Spring 2019

The Impact of New NIH Requirements on the Preclinical Research Sex Disparity --- A Meta-Analysis

Nicole McGrath
University of Maine

THE IMPACT OF NEW NIH REQUIREMENTS ON THE PRECLINICAL RESEARCH SEX DISPARITY—A META-ANALYSIS

by

Nicole McGrath

A Thesis Submitted in Partial Fulfillment
of the Requirements for a Degree with Honors
(Biology)

The Honors College

University of Maine

May 2019

Advisory Committee:
Lynn Atkins, Instructor of Biology, Advisor
Michelle Goody, Research Assistant Professor in the School of Biology and Ecology
Sandra Haggard, Associate Professor of Biological Sciences at UMA and
Collaborating Associate Professor of Women's, Gender, and Sexuality Studies
Mark Haggerty, Rezendes Preceptor for Civic Engagement in the Honors College
Alan Rosenwasser, Professor of Psychology

ABSTRACT

Historically, women as subjects have been underrepresented in clinical research. Due to this problem, legislation was enacted by Congress in 1993 to require inclusion of women in NIH funded clinical trials. Female animals are also underrepresented in preclinical research and need to be included to ensure safe and effective drugs. Studies exclude female mammals under the assumption that the estrus cycle contributes to variability (Beery, 2011). This notion has been contradicted by several studies (Prendergast, 2014; Becker, 2016). New requirements of NIH funded researchers to consider sex as a basic biological variable were announced in 2014 (Clayton, 2014). As of June 5. 2016, all NIH grant applications must include plans to use equal numbers of both sexes and to perform statistical analysis for possible sex differences (NIH, 2016).

This study examined the impact that these requirements have had on the inclusion of both sexes and the analysis of sex differences in preclinical research. The fields of neuroscience, pharmacology, and immunology were chosen for analysis based on research indicating that they had the lowest rates of analyzing sex differences prior to the mandate (Beery, 2011). A significant increase in the inclusion of both sexes was found in all fields ($p<0.001$), along with a 3.58 fold increase in the proportion of articles that analyzed sex differences. NIH funded pharmacology research was more likely to include both sexes and report sex difference analyses post-mandate. However, articles still must analyze sex differences and include both sexes at a much higher rate than the current 2018 statistics calculated in this analysis (16.4% and 36.4%).Due to the recentness of the mandate, it is recommended that a follow-up study be conducted. The increases in female inclusion and sex differences analysis are promising signs of future improvement.

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor Lynn Atkins and my committee, Dr. Alan Rosenwasser, Dr. Sandra Haggard, Dr. Michelle Goody, and Dr. Mark Haggerty. I would also like to thank Phoenix Throckmorton for writing the program required to extract data. This work was made possible in part by the support of the Rendle A. Jones ’65 and Patricia K. Jones ’65 Honors Thesis Fellowship.

TABLE OF CONTENTS

TABLE OF FIGURES

INTRODUCTION

Historically, female mammals are excluded from preclinical research due to the assumption that results from males apply to females, and concerns that hormonal cycles increase variability in samples and confound experimental manipulations (Beery, 2011). It is not only female animals that have been historically excluded from research. In 1977, the FDA advised the exclusion of women of child bearing age from drug trials due to possible effects of experimental medications on fertility and pregnancy (US Food and Drug Administration). These recommendations led to underrepresentation of women in clinical research for decades (Beery, 2011). Of drugs withdrawn from the US market from 1997–2000, the US Government Accountability Office reported that 8 out of 10 drugs taken off the market had greater adverse effects in women (Simon, 2005). Without proper representation of women and female animals in research, the general public is put at risk.

The notion of the 78 kg white man as the norm for research is harmful to women and minorities. For example, crash test dummies have been based off the biometrics to the average male for the past thirty years (Bose, 2011). As a result, seat-belted women are 47 percent more likely to have serious injuries in accidents (Bose, 2011). Basing research off of just men has also led to bias in the way that female patients are diagnosed in diseases like coronary heart disease (CHD) and Attention-Deficit/Hyperactivity Disorder (ADHD) (Mikhail, 2005;Walters, 2018). In women, coronary heart disease presents differently than the clinical symptoms that have been classified in men (Mikhail, 2005). CHD kills an equal or higher percentage of women compared to men, yet the general

public still refers to heart health as a predominantly male disease (Mikhail, 2005). Sources also indicate that up to three quarters of women and girls with ADHD go undiagnosed due to symptoms mischaracterized as laziness or introversion (Walters, 2018).

The bias of using men as a model for human disease has also affected the dosage of certain drugs and vaccines (Klein, 2016). Legislation has been in place to increase diversity in US clinical trials since the 1993 NIH Revitalization Act passed by Congress. This legislation has caused the proportion of women in clinical trials to increase to 50% (Beery, 2011). Likewise, as new rules such as the NIH mandate are instituted the hope is that female animal representation in preclinical research will reach 50 percent. The plan to enforce these rules revolves around data mining techniques by the NIH. The NIH aims to keep researchers accountable by monitoring when studies do not include both sexes or analyze sex differences; however, no repercussions for this have been indicated (Clayton, 2014). If the mandate is not enough to increase female inclusion, it may be recommended that legislation be enacted.

In May 2014, the National Institutes of Health announced that the agency planned to ensure that investigators account for sex as a basic biological variable (SABV) in NIH-funded preclinical research (Clayton, 2014). As of June 2016, all NIH grant applications must include plans to use equal numbers of each sex and to perform statistical analysis for possible sex differences (NIH, 2016). This mandate was influenced by a 2011 meta-analysis which revealed a male bias in 8 out of 10 biological disciplines (Beery, 2011). To address the issue of underrepresentation of female animals and analyze the effects of the NIH mandate, a meta-analysis of the most disparate biomedical fields was conducted.

The fields of neuroscience, pharmacology, and immunology were chosen for analysis based on research indicating that they had the lowest rates of analyzing sex differences prior to the mandate (Beery, 2011). It was hypothesized that research funded by the NIH would be more inclusive of sex and have a larger increase in the inclusion of females than research with other sources of funding.

The plan to include sex as a basic biological variable (SABV) in all grant applications was officially implemented for the fiscal year of 2017. The SABV policy requires researchers to factor sex into the design, analysis, and reporting of vertebrate animal and human studies (NIH, 2016). This meta-analysis aimed to assess the current state of sex inclusion in biomedical research. Three of the previously most disparate disciplines identified by Beery (2011) were analyzed: immunology, pharmacology, and neuroscience. According to data from 2009, these three disciplines analyzed for sex differences less than ten percent of the time, the least out of ten biomedical fields analyzed (Beery, 2011). Furthermore, males were specified in 65 percent of pharmacology articles while females were included in only 20 percent, with 15 percent of articles not specifying sex (Beery, 2011). Less than 10 percent of articles in these fields analyzed sex differences (Beery, 2011). Articles in immunology use only females more often than only males. Over 60 percent of articles did not specify what sexes were used (Beery, 2011). For neuroscience, 25 percent of articles did not specify sex and 55 percent of articles used only males (Beery, 2011).

The consequence of using only one sex in preclinical research is that important differences between male and female organisms are being neglected. Estrogen and testosterone play important roles in both sexes' bodies, however the levels and roles of

these hormones are different between men and women. For example, in males testosterone is first converted into estradiol by aromatase to masculinize the brain. Estrogen affects a wide variety of processes including neuronal differentiation and innate immunity. Due to the effects of estrogen, women have more reactive innate immune systems than men, leading to more adverse reactions to vaccines (Klein, 2016).

The pressure on researchers to produce statistically significant results may have an impact on how they approach the topic of choosing animals to include in their analysis. Some have argued that using only male animals increases reproducibility due to the estrus cycle of female mammals possibly adding variance in results. This notion has been contradicted by four studies of mice and rats (Becker, 2016; Prendergast, 2014; Itoh, 2015; Meziane, 2007). A meta-analysis of neuroscience articles indicated that studies that used both male and female rats, with no regard to estrus staging, were no more variable than studies that used only males (Becker, 2016). P values ranged from 0.6 to 0.95 when conducting a standard t-test on the trait variance of neurobehavioral measures in males versus females (Becker, 2016).

A meta-analysis of whether monitoring the estrous cycle reduces variability of results found that females at any point in estrus were no more variable than males. This opposes the idea that if females are included in research, they need to be studied at the four different phases of estrous (Prendergast, 2014). This study also found that group housing increased trait variability by 37 percent. This is a much larger amount of variability added to experimental designs than including both sexes. Furthermore, it should be noted that individual differences create variability in experimental results. Studying both sexes is just one way to increase the generalizability to a population. For

example, when studying certain antidepressants in males, the results do not have any predictive relevance to the clinical outcomes of females (Koren, 2014).

Lastly, a meta-analysis of microarray gene expression datasets (5 million probes), found that in both mice and humans females were very slightly less variable than males (Itoh, 2015). However, when looking a tissue specific microarrays, males were more variable in spleen tissue and females were more variable in adrenal tissue. These findings support the opinion that one sex is not more variable than the other. On the contrary, sex differences are ubiquitous and need to be examined on a molecular, cellular and organismal level.

Sex-based differences in disease are consequences of X chromosome inactivation, differences in quantities of hormones, and differences in anatomy. Immunology may be the field with the most pronounced sex differences out of the three. There are 60 known genes carried on the X chromosome that are involved in immune regulation (Klein, 2016). X chromosome inactivation likely is the largest factor when it comes to sex differences in immunity (Klein, 2016). Females have higher expression of genes on the X-chromosome which include immune markers like FoxP3 and CD40L. Females produce higher Th2 response and antibodies. This leads to better protection from infections but their hyperimmune responses increase susceptibility to autoimmune diseases (Taneja, 2018).

Klinefelter's syndrome, which occurs when males have an extra X chromosome, leads to many immunological changes. This condition results in low testosterone, increased gonadotrophins, and elevated estrogen concentrations (Ko'ar, 2000). Due to these hormonal and cellular differences, men with Klinefelter's syndrome respond with

higher immunoglobulin concentrations, $CD4^+$ T cell numbers, CD4/CD8 T cell ratios, and B cell numbers than XY males (Ko'ar, 2000). Furthermore, the immunological effects of Klinefelter's syndrome are reversed by testosterone therapy. This illustrates the important role of hormones in immunity. Moreover, women with Turner syndrome (nondisjunction error resulting in X0 instead of XX) have lower IgG and IgM levels and lower T cell and B cell levels compared to XX females (Klein, 2016). Interestingly, both patients with Klinefelter's syndrome and patients with Turner syndrome show increased development of autoimmune disease. This illustrates the major role of the X chromosome in susceptibility to autoimmune diseases (Klein, 2016). Estrogen also plays a major role in autoimmune diseases such as Grave's disease, systemic lupus erythematosus, and multiple sclerosis (Klein, 2016).

In general, both the proportion of individuals infected and the severity of infection are higher in males than females for viral, bacterial, fungal, and parasitic diseases (Klein, 2016). Pro-inflammatory cytokine responses, T cell proliferation, and antibody responses are also greater in female mice than male mice (Klein, 2016). Furthermore, there are variations in estrogen and progesterone levels during the different phases of the menstrual cycle. These hormones influence t-helper 1, t-helper 2, and t- regulatory cell populations. T-reg and T_H2 cells are associated with peaks in estrogen while T_H1 cell levels drop during estrogen peaks (Klein, 2016). However, predictable hormonal changes do not make females any more variable than males (Becker, 2016). The sex differences in immunology are diverse and ubiquitous throughout the body. Research on both sexes is imperative to fully understand these differences.

Sex differences exist in every part of the brain, including the hippocampus, amygdala, and neocortex (Andreano, 2009). Functional cerebral asymmetries are more common in males than in females, meaning that functions such as speech or facial recognition are lateralized to one side (Killgore, 2001). When men perceive happy faces versus fearful faces, amygdala activation is lateralized to either side according to fMRI data. However, when women see the same image there is no significant difference in the areas of the amygdala activated (Killgore, 2001).

It is no secret that there are psychosocial, language, and memory differences between men and women (Andreano, 2009). Although some of these differences can be attributed to hormones, there are ubiquitous sex differences throughout the nervous system, from the anatomical to subcellular level. The gonadal steroid estrogen has been shown to affect neuronal growth, differentiation and survival at every point in development (Abel, 2010). Research has also shown that estrogen protects cortical neurons from glutamate toxicity (Singer, 1996).

The male brain is masculinized by testosterone being converted to estradiol and then crossing the blood brain barrier. Mechanisms for sex differences in the brain have been proposed for both hormonal and molecular factors. Sex differences have been found at the transcriptional level in zebrafish brains (Lee, 2018). Brain aromatase, prostaglandin 3a synthase, and prostaglandin reductase 1 are among the genes with sexually dimorphic expression patterns. Furthermore, seven mouse genes have been found to show differential expression between the developing brains of male and female mice at stage 10.5 days post coitum (dpc), before any gonadal hormone influence (Dewing, 2003). It is believed that those genes are integral in brain sexual differentiation as determined by

chromosomal sex. While my meta- analysis focused on in-vivo sex differences, it is important to remember that sex differences can be apparent in in-vitro tissues and cell cultures as illustrated in the prior study.

The consequences of failing to include sex-based differences in study design and analyses has effectively led to treatment regimens that are identical for both men and women. As a result of this bias, differences in drug efficacy and adverse effects reportedly led to the withdrawal of eight out of ten prescription drugs from the United States market in 2005, specifically owing to health issues in women (Simon, 2005). Critics of sex-specific analysis claim that conducting scientifically rigorous trials with enough statistical power to detect sex differences is inefficient in terms of time as well as cost. Nevertheless, when prescription drugs are withdrawn from the U.S. market because they cause greater health risks for women than men, the cost of not doing such analyses becomes a greater liability for drug companies.

Arrhythmias due to atypical antipsychotics were found to be much more common in women (Aichhorn, 2007). According to the FDA's good laboratory practices for preclinical research, which provides the basis for toxicity and dosing, it is not required that research be conducted on both sexes of animals to be brought to clinical trials. Only guidance for the 'Animal Rule'—which allows experimental drugs to surpass the clinical trial process if it is unethical to test on humans—requires testing on both sexes. However, in the normal industry standards, there is no requirement for this. Not testing both sexes can lead to adverse outcomes in the understudied sex during and after the clinical trial process.

There are sex differences other than adverse effects caused by drugs. Differences in drug metabolism and absorption rate of certain drugs have been found. This has been hypothesized to be due to X chromosome inactivation of certain CYP genes, which play a major role in drug metabolism. The absorption rate and extent of a drug are drug-specific so it is difficult to pick out blanket mechanisms different in men and women. Examples of drugs that illustrate sex differences in drug absorption include, rifampicin, and IM cephradine (Soldin, 2009). Increased absorption of rifampicin is seen in women (Gorksi, 2009). Lower bioavailability and absorption of intramuscular cephadrine was observed in women. Furthermore, it is postulated that women, due to possessing larger amounts of subcutaneous lipid content, receive different doses of transdermally administered drugs. Additionally, women have greater respiratory minute ventilation and lower tidal volume, which may result in decreased ingestion of inhaled aerosol drugs (Soldin, 2009).

Female enterocytes express significantly different levels of CYP3A isoenzymes than males, which contribute the metabolism of many orally administered drugs (Soldin, 2009). Drug rate of absorption is influenced by multiple factors, including gut transit times, lipid solubility and molecular weight of the drug, along with pH and motility of the gut. Gastric fluids are more acidic in males than females. Reduced pH results in decreased absorption of weak acids and increased absorption of weak bases. Transit times are significantly different in men and women. Mean transit times are shorter in men than in women; 44 hours compared to 92 hours (Soldin, 2009). It has also been found that studying the dosage of certain antidepressants in males is completely irrelevant to the clinical outcomes in females (Koren, 2014). Studying sex differences in all pharmacological studies could lead to better clinical outcomes for women.

In summary, a meta-analysis of three disparate fields identified by Beery was conducted. Neuroscience, immunology and pharmacology all have widespread sex differences which are worth analyzing in papers that are relevant to both men and women. It was hypothesized that the inclusion of both sexes and sex differences analysis would increase after the enactment of the mandate in 2016. Furthermore, it was expected that NIH funded research would include both sexes and analyze sex differences more often than research with other sources of funding.

METHODS

Published papers in nine peer reviewed journals were evaluated for several variables (See Table 1). The fields neuroscience, pharmacology, and immunology were chosen to be the targets of the research based on Beery and Zucker's 2011 paper illustrating that these three fields had the largest discrepancies in the use of female versus male animals. Three major journals from each field were selected for review. Then the choice was made to use solely mice papers because mice are the most common model organism for studying human disease (Spencer, 2002). Only papers dated from 2009 to 2018 were used.

Table 1: Number of Articles per Field

Field	Neuroscience	Pharmacology	Immunology
Journals	Neuron: 471 Neuroscience: 357 Nature Neuroscience: 258	Journal of Medicinal Chemistry: 229 British Journal of Pharmacology: 81 Neuropsychopharmacology: 169	PloS Pathogens: 324 Vaccine: 169 Nature Immunology: 275
Total	1082	474	766

Mice were chosen to be the organism of interest for this study due to immunology using primarily mice as subjects (Beery, 2011). Furthermore, the use of mice as a model organism has drastically risen since 1990. The mouse has surpassed the rat as the most popular animal model (Spencer, 2002). Because in this meta-analysis, there was particular attention paid to the role of hormones and their epigenetic consequences, studies sacrificing embryos and mice younger than 4 days old were not included in the

analysis due to minimal sex characteristics present (Schlomer, 2013). Mice are altricial which means that they do not start secreting hormones until after birth, unlike humans which sexually differentiate in-utero. Therefore, it can be assumed that most sex differences other than on the cellular level will not be apparent. It can also be inferred that embryonic research uses both sexes due to the lack of most defining sexual characteristics until postnatal day 4-5 (Schlomer, 2013).

The following inclusionary and exclusionary criteria was then applied to the search (see Table 2). Data for the nine journals were mined from PubMed. A targeted search was performed to find articles that met the inclusionary and exclusionary criteria listed in Table 2. The search can be found in Appendix C. A program was written to extract the title, data, authors, and affiliation of the first author. Another program was written to separate American articles from international articles, in order to control for other policies that would affect the proportion of sexes used. After the valid articles' information was imported into an excel sheet, each was analyzed by hand for possible excluding characteristics and for the three variables gathered.

Table 2: Inclusionary and Exclusionary Criteria.

Inclusionary	*Exclusionary*
• USA affiliated first author • Uses mice • Relevant to human health • In-vivo and ex-vivo experiments	• Used animals other than mice • ≥90% one gender affected (i.e. breast cancer) • Conditions involving the gonads or genitals • X-linked conditions • BALB/c and Nu/Nu mice • Mice under 5 days old

The search was limited to the United States due to the NIH only funding US researchers. In-vivo and ex-vivo experiments were particularly of interest due to hormonal factors affecting cells being the primary reasoning of researchers to not use females. BALB/c mice were excluded due to unusual male aggression seen in that strain of mice leading to more females used. Nude mice were also excluded due to diminished reproductive capability of homozygous females, leading to more readily available female animals. Embryonic and neonatal mice under the age of P5 were excluded due to most sexual differentiation not starting until postnatal day 5 due to their lack of precocial hormone secretions.

Papers that did not include any in-vivo experiments, studied diseases that affected primarily one gender, involved reproduction, involved X-linked diseases or mutations, or researched urogenital diseases were all excluded from the PubMed search. The search can be found in Appendix B along with the program used to extract data. The population of papers using mice from the last ten years in the 9 journals totaled 14,819 articles (Figure 1). The full population of biomedical research using mice between 2009 and 2018 was narrowed down to 2322 articles across nine journals and 3 fields known for their disparate use of sexes (Figure 1).

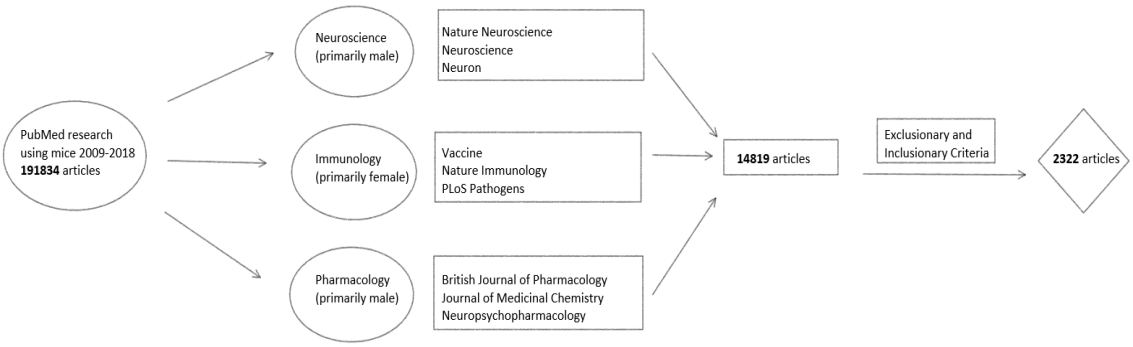


Figure 1: Applications of Search Criteria.

According to the NIH's new policies: "Applicants must provide strong justification for applications proposing to study only one sex. Such justification may include the study of sex-specific conditions or phenomena (e.g., ovarian or prostate cancer), acutely scarce resources (e.g., non-human primates), or investigations in which the study of one sex is scientifically appropriate. The absence of evidence regarding sex differences in an area of research does not constitute strong justification to study only one sex." These guidelines served as a basis for whether or not a study using only one sex would be included. Statements such as, "prior research showed that female animals consume more alcohol and therefore were excluded" were not considered to be valid justification.

Due to availability of athymic mice and greater variation in group housed BALB/c male mice, immunology is generally biased to use more females. BALB/c mice are well known for their ability to produce plasmocytomas upon injection with mineral oil. This is a reliable way to study monoclonal antibodies. Because of this unique characteristic and the prevalence of abnormal male aggression skewing results, articles using BALB/c strains were considered to have sound justification for using one sex and therefore were excluded from analysis.

Furthermore athymic (nu/nu) mice were excluded from this study. This was due to the fact that all athymic nude females have extremely reduced reproductive capability. Moreover, the most humane way to study T cells requires the use of this particular breed of mice. In turn, this leads to a surplus of homozygous females to use because the most effective way to breed would require a homozygous male and a heterozygous female. This unfortunately limits the usefulness of data obtained in immunology due to two of the most commonly used strains being excluded. The results likely are skewed to favor more sex differences analyses and more inclusion of male mice in studies. The categorization of variables veered from Beery's paper in that a definite cutoff was proposed for the use of sexes to be roughly equal rather than assuming equal representation if they used both sexes at all. A cutoff was also set for specification but this rarely mattered as most studies either specified sex for all of their experiments or for none at all (Table 3). Furthermore, the exact percentage of studies that did not uniformly use the same sexes throughout experiments is not available, but can be assumed to be roughly 8 percent of articles based on data from the journal Nature Neuroscience.

Table 3: Description of Variables Analyzed and Classification.

Definition of Variables

Variable	*Categorization*	*Definition*
NIH Funding	Yes/No	Determined by the acknowledgement of such in the article.
Analysis	Yes/No	Indicated that there were no observed sex differences or described statistical tests performed on sex
Sex	Male	Used only males in 25% or more of the experiments and both sexes or unspecified in the other 75% Used males as controls and both as experimental animals or the opposite
	Female	Used only females in 25% or more of the experiments and both sexes or unspecified in the other 75% Used females as controls and both as experimental animals or the opposite
	Both	Used roughly equal parts males and females in more than 75% of experiments Used mostly one sex but repeated the same exact study with a smaller N of the other sex
	Unspecified	Did not specify sex in more than 25% of experiment types and used both in the other experiments Unspecified in more than 75% of experiments and used only males or only females in in less than 25%
	Other	Any instances that did not fit the above categories. Such as, a study used males as controls and females as experimental animals

For the large majority of articles (91.8 percent of Nature Neuroscience articles), the sexes used would be specified in the methods section and would be uniform throughout the study. A typical phrase found in the methods section would be: “Males and females were used throughout this study.” However, females sometimes were excluded for particular types of experiments such as behavioral measures. This is why

there was a need for a specific cutoff rather than marking a study as ‘both’ if it included both males and females at some point in the study.

Table 3 describes the cutoff percentage of experiments that used both sexes for it to be considered approximately equal representation. The percentage of experiments as stated above was only based on experimental types, not the individual N of males and females. Only in-vivo experiments were included, if the study also performed in-vitro experiments. Note that the ‘other’ category only contained 10 out of 2322 articles, mostly within the pharmacology category.

Articles that met the inclusionary and exclusionary criteria analyzed for the following variables: NIH funding, what sexes were used, whether or not they analyzed for possible sex differences, and if they were published before or after the mandate. Chi-square tests were performed in SPSS (Statistical Package for the Social Sciences by IBM) along with Pearson correlations in Excel.

RESULTS

There have been significant improvements across fields in respect to reporting sex differences analyses and to including both males and females. Figure 2 illustrates the proportional differences in what sexes were used prior to the mandate versus how the proportions changed after it was issued. Neuroscience, immunology, and pharmacology all had significant increases in the amount of articles that used both males and females after the mandate was issued ($p<0.001$). Note that chi square statistical analyses were used. A limitation of this type of statistical test is that with a very large N, results may be reported as very significant with a relatively small change.

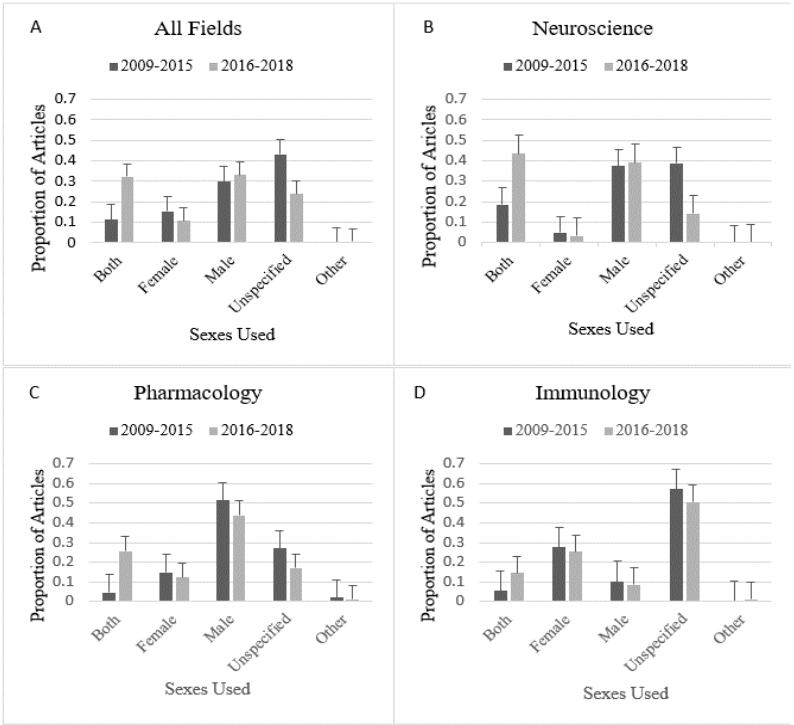


Figure 2: Proportion of Each Sex Used by Field Pre- versus Post-Mandate
A. The figure indicates a significant increase in the use of both sexes in a given article ($p<0.001$) and a significant decrease in the amount of articles that do not specify sex ($p<0.001$). This data combines all three fields. B. In neuroscience there was a significant increase in the use of both sexes and a significant decrease in the amount of unspecified carticles ($p<0.001$). C. In pharmacology there was a significant increase in the amount of articles that used both males and females ($p<0.001$) D. There was also a significant increase in the use of both sexes in immunology ($p<0.001$).

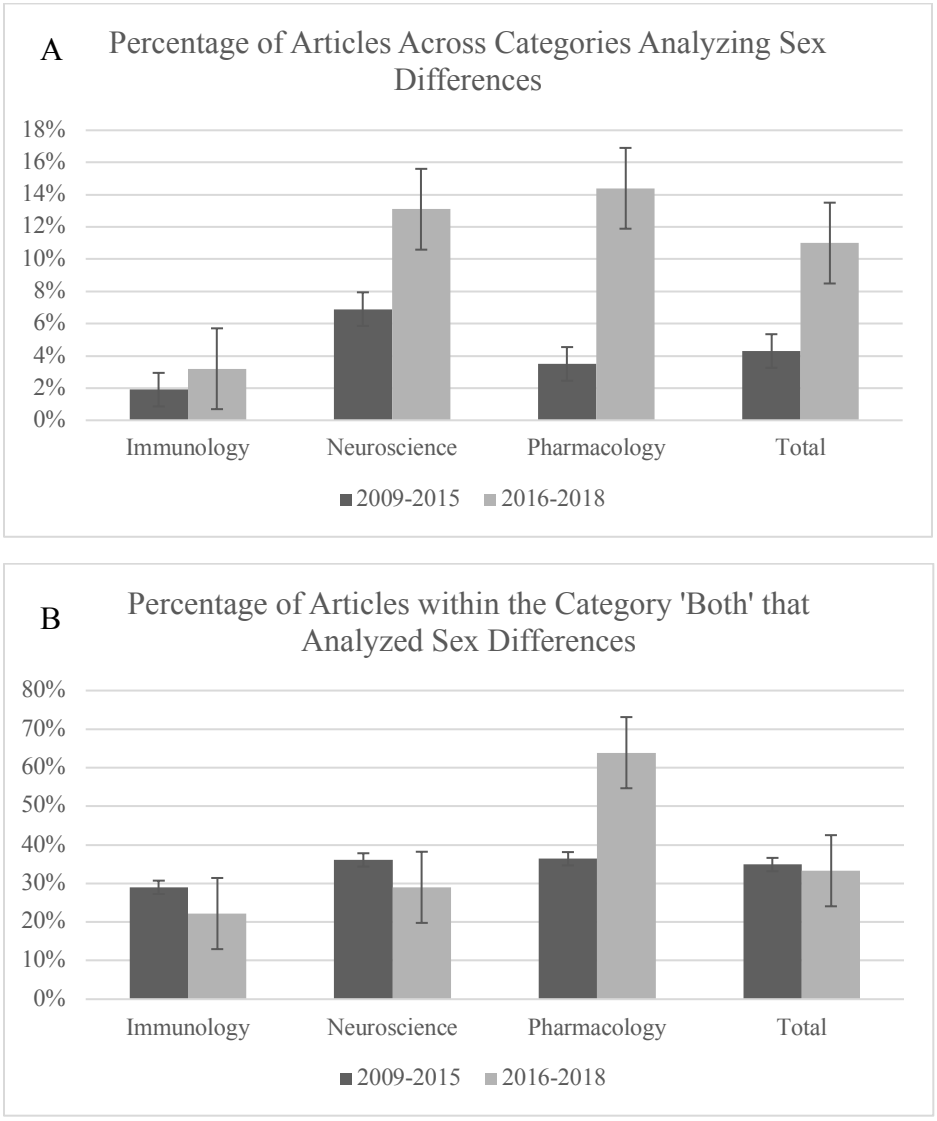


Figure 3: Percentage of Articles Analyzing Sex Differences

A. This figure indicates the percentage of articles per field that analyzed sex differences. This is subdivided into prior to 2016, when the regulations enacted, and 2016-2018. Immunology articles analyze sex differences at a significantly lower rate than neuroscience and pharmacology ($p<0.001$). Neuroscience and pharmacology both saw significant increases in the percentage of articles that analyzed sex differences while Immunology only increased by 1 percent.

B. This graph only takes into account articles that specified that they used both sexes in roughly equal proportions. There isn't a significant change overall or within neuroscience and immunology. However, pharmacology showed a remarkable increase in sex differences analyses after 2015 ($p<0.001$). Note that the percentages in B are proportional to the percentages in A. However the number of articles that used both males and females is smaller than the total within each field, thus increasing the percentages found in B.

Figure 3 compares the proportion of articles that analyze sex differences prior to the mandate versus after its enactment. There was a significant increase in the overall number of articles that analyzed sex differences after the mandate ($p<0.001$) (Figure 3A). However, when looking only at articles that used both males and females, the percentage of articles that analyzed sex differences did not change in any field but pharmacology ($p<0.001$) (Figure 3B).

The difference between 3A and 3B is that 3A analyzed all articles within a given field while 3B analyzes the subset of articles that are labeled as 'both'. For example in neuroscience for 3A, 47/653 neuroscience articles published before 2016 analyzed sex differences and 57/ 429 articles published between 2016 and 2018 analyzed sex differences. Meanwhile in 3B, 44/122 articles that used both males and females analyzed sex differences articles prior to 2016. 54/186 articles that used both males and females analyzed sex differences in 2016 onward. It is speculated that the difference between 3A and 3B is due to the increase in proportion of articles that use both males and females in neuroscience (seen in figure 2B). The proportion of neuroscience articles labeled as 'both' increased from 18.6% to 43.4% (Figure 2B). The overall increase in the percentage of articles that used both sexes is strongly correlated with the increase in the overall percentage of articles that analyzed sex differences ($r= 0.822$).

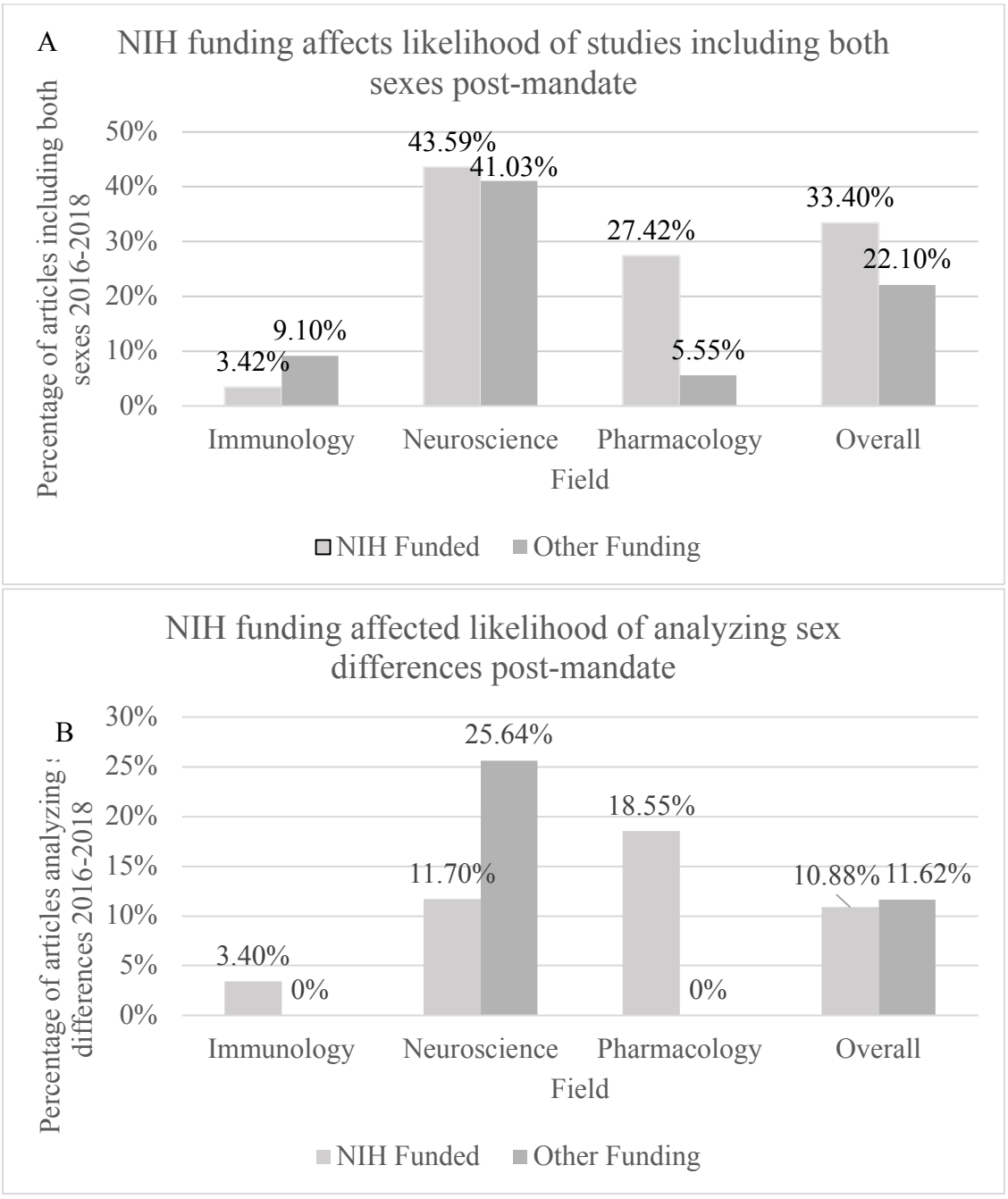


Figure 4: NIH funding Increased the Likelihood of Sex Differences Analysis and Sex Inclusion in Pharmacology

A. A slightly larger percentage of NIH funded neuroscience articles included both males and females. A significantly larger proportion of NIH funded pharmacology articles included both males and females ($p=0.001$)

B. NIH funded studies had a larger amount of studies that analyzed sex differences in immunology ($p>0.05$) and pharmacology ($p=0.005$). However in neuroscience, the inverse was seen ($p=0.014$).

It was hypothesized that NIH funding would affect the inclusion of sexes and analysis of sex differences. These relationships are indicated in Figure 4A and 4B. Pharmacology seems to be the field that was the most positively affected by the NIH mandate. In Figure 4B, significantly more NIH funded articles analyzed sex differences in pharmacology after the 2016 mandate ($p= 0.001$). This relationship was not seen for any field in articles published in 2009-2015 (not shown in a figure). However, for neuroscience a negative relationship was seen between NIH funding and sex differences analysis. This might be due to a much smaller number of articles lacking NIH funding in Neuroscience compared to the other two fields, especially post 2016. For Figure 4A, a slightly larger percentage of NIH funded neuroscience articles included both males and females. A significantly larger proportion of NIH funded pharmacology articles included both males and females ($p=0.001$).

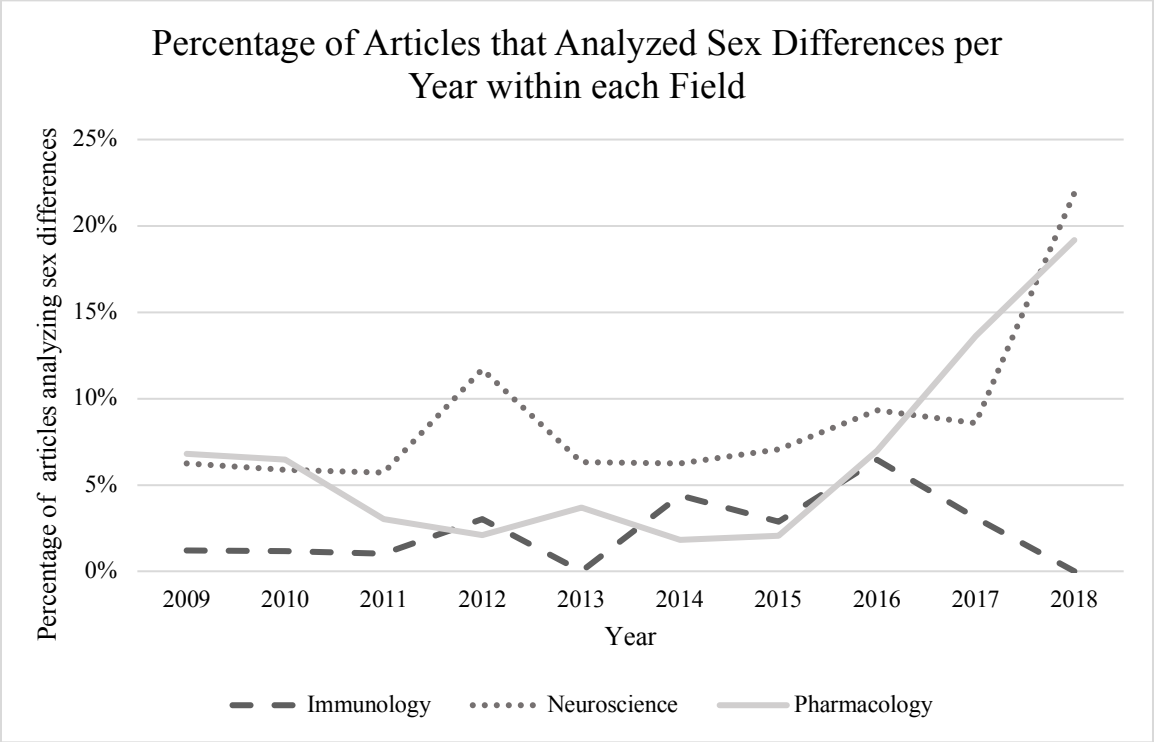


Figure 5: Percentage of Articles that Analyzed Sex Differences per Year per Field

The percentage of papers that analyzed for sex increased by 3.56 fold between 2015 and 2018 and overall by 3.78 fold. There is a visible peak in sex differences analysis in 2012, the year after Beery's publication, and there is also an increase in analysis in 2016, the year that the plan was announced. Sex analysis doubled in 2018 compared to 2017. However there was no overall change in the proportion of articles that used both sexes to analyze sex differences. The large increase in Figure 2A is likely due to a generalized increase in the use of both sexes.

The percentage of papers that analyzed for sex increased by 3.56 fold between 2015 and 2018 and overall by 3.78 fold, which can be seen in Figure 5. There is a visible peak in sex differences analysis in 2012, the year after Beery's publication, and there is also an increase in analysis in 2016, the year that the mandate was enacted. Furthermore, sex analysis doubled in 2018 compared to 2017. The upward trend seen in Figure 5 provides promise that there will be even more progress in the coming years as more articles affected by the mandate are published. It can be speculated that much of the change seen after 2016 was done under researchers' own volition rather than in

accordance to the mandate. The large majority of research directly affected by the FY 2017 regulations had not yet been published which indicates that researchers heard about the new mandate and followed it even if it was not directly applicable to their research.

DISCUSSION

In 1977, the FDA advised to exclude women of child bearing age from drug trials due to possible effects of experimental medications on fertility and pregnancy. These recommendations led to underrepresentation of women in clinical research into the 1990's. Prior to the NIH Revitalization Act in 1993, clinical trials were not required to study the effects of a medication on both men and women. After this Act was passed, inclusion of women has increased to fifty percent (Department of Health and Human Services, 2012). Of drugs withdrawn from the US market from 1997–2000, the US Government Accountability Office (GAO) reported that 8 out of 10 drugs taken off the market had greater adverse effects in women. The Food and Drug Administration does not require researchers to perform preclinical research on both sexes of animals to bring a drug to trial. This is an important topic to monitor as industries try to get the most significant results in the most cost effective manner.

The NIH first announced plans to require researchers to analyze sex differences and include both sexes in 2014. The official requirements were rolled out June 5, 2016. Before this meta-analysis, there had not been a study that analyzed the relationship between National Institutes of Health funding and the inclusion of both sexes in preclinical research. Furthermore, a follow up to Beery's 2011 paper analyzing sex inclusion across fields has not been published. This meta-analysis aimed to quantify the progress that has been made since Beery's publication of 2009 data and define a relationship between government funding and sex difference analysis. It was hypothesized that there would be an increase in the inclusion of both sexes and increased

analyses of sex differences over time. Furthermore, it was predicted that NIH funded articles would be more likely to be inclusive of both sexes and analyze for sex differences.

There are widespread sex differences in humans, many of which are conserved across species. It is important for scientists conducting preclinical research to take into account these differences when applying their research to humans. To understand the biology of women or develop safe treatments for diseases of women one must do more than study men and male animals. Pharmacology, immunology, and neuroscience in particular are fields with important sex differences that need to be studied.

Studying sex differences is important for neuroscience due to their presence across cellular, anatomical, and behavioral levels. These differences include increased lateralization in males, differential gene expression independent of hormones in the brain, and neuroprotective effects of the gonadal steroid estrogen (Killgore, 2001; Cahill, 2006; Koren, 2014). Furthermore, women are predisposed to the neurodegenerative disease multiple sclerosis (MS) which falls under immunology and neuroscience (Klein, 2016). Estrogen has been hypothesized to play an important role in autoimmune diseases such as MS and Grave's disease (Klein, 2016). Furthermore, women have more active immune systems which can lead to adverse effects of vaccines and other immunological drugs. In Pharmacology, many metabolism genes are differentially expressed based on cellular sex (Soldin, 2011). Lower gastric pH is also common in men which can contribute to enhanced or reduced drug absorption (Soldin, 2009). In summary, sex differences are due to X chromosome inactivation, anatomical differences, and exposure to different levels of

sex hormones. The differences previously mentioned have the capacity to affect the health of women and men if they are not properly analyzed in preclinical research.

According to the performed analyses, there was a significant overall increase in the number of articles that analyzed sex differences after 2015 ($p<0.001$, Figure 3A). However, when only including articles that included both sexes, there was only an increase in sex differences analyses for pharmacology (Figure 3B). Furthermore, there was an overall increase in the inclusion and specification of sexes ($p<0.001$). The inclusion of sexes and analysis of sex differences varied across the three examined fields.

All fields showed significant improvements in the inclusion of both sexes after the mandate was enacted (Figure 2 A-D). Neuroscience also showed a significant increase in the number of articles that specified sex (Figure 2B). While it is promising that these fields have moved in the right direction in terms of sex inclusion, immunology still only included both sexes in 14.5 percent of research between 2016 and 2018 (Figure 2D). Neuroscience showed the largest improvement in inclusiveness, and used both sexes in 42 percent of research between 2016 and 2018.

Neuroscience and pharmacology improved in regard to sex differences analyses over time (Figure 5). The amount of articles that analyzed sex differences in these two fields doubled between 2017 and 2018. In respect to the proportion of articles that analyzed sex differences within the 'both' category, only pharmacology articles showed an improvement ($p<0.001$; Figure 3A-B).

Immunology did not improve in respect to specification of sex or in analysis of sex differences even though the field has started to include more males. It is possible that

because the field primarily uses females, researchers do not believe that the mandate applies to them. Furthermore, pharmacology was the only field that had an interaction between NIH funding and the likelihood of analyzing sex differences and including both sexes post 2015 (Figure 4A and B; $p<0.001$). This could mean that pharmacology, in respect to sex analysis, was most affected by the mandate.

While both pharmacology and neuroscience have improved dramatically in comparison to 2009, there is still a long way to go. For 2018, the overall rate of articles that analyzed for sex differences was only 16.4 percent. Increasing that statistic to near one hundred percent in relevant articles would be ideal. Furthermore, almost seventy percent of articles are still using either one sex or not specifying what sexes were used. This is unacceptable unless this statistic continues in the positive direction. The only current repercussion issued by the NIH is that grant applications will be rejected if they do not address sex differences in their proposed experimental design. I propose that further repercussions such as grant repayment, and future disqualification from funding should be taken if it is found that a researcher failed to address sex differences in their publication as it was stated in their grant application.

Several limitations are present in this data. This meta- analysis only focused on papers that included in-vivo experiments. However, it is important to acknowledge cellular and genetic sex differences found in animals that are independent of hormones. Furthermore, while mice are the most commonly used model organism, many pharmaceutical companies use multiple species, leading to a possibly less representative population of articles. Mice also comprise fifty percent of the model organisms used, so this research may not be representative of the other half of preclinical research. Lastly,

this analysis only accounts for two years after the implementation of the new NIH policy, so it is possible that research conducted prior to fiscal year 2017 is being published with no requirements to include both sexes. Furthermore, several discrepancies in 2009 data were found between this meta-analyisis and Beery's paper. According to her paper, the rate of specification of sex in 2009 was 75, 80, and 35 percent for neuroscience, pharmacology, and immunology respectively (Figure 8; Beery, 2011). My analysis resulted in a specification rate of 50, 75, and 40 for these fields in 2009. This is justifiable since hers study used more types of animals than just mice and a less strict classification system than the one used in this study (Beery, 2011).

It was hypothesized that an increase in the inclusion of both sexes and sex differences analysis would increase after the enactment of the mandate in 2016. Furthermore, it was expected that NIH funded research would include both sexes and analyze sex differences more often than research with other sources of funding. All fields did significantly increase in the proportion of articles that used both sexes post-mandate (Figure 2; $p<0.001$). No relationship was found between NIH funding and sex differences analysis or sex inclusion in immunology. For pharmacology, the presence of NIH funding post 2015, increased the likelihood that a given article would analyze sex differences and include both sexes (Figure 4A-B). This relationship was not seen in pharmacology for articles published between 2009 and 2015. This finding supports the hypothesis that an increase in sex differences analysis and sex inclusion would be seen after the mandate's enactment.

While there were significant improvements seen in the three fields studied, it should be noted that the mandate itself reflects requirements for grants rather than papers

published. Furthermore, most research funded by grants accepted during the fiscal year of 2017 has not been published yet. It could be that the improvements seen throughout the years are due to a changing opinion within the scientific community of studying sex differences. This is supported by the finding that more research funded by other sources than the NIH in the field of neuroscience analyzed sex differences post 2015 (Figure 4B). If the new results are due to a changing environment, hopefully in the coming years, as more research directly affected by the mandate is published, there will be even larger increases in sex differences analyses and sex inclusion.

In conclusion, there are many widespread sex differences present in the fields of pharmacology, immunology, and neuroscience. In order to ensure proper medical treatment of women, these differences should be addressed in preclinical research. This analysis indicates that the number of articles that use both sexes in their research has increased since the implementation of the new policies. There have also been increases in the specification of sex and the analysis of sex differences. The current statistics of articles analyzing sex differences (16.4 percent), and articles including both sexes (36.4 percent) are unacceptable and need to increase in the coming years. A positive trend could continue into 2019 and beyond, due to the fact that researchers funded in 2016 and after likely have not yet published their research. It is possible that there could be a significant uptick in articles using both sexes in the coming years. This should be monitored and it is recommended that a similar study be conducted by 2022.

REFERENCES

Abel, K. M., Drake, R., & Goldstein, J. M. (2010). Sex differences in schizophrenia. *International Review of Psychiatry*, *22*(5), 417–428. https://doi.org/10.3109/09540261.2010.515205

Aichhorn, W., Whitworth, A. B., Weiss, E. M., Hinterhuber, H., & Marksteiner, J. (2007). Differences between men and women in side effects of second-generation antipsychotics. *Der Nervenarzt*, *78*(1), 45–52. https://doi.org/10.1007/s00115-006-2112-0

Andreano, J. M., & Cahill, L. (2009). Sex influences on the neurobiology of learning and memory. *Learning & Memory*, *16*(4), 248–266. https://doi.org/10.1101/lm.918309

Becker, J. B., Prendergast, B. J., & Liang, J. W. (2016). Female rats are not more variable than male rats: a meta-analysis of neuroscience studies. *Biology of Sex Differences*, *7*(1), 34. https://doi.org/10.1186/s13293-016-0087-5

Beery, A. K., & Zucker, I. (2011). Sex bias in neuroscience and biomedical research. *Neuroscience & Biobehavioral Reviews*, *35*(3), 565–572. https://doi.org/10.1016/J.NEUBIOREV.2010.07.002

Bose, D., Segui-Gomez, M., & Crandall, J. R. (2011). Vulnerability of female drivers involved in motor vehicle crashes: an analysis of US population at risk. *American Journal of Public Health*, *101*(12), 2368–2373. https://doi.org/10.2105/AJPH.2011.300275

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience*, *7*(6), 477–484. https://doi.org/10.1038/nrn1909

Clayton, J. A., & Collins, F. S. (2014). Policy: NIH to balance sex in cell and animal studies. *Nature*, *509*(7500), 282–283. https://doi.org/10.1038/509282a

Department of Health and Human Services. (2012). *Monitoring Adherence to the NIH Policy on the Inclusion of Women and Minorities as Subjects in Clinical Research*. https://orwh.od.nih.gov/sites/orwh/files/docs/Inclusion-ComprehensiveReport-FY-2011-2012.pdf

Freeman, A., Stanko, P., Berkowitz, L. N., Parnell, N., Zuppe, A., Bale, T. L., … Epperson, C. N. (2017). Inclusion of sex and gender in biomedical research: survey of clinical research proposed at the University of Pennsylvania. *Biology of Sex Differences*, *8*(1), 22. https://doi.org/10.1186/s13293-017-0139-5

Institute of Medicine. (1994). *Women and Health Research*. https://doi.org/10.17226/2304

Itoh, Y., & Arnold, A. P. (2015). Are females more variable than males in gene expression? Meta-analysis of microarray datasets. *Biology of Sex Differences*, *6*(1), 18. https://doi.org/10.1186/s13293-015-0036-8

Jackson Laboratory. (n.d.). 000651 - BALB/cJ. https://www.jax.org/strain/000651

Killgore, W., & Yurgelun-Todd, D. A. (2001). Sex differences in amygdala activationduring the perception of facial affect. *Neuroreport*, *12*(11), 2543–2547. Retrieved from https://oce-ovid-com.prxy4.ursus.maine.edu/article/00001756-200108080-00050/HTML

Klein, S. L., & Flanagan, K. L. (2016). Sex differences in immune responses. *Nature Reviews Immunology*, *16*(10), 626–638. https://doi.org/10.1038/nri.2016.90

Ko'ar, I. H., Yesilova, Z., Ozata, M., Turan, M., Sengul, A., & Ozdemir, I. ?. (2000). The effect of testosterone replacement treatment on immunological features of patients with Klinefelter's syndrome. *Clinical and Experimental Immunology*, *121*(3), 448–452. https://doi.org/10.1046/j.1365-2249.2000.01329.x

Koren, G., Vranderick, M., Gill, S., Farine, D., & MacLeod, S. (2014). Sex Differences in Pharmacokinetics of Doxylamine–Pyridoxine Combination. *Obstetrics & Gynecology*, *123*, 151S. https://doi.org/10.1097/01.AOG.0000447140.77554.9c

Lee, S. L. J., Horsfield, J. A., Black, M. A., Rutherford, K., & Gemmell, N. J. (2018). Identification of sex differences in zebrafish (Danio rerio) brains during early sexual differentiation and masculinization using 17α-methyltestoterone†. *Biology of Reproduction*, *99*(2), 446–460. https://doi.org/10.1093/biolre/iox175

Liu, K. A., & Mager, N. A. D. (2016). Women's involvement in clinical trials: historical perspective and future implications. *Pharmacy Practice*, *14*(1), 708. https://doi.org/10.18549/PharmPract.2016.01.708

Mastroianni, A. C., Faden, R. R., Federman, D. D., & Institute of Medicine (U.S.). Committee on the Ethical and Legal Issues Relating to the Inclusion of Women in Clinical Studies. (1994). *Women and health research. Vol. 1 : ethical and legal issues of including women in clinical studies*. National Academy Press.

Meziane, H., Ouagazzal, A.-M., Aubert, L., Wietrzych, M., & Krezel, W. (2007). Estrous cycle effects on behavior of C57BL/6J and BALB/cByJ female mice: implications for phenotyping strategies. *Genes, Brain and Behavior*, *6*(2), 192–200. https://doi.org/10.1111/j.1601-183X.2006.00249.x

Mikhail, G. W. (2005). Coronary heart disease in women. *BMJ (Clinical Research Ed.)*, *331*(7515), 467–468. https://doi.org/10.1136/bmj.331.7515.467

Prendergast, B. J., Onishi, K. G., & Zucker, I. (2014). Female mice liberated for inclusion in neuroscience and biomedical research. *Neuroscience & Biobehavioral Reviews*, *40*, 1–5. https://doi.org/10.1016/j.neubiorev.2014.01.001

Schlomer, B. J., Feretti, M., Rodriguez, E., Blaschko, S., Cunha, G., & Baskin, L. (2013). Sexual Differentiation in the Male and Female Mouse from Days 0 to 21: A Detailed and Novel Morphometric Description. *Journal of Urology*, *190*(4S), 1610–1617. https://doi.org/10.1016/j.juro.2013.02.3198

Simon, V. (2005). Wanted: women in clinical trials. *Science*, *308*(5728), 1517–1518. http://go.galegroup.com.prxy4.ursus.maine.edu/ps/i.do?p=AONE&u=maine_orono&id=GALE%7CA133569531&v=2.1&it=r&sid=summon

Singer, C. A., Rogers, K. L., Strickland, T. M., & Dorsa, D. M. (1996). Estrogen protects primary cortical neurons from glutamate toxicity. *Neuroscience Letters*, *212*(1), 13–16. https://doi.org/10.1016/0304-3940(96)12760-9

Soldin, O. P., Chung, S. H., & Mattison, D. R. (2011). Sex Differences in Drug Disposition. *Journal of Biomedicine and Biotechnology*, *2011*, 1–14. https://doi.org/10.1155/2011/187103

Soldin, O. P., & Mattison, D. R. (2009). Sex differences in pharmacokinetics and pharmacodynamics. *Clinical Pharmacokinetics*, *48*(3), 143–157. https://doi.org/10.2165/00003088-200948030-00001

Spencer, G (2002). Background on Mouse as a Model Organism - National Human Genome Research Institute (NHGRI). Retrieved March 30, 2019, from https://www.genome.gov/10005834/background-on-mouse-as-a-model-organism/

Walters, A. (2018). Girls with ADHD: Underdiagnosed and untreated. *The Brown University Child and Adolescent Behavior Letter*, *34*(11), 8–8. https://doi.org/10.1002/cbl.30337

Yeung, A. W. K., Goto, T. K., & Leung, W. K. (2017). At the Leading Front of Neuroscience: A Bibliometric Study of the 100 Most-Cited Articles. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00363

APPENDICES

APPENDIX A: SUPPLEMENTAL FIGURES

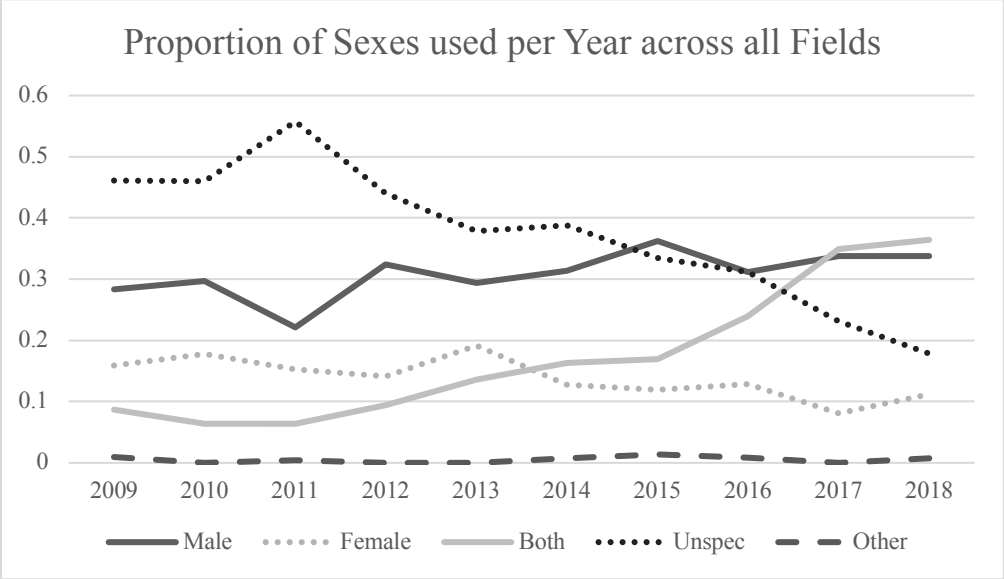
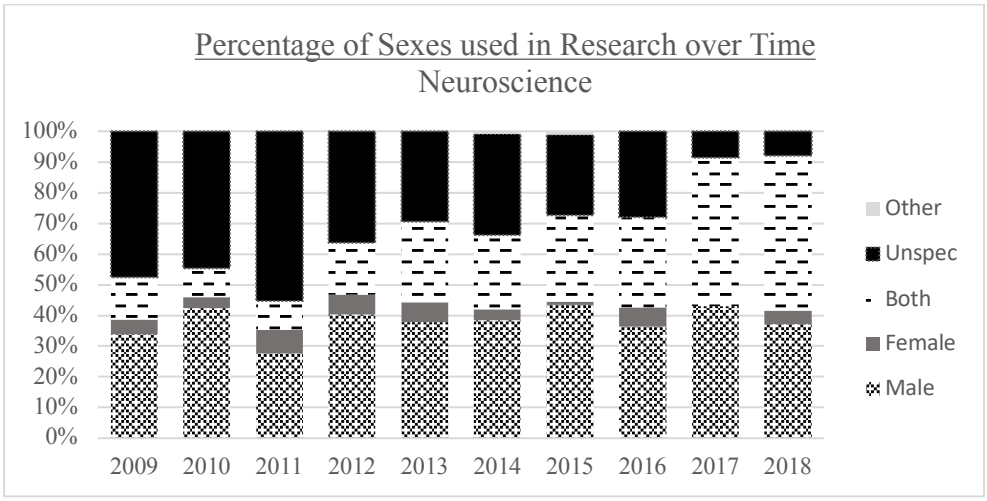


Figure 6: Proportion of Sexes used per Year across all Fields

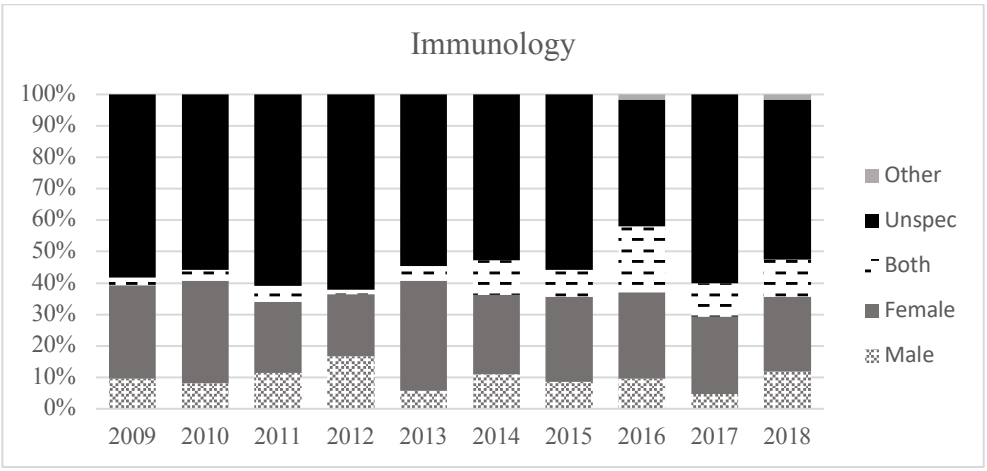
Indicates the proportion of sexes used each year across fields.

	Male	*Female*	*Both*	*Unspecified*	*Other*
Male	1				
Female	-0.57951	1			
Both	0.596579	-0.78454	1		
Unspecified	-0.76486	0.685571	-0.95563	1	
Other	0.243499	-0.27744	0.113465	-0.15736	1

A.



B.



C.

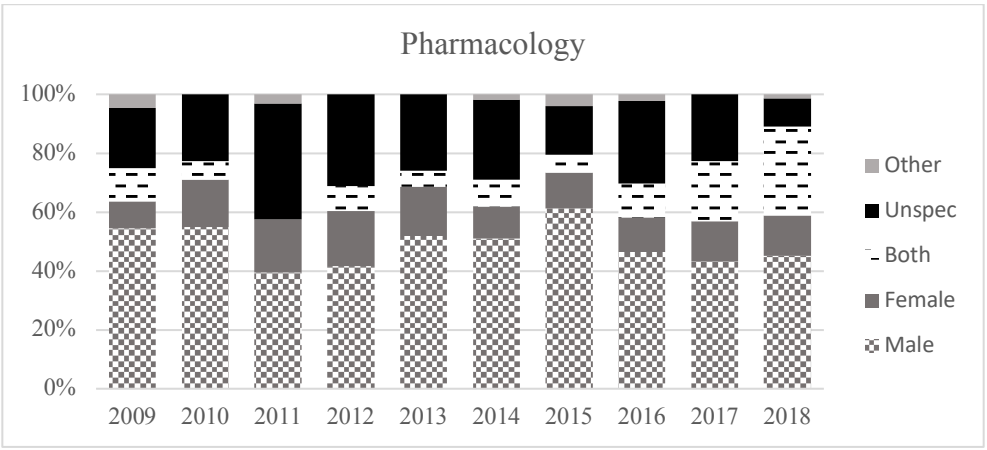
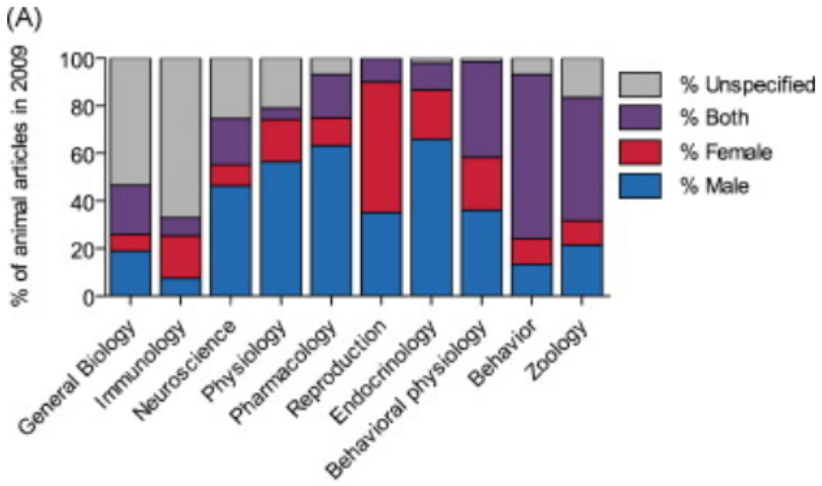


Figure 7: Percentage of Sexes used per Field 2009-2018.

A. Neuroscience has increased in specification by 40 percent while increasing the inclusion of both sexes by the same amount. B. There has been a slight increase in the amount of articles that used both sexes in Immunology. C. The inclusion of both sexes increased by 25 percent between 2015 and 2018.



(Taken from Annaliese Beery: Sex bias in neuroscience and biomedical research.)

Figure 8: Beery 2009 Data by Field

Data from 2009 indicated that immunology analyzed sex differences less than 5 percent of the time while neuroscience and pharmacology analyzed them at rates of 20 and 30 percent.

APPENDIX B: SUPPLEMENTAL METHODS

Search conducted on PubMed:

("mice"[Mesh terms] OR "mouse"[All Fields] OR "mice"[All fields] NOT "Genetic Diseases, X-Linked"[MeSH Terms] AND ("2009/01/01"[PDAT] : "2018/12/31"[PDAT]) NOT "In-vitro Techniques"[mesh terms] NOT "Genitalia"[mesh terms] NOT "breast"[mesh terms] NOT Review[ptyp] NOT "Reproduction"[mesh terms] NOT "veterinary"[all fields] NOT "Genetic Diseases, Y-Linked"[Mesh terms] NOT "Embryo"[all fields] NOT "Female Urogenital Diseases"[mesh terms] NOT "Mental Retardation, X-Linked"[mesh terms] NOT "Male Urogenital Diseases" NOT "Herpes Genitalis"[mesh terms] NOT "Alphapapillomavirus"[mesh terms] AND ("USA"[affiliation] OR "United States"[Affiliation] OR United States of America[Affiliation]) NOT "Animal Diseases"[mesh terms] AND "loattrfull text"[sb] NOT Cercopithecidae NOT Artiodactyla NOT Carnivora NOT Cetacea NOT Lagomorpha NOT Perissodactyla NOT Guinea Pigs NOT Chinchilla NOT Cuniculidae NOT Rats NOT Gerbillinae NOT Mole Rats NOT Myoxidae NOT Octodon NOT Porcupines NOT Sciuridae NOT Chiroptera NOT scandentia NOT Sirenia NOT Hyraxes NOT Insectivora NOT Marsupialia NOT Monotremata NOT Proboscidea Mammal NOT Xenarthra NOT Birds NOT Amphibians NOT Fishes NOT Reptiles NOT Invertebrates NOT Chlamydia NOT hamsters NOT rabbit NOT dog NOT cat NOT maternal NOT HPV NOT "Mice, Inbred BALB C"[Mesh terms] NOT Mice, Nude) AND "Plos pathogens"[Journal]

Program

The following is the program that was written to extract the following from the results of the above PubMed search: title, first author, first author affiliation, year of publication of each article. The articles were then grouped into USA affiliated and non-USA affiliated first authors, only the former was used in analysis.

```
%% Written by:      Phoenix Throckmorton
%%Purpose: This script will load string data from a .txt file and
extract
%%important information. This information will be filtered and
organized
%%into an excel spread sheet for presentation.
tic
%The script is organized as follows:
%   1. INPUT VARIABLES
%   2. Variable Allocation
%   3. Extracting Data
%   4. Post Processing
```

```
%   5. Outputting Data
clear all;
close all;


%% 1.INPUT VARIABLES (Edit this)
%%Change these as needed. Do not edit any other part of the script

%enter directory where ALL files are located:
file_directory = 'C:\Users\pthro\Desktop\NicolesProgram';

%enter name of .txt file, including .txt extension
txt = 'Neuron.txt';

%enter the name of the exporting excel spreadsheet, including .xlsx
%extension
excel = 'USA_Neuron.xlsx';

%enter the name of the NON USA exporting spreadsheet, including .xlsx
NONUSA_excel = 'NONUSA_Neuron.xlsx';

%% 2.Variable Allocation
%This section opens files and prepares/converts variables for data
transfer

%set appropriate file directories and names
cd(file_directory);
txt_full = strcat(file_directory, '\', txt);
excel_full = strcat(file_directory, '\', excel);

%Prepares the .txt for file extraction, string by string
ID = fopen(txt_full);
data = textscan(ID, '%s');
size = length(data{1});

%%Preallocate other variables for data extraction

%%title
%counts how many times the initialization and termination terms occur
initial_title_count = 0;
final_title_count = 0;
%keeps track of the indices for each occurence
title_index = 0;

%%year
%counts how many times the initialization and termination terms occur
initial_year_count = 0;
final_year_count = 0;
%keeps track of the indices for each occurence
year_index = 0;

%%firstname
%counts how many times the initialization and termination terms occur
initial_firstname_count = 0;
```

```
final_firstname_count = 0;
%keeps track of the indices for each occurence
firstname_index = 0;

%%lastname
%counts how many times the initialization and termination terms occur
initial_lastname_count = 0;
final_lastname_count = 0;
%keeps track of the indices for each occurence
lastname_index = 0;

%%affiliation
%counts how many times the initialization and termination terms occur
initial_affiliation_count = 0;
final_affiliation_count = 0;
%keeps track of the indices for each occurence
affiliation_index = 0;

%% 3.Extracting data Index

%get the locations of all of the article titles, year, first name, last
%name, and affiliation

%This loop is performed first since only one instance of each term
appears
%for each article
for i = 1:size

   %check if an article title is to be shown, and save bounds
    if contains(data{1}{i},'<ArticleTitle>') == 1
        initial_title_count = initial_title_count + 1;
        title_index(initial_title_count,1) = i;
    end

    if contains(data{1}{i},'</ArticleTitle>') == 1
        final_title_count = final_title_count + 1;
        title_index(final_title_count,2) = i;
    end

    %check if year is to be shown, and save bounds
    if contains(data{1}{i},'<PubDate>') == 1
        initial_year_count = initial_year_count + 1;
        year_index(initial_year_count,1) = i;
    end

    if contains(data{1}{i},'</PubDate>') == 1
        final_year_count = final_year_count + 1;
        year_index(final_year_count,2) = i;
    end
end

%initialize synchronization counters for initial and final bounds
size_sync = length(year_index(:,1));
```

```
%now that the total number of articles is known, run through again
between
%each bound and grab the first author and affiliation and pair them
for j = 1:size_sync
    if j == size_sync
        loop_condition = size;   %prevents strange behavior at end of
loop
    else
        loop_condition = title_index(j+1,1);
    end

    %check if author lastname is to be shown, and save bounds
    for k = title_index(j,1): loop_condition
        if contains(data{1}{k},'<LastName>') == 1
            initial_lastname_count = initial_lastname_count + 1;
            lastname_index(initial_lastname_count,1) = k;
        end
        if contains(data{1}{k},'</LastName>') == 1
            final_lastname_count = final_lastname_count + 1;
            lastname_index(final_lastname_count,2) = k;
            break; %TERMINATES AFTER FIRST OCCURENCE
        end
    end

    %check if author firstname is to be shown, and save bounds
    for k = title_index(j,1): loop_condition
        if contains(data{1}{k},'<ForeName>') == 1
            initial_firstname_count = initial_firstname_count + 1;
            firstname_index(initial_firstname_count,1) = k;
        end

        if contains(data{1}{k},'</ForeName>') == 1
            final_firstname_count = final_firstname_count + 1;
            firstname_index(final_firstname_count,2) = k;
            break %TERMINATES AFTER FIRST OCCURENCE
        end
    end

    %check if affiliation is to be shown, and save bounds
    for k = title_index(j,1): loop_condition
        if contains(data{1}{k},'<AffiliationInfo>') == 1
            initial_affiliation_count = initial_affiliation_count + 1;
            affiliation_index(initial_affiliation_count,1) = k;
        end

        if contains(data{1}{k},'</AffiliationInfo>') == 1
            final_affiliation_count = final_affiliation_count + 1;
            affiliation_index(final_affiliation_count,2) = k;
            break %TERMINATES AFTER FIRST OCCURENCE
        end
    end


end
```

```
%do a second pass to truly filter out the year locations
for j = 1:size_sync
    for k = year_index(j,1): year_index(j,2)
        if contains(data{1}{k},'<Year>') == 1
            year_index(j,1) = k;
        end

        if contains(data{1}{k},'</Year>') == 1
            year_index(j,2) = k;
        end
    end
end

%%3.B Raw Data Extraction
%Grab the data using the determined indeces

%preallocate data storing variables based on now known number of
articles
title = strings(size_sync,1);
year = strings(size_sync,1);
first = strings(size_sync,1);
last = strings(size_sync,1);
affiliation = strings(size_sync,1);
institution = strings(size_sync,1);

for j = 1:size_sync
    %For article titles
    for k = title_index(j,1): title_index(j,2)
       title(j) = strcat(title(j),{' '},data{1}{k}); %adds the entire
title line
    end
    %For years
    for k = year_index(j,1): year_index(j,2)
       year(j) = strcat(year(j),{' '},data{1}{k}); %adds the entire
year line
    end
    %For lastnames
    for k = lastname_index(j,1): lastname_index(j,2)
       last(j) = strcat(last(j),{' '},data{1}{k}); %adds the entire
lastname line
    end
    %For firstnames
    for k = firstname_index(j,1): firstname_index(j,2)
       first(j) = strcat(first(j),{' '},data{1}{k}); %adds the entire
firstname line
    end
    %For affiliations
    for k = affiliation_index(j,1): affiliation_index(j,2)
       affiliation(j) = strcat(affiliation(j),{' '},data{1}{k}); %adds
the entire affiliation line
    end

    %Determination of university
```

```
    if contains(affiliation(j), 'inc.') ||
contains(affiliation(j),'incorporated') ...
            || contains(affiliation(j), 'Inc.') ||
contains(affiliation(j), 'corporation') ...
            || contains(affiliation(j), 'Corporation')

        institution(j) = 'Biotech';
    elseif contains(affiliation(j), 'Hospital')
        institution(j) = 'Hospital';
    elseif contains(affiliation(j), 'University')
        institution(j) = 'University';
    else
        institution(j) = ' ';
    end
end


%% 4.Post Processing

%%Remove any articles that were published outside of the US

%preallocate based on total files
junk_title = strings(size_sync,1);
junk_year = strings(size_sync,1);
junk_first = strings(size_sync,1);
junk_last = strings(size_sync,1);
junk_affiliation = strings(size_sync,1);
junk_institution = strings(size_sync,1);
NONUSA_counter = zeros(size_sync,1);

%Remove any information that is not desired for the final output

%for title
title(1:size_sync) = erase(title(1:size_sync), ' <ArticleTitle>');
title(1:size_sync) = erase(title(1:size_sync), '<ArticleTitle> ');
title(1:size_sync) = erase(title(1:size_sync), '<ArticleTitle>');
title(1:size_sync) = erase(title(1:size_sync), ' </ArticleTitle>');
title(1:size_sync) = erase(title(1:size_sync), '</ArticleTitle> ');
title(1:size_sync) = erase(title(1:size_sync), '</ArticleTitle>');

%for first name
first(1:size_sync) = erase(first(1:size_sync), ' <ForeName>');
first(1:size_sync) = erase(first(1:size_sync), '<ForeName> ');
first(1:size_sync) = erase(first(1:size_sync), '<ForeName>');
first(1:size_sync) = erase(first(1:size_sync), ' </ForeName>');
first(1:size_sync) = erase(first(1:size_sync), '</ForeName> ');
first(1:size_sync) = erase(first(1:size_sync), '</ForeName>');

%for lastname
last(1:size_sync) = erase(last(1:size_sync), ' <LastName>');
last(1:size_sync) = erase(last(1:size_sync), '<LastName> ');
last(1:size_sync) = erase(last(1:size_sync), '<LastName>');
last(1:size_sync) = erase(last(1:size_sync), ' </LastName>');
last(1:size_sync) = erase(last(1:size_sync), '</LastName> ');
```

```
last(1:size_sync) = erase(last(1:size_sync), '</LastName>');

%for year
year(1:size_sync) = erase(year(1:size_sync), ' <Year>');
year(1:size_sync) = erase(year(1:size_sync), '<Year> ');
year(1:size_sync) = erase(year(1:size_sync), '<Year>');
year(1:size_sync) = erase(year(1:size_sync), ' </Year>');
year(1:size_sync) = erase(year(1:size_sync), '</Year> ');
year(1:size_sync) = erase(year(1:size_sync), '</Year>');

%for affiliation
affiliation(1:size_sync) = erase(affiliation(1:size_sync), '
<AffiliationInfo>');
affiliation(1:size_sync) = erase(affiliation(1:size_sync),
'<AffiliationInfo> ');
affiliation(1:size_sync) = erase(affiliation(1:size_sync),
'<AffiliationInfo>');
affiliation(1:size_sync) = erase(affiliation(1:size_sync), '
</AffiliationInfo>');
affiliation(1:size_sync) = erase(affiliation(1:size_sync),
'</AffiliationInfo> ');
affiliation(1:size_sync) = erase(affiliation(1:size_sync),
'</AffiliationInfo>');
affiliation(1:size_sync) = erase(affiliation(1:size_sync), '
<Affiliation>');
affiliation(1:size_sync) = erase(affiliation(1:size_sync),
'<Affiliation> ');
affiliation(1:size_sync) = erase(affiliation(1:size_sync),
'<Affiliation>');
affiliation(1:size_sync) = erase(affiliation(1:size_sync), '
</Affiliation>');
affiliation(1:size_sync) = erase(affiliation(1:size_sync),
'</Affiliation> ');
affiliation(1:size_sync) = erase(affiliation(1:size_sync),
'</Affiliation>');

for j = 1:size_sync
    %check to see if the article is US affiliated
    if contains(affiliation(j), 'USA') || contains(affiliation(j),
'United States')

    else %keeps track of where non US files are
        NONUSA_counter(j) = 1;
    end
end

%determines where NONUSA Data occurs
c = 0;
for j = 1:size_sync
    if NONUSA_counter(j) == 1
        c = c + 1;
        NONUSA_counter(j) = 1;
        junk_title(c) = title(j);
        junk_first(c) = first(j);
        junk_last(c) = last(j);
```

```
        junk_affiliation(c) = affiliation(j);
        junk_institution(c) = institution(j);
        junk_year(c) = year(j);
    end
end

%deletes data where appropriate

%convert to logical indexing format
NONUSA_counter = NONUSA_counter == 1;

%apply conditions
title(NONUSA_counter,:) = [];
year(NONUSA_counter,:) = [];
first(NONUSA_counter,:) = [];
last(NONUSA_counter,:) = [];
affiliation(NONUSA_counter,:) = [];
institution(NONUSA_counter,:) = [];
%% 5. Data Output

%create the headings for excel data:
heading = strings(1,6);
heading(1) = 'Article Title';
heading(2) = 'Year';
heading(3) = 'Last Name';
heading(4) = 'First Name';
heading(5) = 'Affiliation';
heading(6) = 'Institution';

final_data = horzcat(title, year, last, first, affiliation,
institution);
final_data = vertcat(heading,final_data);
xlswrite(excel, final_data);

NON_USA_data = horzcat(junk_title, junk_year, junk_last, junk_first,
...
    junk_affiliation, junk_institution);
NON_USA_data = vertcat(heading,NON_USA_data);
xlswrite(NONUSA_excel, NON_USA_data);

toc
```

AUTHOR'S BIOGRAPHY

Nicole was raised in Billerica, Massachusetts and attended Shawsheen Valley Vocational Technical High School. She studied Medical Assisting during her time there. She majored in Biology with minors in Neuroscience and Psychology at University of Maine. Her plans after graduation include attending a post baccalaureate program at the NIH before obtaining a PhD in Neurobiology.