The importance of sex and estrus as biological variables in neuroscience research

Anna Püsök, Kendall Moore, Katharine M. Cammack
The University of the South, Sewanee, Tennessee 37383

Females have been historically underrepresented in preclinical literature and sex differences remain underexplored in neuroscience, pharmacology, and behavioral research, particularly in rodent models (Beery and Zucker, 2011; Will et al., 2017; Mamlouk et al., 2020). A major reason for the omission of females is a concern over the natural fluctuations in gonadal hormone levels across their reproductive (estrus) cycle, which can affect behavior and neuroanatomy. This review presents an argument for examining sex and estrus stage as biological variables and discusses ways in which neurophysiological and behavioral data may (or may not) be impacted by females’ estrus stage. Evaluation of existing estrus staging protocols, approaches to minimize potential variables, and ways to optimize estrus reliability and replicability in undergraduate research experiences are presented as well. Finally, we included a small pilot study during laboratory members’ training to assess the inter- and intra-rater reliability of estrus identifications and further refine our protocol. We discuss the importance of training and education in undergraduate research laboratories that use preclinical models, as well as potential implications related to equity and the reproducibility crisis.


Keywords: female, sex difference, estrus cycle, mice, behavior, protocol, undergraduate research

Background

Sex bias & omission in neuroscience research

For decades, neuroscience research has reflected a strong sex bias, defined as more frequent use of male over female research animals (Beery and Zucker, 2011; Will et al., 2017; Mamlouk et al., 2020). In 2011, Beery and Zucker evaluated instances of sex bias in studies from 10 major biological disciplines, including neuroscience, that used preclinical (non-human mammalian) models and were published in scientific journals (Zucker and Beery, 2010; Beery and Zucker, 2011). They identified sex bias by calculating the ratio of studies reporting only male data versus only female data within the studies from each discipline; a male bias was identified in 80% of the disciplines, including neuroscience. These notable sex biases have been attributed, in part, to the increased use of rodents, particularly mice (e.g., transgenic models) in preclinical research over the past century (Beery and Zucker, 2011; Mamlouk et al., 2020). Notably, in the same journals, this sex bias was less prominent within the human studies in neuroscience. This discrepancy has important implications for understanding psychiatric disorders (e.g., anxiety, depression) and neurological diseases (e.g., multiple sclerosis) that are more common in women than men (Beery and Zucker, 2011).

Sex omission, defined as a failure to incorporate sex as a biological variable, is also common in preclinical neuroscience research (Mamlouk et al., 2020; Will et al., 2017; Zucker and Beery, 2010). Amongst a sample of neuroscience articles published in 2009, only ~20% of studies that included both sexes
performed statistical analyses to identify sex differences (Beery and Zucker, 2011). Subsequent studies (Mamlouk et al., 2020) found that sex omission in rat and mouse studies decreased between 2010 (47.1%) and 2017 (10.3%). However, across neuroscience studies published in 2017, over half of the sampled mouse studies included both males and females, but only 16% of those studies analyzed data by sex. Far fewer rat studies (~26%) included both males and females, likely due to cost and husbandry considerations; a similar proportion (~18%) of those studies analyzed data by sex (Mamlouk et al., 2020). Together, these findings suggest that, while more rodent studies include both males and females, few of these studies explore sex as a biological variable, contributing to persistent sex omission within neuroscience research.

In an effort to reduce sex bias/omission in preclinical research, in 2015 the National Institutes of Health (NIH) instituted a policy (NOT-OD-15-102) to promote the inclusion of sex as a biological variable in study designs, data analyses, and reporting (NIH 2015). However, Mamlouk and colleagues (2020) found that neither sex bias nor omission varied based on NIH funding status, suggesting that these issues persisted despite NIH funding status. Indeed, the proportion of neuroscience articles that only used male rats or mice remained fairly consistent and high (~30-40% between 2010–2017; Mamlouk et al., 2020), suggesting a persistent sex bias within neuroscience research using rodent models.

While it is possible that the data reported in these studies (published in 2017) were collected prior to the NIH recommendations, the larger scientific community has cited the additional costs, necessary training, perceived variability of data collected from female rodents, and assumed generalizability of findings as reasons that they have continued to rely exclusively on male rodents. The additional costs associated with including both sexes (rather than one) in a rodent study (including, but not limited to, animal housing, experimenter time and labor, and expendable materials) can be non-trivial and impact the scope and timeline of a proposed set of experiments. Three potential concerns – experimenter training, perceived variability of females’ data, and perceived generalizability from males to females – will be addressed in this review.

Scientists can study sex differences in preclinical (rodent) models in a variety of ways, including qualitative differences, quantitative differences, population differences, and differences in underlying mechanisms (Becker and Koob, 2016). While a comprehensive discussion of these approaches is beyond the scope of this review, qualitative and quantitative differences in females’ hormone levels are important to consider, particularly with reference to the above rationale for sex bias and omission. In this review, we will summarize key aspects of females’ reproductive biology (focusing on gonadal hormone levels), discuss ways that these differences seem to impact behavioral findings and ways that they do not, and describe easy and cost-effective ways to incorporate sex as a biological variable into neuroscience research. It is our belief that incorporating these reproductive cycle protocols into more preclinical experiments will help reduce sex biases in scientific research.

Females’ hormone levels fluctuate across the reproductive cycle

In mammals, the hypothalamic-pituitary-gonadal (HPG) axis regulates fluctuating levels of major reproductive hormones across the females’ reproductive cycle (Becker et al., 1992). In both rodents and primates, hypothalamic neurons release gonadotropin-releasing hormone (GnRH) into the anterior pituitary. In response, pituitary cells release FSH into the general circulation, which travels to ovarian tissue to stimulate the growth of ovarian follicles, which contain ova. As FSH levels increase, these follicles increase in size (helping the ova to mature) and release estrogen (estradiol) into the general circulation. Increasing levels of estrogen eventually cause the pituitary to release a large surge of LH, which causes the now-mature follicle to rupture and ovulation to occur. Estrogen levels subsequently decline, and the resulting corpus luteum of the ovary produces progesterone to help support possible pregnancy. If the ovum is not fertilized and/or fails to implant in the uterus, progesterone levels drop and the cycle begins again. Notably, rodents will reabsorb the uterine lining that builds up during the latter half of this cycle while humans will
typically menstruate. Thus, in primates, this cycle is often referred to as the menstrual cycle; in rodents, this cycle is referred to as the estrus cycle.

**The estrus cycle describes sequential stages of female rodents’ reproductive cycle**

In rodents, the female reproductive cycle—the estrus cycle—consists of four stages: proestrus, estrus, metestrus (or diestrus I), and diestrus (or diestrus II; Butcher et al., 1974). Each stage is associated with different levels of hormones within the HPG axis (Butcher et al., 1974). For example, during the proestrus stage of the cycle, 17-β-estradiol, luteinizing (LH) and follicle-stimulating hormone (FSH) levels reach their peak levels in blood plasma, while progesterone levels are relatively low. Ova maturation and ovulation occur during proestrus and estrus (Butcher et al., 1974). During metestrus, estradiol and progesterone levels increase, and progesterone levels peak during the last phase of the cycle, diestrus (Marcondes et al., 2002).

The first estrus cycle generally coincides with the onset of puberty and vaginal opening ~26 days after birth (Caligioni et al., 2009; Ajayi and Akhigbe, 2020). The length of the estrus cycle is typically shorter (four to five days) in mice and rats than in other mammals (18-24 days in cattle; Spornitz et al., 1999; Marcondes et al., 2002; Caligioni, 2009; Crown, 2022). While each estrus stage typically lasts ~24hrs, there is some variability in the duration of each cycle stage. For instance, Caligioni and colleagues (2009) tested mice’s estrus stage via vaginal lavage (see below) daily for 19 days; despite regular testing, the same stage was identified multiple days in a row at various points for each mouse, all four stages were not identified within each full cycle, and a varying number of full cycles (two to four) were identified over the 19 days. While some of these differences may have been due, in part, to variables such as experimenter or timing of sample collection, these data nevertheless suggest that estrus stage on a given day cannot be assumed based on the estrus stage identified on the previous or subsequent day.

Are data collected from females more variable due to estrus status?

The endocrine system, including the HPG axis and gonadal hormones, is widespread and interacts with a number of nervous system structures (Becker et al., 1992), suggesting that estrus status may influence neuroscientific data. Indeed, researchers have identified select changes in females’ neuroanatomy and behavior across the different stages of rodents’ reproductive cycle. For example, the number of dendritic spines on hippocampal neurons fluctuates across the rat estrus cycle, with significantly fewer spines in estrus than proestrus (Woolley et al., 1990). This anatomical difference was notable given that these two stages are sequential and typically separated by only 24-48hrs. As the hippocampus is important for learning, memory, and spatial navigation (Lisman et al., 2017), this potential confound may have contributed to some of the sex bias in preclinical research.

Estrus stage can also impact anxiety levels. Marcondes and colleagues (2001) found that proestrus females spent more time on the open arms of the elevated plus-maze and that this reduction in anxiety coincided with the highest serum estradiol levels. While diestrus females spent the least amount of time in the open arms, their open-arm time increased to proestrus levels following estradiol treatment. A similar reduction in anxiety during proestrus and estrus was identified by Mora and colleagues (1996), but only in low-light conditions. Together, these results suggest that fluctuating ovarian hormone levels across the estrus cycle may impact certain affective behaviors.

These findings, alongside others, contributed to the perception that females’ data might be more variable than males’ data. To explore this question, Prendergast and colleagues (2014) calculated the coefficient of variation (CV) as a relative measure of variability across various measures in a range of trait categories, including many relevant to neuroscience (e.g., neuroanatomy, neurotransmitters, learning, and emotional behaviors, electrophysiology). They found that data collected from unstaged (no estrus stage identified) female mice were no more variable than data collected from male mice (Prendergast et al., 2014). These findings suggest that the amount of variability within a dataset
collected from females is similar to the amount of variability within a dataset collected from males.

**Estrus status impacts reward-related brain circuits**

While Prendergast’s meta-analysis suggested that overall variability in data did not differ between females and males, growing evidence suggests that select reward-related behaviors are modulated by gonadal hormones and thus differ across estrus stages. We will focus on the mesocorticolimbic circuits involved in reward-related processes.

Dopamine (DA) is a catecholamine that acts as a chemical signal at the synapse. Dopaminergic (dopamine-containing) neurons are located in two midbrain structures: the ventral tegmental area (VTA) and the substantia nigra (SN). The VTA DA neurons project to the nucleus accumbens (nAc); this pathway, known as the mesolimbic DA system, is involved in reward- and motivation-related processes (Berger, 2012; Salamone and Correa, 2012; Salamone et al., 2018). Extracellular DA levels in the nAc change in response to the presentation of a variety of rewarding and aversive stimuli (e.g., Joseph et al., 2003; Young, 2004; Salamone et al., 2015). Most abused drugs increase DA levels in the nAc (DiChiara and Imperato, 1988), indicating that the drug-induced activation (whether direct or indirect) of these DA neurons may serve as a common pathway for the incentive-motivational value of these drugs (Berridge, 2012; Salamone and Correa, 2012; Salamone et al., 2018).

The baseline activity of VTA DA neurons is modulated estrus stage. The basal firing activity of these DA neurons is higher during estrus (Zhang et al., 2008; Calipari et al., 2017) than proestrus (Zhang et al., 2008) or diestrus (Calipari et al., 2017). Basal levels of DA in the nAc are highest during estrus, consistent with the above findings, and this peak coincides with high estrogen levels (Calipari et al., 2017).

Additionally, drugs impact the mesolimbic DA system differently across the estrus cycle. Calipari and colleagues (2017) found that cocaine potency, assessed through concentration-response curves, was higher in estrus females than diestrus females or males. Cocaine’s ability to inhibit the dopamine reuptake transporter (DAT) also seemed to be potentiated during estrus, suggesting that cocaine had greater affinity for DAT and thus greater impact on DA clearance in the nAc. This functional difference coincided with a great proportion of phosphorylated DAT and extracellular signal-regulated kinase (ERK) during estrus compared to diestrus and provides a potential explanation for the enhanced uptake function. Further sex differences in addiction have been reviewed elsewhere (Yoest et al., 2018; Becker, 2016).

**Estrus status can influence drug preference**

Extensive literature suggests that gonadal hormones, particularly estrogen, contribute to the reward value of certain drugs and aspects of addiction-like behavior in rodents (for review, see Becker and Koob, 2016; Becker and Chartoff, 2019). For instance, when estrogen is high, female rats are more likely to initiate cocaine use, escalate their drug-taking more rapidly, experience more severe withdrawal signs, and relapse after being drug-free. It is thus likely that the incentive-motivational value of drugs may also change across the estrus cycle.

A common way to assess the incentive-motivational value of drugs is using a conditioned place preference (CPP) procedure (Bardo et al., 1995; Tzschentke, 1998, 2007; Cunningham et al., 2006). Briefly, a CPP apparatus typically consists of a black and a white (with some of them containing an additional middle gray) compartment. During the conditioning period, one context is paired with drug injections and the other is paired with vehicle (saline) injections. These chamber-injection pairings occur once per day for eight days, on alternating days. During a drug-free subsequent testing session, the animal model is able to move between the sides, and choose a preferred side. The amount of time that mice spent in each chamber reflected their preference for that chamber and therefore for the drug. Of course, many factors—including but not limited to species/strain, animal age, drug dose, route of administration, length of conditioning session, and drug history—can affect the strength of CPP and may reveal sex and/or estrus-related differences (Bardo et al., 1995; Tzschentke, 1998, 2007; Cunningham et al., 2006).
Calipari and colleagues (2017) explored cocaine CPP in mice, with respect to sex and estrus. Mice received only two cocaine-chamber and two saline-chamber pairings, a subthreshold conditioning procedure that does not typically induce strong CPP. However, estrus females displayed stronger CPP than males or diestrus females. Using fiber photometry to measure VTA cell responses with high temporal resolution, researchers found that the activity of VTA neurons increased prior to entry to the cocaine-paired chambers to a greater extent in estrus females than the other two groups. Activity at VTA neuron terminals within the nAc also showed increased activity upon entry into the cocaine-paired chamber, and this activity correlated positively with the strength of cocaine CPP expressed by the mice. Together, these findings suggest that the mesolimbic DA system functions differently as a result of estrus stage and that these differences contribute to the incentive-motivational value of cocaine.

Incorporating estrus status as a biological variable in undergraduate neuroscience research

For equity and transparency, it is important for undergraduate researchers in neuroscience laboratories to consider sex and estrus status in preclinical experiments using rodents.

How to identify females’ estrus stage

Estrus stage can be identified using three techniques. First, blood samples can provide quantitative hormone levels (e.g., Butcher et al., 1974). This procedure can require specific materials (e.g., needles), large equipment (e.g., ELISA assays for hormone levels), and/or survival surgeries (e.g., for IV-line placement), and thus specialized experimenter training. These procedures, even when performed correctly (e.g., tail vein sampling), can cause distress to the rodent, particularly if restraint is involved (Guide for the Care and Use of Laboratory Animals, 2011). Furthermore, this approach can alter plasma hormone and metabolite concentrations (Arnold and Langhans, 2010), likely due to the stress induced by blood sampling procedures.

A second technique is the visual observation of the vaginal opening (Byers et al., 2012). Some stages (estrus) are more amenable to visual identification than others, and the room lighting and mouse strain (e.g., coat color) can be variables. Nevertheless, this technique is low-cost, minimally distressing, and requires no specialized equipment.

A third technique is vaginal lavage and cytology. Vaginal lavage involves the non-invasive collection of a sample of sloughed cells from the vaginal canal via droplet aspiration (Caligioni, 2009; Marcondes et al., 2002; McLean et al., 2012; Byers et al., 2012). Basic cytology is used to identify different cell types prominent at distinct stages of the estrus cycle (Caligioni, 2009; Marcondes et al., 2002; McLean et al., 2012; Byers et al., 2012). Briefly, proestrus is defined by small, round, nucleated epithelial cells. During the estrus stages, predominantly anucleated keratinized epithelial cells are present. Metestrus is characterized by the combination of anucleated keratinized epithelial cells and neutrophils. Diestrus is defined by very low cellularity, as well as a decrease in the number of anucleated keratinized epithelial cells and neutrophils (Cora et al., 2015). While cytology is correlated with hormonal status and cannot provide concrete information about hormone levels (Caligioni, 2009; Marcondes et al., 2002; McLean et al., 2012), this technique requires little specialized equipment and training, making it a scalable, cost-effective, and efficient way to identify estrus stages.

Timing & frequency of vaginal lavage

The estrus cycle is transitional (Cora et al., 2015), hormone levels can vary within a given stage (Butcher et al., 1974), and there can be individual variability in stage length/progression (e.g., Caligioni et al., 2009). Thus, the timing and frequency of vaginal lavage must be considered carefully.

Repeated lavage may impact behavioral assays commonly used in behavioral neuroscience research. For instance, repeated (7 – 10 days) lavage, prior to an open-field locomotor test, reduces cocaine-induced locomotion in female rats (Walker et al., 2002). As repeated lavage can induce pseudopregnancy (McLean et al., 2012), it is possible that altered...
hormones and/or stress may contribute to this effect. In contrast, a single lavage prior to the open-field session did not affect locomotor activity (Walker et al., 2002).

The timing of lavage can also impact behavioral outcomes. For instance, in a modified CPP study, vaginal lavage was performed immediately before or after female rats were placed in an empty CPP conditioning chamber (Walker et al., 2002). After repeated “pairings,” rats expressed a conditioned preference for the chamber if lavage preceded the conditioning sessions but not if lavage followed the conditioning sessions. Touch to the rats’ flank prior to conditioning sessions did not induce preference, suggesting that the vaginal stimulation associated with lavage may impact CPP, a common behavioral assay used in neuroscience research.

Given these findings, it seems valuable to perform vaginal lavage at least 1hr following the completion of any behavioral assays. If no behavioral assays are performed, lavage should be performed at the same time each day, preferably early in the morning. This could potentially further reduce variability among samples (Cora et al., 2015).

Protocols for vaginal lavage & cytological analysis

There are many outstanding protocols and scientific papers on vaginal lavage and cytological analysis. McLean and colleagues’ protocol (2002) is highly cited and associated with instructive video tutorials (https://www.jove.com/v/4389/performing-vaginal-lavage-crystal-violet-staining-vaginal-cytological). The cytological images presented by Cora and colleagues (2015) are comprehensive and include valuable images to assist with training and troubleshooting, including insufficient samples and common contaminants. The schematic estrus cycle stage identification tools for rats and mice (Byers et al., 2012; Cora et al., 2015) are particularly helpful in conveying the relative ratios of different cell types present at each stage.

Our laboratory adapted and combined these resources for our own estrus staging protocol. Briefly, a disposable transfer pipette is filled with <0.5ml of double distilled water (ddH20) and checked to ensure that bubbles are not present. The female mouse is placed on the wire cage top; she will typically hold onto the bars with her forepaws, allowing the proximal end of her tail to be gently grasped and lifted to expose the vaginal opening. The tip of the pipette is placed just outside the vaginal opening and the water is gently pipetted in and out four to five times. The sample is placed immediately on a clean microscope slide. Sometimes a water droplet remains visible at the opening of the vaginal canal and can be aspirated using the pipet and added to the sample already on the slide. The sample may look slightly cloudy. The slides are air-dried overnight in a slide folder to protect them from dust, then stained with a 0.1% crystal violet solution as described by McLean and colleagues (2012). Samples can be analyzed using a brightfield microscope at 10x-40x magnification; we have used large Zeiss microscopes and small teaching microscopes with similar results. Samples are analyzed with reference to exemplar images (McLean et al., 2012; Cora et al., 2015).

The importance of training

Adequate training in cytological analysis is important for accurate and reliable estrus stage identification. First, students can start by reading existing protocols and scientific papers (McLean et al., 2002; Byers et al., 2012; Cora et al., 2015), and applying their understanding to exemplar samples and/or images. Subsequent training with an experienced researcher on identifying cell types on “real-world” samples that are less clear to stage and/or contain contaminants may help students identify areas of confusion and troubleshoot difficult samples. For instance, leukocytes are smaller than the epithelial cells and have multilobular nuclei, which can make them more difficult to distinguish from slide contaminants (e.g., dust, crystal violet granules). Finally, trained students can share sample identifications with each other, pointing out aspects of each sample that they observed, and thought was important.

It is also important to emphasize the transitional nature of estrus cycle stages. Each sample represents a one-time point within a particular stage, and each stage is characterized by shifting ratios of the different cell types. For
example, in mice, late metestrus and early diestrus each contain all three cell types at fairly similar ratios, as reflected in the mouse estrus cycle stage identification tool (Byers et al., 2012; Cora et al., 2015). Conversely, stages characterized by predominantly one cell type (e.g., estrus) are generally easier to identify.

Finally, as many undergraduate labs have many student researchers, it may also be helpful to have one-two “resident experts” in estrus staging that can lead estrus identification in multiple projects. Limiting the number of students analyzing each sample may enhance consistency and reduce variability within and across laboratory projects.

**Reliability & reproducibility are critical for sound science**

An important part of the scientific process is reproducibility, or the ability to reliably repeat an experiment/observation to produce consistent data. Reproducibility helps establish lab protocols, verify findings, and build upon others’ work. However, approximately 90% of researchers believe there is a reproducibility crisis in science (Baker 2016). Over 70% of researchers have been unable to replicate results from other scientists’ studies, and more than half of researchers have been unable to replicate their own results (Baker 2016). Many factors prevent researchers from reproducing their own/other’s data, including insufficient mentoring/oversight, problems with experimental design, and time pressure (Baker 2016), all issues that can be found in undergraduate neuroscience laboratories. It is therefore important to acknowledge the importance of, as well as include opportunities for, scientific replication and reproduction in undergraduate laboratory experiences. Reliability and reproducibility is especially important in sex difference related research. If skewed data is presented due to not reporting both sexes equally or sex differences in studies not being addressed, that often leads to the loss of reproducibility (Gulinello et al., 2019).

**Assessing reliable & replicable estrus staging protocols in our laboratory**

We performed a small pilot study during laboratory members’ training to assess the inter- and intra-rater reliability of estrus identifications and refine our stage identification protocol.

**Materials & Method**

Mice were ~60-90 days old, C57BL/6J strain (Jackson Laboratories), and singly housed on a reverse 12hr:12hr light-dark cycle (lights on at 0700) with *ad libitum* access to food and water. Animal care and use procedures were approved by the Institutional Animal Care & Use Committee (IACUC #6-2022). Briefly, 12 cytological samples were collected via vaginal lavage between 0900-1300 by the laboratory head (K. Cammack). After lavage, a drop of each sample is placed on a slide in a thin layer and air-dries. The samples were stained with crystal violet and stored at room temperature.

Cytology was analyzed using a bright-field microscope at 10-40x magnification. The presence of each of three cell types (nucleated epithelial cells, cornified epithelial cells, and leukocytes) was recorded using a four-part scale (0 = no cells present; + = few cells present; ++ = moderate number of cells present; +++ = many cells present) adapted from existing work (Cora et al., 2015). These qualitative data allowed an estimate of the ratio of these cells, which was used to determine estrus status, with reference to the mouse estrus cycle stage identification tool (Byers et al., 2012; Cora et al., 2015).

Nine undergraduate student researchers in our Behavioral Neuroscience Laboratory analyzed each of the 12 samples three times within 14 days. We defined their first observation as “novice” and their third observation as “experienced.” Students were blind to their own previous ratings, as a way to assess overall training as well as intra-rater reliability. Students were also blind to each other’s observations, to assess inter-rater reliability. All results are descriptive.

**Results**

On some samples, students were fairly consistent between their first and third observation sessions (Figure 1). More inconsistency reflected the importance of training and value of the experimenter’s experience. The consistency of stage identification did not seem to be related to the amount of biology background, year at college, or length of time as
a laboratory member, but a larger sample would be needed to verify this while still retaining researchers’ anonymity.

![Image](image.png)

**Figure 1.** Proportion of cytological samples identified as the same cycle stage (“consistent”) or a different cycle stage (“inconsistent”) by each student researcher (A-H) during their first (novice) and third (experienced” observation session. One research did not complete both sessions and was omitted from analyses.

Identification of certain samples not only benefited from experience but also differed fairly reliably between student researchers. For instance, a similar proportion of student researchers identified Female 10’s sample as estrus and metestrus (Figure 2), which lowered the interrater reliability. Leukocytes seemed more difficult to identify, so this may have contributed to the differing identifications. In general, stages characterized by predominantly one cell type (e.g., estrus) were easier to identify within and across experimenters. Finally, inconsistencies within and between researchers may have also been influenced by response fatigue, where

![Image](image.png)

**Figure 2.** Proportion of student researchers who identified a cytological sample (n=10) as one of three estrus stages during their first (“novice”) and third (“experienced”) analysis.

Future directions to our approaches in the lab include exploring the idea of ceiling effect (the number of observations reach the plateau phase, where new observations are not helpful with the identification of the stages). Furthermore, we would like to explore whether there is a correct number of observations with which we can get the most accurate identification of the stages. The creation of a standard, replicable estrus staging protocol could further encourage researchers to use both female and male subjects in their research and explore sex effects in their studies.

### Discussion

Females have been underrepresented in preclinical research in a variety of fields, including neuroscience, for a long time (Beery & Zucker 2011; Mamlouk et al., 2020). While recent trends are somewhat encouraging (Mamlouk et al., 2020), many research studies still do not analyze sex differences and/or account for females’ estrus status. Our scientific understanding of basic nervous system function (e.g., mesolimbic DA system) and of neurological and psychiatric conditions within females will continue to be limited without greater attention to sex bias and omission.

There are promising changes in the research community. Large meta-analyses have provided valuable context and reassurance to researchers wary of using females (Prendergast et al., 2014). Further, the formation of specialized journals, like the Organization for the Study of Sex Differences (https://www.ossdweb.org/) hold promise for promoting research of female neurobiology and pharmacology. Open-access resources, such as publicly available workshops, on the study of sex as a biological variable (e.g., https://commonfund.nih.gov/sexdifferences/workshop) may also help get researchers started.

At an undergraduate level, our pilot study demonstrates a low-cost, trainable approach to identifying estrus status in female mice. Our data on inter-rater and intra-rater reliability underscore the importance of adequate training in
reproducibility, or the ability to reliably repeat an experiment/observation to produce consistent data. Reproducibility is critical for sound science, including establishing lab protocols, verifying findings, and building upon others’ work (Baker 2016), and is an important aspect of undergraduate research experiences both in and outside of the classroom.

Working to reduce sex bias and omission may also contribute importantly to broader science citizenship. Scientific data are not always interpreted and communicated accurately in lay media, and miscommunications can exacerbate existing biases or misinformation about the nature, scope and relevance of sex differences. Maney (2015) addresses a few ways to address the issue. First, a shift in the narrative of sex difference may perpetuate the idea that there are dichotomous categories. Instead, the term “sex effect” may better include females in the narrative without necessarily indicating a difference that impacts behavior meaningfully. Maney also suggests focusing on the factors that covary with sex, such as genes or hormones, which shape the development of the brain. Finally, Maney recommends presenting the data in a way that includes the overlaps between sexes instead of highlighting the differences only so that people understand the effect of sex better. We believe that scientists have a responsibility to help address and mitigate these issues. Attention to sex in neuroscience will be maximally beneficial if the narrative is accurate and policies are made with compelling and complete scientific information.

Future neuroscience studies should carefully consider the rationale for studying females in preclinical studies. It is often difficult to change decades of practice, but it is important to increase representation of females in research and identify unexplored sex effects in these fields. It is our belief that proper education and training in sex differences will help mitigate sex bias and omission, which will subsequently advance our understanding of the role sex plays in neurotypical and atypical conditions.

Acknowledgments

We thank our fellow members of the Behavioral Neuroscience Laboratory for contributing to the datasets presented in this manuscript: Revie Atkinson, Laura Botros, AbdAlrahman Najjar, Tyler Reese, Yoshi Ruffin-Douglas, Rob Sullivan, and Yana Van Den Abbeele. We also thank the Office of Undergraduate Research for summer funding for Anna Püsök, and the Psychology Department & Neuroscience Program at The University of the South for their support.

Corresponding Authors

Anna Püsök
The University of the South
735 University Avenue, Sewanee, TN 37383
pska0@sewanee.edu

Kendall Moore
The University of the South
735 University Avenue, Sewanee, TN 37383
mooreke0@sewanee.edu

References

Mamlouk GM, Dorris DM, Barrett LR, Meitzen J (2020) Sex bias and omission in neuroscience research is influenced by research model and journal, but not reported NIH funding. Front Neuroendocrinol 57:100835.


Wagner MECV (n.d.) OSSD. Organization for the Study of Sex Differences Available at: https://www.ossdweb.org/ [Accessed April 25, 2023].


