The National Institutes of Health have recently introduced policies to increase the proportion of studies in which both female and male cells and organisms are used.1 This ongoing emphasis on the study of female cells and organisms2 directly responds to the overwhelming male bias in preclinical scientific research.3,4 This bias limits the translation potential of scientific discovery for the medical care of women, sometimes resulting in poor clinical outcomes.5 For example, of the US Food and Drug Administration–approved drugs withdrawn from the market between 1997 and 2001, 80% were withdrawn because of greater health risk in women than in men6; not surprising considering the use of male to female animals in pharmacological research is almost 6:1.7 Moreover, nonharmful drugs are often differentially effective in women and men, further illustrating the consequences of sex bias in clinical and preclinical research.8

One of the most prominently believed rationales for male bias in research is that the estrous cycle significantly complicates the study of physiology in females. It is well known that many phenotypic measurements are significantly affected by estrogen and other circulating hormones. It is, therefore, reasonable to suppose that varying hormone levels during the estrous cycle would confound the study of these phenotypes in female organisms by increasing the variance in uncontrolled female experimental groups. If present, this increased variance would mean that greater numbers of female animals than male would be required to detect a difference in magnitude. Approaches to counteract these effects include precise estrous staging,9,10 incorporation of gonadectomized rat controls, 11,12 or increasing the size of experimental groups to counteract the variance added by the estrous cycle.

A more robust understanding of sex differences in variability within specific model systems is clearly needed to guide experimental design. Although it is undeniable that specific phenotypes can be profoundly affected by estrous stage,9,10,13 recent data suggest that the effects of estrous cycles and experimental design

Breaking the Cycle

Estrous Variation Does Not Require Increased Sample Size in the Study of Female Rats

Alex Dayton,* Eric C. Exner,* John D. Bukowy, Timothy J. Stodola, Theresa Kurth, Meredith Skelton, Andrew S. Greene, Allen W. Cowley Jr

Abstract—Despite the striking differences between male and female physiology, female physiology is understudied. In response, the National Institutes of Health is promulgating new policies to increase the use of female organisms in preclinical research. Females are commonly believed to have greater variability than males because of the estrous cycle, but recent studies call this belief into question. Effects of estrous cycling on mean arterial pressure were assessed in female Dahl S rats using telemetry and vaginal cytometry and found that estrous cycling did not affect mean arterial pressure magnitude or variance. Data from the PhysGen arm of the Program for Genomic Applications was used to compare male and female variance and coefficient of variation in 142 heart, lung, vascular, kidney, and blood phenotypes, each measured in hundreds to thousands of individual rats from over 50 inbred strains. Seventy-four of 142 phenotypes from this data set demonstrated a sex difference in variance; however, 59% of these phenotypes exhibited greater variance in male rats rather than female. Remarkably, a retrospective power analysis demonstrated that only 16 of 74 differentially variable phenotypes would be detected when using an experimental cohort large enough to detect a difference in magnitude. No overall difference in coefficient of variation between male and female rats was detected when analyzing these 142 phenotypes. We conclude that variability of 142 traits in male and female rats is similar, suggesting that differential treatment of males and females for the purposes of experimental design is unnecessary until proven otherwise, rather than the other way around. (Hypertension. 2016;68:1139-1144. DOI: 10.1161/HYPERTENSIONAHA.116.08207.) • Online Data Supplement

Key Words: estrous cycle ■ experimental design ■ experimental models ■ sample size ■ sex ■ variation

Received July 22, 2016; first decision August 4, 2016; revision accepted August 11, 2016.
From the Department of Physiology (A.D., E.C.E., J.D.B., T.J.S., T.K., M.S., A.S.G., A.W.C.) and Department of Biomedical Engineering (A.S.G.), Medical College of Wisconsin, Milwaukee.
*These authors contributed equally to this work.
The online-only Data Supplement is available with this article at http://hyper.ahajournals.org/lookup/suppl/doi:10.1161/HYPERTENSIONAHA.116.08207/-/DC1.
Correspondence to Allen W. Cowley Jr, Medical College of Wisconsin, 8701 Watertown Plank Rd, Milwaukee, WI 53226. E-mail cowley@mcw.edu
© 2016 American Heart Association, Inc.
Hypertension is available at http://hyper.ahajournals.org DOI: 10.1161/HYPERTENSIONAHA.116.08207

1139
cycling on cohort variance are not as confounding as was assumed. A meta-analysis of 293 publications in which male and female mice were studied found no difference in the distribution of coefficients of variation (CVs) between the male and female groups. However, the actual variances and CVs of individual phenotypes were not compared; rather, difference was detected en bloc. Here, we directly compare the variance and CV between male and female rats in 142 physiological phenotypes that are most specifically related to physiology and pathophysiology of diseases related to the heart, lung, vasculature, blood, and kidney, that is, tissues of specific relevance to the National Heart, Lung, and Blood Institute. Unlike a meta-analytic approach, the current study examines a single data set. This eliminates interexperimenter noise, as well as avoiding statistical artifacts that can present themselves when compiling many experiments, each with small N.

Methods

Experimental Animals

Female rats were obtained at weaning from colonies developed and maintained at the Medical College of Wisconsin under controlled environmental conditions, with parents and offspring fed a purified AIN-76A rodent food (Dyets, Bethlehem, PA) containing 0.4% NaCl with water provided ad libitum until the experimental period of 8.0% NaCl diet (high salt; Dyets, Bethlehem, PA). All experimental protocols were approved by the Medical College of Wisconsin Institutional Animal Care and Use Committee.

Measurement of Blood Pressure

Mean arterial pressure (MAP) was measured by radiotelemetry in adult female rats. At 9 weeks of age, rats were surgically prepared, as we have described previously, and then given a 5- to 7-day recovery period before recording blood pressure. MAP was recorded for 3 hours in the morning each day. Rats continued to be maintained on the 0.4% NaCl diet for 4 days during baseline MAP measurement, after which diet was switched to an 8% NaCl diet for 13 days. Rats were assessed for estrous stage each day, as described, and the date of estrus for each rat was noted. Rats were retrospectively split into 2 groups, indicating those in which estrus was detected 3 or more times during the 16-day period (cycling, N=8) and those in which estrus was detected 2 or fewer times during the 16-day period (noncycling, N=6). The MAP of these rats was compared using a 2-way, repeated measures analysis of variance (ANOVA), and the variance of the pressure data was compared using a repeated measures Brown–Forsythe test, each performed using SigmaPlot 12.

PhysGen Program for Genomic Applications Data Set

Data were downloaded from the PhysGen Program for Genomic Applications website (pga.mcw.edu) on February 29, 2016. Phenotypes were removed from the data set if they were calculated from 2 or more other data points within the data set, for example, ΔMAP after norepinephrine infusion was excluded from further analysis because both pre-norepinephrine MAP and post-norepinephrine MAP were also included in the data set. This exclusion reduced the number of phenotypes from 211 to 160. Further exclusions of phenotypes necessary for the performance of specific statistical tests are noted in the sections describing those tests.

Variance Testing

Differences in variance between males and females were assessed on a phenotype-by-phenotype basis by using the Brown–Forsythe test, a 1-way ANOVA on the set of absolute deviations from the median. Differences in magnitude were assessed on a phenotype-by-phenotype basis by use of heteroscedastic t test. Those phenotypes that exhibited a difference in both magnitude and variance, or a difference in variance alone, were then subjected to a retrospective power analysis to determine the sample size necessary to detect the observed differences. A 1-way ANOVA power analysis was used for both magnitude and variance, most likely resulting in more conservative (ie, greater) estimates of sample size needed to detect differences in magnitude. The power analyses used α=0.05, β=0.2. Between-group variance was calculated as the variance between the male and female means, and within-group variance was taken as the greater of the male and female variance, again resulting in more conservative estimates for sample size. Phenotypes measured in multiple phenotypic categories were consolidated to one point for the All Phenotypes Deming regression by keeping the measurement with lowest N, and excluding the others. All measurements were used for the analysis of phenotypes by strain; however, an additional exclusion criteria was applied: phenotypes within a single strain that had sample size of 2 or fewer were removed. Brown–Forsythe and t tests were performed using functions within the Matlab R2014b Statistics and Machine Learning Toolbox. ANOVA power analyses were performed using R 3.0.1, Stats package.

Coefficient of Variation

CVs were calculated for each phenotype and sex as standard deviation divided by the mean, multiplied by 100%. Because this metric can be heavily skewed by a mean close to 0, any phenotypes whose magnitude did not significantly differ from 0 were excluded from this analysis, resulting in the exclusion of 1 phenotype (pre- to postcontrol ΔHR after angiotensin II [bpm]). An unweighted Deming linear regression was then performed on the set of all male–female pairs of CVs. 95% confidence intervals were calculated for the slope and intercept of the regression line. If the hypothesis that females are more variable than males were true, the regression line would be deflected toward the female axis. The slope of the regression line was compared with a null hypothesis slope of 1, a slope indicating equal CVs for male and female phenotypes. This analysis was repeated on the specific subsets of data as specified by the phenotypic categories given by the PhysGen Program for Genomic Applications, consisting of biochemistry, cardiac, lung, respiratory, renal (consisting of renal A and renal B), and vascular. Phenotypes measured in multiple phenotypic categories were consolidated to one point for the All Phenotypes Deming regression by keeping
the measurement with the greatest ratio of female to male CV (ie, the most female-variable data) and excluding the others; elimination of these phenotypes reduced the number of phenotypes analyzed from 160 to 141. All measurements were used for the Deming regression on the specific phenotypic categories. Deming regression and slope testing were performed in SigmaPlot 12.

**Statistical Analysis**

Individual statistical analyses are described in their respective sections of the methods. Methods were implemented in R 3.0.1, Matlab R2014b, and SigmaPlot12 as indicated. Unless otherwise noted, data are presented as mean±SEM.

**Results**

First, we sought to determine whether estrous cycling per se was necessary for maintenance of normal phenotypic outcomes and variance for MAP, a phenotype of great clinical interest. To determine whether disruption of the estrous cycle caused by surgical procedures (arterial catheter implantation) affects variability, we examined high-salt diet–induced changes in MAP in female Dahl salt-sensitive rats via implanted telemeters. Using vaginal cytometry to identify the stage in the estrous cycle, we partitioned a group of 14 rats into those that cycled 3 to 5 times and rats that cycled 2 or fewer times over 18 days (Figure 1). Using a repeated measures Brown–Forsythe test, we found no significant difference in variance between the 2 groups, nor did we find a significant difference in magnitude (Figure 1; Table S4).

To more thoroughly examine the effects of sex on phenotypic variability, we used publicly available data collected by our department as part of the PhysGen Program for Genomic Applications (pga.mcw.edu). The phenotyping component of PhysGen Program for Genomic Applications provided high throughput, detailed phenotypic data on many consomic and mutant rat strains as well as their respective parental strains. Here, we used data from the consomic panel, which includes 46 consomic strains and 11 commonly used inbred parental strains.

The Brown–Forsythe test was used to examine differences in variance for each of the 142 phenotypes remaining after elimination of values mathematically derived from phenotypes otherwise accounted for in the data. Notably, we chose not to correct for multiple testing, to increase the sensitivity of the approach, and more accurately represent the potential result of a future investigator who will most likely not be measuring hundreds of phenotypes in a single experiment. We found that 74 phenotypes of the 142 tested had a significant sex difference in variance (P<0.05; Table S1 in the online-only Data Supplement). Note that variance was greater in males for 59% (44/74) of these phenotypes, despite measurements being taken from freely cycling females (Table S1). Importantly, this trend is maintained within individual strain comparisons, with greater variance in males in 399 of 601 significantly different phenotypes (Table S3).

Of the 74 traits with significant differences in variance, 62 also had a significant difference in magnitude. To assess the potential impact of variation on experimental design, we performed retrospective power analyses on these 62 traits, with a significant sex difference in both variance and magnitude. For the majority of these phenotypes, the number of rats necessary to detect the sex difference in variance (N_\text{V}) was much larger than that needed to detect the difference in magnitude (N_\text{M}) of the trait (Figure 2). For instance, based on data from the respiratory protocol, cohorts of 34 male animals and 34 female animals would be necessary to detect a difference in variance in body weight with an \(\alpha\) of 0.05 (corresponding to the standard P<0.05 for significance) and a power of 0.8 (indicating a probability of 0.8 for discovering a true difference when such a difference exists). By contrast, only 7 animals were necessary to detect a significant difference in magnitude of body weight.

In total, we found 16 traits that either displayed differences in variance at lower N than differences in magnitude (6) or displayed differences in variance with no corresponding difference in magnitude detected (10; Figure 2). Importantly, the relationship between N_\text{V} and N_\text{M} is maintained within the individual strain comparisons, with only 81 out of 341 strain–phenotype combinations with significant sex differences in both magnitude and variance having smaller N_\text{V} than N_\text{M} (Table S4).

We next sought to determine whether the detected sex differences in phenotypic variance were a consequence of the differences in magnitude of each trait. For example, note that the phenotype of body weight exhibited a significantly different variance between males and females. However, this increase in variance corresponded to an increase in magnitude of the measured trait, that is, male rats were larger than female. This relationship can be described using the CV, defined as the standard deviation for a trait divided by the mean value for that trait multiplied by 100%. Using this statistic, we found that the CVs were similar between male and female rats (9.52% versus 9.99%), suggesting that the difference in variance may be explained by the difference in magnitude.
For each phenotype, we calculated the CV for both males and females. We then plotted each of these pairs and performed an orthogonal Deming regression on the CVs (Figure 3). The slope of this regression line between males and females was not significantly different from 1 ($P=0.58$), which does not support the hypothesis that the female rats display greater variability.

To further examine sex differences in CV, we examined subgroups of phenotypic data based on physiological categories from the PhysGen project (Figure S1). We did not find statistically significant evidence that the slopes of the regressions in renal, respiratory, vascular, and cardiovascular phenotype categories differed from 1. There was a statistically significant difference of the slope of the regression line from 1 in the lung ($P=0.0374$, trending toward female) and biochemistry ($P=0.0001$, trending toward male) phenotype categories.

**Discussion**

Although the scientific community continues to combat male bias in preclinical research, representation of women in late-phase clinical trials has improved markedly in recent years.21 The National Institutes of Health initiative to increase the use of female organisms and cells in preclinical research ensures that the scientific community as a whole will grow to fully address the health of both men and women. A major obstacle to the incorporation of female animals into studies, however, has been the belief that the rodent estrous cycle would render the study of these animals prohibitively expensive or difficult. Our study shows for the first time in a single large data set of physiological measurements that this belief is not accurate.

Our findings show that sex differences in CV among rats are small despite hormone fluctuations because of the estrous cycle. Moreover, for the majority of phenotypes with statistically significant sex differences in both variance and magnitude, $N_M$ was smaller than $N_V$ (56/62). Six traits had $N_V$ greater than $N_M$, and are listed in the table along with the 10 traits that had a difference in variance detected along with no corresponding difference in mean. These traits may form a heretofore understudied scientific field: phenotypes that differ only, or mostly, in terms of variance between the sexes.
of these phenotypes. Indeed, this finding may be because of the fact that male rats are more variable than is commonly appreciated, as pointed out by Prendergast in his excellent meta-analysis. Clearly, the true nature of this variability is a worthy topic of further study. However, this analysis forms the most comprehensive evidence, thus far, that study designs currently acceptable for male animals can also be applied to female animals, aside from a few noted exceptions. This finding will allow us to design studies to find true differences in magnitude of traits between the sexes, which undeniably exist.

As for the phenotypes that differ only, or mostly, in terms of variance between the sexes, these may be of particular experimental interest because they form a heretofore understudied scientific field. This could point to possible differential control systems underlying the sex-specific expression of these phenotypes. As such, they may represent unique opportunities to probe the underlying genetic or molecular systems that control physiological traits. For example, several of the traits related to angiotensin II stimulation showed differences in variance between the sexes. It has been shown that estrogen modifies the expression of both type 1 and angiotensin receptors changes throughout the estrous cycle, suggesting that changing estrogen levels could lead to the increased variability observed here. However, the finding that several traits are more variable in males than in females suggests that this list is not simply a surrogate for traits affected by the estrous cycle. Clearly, further study is needed to understand the differences in variability observed here.

Perspectives

To contribute to male–female equity in preclinical research, which is both a goal and mandate of the scientific community, we sought to test the hypothesis that freely cycling female rats are more variable than male rats. Historically, female animals have been excluded from large portions of the scientific literature on the assumption that females are more variable than their male counterparts. Here, we show that variability of 142 traits in male and female rats is similar, suggesting that differential treatment of males and females for the purposes of experimental design is unnecessary until proven otherwise, rather than the other way around.

Acknowledgments

Alex Dayton and Eric C. Exner performed the statistical design, data analysis, and interpretation and drafted the article. John D. Bukowy, Meredith Skelton, and Timothy J. Stodola contributed to statistical analysis and article editing. Theresa Kurth collected the blood pressure data and performed the estrous cycle cytology. Andrew S. Greene performed statistical design and article editing. Allen W. Cowley edited the article and participated in data interpretation.

Sources of Funding

This work was supported by the National Institutes of Health (NIH) via the National Heart, Lung, and Blood Institute (grant IDs: HL-116264, HL-082798, HL-122662).

Disclosures

None.

References

6. (GAO) USGAO. Drug safety: Most drugs withdrawn in recent years had greater health risks for women. 2001.


Breaking the Cycle: Estrous Variation Does Not Require Increased Sample Size in the Study of Female Rats
Alex Dayton, Eric C. Exner, John D. Bukowy, Timothy J. Stodola, Theresa Kurth, Meredith Skelton, Andrew S. Greene and Allen W. Cowley, Jr

Hypertension. 2016;68:1139-1144; originally published online September 26, 2016; doi: 10.1161/HYPERTENSIONAHA.116.08207
Hypertension is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2016 American Heart Association, Inc. All rights reserved.
Print ISSN: 0194-911X. Online ISSN: 1524-4563

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://hyper.ahajournals.org/content/68/5/1139

Data Supplement (unedited) at:
http://hyper.ahajournals.org/content/suppl/2016/09/26/HYPERTENSIONAHA.116.08207.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Hypertension can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Hypertension is online at:
http://hyper.ahajournals.org//subscriptions/