Sex differences in PTSD resilience and susceptibility: Challenges for animal models of fear learning. Neurobiol. Stress *1*, 60–65.
48. Morrison, K.E. (2020). Animal models built for women's brain health: Progress and potential. Front. Neuroendocrinol. *59*, 100872.
49. Office of Research on Women's Health (ORWH). Sex as a Biological Variable: A Primer. https://orwh.od.nih.gov/career-development-education/e-learning/sabv-primer.
50. Wizemann, X., and Pardue, X. (2001). Does Sex Matter? (Institute of Medicine).
51. Ritz, S.A. (2017). Complexities of Addressing Sex in Cell Culture Research. Signs: J. Women Cult. Soc. *42*, 307–327.
52. Kalueff, A.V., Minasyan, A., Keisala, T., Shah, Z.H., and Tuohimaa, P. (2006). Hair barbering in mice: Implications for neurobehavioural research. Behav. Processes *71*, 8–15.

53. Prendergast, B.J., Onishi, K.G., and Zucker, I. (2014). Female mice liberated for inclusion in neuroscience and biomedical research. Neurosci. Biobehav. Rev. *40*, 1–5.

54. Weber, E.M., Zidar, J., Ewaldsson, B., Askevik, K., Udén, E., Svensk, E., and Törnqvist, E. (2022). Aggression in Group-Housed Male Mice: A Systematic Review. Animals (Basel) *13*, 143.

55. Joel, D., and McCarthy, M.M. (2017). Incorporating Sex As a Biological Variable in Neuropsychiatric Research: Where Are We Now and Where Should We Be? Neuropsychopharmacology *42*, 379–385.

56. Mauvais-Jarvis, F., Arnold, A.P., and Reue, K. (2017). A Guide for the Design of Pre-clinical Studies on Sex Differences in Metabolism. Cell Metab. *25*, 1216–1230.

57. Oudshoorn, N. (1994). Beyond the Natural Body: an Archaeology of Sex Hormones (Routledge).

58. Fujimura, J.H. (2006). Sex Genes: A Critical Sociomaterial Approach to the Politics and Molecular Genetics of Sex Determination. Signs *32*, 49–82.

59. Richardson, S. (2013). Sex Itself: The Search for Male and Female in the Human Genome (University of Chicago Press).

60. Marsh, D.M., and Hanlon, T.J. (2007). Seeing What We Want to See: Confirmation Bias in Animal Behavior Research. Ethology *113*, 1089–1098.

61. Marsh, D.M., and Hanlon, T.J. (2004). Observer gender and observation bias in animal behaviour research: experimental tests with red-backed salamanders. Anim. Behav. *68*, 1425–1433.

62. Office of Research on Women's Health (ORWH) (2018). Infographic: how sex/gender influence health & disease. https://orwh.od.nih.gov/sites/orwh/files/docs/SexGenderInfographic_11x17_508.pdf.

63. Aragonès, E., Piñol, J.L., and Labad, A. (2006). The overdiagnosis of depression in non-depressed patients in primary care. Fam. Pract. *23*, 363–368.

64. Wolfe, F., Walitt, B., Perrot, S., Rasker, J.J., and Häuser, W. (2018). Fibromyalgia diagnosis and biased assessment: Sex, prevalence and bias. PLoS One *13*, e0203755.

65. Wallis, K.E., Guthrie, W., Bennett, A.E., Gerdes, M., Levy, S.E., Mandell, D.S., and Miller, J.S. (2020). Adherence to screening and referral guidelines for autism spectrum disorder in toddlers in pediatric primary care. PLoS One *15*, e0232335.

66. Tajfel, H. (1969). Cognitive Aspects of Prejudice. J. Soc. Issues *25*, 79–97.

67. Kerr, N.L. (1998). HARKing: Hypothesizing After the Results are Known. Pers. Soc. Psychol. Rev. *2*, 196–217.

68. Maney, D.L., Karkazis, K., and Hagen, K.B.S. (2023). Considering Sex as a Variable at a Research University: Knowledge, Attitudes, and Practices. J. Womens. Health (Larchmt) *32*, 843–851.

69. Rechlin, R.K., Splinter, T.F.L., Hodges, T.E., Albert, A.Y., and Galea, L.A.M. (2022). An analysis of neuroscience and psychiatry papers published from 2009 and 2019 outlines opportunities for increasing discovery of sex differences. Nat. Commun. *13*, 2137.

70. Allison, D.B., Brown, A.W., George, B.J., and Kaiser, K.A. (2016). Reproducibility: A tragedy of errors. Nature *530*, 27–29.

71. Makin, T.R., and Orban de Xivry, J.J. (2019). Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. eLife *8*, e48175.

72. Sainani, K. (2010). Misleading comparisons: the fallacy of comparing statistical significance. PM R *2*, 559–562.

73. Nieuwenhuis, S., Forstmann, B.U., and Wagenmakers, E.J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. Nat. Neurosci. *14*, 1105–1107.

74. Canadian Institutes of Health Research (CIHR) (2023). Online Training Modules: Integrating Sex & Gender in Health Research. https://cihr-irsc.gc.ca/e/49347.html.

75. LIBRA. The Sex and Gender Dimension in Biomedical Research. Online training. https://www.libra-sgr.eu/libra/.

76. Clayton, J.A. (2018). Applying the new SABV (sex as a biological variable) policy to research and clinical care. Physiol. Behav. *187*, 2–5.

77. Tannenbaum, C., Ellis, R.P., Eyssel, F., Zou, J., and Schiebinger, L. (2019). Sex and gender analysis improves science and engineering. Nature *575*, 137–146.

78. Shansky, R.M., and Murphy, A.Z. (2021). Considering sex as a biological variable will require a global shift in science culture. Nat. Neurosci. *24*, 457–464.

79. McLaughlin, J.F., Brock, K.M., Gates, I., Pethkar, A., Piattoni, M., Rossi, A., and Lipshutz, S.E. (2023). Multivariate Models of Animal Sex: Breaking Binaries Leads to a Better Understanding of Ecology and Evolution. Integr. Comp. Biol. *63*, 891–906.

80. Zhao, H., DiMarco, M., Ichikawa, K., Boulicault, M., Perret, M., Jillson, K., Fair, A., DeJesus, K., and Richardson, S.S. (2023). Making a 'sex-difference fact': Ambien dosing at the interface of policy, regulation, women's health, and biology. Soc. Stud. Sci. *53*, 475–494.

81. Karkazis, K. (2019). The misuses of "biological sex.". Lancet *394*, 1898–1899.

82. WomenHeart, Society for Women's Health Research (2011). 10Q Report: Advancing Women's Heart Health through Improved Research, Diagnosis and Treatment. https://swhr.org/swhr_resource/10q-report-advancing-womens-heart-health-through-improved-research-diagnosis-and-treatment/.

83. Rubini Gimenez, M., Reiter, M., Twerenbold, R., Reichlin, T., Wildi, K., Haaf, P., Wicki, K., Zellweger, C., Hoeller, R., Moehring, B., et al. (2014). Sex-specific chest pain characteristics in the early diagnosis of acute myocardial infarction. JAMA Intern. Med. *174*, 241–249.

84. Pilote, L. (2014). Chest pain in acute myocardial infarction: are men from Mars and women from Venus? JAMA Intern. Med. *174*, 249–250.

85. Greenblatt, D.J., Harmatz, J.S., and Roth, T. (2019). Zolpidem and Gender: Are Women Really At Risk? J. Clin. Psychopharmacol. *39*, 189–199.

86. McCarthy, M.M., Arnold, A.P., Ball, G.F., Blaustein, J.D., and De Vries, G.J. (2012). Sex differences in the brain: the not so inconvenient truth. J. Neurosci. *32*, 2241–2247.

87. Eliot, L., Beery, A.K., Jacobs, E.G., LeBlanc, H.F., Maney, D.L., and McCarthy, M.M. (2023). Why and How to Account for Sex and Gender in Brain and Behavioral Research. J. Neurosci. *43*, 6344–6356.

88. Sayad, A., Ghafouri-Fard, S., Shams, B., Arsang-Jang, S., Gholami, L., and Taheri, M. (2020). Sex-specific up-regulation of p50-associated COX-2 extragenic RNA (PACER) lncRNA in periodontitis. Heliyon *6*, e03897.

89. National Academies of Sciences, Engineering, and Medicine (2022). Measuring Sex, Gender Identity, And Sexual Orientation (The National Academies Press).

90. Laurie, C.C., Doheny, K.F., Mirel, D.B., Pugh, E.W., Bierut, L.J., Bhangale, T., Boehm, F., Caporaso, N.E., Cornelis, M.C., Edenberg, H.J., et al. (2010). Quality control and quality assurance in genotypic data for genome-wide association studies. Genet. Epidemiol. *34*, 591–602.

91. Danielsen, A.C., Lee, K.M., Boulicault, M., Rushovich, T., Gompers, A., Tarrant, A., Reiches, M., Shattuck-Heidorn, H., Miratrix, L.W., and Richardson, S.S. (2022). Sex disparities in COVID-19 outcomes in the United States: Quantifying and contextualizing variation. Soc. Sci. Med. *294*, 114716.

92. Fausto-Sterling, A. (2005). The bare bones of sex: Part 1—sex and gender. Signs: J. Women Cult. Soc. *30*, 1491–1527.

93. Krieger, N. (2003). Genders, sexes, and health: what are the connections—and why does it matter? Int. J. Epidemiol. *32*, 652–657.

94. Homan, P. (2019). Structural Sexism and Health in the United States: A New Perspective on Health Inequality and the Gender System. Am. Sociol. Rev. *84*, 486–516.