



## COMMENT

## Analysis of sex differences in pre-clinical and clinical data sets

Adriene M. Beltz <sup>1</sup>, Annaliese K. Beery<sup>2,3</sup> and Jill B. Becker <sup>1,4</sup>*Neuropsychopharmacology* (2019) 44:2155–2158; <https://doi.org/10.1038/s41386-019-0524-3>

## INTRODUCTION

As more studies begin to address the role of sex as a biological variable (SABV), it has become increasingly important to understand how to collect and analyze the data so that the presence or absence of sex differences can be assessed. This has led to some concerns about how to conduct statistical analyses. In this brief commentary we do not attempt to review the field of sex differences, but provide a conceptual guide to the statistical analysis of sex differences in research with both animal and human subjects.

## STATUS QUO

The use of predominantly male subjects has been documented across myriad research domains—both historically and at present—with the extent of bias varying by subdiscipline [1, 2]. Females have historically been viewed as more variable than males because of the presence of estrous or menstrual cycles in many species, although this belief has been discredited across a wide range of animal models (reviewed in [3, 4]). Noting the potential negative consequences of subject sex bias for women's health, the National Institutes of Health issued a 1993 act to include women in clinical research, followed by a 2014 policy to encourage balanced use of male and female subjects and tissues by considering SABV during research design and analysis. The mere inclusion of females, however, does not provide insight into the role of sex/gender in physiological, behavioral, and psychological traits, and the majority of studies using male and female subjects fail to report on whether sex differences are present [2, 4].

## TYPES OF SEX DIFFERENCES

In order to analyze the role of SABV, it is important to understand that there are different types of sex differences, and that these may require different analysis strategies. These categories are not mutually exclusive—more than one type of sex difference may be involved in any given trait. Figure 1 provides a graphical representation of four types of sex differences, as well as guidance on conducting analyses based on the type of sex difference. Specifically, major sex differences that prevent the analysis of males and females on the same scale or metric (Fig. 1a) are qualitative sex differences. Qualitative differences are also present if variables are not related to each other in the same way across the sexes (Fig. 1b). In the case of qualitative sex differences, analyses should be conducted independently by sex and the data reported as though they are two independent experiments. When

the average or mean of a dependent variable is different for males and females, these are quantitative sex differences (Fig. 1e) and analyses should be conducted with sex as a factor. There are also sex differences where one aspect of a trait is the same for males and females, but the mechanisms underlying the trait are different, or emerge only under certain conditions: these are latent sex differences (Fig. 1c; also referred to as mechanistic, convergent or divergent). Finally, when the proportions of males and females that exhibit particular traits in response to the independent variable are different, these are population sex differences (Fig. 1d). For latent and population sex differences, the types of analyses will vary based on the data. See the Fig. 1 legend for discussion. For additional examples and discussion of these different types of sex differences refer to [5]. It is important to note that many assessments were developed and validated using exclusively male samples, and when extending these tests to females it is not always clear whether they are operationalizing the same trait in females - a topic which requires further consideration [6].

## BEST PRACTICES

How should a researcher approach data analysis in light of potential sex differences of various types? While it is beyond the scope of this brief commentary to give prescriptive guidance for all research involving both males and females, we hope to provide some guiding concepts the reader can use when evaluating their data. First, one should consider what is already known about sex differences in the context of the study topic and consider the data that have been obtained (Fig. 1). For a process that only occurs in one sex or varies fundamentally by sex (i.e., qualitative differences), the analysis must proceed separately by sex or only in the sex in which the process occurs (e.g., specific reproductive functions). When sex should be included as a variable in analyses (i.e., if differences are quantitative) descriptions of the effect size or magnitude of any sex differences may help determine the practical significance of the differences, regardless of their statistical significance, and confirmation of the effect is essential [7]. In yet other cases, sex may not appear to be an important contributor to the final results; in these instances, the lack of sex differences should be documented and reported, and consideration should be given to the ability of the data set to identify sex differences (e.g., statistical power). Specific examples of how such analyses might proceed appear in Table 1. In general, we advocate for reporting as much data as feasible for each sex, analyzed in different ways if appropriate. This avoids selecting data that

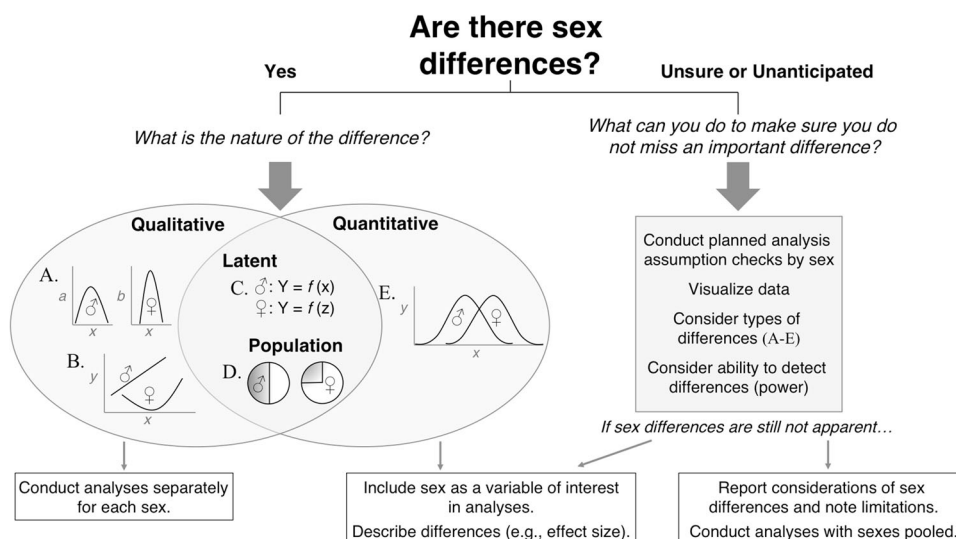
<sup>1</sup>Department of Psychology, University of Michigan, Ann Arbor, MI 48109, USA; <sup>2</sup>Department of Psychology, Department of Biology, Neuroscience Program, Smith College, Northampton, MA, USA; <sup>3</sup>Neuroscience and Behavior Graduate Program, University of Massachusetts, Amherst, MA, USA and <sup>4</sup>Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, Michigan 48109, USA

Correspondence: Jill B. Becker (jbbecker@umich.edu)

These authors contributed equally as co-first authors: Adriene M. Beltz, Annaliese K. Beery

Received: 22 May 2019 Revised: 30 August 2019 Accepted: 4 September 2019

Published online: 17 September 2019



**Fig. 1** Flowchart of questions to ask, possible answers, and suggested analysis approaches when examining empirical data from male and female subjects. Although informed by the broader literature, responses to questions should be made with respect to the particular data set being analyzed. If a sex difference is expected or present (“Yes”), then the nature of the difference will influence the type of analysis to be conducted. Qualitative differences, in which different traits are examined in the sexes (“A”) or in which the functional form of a relation varies by sex (“B”), require analyses to be conducted separately for males and females; that is, data should be analyzed as if the males and females are from two independent experiments. Quantitative differences, in which males and females show the same pattern of traits, but to different levels, extents, or degrees (“E”), indicate analyses that include sex as a variable of interest (e.g., using sex as a term in a factorial analysis). Latent differences mark sex differences in the mechanisms underlying a trait (“C”), and population differences reflect sex differences in the prevalence of different, non-overlapping traits (“D”). Both can be evaluated in analyses conducted separately by sex, or by including sex as a variable of interest in planned analyses; the decision depends upon the research question, and whether the specific difference of interest is qualitative or quantitative. Thus, a given behavior can reflect multiple types of sex differences. In addition to specifying the statistical significance of sex differences, it is also important to describe their direction and practical importance (e.g., effect size, such as Cohen’s *d* or percent difference). If it is unclear whether a sex difference is present or not (“Unsure or Unanticipated”), then sex should be included in all assumption checks performed on the data prior to conducting inferential analyses; this may include consideration of all types of differences (“A” thru “E”) with descriptive statistics (e.g., distributions), variance and normality evaluations by sex, and data visualizations (e.g., scatterplots). The sequence and nature of these checks may differ by data set, and may also be done prior to analyzing data in which sex differences are expected. Results can then guide the response to the initial question (i.e., “What can you do to make sure you do not miss an important difference?”). If still unsure whether there are sex differences, then sex can be included as a variable of interest in planned analyses, with the knowledge that analyses will only capture quantitative and some latent and population differences, or analyses may pool subjects across sex. When pooling subjects, effect size should be considered in addition to statistical power, as samples may be too small to detect differences even if they are present (although increases in sample size required to examine main and even interaction effects of sex are modest; see [5]). It may also be important to note when examinations of sex differences are incomplete (e.g., if all types of differences were not studied) or inconclusive (e.g., if power was limited to detect effects). We do not recommend controlling for gender or sex differences as these are non-random differences [8]. Other types of analyses may require examination of individual differences in addition to sex differences [9–12]. See Table 2 for more information.

**Table 1.** Sample analysis paths for different research scenarios concerning the study of sex differences

Sample research scenario	Sample analysis paths
The process of oogenesis and ovulation in women is qualitatively different from spermatogenesis in men. Women ovulate once each ~28 day cycle, whereas men produce sperm continuously. (Many examples of qualitative sex differences involve reproductive pathways.)	Unless comparison of specific aspects of gametogenesis between males and females is the focus, analyze data separately by sex using statistical tests appropriate for the research question.
Pavlovian fear conditioning is conducted in males by quantifying freezing, but when females are tested their ‘deficits’ in a fear response turn out to be an alternative strategy—females also exhibit escape responses [6]. This is a qualitative difference and a quantitative difference.	A qualitative sex difference in response to fear conditioning is first demonstrated. Then, individual behaviors (darting, velocity, and freezing) are scored and average values for each behavior for each group are analyzed for each session using two-way repeated measured ANOVA [6].
The link between pubertal timing and depression is qualitatively different in human males and females. In males, there is a linear relation, with late maturing males reporting greater symptoms than early maturing males. In females, there is a quadratic relation, with early and late maturing individuals reporting greater symptoms than those who are on-time. This also suggests a latent difference in the puberty-depression relationship for the sexes.	Analyze data separately for males and females. Regression analyses (with polynomial functions) have been used [10]. Other possible approaches include growth curve models and structural equation models [12].
In humans there are sex differences in spatial skills; in fact, the largest cognitive sex difference is in three dimensional mental rotations. This is a quantitative difference, in that—on average—males perform the task more accurately and rapidly than females.	Include sex as a variable of interest in planned analyses. Independent samples <i>t</i> tests and factorial analyses can be used; in the latter, sex should be considered as a main and interactive (with other variables of interest) effect.
Female rats are more likely than males to prefer cocaine to a food reward, but cocaine-preferring males and females are similar to each other in terms of behavioral responding and changes in nucleus accumbens dopamine release. This is a population difference.	It might be misleading to include sex as a variable of interest in analyses, as frequency (or another nonparametric) analysis points to the nature of the sex difference. But, males and females of similar preference phenotypes may be pooled and statistical tests appropriate for the research question used [9].
Each subtype of red-green “color blindness” is more common in men than women (women require mutations on both X chromosomes versus the single X chromosome). This is a population difference.	Phenotypes are similar within each subtype of red-green color blindness, so males and females could be studied together within specific mutation/dichromacy variants, and/or as controls, and sex could be used as a factor in analysis. For inheritance-related research

**Table 1** continued

Sample research scenario	Sample analysis paths
<p>Pain hypersensitivity following nerve injury exhibits a sex difference in signaling pathways. Males exhibit microglia-mediated signaling, while female signaling is dependent on T cells. Cross-over between these pathways is observed in males lacking testosterone, and in pregnant females or those lacking T cells. This suggests a latent difference.</p> <p>Mice are being tested on a particular behavioral test for the first time—e.g., a same-sex version of the partner preference test developed in voles—and potential sex differences are unknown.</p>	<p>questions, non-parametric analyses of population frequency (e.g., chi-square or logistic regression) or analyses conducted separately by sex are needed.</p> <p>Simultaneous study of males and females and analysis with sex as a variable of interest is necessary to identify these and other differences in pain processing [13]. For the study of sex-specific signaling (e.g., the role of pregnancy in females), data would be analyzed in one sex using statistical tests appropriate for the research questions.</p> <p>Include males and females as study subjects. If qualitative sex differences are present, analyze females and males separately. If differences are quantitative, use sex as a factor in analyses. If not present, use sex as a factor, or pool across sexes. In any of these cases, report on the presence/absence and magnitude of sex differences.</p>

**Table 2.** Responses to frequently asked questions about the study of sex differences

<p>What if data from males and females are collected in different cohorts or at different times?</p>	<p>Under these conditions it is not appropriate to directly compare males and females. Similarities and differences between the samples may be discussed at a conceptual level. Potential sex differences should be investigated with simultaneous testing of both sexes in future studies.</p>
<p>What if the trait of interest occurs predominantly or only in one sex? Should both sexes be studied?</p>	<p>After a sex difference in the incidence of a trait (i.e., a population difference) has been established, additional studies may be sex-specific. Especially for biomedical research, mechanisms underlying relevant outcomes in both male and female subjects should be studied.</p>
<p>Is it always necessary to compare females across the estrous or menstrual cycle because of potential cycle-dependent variability in females?</p>	<p>No. Female rodents are not more variable than males, and ovarian cycles do not introduce greater variability in females. If estrous or menstrual cyclicity is part of the research question, then cycle stage is a valid variable to consider. Ways to incorporate these variables into a scientific design have been discussed previously [14].</p>
<p>Is it appropriate to “control” for sex in studies with humans?</p>	<p>No. In the human literature, it is not uncommon to include both males and females in studies, yet discount or disregard effects of sex in analyses. This is typically done by “controlling for” sex or using sex as a covariate (of no interest) in statistical tests. Doing this relies on two assumptions: (1) once sex-related variance is removed, sex does not “confound” results; and (2) study findings are then valid for both sexes. Both assumptions, however, are faulty (for a discussion of the perils of statistically controlling for non-random group differences in analyses, see [8]). The first assumes that sex differences are quantitative, linear, and without interactions, which is rarely the case (discussed above and in Becker et al. [15]). The second assumes individuals are “averagely-sexed” or “sex-less,” and very rarely is this true of the populations to which generalizations are being made. Instead, sex should be explicitly considered as a variable of interest (see Fig. 1).</p>
<p>Can sex differences change over time? How should repeated measures data related to sex be analyzed?</p>	<p>Yes. All types of sex differences can change over proximal or distal time. For instance, a sex difference may be present at one measurement occasion, but not the next, or there may be sex differences in patterns of change, including variability. Depending upon the type of difference, data can be analyzed separately by sex or by including sex as a variable of interest in planned analyses. These differences can be detected using longitudinal analysis approaches, such as repeated measures analyses of variance (e.g., for quantitative differences), growth curve analyses (e.g., for qualitative and latent differences), and time series analyses (e.g., of intensive longitudinal data to examine differences in variability and patterns of covariation among variables of interest).</p>
<p>What if some individuals are not well-represented by the average for their sex? For instance, what if a female exhibits a “male-like” trait (e.g., high spatial ability) or a male exhibits a “female-like” trait (e.g., major depression)?</p>	<p>Most studies of sex differences and between-subject analyses (e.g., <i>t</i> tests, regression, factorial designs) assume homogeneity within each sex, so that results can be generalized to all subjects of that sex. This assumption is often violated, especially in human research or research on multi-determined traits. To address this, <i>individuals</i> should be studied using within-subject analyses (e.g., based on intensive longitudinal designs, or studies in which many data points are collected from the same individual on the same variables) that characterize subject-specific patterns (see [11]). This also suggests that analysis of population differences should be investigated for the measure.</p>
<p>I am confused about the different types of sex differences. How do I tell them apart?</p>	<p>The easiest sex differences to discern are qualitative differences, wherein males and females cannot be measured on the same scale; these differences can emerge as a function of sexual differentiation. Population differences resemble qualitative differences in that the traits measured are not overlapping, but both males and females exhibit the traits, albeit in different proportions. These traits may emerge as a function of life events (e.g., early life stress or experience with drugs of abuse). For quantitative differences the trait in males and females is measured on the same scale but the normal distribution around the mean is different for the two sexes. Quantitative and population differences may seem difficult to dissociate, especially if the trait is measured on a scale developed for only one sex. If the distribution of values for the one or both of the sexes is skewed with a disproportion of values at one end of the scale, for example, this may indicate that there is an underlying population difference that could be explored. Finally, latent sex differences in a trait are differences between the sexes that emerge as a function of challenge, such as stress, or when investigating mechanism (e.g., intracellular signaling or gene expression); in this case, studies of a trait might identify quantitative differences, while studies of the mechanism underlying it might identify qualitative differences.</p>

support a hypothesis or defaulting to the assumption that the differences between the sexes are quantitative. Responses to frequently asked questions concerning the analysis of sex differences also appear in Table 2.

## CONCLUSIONS

Our goal has been to lay out a strategy for analysis of data collected from both female and male subjects. We emphasize that investigators should assess and consider the types of sex differences present in the data to guide analyses, and to ensure that meaningful results will be obtained.

## FUNDING AND DISCLOSURES

No funding sources were used in the creation of this commentary. The remaining authors have nothing to disclose.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

1. Sugimoto CR, Ahn Y-Y, Smith E, Macaluso B, Larivière V. Factors affecting sex-related reporting in medical research: a cross-disciplinary bibliometric analysis. *Lancet*. 2019;393:550–9.
2. Beery AK, Zucker I. Sex bias in neuroscience and biomedical research. *Neurosci Biobehav Rev*. 2011;35:565–72.
3. Shansky RM. Are hormones a “female problem” for animal research? *Science*. 2019;364:825–6.
4. Beery AK. Inclusion of females does not increase variability in rodent research studies. *Curr Opin Behav Sci*. 2018;23:143–9.
5. Becker JB, Chartoff E. Sex differences in neural mechanisms mediating reward and addiction. *Neuropsychopharmacology*. 2019;44:166–83.
6. Gruene, TM, Flick, K, Stefano, A, Shea, SD, Shansky, RM. Sexually divergent expression of active and passive conditioned fear responses in rats. *eLife*. 2015; 4.
7. Mogil JS, Macleod MR. No publication without confirmation. *Nature*. 2017; 542:409–11.
8. Miller GA, Chapman JP. Misunderstanding analysis of covariance. *J Abnorm Psychol*. 2001;110:40–48.
9. Perry AN, Westenbroek C, Becker JB. The development of a preference for cocaine over food identifies individual rats with addiction-like behaviors. *Plos ONE*. 2013;8:e79465.
10. Beltz AM. Gendered mechanisms underlie the relation between pubertal timing and adult depressive symptoms. *J Adolesc Health*. 2018;62:722–8.
11. Beltz AM, Wright AGC, Sprague BN, Molenaar PCM. Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment*. 2016;23:447–58.
12. Mendle J, Harden KP, Brooks-Gunn J, Graber JA. Development's tortoise and hare: pubertal timing, pubertal tempo, and depressive symptoms in boys and girls. *Developmental Psychol*. 2010;46:1341–53.
13. Sorge RE, Martin LJ, Isbester KA, Sotocinal SG, Rosen S, Tuttle AH. et al. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nature Methods*. 2014;11, pp. 629–632.
14. Becker JB, Arnold AP, Berkley KJ, Blaustein JD, Eckel LA, Hampson E. et al. Strategies and methods for research on sex differences in brain and behavior. *Endocrinology*. 2005;146, pp. 1650–1673.
15. Becker JB, McClellan ML, Reed BG. Sex differences, gender and addiction. *Journal of Neuroscience Research*. 2017;95, pp. 136–147.