

Why and How to Account for Sex and Gender in Brain and Behavioral Research

Lise Eliot,^{1*} Annaliese K. Beery,² Emily G. Jacobs,³ Hannah LeBlanc,⁴ Donna L. Maney,⁵ Margaret M. McCarthy⁶

¹Stanson Toshok Center for Brain Function and Repair
Chicago Medical School
Rosalind Franklin University of Medicine & Science
North Chicago, IL 60064

²Department of Integrative Biology
University of California, Berkeley
Berkeley, CA 94720

³Department of Psychological & Brain Sciences
University of California, Santa Barbara
Santa Barbara, CA 93106

⁴Division of the Humanities & Social Sciences
California Institute of Technology
Pasadena, CA 91125

⁵Department of Psychology
Emory University
Atlanta, GA 30322

⁶Department of Pharmacology
University of Maryland School of Medicine
Baltimore, MD 21201

*Address correspondence to: lise.eliot@rosalindfranklin.edu (847.612.3496)

ABSTRACT

Long overlooked in neuroscience research, sex and gender are increasingly included as key variables potentially impacting all levels of neurobehavioral analysis. Still, many neuroscientists do not understand the difference between the terms “sex” and “gender,” the complexity and nuance of each, or how to best include them as variables in research designs. This TechSights article outlines rationales for considering the influence of sex and gender across taxa, and provides technical guidance for strengthening the rigor and reproducibility of such analyses. This guidance includes the use of appropriate statistical methods for comparing groups as well as controls for key covariates of sex (e.g., total intracranial volume) and gender (e.g., income, caregiver stress, bias). We also recommend approaches for interpreting and communicating sex and gender-related findings about the brain, which have often been misconstrued by neuroscientists and the lay public alike.

1.0. INTRODUCTION: WHY ACCOUNT FOR SEX AND GENDER?

Sex and gender differences in the brain and behavior are of great interest to society and have the potential to impact the diagnosis and treatment of neuropsychiatric and neurological disorders. But there remains considerable misunderstanding about the difference in meaning between the two terms, the complexity that arises by treating each as a research variable, and thus, uncertainty about how to detect and interpret sex and gender influences when conducting neuroscientific research.

According to the NIH Office of Research on Women's Health (U.S. Dept. Health & Human Services, n.d.), "sex" is defined as:

A multidimensional construct based on a cluster of anatomical and physiological traits, that include external genitalia, secondary sex characteristics, gonads, chromosomes, and hormones.

And "gender" is defined as:

A social and cultural variable that encompasses several domains, each of which influences health: gender identity and expression, gender roles and norms, gender relations, structural sexism, power, and equality and equity.

Note that neither variable is strictly binary and each is composed of multiple dimensions that may or may not "match" within a single individual. Note also that the two constructs are inextricably linked with each other such that it is almost never possible, particularly in studies on humans, to experimentally separate them. For that reason, in this article we will sometimes refer to them together as 'sex/gender' to acknowledge their entanglement.

Many human psychological traits show gender differences, albeit most are minor (Hyde, 2005; Zell et al., 2015). Each of these could be shaped by sex-related factors (e.g., presence or absence of the Y chromosome in mammals), gender-related factors (e.g., occupational experience), or both. This interplay is even more critical when neuroscientists seek to explain gender disparities in mental health and neurological disorders: diagnoses such as depression, anxiety, eating disorders, and dementia are at least 50% more common in women, whereas ADHD, substance use disorders, dyslexia, and autism spectrum disorders are at least twice as common in men (Figure 1).

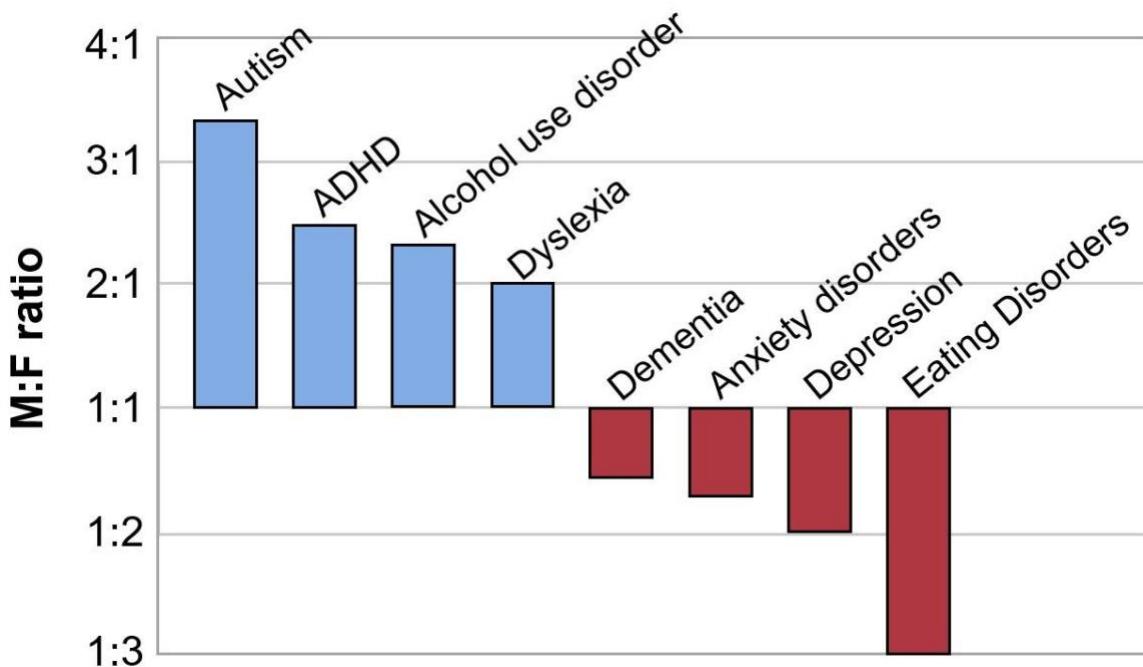


Figure 1: Gender prevalence ratios for common neuropsychiatric disorders. Modified from Eliot et al. (2021). SOURCES: Autism (Loomes et al., 2017); ADHD (Polanczyk et al., 2007); alcohol use disorder (Grant et al., 2004); dyslexia (Rutter et al., 2004); dementia (Buckley et al., 2019); anxiety disorders (McLean et al., 2011); depression (Salk et al., 2017); eating disorders (Hudson et al., 2007).

Motivated in part by such disparities, many health research funding agencies have begun mandating the consideration of sex and gender in biomedical research. In the U.S., this process began with the 1993 NIH Revitalization Act which required that all clinical trials funded through the agency include women and minority groups when clinically relevant. In neuroscience, gender parity has largely been achieved in imaging studies of the human brain (Eliot et al., 2021) but women continue to be under-represented in research on certain disorders, including stroke (Carcel et al., 2021). A greater problem is the failure of researchers to report data disaggregated by sex, even when women and men are included in comparable numbers (Geller et al., 2018). This omission creates a lost opportunity, considering the large sample size of many human studies and short-changes systematic reviews and meta-analyses.

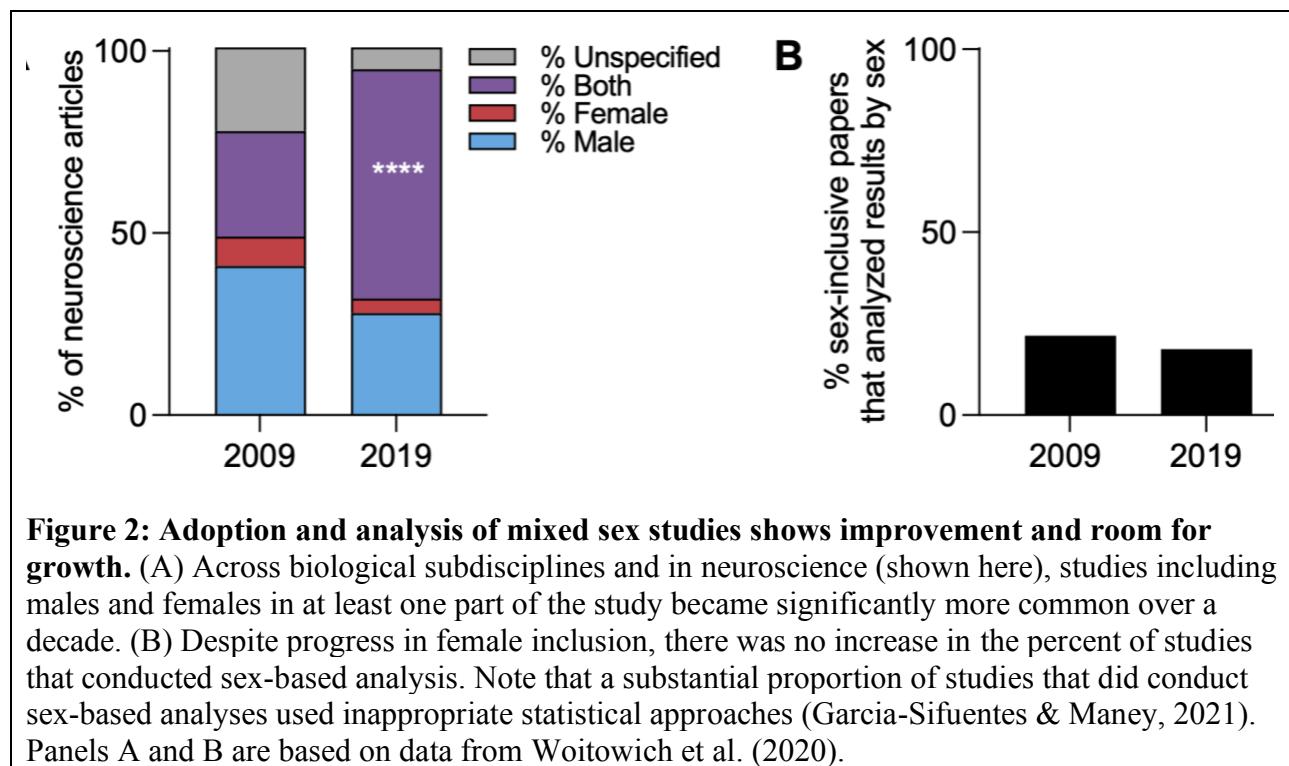
With regard to preclinical research involving animals and cells, it was not until 2016 that the U.S. National Institutes of Health mandated that sex be considered, through its “Sex as a Biological Variable” (SABV) policy. In many fields, including neuroscience, female non-human animals had been systematically excluded under the mistaken impression that the estrus cycle adds unacceptable variability (Beery and Zucker, 2011). However, meta-analyses of trait variability in male versus unstaged female mice (Prendergast et al., 2014) and rats (Becker et al., 2016) have demonstrated that while certain traits can be more variable in one sex or the other, females of these species are no more variable than males overall and are significantly less variable in spontaneous behavior, both within and across individuals (Levy et al., 2023). Similarly among humans, the coefficient of variation for cognitive, mental and physical health does not differ

between men and women during the reproductively active years (Smarr et al. 2021; Pritschet et al., 2022). Imaging studies of the brain demonstrate that structural volumes and surface area are modestly more variable in males (Forde et al., 2020; Wierenga et al., 2020). Thus, while pregnancy remains a concern in certain clinical trials (Blehar et al., 2013), in most cases there is no reason to exclude females from basic science or clinical research.

Single-sex studies still have an important place in certain domains of research, but for most biomedical topics, including neuroscience, there are ample reasons to include equal numbers of females and males. In addition to exploring the neural correlates of gendered behavioral differences and health disparities, inclusion of both sexes can uncover fundamental mechanisms of CNS function that may not be evident at the behavioral or even cellular level. Differences in underlying cellular and molecular mechanisms, known as “latent sex differences” (Jain et al., 2019) could potentially explain earlier inconsistencies in the scientific literature when sex was not specified, and may also lead to different treatments of diagnostic criteria for boys versus girls, or men versus women. Some more prominent examples of latent sex differences in animal studies include the cellular and molecular mediators of pain (Sorge et al., 2018), social attachment (De Vries 2004), reward (Becker and Koob, 2016), and hippocampal endocannabinoid signaling (Tabatadze et al., 2015).

By contrast, other sex effects can be divergent, where a specific context or environmental perturbation invokes a sex difference in behavior from an otherwise similar neurobehavioral profile. In rodents, for example, stress can have opposite effects on learning and memory in males and females (Shors et al., 2001) and behavioral strategies in response to foot shock can diverge between the sexes (Gruene et al., 2015). In such cases, studies that exclude one or the other sex provide a distorted view of such phenomena, and pooling male and female animals may even cancel out an effect of the treatment or exposure (Beery 2018). The only way to uncover such mechanisms that may differ by sex is to include sufficient numbers of male and female animals and to include sex as a factor in statistical analyses.

For U.S.-based researchers, implementation of the SABV policy and changing awareness of the importance of female inclusion have produced a substantial increase in the proportion of studies including both male and female animals. Over the decade from 2009-2019, studies using males and females increased across biological disciplines from 23% to 49% (Beery and Zucker 2011; Woitowich et al. 2020; Figure 2A). But while inclusion has improved, other aspects of reporting and statistical analysis remain problematic. Across 10 domains of preclinical biology using animal subjects, only 50% of mixed-sex studies in 2009 considered sex in their analysis; ten years later, this number *fell* to 42% of mixed-sex studies reporting sex-based analyses (Woitowich et al. 2020). In neuroscience, these numbers were 22% and 18% respectively (Figure 2B), and only a fraction of those used appropriate analyses to detect sex differences (Garcia-Sifuentes & Maney, 2021). An in-depth analysis of neuroscience and psychiatry papers over the same time interval replicated the finding that most studies did not use an optimal design for discovering sex differences (Rechlin et al. 2022). The authors documented frequent omission of sample size, imbalanced use of males and females, inappropriate statistical approaches, and use of males and females in some but not all parts of a study, revealing additional room for improvement. Thus, researchers would benefit from greater awareness of the context and components of sex-based analysis to fully achieve the aims of the SABV policy.



For studies of humans, the influence of environmental factors is multiplied many-fold due to the contribution of gender. All human societies are organized around gender roles and categorizations that can have profound effects on the brain and behavior throughout the lifespan. Thus, in addition to accounting for sex, human neuroscience research must take account of gender—that is, factors such as gender norms, power relations, economic security, and other inequalities that may favor men or women in different circumstances. Such approaches are now advocated by public funding agencies including the Canadian Institutes of Health and the European Commission (White et al., 2021). Gender analysis is recommended, though not yet required, by the U.S. National Institutes of Health ([U.S. Department of Health and Human Services, n.d.](#)).

Given the potential importance of sex and gender to brain and behavioral function, and the opportunities they provide to uncover novel neurobiological processes and foster new avenues for translational research, neuroscientists should invest greater effort in understanding and accounting for their effects. Doing so, however, requires more than simply including equal numbers of male and female subjects. In the remainder of this article, we will address the “how” of studying sex- and gender-related influences on the brain and behavior, including considerations of sample size and statistical comparisons, key covariates of both sex and gender, strategies for analyzing mechanisms underlying male/female difference, and how to operationalize “gender” in human studies. We end with a focus on how sex- and gender-related findings are interpreted and communicated to the non-scientific public, an often-fraught enterprise in need of greater thoughtfulness on the part of researchers (Eliot, 2011; Maney 2014; Rippon et al., 2021). In each main section, we conclude with specific recommendations to aid neuroscientists in their analyses of sex and/or gender effects..

1.1 Recommendations

- When possible, include equal and sufficient numbers of females and males in experiments.
- Analyze and report data both combined and separated by sex.
- Consider the range of explanations for sex and gender differences

2.0. UNDERSTANDING SEX EFFECTS

2.1. Operationalization of sex

Although we often think of sex as a single binary variable, in fact, it is a complex phenotype composed of many physiological elements that can change dramatically at different stages of development and in different environmental contexts. Thus, because sex is not a unitary variable, the first step in accounting for sex is to operationalize it using a stated measure (e.g. karyotype, anogenital distance, self-report) on which data are subsequently collected and published. The measure chosen to represent sex does not have to be the same across labs or even across studies within a lab; for reasons of practicality, sex is typically operationalized using different measures for cells in culture vs. mice ordered from a vendor vs. patients in a clinical trial. In any of these cases, the measure used should be transparently reported and consistent within a study (Richardson, 2022).

Once it is decided which variable will represent sex in a study, it is equally important to recognize variables that covary with that one; these additional variables have potential to confound the interpretation of results. In the sections below, we review some of the mechanisms that are typically included in the construct of sex, plus others not typically included that can nonetheless give rise to brain and bodily sex differences. Our goal in these sections is to raise researchers' awareness of proximal causes of sex differences as well as covariates of sex that could be addressed. Sex differences in the brain and behavior are never a purely direct effect of genes, gonads, or any other variable used to operationalize sex; rather, they involve intermediary steps. In rodents, for example, sex differences can result from differential transcription of genes, apoptosis of cells in specific brain regions, and differential responses of a parent to male and female offspring (McCarthy and Arnold, 2011). So it is typically these underlying, sex-related variables or processes that are more informative for elucidating the neural basis of male-female difference, along with phenotypic variance within sex. Sometimes, sex-related variables such as body size or longevity may provide simple explanations for apparent neural or clinical differences between males and females; in other situations, cellular and molecular processes may differ, opening up unique opportunities for advances in basic and clinical science.

2.2. Effects of genes

The determination of gonadal phenotype in mammals is controlled by a single gene on the Y-chromosome, the Sry gene (sex-determining region of the Y chromosome), which codes for a

transcription factor that initiates a gene expression cascade that drives the undifferentiated gonad towards a testis. If Sry is not present, the gonad develops into an ovary (Goodfellow and Lovell-Badge, 1993). The presence of testes and subsequent prenatal testosterone production is the key initiator of masculinization of the brain and behavior. For these reasons, the presence of the Y chromosome is a common sex-related factor used to operationalize sex. But are there genes on the X or Y chromosome that could directly differentiate the brain and/or impact sex differences in behavior? The advent of a genetically modified mouse in which the Sry gene is re-located to an autosome allowed for this question to be asked definitively, by permitting the comparison of wild-type animals to XX-males (i.e. with testis) and XY-females (i.e. with ovaries). Known as the 4-core-genotype (Arnold & Burgoine, 2004), this model is being employed to dissect out the relative influence of gonadal steroids hormones and X or Y chromosome genes. For example, certain features of aggressive and parental behaviors in mice, along with vasopressin immunoreactivity in the lateral septum, were found to be influenced by X or Y chromosomes, independent of gonadal sex (Gatewood et al., 2006). Similarly using this model, susceptibility to an experimental form of autoimmune encephalomyelitis was found to be greater in XX, compared to XY males (Smith-Bouvier et al., 2008), which may be pertinent to the gender difference in multiple sclerosis prevalence. Importantly, when an effect of X or Y complement is detected, it does not necessarily mean the effects are directly due to genes on one of those chromosomes, as both contain genes with broad epigenetic effects that can alter expression on the autosomes (Arnold and Chen, 2009).

Expression of genes throughout the genome can now be quantified using deep sequencing of the transcriptome of tissue samples or even single cells, revealing unprecedented levels of specification within what were once considered simple phenotypes, such as inhibitory versus excitatory neurons. The importance of transcriptional networks, as opposed to single genes, further expands the complexity of gene regulatory control and modifiability by hormones and other factors which may or may not be sex-specific (McCarthy 2020). Such approaches are uncovering highly-specific nodes of specialized cells involved in the control of reproductive and other social behaviors (Knoedler et al., 2022; Raam and Hong, 2021). From a disease perspective, transcriptomics has revealed vast and unpredicted patterns of sex-specific gene expression associated with the same diagnostic phenotype in males and females (Labonte et al., 2017; Massey et al. 2021). However, a word of caution is warranted when unbiased transcriptomic analysis is benchmarked against databases that themselves may be sex-biased. Most of the publicly available -omics resources do not annotate entries by sex. Given that the bulk of biomedical research over the last few decades has been conducted on male subjects, there is a strong potential for an embedded bias in databases such as Gene Ontology, KEGG or DAVID (Bond et al 2021). This bias could result in distorted conclusions by either the omission or underestimation of important sex differences.

2.3. Effects of hormones

The NIH definition includes hormones in a cluster of physiological traits that contribute to the construct of sex. The hormones likely to be most relevant in this context are those produced by the ovary and testis, such as testosterone and estradiol. In the past, these hormones have been

referred to as “sex steroids” or “gonadal steroids.” These terms are falling out of use, first because there is no hormone that itself defines a sex; testosterone and estradiol are produced in all female and male bodies. Further, tissues other than the gonads, such as the brain, can synthesize these steroids *de novo* (Azcoitia et al., 2018). Here, we will call these hormones “steroid hormones” with the caveat that this label can also refer to stress hormones, e.g. glucocorticoids. We recommend using the specific name of the hormone when possible and relevant, e.g. “testosterone,” or the more general terms “androgens” or “estrogens.”

Steroid hormones are powerful modulators of biological processes at all levels, often independently of other markers of an animal’s sex (McCarthy and Konkle, 2005). Nonetheless, because plasma levels of androgens and estrogens can differ dramatically between males and females, they are often the default hypothesized mechanism for findings of sex differences (Maney et al., 2023), particularly when the hypothesizing is done post-hoc. As noted below, however, steroid hormones are just one of many potential contributors to sex differences, and care should be taken to consider less obvious factors, some of which are discussed below in sections 2.4 – 2.6.

2.31. Prenatal and neonatal hormones

In rodents, many sex differences in the brain and subsequent behavior arise from differences in pre- or perinatal exposure to testosterone and its major metabolite, estradiol (Figure 3). In both rats and mice, the fetal testis begins copious androgen production during late gestation, 3-5 days prior to birth, and continuing through birth but dropping precipitously thereafter and staying low until puberty. In females, the ovaries remain quiescent, and there is far less exposure to steroids developmentally. Much of what we understand about steroid-mediated sexual differentiation of the brain comes from the fortuitous fact that newborn female rodents remain sensitive to the actions of testosterone; if it is administered exogenously, the process of masculinization is initiated and the endpoints that are normally achieved in males prenatally can now be triggered in females postnatally, making them an interesting model in which to test the effects of androgens across the lifespan (McCarthy et al., 2018). In humans and non-human primates, steroid-mediated brain sexual differentiation begins during the second trimester and this surge is largely complete by birth, greatly challenging the ability to interrogate mechanisms. Regardless, certain behavioral sex differences, along with the sensitivity to steroid hormones in adulthood, appear to be influenced by this prenatal exposure in primates (Breedlove 1994; Wallen & Baum 2002), making it an important period in brain sexual differentiation.

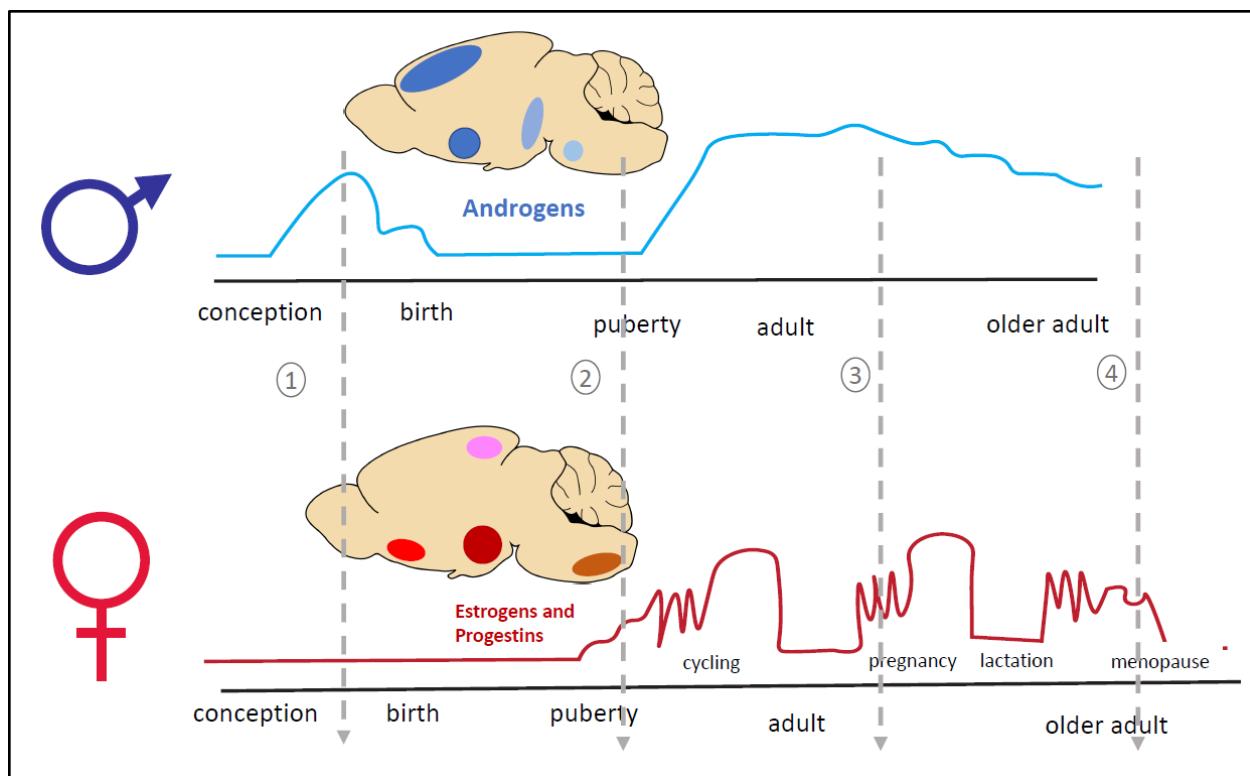


Figure 3: Unique endocrine exposures in male and female rodents over the lifespan.

Directly comparing male and female rodents at a given life stage must take into account underlying differences in past or current hormonal milieu and brain maturation. 1) Prior to birth, males are exposed to relatively high testosterone of their own making, whereas in females, brain exposure to testosterone, estrogens and progestins remain low during this time. 2) Females enter puberty sooner than males and 3) during the reproductive years, females may be cycling, pregnant, or lactating, all of which are physiological states distinct from the relatively stable and high levels of testosterone in males. 4) Later in life, females spend considerable time in a post-reproductive state with very low steroid hormone levels. Note that this figure illustrates general hormonal changes across the rodent lifespan, not precise levels, and omits the exposure of males to estrogens and progestins and females to androgens. Colored ovals in the brain represent that regional sex differences vary in magnitude and location over time. Observations of sex differences are appropriately framed when all of these variables are considered.

2.32. Rhythmic changes in gonadal hormone production

A central feature of the mammalian endocrine system is that hormone production varies over time. Many of these endogenous endocrine rhythms occur across sexes. For example, gonadotropin and steroid hormones exhibit circadian rhythms tied to the sleep-wake cycle in males and females. Testosterone and cortisol production peak in the A.M. and decline in the P.M., a circadian pattern that is common to males and females. In other instances, the frequency, duration, and intensity of these endocrine rhythms can differ by sex. For example, ovarian cycles

such as the 4–5 day estrous cycle in rodents and ~28-day menstrual cycle in humans produce cyclic changes in circulating estrogens and progesterone. During pregnancy, serum estrogens and progesterone concentrations increase ~100-fold and then decline precipitously at parturition. Each of these neuroendocrine transitions drive physiological changes throughout the body, in some cases including the brain. For example, in a series of foundational studies in rats, Woolley and colleagues (1992) identified changes in hippocampal spine density in CA1 neurons that are tied to stage of the estrous cycle, with a ~30% increase in spines during proestrus (the day of ovulation with estradiol peaks) relative to estrus 24h later. Similarly rapid effects of steroid hormones on hippocampal morphology and the functional connectome have been observed over the human menstrual cycle (Taylor et al., 2020; Pritschet et al., 2020; Pritschet et al., 2021). A recent longitudinal study by Zsido and colleagues (2023) used ultra high-field imaging to examine the brain at six time points across the human menstrual cycle. High estradiol/low progesterone conditions were associated with volumetric increases in hippocampal CA1 in women, mirroring the findings of Woolley et al (1992) albeit in a different species and different spatial scale. During pregnancy, organizational effects of steroid hormones are likely responsible for brain morphological changes that have been observed between pre- and post-partum phases in women (Hoeckzema et al., 2017). Neuroanatomical adaptations are also evident in new fathers during the transition to parenthood (Martinez-Garcia et al., 2023), though their endocrine drivers and time-course likely differ from the neuronal changes observed in mothers.

2.4. Effects of age

Depending on species, males and females may develop, mature, and age at different rates. In humans, the sex difference in longevity is one of the more robust findings and is not seen in many other mammals (Austad & Fischer, 2016). Although its basis is not well-understood, this longevity gap in humans has a major impact on neurodegenerative disorders, notably Alzheimer's disease, where it largely explains the nearly double number of women as men (Figure 1) living with the disease (Fiest et al., 2016; Buckley et al. 2019). Maturational sex differences are also important earlier in life, as girls undergo puberty and reach their peak brain volume, gray matter density, and cortical surface area 1–2 years earlier than boys over adolescence (Kaczkurkin et al., 2019). Such asynchrony is pertinent when comparing adolescent brain or behavioral measures between sexes, since what looks like a sex difference at a particular age may be better explained by a difference in pubertal timing. Conversely, when comparing boys and girls at similar pubertal stages, their mismatch in age may explain any brain or behavioral differences better than differences in steroid hormone levels (Vijayakumar et al., 2018).

Indeed, age is often an important variable when considering the effects of hormones. In rodents, marked differences in hormonal profiles of males and females occur across the lifespan, beginning prior to birth, diverging at puberty, and then again after reproductive senescence or menopause in females. Moreover, females experience unique hormonal fluctuations (e.g., the estrus cycle, pregnancy), in addition to the diurnal changes in hormone production experienced by both males and females. This variation presents a challenge when one is directly comparing males and females as they may differ by both maturational status and hormonal profile (Figure 3). There is no simple answer to this conundrum, but awareness of what is actually being compared is essential. For example, an insult to the nervous system in a 35-days-old mouse or rat will coincide

with a pre-pubertal window in males but a post-pubertal window in females, due to differences in the timing of the onset of puberty. Likewise, aged males are likely to have circulating testosterone similar to that of a younger adult, whereas aged females will have minimal circulating estrogens or progestins.

In humans, reproductive aging (defined as changes in steroid hormone production that occur with age) is a major contributor to the physiological changes that take place during middle age in both men and women. For men, age-related changes in sex hormone production are linear and protracted, gradually declining beginning in the mid-30s. In contrast, women undergo a more complex transition at mid-life, marked by high fluctuations in steroid hormone production during the perimenopausal phase, and culminating in ovarian senescence at the end of menopause—with the median age of complete reproductive senescence at 52.4 years in the U.S. (Gold et al., 2013). The menopausal transition results in a substantial decline in ovarian hormone production—up to 90% for both estradiol and progesterone. Human studies of the aging brain should therefore distinguish between the effects of chronological versus endocrine aging, though many do not. One approach is to capitalize on variation in the timing of the menopausal transition, matching participants on chronological age while allowing reproductive stage to vary (Jacobs et al., 2016). Another approach is to study the effects of pharmacological manipulations that temporarily and reversibly induce a menopause-like state (Berman et al., 1997).

2.5. Effects of body size

In many of the commonly studied mammals, males have larger bodies than females. While there are many sex differences that size cannot account for, body weight makes an important contribution to many traits (Wilson et al. 2022). This size difference is indeed reflected in larger volumes of males' brains, with differences ranging from 2.5% larger male brain volume in C57BL/6J mice (Spring et al., 2007) to 10% larger in Fischer 344 rats (Gorezen et al., 2020) and 11% larger in humans (Williams et al., 2021). For studies exploring regional brain differences between males and females, it is therefore necessary to control for total brain volume (TBV) or intracranial volume (ICV), which largely eliminates regional volume differences in humans (Jäncke et al., 2015; Ritchie et al., 2018). The small (1-2%) volumetric sex differences that remain after controlling for TBV or ICV have proven quite variable and dependent on image processing pipeline and the method of size normalization (Eliot et al., 2021). Absolute brain size also affects other features of brain architecture, such as the ratio of white:gray matter and the ratio of inter-hemispheric to intra-hemispheric connections. Larger brains need more or thicker white matter pathways to interconnect more distant regions, and inter-hemispheric traffic grows inefficient with increases in brain size (Zhang and Sejnowski, 2000). Hence, men's brains have about 6% higher white:gray matter ratio (Pintzka et al., 2015; Ritchie et al., 2018) and a lower ratio of inter- to intra-hemispheric connections (Ingalhalikar et al., 2014) than women's brains, both of which are attributable to brain size, not sex per se (Lewis et al., 2009; Hänggi et al., 2014). In other words, such measures differ between large- and small-headed men as much as between men and women, so are unlikely relevant to behavioral or clinical gender differences. These findings highlight the importance of controlling for individual ICV or TBV when conducting any study of individual brain structural difference, gender-related or otherwise (Jäncke et al., 2015), using

correction factors that may differ for different brain compartments (Jäncke et al., 2019; Williams et al., 2021).

2.6. Effects of environment

Still other important, but minimally studied contributors to brain sex differences are social and environmental factors. Although the term “gender” is usually restricted to humans, male and female animals often inhabit different social spaces and are faced with different experiences that can alter their brains and behavior. Such differential experience can begin even before birth; for example, females’ exposure to a male co-twin or adjacent male littermate may shift certain features of physiology and behavior in a male-typical direction, presumably as a result of prenatal testosterone exposure (Ryan and Vandenberg, 2002; Tapp et al., 2011).

After birth, the opportunity for sex- or gender-differentiated experience is magnified many-fold. Although we tend to focus on such differential experiences primarily in humans, the social environment can be highly differentiating in other animals as well. For example, rat dams treat male and female pups differently from the first days of life, licking and grooming males more than females, a difference that acts to shape the neural circuit underlying male copulatory behavior (Moore, 1992). Since licking-and-grooming, acting through epigenetic mechanisms, is also known to influence stress regulation, exploratory behavior, and spatial learning (Champagne and Curley, 2009), such sex-specific rearing differences could contribute to brain and behavioral sex differences. In non-human primates, too, the complex social environment is known to foster sex differences in behavior; for example, wild chimpanzees fish for termites with flexible sticks, a learned behavior that females in some locales master earlier than males because the females spend more time than males watching their mothers “fish” in early life (Lonsdorf et al., 2004).

One environmental factor that has been widely studied is stress, which can exert different effects on male versus female brains and behavior (McCarthy, 2016). For example, Bohacek et al. (2015) found that a sex difference in mouse hippocampal gene expression is induced by certain stressors (cold water) but not others (restraint). Research in humans and non-human animals has now documented the impact of a variety of stressors, occurring at various times across the lifespan, on the emergence of brain and behavioral sex differences, especially in psychopathology (Bale and Epperson, 2015). Such findings illustrate the importance that environmental context can play in generating brain and behavioral sex differences, and may be especially impactful given the different stressors that are often exerted on male versus female bodies (Hyde and Mezulis, 2020).

2.7 Recommendations

- Clearly operationalize sex using a stated measure and collect data on that measure.
- Recognize that factors contributing to sex differences can include genes, hormones, body size, gendered environments, and stress.
- Because all of these factors can interact with each other and change over time, it is important to consider stage of development in any sex comparisons.

3.0. UNDERSTANDING GENDER EFFECTS

This brings us to a preeminent part of human experience that has thus far been largely ignored in our effort to understand male-female brain differences: the impact of gender. As defined above, gender is a complex psychosocial construct that can have profound effects on experiences and behavior and therefore, the brain. In some cases, gender acts independently of sex, and in others sex and gender interact over the lifespan (Hyde et al., 2019; Springer et al., 2012; Krieger, 2003).

Researchers outside of neuroscience have found that sex and gender can have independent effects on health and can do so even in cases where there is no obvious difference between women and men. For example, among younger patients diagnosed with Acute Coronary Syndrome (ACS), Pelletier et al. (2016) observed no difference in the risk of recurrence between women and men, but instead found that patients with a higher “femininity score” were more likely to experience a recurrence of ACS, regardless of their sex.

Operationalizing gender as separate from sex shows promise in brain imaging studies as well. In a recent review, Rauch and Eliot (2022) identified eight neuroimaging studies that have assessed brain parameters using gender as an independent variable. While still preliminary, this work provides evidence that gender attributes can explain some of the individual variance in regional brain structure and functional circuitry, apart from or in addition to binary sex effects.

Despite these encouraging efforts, research using gender as a variable of interest is still in its infancy, in part due to the inadequacy of existing instruments for its operationalization. Existing measures largely fail to capture gender’s nonbinary, multidimensional nature (Horstmann et al., 2022). Research that operationalizes gender as either one-dimensional (a bipolar continuum that ranges from “masculine” to “feminine”) or two-dimensional (with masculinity and femininity as two distinct traits) cannot identify the specific gendered factors that affect health or the brain (Nielsen et al., 2021; Horstmann et al., 2022). Moreover, such binary conceptualizations of gender often exclude the experiences of trans, non-binary, and gender-queer individuals (Hyde et al., 2019; Miani et al., 2021).

In addition to being nonbinary, gender operates on many scales: from individuals’ gender identities, to the structure of labor markets, to expectations around caregiving work, to experiences of discrimination (Bolte et al., 2021; Miani et al., 2021). While these different dimensions often correlate with one another, any single dimension (e.g., gender identity) may be a poor proxy for any other dimension (e.g., the number of hours spent caring for a family member). Gender remains a primary category of social organization, but any individual will exhibit a unique constellation of these various factors, which also intersect with other axes of social inequality in ways that can affect brain health.

To address these problems and create a modern tool for operationalizing gender, an international team recently developed a new instrument, called the Stanford Gender-related Variables for Health Research (GVHR; Nielsen et al., 2021). The GVHR conceptualizes gender across three domains: gender norms, gender-related traits, and gender relations. Gender norms are spoken and unspoken rules enforced through social institutions like the workplace, family, and broader culture. Gender-related traits refer to individuals’ behaviors and tendencies, which may or may not conform to their gender norms. Gender relations refer to the social relationships between

individuals, whose own gender identities and related traits operate within a system of gender norms. Gender norms, traits, and relations are culture- and context-specific and change over time.

The GVHR is a questionnaire of 25 items that measure 7 gender-related variables across these three domains. The variables are: caregiver strain and work strain (gender norms); independence, risk-taking, and emotional intelligence (gender-related traits), and social support and discrimination (gender relations). The questionnaire was validated in three online samples totaling over 4,000 diverse respondents.

The GVHR has several advantages for neuroscience researchers. First, the scales make no assumptions about whether traits are “masculine” or “feminine,” as such labels are often normative and imprecise. Second, each of the variables is scored separately, rather than being collapsed into a single gender score. This scoring system improves the precision of the measure, gives researchers the opportunity to select variables that are most relevant to their scientific question , and makes it appropriate for use with non-binary and trans populations. The GVHR does not replace the independent variable of gender identity (e.g., woman, man, genderqueer, genderfluid, etc.), but may be used in combination with it to better characterize participants within the gender space. Using the GVHR or similar instruments in combination with large-scale neuroimaging should enhance our understanding of the neural basis of gender-related behaviors and psychopathologies.

3.1 Recommendations

- When possible in human studies, operationalize gender as separate from sex.
- Appreciate the key components of gender beyond identity that may impact brain health and incorporate such measures as appropriate.

4.0. ANALYTIC APPROACHES

Whether one is studying animals or humans, accounting for sex or gender adds complexity to statistical analyses. In this section, we address issues of sample size and analytical approaches. In alignment with the existing SABV policy, we focus on analyses of binary group differences, recognizing that more complex statistical models are needed for analyses that use continuous measures of sex or gender, such as the GVHR, as the independent variable.

4.1. Design considerations

Mere inclusion of females and males is insufficient to provide insight into the role of sex/gender in trait variability. Having rejected concerns that females are more variable than males (Becker et al., 2016; Prendergast et al., 2014), we must still consider the possibility of differences between males and females and their potential impacts.

When sex-inclusive data are analyzed *without* sex as a factor, sex differences have the potential to increase the variability of the sample as a whole. Fortunately, inclusion of sex as a factor in analysis recovers statistical power in many instances, whether sex differences are present or absent (Figure 4). Thus, the “low hanging fruit” is to analyze results of studies on males and females with sex (formally operationalized) as a factor in the analysis, and to appropriately

balance sample sizes. Of note, some studies treat sex/gender as a nuisance variable in their models, accounting or “controlling” for it without noting possible effects (Beltz et al. 2019). Whether or not sex contributes to the final results, the presence or absence of sex differences and interactions between sex and treatment should be reported, descriptive data should be disaggregated and reported by sex, and consideration should be given to the statistical power of the data set to identify sex differences.

4.2. Exploratory vs. confirmatory studies

In their guidance on how to incorporate SABV into research designs and analysis, NIH distinguishes between studies “intended” and “not intended” to detect sex differences (NIH, 2015). In the former, testing for sex differences is an *a priori*, stated goal of the study and included among its hypotheses. This sort of research, which we will call “confirmatory” (as opposed to “exploratory,” see below), must be powered to test hypotheses about sex. Depending on the question, the sexes might simply be compared with each and no other factors, such as a drug treatment, need be considered. In most cases, however, authors may be testing the extent to which males and females respond differently to a treatment or exposure. For such “factorial” designs, one can test for effects of one factor (for example, treatment) and effects of another factor (for example, sex) in the same statistical model. These “main effects” can indicate whether the treatment had an effect overall, regardless of sex (Figure 4A) and whether there was a sex difference in the outcome measure regardless of treatment (Figure 4B). Note, however, that a main effect of sex does not indicate that females and males responded differently to the treatment. For example, in a study examining the effect of an intervention on stress hormone levels, a main effect of sex would indicate a significant sex difference in stress hormone levels, but does not provide evidence that males and females responded differently to the intervention.

For studies with factorial designs, testing whether males and females responded differently to a drug or intervention can be done by including an interaction term in the statistical model, for example Sex*Treatment (Figure 4C). A significant interaction would indicate that males and females did indeed respond differently to the treatment. Such interactions can require substantial power to detect. Thus, studies intended to reveal “sex-specific” effects of another factor typically require a larger sample size than is required to detect the factor’s effect in a single-sex study. Other authors have provided more detailed guidance on determining sample sizes when the goal is to detect a sex difference in the effect of another factor (e.g., Buch et al., 2019; Galea et al., 2020).

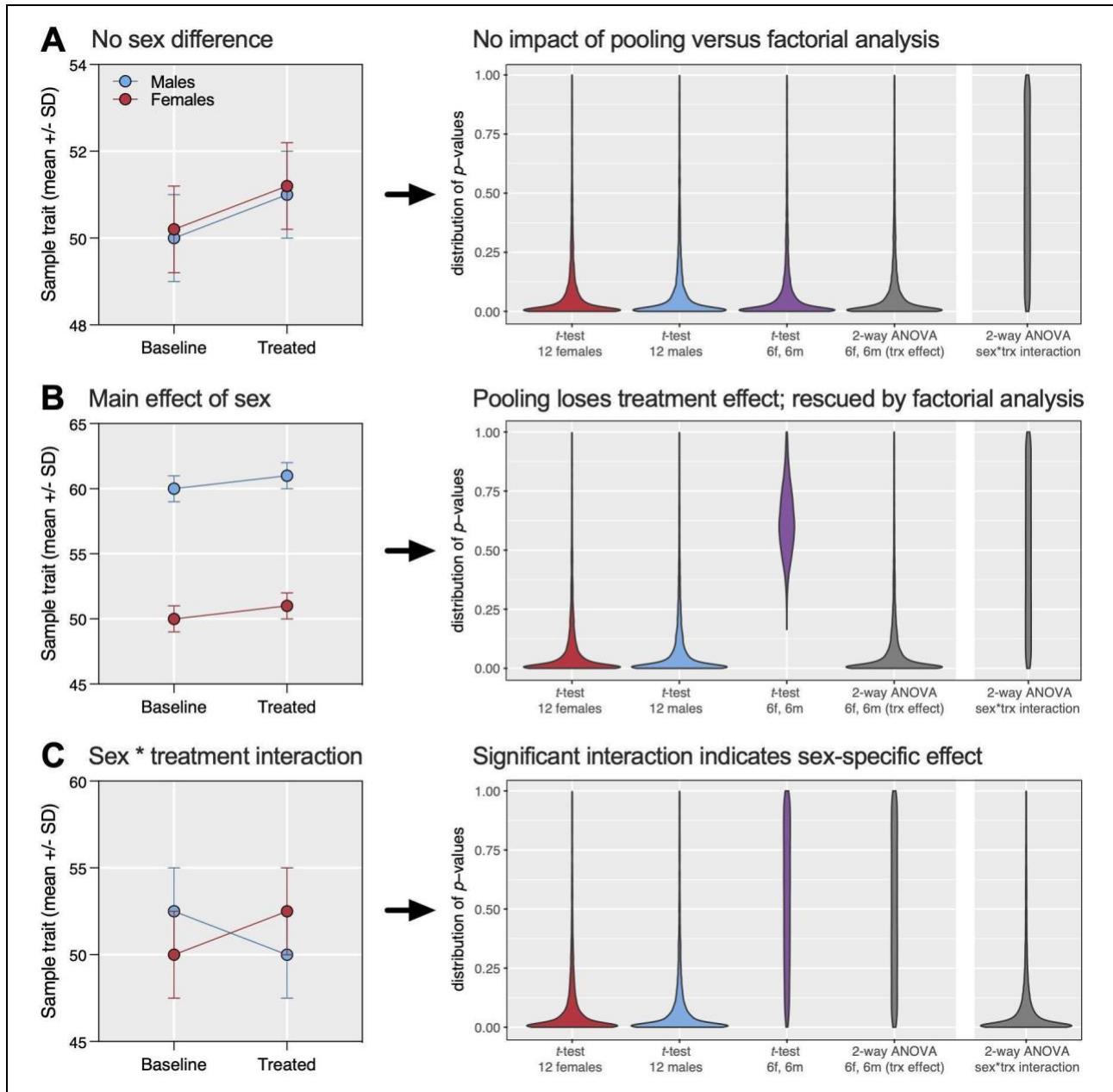


Figure 4. Example data scenarios and impact of analysis as single-sex samples (only females or only males tested), pooled mixed-sex samples, or factorial analysis of mixed-sex samples (2-way ANOVA assessing effects of sex, treatment, and their interaction). Left panels show mean and standard deviation of each group (12 females or males in single-sex studies, or 6 males and 6 females in mixed-sex groups). Effect sizes in each sex were matched (Cohen's $d=1.0$) and 10,000 simulated groups of each type were generated for comparisons. Right panels display the statistical consequences of different analysis paths as violin plots of p -values generated for each test across 10,000 samples. Distributions weighted closer to zero (e.g. red distributions) yielded “significant” treatment differences in more simulated comparisons. **Red:** effect of treatment across all-female comparisons. **Blue:** effect of treatment in all-male comparisons. **Purple:** effect of treatment when males and females are pooled and sample size is not increased. **Gray:** treatment and

sex*interaction effects in a 2-way ANOVA. (A) Pooling has no impact in the absence of sex differences, but (B) obscures the treatment effect in the presence of a main effect of sex, or (C) when there is an interaction between sex and treatment (i.e., the treatment affected the sexes differently). Factorial analysis rescues the treatment signal when there is a main effect of sex (B), and reveals a potential sex*treatment interaction (C). If the sample size is not increased for mixed-sex exploratory studies, a finding such as C would call for a future follow-up with a larger sample of males and females, and potential future studies of mechanisms in each sex. Redrawn from and based on simulations performed by Beery (2018).

Sometimes, researchers are not focused on sex differences—they are including males and females to increase the generalizability of their findings, or to make their work more inclusive, or simply to comply with the SABV policy. In these “exploratory” studies, per NIH guidance, researchers need not increase the sample size beyond that necessary to detect the main effects of the other variables of interest, such as exposure, time or treatment (NIH, 2020). Most main effects that can be detected with a certain number of men or males, for example, should theoretically be detectable using a similarly-sized sample of mixed sex or gender (Buch et al., 2019, NIH, 2020). Notably, NIH expects researchers to consider sex even in exploratory studies that are not powered to compare the sexes (Clayton, 2018; Clayton & Tannenbaum, 2016; Cornelison & Clayton, 2017; NIH 2015; 2020). However, the latter practice can potentially lead to inaccurate pronouncements of sex or gender differences (false positives) and conversely, real sex or gender differences may not be detectable with small sample sizes, leading to the reporting of false negatives. In the next sections, we discuss some common mistakes and how to avoid them.

4.3. How to test for a sex-specific effect

NIH recommends that researchers analyze and report data separately for males and females (Clayton, 2018; Clayton & Tannenbaum, 2016; Cornelison & Clayton, 2017; NIH, 2020). Many journals offer similar guidance (Haidari et al., 2016). Although this reporting is important to promote transparency and future meta-analyses, the analysis of disaggregated data can lead to a well-known analytical error. If two independent samples are analyzed separately and a significant effect is found in one but not the other (e.g., $p = 0.02$ vs. 0.20), such a result does not constitute evidence that the effect in one sample is reliably larger or smaller than in the other (Gelman & Stern, 2006; Makin et al., 2019; Nieuwenhuis et al., 2011; Sainani, 2010). The problem is that this approach does not test whether an effect in one group is stronger than in another. For example, to show that an effect is significantly different between two sexes, we must directly compare the two effects by testing for an interaction between the factor in question and sex (Bland et al., 2015; Gelman & Stern, 2006; George et al., 2016; Makin et al., 2019; Nieuwenhuis et al., 2011; Sainani, 2010; Vorland, 2021; Vorland et al., 2021).

Failing to test formally for sex-specific effects is so rampant an error that it has earned its own acronym: the Difference in Sex-Specific Significance, or DISS error (Maney & Rich-Edwards, 2023). DISS errors most often result in Type I errors (Bland et al., 2015), biasing researchers toward positive findings. But DISS approaches can also lead to Type II errors, or false negatives,

where sex-specific effects are missed (Vorland et al., 2023). Both types of errors obviously hamper reproducibility.

When assessing sex-specific effects, it is therefore critical to test for interactions between sex and other factors of interest (e.g., Figure 4, gray plots) and not simply compare the statistical significance of that effect within each sex. Testing for interactions will increase power to detect effects of the other factors (because all of the samples can be considered), while at the same time ensuring that exploratory testing for potential sex differences is done in a statistically valid, rigorous way. A rigorous approach is *especially* important in studies not powered to detect small or medium-sized interactions, since the false positive rate of the DISS approach can reach 50% in such cases (Allison et al., 2016). If the interaction with sex is significant, then one can follow up with post-hoc tests within each sex. If the interaction is not significant, however, the post-hoc tests are not warranted and should not be performed; authors should instead report that there is no evidence that the sexes responded differently, with the caveat that the study may be underpowered. In these cases it may be appropriate to call for follow-up studies to test for sex-specific effects of the treatment or exposure.

Note that, before statistically comparing the sexes, it is important to discern whether the sexes are actually in a comparable state. In Section 2.4, we pointed out that same-age males and females can be at very different stages of pubertal development or reproductive senescence. In such cases, when sex is completely confounded with another relevant variable, it may be appropriate to test hypotheses only within-sex rather than including males and females in the same statistical model. This approach does not compare the sexes with each other and thus no claim can be made about sex difference *per se*.

Regardless of the analytical approach, whenever mixed-sex studies are performed, it is imperative to make all raw data, disaggregated by sex, available for future investigations. This can be done either in a main paper or in its Supplemental Material, and will facilitate subsequent meta-analyses and help avoid “file drawer” omissions (Rosenthal, 1979).

4.4 Recommendations

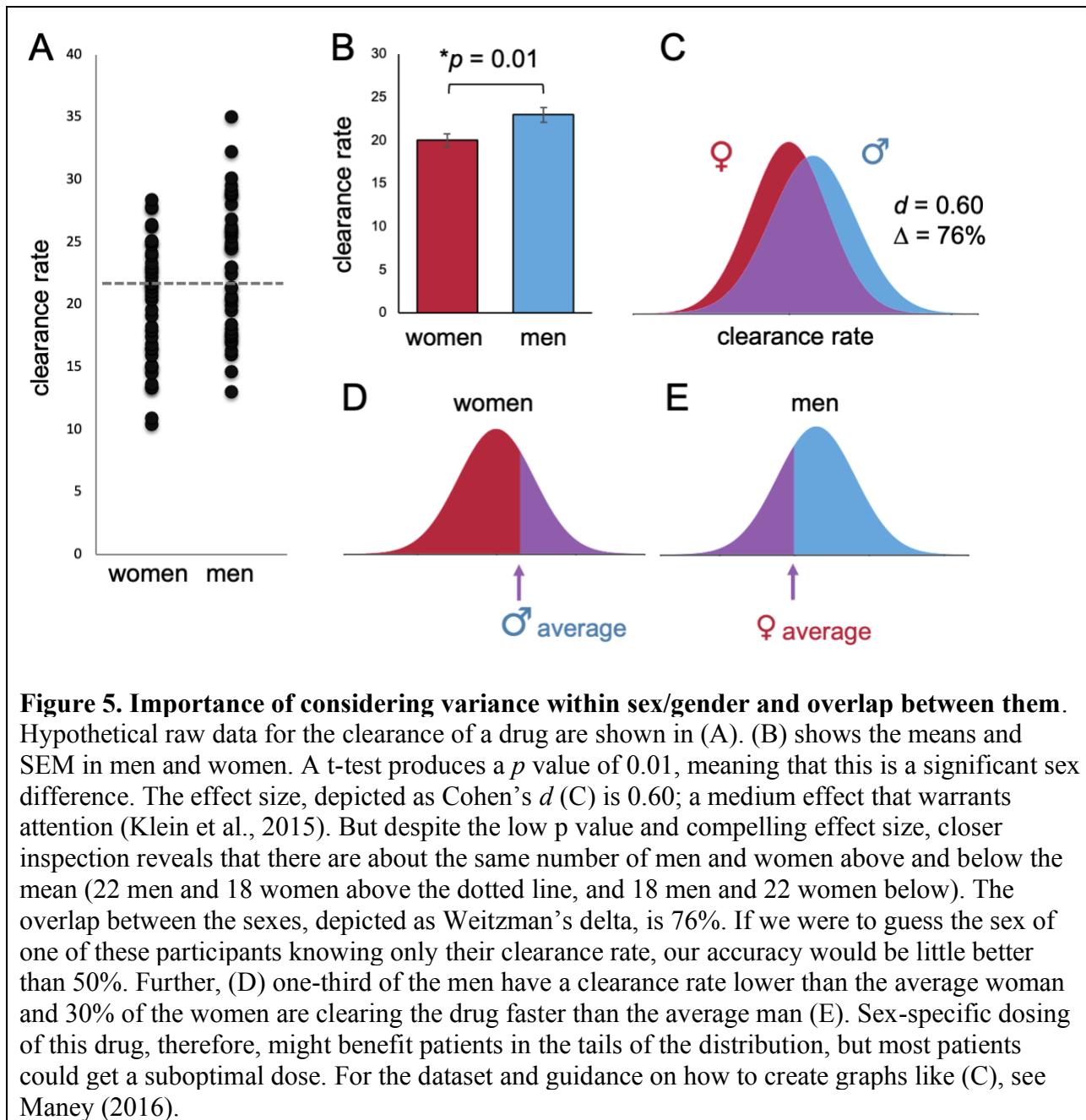
- Analyze results of studies on males and females with sex as an explicit factor in the analysis, and don’t “control for” sex.
- Balance sample sizes as appropriate.
- Report all statistics relevant to sex comparisons, including main effects and interactions.
- For exploratory, underpowered studies, acknowledge the limitations of the approach without relaxing statistical rigor.
- Regardless of the outcome of any sex comparison, publish the sex-disaggregated data for each sample either in the main paper or supplemental material.

5.0. INTERPRETATION AND COMMUNICATION

Because the framing of male/female brain differences has broad impact on issues related to gender equity, health disparities, and the rights of sexual and gender minorities, we end with guidance on the interpretation and communication of sex/gender differences. Neuroscientific

findings hold great appeal in the public imagination, particularly as they relate to differences between men and women or boys and girls. But too often, complex and unreplicated findings about male/female brain differences are simplified and over-hyped by popular media, propagating stereotypes and distorting the real science they are purporting to convey (Eliot, 2011; Maney, 2014; 2016; Rippon et al., 2021). When it comes to research on sex and gender difference, neuroscientists should therefore pay closer attention to: 1) negative findings, which often do not get mentioned in studies including both males and females; 2) the actual magnitude (effect size) of group-level sex/gender differences, as opposed to mere statistical significance; 3) the practical relevance of sex/gender difference findings to biomedical treatment or policy making; and 4) the multifactorial determinants (e.g., genetic, morphological, psychological, social) that may underlie sex/gender differences, before attempting to speculate on their clinical or behavioral relevance. Just as the National Academy of Sciences (2023) recently warned against “typological thinking” about race and ancestry in the dissemination of genetics and genomics research, sex/gender difference researchers should be cautious about interpreting group-level differences as fixed and attributable to “biological” processes, especially when the underlying mechanisms are largely unknown.

As a baseline, neuroscientists are advised to avoid the term “dimorphism” to describe statistical differences between male and female brains (McCarthy et al., 2012), which rarely satisfy the proper definition of “two forms.” The term “dimorphism” refers to two distinct types, or distinct structures present in only one, but not the other sex. To-date, dimorphic structures have not been found in the central nervous systems of vertebrates, with the possible exception of striatal Area X in some songbirds (Nottebohm and Arnold, 1976) or the bulbocavernosus spinal motor nucleus in some mammals (Breedlove and Arnold, 1981). Like racial categorization, the term “dimorphism” promotes typological or essentialist thinking, whereas most brain features are more appropriately described as “monomorphic” across the sexes of humans and common experimental animals (Eliot et al., 2021). Indeed, for most measures of mammalian brains and behavior, there is considerable overlap between the sexes such that the sex of any individual cannot be reliably predicted on the basis of that measure alone. (However, multivariate machine-learning algorithms can predict an individual’s sex with good accuracy based on hundreds of small morphometric differences, especially when brain size is not controlled; Sanchis-Segura et al., 2020). Similarly, a sex difference requiring hundreds of individuals to detect cannot be used to predict the best course of treatment for any single individual (Figure 5).



As the foregoing discussion makes clear, sex and gender are pertinent to many issues concerning the brain and behavior. However, sex and gender themselves comprise many components and covariates that will always better account for the differences between groups of males and females or variance across all genders (Richardson, 2022). Thus, precise operationalization and attention to confounds is key, and as researchers increasingly incorporate comparisons of males and females, it is important to note this nuance and evaluate these confounds accordingly. For example, if a certain drug induces more side effects in women than men, should the dosing be adjusted by sex or by body weight, fat-free mass, circulating steroid hormone levels, or some

other continuous physiological variable (Özdemir et al., 2022)? Similarly, when differences are found in neuropsychiatric disease risk, might they be explained by covariates of gender or sex, such as caregiver stress, occupational attainment, or steroid hormone levels (Mielke et al., 2018)? These questions are challenging to answer, but need to be considered for any drug, treatment, or condition that shows group-level differences between men and women, or boys and girls.

Finally, it is important for anyone commenting on brain sex differences to be mindful of the sexist history that has prefaced research on this topic. Across the many centuries that philosophers and scientists have been addressing brain sex/gender differences, their search has largely focused on the neural basis of women's presumed intellectual and creative inferiority (Saini, 2017; Schiebinger, 1993). Although recent years have seen a decline in overt misogyny, the interpretation of brain sex differences is often still posed within a stereotyped frame, such as assumptions about men's propensity for "action" versus women's "intuition" (Ingalhalikar et al., 2014) that demonstrably influence public understanding of gender and reinforce benevolent sexism (O'Conner & Joffee, 2014). More broadly, typological thinking neglects trans and gender diverse people. Most of this concern is focused on human studies, but as preclinical studies are funded on the basis of their relevance to human health, such considerations are also important when translating findings from non-human animals.

As part of an antisexist framework, there is also a renewed call to look beyond sex differences, to study key aspects of the human condition largely ignored by prior generations of research. Neuroscientists know little of how menopause, pregnancy, the menstrual cycle, and hormone-based medications influence the brain (Taylor et al., 2021). Even less of the research on these topics has addressed people across diverse ethnic, socioeconomic and geographic origins (Peterson et al., 2023). These intersectional blind spots have limited progress on human health generally, and brain health in particular (Taylor et al., 2019).

Thus, special caution applies to the interpretation of neuroscientific findings based on the predominantly White and (presumed) cisgender populations that comprise most of the large and highly-utilized brain/behavioral databases. As we seek to understand the causes of sex/gender behavioral and mental health disparities, it is important to ask, "which men?" and "which women?" and to appreciate that sex and gender lie within an intersectional framework of social identities, all of which interact to shape neural function and behavior (Duchesne and Trujillo, 2021). Merely averaging brain measures in one population, no matter how large, may not lead to a better understanding of average male-female behavioral differences and risks endorsing essentialist interpretations that are already the default among many clinicians and the lay public. Little attention has thus far been paid to cross-cultural or ethnic variation in gender expression, to understanding the development and environmental modulation of gender expression, or—more broadly—to understanding the neural basis of gender in any non-binary way (Rauch and Eliot, 2022). The danger is especially great at present, as legislative bodies increasingly co-opt biomedical findings to define individual rights based on "genetics," a "person's chromosomes" or "anatomy" as assessed at birth. By embracing, rather than over-simplifying, the nuance in sex- and gender-related neurobehavioral data, neuroscientists have the opportunity to challenge such essentialist interpretations that are presently aimed at restricting the rights of sexual and gender minorities (Sudai et al., 2022).

5.1 Recommendations

- Report negative as well as positive findings.
- Consider effect sizes and practical relevance of sex differences.
- Consider environmental explanations for sex differences, even when the variables measured are physiological in nature.
- Be sensitive to researcher bias and the sexist history of sex difference research.
- Consider whether the study population is properly representative of human diversity.

REFERENCES

Allison DB, Brown AW, George BJ, Kaiser KA (2016) Reproducibility: A tragedy of errors. *Nature*, 530:27–29.

Austad SN, Fischer KE (2016) Sex differences in lifespan. *Cell Metabolism*, 23:1022-1033.

Arnold AP, Burgoyne PS (2004) Are XX and XY brain cells intrinsically different? *Trends in Endocrinology and Metabolism*, 15: 6-11.

Arnold AP, Chen X (2009) What does the "four core genotypes" mouse model tell us about sex differences in the brain and other tissues? *Frontiers in Neuroendocrinology*, 30: 1-9.

Azcoitia I, Arevalo MA, Garcia-Segura LM (2018) Neural-derived estradiol regulates brain plasticity. *Journal of Chemical Neuroanatomy*, 89:53-9.

Bale TL, Epperson CN (2015) Sex differences and stress across the lifespan. *Nature Neuroscience*, 18:1413-1420.

Becker JB, Koob GF (2016) Sex differences in animal models: Focus on addiction. *Pharmacological Reviews*, 68:242-63.

Becker JB, Prendergast BJ, Liang JW (2016) Female rats are not more variable than male rats: a meta-analysis of neuroscience studies. *Biology of Sex Differences*, 7:1-7.

Beltz AM, Beery AK, Becker JB (2019) Analysis of sex differences in pre-clinical and clinical data sets. *Neuropsychopharmacology*, 44:2155–2158.

Beery AK (2018) Inclusion of females does not increase variability in rodent research studies. *Current Opinion in Behavioral Sciences*, 23:143–149.

Berman KF, Schmidt PJ, Rubinow DR, Danaceau MA, Van Horn JD, Esposito G, Ostrem JL, Weinberger DR (1997) Modulation of cognition-specific cortical activity by gonadal steroids: A positron-emission tomography study in women. *Proceedings of the National Academy of Sciences*, 94:8836-41.

Bland JM, Altman, DG (2015) Best (but oft forgotten) practices: Testing for treatment effects in randomized trials by separate analyses of changes from baseline in each group is a misleading approach. *The American Journal of Clinical Nutrition*, 102:991-994.

Blehar MC, Spong C, Grady C, Goldkind SF, Sahin L, Clayton JA (2013) Enrolling pregnant women: issues in clinical research. *Women's Health Issues*, 23:e39-45.

Bohacek J, Manuella F, Roszkowski M, Mansuy IM (2015) Hippocampal gene expression induced by cold swim stress depends on sex and handling. *Psychoneuroendocrinology*, 52:1-12.

Bolte G, Jacke K, Groth K, Kraus U, Dandolo L, Fiedel L, Debiak M, Kolossa-Gehring M, Schneider A, Palm K (2021) Integrating sex/gender into environmental health research: Development of a conceptual framework. *International Journal of Environmental Research and Public Health* 18:12118.

Bond KM., McCarthy MM, Rubin JB, Swanson KR (2021) Molecular omics resources should require sex annotation: a call for action. *Nature Methods*, 18:585–588.

Breedlove SM (1994) Sexual differentiation of the human nervous system. *Annual Review of Psychology*, 45:389-418.

Breedlove SM, Arnold AP (1981) Sexually dimorphic motor nucleus in the rat lumbar spinal cord: response to adult hormone manipulation, absence in androgen-insensitive rats. *Brain Research*, 225:297-307.

Buch T, Moos K, Ferreira FM, Frölich H, Gebhard C, Tresch A (2019) Benefits of a factorial design focusing on inclusion of female and male animals in one experiment. *Journal of Molecular Medicine* 97: 871-877.

Buckley, RF, Waller M, Masters CL, Dobson A (2019) To what extent does age at death account for sex differences in rates of mortality from Alzheimer disease? *American Journal of Epidemiology*, 188:1213-1223.

Carcel C, Harris K, Peters SA, Sandset EC, Balicki G, Bushnell CD, Woodward M (2021) Representation of women in stroke clinical trials: a review of 281 trials involving more than 500,000 participants. *Neurology*, 97: e1768-e1774.

Champagne FA, Curley JP (2009) Epigenetic mechanisms mediating the long-term effects of maternal care on development. *Neuroscience & Biobehavioral Reviews*, 33:593-600.

Clayton JA (2018) Applying the new SABV (sex as a biological variable) policy to research and clinical care. *Physiology & Behavior*, 187: 2-5.

Clayton JA, Tannenbaum C (2016) Reporting sex, gender, or both in clinical research? *JAMA*, 316:1863-1864.

Cornelison TL, Clayton JA (2017) Article commentary: considering sex as a biological variable in biomedical research. *Gender and the Genome*, 1:89-93.

De Vries GJ (2004) Minireview: Sex differences in adult and developing brains: Compensation, compensation, compensation. *Endocrinology*, 145:1063–1068.

Duchesne A, Kaiser Trujillo A (2021) Reflections on neurofeminism and intersectionality using insights from psychology. *Frontiers in Human Neuroscience*, 15:684412.

Eliot L (2011) The trouble with sex differences. *Neuron*, 72:895-98.

Eliot L, Ahmed A, Khan H, Patel J (2021) Dump the “dimorphism”: Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neuroscience and Biobehavioral Reviews*, 125:667-97.

Fiest KM, Roberts JI, Maxwell CJ, Hogan DB, Smith EE, Frolikis A, Cohen A, Kirk A, Pearson D, Pringsheim T, Venegas-Torres A, Jetté N (2016) The prevalence and incidence of dementia due to Alzheimer's disease: A systematic review and meta-analysis. *Canadian Journal of Neurological Science*, 43 Suppl. 1:S51-82.

Forde NJ, Jeyachandra J, Joseph M, Jacobs GR, Dickie E, Satterthwaite TD, Voineskos AN (2020) Sex differences in variability of brain structure across the lifespan. *Cerebral Cortex*, 30:5420-5430.

Galea LAM, Choleris E, Albert AYK, McCarthy MM, Sohrabji F (2019) The promises and pitfalls of sex difference research. *Frontiers in Neuroendocrinology*, 56:100817.

Garcia-Sifuentes Y, Maney DL (2021) Reporting and misreporting of sex differences in the biological sciences. *eLife*, 70817.

Gatewood JD, Wills A, Shetty S, Xu J, Arnold AP, Burgoyne PS, Rissman EF (2006) Sex chromosome complement and gonadal sex influence aggressive and parental behaviors in mice. *Journal of Neuroscience*, 26:2335-42.

Geller SE, Koch AR, Roesch P, Filut A, Hallgren E, Carnes M (2018) The more things change, the more they stay the same: a study to evaluate compliance with inclusion and assessment of women and minorities in randomized controlled trials. *Academic Medicine*, 93: 630-635.

Gelman A, Stern H (2006) The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician*. 60:328-31.

George BJ, Beasley TM, Brown AW, Dawson J, Dimova R, Divers J, Goldsby TU, Heo M, Kaiser KA, Keith SW, Kim MY, Li P, Mehta T, Oakes JM, Skinner A, Stuart E, Allison DB (2016) Common scientific and statistical errors in obesity research. *Obesity* 24:781-790.

Goerzen D, Fowler C, Devenyi GA, Germann J, Madularu D, Chakravarty MM, Near J (2020) An MRI-derived neuroanatomical atlas of the Fischer 344 rat brain. *Scientific Reports*, 10:6952.

Gold EB, Crawford SL, Avis NE, Crandall CJ, Matthews KA, Waetjen LE, Lee JS, Thurston R, Vuga M, Harlow SD (2013) Factors related to age at natural menopause: longitudinal analyses from SWAN. *American Journal of Epidemiology*, 178:70-83.

Goodfellow PN, Lovell-Badge R (1993) SRY and sex determination in mammals. *Annual Review of Genetics* 27:71-92.

Grant BF, Dawson DA, Stinson FS, Chou SP, Dufour MC, Pickering RP (2004) The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991–1992 and 2001–2002. *Drug and Alcohol Dependence*, 74:223–234.

Gruene TM, Flick K, Stefano A, Shea SD, Shansky RM (2015) Sexually divergent expression of active and passive conditioned fear responses in rats. *Elife*, 4:e11352.

Hänggi J, Fövenyi L, Liem F, Meyer M, Jäncke L (2014) The hypothesis of neuronal interconnectivity as a function of brain size-a general organization principle of the human connectome. *Frontiers in Human Neuroscience*, 8:915.

Heidari S, Babor TF, 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*, 1:1-9.

Hoekzema E, Barba-Müller E, Pozzobon C, Picado M, Lucco F, García-García D, ... Vilarroya O (2017) Pregnancy leads to long-lasting changes in human brain structure. *Nature Neuroscience*, 20:287-296.

Horstmann S, Schmeche C, Palm K, Oertelt-Prigione S, Bolte G (2022) The operationalisation of sex and gender in quantitative health-related research: A scoping review. *International Journal of Environmental Research and Public Health*, 19:7493.

Hudson JI, Hiripi E, Pope HG, Kessler RC (2007) The prevalence and correlates of eating disorders in the national comorbidity survey replication. *Biological Psychiatry*, 61:348–358.

Hyde JS (2005) The gender similarities hypothesis. *American Psychologist*, 60:581-592.

Hyde JS, Bigler RS, Joel D, Tate CC, van Anders SM (2019) The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, 74:171–193.

Hyde JS, Mezulis AH. (2020) Gender differences in depression: biological, affective, cognitive, and sociocultural factors. *Harvard Review of Psychiatry*, 28:4-13.

Ingallalikar M, Smith A, Parker D, Satterthwaite TD, Elliott MA, Ruparel K, Hakonarson H, Gur RE, Gur RC, & Verma R (2014) Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences USA* 111:823–828.

Jacobs EG, Weiss BK, Makris N, Whitfield-Gabrieli S, Buka SL, Klibanski A, Goldstein JM (2016) Impact of sex and menopausal status on episodic memory circuitry in early midlife. *Journal of Neuroscience*, 36:10163-73.

Jäncke L (2018) Sex/gender differences in cognition, neurophysiology, and neuroanatomy. *F1000 Faculty Review*, 7:805.

Jäncke L, Mérillat S, Liem F, Hänggi J (2015) Brain size, sex, and the aging brain. *Human Brain Mapping*, 36(1):150-169.

Jain A, Huang GZ, Woolley CS (2019) Latent sex differences in molecular signaling that underlies excitatory synaptic potentiation in the hippocampus. *Journal of Neuroscience*, 39:1552-1565.

Kaczkurkin AN, Raznahan A, Satterthwaite TD (2019) Sex differences in the developing brain: insights from multimodal neuroimaging. *Neuropsychopharmacology*, 44:71–85.

Klein SL, Schiebinger L, Stefanick ML, Cahill L, Danska J, De Vries GJ, ... & Zucker I (2015) Sex inclusion in basic research drives discovery. *Proceedings of the National Academy of Sciences*, 112:5257-5258.

Knoedler JR, Inoue S, Bayless DW, Yang T, Tantry A, et al. (2022) A functional cellular framework for sex and estrous cycle-dependent gene expression and behavior. *Cell* 185:654-71 e22

Krieger N (2003) Genders, sexes, and health: what are the connections—and why does it matter? *International Journal of Epidemiology*, 32:652–657.

Labonté B, Engmann O, Purushothaman I, Menard C, Wang J, Tan C, Scarpa JR, Moy G, Loh YE, Cahill M, Lorsch ZS, Hamilton PJ, Calipari ES, Hodes GE, Issler O, Kronman H, Pfau M, Obradovic ALJ, Dong Y, Neve RL, ... Nestler EJ (2017) Sex-specific transcriptional signatures in human depression. *Nature Medicine*, 23:1102–1111.

Lewis JD, Theilmann RJ, Sereno MI, Townsend J (2009) The relation between connection length and degree of connectivity in young adults: a DTI analysis. *Cerebral Cortex*, 19:554-562.

Levy DR, Hunter N, Lin S, Robinson EM, Gillis W, Conlin EB, ... Datta SR (2023) Mouse spontaneous behavior reflects individual variation rather than estrous state. *Current Biology*.

Lonsdorf EV, Eberly LE, Pusey AE (2004) Sex differences in learning in chimpanzees. *Nature*, 428:715-716.

Loomes R, Hull L, Mandy WPL (2017) What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, 56:466–474

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

Maney DL (2014) Just like a circus: The public consumption of sex differences. *Current Topics in Behavioral Neuroscience*, 19:279-296.

Maney DL (2016) Perils and pitfalls of reporting sex differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1688):20150119.

Maney DL, Rich-Edwards JW (2023) Sex-inclusive biomedicine: Are new policies increasing rigor and reproducibility? *Women's Health Issues*. In press. doi.org/10.1016/j.whi.2023.03.004

Maney DL, Karkazis K, Hagen, KSB (2023) Considering sex as a variable at a research university: Knowledge, attitudes, and practice. *Journal of Women's Health*.

Martínez-García M, Paternina-Die M, Cardenas SI, Vilarroya O, Desco M, Carmona S, Saxbe DE (2023) First-time fathers show longitudinal gray matter cortical volume reductions: evidence from two international samples. *Cerebral Cortex*, 33:4156-63.

Massey SC, Whitmire P, Doyle TE, Ippolito JE, Mrugala MM, et al. (2021) Sex differences in health and disease: A review of biological sex differences relevant to cancer with a spotlight on glioma. *Cancer Lett* 498:178-87.

McCarthy MM (2016) Multifaceted origins of sex differences in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371:20150106.

McCarthy MM (2020) A new view of sexual differentiation of mammalian brain. *Journal of Comparative Physiology. A, Neuroethology, Sensory, Neural, and Behavioral Physiology*, 206:369-78.

McCarthy MM, Arnold AP (2011) Reframing sexual differentiation of the brain. *Nature Neuroscience*, 14(6):677-83. doi: 10.1038/nn.2834.

McCarthy MM, Arnold AP, Ball GF, Blaustein JD, De Vries GJ (2012) Sex differences in the brain: The not so inconvenient truth. *Journal of Neuroscience*, 32(7), 2241-2247.

McCarthy MM, Herold K, Stockman SL (2018) Fast, furious and enduring: Sensitive versus critical periods in sexual differentiation of the brain. *Physiology & Behavior*, 187: 13-19.

McCarthy MM, Konkle AT (2005) When is a sex difference not a sex difference? *Front Neuroendocrinol.* 26(2):85-102. doi: 10.1016/j.yfrne.2005.06.001.

McLean CP, Asnaani A, Litz BT, Hofmann SG (2011) Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *Journal of Psychiatric Research*, 45:1027-1035.

Miani C, Wandschneider L, Niemann J, Batram-Zantvoort S, Razum O (2021) Measurement of gender as a social determinant of health in epidemiology—A scoping review. *PLoS One*, 16:e0259223.

Mielke MM, Ferretti MT, Iulita MF, Hayden K, Khachaturian AS (2018) Sex and gender in Alzheimer's disease - does it matter? *Alzheimer's & Dementia*, 14:1101-1103.

Moore CL (1992) The role of maternal stimulation in the development of sexual behavior and its neural basis. *Annals of the New York Academy of Sciences*, 662(1), 160-177.

National Academies of Sciences, Engineering, and Medicine (2023) *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/26902>.

NIH (2015) Reviewer guidance to evaluate sex as a biological variable (SABV). Retrieved from: https://grants.nih.gov/grants/peer/guidelines_general/SABV_Decision_Tree_for_Reviewers.pdf. Accessed October 12, 2022.

NIH (2020) Sex as a biological variable: A primer. Retrieved from: <https://orwh.od.nih.gov/career-development-education/e-learning/sabv-primer>. Accessed October 12, 2022.

Nielsen MW, Stefanick ML, Peragine D, Neilands TB, Ioannidis J, Pilote L, Prochaska JJ, Cullen MR, Einstein G, Klinge I, LeBlanc H, et al. (2021) Gender-related variables for health research. *Biology of Sex Differences*, 12:1-16.

Nieuwenhuis S, Forstmann BU, Wagenmakers E-J (2011) Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience*, 14:1105-7.

Nottebohm F, Arnold AP (1976) Sexual dimorphism in vocal control areas of the songbird brain. *Science*, 194:211-213.

O'Connor C, Joffe H (2014) Gender on the brain: A case study of science communication in the new media environment. *PLoS ONE*, 9:e110830.

Özdemir BC, Gerard CL, Espinosa da Silva C (2022) Sex and gender differences in anticancer treatment toxicity: a call for revisiting drug dosing in oncology. *Endocrinology*, 163:bqac058.

Pelletier R, Khan NA, Cox J, Daskalopoulou SS, Eisenberg MJ, Bacon SL, Lavoie KL, Daskupta K, Rabi D, Humphries KH, Norris CM, Thanassoulis G, Behlouli H, Pilote L (2016) Sex versus gender-related characteristics: which predicts outcome after acute coronary syndrome in the young? *Journal of the American College of Cardiology*, 67:127–135.

Peterson N, Beltz AM, Casto KV, Taylor CM, Jacobs EG, Sundström-Poromaa I, Pletzer B (2023) Toward a more comprehensive neuroscience of hormonal contraception. *Nature Neuroscience*, 26:529-531.

Pintzka CW, Hansen TI, Evensmoen HR, Håberg AK (2015) Marked effects of intracranial volume correction methods on sex differences in neuroanatomical structures: a HUNT MRI study. *Frontiers in Neuroscience*, 9:238.

Polanczyk G, De Lima MS, Horta BL, Biederman J, Rohde LA (2007) The worldwide prevalence of ADHD: A systematic review and metaregression analysis. *American Journal of Psychiatry*, 164:942–948.

Prendergast BJ, Onishi KG, Zucker I (2014) Female mice liberated for inclusion in neuroscience and biomedical research. *Neuroscience & Biobehavioral Reviews*, 40:1-5.

Pritschet L, Santander T, Taylor CM, Layher E, Yu S, Miller MB, Grafton ST, Jacobs EG (2020) Functional reorganization of brain networks across the human menstrual cycle. *NeuroImage*, 220:117091.

Pritschet L, Taylor C, Santander T, Jacobs EG (2021) Applying dense-sampling methods to reveal dynamic endocrine modulation of the nervous system. *Current Opinions in Behavioral Science*, 40:72-78.

Pritschet L (2022) The scientific body of knowledge – whose body does it serve? A spotlight on human neuroscience. NIH Science Policy Scholar Talk presented at the Organization for the Study of Sex Differences, Marina Del Ray, CA.

Raam T, Hong W (2021) Organization of neural circuits underlying social behavior: A consideration of the medial amygdala. *Current Opinion in Neurobiology*, 68:124-36.

Rauch JM, Eliot L (2022) Breaking the binary: Gender versus sex analysis in human brain imaging. *NeuroImage*, 264:119732.

Rechlin RK, Splinter TF, Hodges TE, Albert AY, Galea LA (2022) An analysis of neuroscience and psychiatry papers published from 2009 and 2019 outlines opportunities for increasing discovery of sex differences. *Nature Communications*, 13:1-14.

Richardson SS (2022) Sex contextualism. *Philosophy, Theory, and Practice in Biology*, 3: 100550.

Rippon G, Eliot L, Genon S, Joel D (2021) How hype and hyperbole distort the neuroscience of sex differences. *PLoS Biology*, 19: e3001253.

Ritchie SJ, Cox SR, Shen X, Lombardo MV, Reus LM, Alloza C, Harris MA, Alderson HL, Hunter S, Neilson E (2018) Sex differences in the adult human brain: evidence from 5216 UK Biobank participants. *Cerebral Cortex*, 28:2959–2975.

Rosenthal, R (1979) The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86: 638.

Rutter M, Caspi A, Fergusson D, Horwood LJ, Goodman R, Maughan B, Moffitt TE, Meltzer H, Carroll J (2004) Sex differences in developmental reading disability: new findings from 4 epidemiological studies. *JAMA*, 291:2007–2012.

Ryan BC, Vandenberg JG (2002) Intrauterine position effects. *Neuroscience and Biobehavioral Reviews*, 26:665-678.

Sanchis-Segura C, Ibañez-Gual MV, Aguirre N, Cruz-Gómez ÁJ, Forn C (2020) Effects of different intracranial volume correction methods on univariate sex differences in grey matter volume and multivariate sex prediction. *Scientific Reports*, 10:12953.

Sainani K (2010) Misleading comparisons: The fallacy of comparing statistical significance. *PM&R* 2:559-562.

Saini A (2017) *Inferior: How Science Got Women Wrong*. Boston: Beacon Press.

Salk RH, Hyde JS, Abramson LY (2017) Gender differences in depression in representative national samples: meta-analyses of diagnoses and symptoms. *Psychological Bulletin*, 143:783–822.

Schiebinger L (1993) *Nature's Body: Gender in the Making of Modern Science*. Boston: Beacon Press.

Shors TJ, Chua C, Falduto J (2001) Sex differences and opposite effects of stress on dendritic spine density in the male versus female hippocampus. *Journal of Neuroscience*, 21:6292-6297.

Smarr BL, Ishami AL, Schirmer AE (2021) Lower variability in female students than male students at multiple timescales supports the use of sex as a biological variable in human studies. *Biology of Sex Differences*, 12:32.

Smith-Bouvier DL, Divekar AA, Sasidhar M, Du S, Tiwari-Woodruff SK, King JK, Arnold AP, Singh RR, Voskuhl RR. A role for sex chromosome complement in the female bias in autoimmune disease. *Journal of Experimental Medicine*, 205:1099-108.

Sorge RE, Strath LJ (2018) Sex differences in pain responses. *Current Opinion in Physiology*, 6:75-81.

Spring S, Lerch JP, Henkelman RM (2007) Sexual dimorphism revealed in the structure of the mouse brain using three-dimensional magnetic resonance imaging. *Neuroimage*, 35:1424-33.

Springer KW, Mager Stellman J, Jordan-Young RM (2012) Beyond a catalogue of differences: a theoretical frame and good practice guidelines for researching sex/gender in human health. *Social Science Medicine*, 74:1817-1824.

Sudai M, Borsa A, Ichikawa K, Shattuck-Heidorn H, Zhao H, Richardson SS (2022) Law, policy, biology, and sex: Critical issues for researchers. *Science*, 376:802-804.

Tabatadze N, Huang G, May RM, Jain A, Woolley CS (2015) Sex differences in molecular signaling at inhibitory synapses in the hippocampus. *Journal of Neuroscience*, 35:11252-11265.

Tapp A, Maybery MT, Whitehouse AJ (2011) Evaluating the twin testosterone transfer hypothesis: a review of the empirical evidence. *Hormones and Behavior*, 60:713-722.

Taylor C, Pritschet L, Yu S, Jacobs EG (2019) Applying a women's health lens to the study of the aging brain. *Frontiers in Human Neuroscience*, 13:224.

Taylor C, Pritschet L, Olsen R, Layher E, Santander T, Grafton ST, Jacobs EG (2020) Progesterone shapes medial temporal lobe volume across the human menstrual cycle. *NeuroImage*, 220:117125.

Taylor C, Pritschet L, Jacobs EG (2021) The scientific body of knowledge – whose body does it serve? A spotlight on oral contraceptives and women's health factors in neuroimaging. *Frontiers in Neuroendocrinology*, 60:100874.

U.S. Department of Health and Human Services, n.d. Sex & gender. National Institutes of Health Retrieved from <https://orwh.od.nih.gov/sex-gender>.

Vijayakumar N, de Macks ZO, Shirtcliff EA, Pfeifer JH (2018) Puberty and the human brain: Insights into adolescent development. *Neuroscience & Biobehavioral Reviews*, 92:417-436.

Vorland CJ (2021) Sex difference analyses under scrutiny. *eLife*, 10:e74135.

Vorland CJ, Brown AW, Dawson JA, Dickinson SL, Golzarri-Arroyo L, Hannon BA, Heo M, Heymsfield SB, Jayawardene WP, Kahathuduwa CN (2021) Errors in the implementation, analysis, and reporting of randomization within obesity and nutrition research: a guide to their avoidance. *International Journal of Obesity*, 45:2335-2346.

Vorland CJ, Golzarri-Arroyo L, Maney DL, Brown AW (2023) Determination of sex differences requires formal test for differences: Comment on 'Lipoprotein-subclass particle numbers in children with abdominal obesity.' *Pediatrics International*, 65:e15489.

Wallen K, Baum MJ (2002) Masculinization and defeminization in altricial and precocial mammals: Comparative aspects of steroid hormone action. In *Hormones Brain and Behavior*, ed. D Pfaff, pp. 385-424. London, UK: Academic Press.

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. *Journal of Clinical Endocrinology & Metabolism*, 106:3034–3048.

Wierenga LM, Doucet GE, Dima D, Agartz I, Aghajani M, Akudjedu TN, ... Wittfeld K (2022) Greater male than female variability in regional brain structure across the lifespan. *Human Brain Mapping*, 43:470-499.

Williams CM, Peyre H, Toro R, Ramus F (2021) Neuroanatomical norms in the UK Biobank: The impact of allometric scaling, sex, and age. *Human Brain Mapping*, 42:4623-4642.

Wilson LAB, Zajitschek SRK, Lagisz M, Mason J, Haselimashhadi H, Nakagawa S (2022) Sex differences in allometry for phenotypic traits in mice indicate that females are not scaled males. *Nature Communications*, 13:7502.

Woitowich NC, Beery A, Woodruff T (2020) Meta-research: a 10-year follow-up study of sex inclusion in the biological sciences. *eLife*, 9:e56344.

Woolley CS, Gould E, Frankfurt M, McEwen BS (1990) Naturally occurring fluctuation in dendritic spine density on adult hippocampal pyramidal neurons. *Journal of Neuroscience*, 10:4035–4039.

Zell E, Krizan Z, Teeter SR (2015) Evaluating gender similarities and differences using metasynthesis. *American Psychologist*, 70:10-20.

Zhang K, Sejnowski TJ (2000) A universal scaling law between gray matter and white matter of cerebral cortex. *Proceedings of the National Academy of Sciences*, 97:5621-5626.