

Sex and Gender Differences Research Design for Basic, Clinical, and Population Studies: Essentials for Investigators

Janet W. Rich-Edwards,^{1,2,3} Ursula B. Kaiser,⁴ Grace L. Chen,² JoAnn E. Manson,^{2,3,5} and Jill M. Goldstein^{1,6,7,8,9}

¹Division of Women's Health, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115; ²Connors Center for Women's Health and Gender Biology, Brigham and Women's Hospital, Boston, Massachusetts 02115; ³Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115; ⁴Division of Endocrinology, Diabetes and Hypertension, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115; ⁵Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115; ⁶Department of Psychiatry, Harvard Medical School, Boston, Massachusetts 02115; ⁷Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115; ⁸Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114; and ⁹Department of Obstetrics and Gynecology, Massachusetts General Hospital, Boston, Massachusetts 02114

ABSTRACT A sex- and gender-informed perspective increases rigor, promotes discovery, and expands the relevance of biomedical research. In the current era of accountability to present data for males and females, thoughtful and deliberate methodology can improve study design and inference in sex and gender differences research. We address issues of motivation, subject selection, sample size, data collection, analysis, and interpretation, considering implications for basic, clinical, and population research. In particular, we focus on methods to test sex/gender differences as effect modification or interaction, and discuss why some inferences from sex-stratified data should be viewed with caution. Without careful methodology, the pursuit of sex difference research, despite a mandate from funding agencies, will result in a literature of contradiction. However, given the historic lack of attention to sex differences, the absence of evidence for sex differences is not necessarily evidence of the absence of sex differences. Thoughtfully conceived and conducted sex and gender differences research is needed to drive scientific and therapeutic discovery for all sexes and genders. (*Endocrine Reviews* 39: 424 – 439, 2018)

Many compelling publications have argued why sex and gender should be considered in preclinical, clinical, and population research (1–4). Both sex (the biological attributes of females and males) and gender (socially constructed roles, behaviors, and identities in a spectrum, including femininity and masculinity) affect molecular and cellular processes, clinical traits, response to treatments, health, and disease (1). Since 2010, the Canadian Institutes of Health Research has mandated that all grant applicants address whether they had considered sex and/or gender in their applications (5). In 2014, the European Commission issued the Horizon 2020 guideline, which makes explicit the rules for sex and gender inclusion as elements of European Union grant evaluation and monitoring (6, 7). Although the 1993 National Institutes of Health (NIH) Revitalization Act

required the inclusion of women in NIH-funded clinical research, it was not until 2015 that the NIH announced policies requiring the consideration of sex as a biological variable in study design, analysis, and reporting (1, 8–10). Such mandates to include females are not mere political correctness (11). A sex-informed and gender-informed perspective is essential to increase rigor, promote discovery, expand the relevance of research, and improve patient care. At the very least, it will allow readers of the scientific literature to critically assess the validity of what they read.

Investigators who wish to—or now find themselves required to—include both sexes in their studies are faced with a number of methodological questions, including issues of motivation, subject selection, sample size, data collection, analysis, and interpretation. We provide an overview of these issues in this review as

ESSENTIAL POINTS

- A sex- and gender-informed perspective increases rigor, promotes discovery, and expands the relevance of biomedical research
- Methods exist to test sex and gender differences as interactions; inference from sex- and gender-stratified data should be viewed with caution
- Without careful methodology, the pursuit of sex and gender difference research as a poorly considered mandate will result in a literature of contradiction
- However, given the paucity of sex and gender differences research, the absence of evidence for differences is not necessarily evidence of the absence of differences

they pertain to basic, clinical, and population research (Table 1). This review builds on earlier discussions of sex differences research methodology (11–18) in several ways: we consider gender as well as sex differences; we examine the entire research process, from motivation to analysis and presentation; and we discuss nuances of statistical design and interpretation,

particularly how to plan robust tests of sex or gender interactions that can help minimize statistical artifacts. Rather than assume ubiquitous sex and gender differences in biology, health, and disease, we propose methods and interpretation that will increase the likelihood of detecting true differences where they exist.

Study Motivation

There is ample evidence of sex differences—at the level of the cell, organism, and population—to motivate sex differences research. Sex chromosomes encode sexual differentiation through three mechanisms: (1) presence of Y genes; (2) increased dose of X genes in XX vs XY cells; and (3) X chromosome inactivation and imprinting (12). These primary chromosomal differences lead to sexual differentiation and the somatic and gonadal expressions of sex (19). The resulting “sexome” produces differences in all organ systems and across the lifespan, influencing how our bodies interact with the environment to determine health (20). The sex-informed framework considers sex differences in anatomy and physiology, understood within a lifespan perspective of sensitive periods of fetal and childhood development, differential pace and timing of puberty, reproductive events, and senescence. This is critical given that timing is everything when it comes to identifying developmental sex effects (14, 21, 22). Furthermore, sex differences in treatment abound: pharmacokinetics and pharmacodynamics of medications often vary by sex, as may effects of other treatment modalities (23).

Gender, too, is a determinant of health, influencing the physical and social environments to which individuals are exposed, their access to resources that affect health, their agency to seek health care and receive treatment, and the equitability of research that drives medical discovery (14, 17, 24–26).

This sex- and gender-informed perspective is necessitated by widespread differences in disease incidence, prevalence, and survival that have been reviewed elsewhere (2, 26–28). There are sex and

gender differences in symptoms and clinical presentations of illness, reliability of diagnostic tests, and response to treatment. There is “sex bias” down to the level of epigenetic marking and gene expression (28). In short, the rigor of research depends on researchers’ understanding of the ways in which sex and gender influence the biologic systems they study.

Investigators seeking to construct a sex- and gender-informed framework for their research may be disappointed by a lack of systematic evidence regarding sex and gender differences in the literature. There is a particular dearth of true gender-difference studies; in fact, literature searches on “gender differences” largely turn up studies on sex differences that have used the term “gender” to refer to biologic sex. The historic neglect of women in clinical studies and the sex of animals and cells in basic research should be kept in mind when gathering evidence of sex and gender differences. Although data to interrogate sex differences may exist in some studies, they have yet to be examined. In other cases, sex-informed questions have yet to be posed. Furthermore, as argued below, the proliferation of ill-considered and often unplanned sex difference inquiries leads to a literature of contradictions. Thus, the absence of evidence for sex differences is not necessarily evidence of the absence of sex differences.

Study Design

Overarching study design

In most cases, the choice of overarching study design, whether experimental or observational, is little affected by considerations of sex and gender.

Table 1. Methodological Considerations in Investigations of Sex and Gender Differences

Research Step	Best Practices
Motivation	<p>Consider known sex differences in disease incidence, prevalence, and survival.</p> <p>Review existing literature on sex and gender differences, alert to the fact that many hypotheses have not been well tested. Read carefully to consider likelihood of false-positives (especially in context of multiple testing) and false-negatives (especially where statistical power is low).</p> <p>Apply a life course perspective to consider the timing of exposures that might interact with sex and gender in specific developmental windows.</p>
Subject selection	<p>Consider sex-specific age incidence of disease to maximize statistical power.</p> <p>Consider reproductive stages and cycles, particularly where they may modify the impact of the main exposure being investigated.</p> <p>Consider the impact of gendered social environment for the distribution of factors that may interact with the main exposure.</p> <p>For basic and preclinical studies, review options for classical gonadectomy, knockouts, or four-core genotype experiments.</p> <p>Consider whether sex of cell lines is known, relevant, and generalizable.</p>
Randomization (if applicable)	In smaller studies, stratified randomization by sex or gender will ensure balance, even if different numbers of males and females are included.
Sample size	<p>True tests of sex differences need to be large enough to test interaction between sex and the main exposure or treatment; such tests typically require several times the sample size to be adequately powered, compared with studies of main effects.</p> <p>Studies too small to detect interaction can still report the main effects of the exposure or treatment by sex; however, they cannot claim to have tested a sex difference. Be alert to the risk of false-negatives in underpowered sex strata.</p> <p>Studies too small to detect even the main effects of sex can provide sex-specific data to generate hypotheses or contribute to meta-analyses of sex differences.</p> <p>“Big data” studies, where the variable of sex is often available, need to be conducted thoughtfully to avoid contributing false-positives to the sex difference literature.</p>
Data collection	<p>Consider sex and gender differences in disease presentation.</p> <p>Consider whether exposures mean the same thing in both sexes and genders.</p> <p>Be aware of sex and gender differences in pharmacokinetics and pharmacodynamics; the same dose may have different impact in males and females or may vary by body size.</p> <p>Collect data on exogenous hormones: contraceptives, menopausal hormone therapy, testosterone, and other steroid use.</p> <p>Consider recording data on reproductive cycle (follicular/luteal), and stage (prepuberty, puberty, pregnancy, lactation, premenopause and postmenopause).</p> <p>Collect data on influential covariates that may vary by sex and gender in the study population.</p>
Analysis, reporting, and interpretation	<p>Prespecify tests of sex differences to reduce type I error.</p> <p>Account for confounding by factors associated with sex and gender.</p> <p>Investigate intermediate “pathway” variables to understand apparent sex differences.</p> <p>Admit when sex differences were tested as <i>post hoc</i> exploratory analyses.</p> <p>Make opportunities to replicate sex difference findings.</p> <p>Interpret apparent sex and gender differences in the light of biological plausibility and social context.</p>

Exceptions to this are experiments precluded by ethical considerations, such as inclusion of pregnant women for trials of potentially teratogenic drugs. However, nearly every other feature of study design necessitates a sex-informed perspective, including subject selection, randomization, sample size, and data collection.

Subject selection

Inclusion of both sexes is more nuanced than deciding that the sample should be equally divided by sex. Sex-specific age incidence of disease, reproductive stage, reproductive cycle, and environment need to be considered to optimize the validity, generalizability, and efficiency of a study sample. More often than not,

investigators have to compromise between competing goals of validity (by narrowing subject selection to increase the likelihood that findings are true for a specific population) and generalizability (by widening subject selection to make broad inference at the risk of overgeneralizing across true differences between groups.) There are also compromises between large scientific goals and restricted available funds. Such trade-offs are best made as choices informed by already known sex and gender differences. The most efficient subject selection will pick the minimum number of each sex or gender necessary to make valid inferences about sex and gender differences; a 50/50 split between males and females may not be the most efficient, as discussed below.

Sex-specific incidence of disease

Sex differences in incidence and age-incidence trajectories are important considerations in subject selection. For example, at ages 55 to 64 men have more than double the rate of coronary heart disease (CHD) of women. By ages 85 to 94, male CHD rates are only 10% higher than those of females (20). Thus, an investigator wishing to enroll a cohort of 50-year-olds to study CHD incidence will need to enroll two to three times as many women as men to ensure equivalent statistical power, or consider selecting older women. For example, the Vitamin D and Omega-3 Trial study of dietary supplements to reduce heart disease, stroke, and cancer includes women aged 55 or older and men aged 50 or older to account for the later onset of disease in women (29). Sex differences in disease incidence exist in animals as well. For example, in the nonobese diabetic mouse, diabetes is more prevalent in females, so that more male mice must be included to yield the same number of affected animals of each sex (30).

Sex-specific differences in aging

Females outlive males in most vertebrate species (31). In mammals, the heterogametic (XY) sex may have a shorter lifespan because of the unguarded expression of harmful recessive alleles on the Y sex chromosome. Similarly, the homogametic (XX) sex may be protected by the stochastic X-inactivation that creates mosaics of females; although female neonates are a 1:1 mosaic of maternal and paternal allele expression, over time that ratio becomes skewed to favor the cellular population whose active X presumably confers a survival advantage (32–35).

Sex differences in the rate of aging and the incidence of disease onset are reflected at the cellular level. For example, there are sex differences in the length of telomeres, noncoding DNA sequences that cap and protect chromosomes, the length of which are correlated with longevity. Although similar at birth, male telomeres shorten faster during the lifespan than do female telomeres (35). This difference could be the

result of sex or gender; most likely, it is a combination of biological sex differences and gendered experiences (such as smoking) (36). Similarly, although stem cell populations decline with aging, this loss is earlier and more rapid in male than in female mice (37). Methylation patterns also differ between the sexes, likely influencing DNA expression over the life course (38, 39). As research further clarifies sex-specific or sex-dependent mechanisms of senescence, investigators may want to consider sex differences in the cellular age and methylation patterns of their subjects, be they cells, animals, or people.

Reproductive stages and cycles

All animals, regardless of sex or species, go through a process of reproductive maturation whose timing, duration, and outcome are subject to physical and social cues from the environment. In mammals, puberty involves sex-specific, but variable, changes in central neural systems, gonadal steroid production, and the emergence of secondary sexual characteristics, including behaviors. When a study investigates adolescence or young adulthood, accounting for sex differences in the pace and timing of puberty will be critical for identifying sex effects (14).

Mature mammals of both sexes have variations in gonadal steroid levels that may affect subject selection. In males, testosterone levels have circadian and perhaps seasonal variations and vary with age, physical activity, and energy homeostasis (40, 41). Reproductive age females have menstrual or estrous cycles. On top of natural variability, women may use hormonal contraceptives or menopausal hormone therapy; many men use exogenous androgens and anabolic steroids. These factors are important in subject selection if an investigator wants to understand how the exposure-outcome associations under study are impacted by sex hormones. Researchers may decide to include a representative range of reproductive phases or cycles. For example, cyclical patterns of DNA synthesis and rates of cell division and death would not have been discovered if females in different cycle phases had not been studied (42, 43). The knowledge that natural killer cell activity peaks during the luteal phase came from studies of cycling women (44). Understanding of the roles of neurokinin B and kisspeptin in reproduction has been facilitated by studying male and female animals at varying reproductive stages, with and without gonadectomy (45).

Sex differences in physiology and behavior have been observed even in the prepubertal and peripubertal periods, before the pubertal activation of the hypothalamic–pituitary–gonadal axis and production of gonadal sex steroid hormones. These prepubertal sex differences have been largely attributed to the effects of prenatal and perinatal activity of the hypothalamic–pituitary–gonadal axis and resultant sex steroid hormone production and actions. Among

the best described effects are the so-called activational and organizational effects of gonadal hormones on brain development (46). The first robust sex difference described in the mammalian brain was the sexually dimorphic nucleus of the preoptic area (47). More recently, a sexually dimorphic population of kisspeptin neurons was identified that is present in higher numbers in the anteroventral periventricular nucleus in prepubertal females than in males, to which the sexually dimorphic preovulatory luteinizing hormone surge that occurs in adult females but not males is attributed (48).

Thus, to the extent that hormone levels affect study outcomes, researchers may need to examine subjects who are premenopausal or postmenopausal, in the follicular or luteal phase, and with or without hysterectomy or gonadectomy. Including or excluding participants using hormonal therapies, such as contraceptives and female and male hormone replacement or suppression therapies, is another potentially important design choice. To fully capture between-sex variability, it may be of use to compare men to two or more groups of women. For example, a study of brain activity in the stress response circuitry found few differences between healthy men and women in the early follicular phase, but striking differences between men and the same women at midcycle (49). Furthermore, sex differences in brain activity in memory circuitry were statistically significant in premenopausal and perimenopausal women, but attenuated in postmenopausal women compared with men (50). To capture within-sex variability, studies compare the same females at different cyclical stages, perhaps in crossover fashion.

Note that the effects of sex steroid hormones extend beyond estradiol and testosterone. There are multiple types of estrogens produced by the ovaries and other tissues, as well as multiple androgens beyond testosterone. Progesterone levels also need to be considered. Furthermore, there is target tissue specificity in the actions of estrogens, which can be attributed to tissue-specific expression patterns of estrogen receptors (ERs), including ER α , ER β , and estrogen membrane receptors such as membrane ER α and the G protein-coupled receptor GPER1/GPR30 (51, 52).

In addition to the multiple ERs, tissue-specific responses to estrogens can occur through the presence of modulating proteins such as coactivators and corepressors, among others. Varying tissue-specific responses are exemplified by the action of synthetic agonists and antagonists such as the selective ER modulators, including tamoxifen, raloxifene, and toremifene. These selective ER modulators are competitive inhibitors of estrogen binding to ERs, with mixed agonist and antagonist activity, depending on the target tissue (53). For example, tamoxifen is used in the prevention and treatment of breast cancer as an ER

antagonist, but it has ER agonist activity in some other tissues such as bone and endometrium. Progesterone also acts through multiple receptors, which are generated as splice variants from a single gene (54). The actions of testosterone, through the androgen receptor, are modulated at the local tissue level through local activity of the enzyme 5 α -reductase, which catalyzes the formation of the more potent androgen receptor agonist, dihydrotestosterone (55).

Gender and subject selection

Many determinants of disease, both physical and social, are differentially distributed by gender. Some of these factors may confound experiments if not carefully accounted for in study design and analysis. For example, in many societies, women are more likely to have vitamin D deficiency (56), affecting multiple tissues and systems, and men are more likely to smoke cigarettes and drink alcohol. Men and women are exposed differentially to types of violence and trauma (57–61). Such stressors may affect gonadal steroid secretion in a sex- and hormone-dependent fashion (12). In the case of powerful covariates strongly associated with gender or sex, investigators may want to select participants to ensure these covariates are balanced in male and female samples.

Special considerations regarding subject selection for basic studies

Historical reliance on male animal models (e.g., mice, rats) has resulted in incomplete data to guide human subject research in both men and women. Basic studies can complement clinical studies by investigating mechanisms of sex-dependent or sex-specific processes in greater depth by manipulating genotypic and phenotypic sex experimentally (12). Beyond simply studying both male and female animals as they age naturally, studies can include classic gonadectomy with or without hormone replacement: prenatally and perinatally to address developmental effects; in juvenile animals to study postnatal developmental and differentiation effects; in adults to assess the effects of sex steroid hormones at the time of testing; and in aging animals to study effects of sex steroids in models of aging. Several new genetic and epigenetic animal models have increasing translational validity to represent human ovarian failure and menopause (62). Some alternative models of menopause or ovarian failure include Foxl2-deficient mice with accelerated rates of decline in ovarian reserve (63).

Another frequent approach is to study “knockout” mice (or other species) that lack a specific sex steroid receptor. ER α knockout mice have shown that the absence of ER α promotes adiposity in male and female animals and, in turn, the progression of breast cancer in females (64). Animals with “conditional knockout” or “conditional knockin” of a specific sex steroid

receptor can be used to target specific tissues or life stages.

Additionally, targeted mutagenesis can be used to address the role of specific domains or specific functions of a sex steroid receptor. For example, although ER α has traditionally been thought of as a nuclear, ligand-dependent transcription factor acting through estrogen response elements in gene promoters, the molecular mechanisms of action are more complex. Estradiol actions can be mediated by other “nonclassical” ER α pathways: (1) ligand-independent ER α signaling, in which gene activation alters phosphorylation of ERs via second-messenger pathways that affect intracellular kinase and phosphatase activity; (2) rapid, nongenomic effects through a membrane-associated ER; and (3) genomic, estrogen response element-independent signaling, in which ER α regulates genes via protein–protein interaction with other transcription factors, including c-Fos/c-Jun B (AP-1), Sp1, and nuclear factor κ B (65). For example, as noted above, estradiol is critical to the regulation of energy balance and body weight. In an experiment with female mice, ER α -null mutant mice become obese, with decreased energy expenditure and locomotion, increased adiposity, hyperleptinemia, and altered glucose homeostasis, characteristics similar to the propensity of postmenopausal women to develop obesity and type 2 diabetes. Interestingly, knockin mice that express a mutant ER α that can signal only through a nonclassical pathway (*i.e.*, without direct estrogen response element binding) restored the metabolic parameters to normal or near-normal values, including energy expenditure. These findings indicate that nonclassical ER α signaling mediates major effects of estradiol on energy balance, raising the possibility that selective ER α modulators may be developed to reduce the risks of obesity and metabolic disturbances in postmenopausal women (66).

The “four core genotypes” mouse model has emerged as a major model in which investigators can vary the sex chromosome complement (XX vs XY) to assess whether sex differences in phenotypes are caused by the sex chromosome complement, gonadal hormones, or both. In this model, the testis-determining gene *Sry* is deleted from the Y chromosome and inserted onto an autosome. Four genotypes are produced: XY[−]*Sry* mice with testes, XX*Sry* mice with testes (with the *Sry* gene on an autosome), XX mice with ovaries, and XY[−] mice with ovaries. In this manner, XX and XY mice with the same type of gonad can be compared to assess phenotypic effects of the sex chromosome complement in cells and tissues (12, 67). These models permit investigators to observe, for the most part, the independent effects of sex chromosomes and hormones on physiologic and pathophysiologic processes. Four core genotypes methods have shown sex chromosomes to affect body weight and metabolism

independent of hormonal pathways (68) and have isolated effects of sex chromosomes on cardiac ischemia/reperfusion injury from those of gonadal hormones (69).

Sex of cells

Although it is facile to insist that basic researchers use and report on both XX and XY cells in their experiments, this is not always possible (11). In fact, cell lines are a poor model with which to study sex differences, even when the sex of the lineage is specified. By definition, immortalized cell lines, chosen for their peculiarities and derived from a single organism, may be inherently impossible to cull or create from a second organism of any sex. Even where it is possible to create cell lines from a male and female similar enough to interrogate a particular question, inferences about sex differences cannot reliably be made. As with a clinical study with $n = 2$, a comparison of a male and a female cell line, because each is derived from a single individual, cannot distinguish sex differences from other genetic, epigenetic, or environmental characteristics of the founding individuals from which they were derived. Cell lines may have sex-dependent features other than the sex chromosome complement, including differences in hormone production or hormone responses related to variation in steroidogenic enzyme expression or expression of sex steroid hormone receptors. There may also be differences in expression of other genes related to imprinting or epigenetic differences. Moreover, each cell line is clonal in origin and has unique characteristics based on the experimental conditions in which it was derived and propagated—even two cell lines derived from the same organism may have different characteristics.

It is more reasonable to request that investigators specify the sex of a cell line used in a study (*i.e.*, derived from a male vs female, or XX vs XY in sex chromosome complement), as the sex of many cell lines has been established (70). However, even this is not always possible, as cell lines can lose their sex chromosome complement over time (11). Although primary cultures can isolate cells directly from the body, permitting the creation of a small population of male or female cells, the procedure may be technically difficult and time-consuming, and the cells may be short-lived, limited in number, difficult to manipulate, and can change their characteristics over time in culture. Furthermore, Miller *et al.* (14) caution that the hormonal environment of cultured cells, including some media, can affect experimental outcomes. Finally, comparisons of isolated male and female cells oversimplify the question of sex, let alone gender, because such cells are removed from the complex interactions with other cells, hormones, neurotransmitters, nutrients, pathogens, and environmental exposures, which themselves vary in living organisms by sex and gender (11). In such cases, the absence of evidence for sex

“The hormonal environment of cultured cells...can affect experimental outcomes.”

differences *in vitro* may well be absence of any evidence at all, a straw man (or woman) of an experiment purportedly about sex.

Randomization by sex and/or gender

Experimentalists, particularly those conducting studies with >100 subjects, may wish to randomize the sexes separately to ensure similar distributions of treated and untreated males and females. In preclinical experiments, this is known as a factorial design (15). Such stratified randomization retains the advantages of standard random allocation, effectively creating a mini-trial within each sex stratum (71). Stratified randomization can accommodate a study plan with unequal numbers of male and female subjects, especially helpful when men and women join a study at different rates or in different time periods. Stratified randomization can also be used to balance follicular vs luteal phase participants, or any other marker of sex or gender.

Sample size considerations for studies including males and females

Most studies are planned from the outset with a sample size just large enough to afford 80% statistical power to detect the main effect of the primary exposure. Unless preplanned, most studies are underpowered to examine associations separately for males and females. This is particularly true of secondary data analyses of studies never designed to examine subgroup differences. This lack of statistical power to detect sex and gender differences can lead to the premature conclusion that such differences do not exist; in fact, most studies are simply too small to fairly test all but the most pronounced sex and gender differences. In the current era of accountability to analyze and present sex-stratified data, it is worth considering ideal practice and reality with respect to power and sample sizes to detect sex differences. Although most researchers will find that limited samples and funds constrain their ability to investigate sex and gender differences, we will also address the special case of “big data,” where problems may ensue from an abundance of statistical power to detect trivial differences, rather than too little power to detect meaningful differences.

Effect modification and interaction by sex

Epidemiologists and clinical researchers are familiar with the concepts of effect modification and interaction, although the terminology may differ between disciplines. However, basic investigators, whose aim is usually to limit all variation other than the exposure under examination, may be less familiar with these issues. “Effect modification” refers to the ability of a third variable (here, sex) to modify or interact with the “main effect” of the exposure (say, treatment) on outcome (usually, disease). For example, the

association of diabetes with cardiovascular disease (CVD) is stronger for women than men (72, 73); it is said that sex “interacts” with diabetes to cause CVD or that sex “modifies” the diabetes–CVD association.

Although stratifying data by sex to examine the exposure–disease association separately for males and females allows the investigator to eyeball effect modification by sex, such estimation gives no indication of the extent to which any observed sex differences are due to chance. To gauge this likelihood, many researchers test the statistical significance of sex differences by incorporating into their statistical models a (usually multiplicative) “interaction” term that represents the intersection of exposure and sex. For example, if the main effect of treatment is represented as a binary variable (0 if untreated; 1 if treated) and the main effect of sex as a binary variable (0 if male; 1 if female), then an interaction term (treatment \times sex) which equals 1 only for treated females will, when modeled with the main effects of treatment and sex, capture the additional increment or decrement in the risk of the outcome that is attributable to both treatment and female sex, that is, the sex difference in the association between treatment and disease. By convention, P values <0.05 for such interaction terms are indicative of a statistically significant sex difference, one that is unlikely due to chance alone. Such tests of effect modification or interaction by sex can (and should) be as easily incorporated into basic research as in clinical and population research. The difficulty is having the statistical power to do so.

Ideal: statistical power to detect interaction by sex

The sample size required to detect statistically significant sex differences (interactions by sex) is considerably larger than that required to detect the main effects of treatment or sex alone. Statistical power to detect a sex difference depends on the prevalence of the exposure, outcome, and sex, as well as the strength of the associations between them. Software is freely available to calculate sample sizes to detect interactions (74, 75). However, the rule of thumb is that it takes fourfold the sample size to detect an interaction than it does to detect main effects (*i.e.*, treatment or sex alone) (76). Investigators need to take into account differential disease rate by sex and the expected magnitude of the main effect in each sex; statistical power to detect either main effects or a sex interaction may not be optimized by recruiting half women and half men. In planning, investigators may have to make “best guesses” at the magnitude of expected sex differences, based on the literature and biologic understanding. As with any power calculation, it is best to input a range of likely main effects and interactions to evaluate the impact of sample size on the ability to detect a sex interaction.

A study that is large enough to detect a sex interaction, if one exists, represents the “ideal” in sex

difference studies. Few studies are planned with the power to detect statistically significant sex differences. Many studies that have attempted to test interactions by sex have been woefully underpowered to do so. Unfortunately, researchers easily forget that an interaction P value >0.05 often says as much about the design and size of the study as it does about the presence or absence of a sex difference.

Next best: statistical power to detect main effects within sex strata

Even where a study is too small to test for sex interaction effects, it may still have enough statistical power to examine the main effects of exposure within sex strata. This is simple sex stratification to examine exposure–disease associations for each sex. (Does diabetes predict CVD among males? Does diabetes predict CVD among females?) A study may find a statistically significant beneficial impact of treatment on disease among males and fail to find a significant effect of treatment among females (or, in extreme cases, find statistically significant benefits or harms that vary by sex). However, if the study lacks power to test an interaction by sex, investigators cannot claim that they have detected a difference between males and females that meets conventional standards for ruling out chance. As discussed, the detection of a statistically significant interaction by sex is a high bar. However, apparent contrasts in sex-stratified data—differential main effects of treatment by sex—can suggest the presence of sex differences. At the least, they provide a rationale for larger studies powered to detect sex interactions, or incentivize data collection across studies for meta-analyses of interactions by sex (15).

To plan a study with adequate statistical power to detect main effects by sex is straightforward: simply calculate sample sizes needed to detect main effects in men and women separately (and add them up), taking into account sex differences in rates of disease, expected size of impact of exposure, and, for observational studies, expected prevalence of exposure.

Many studies analyze their data by sex as an afterthought. Such subgroup analyses of main effects stratified by sex are often underpowered, which heightens the risk of type II error, or false-negative results. This is true even when the original analysis, in which all subjects are analyzed together, regardless of sex, reports a statistically significant association of exposure with disease. For example, in a study in which the exposure–disease association approaches statistical significance (say, a 2 standard error difference in outcome between study arms), splitting subjects into two groups of similar size will yield a one in three chance that the association will be sizeable and statistically significant ($P < 0.05$) in one group and inconsequential in the other (less than a standard error difference) (77).

The danger of such underpowered comparison of “main effects” among males and females is that, by chance alone, one out of every six studies reporting an overall effect in the combined sample will, on sex stratification, report a significant effect in females (but not males) and one out of six will report a significant effect among males (but not females). Add to this the additional risk that studies that are null in the aggregate may also report statistically significant findings exclusive to males or females. This, of course, leads to a literature of contradictions.

Better than nothing: representation of sex

Studies underpowered to detect even sex-stratified main effects can still make available data and/or analyses stratified by sex, particularly in supplemental material, without making inferences regarding sex differences *per se*. Such data may serve as preliminary analyses for future studies adequately powered to detect sex differences and may be used in meta-analyses.

Special considerations for “big data”

We have entered an era in which enormous datasets are increasingly available. Many of them include the variable “sex.” Two cautions are important to emphasize. First, such datasets, while deep in sample size, are often narrow in breadth, lacking the variables (discussed below) helpful to contextualize and understand sex differences. Second, the temptation in very large datasets to stratify by any variable is strong, as it is easy to detect statistically significant interactions, including sex differences, of clinically trivial and meaningless magnitude (78). Sound motivation to test sex differences, discussed earlier, is essential. So is conservative interpretation of statistically significant findings. To whom much data are given, much common sense is demanded: extra caution needs to be exercised in interpreting studies with enormous statistical power to detect minute differences between subgroups.

“Defining gender in human studies is both difficult and controversial.”

Data Collection

Truly sex-informed research is more than just stratifying by sex or gender. Researchers should collect the data to characterize the ways in which exposures, diseases, and contributing environmental factors vary by sex and gender.

Defining and measuring sex and gender

Although it can be difficult to determine the sex of subjects in some species, for the most part, the sex of humans and nonhuman subjects in biomedical research is known. Categories of sex include males, females, intersexual individuals born with male and female characteristics, and people who undergo

interventions to reassign their sex (25). In some instances, syndromes resulting from atypical sexual development can complicate categorization of sex (79).

Defining gender in human studies is both difficult and controversial. Indeed, some have argued that sex and gender are “irreducibly entangled,” and that even the most seemingly straightforward presentation of sex as a biological variable in human studies is inevitably a mix of sex and gender (24, 25, 80). Sociologists Westbrook and Saperstein (81), observing the tendency of large surveys to conflate sex and gender, call the state of measurement a “conceptual muddle” that is fraught with essentialist treatment of sex and gender as synonymous, obvious, easily determined by others, and unchanging over the life course.

The very concept of gender is subtle, complex, and shifting. It has been suggested that gender comprises at least three distinct, but interrelated components, the “three dimensions of gender” (82). These include: (1) our physical bodies, how we experience them, and how others interact with them; (2) our gender identity, our internal sense of ourselves as female, male, a blend of both, or neither; and (3) our gender expression, how we present our gender and how society interacts with the gender we present. Note that these dimensions are independent of sexual orientation. We are likely to see new measures of gender emerge; however, at present, there are few studies that have attempted to relate nuanced dimensions of gender to health and disease (81, 83).

In the meantime, some researchers have ventured measures of gender that are intended to be distinct from sex (84–86). For example, several gendered factors correlated with poor health among women have been proposed as proxies for gender influences on health, including income, education, labor force participation, single-headed household, unpaid child and elder care, unpaid housework, political participation, and access to education or health (86). Particularly problematic has been the identification of proxies for male gender that might influence men’s health. The prevalence of gun ownership, for example, has been proposed (84). Such measures of gender are often measures of gender inequality. Many times they are based on national or state-level statistics, rather than more granular individual or household data (86).

Pelletier *et al.* (85) have proposed a method to measure individual-level gender as “psychosocial sex,” in contrast to “biological sex.” They argue that, as gender roles and attitudes—components that might comprise a gender index—depend on culture, age, and era, no single gender scoring system is broadly applicable. Rather, a method for defining gender within a study population is a better approach to measure gender. Drawing from extensive questionnaires completed by their study participants, the researchers identified a set of seven variables (including income, hours doing housework, and scores on a sex role

inventory survey) that resulted in a continuous gender score ranging from masculine to feminine characteristics. Independent of sex, a high gender score (more feminine characteristics) was associated with increased risk of diabetes, hypertension, and depression and anxiety symptoms (85). In fact, once gender was accounted for, sex no longer predicted these health outcomes. Although the study was not large enough to exclude a modest interaction between sex and gender (*i.e.*, did the gender score predict outcomes more among males or among females?), the authors observed that the higher femininity score appeared to predict outcomes for men as well as for women (Louise Pilote, personal communication). This study was possible only because of the extensive collection of economic and psychosocial covariates related to gender. To the extent possible, studies should be designed to collect data on gender. However, lack of data with which to construct a comprehensive gender measure does not absolve investigators of considering gender in their interpretation of data regarding sex differences.

Is gender relevant to animal studies? If it is hard to measure gender in human beings, it would seem entirely alien to do so in other species. However, a few investigators have attempted to design exposures that mimic human gendered experiences. For example, Shors *et al.* (87) developed an animal model (sexual conspecific aggressive response, or SCAR) to examine the effects of sexual aggression on the brain and learned behaviors. Pubescent female rodents are paired with sexually experienced adult males. The female releases high levels of adrenal stress hormones. Her ability to learn, including to learn maternal caring behavior, was suppressed. The authors suggested that such experiments are aimed at understanding how sexual trauma impacts mechanisms that shape the female brain. Although other interpretations of that animal model are possible, studies have reported that women with a history of childhood sexual trauma exhibit changes in brain and associated physiology. Women with a history of childhood sexual trauma, a highly prevalent exposure, have irregularities to cortical and subcortical tissue and long-term alterations to their hypothalamic–pituitary adrenal axis, compared with women without childhood sexual trauma (88, 89). Sexual assault occurs to all sexes and genders, but considerably more often to girls and women, and therefore constitutes a gendered exposure (58–61). National surveys show that physical child abuse is also common, often more so for boys than girls (59, 60). Other violent exposures, such as combat casualties and war-time trauma, also have gendered distributions and implications (90, 91).

Measuring sex and gender differences in disease presentation

It is essential to capture outcomes in sufficient detail to detect sex and gender differences in disease

presentation. The classic example in clinical research is CHD, one of the leading causes of morbidity and mortality for men and women in the United States. Traditionally, myocardial infarction was characterized as the result of obstruction of the large coronary arteries. However, up to one third of women with a myocardial infarction and two thirds of women with chest pain had unobstructed arteries upon angiography (92, 93). It is now recognized that myocardial ischemia may result from disease of the coronary microvessels. The Women's Ischemia Syndrome Evaluation study reported that roughly half of the women with angina and ischemia without coronary artery obstruction evidenced microvascular dysfunction (94). Instead of the classic chest-crushing sensation of coronary artery obstruction, women with microvessel disease may present with shortness of breath and fatigue, nonspecific symptoms easy to misdiagnose and often dismissed. Thus, an investigator studying CHD in both sexes needs to consider the symptoms and diagnostic tests that will capture the presentation of disease in women and men (95).

Another example of differential disease presentation by sex is the tendency for prolactinomas to be detected as microadenomas among women, but macroadenomas among men. This results, at least in part, from the earlier detection among women, in whom small elevations in prolactin can cause infertility, menstrual disturbances, and/or galactorrhea. In contrast, in men, prolactinomas may progress to macroadenomas before they become symptomatic with headaches, double vision, or vision loss resulting from the mass of the tumor pressing on neurologic structures in the brain. Although this difference between the sexes is largely attributed to differences in diagnostic timing, the possibility that prolactinomas are more aggressive in men has not been entirely excluded (96, 97). In animal models, sex differences in the expression and activity of pituitary transforming growth factor $\beta 1$ may contribute to sex differences in prolactinoma incidence (98).

Sex and gender differences in drug exposures and metabolism

Studies of drugs need to account for differential impacts of dose on males and females, variations that may stem from disparities in body size or volume of distribution or other sex differences in pharmacokinetics or pharmacodynamics. For example, women have less intestinal enzymatic activity and slower renal clearance that may impact the metabolism of many drugs. Recently in the news is the sleeping medication zolpidem (the active ingredient in Ambien), which lingers in the bloodstream for hours longer in women than men given the same dose (99). This may be related to a modulating impact of testosterone on

CYP3A4 activity, the principal enzyme responsible for zolpidem metabolism; lower levels of free testosterone may explain the slower clearance and higher plasma concentrations of zolpidem in women and elderly men (100, 101). The Food and Drug Administration has recommended that the dose of Ambien be halved for women (102). So, too, men achieve lower blood levels of alcohol for the same ethanol consumption, due to the enhanced activity of alcohol dehydrogenase among males. The dynamics of drugs may vary, too: β -blockers, opioids, selective serotonin reuptake inhibitors, and several antipsychotic drugs may attain higher blood levels and/or have different efficacy in women (2, 23). In animal studies, too, the dynamics of drug metabolism may vary by sex; for example, female mice have higher fat mass and lower lean body mass than do males, something that should be considered in dose calculation of drugs.

The use of exogenous hormones, such as oral contraceptives, menopausal hormone therapy, testosterone, and anabolic steroids, is particularly important to document. Taken systemically, by mouth, injection, or patch, such drugs affect reproductive and nonreproductive systems throughout the body, and they could be important to investigations in which the exposure–disease relationship could be affected by sex hormones. In the United States, sales of testosterone, available as oral medicine, gel, patch, or injection, grew 77% from 2010 to 2013, with 2.3 million prescriptions filled (103). It is estimated that 2.9 to 4 million Americans, largely men, have used anabolic steroids in their lifetime (104).

In addition to their direct impact on brain, bone, muscle, metabolism, immune, cardiovascular, and reproductive function (105–107), reproductive steroids often interact with other drugs. For example, among patients with growth hormone deficiency, women taking oral estrogen require twice as much growth hormone as men or women not taking oral estrogen to achieve the same levels of insulin-like growth factor 1 (108); current guidelines for treatment of adult growth hormone deficiency now recommend the consideration of estrogen status in dosing (109). Basic scientists may find it illuminating to vary the levels of exogenous hormone exposures in their experiments to mimic widespread human exposures.

In addition to intentional exogenous hormone exposure, there is an increasing body of literature suggesting that exposure to endocrine disrupting chemicals in the environment may affect human health. For example, phthalates are a nearly ubiquitous class of chemicals used in the manufacturing of household products, including food packaging and personal care products such as cosmetics and nail polish. Exposure to phthalates may depend on occupation and use of personal care products; higher urine concentrations of phthalate metabolites have

"The failure to consider exogenous hormone use...may contribute to the lack of reproducibility in many studies."

been reported among women compared with men (110, 111). Phthalate exposures have been associated with insulin receptor and glucose oxidation in the Chang liver cell line (unspecified sex) (112), signs of diabetes and endocrine disruption in female rats (113), and with insulin resistance and diabetes in men and women (111, 114, 115).

The failure to consider exogenous hormone use, endogenous hormones, and/or markers of hormonal status (such as menopause) may contribute to the lack of reproducibility in many studies. For example, investigators found that the serum concentrations of 68% of 171 serum biomarkers associated with chronic disease were affected by sex, oral contraceptive use, menstrual phase, or menopausal status (116). They estimated up to 40% false discoveries in biomarkers when sex was ignored and up to 41% false discoveries when oral contraceptive use was ignored. Heeding this caution, investigators may want to collect data on menstrual or estrus phase, menopausal status, use of exogenous hormones, and/or levels of circulating hormones (14).

Measuring reproductive cycle and phase

There are nuances to asking women to report their menstrual cycle and menopausal status. For example, as the duration of the luteal phase varies less than that of the follicular phase, menstrual cycle timing is best recorded retrospectively from the first day of the next menstrual period (117). As menstrual cycles may be suppressed or dictated by hormonal contraceptives (including hormonal intrauterine devices), or breastfeeding, it is useful to record these variables when assessing menstrual timing. Menopause may occur naturally or may result from hysterectomy, oophorectomy, or chemotherapy, and it cannot be determined until a year after the last menstrual period. Measurement of menstrual cycle phase and menopausal transition are covered elsewhere (117–119). Archived biospecimens should include information about such variables (e.g., time of blood draw, day of menstrual cycle).

Influential covariates that vary by sex and gender

Some factors that vary by sex or gender can influence the exposure–disease association under study, either as confounders (easily mistaken for sex differences) or effect modifiers (covariates that interact with sex to change outcome). Some of these sex-dependent covariates are obvious, such as parity. Other aspects of reproductive history may affect nonreproductive systems under study (120). For example, history of the hypertensive pregnancy disorder preeclampsia predicts twofold higher risk of CVD in women affected by the disorder (121). A woman's history of preeclampsia might modify the impact of an antihypertensive drug. Exposures to exogenous endocrine drugs, such as those administered in the course of assisted reproductive

technologies such as *in vitro* fertilization, might affect systems under study.

The degree to which other exposures, such as cigarette smoking, alcohol use, physical activity, socioeconomic position, caretaking responsibilities, and medication use, vary by sex and gender will depend on the population under study. For example, in some, but not all, cultures and climates, circulating 25-hydroxyvitamin D concentrations may differ considerably between men and women as a result of gender differences in factors such as clothing, time spent outdoors, and supplement use (122–124); depending on the country, these differences may be equalized by dietary vitamin D intake, particularly of fortified foods. Additionally, as shown in a study in the Netherlands, lower 25-hydroxyvitamin D levels among women may be explained by their higher adiposity levels, a difference that could be attributed either to sex (a biological difference) or gender (as many social determinants of adiposity are gendered) (125).

Particularly important to consider are sex or gender differences in the distribution of comorbidities that might influence an exposure–disease association. For example, compared with diabetic women, diabetic men have lower prevalence of depression and anxiety, gendered psychosocial factors that impede self-care activity and treatment success (126). Thus, it would be wise for studies examining sex differences in CVD to account for major depression history, especially when depression is associated with the main exposure under consideration (21).

Analysis, Reporting, and Interpretation of Sex Differences

The problem of stratifying everything by sex

Subgroup analyses from studies thoughtfully designed to query sex differences, particularly once replicated, can provide sound evidence of benefit or protection from harm for women and men. Alternatively, *post hoc* sex difference analyses, devoid of theoretical basis and sound construction, may create more noise than light. A recent analysis of sex differences presented in Cochrane reviews of clinical trials suggested that few met the stringent criteria of documenting statistically significant interactions by sex (127); this criticism of sex differences research is cautionary. Whether the absence of sex differences reflects fact, indiscriminate testing, lack of sample size, or the decades it takes for sex differences observed in basic or population research to reach clinical testing remains to be seen (128).

The risk of a blanket mandate to require all studies to stratify all results by sex is that the literature-wide type I error, that is, the risk of detecting false sex differences, will skyrocket. Furthermore, as we increase the size and statistical power of our studies to detect

true sex interactions (minimizing type II error), we court the risk of finding sex differences where none exist (type I error). As mentioned earlier, type I error is a particular hazard of a theoretical “big data” analysis.

If we pursue sex difference analysis as a poorly considered mandate, a literature of contradiction will follow. The field of sex differences research risks discredit from unthinking and profligate enthusiasm. How, then, can we encourage sex differences research that is thoughtful, conservative, and consistent over time?

Prespecified hypothesis tests

In any subgroup analysis, including sex and gender, tests of interaction should be limited and prespecified in statistical analytic plans (129). Although this does not guarantee that tests of hypotheses will be well constructed, it does help to protect against *post hoc* fishing and data-derived hypothesis testing (themselves self-fulfilling prophecies). “Surprise” subgroup findings should be presented as such, and interpreted with caution—the basis for further study, not for instant translation to clinic or policy. There seems almost a reflexive tendency of researchers to view male and female as the fundamental dichotomy of the biologic world (24). We need to approach the question of sex differences with curiosity and skepticism, rather than unquestioning assumption.

In creating an *a priori* hypothesis, it is best practice to prespecify the expected direction and magnitude of the sex difference. Should there be subgroups within sex, such as nulliparous vs parous, or premenopausal vs postmenopausal? Careful *a priori* hypothesizing is important for observational studies, experiments, and trials, and it serves to maintain scientific transparency. In large studies, some statistically significant sex differences may arise by chance. Thus, prespecifying the form of the expected interaction helps guard against indiscriminate *post hoc* scrambling.

Accounting for confounding

Here, careful data collection bears fruit. Where a third factor (say, smoking) is associated both with sex (say, male sex) and outcome (e.g., lung cancer), it could be easy to misattribute the disease outcome to sex, for example, concluding that lung cancer is a male disease. There are several techniques for adjusting for measured confounding factors, discussed elsewhere (130). Briefly, typical methods to reduce the likelihood that an exposure–outcome association is confounded by a third variable include: adding to the basic exposure–outcome model additional independent covariates that represent the potential confounder (e.g., a term to represent number of cigarettes per day or a set of terms to indicate nonsmokers, past smokers, and present smokers); matching on the potential confounder using careful participant selection and conditional modeling

(e.g., smokers can be compared only with other smokers, and nonsmokers with other nonsmokers); or conditioning on a propensity score that estimates from a set of covariates the likelihood that an individual was exposed, thus reducing or eliminating the unwanted impact of those covariates on the estimated exposure–outcome association (131). Where data on potential confounders are unavailable (often the case with “big data”), simulation analyses can be used to gauge the impact of unmeasured confounders on the association of sex or gender with disease or exposure–disease association (132, 133).

Variables on the pathway between sex or gender and outcome

Some variables may be intermediates between sex or gender and the outcomes under study, and their treatment in analysis deserves special consideration. For example, sex and gender are two of many factors that determine body size and composition. Not only are body size and adiposity determined by sex steroids, sex steroid receptors, adipokines, and other differences between males and females (134), but gender differences in physical activity also affect body size and composition (135). [Interestingly, there are also sex differences in the voluntary physical activity of rodents: females exercise more than males. (135)] Whether to control for body size in studies of sex and gender effects is a nuanced decision.

For example, in a study of differences between men and women in the impact of Ambien and impaired driving, should the investigator adjust for body size in evaluating sex or gender differences? Alternatively, body size is (at least partially) a product of sex and gender, and adjusting for it might obscure the most important pathway (body size) through which sex and gender impact the metabolism of the drug. Additionally, adjustment for body size might reveal other mechanisms (some of which are discussed in “Sex and gender differences in drug exposures and metabolism” above) through which sex or gender affect drug clearance. Statistical methods can help segregate or ‘decompose’ the impact of such intermediate variables, also known as mediators (136, 137).

“...Researchers will be collecting and analyzing data by sex, but the onus is on investigators to address this adequately and at all levels of basic, clinical, and population research.”

Replication

The critical importance of replication has been addressed by others (78). This is particularly true in a scientific climate that encourages, or even mandates, subgroup analysis.

Interpretation

As with any study finding, apparent sex differences (or lack thereof) need to be interpreted with caution (138). The magnitude and direction of apparent sex effects need to be placed in the context of prior knowledge. Biological mechanisms, hopefully outlined *a priori*, need to be discussed. The adequacy of the study to rule

out bias, confounding, and chance needs to be frankly addressed. Even statistically significant sex differences may be due to chance or bias, instead of true heterogeneity of exposure–disease associations or of treatment effects (129).

This is especially true when interpreting main effects stratified by sex, where a study lacks statistical power to test interaction by sex. In this case, the play of chance is often overlooked and findings are overinterpreted. As an example, Assman *et al.* (139) cite a subgroup analysis of a trial that followed myocardial infarction survivors (140). The investigators, laudably, considered the differential impact of treatment on the mortality of men and women, including both sexes and prespecifying stratification by sex in the analyses. However, they did not plan to test an interaction by sex. Their intervention had no overall impact on mortality. There was no association of treatment with cardiac mortality among men ($P = 0.94$). However, in women, the authors observed what they interpreted as a “possible harmful impact of the intervention” on women’s cardiac mortality ($P = 0.06$). Later, Assman *et al.* used their data to calculate a proper test of interaction by sex, which revealed no statistically significant evidence of an interaction between treatment and sex (P for interaction = 0.21), indicating that chance could well explain the seeming sex effect. Thus, despite the disparate associations and P values in the two sex strata, the study was simply too small to test whether the impact of the treatment on cardiac mortality truly differed by sex.

Most importantly, surprise subgroup findings need to be acknowledged as such. Sex differences that were discovered as the result of *post hoc* poking around in the data need to be treated with caution until they are replicated as *pre hoc* tests in other studies. In this event, supplemental tables stratified by sex, data repositories, and meta-analyses may extend the impact of any single study. In other words, investigators can make available sex-stratified data to spur the generation of new hypotheses, without presenting sex-stratified analyses that overreach the intent and design of their original study.

Conclusions

New governmental mandates mean that researchers will be collecting and analyzing data by sex, but the onus is on investigators to address this adequately and at all levels of basic, clinical, and population research. If we fail, the “noise” created by multiple testing across all our datasets may drown out the signal of true sex differences. Furthermore, in human studies it is important to investigate the impact of both sex and gender to illuminate fundamental, modifiable causes of disease and avoid a reflexive attribution of seeming sex differences solely to biology. If we address these design and analytic issues skillfully, then we have the chance for new insights for men and women that will be critical for the next generation of scientific and therapeutic discoveries in this age of precision medicine.

References

1. National Institutes of Health. Consideration of sex as a biological variable in NIH-funded research. Available at: grants.nih.gov/grants/guide/notice-files/NOT-OD-15-102.html. Accessed 17 April 2017.
2. Legato MJ, Johnson PA, Manson JE. Consideration of sex differences in medicine to improve health care and patient outcomes. *JAMA*. 2016;316(18):1865–1866.
3. Goldstein JM, Holsen L, Handa R, Tobet S. Fetal hormonal programming of sex differences in depression: linking women’s mental health with sex differences in the brain across the lifespan. *Front Neurosci*. 2014;8:247.
4. Institute of Medicine. *Exploring the Biological Contributions to Human Health: Does Sex Matter?* Washington, DC: National Academies Press; 2001.
5. Johnson J, Sharman Z, Vissandjee B, Stewart DE. Does a change in health research funding policy related to the integration of sex and gender have an impact? *PLoS One*. 2014;9(6):e99900.
6. European Commission Directorate-General for Research and Innovation. H2020 programme: guidance on gender equality in Horizon 2020. Available at: eige.europa.eu/sites/default/files/h2020-hi-guide-gender_en.pdf. Accessed 5 April 2018.
7. Rabesandratana T. Adding sex-and-gender dimensions to your research. Available at: www.sciencemag.org/careers/2014/03/adding-sex-and-gender-dimensions-your-research. Accessed 12 February 2018.
8. National Institutes of Health. Enhancing reproducibility through rigor and transparency. Available at: grants.nih.gov/grants/guide/notice-files/NOT-OD-15-103.html. Accessed 17 April 2017.
9. National Institutes of Health/Agency for Healthcare Research and Quality. Implementing rigor and transparency in NIH & AHRQ research grant applications. Available at: grants.nih.gov/grants/guide/notice-files/NOT-OD-16-011.html. Accessed 17 April 2017.
10. National Institutes of Health/Agency for Healthcare Research and Quality. Implementing rigor and transparency in NIH & AHRQ career development award applications. Available at: grants.nih.gov/grants/guide/notice-files/NOT-OD-16-012.html. Accessed 17 April 2017.
11. Ritz SA, Antle DM, Côté J, Deroy K, Fraleigh N, Messing K, Parent L, St-Pierre J, Vaillancourt C, Mergler D. First steps for integrating sex and gender considerations into basic experimental biomedical research. *FASEB J*. 2014;28(1):4–13.
12. Becker JB, Arnold AP, Berkley KJ, Blaustein JD, Eckel LA, Hampson E, Herman JP, Marts S, Sadee W, Steiner M, Taylor J, Young E. Strategies and methods for research on sex differences in brain and behavior. *Endocrinology*. 2005;146(4):1650–1673.
13. Ouyang P, Wenger NK, Taylor D, Rich-Edwards JW, Steiner M, Shaw LJ, Berga SL, Miller VM, Merz NB. Strategies and methods to study female-specific cardiovascular health and disease: a guide for clinical scientists. *Biol Sex Differ*. 2016;7(1):19.
14. Miller VM, Kaplan JR, Schork NJ, Ouyang P, Berga SL, Wenger NK, Shaw LJ, Webb RC, Mallampalli M, Steiner M, Taylor DA, Merz CN, Reckelhoff JF. Strategies and methods to study sex differences in cardiovascular structure and function: a guide for basic scientists. *Biol Sex Differ*. 2011;2(1):14.
15. Miller LR, Marks C, Becker JB, Hurn PD, Chen WJ, Woodruff T, McCarthy MM, Sohrabji F, Schiebinger L, Wetherington CL, Makris S, Arnold AP, Einstein G, Miller VM, Sandberg K, Maier S, Cornelison TL, Clayton JA. Considering sex as a biological variable in preclinical research. *FASEB J*. 2017;31(1):29–34.

16. Cornelison TL, Clayton JA. Considering sex as a biological variable in biomedical research. *Gender and the Genome*. 2017;1(2):89–93.

17. Nieuwenhoven L, Klinge I. Scientific excellence in applying sex- and gender-sensitive methods in biomedical and health research. *J Womens Health (Larchmt)*. 2010;19(2):313–321.

18. Legato MJ. *Principles of Gender-Specific Medicine: Gender in the Genomic Era*. 3rd ed. London, UK: Academic Press; 2017.

19. Arnold AP, Lusis AJ. Understanding the sexome: measuring and reporting sex differences in gene systems. *Endocrinology*. 2012;153(6):2551–2555.

20. Mosca L, Barrett-Connor E, Wenger NK. Sex/gender differences in cardiovascular disease prevention: what a difference a decade makes. *Circulation*. 2011;124(19):2145–2154.

21. Tober SA, Handa RJ, Goldstein JM. Sex-dependent pathophysiology as predictors of comorbidity of major depressive disorder and cardiovascular disease. *Pflugers Arch*. 2013;465(5):585–594.

22. Anastasio M, Salafia CM, Fitzmaurice G, Goldstein JM. Impact of fetal versus perinatal hypoxia on sex differences in childhood outcomes: developmental timing matters. *Soc Psychiatry Psychiatr Epidemiol*. 2012;47(3):455–464.

23. Whitley H, Lindsey W. Sex-based differences in drug activity. *Am Fam Physician*. 2009;80(11):1254–1258.

24. Springer KW, Mager Stellman J, Jordan-Young RM. Beyond a catalogue of differences: a theoretical frame and good practice guidelines for researching sex/gender in human health. *Soc Sci Med*. 2012;74(11):1817–1824.

25. Krieger N. Genders, sexes, and health: what are the connections—and why does it matter? *Int J Epidemiol*. 2003;32(4):652–657.

26. Sen G, Ostlin P. Gender inequity in health: why it exists and how we can change it. *Glob Public Health*. 2008;3(sup1, Suppl 1):1–12.

27. Garcia M, Mulvagh SL, Merz CN, Buring JE, Manson JE. Cardiovascular disease in women: clinical perspectives. *Circ Res*. 2016;118(8):1273–1293.

28. Jazin E, Cahill L. Sex differences in molecular neuroscience: from fruit flies to humans. *Nat Rev Neurosci*. 2010;11(1):9–17.

29. Manson JE, Bassuk SS, Lee IM, Cook NR, Albert MA, Gordon D, Zaharris E, MacFadyen JG, Danielson E, Lin J, Zhang SM, Buring JE. The VITamin D and OmegA-3 Trial (VITAL): rationale and design of a large randomized controlled trial of vitamin D and marine omega-3 fatty acid supplements for the primary prevention of cancer and cardiovascular disease. *Contemp Clin Trials*. 2012;33(1):159–171.

30. King AJ. The use of animal models in diabetes research. *Br J Pharmacol*. 2012;166(3):877–894.

31. Clutton-Brock TH, Ivarsson K. Sex differences in ageing in natural populations of vertebrates. *Proc Biol Sci*. 2007;274(1629):3097–3104.

32. Lyon MF. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature*. 1961;190(4773):372–373.

33. Busque L, Mio R, Mattioli J, Brais E, Blais N, Lalonde Y, Maragh M, Gilliland DG. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood*. 1996;88(1):59–65.

34. Abkowitz JL, Taboada M, Shelton GH, Catlin SN, Guttrop P, Kiklevich JV. An X chromosome gene regulates hematopoietic stem cell kinetics. *Proc Natl Acad Sci USA*. 1998;95(7):3862–3866.

35. Barrett EL, Richardson DS. Sex differences in telomeres and lifespan. *Aging Cell*. 2011;10(6):913–921.

36. Astuti Y, Wardhana A, Watkins J, Wulaningsih W; PILAR Research Network. Cigarette smoking and telomere length: a systematic review of 84 studies and meta-analysis. *Environ Res*. 2017;158:480–489.

37. Resende MM, Taylor DA. Building solutions for cardiovascular disease in women. *Tex Heart Inst J*. 2013;40(3):285–287.

38. El-Maari O, Becker T, Junen J, Manzoor SS, Diaz-Lacava A, Schwaab R, Wienker T, Oldenburg J. Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. *Hum Genet*. 2007;122(5):505–514.

39. Zhu ZZ, Hou L, Bollati V, Tarantini L, Marinelli B, Cantone L, Yang AS, Vokonas P, Lissowska J, Fustini S, Pesatori AC, Bonzini M, Apostoli P, Costa G, Bertazzi PA, Chow WH, Schwartz J, Baccarelli A. Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis. *Int J Epidemiol*. 2012;41(1):126–139.

40. Smith RP, Coward RM, Kovac JR, Lipshultz LI. The evidence for seasonal variations of testosterone in men. *Maturitas*. 2013;74(3):208–212.

41. Winston AP, Wijeratne S. Hypogonadism, hypo-leptinaemia and osteoporosis in males with eating disorders. *Clin Endocrinol (Oxf)*. 2009;71(6):897–898.

42. Masters JR, Drife JO, Scarisbrick JJ. Cyclic variation of DNA synthesis in human breast epithelium. *J Natl Cancer Inst*. 1977;58(5):1263–1265.

43. Anderson TJ. Mitotic activity in the breast. *J Obstet Gynaecol*. 1984;4(Suppl 2):S114–S118.

44. Sulke AN, Jones DB, Wood PJ. Variation in natural killer activity in peripheral blood during the menstrual cycle. *Br Med J (Clin Res Ed)*. 1985;290(6472):884–886.

45. Ruiz-Pino F, Navarro VM, Bentzen AH, Garcia-Galiano D, Sanchez-Garrido MA, Ciofi P, Steiner RA, Mikkelsen JD, Pinilla L, Tena-Sempere M. Neurokinin B and the control of the gonadotropin axis in the rat: developmental changes, sexual dimorphism, and regulation by gonadal steroids. *Endocrinology*. 2012;153(10):4818–4829.

46. de Vries CJ, Södersten P. Sex differences in the brain: the relation between structure and function. *Horm Behav*. 2009;55(5):589–596.

47. McCarthy MM, Pickett LA, VanRyzin JW, Kight KE. Surprising origins of sex differences in the brain. *Horm Behav*. 2015;76:3–10.

48. Semaan SJ, Tolson KP, Kauffman AS. The development of kisspeptin circuits in the mammalian brain. In: Kauffman AS, Smith JT, eds. *Kisspeptin Signaling in Reproductive Biology*. New York, NY: Springer; 2013:221–252.

49. Holsen LM, Lancaster K, Klibanski A, Whitfield-Gabrieli S, Cherkerzian S, Buka S, Goldstein JM. HPA-axis hormone modulation of stress response circuitry activity in women with remitted major depression. *Neuroscience*. 2013;250:733–742.

50. Jacobs EG, Weiss BK, Makris N, Whitfield-Gabrieli S, Buka SL, Klibanski A, Goldstein JM. Impact of sex and menopausal status on episodic memory circuitry in early midlife. *J Neurosci*. 2016;36(39):10163–10173.

51. Heldring N, Pike A, Andersson S, Matthews J, Cheng G, Hartman J, Tujague M, Ström A, Treuter E, Warner M, Gustafsson JA. Estrogen receptors: how do they signal and what are their targets. *Physiol Rev*. 2007;87(3):905–931.

52. Barton M, Filardo EJ, Lolait SJ, Thomas P, Maggiolini M, Prossnitz ER. Twenty years of the G protein-coupled estrogen receptor GPER: historical and personal perspectives. *J Steroid Biochem Mol Biol*. 2018;176:4–15.

53. Cosman F, Lindsay R. Selective estrogen receptor modulators: clinical spectrum. *Endocr Rev*. 1999;20(3):418–434.

54. Jacobsen BM, Horwitz KB. Progesterone receptors, their isoforms and progesterone regulated transcription. *Mol Cell Endocrinol*. 2012;357(1-2):18–29.

55. Marks LS. 5α-Reductase: history and clinical importance. *Rev Urol*. 2004;6(Suppl 9):S11–S21.

56. Looker AC, Johnson CL, Lacher DA, Pfeiffer CM, Schleicher RL, Sempos CT. Vitamin D status: United States, 2001–2006. *NCHS Data Brief*. 2011;59(59):1–8.

57. Tolin DF, Foa EB. Sex differences in trauma and posttraumatic stress disorder: a quantitative review of 25 years of research. *Psychol Bull*. 2006;132(6):959–992.

58. Breiding MJ, Smith SG, Basile KC, Walters ML, Chen J, Merrick MT. Prevalence and characteristics of sexual violence, stalking, and intimate partner violence—national intimate partner and sexual violence survey, United States, 2011. *MMWR Surveill Summ*. 2014;63(8):1–18.

59. Black MC, Basile KC, Breiding MJ, Smith SG, Walters ML, Merrick MT, Chen J, Stevens MR. *The National Intimate Partner and Sexual Violence Survey (NISVS): 2010 Summary Report*. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention; 2011.

60. Afifi TO, MacMillan HL, Boyle M, Taillieu T, Cheung K, Sareen J. Child abuse and mental disorders in Canada. *CMAJ*. 2014;186(9):E324–E332.

61. World Health Organization. *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence*. Geneva, Switzerland: World Health Organization; 2013.

62. Diaz Brinton R. Minireview: translational animal models of human menopause: challenges and emerging opportunities. *Endocrinology*. 2012;153(8):3571–3578.

63. Uhlenhaut NH, Treier M. Foxl2 function in ovarian development. *Mol Genet Metab*. 2006;88(3):225–234.

64. Drew BG, Hamidi H, Zhou Z, Villanueva CJ, Krum SA, Calkin AC, Parks BW, Ribas V, Kalajian NY, Phun J, Daraei P, Christofk HR, Hewitt SC, Korach KS, Tontonoz P, Lusis AJ, Slamon DJ, Hurvitz SA, Heveren AL. Estrogen receptor (ER)α-regulated lipocalin 2 expression in adipose tissue links obesity with breast cancer progression. *J Biol Chem*. 2015;290(9):5566–5581.

65. McDevitt MA, Glidewell-Kenney C, Jimenez MA, Ahearn PC, Weiss J, Jameson JL, Levine JE. New insights into the classical and non-classical actions of estrogen: evidence from estrogen receptor knock-out and knock-in mice. *Mol Cell Endocrinol*. 2008;290(1-2):24–30.

66. Park CJ, Zhao Z, Glidewell-Kenney C, Lazic M, Chamblon P, Krust A, Weiss J, Clegg DJ, Dunaif A, Jameson JL, Levine JE. Genetic rescue of nonclassical ERα signaling normalizes energy balance in obese Erα-null mutant mice. *J Clin Invest*. 2011;121(2):604–612.

67. Itoh Y, Mackie R, Kampf K, Domadia S, Brown JD, O'Neill R, Arnold AP. Four core genotypes mouse model: localization of the Sry transgene and bioassay for testicular hormone levels. *BMC Res Notes*. 2015;8(1):69.

68. Chen X, McClusky R, Chen J, Beaven SW, Tontonoz P, Arnold AP, Reue K. The number of X chromosomes causes sex differences in adiposity in mice. *PLoS Genet*. 2012;8(5):e1002709.

69. Li J, Chen X, McClusky R, Ruiz-Sundstrom M, Itoh Y, Umar S, Arnold AP, Eghbali M. The number of X chromosomes influences protection from cardiac ischaemia/reperfusion injury in mice: one X is better than two. *Cardiovasc Res*. 2014;102(3):375–384.

70. Shah K, McCormack CE, Bradbury NA. Do you know the sex of your cells? *Am J Physiol Cell Physiol*. 2014;306(1):C3–C18.

71. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *J Clin Epidemiol*. 1999;52(1):19–26.

72. Peters SA, Huxley RR, Woodward M. Diabetes as risk factor for incident coronary heart disease in women compared with men: a systematic review and meta-analysis of 64 cohorts including 858,507 individuals and 28,203 coronary events. *Diabetologia*. 2014; 57(8):1542–1551.

73. Huxley RR, Peters SA, Mishra GD, Woodward M. Risk of all-cause mortality and vascular events in women versus men with type 1 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol*. 2015;3(3):198–206.

74. Garcia-Closas M, Lubin JH. POWER V3.0 software. Available at: dceg.cancer.gov/tools/design/power. Accessed 5 April 2018.

75. VanderWeele TJ. Tools and tutorials. Available at: www.hsph.harvard.edu/tyler-vanderweele/tools-and-tutorials/. Accessed 5 April 2018.

76. Brookes ST, Whitley E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol*. 2004;57(3):229–236.

77. Petro R. Statistical aspects of cancer trials. In: Halnan KE, ed. *Treatment of Cancer*. London, UK: Chapman and Hall; 1982:867–871.

78. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.

79. MacLaughlin DT, Donahoe PK. Sex determination and differentiation. *N Engl J Med*. 2004;350(4): 367–378.

80. Hankivsky O, Doyal L, Einstein G, Kelly U, Shim J, Weber L, Repta A. The odd couple: using biomedical and intersectional approaches to address health inequities. *Global Health Action*. 2017; 10(Suppl 2):1326686.

81. Westbrook L, Saperstein A. New categories are not enough: rethinking the measurement of sex and gender in social surveys. *Gend Soc*. 2015;29(4): 534–560.

82. Gender Spectrum. Understanding gender. Available at: www.genderspectrum.org/quick-links/understanding-gender/. Accessed 4 April 2018.

83. Gender Identity in U.S. Surveillance (GenlUSS) Group. Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys. Herman JL, ed. Los Angeles, CA: The Williams Institute; 2014.

84. Phillips SP. Defining and measuring gender: a social determinant of health whose time has come. *Int J Equity Health*. 2005;4(1):11.

85. Pelletier R, Ditto B, Pilote L. A composite measure of gender and its association with risk factors in patients with premature acute coronary syndrome. *Psychosom Med*. 2015;77(5):517–526.

86. Tamambang L, Auger N, Lo E, Raynault M-F. Measurement of gender inequality in neighbourhoods of Québec, Canada. *Int J Equity Health*. 2011;10(1):52.

87. Shors TJ, Tobón K, DiFeo G, Durham DM, Chang HYM. Sexual conspecific aggressive response (SCAR): a model of sexual trauma that disrupts maternal learning and plasticity in the female brain. *Sci Rep*. 2016;6(1):18960.

88. Blanco L, Nydegger LA, Camarillo G, Trinidad DR, Schramm E, Ames SL. Neurological changes in brain structure and functions among individuals with a history of childhood sexual abuse: a review. *Neurosci Biobehav Rev*. 2015;57:63–69.

89. Heim C, Nemeroff CB. The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biol Psychiatry*. 2001;49(12):1023–1039.

90. Tepe V, Yarnell A, Nindl BC, Van Arsdale S, Deuster PA. Women in combat: summary of findings and a way ahead. *Mil Med*. 2016;181(Suppl 1):109–118.

91. Cross JD, Johnson AE, Wenke JC, Bosse MJ, Ficke JR. Mortality in female war veterans of operations enduring freedom and Iraqi freedom. *Clin Orthop Relat Res*. 2011;469(7):1956–1961.

92. Chokshi NP, Iqbal SN, Berger RL, Hochman JS, Feit F, Slater JN, Pena-Singh I, Yatskar L, Keller NM, Babaev A, Attubato MJ, Reynolds HR. Sex and race are associated with the absence of epicardial coronary artery obstructive disease at angiography in patients with acute coronary syndromes. *Clin Cardiol*. 2010;33(8):495–501.

93. Sharaf BL, Pepine CJ, Kerensky RA, Reis SE, Reichek N, Rogers WJ, Sopko G, Kelsey SF, Holubkov R, Olson M, Miele NJ, Williams DO, Merz CN; WISE Study Group. Detailed angiographic analysis of women with suspected ischemic chest pain (pilot phase data from the NHLBI-sponsored Women's Ischemia Syndrome Evaluation [WISE] Study Angiographic Core Laboratory). *Am J Cardiol*. 2001; 87(8):937–941.

94. Reis SE, Holubkov R, Conrad Smith AJ, Kelsey SF, Sharaf BL, Reichek N, Rogers WJ, Merz CN, Sopko G, Pepine CJ; WISE Investigators. Coronary microvascular dysfunction is highly prevalent in women with chest pain in the absence of coronary artery disease: results from the NHLBI WISE study. *Am Heart J*. 2001;141(5):735–741.

95. Sanghavi M, Gulati M. Sex differences in the pathophysiology, treatment, and outcomes in IHD. *Curr Atheroscler Rep*. 2015;17(6):34.

96. Ciccarelli A, Daly AF, Beckers A. The epidemiology of prolactinomas. *Pituitary*. 2005;8(1):3–6.

97. Klibanski A. Prolactinomas. *N Engl J Med*. 2010; 362(13):1219–1226.

98. Recouvreux MV, Faraoni EY, Camilletti MA, Ratner L, Abeledo-Machado A, Rulli SB, Diaz-Torga G. Sex differences in the pituitary TGF β 1 system: the role of TGF β 1 in prolactinoma development. *Front Neuroendocrinol*. 2017;50:91–3022(17):30063–8.

99. Greenblatt DJ, Harmatz JS, Singh NN, Steinberg F, Roth T, Molina ML, Harris SC, Kapil RP. Gender differences in pharmacokinetics and pharmacodynamics of zolpidem following sublingual administration. *J Clin Pharmacol*. 2014;54(3):282–290.

100. Cubala WJ, Wiglusz M, Burkiewicz A, Galuszko-Wegielniak M. Zolpidem pharmacokinetics and pharmacodynamics in metabolic interactions involving CYP3A: sex as a differentiating factor. *Eur J Clin Pharmacol*. 2010;66(9):955.

101. Olubodun JO, Ochs HR, von Moltke LL, Roubenoff R, Hesse LM, Harmatz JS, Shader RI, Greenblatt DJ. Pharmacokinetic properties of zolpidem in elderly and young adults: possible modulation by testosterone in men. *Br J Clin Pharmacol*. 2003;56(3): 297–304.

102. U.S. Food and Drug Administration. Questions and answers: risk of next-morning impairment after use of insomnia drugs: FDA requires lower recommended doses for certain drugs containing zolpidem (Ambien, Ambien CR, Edluar, and Zolpimist). Available at: www.fda.gov/Drugs/DrugSafety/ucm334041.htm#q6. Accessed 17 April 2017.

103. U.S. Food and Drug Administration. FDA drug safety communication. FDA cautions about using testosterone products for low testosterone due to aging: requires labeling change to inform of possible increased risk of heart attack and stroke with use. Available at: www.fda.gov/Drugs/DrugSafety/ucm436259.htm. Accessed 17 April 2017.

104. Pope HG Jr, Kanayama G, Athey A, Ryan E, Hudson JL, Baggish A. The lifetime prevalence of anabolic-androgenic steroid use and dependence in Americans: current best estimates. *Am J Addict*. 2014;23(4):371–377.

105. Fish EN. The X-files in immunity: sex-based differences predispose immune responses. *Nat Rev Immunol*. 2008;8(9):737–744.

106. Power ML, Schulkin J. Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins. *Br J Nutr*. 2008;99(5): 931–940.

107. Mendelsohn ME, Karas RH. Molecular and cellular basis of cardiovascular gender differences. *Science*. 2005;308(5728):1583–1587.

108. Cook DM, Ludlam WH, Cook MB. Route of estrogen administration helps to determine growth hormone (GH) replacement dose in GH-deficient adults. *J Clin Endocrinol Metab*. 1999; 84(11):3956–3960.

109. Molitch ME, Clemons DR, Malozowski S, Merriam GR, Vance ML; Endocrine Society. Evaluation and treatment of adult growth hormone deficiency: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2011;96(6):1587–1609.

110. Silva MJ, Barr DB, Reidy JA, Malek NA, Hodge CC, Caudill SP, Brock JW, Needham LL, Calafat AM. Urinary levels of seven phthalate metabolites in the U.S. population from the National Health and Nutrition Examination Survey (NHANES) 1999–2000. *Environ Health Perspect*. 2004;112(3): 331–338.

111. Huang T, Saxena AR, Isganaitis E, James-Todd T. Gender and racial/ethnic differences in the associations of urinary phthalate metabolites with markers of diabetes risk: National Health and Nutrition Examination Survey 2001–2008. *Environ Health*. 2014;13(1):6.

112. Rengarajan S, Parthasarathy C, Anitha M, Balasubramanian K. Diethylhexyl phthalate impairs insulin binding and glucose oxidation in Chang liver cells. *Toxicol In Vitro*. 2007;21(1):99–102.

113. Gayathri NS, Dhanya CR, Indu AR, Kurup PA. Changes in some hormones by low doses of di(2-ethyl hexyl) phthalate (DEHP), a commonly used plasticizer in PVC blood storage bags & medical tubing. *Indian J Med Res*. 2004;119(4):139–144.

114. Stahlhut RW, van Wijngaarden E, Dye TD, Cook S, Swan SH. Concentrations of urinary phthalate metabolites are associated with increased waist circumference and insulin resistance in adult U.S. males. *Environ Health Perspect*. 2007;115(6): 876–882.

115. James-Todd T, Stahlhut R, Meeker JD, Powell SG, Hauser R, Huang T, Rich-Edwards J. Urinary phthalate metabolite concentrations and diabetes among women in the National Health and Nutrition Examination Survey (NHANES) 2001–2008. *Environ Health Perspect*. 2012;120(9):1307–1313.

116. Ramsey JM, Cooper JD, Penninx BW, Bahn S. Variation in serum biomarkers with sex and female hormonal status: implications for clinical tests. *Sci Rep*. 2016;6(1):26947.

117. Gangestad SW, Haselton MG, Welling LLM, et al. How valid are assessments of conception probability in ovulatory cycle research? Evaluations, recommendations, and theoretical implications. *Evol Hum Behav*. 2016;37(2):85–96.

118. Harlow SD, Gass M, Hall JE, Lobo R, Maki P, Rebar RW, Sherman S, Sluss PM, de Villiers TJ; STRAW + 10 Collaborative Group. Executive summary of the Stages of Reproductive Aging Workshop + 10: addressing the unfinished agenda of staging reproductive aging. *J Clin Endocrinol Metab*. 2012;97(4):1159–1168.

119. Shifren JL, Schiff I. The aging ovary. *J Womens Health Gend Based Med*. 2000;9(Suppl 1):S3–S7.

120. Rich-Edwards JW, Fraser A, Lawlor DA, Catov JM. Pregnancy characteristics and women's future cardiovascular health: an underused opportunity to improve women's health? *Epidemiol Rev*. 2014;36(1):57–70.

121. Bellamy L, Casas J-P, Hingorani AD, Williams DJ. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. *BMJ*. 2007;335(7627):974.

122. Holvik K, Meyer HE, Haug E, Brunvand L. Prevalence and predictors of vitamin D deficiency in five immigrant groups living in Oslo, Norway: the Oslo Immigrant Health Study. *Eur J Clin Nutr*. 2005;59(1):57–63.

123. Calvo MS, Whiting SJ, Barton CN. Vitamin D intake: a global perspective of current status. *J Nutr*. 2005;135(2):310–316.

124. Hilger J, Friedel A, Herr R, Rausch T, Roos F, Wahl DA, Pierroz DD, Weber P, Hoffmann K. A systematic review of vitamin D status in populations worldwide. *Br J Nutr*. 2014;111(1):23–45.

125. van Dam RM, Snijder MB, Dekker JM, Stehouwer CD, Bouter LM, Heine RJ, Lips P. Potentially modifiable determinants of vitamin D status in an older population in the Netherlands: the Hoorn Study. *Am J Clin Nutr*. 2007;85(3):755–761.

126. Kautzky-Willer A, Harreiter J. Sex and gender differences in therapy of type 2 diabetes. *Diabetes Res Clin Pract*. 2017;131:230–241.

127. Wallach JD, Sullivan PG, Trepanowski JF, Steyerberg EW, Ioannidis JP. Sex based subgroup differences in randomized controlled trials: empirical evidence from Cochrane meta-analyses. *BMJ*. 2016;353:i5826.

128. Miller VM, Tannenbaum C, Regitz-Zagrosek V. Sex based subgroup differences in randomized controlled trials: empirical evidence from Cochrane meta-analyses; response to authors comment. *BMJ*. 2016;353:i5826.

129. Alesh M, Fritsch K, Huque M, et al. Statistical considerations on subgroup analysis in clinical trials. *Stat Biopharm Res*. 2015;7(4):286–303.

130. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.

131. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46(3):399–424.

132. Fewell Z, Davey Smith G, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol*. 2007;166(6):646–655.

133. Ding P, VanderWeele TJ. Sensitivity analysis without assumptions. *Epidemiology*. 2016;27(3):368–377.

134. Shi H, Seeley RJ, Clegg DJ. Sexual differences in the control of energy homeostasis. *Front Neuroendocrinol*. 2009;30(3):396–404.

135. Rosenfeld CS. Sex-dependent differences in voluntary physical activity. *J Neurosci Res*. 2017;95(1-2):279–290.

136. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure–mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods*. 2013;18(2):137–150.

137. Richiardi L, Bellocchio R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol*. 2013;42(5):1511–1519.

138. Buyse ME. Analysis of clinical trial outcomes: some comments on subgroup analyses. *Control Clin Trials*. 1989;10(4, Suppl):1875–1945.

139. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000;355(9209):1064–1069.

140. Frasure-Smith N, Lespérance F, Prince RH, Verrier P, Garber RA, Juneau M, Wolfson C, Bourassa MG. Randomised trial of home-based psychosocial nursing intervention for patients recovering from myocardial infarction. *Lancet*. 1997;350(9076):473–479.

Acknowledgments

Correspondence and Reprint Requests: Janet W. Rich-Edwards, PhD, Brigham and Women's Hospital, 1620 Tremont Street, Boston, Massachusetts 02120. E-mail: jr33@partners.org

Disclosure Summary: The authors have nothing to disclose.

Abbreviations

CHD, coronary heart disease; CVD, cardiovascular disease; ER, estrogen receptor; NIH, National Institutes of Health.