Neuroscience and Biobehavioral Reviews 40 (2014) 1–5

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review

Female mice liberated for inclusion in neuroscience and biomedical research

Brian J. Prendergast[a], Kenneth G. Onishi[a], Irving Zucker[b,c,*]

[a] Department of Psychology and Institute for Mind and Biology, University of Chicago, United States
[b] Department of Psychology, University of California, Berkeley, United States
[c] Department of Integrative Biology, University of California, Berkeley, United States

ARTICLE INFO

Article history:
Received 26 September 2013
Received in revised form
26 December 2013
Accepted 7 January 2014

Dedicated to Elliot Valenstein on the occasion of his 90th birthday.

Keywords:
Mice
Females
Males
Sex bias
Sex differences

ABSTRACT

The underrepresentation of female mice in neuroscience and biomedical research is based on the assumption that females are intrinsically more variable than males and must be tested at each of four stages of the estrous cycle to generate reliable data. Neither belief is empirically based. In a meta-analysis of 293 articles, behavioral, morphological, physiological, and molecular traits were monitored in male mice and females tested without regard to estrous cycle stage; variability was not significantly greater in females than males for any endpoint and was substantially greater in males for several traits. Group housing of mice increased variability in both males and females by 37%. Utilization of female mice in neuroscience research does not require monitoring of the estrous cycle. The prevalence of sex differences at all levels of biological organization, and limitations in generalizing findings obtained with males to females, argue for the routine inclusion of female rodents in most research protocols.

Contents

1. Introduction

Despite well-established sex differences in many aspects of human biology, women still remain underrepresented in clinical trials (Wizemann, 2012; Nature editorial, 2010). Many hold that beyond the reproductive system sex differences in cellular and molecular processes either do not exist or are irrelevant (Wizemann and Pardue, 2001), diminishing the theoretical importance of studying females. Yet, every cell has a sex (in wild-type mice, either XX or XY). There are multiple sex differences in basic genetic, cellular and biochemical organization, some endogenous, others of epigenetic origin; many unrelated to gonadal hormones are commonplace (e.g., Gabory et al., 2009). The prevalence of sex

* Corresponding author at: Department Psychology, University of California, 3210 Tolman Hall (MC 1650), Berkeley, CA 94720-1650, United States.
Tel.: +1 510 642 7136; fax: +1 510 642 5293.
E-mail address: irvzuck@berkeley.edu (I. Zucker).

0149-7634/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neubiorev.2014.01.001

differences in the nervous system and behavior and the immune system is well established. The investigation of both sexes increases heterogeneity of study populations, and results are more likely to generalize to other similarly diverse populations. Cahill (2012) has argued that sex can modify, negate, or even reverse findings; also that phenotypic effects of a gene knockout in one sex may not exist at all or even be reversed in the other sex given the same knockout. This argues strongly for a female presence in virtually all protocols.

The exclusion of females from animal research has been repeatedly highlighted (Berkley, 1992; Mogil and Chanda, 2005; Becker et al., 2005; Hughes, 2007; Beery and Zucker, 2011). A 2009 survey documented male bias in 8 disciplines, with ratios of male-only versus female-only studies ranging from 3.7:1 in physiology to 5:1 in pharmacology and neuroscience (Beery and Zucker, 2011). Hughes (2007) noted that despite repeated attempts to draw attention to sex-dependent drug effects the vast majority of rodent researchers continue to use males exclusively in drug studies. This is problematic given adverse effects of various drugs are more common or severe in women than men (Rogers and Ballantyne, 2008). The tendency to ignore females typifies all types of research, from studies of cell lines to those of higher order behaviors, and everything in between (Beery and Zucker, 2011) and remains uncorrected in 2014. Reforms undertaken by the NIH 20 years ago to counteract limitations inherent in generalizing results of epidemiological and clinical studies of men to women have offered some relief, but males still remain overrepresented and even when females and males are included in human trials, both sexes are not equally represented. A striking male bias also characterizes many animal models of human diseases and traits (Zucker and Beery, 2010).

A major impediment to reversing sex bias in rodent research is the widespread belief that the 4-day estrous cycle of rats and mice requires daily tracking of vaginal cytology, viewed as a time-consuming undertaking in experiments with females. Some maintain that interpretable data emerge only when females are monitored at each of the four stages of the estrous cycle, significantly multiplying the cost of research (Wald and Wu, 2010). Until recently, funding agencies have been unwilling to provide additional funds to cover projected costs of incorporating estrous cycle-staged females (Wald and Wu, 2010).

The decades-long assumption that the estrous cycle renders female rodents intrinsically more variable than males may be the single greatest barrier to eliminating male sex bias in animal research. With one exception (Mogil and Chanda, 2005), the issue of whether freely-cycling rodents, chosen without regard of the stage of the estrous cycle are more variable than males, has not been addressed. A report that assessed nociceptive traits in more than 8000 individual measurements, collected from 40 different mouse strains in 3 laboratories, noted that females tested at random points in their estrous cycles were no more variable than males and that estrous cyclicity either does not add measurably to variability and/or male mice feature their own sex-specific variability (Mogil and Chanda, 2005). Whether this conclusion extends beyond measures of pain is unknown. Arnold and Lusis (2012) maintained that for biomedical science to be relevant to everyone, greater emphasis must be placed on studying females and direct comparison of the two sexes is desirable.

Our goal was to compare male variability on a wide array of traits to that of females tested at random stages of the estrous cycle. We now report that female variability is no greater than that of males, thereby removing a barrier that has contributed to the underrepresentation of females in biomedical research. We also document that variability in both sexes increases in group- compared to individually housed mice

2. Literature survey

2.1. Inclusion criteria

To quantify and compare intrinsic variability of females and males, a range of behavioral, morphological, physiological, and molecular traits was examined in mice tested without regard to the stage of the estrous cycle; analyses were restricted to mice because they dominate contemporary biomedical research; more than 45% of studies published in 2009 in four well-established neuroscience journals employed mice (Beery and Zucker, 2011).

Three separate searches of the ISI Web of Knowledge (Thompson Scientific) with the terms “mice and sex differences”, “mice and estrous cycle” and “mice and estrus” for the years 2009–2012 that were completed in June 2012 yielded 1694 articles,with considerable overlap, each of which was examined for relevance. Excluded from further consideration were review articles, those in which female mice were tested only at specific stages of a monitored estrous cycle, or studies of embryos and fetuses. Also excluded were articles that did not use *Mus musculus* as the model species, those in which only one sex was examined, and articles in which only transgenic mice (absent wild-type controls) were used. The final dataset was comprised of 293 articles, in which female mice tested at random stages of the estrous cycle were compared with males, yielding 9932 trait measurements, spanning 30 broad categories (Table S1).

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.neubiorev.2014.01.001.

2.2. Data extraction

The Methods and Results sections of each article were examined, and the mean value and standard deviation or standard error of the mean were recorded for every dependent variable. In 105 of 9932 instances (1.06% of the data), S.D./S.E.M. were not available, and the range was used as a measure of variance; in 17 instances (0.17%), means were not presented, and the median was used as a measure of central tendency.

Because the majority of data points examined (69.0%) appeared in graphs, mean and SD data were extracted digitally on a Macintosh computer (OSX 10.8) from high-resolution image files, each of which was generated from screenshots of article PDFs (using Grab, version 1.7). A vector graphics software program (Adobe Illustrator CS, version 11.0) was used to quantify the mean and SEM or SD values directly from each graph. Briefly, each graph was overlaid with rectangles, and the length or width (as appropriate) of the rectangles (in mm) provided a relative measure of the mean and SD/SEM for each trait. In addition a rectangle was laid over the dependent variable axis of the graph; the dimensions of this rectangle permitted conversion of measurements from arbitrary mm units into units specific to the dependent variable being measured. In articles that utilized transgenic or knockout mice, data were only extracted from wild-type controls. Graphical data extractions were performed by 2 trained researchers. Inter-rater reliability was high ($R^2 = 0.994$, $P < 0.001$), as assessed using a training set of unfamiliar bar and line graphs ($n = 556$); therefore, measured values obtained from the two researchers were averaged for the final analysis. In some instances ($n = 157$), data were presented both in a graph and in a table (or in the text), which afforded an opportunity to estimate accuracy of the graphical data extractions: extracted values were positively correlated with actual (table/text) trait values ($R^2 > 0.999$, $P < 0.001$), indicating a high degree of measurement accuracy. A minority of the data (29.8%) that appeared exclusively in tables or text of the methods or results sections were transcribed directly.