Beyond a catalogue of differences: A theoretical frame and good practice guidelines for researching sex/gender in human health

Kristen W. Springer [a,*], Jeanne Mager Stellman [b], Rebecca M. Jordan-Young [c]

[a] Department of Sociology and Institute for Health, Health Care Policy and Aging Research, Rutgers University, 26 Nichol Avenue, New Brunswick, NJ 08901, United States
[b] Department of Health Policy and Management, Mailman School of Public Health, Columbia University, United States
[c] Department of Women's, Gender, and Sexuality Studies, Barnard College, Columbia University, United States

ABSTRACT

Extensive medical, public health, and social science research have focused on cataloguing male–female differences in human health. Unfortunately, much of this research unscientifically and unquestionably attributes these differences to biological causes — as exemplified in the Institute of Medicine's conclusion that "every cell has a sex." In this manuscript we theorize the entanglement of sex and gender in human health research and articulate good practice guidelines for assessing the role of biological processes — along with social and biosocial processes — in the production of non-reproductive health differences between and among men and women. There are two basic tenets underlying this project. The first is that sex itself is not a biological mechanism and the second is that "sex" and "gender" are entangled, and analyses should proceed by assuming that measures of sex are not pristine, but include effects of gender. Building from these tenets — and using cardiovascular disease as a consistent example — we articulate a process that scientists and researchers can use to seriously and systematically assess the role of biology and social environment in the production of health among men and women. We hope that this intervention will be one further step toward understanding the complexity and nuance of health outcomes, and that this increased knowledge can be used to improve human health.

© 2011 Elsevier Ltd. All rights reserved.

Introduction

Social science and health research ubiquitously treat sex as a dichotomized independent variable. The United States Institute of Medicine (IOM) Committee on Understanding the Biology of Sex and Gender Differences drew an 'overarching conclusion' that "Sex … should be considered when designing and analyzing studies in all areas and at all levels of biomedical and health-related research. …" (xix) and that "every cell has a sex" ((Wizemann & Pardue, 2001), p.4; see also pp.28–44). Indeed, the IOM not only reified sex as a master independent variable, but also focused its own committee work on "sex-based differences, versus similarities, as …more likely to … lead to greater understanding of the significance of sex in human biology and health" (2). The IOM study did acknowledge that social and cultural factors could contribute to observed male-female differences in anatomy and physiology (see especially pp.18–19). Generally, though, the committee's analysis departed from its official openness to such interactive models, and

from its designation of "gender" as a continuum (p.17). Instead, as two dissenting committee members noted, the committee's overall approach suggests a "predominance of physiology, with a subsequent fine-tuning by environment" (p.18), and moreover, as we argue below, such an approach incorrectly implies that scientists can completely "control" for gender to identify the pure substrate of sex biology that underlies observed male-female differences.

The IOM, drawing especially on definitions advanced by the World Health Organization (WHO) and the style manual of the Journal of the American Medical Association (JAMA), defines sex as "The classification of living things, generally as male or female according to their reproductive organs and functions assigned by chromosomal complement" and gender as "A person's self-representation as male or female, or how that person is responded to by social institutions based on the individual's gender presentation. Gender is rooted in biology and shaped by environment and experience" ((Wizemann & Pardue, 2001) p.17). These definitions lend a superficial sense that sex and gender are distinct domains, even as they give causal and temporal priority to biology ("gender is rooted in biology" but sex is presumably pristine and emerges regardless of environment and experience).

* Corresponding author. Tel.: +1 848 932 7516.
E-mail address: kspringer@sociology.rutgers.edu (K.W. Springer).

Entanglement of sex/gender in human health

In this paper we build on prior research to provide an alternative to IOM-type mandates for health research where "sex" is treated as the main explanatory variable for proposed male–female health differences in human health (Bird & Rieker, 1999; Epstein, 2007; Fausto-Sterling, 2005). We aim to articulate a theoretical frame and good practice guidelines for exploring the role of sex-related biology in human health, especially as it may influence non-reproductive differences between males and females. We are explicitly engaging with sex difference research – so we take a frame of male–female difference in much of our discussion, although we also focus on the importance of within-sex variation. There are two basic tenets to our argument. First, sex is not a biological mechanism and its use as a proxy for other measures does little to further the understanding of health-related research questions. Second, in the vast majority of health research, "sex" and "gender" are entangled and analyses should proceed by assuming that measures of sex are not pristine, but include effects of gender. We are not arguing against any biologically based male–female differences. Rather we are arguing that the vast majority of male–female health differences are due to the effects of the irreducibly entangled phenomenon of "sex/gender," and therefore this entanglement should be theorized, modeled, and assumed until proven otherwise.

We borrow the concept of entanglement from multiple other fields including quantum physics and science/technology studies, where entanglement has been used to describe inextricably interwoven factors (Callon & Rabeharisoa, 2003; Papp et al., 2009). For example, according to quantum physics, no individual element can fully describe a quantum mechanical state without incorporation of other elements, and any attempt to measure a single factor inevitably perturbs the remaining system rendering the results inaccurate. Of course, physicists do measure physical properties, and do accept a certain level of inaccuracy and imprecision in their work by doing so – just as social scientists define and measure sex and gender in research on health outcomes. Nonetheless, we maintain that a true and full appreciation of sex/gender in human health cannot rely on the dichotomization of people into males and females without attention to how sex and gender may be – and almost always are – entangled.

The idea of sex/gender entanglement is not new and a growing body of research demonstrates the problematic task of distinguishing between sex and gender in practice (Fausto-Sterling, 2005; Kessler, 1998; Oudshoorn, 1994). Feminist epidemiologists and biologists, among other scientists, increasingly replace the discrete concepts of "sex" and "gender" with more complex formulations, such as Nancy Krieger's notions of "biologic expressions of gender" and "gendered expressions of biology" ((Krieger, 2003), p. 653). As Kaiser, Haller, Schmitz, and Nitsch (2009) recently observed, "sex is not a pure bodily and material fact, but is deeply interwoven with social and cultural constructions of gender" ((Kaiser et al., 2009), p.50). Fausto-Sterling (2005) has shown in the case of bone development and change, and Jordan-Young and Rumiati (forthcoming) have argued in the case of brain structure and function, that gendered life experiences have material effects on the body. These effects show up, in turn, as biologically based "sex differences." Indeed, the social structure of gender even creates biological differences for boys and girls starting in the womb (e.g. birth weight) – as evidenced by lower levels of prenatal care for female fetuses in countries with a preference for boys (Al-Akour, 2008).

While the causal link between "sex" and "gender" is often thought to flow automatically from biological to social difference, recent research has forcefully demonstrated that the influence often operates in the other direction. A strong example can be found in social neuroendocrinology where research has confirmed that mood states, social interactions, and status differentials affect neuroendocrine production and function (Booth, Granger, Mazur, & Kivlighan, 2006; Haneishi et al., 2007; Sapolsky, 1997; van Anders & Watson, 2006). One obvious conclusion is that sex-linked outcomes related to neuroendocrine function must be conceived as flowing, at least in part, from social variables.

The potential implications of recognizing "sex endocrinology" as a biosocial complex are profound. Consider the relationship of neuroendocrine factors to risks for cardiovascular disease (CVD), which we will continue to use as an example throughout because of its epidemiological importance as a primary cause of death and because of the historical importance of CVD research originally ignoring women and then reacting by automatically focusing on male-female differences in sex-linked biology (Epstein, 2007; Mosca et al., 2004).

A multidisciplinary team working on psychosocial factors in CVD has noted, for example, that "psychosocial stress reliably induces ovarian dysfunction, hypercortisolemia, and excessive adrenergic activation in pre-menopausal females, leading to accelerated atherosclerosis" (Rozanski, Blumenthal, & Kaplan, 1999: p.2192). These results of psychosocial stress constitute variables generally characterized as aspects of "sex biology." However, since psychosocial stress is itself the result of processes that are profoundly gendered (e.g., uneven family obligations, gender-specific forms of harassment or discrimination, and greater levels of poverty and wage insecurity), these "biological" factors are already inextricably bound up with the social world of gender (Rozanski et al., 1999). Despite some progress, we still know relatively little about specific social-biological pathways through which gendered arrangements become embodied as within- or between-sex differences. We therefore propose that the general dictum should be that sex and gender are entangled, rather than the opposite dictum as set forth by the IOM.

The difficulty in identifying the true directionality of the "causal" links leads us to follow Kaiser et al. (2009), and use the term *sex/gender* "to highlight this socio-biological intertwinement whenever possible" ((Kaiser et al., 2009), p.50). This view of *sex/gender* fits very well with the contemporary focus on intersectional approaches to health – whereby aspects of social status (e.g., gender, race, socioeconomic status, and sexuality) are understood to affect health outcomes in complex, multiplicative ways that can never properly be captured by attempts to parcel out the individual contributions of single social domains (Hankivsky & Christoffersen, 2008; Jackson & Williams, 2006).

We approach sex/gender in this intersectional light and conceptualize sex/gender as a domain of complex phenomena that are simultaneously biological and social, rather than a domain in which the social and biological "overlap." We use the term *gender* (read: gender-not-sex) to draw attention to specifically social and structural factors, such as patterns in the distribution of family responsibilities, formal and informal sanctions for gender non-conforming behavior, and so on (For recent, in-depth analyses that support the claim that these patterns cannot be attributed to "original" differences between males and females, see (Fine, 2010; Jordan-Young, 2010; Jordan-Young & Rumiati, forthcoming)). We avoid the use of the term *sex* (read: "sex-not-gender") as a stand-alone indicator of biology, or as a broad reference to males and females (as in "sex differences") because it is rarely specific enough to guide particular investigations, and because it is too easily confused with the more accurate composite phenomenon of *sex/gender* that we use here.

Further, the direction of influence should not be *assumed* to flow from an idealized version of "pure" biology outward to the behavioral and social world. Consider, for example, glucose metabolism,

a risk factor for CVD. Male–female differences in glucose metabolism have been reported (Basu et al., 2006; Peterson et al., 2007), and cellular level mechanisms related to glucose–insulin interactions have been hypothesized to explain this difference (Basu et al., 2006). Yet a recent experimental study of response to oral glucose "challenge" found that the male-female difference, comparable in size in this study to other reports, could be entirely explained by difference in skeletal muscle mass between men and women (Rattarasarn, Leelawattana, & Soonthornpun, 2010). Given the extensive data on social influences on the male-female difference in muscle mass (for instance, via differential nutrition, aesthetic norms, occupation, and use of outdoor spaces for exercise), the metabolic differences seen at the cellular level between males and females could be understood as bearing the traces of their gendered lives (Bentley et al., 1999; Courtenay, 2000; Richardson & Mitchell, 2010).

In a case such as this, sex/gender is a preferable notation to "sex" alone, although the exact qualities that sex/gender represent may not be fully specified. Unmeasured aspects of gender will always be present and usually we will be unaware of the specific ways in which gendering of activities, as through nutrition and psychology, affects cellular level processes.

Good practice guidelines for examining male-female human health differences

We build on the theoretical frame of sex/gender entanglement to present an encapsulation of good practice guidelines for researching male-female health differences (Fig. 1). This encapsulation is critical because these good research practices are often ignored in sex/gender and health research, which hampers the ability of research to provide fruitful insights into health inequalities and disparities. These guidelines are organized into three sequential sections (research plan development; a priori comparison plan; and interpretation and presentation of results), but as with all research, looping back through the sections is inevitable and desirable.

Research plan development

Because we posit that sex and gender are entangled, the first step is to assess whether or not biological male-female difference research is both warranted and possible. If yes, the next step is to a priori develop and articulate specific, testable hypotheses, with precise well-defined measures observable in both men and women, and to articulate plausible biological/biosocial mechanisms to explain the biological contribution to male-female health differences. (Biological mechanism is not meant to suggest that biology can be separated from the social environment, but rather to refer to specific mechanisms that can be described in molecular or biochemical terms.)

Fully define outcome and independent measures a priori

An accurate understanding of the biological contributions of male–female differences requires precisely identifying and

RESEARCH PLAN DEVELOPMENT

1. Fully define outcome and independent measures a priori
2. Identify plausible biological mechanism(s) explaining male-female difference
3. Directly test the biological mechanism(s) – if at all possible
4. Develop, operationalize, and test alternative hypotheses that include social and biosocial pathways.
5. Incorporate and test the biological and social/biosocial mechanisms together.

A PRIORI COMPARISON PLAN

1. Include appropriate control groups for ALL hypotheses (biological, social, biosocial) and accurately test for differences
2. Include sufficient measures for confounders and covariates
3. Attend to issues of statistical significance and sample size
4. Conduct sensitivity analyses

INTERPRETATION AND PRESENTATION OF RESULTS

1. Present within group and between group differences
2. Explain effect size: substantive vs. significant difference
3. Report post-hoc tests and adjust significance levels appropriately
4. Describe social/biosocial attenuation *and* the purely social/biosocial findings
5. Describe *similarities* and differences for biological causes
6. Thoroughly report limitations of the study and provide enough information for replication

Fig. 1. Good practice guidelines for researching sex/gender in human health.

operationalizing the outcome of interest. Continuing with the example of CVD, we see that the rubric "cardiovascular disease" itself may be too broad for elucidating observed male-female differences in rates. The International Classification of Diseases (ICD- 9) uses the codes 350 to 459 to classify circulatory system diseases. "Ischemic heart disease" uses more than 20 different codes. The 16th edition of Harrison's Manual of Medicine includes 13 sub-sections on CVD (Kasper, Braunwald, Fauci, Hauser, Longo, & Jameson, 2005). True biological male–female differences in CVD would be expected to vary in important ways depending on the specific aspect of CVD under investigation.

CVD studies may also include a wide range of clinical signs and symptoms like chest or back pain (Berg, Bjorck, Dudas, Lappas, & Rosengren, 2009; Hendrix, Mayhan, Lackland, & Egan, 2005), serum progesterone level (Nilsson, Fransson, & Brismar, 2009), or a psychosocial risk factor like depression (Rozanski et al., 1999). One complication in defining measures is that physicians may differently diagnose and treat the same presentation of symptoms depending on whether the patient is male or female (Enriquez, Pratap, Zbilut, Calvin, & Volgman, 2008).

In addition, studies of the "same" outcome can produce dramatically different results depending on the definitions and measures used to operationalize the outcome. For example, research on male-female differences in the CVD-related outcome of metabolic syndrome shows great variation in the size, shape, and even the presence of male-female differences depending on the definition of the outcome used in the study. As articulated by Regitz-Zagrosek et al. (2006): "Classification according to WHO criteria generally led to a 50% higher estimation of prevalence compared with the EGIR (European Group for the Study of Insulin Resistance) criteria in men. This can mostly be explained by the different cut-off value for central obesity used in the WHO definition. For women, the difference was smaller." (p.137).

Sex/gender differences in a specifically defined outcome may be partially explained by other factors (Naqvi, Naqvi, & Merz, 2005; Rozanski et al., 1999; Suarez, 2006). Duprez et al. (2009) found significant male-female differences in arterial elasticity – an indicator of progression to cardiovascular morbidity and mortality – that were completely accounted for by height differences between men and women. Importantly, the broad variability in height within both sexes means that using sex as a proxy for arterial elasticity results in, at best, unnecessary loss of information and, at worst, gross misclassification errors leading to erroneous results.

Identify plausible biological mechanism(s) explaining male–female difference

Studies should be designed to explicate plausible biological mechanisms and "sex," as such, is not a mechanism. Putative male-female difference must be tied to plausible biological mechanisms and operationalized into measurable variables. A thorough, critical investigation of the literature will assist in understanding biological mechanisms and how they can be operationalized before proceeding with hypothesis development. How the mechanism(s) would be different in men and women – on average – must be articulated. For example, male-female differences in mortality have been linked to male-female variations in metabolic syndrome and diabetes, a CVD risk factor. As explained by Regitz-Zagrosek, Lehmkuhl, and Weickert (2006): "Lipid accumulation patterns differ between women and men. Pre-menopausal women more frequently develop peripheral obesity with subcutaneous fat accumulation, whereas men and post-menopausal women are more prone to central or android obesity. In particular, android obesity is associated with increased cardiovascular mortality and the development of type 2 diabetes" (p.136). Here the proposed biological mechanism for male-female difference is not for *all*

females and males – but for pre-menopausal women compared to men (and post-menopausal women), underscoring the complexity (and generally unacknowledged variability) of the categories "male" and "female." (It is also possible that differential deposition of fat is socially created, as discussed below.)

Directly test the biological mechanism(s) – if at all possible

Identifying a male–female difference in an outcome does not tell us anything about the possible mechanisms (biological or social) that may produce this difference. Therefore, using sex as a variable is an inadequate proxy for testing a proposed biological mechanism. For example, if the proposed biological pathway for CVD mortality is central or android obesity, then it is more useful to directly measure men's and women's obesity, rather than simply including a "sex" variable and then hypothesizing that the effect is due to a difference (on average) in the location of men's and women's fat stores. Even multilevel dichotomous measures, like male, pre-menopausal female, and post-menopausal female, do not measure a mechanism (Regitz-Zagrosek et al., 2006). Sex encompasses many other qualities besides the mechanism in question and will not exhaustively describe the mechanism.

Clearly, theory often will outstrip the ability to measure, and not having access to sufficiently large cohorts and/or sufficient funds to take biological measurements may make it unfeasible to rigorously follow our proposed good practice guidelines (Giannattasio et al., 2007; Regitz-Zagrosek et al., 2006). In such cases, authors must be conservative in interpreting their findings and should avoid overbroad generalization of their findings. Studies in which sex has been used as a proxy for more specific measures can only be considered suggestive, and where possible, should be avoided.

Develop, operationalize, and test alternative hypotheses that include social and biosocial pathways

Most models in health research are incomplete and many relevant factors are omitted for practical reasons of measurement and feasibility. One way to address this problem in sex/gender health research is to develop alternative hypotheses that always include social and biosocial pathways. It is important to consider prior studies to best conceptualize and model these alternative hypotheses. For example, if prior research indicates that the size or direction of male-female difference varies over time, place, or sample characteristics (e.g., by ethnicity, age, occupation, height, menopause, or country of residence) this provides strong evidence of social and/or biosocial effects and offers insights into key variables for alternative hypotheses (Macintyre, Hunt, & Sweeting, 1996). Modeling with relevant social/behavior measures may show that the presumed biological underpinnings of the observed male-female differences in health are spurious.

Indeed, based on a plethora of evidence social causes should be *assumed* and a preponderance of evidence required before arguing for a primarily biological, non-social mechanism. Consider the relationship between fat intake, adiposity, and CVD (Mente, de Koning, Shannon, & Anand, 2009; Regitz-Zagrosek et al., 2006). At least two concurrent secular trends are of importance: 1) relatively high levels of unhealthy dietary fats; and 2) higher male CVD rates compared to pre-menopausal women (although the gap is narrowing) (Towfighi, Zheng, & Ovbiagele, 2009). Possible biological pathways used to explain the male-female difference focus on how men and women process fat and cholesterol in ways that differentially affect CVD (Clifton & Nestel, 1992; Regitz-Zagrosek et al., 2006). Other research suggests that increases in boys' testosterone levels activate hepatic lipase activity, in turn, leading to a decrease in HDL cholesterol levels for boys – but not girls. Clifton and Nestel's (1992) "crossover, randomized, double-blind dietary trial" tested for male-female differences in plasma lipid responses

associated with dietary fat and cholesterol, and provided some evidence for differential processing for males and females. But here, too, alternative hypotheses to explain the observed male-female difference using social pathways are possible. Normative conceptions of men as strong and hearty, compared to women as slender and delicate, shape the types and quantities of food eaten. Specifically, in some cultural contexts, men are more likely to eat red meat and women are more likely to eat fruits and vegetables (Prattala et al., 2007). Sometimes the richest (and most unhealthy) meat is reserved for males when there are variations in quality and/ or insufficient quantity for the whole family (Holm & Mohl, 2000). Gendered social processes could be an alternative plausible mechanism linking the higher male rates of CVD with increased intake of saturated fat, or an alternative hypothesis could be based on the confluence of the social and biological mechanisms: higher male intake of fattier meats and dairy products (social processes) and differential male processing is more harmful for men (biological process).

Incorporate and test biological and social/biosocial mechanisms together

Inclusion of social and biosocial mechanisms *alongside* the originally proposed biological mechanism will improve design because mediation and/or spurious associations are only observable when multiple mechanisms are considered. Alternative mechanisms will help establish whether the biological mechanism is: 1) spurious and can be accounted for by social/structural factors, 2) actually a biosocial mechanism which varies in type, scope, level etc. by social/environmental causes, and/or 3) a substantively meaningful size relative to other social/structural paths. For example, biological contributions from comorbid conditions like diabetes and hypertension cannot completely account for women's disproportionately greater adverse outcomes following an acute myocardial infarction (MI) (Naqvi et al., 2005). However, women with CVD have higher rates of depression, which, in turn, is both socially mediated and a key risk factor for poor cardiovascular outcomes (Naqvi et al., 2005; Whooley et al., 2008). The observed differential recovery from acute MI and the relatively unexplored role of depression in the causal pathway illustrates the importance of incorporating social/biosocial mechanisms within studies exploring the biological contribution to male-female differences in CVD outcomes.

A priori comparison plan

Comparison groups and statistical tests of effects should be developed based on previous empirical evidence and theory prior to proceeding with the research. Carefully planning for comparisons can help avoid the temptation to make atheoretical comparisons simply because the design includes a "sex variable" and a range of interesting outcomes. Such unplanned comparisons almost always yield false positives and can be particularly damaging given the bias towards interpreting male-female differences as 'probably true' and 'probably biological' (Jordan-Young, 2010; Kaiser et al., 2009).

Include appropriate control groups for all hypotheses (biological, social, biosocial) and accurately test for differences

Comparison groups should be strictly comparable. This is particularly important for conducting *statistical tests* of similarity/ difference — without this strict comparability the substantive meaning of tests of statistical difference is unclear. Conclusions based on *ad hoc* comparison of statistical tests conducted *within* groups of men and women are not supportable. Phrases like "more statistically significant" in one group versus another have little practical meaning since such differences could easily be the result of differences in sample sizes, bias in assembling the population, or inadequacy of the measures used. A significant finding in one group and a non-significant finding in the other group is likewise not evidence of a male-female difference. To compare groups it is necessary to test for the significance of a statistical interaction and/or to formally test whether the effect for one group is significantly different from the effect for another group (Aiken & West, 1991).

Include sufficient measures for confounders and covariates

Statistical tests of incomplete models have limited meaning. Models must include confounders and covariates, as well as test for plausible mediation and interactions to avoid spurious findings. For example, Kanaya, Grady, and Barrett-Connor (2002) conducted a meta-analysis to assess the claim that diabetes is a stronger independent risk factor for coronary heart disease mortality in men than in women. The authors found no male-female differences after adjusting for other classic CVD risk factors (Kanaya et al., 2002). In other words, a previously established risk factor (diabetes) for male-female differences in CVD mortality was found to be incorrect when appropriate controls were included in the model.

Attend to issues of statistical significance and sample size

Sample size issues plague many studies but research often proceeds with relatively little attention to how sample size may affect findings and interpretation. The effect of sample size can operate in two directions: very large samples can produce artifactual significant relationships at the $p < 0.05$ level (McCloskey, 1985). Very large samples, as in gene-wide array studies, GWAS, may require a much stricter level of significance (i.e. $p < 10^{-5}$ or smaller) (Caporaso et al., 2009). One common and reasonable strategy is to use Bonferroni corrections based on the baseline p-value determined by the sample size (Rice, 1989). The Bonferroni correction simply involves dividing the baseline significance level (i.e. $p < 0.05$) by the number of comparisons, yielding the new p-value for statistical significance.

Small sample sizes may compromise one's ability to detect an effect. The smaller the sample and the weaker the effect size, the more difficult it will be to achieve sufficient power. Small samples may also make testing interactions difficult if not impossible and biosocial interactions are fundamental to understanding sex/ gender differences. For example, a study with 80% power to detect a main effect will only have 29% power to detect an interaction effect of the same magnitude (Brookes et al., 2004).

Conduct sensitivity analyses

To assess the robustness of the findings and to understand possible variations in effects, an array of sensitivity tests, ideally identified *a priori,* should be conducted. The specific tests depend on the research question(s) but could include: a) testing different operationalization of variables – e.g. a continuous vs. categorical measure of BMI; b) different modeling techniques to assess violations of modeling assumptions; c) alternative treatment of outliers, including transforming skewed variables and removing extreme cases. Sensitivity tests should be fully reported when presenting results – especially, if the outcome of sensitivity tests is *not* consistent with the original results.

Interpretation and presentation of results

Present within group and between group differences

Exploring variation *within* men and *within* women can help illuminate the practical/real significance of a male-female difference. If the observed magnitude of the variation *within* men or *within* women is comparable to the male-female difference, then

this difference may not be any more substantial or relevant than the natural variation within men or women. As with our earlier example, the difference in average heights between men and women in the U.S. is 5 inches (5ft, 8 inches for men and 5ft, 3 inches for women) (McDowell, Fryar, Ogden, & Flegal, 2008), but the height difference between the 5th and 95th percent of height for men and women in the U.S. is 10 inches and 9 inches respectively (McDowell et al., 2008). Such intra-sex height variations can lead to erroneous conclusions about male-female differential CVD risks. Data show that the example of reduction in arterial elasticity as a biological indicator of progression toward cardiovascular morbidity and mortality discussed earlier is dependent on height, regardless of sex/gender (Duprez et al., 2009).

Explain effect size(s): substantive vs. significant difference

Statistical significance is not substantive significance (Miller, 2005). A recent meta-analysis on sex differences in the association between birth weight and total cholesterol (a CVD risk factor) found evidence for a greater effect of low birth on cholesterol among men vs. women, attributed to "different biological processes for females and males" ((Lawlor et al., 2006) p.19.) But the effect size, 0.04 mmol/l in total cholesterol per kg birth weight, was called "unlikely to have important public health effects." (Lawlor et al. 2006, pg 23). The substantiveness, as well as statistical significance of findings, can illuminate the real-life importance (or lack thereof) of any effect.

Report post-hoc tests and adjust significance level appropriately

Interesting findings may emerge after the hypothesized analyses, and *post hoc* statistical testing of these findings is good practice; however, such testing increases the likelihood of false positives. Reporting that tests were *post-hoc*, as well as adjusting significance level for the number of *post-hoc* tests assessed provides a clear account of the procedures and helps assure that significant findings are not statistical artifacts.

Describe social/biosocial attenuation and the purely social/biosocial findings

The results of the biological contribution to male–female differences in health can best be understood when situated in the context of the biosocial and social mechanism results. Describing results from all mechanisms – alternative social and biosocial mechanisms as well as full models with all mechanisms – can help illuminate the relative importance of different social, biological, and biosocial effects. This presentation of the findings allows researchers and readers to compare and theorize the potential for complementary, competing, and interactive social, biosocial, and biological mechanisms.

Describe similarities and differences in biological causes

Similarity may be larger than difference, and also very interesting and useful. Similarity should therefore be considered with the same level of sophistication and attention given to explaining differences. For example, extensive research has focused on steroid hormones as one key cause of male-female difference in CVD risk (Ding, Song, Malik, & Liu, 2006; Perez-Lopez, Larrad-Mur, Kallen, Chedraui, & Taylor, 2010). However, there are also important and interesting similarities in the effects of hormones and CVD risk for men and women. The most striking example of these male-female similarities is the narrowing of CVD risk between men and women post menopause (Mendelsohn & Karas, 2005). Research has identified several hormonal similarities that drive these findings including similar increases in CVD risk for men and post-menopausal women associated with endogenous ratios of estradiol to progesterone (He et al., 2007).

Thoroughly report limitations of the study and provide enough information for replication

Full disclosure of limitations in the design, method, or analyses of the research can help readers understand the scope and significance of the findings – as well as spur new research endeavors to help account for these limitations. Further, replication is the heart of science. Depending on space constraints and journal guidelines, it may be possible to report all information necessary to reproduce the analyses in published manuscripts. Alternatively, or in addition, manuscripts could include a link to a website that houses all information needed to reproduce the findings, including data and codes.

Discussion

We have attempted to present an alternative approach to automatic investigation of "sex differences" in human health studies and we have taken issue with assuming that *sex* is an individual-level variable that should be routinely assessed. Indeed, we propose the IOM mandate for male–female difference research be stood on its head: research should proceed with models based on either entirely social causes or interactive biosocial causes unless there is sound and extensive evidence for a primarily biological mechanism. We further propose that assessments of sex are almost without exception made for individuals whose development, including physical, physiological, and psychological, has been profoundly affected, in ways that we at present only dimly understand, by the operations of gender. Even when so seemingly biological and non-social an entity as a "cell" is assigned a sex, the biography of the individual that the cell came from goes along for the ride, but this is rarely acknowledged, let alone examined, in scientific analyses.

One potentially paradoxical result of our proposal to avoid the broad use of "sex" as a variable is that it could facilitate a more precise investigation of which aspects of biology contribute to male-female differences in health, and how they do so. Messing and Mager Stellman (2006), among others, have urged epidemiologists to avoid the use of the term *sex* as a broad proxy for biology as if the term conveyed a mechanism for generating male–female differences (Messing & Mager Stellman, 2006).

Here we used CVD to illustrate the good practice guidelines, but these guidelines are applicable to other outcomes, as supported by meta-analyses of male-female genetic differences. Patsopolous et al. (2007) found that the vast majority of highly prominent claims of sex-related differences in GWAS were insufficiently documented or spurious. Patsopoulos and colleagues also noted that only a small fraction of supposedly *a priori* considerations for examining male-female differences in genetic associations did, in fact, draw upon "any kind of corroboration history" (Patsopoulos, Tatsioni & Ionnidis 2007, 888–889). While Patsopoulos and colleagues do not make this point, we would emphasize that the habit of viewing male and female as the fundamental division between people makes it all too easy to hold *a priori* expectations of male-female difference even when there is no existing empirical evidence on which to base this expectation.

A final note on the long history and politics of using male-female differences in biology and health to support a wide range of political and social restrictions on women is in order (Barker-Benfield, 1976; Ehrenreich & English, 1978; Hubbard, 1990). In the 1970s, the women's health movement noted a pervasive lack of comprehensive and unbiased knowledge about women's bodies, and decried the tendency for medicine to proceed as if the "normal" or "ideal" patient is a white, middle-class, 160 pound man (Boston Women's Health Book Collective, 1973; Epstein, 2007; Tuana, 2006). We recognize the IOM Report as part of a contemporary

response to the ensuing call for greater attention to women's bodies in medical research and practice, and believe that a dedication to rectifying longstanding male biases in medicine motivates many who currently pursue sex-difference analyses. We applaud that goal, and indeed share it. But for all the reasons we have documented in this essay, we are convinced that the present approach, especially the overarching commitment to identifying and prioritizing difference over similarity, the prioritization of biology, the mandate to address sex as a grouping variable in federal funded research and clinical trials, regardless of biological evidence, and the mistaken belief that biology can be operationally separated from the social environment will not lead to the desired aim. An ever-growing catalogue of differences is not likely to achieve significant health gains for either men or women, and will only continue to distract us from the kind of focused research into specific mechanisms that we need, by offering the rather empty answer of "sex" to all questions we pose about the cause of male–female disparities.

Acknowledgments

References

Aiken, L., & West, S. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage Publications, Inc.

Al-Akour, N. (2008). Knowing the fetal gender and its relationship to seeking prenatal care: results from Jordan. *Maternal and Child Health Journal, 12*(6), 787–792.

van Anders, S. M., & Watson, N. V. (2006). Social neuroendrocrinology – effects on social contexts and behaviors on sex steroids in humans. *Human Nature, 17*(2), 212–237.

Barker-Benfield, G. J. (1976). *The horrors of the half-known life: Male attitudes toward women and sexuality in nineteenth-century America*. New York: Harper and Row.

Basu, R., Dalla Man, C., Campioni, M., Basu, A., Klee, G., Toffolo, G., et al. (2006). Effects of age and sex on postprandial glucose metabolism: differences in glucose turnover, insulin secretion, insulin action, and hepatic insulin extraction. *Diabetes, 55*(7), 2001–2014.

Bentley, G. R., Aunger, R., Harrigan, A. M., Jenike, M., Bailey, R. C., & Ellison, P. T. (1999). Women's strategies to alleviate nutritional stress in a rural African society. *Social Science & Medicine, 48*(2), 149–162.

Berg, J., Bjorck, L., Dudas, K., Lappas, G., & Rosengren, A. (2009). Symptoms of a first acute myocardial infarction in women and men. *Gender Medicine, 6*(3), 454–462.

Bird, C. E., & Rieker, P. P. (1999). Gender matters: an integrated model for understanding men's and women's health. *Social Science & Medicine, 48*(6), 745–755.

Booth, A., Granger, D. A., Mazur, A., & Kivlighan, K. T. (2006). Testosterone and social behavior. *Social Forces, 85*, 167–191.

Boston Women's Health Book Collective. (1973). In *Our bodies, ourselves* New York: Simon and Schuster.

Brookes, S. T., Whitely, E., Egger, M., Smith, G. D., Mulheran, P. A., & Peters, T. J. (2004). Subgroup analyses in randomized trials: risks of subgroup-specific analyses;: power and sample size for the interaction test. *Journal of Clinical Epidemiology, 57*(3), 229–236.

Callon, M., & Rabeharisoa, V. (2003). Research "in the wild" and the shaping of new social identities. *Technology in Society, 25*(2), 193–204.

Caporaso, N., Gu, F., Chatterjee, N., Sheng-Chih, J., Yu, K., Yeager, M., et al. (2009). Genome-wide and candidate gene association study of cigarette smoking behaviors. *PLoS ONE, 4*(2), e4653.

Clifton, P. M., & Nestel, P. J. (1992). Influence of gender, body mass index, and age on response of plasma lipids to dietary fat plus cholesterol. *Arteriosclerosis and Thrombosis, 12*(8), 955–962.

Courtenay, W. H. (2000). Constructions of masculinity and their influence on men's well-being: a theory of gender and health. *Social Science & Medicine, 50*(10), 1385–1401.

Ding, E. L., Song, Y., Malik, V. S., & Liu, S. (2006). Sex differences of endogenous sex hormones and risk of type 2 diabetes: a systematic review and meta-analysis. *The Journal of the American Medical Association, 295*(11), 1288–1299.

Duprez, D. A., Jacobs, D. R., Jr., Lutsey, P. L., Herrington, D., Prime, D., Ouyang, P., et al. (2009). Race/ethnic and sex differences in large and small artery elasticity–results of the multi-ethnic study of atherosclerosis (MESA). *Ethnicity & Disease, 19*(3), 243–250.

Ehrenreich, B., & English, D. (1978). *For her own good: 150 years of the experts' advice to women*. Garden City: Anchor Press.

Enriquez, J. R., Pratap, P., Zbilut, J. P., Calvin, J. E., & Volgman, A. S. (2008). Women tolerate drug therapy for coronary artery disease as well as men do, but are treated less frequently with aspirin, [beta]-blockers, or statins. *Gender Medicine, 5*(1), 53–61.

Epstein, S. (2007). *Inclusion: The politics of difference in medical research*. Chicago: University of Chicago Press.

Fausto-Sterling, A. (2005). The bare bones of sex: part 1-sex and gender. *Signs, 30*(2), 1491–1527.

Fine, C. (2010). *Delusions of gender: How our minds, society, and neurosexism create difference*. New York: W. W. Norton.

Giannattasio, C., Capra, A., Facchetti, R., Viscardi, L., Bianchi, F., Failla, M., et al. (2007). Relationship between arterial distensibility and coronary atherosclerosis in angina patients. *Journal of Hypertension, 25*(3), 593–598.

Haneishi, K., Fry, A. C., Moore, C. A., Schilling, B. K., Li, Y., & Fry, M. D. (2007). Cortisol and stress responses during a game and practice in female collegiate soccer players. *Journal of Strength and Conditioning Research, 21*(2), 583–588.

Hankivsky, O., & Christoffersen, A. (2008). Intersectionality and the determinants of health: a Canadian perspective. *Critical Public Health, 18*(3), 271–283.

He, H., Yang, F., Liu, X., Zeng, X., Hu, Q., Zhu, Q., et al. (2007). Sex hormone ratio changes in men and postmenopausal women with coronary artery disease. *Menopause, 14*(3), 385–390.

Hendrix, K. H., Mayhan, S., Lackland, D. T., & Egan, B. M. (2005). Prevalence, treatment, and control of chest pain syndromes and associated risk factors in hypertensive patients. *American Journal of Hypertension, 18*(8), 1026–1032.

Holm, L., & Mohl, M. (2000). The role of meat in everyday food culture: an analysis of an interview study in Copenhagen. *Appetite, 34*(3), 277–283.

Hubbard, R. (1990). *The politics of women's biology*. New Brunswick, NJ: Rutgers University Press.

Jackson, P. B., & Williams, D. R. (2006). The intersection of race, gender, and SES: health paradoxes. In A. Schultz, & L. Mullings (Eds.), *Gender, race, class, & health: Intersectional approaches*. San Francisco, CA: Jossey-Bass.

Jordan-Young, R. (2010). *Brain storm: The flaws in the science of sex differences*. Cambridge, MA: Harvard University Press.

Jordan-Young, R.M. & Rumiati, R.I. (forthcoming). Hardwired for Sexism? Approaches to Sex/Gender in Neuroscience. *Neuroethics*.

Kaiser, A., Haller, S., Schmitz, S., & Nitsch, C. (2009). On sex/gender related similarities and differences in fMRI language research. *Brain Research Review, 61*(2), 49–59.

Kanaya, A. M., Grady, D., & Barrett-Connor, E. (2002). Explaining the sex difference in coronary heart disease mortality among patients with type 2 diabetes mellitus: a meta-analysis. *Archives of Internal Medicine, 162*(15), 1737–1745.

Kasper, D. L., Braunwald, E., Fauci, A. S., Hauser, S. L., Longo, D. L., & Jameson, J. L. (Eds.). (2005). *Harrison's manual of medicine*. New York: McGraw-Hill.

Kessler, S. J. (1998). *Lessons from the intersexed*. New Brunswick, New Jersey: Rutgers University Press.

Krieger, N. (2003). Genders, sexes, and health: what are the connections–and why does it matter? *International Journal of Epidemiology, 32*(4), 652–657.

Lawlor, D. A., Owen, C. G., Davies, A. A., Whincup, P. H., Ebrahim, S., Cook, D. G., et al. (2006). Sex differences in the association between birth weight and total cholesterol. A meta-analysis. *Annals of Epidemiology, 16*(1), 19–25.

Macintyre, S., Hunt, K., & Sweeting, H. (1996). Gender differences in health: are things really as simple as they seem? *Social Science & Medicine, 42*(4), 617–624.

McCloskey, D. N. (1985). The loss function has been mislaid: the rhetoric of significance tests. *American Economic Review, 75*(2).

McDowell, M. A., Fryar, C. D., Ogden, C. L., & Flegal, K. M. (2008). *Anthropometric reference data for children and adults: United States, 2003–2006*. National health statistics reports. Hyattsville, MD: National Center for Health Statistics.

Mendelsohn, M. E., & Karas, R. H. (2005). Molecular and cellular basis of cardiovascular gender differences. *Science, 308*(5728), 1583–1587.

Mente, A., de Koning, L., Shannon, H., & Anand, S. (2009). A systematic review of the evidence supporting a causal link between dietary factors and coronary heart disease. *Archives of Internal Medicine, 169*(7), 659–669.

Messing, K., & Mager Stellman, J. (2006). Sex, gender and women's occupational health: the importance of considering mechanism. *Environmental Research, 101*(2), 149–162.

Miller, J. (2005). *The Chicago guide to writing about multivariate analysis (Chicago guides to writing, editing, and publishing)*. University Of Chicago Press.

Mosca, L., Appel, L. J., Benjamin, E. J., Berra, K., Chandra-Strobos, N., Fabunmi, R. P., et al. (2004). Evidence-Based guidelines for cardiovascular disease Prevention in women. *Circulation, 109*(5), 672–693.

Naqvi, T. Z., Naqvi, S. S., & Merz, C. N. (2005). Gender differences in the link between depression and cardiovascular disease. *Psychosomatic Medicine, 67*(Suppl 1), S15–S18.

Nilsson, S. E., Fransson, E., & Brismar, K. (2009). Relationship between serum progesterone concentrations and cardiovascular disease, diabetes, and mortality in elderly Swedish men and women: an 8-year prospective study. *Gender Medicine, 6*(3), 433–443.

Oudshoorn, N. (1994). *Beyond the natural body: An archeology of sex hormones*. London and New York: Routledge.

Papp, S. B., Choi, K. S., Deng, H., Lougovski, P., van Enk, S. J., & Kimble, H. J. (2009). Characterization of multipartite entanglement for one photon shared among four optical modes. *Science, 324*(5928), 764–768.

Patsopoulos, N. A., Tatsioni, A., & Ioannidis, J. P. A. (2007). Claims of sex differences: an empirical assessment in genetic associations. *JAMA, 298*(8), 880–893.

Perez-Lopez, F. R., Larrad-Mur, L., Kallen, A., Chedraui, P., & Taylor, H. S. (2010). Gender differences in cardiovascular disease: hormonal and biochemical influences. *Reproductive Sciences, 17*(6), 511–531.

Peterson, L. R., Soto, P. F., Herrero, P., Schechtman, K. B., Dence, C., & Gropler, R. J. (2007). Sex differences in myocardial oxygen and glucose metabolism. *Journal of Nuclear Cardiology, 14*(4), 573–581.

Prattala, R., Paalanen, L., Grinberga, D., Helasoja, V., Kasmel, A., & Petkeviciene, J. (2007). Gender differences in the consumption of meat, fruit and vegetables are similar in Finland and the Baltic countries. *European Journal of Public Health, 17*(5), 520–525.

Rattarasarn, C., Leelawattana, R., & Soonthornpun, S. (2010). Contribution of skeletal muscle mass on sex differences in 2-hour plasma glucose levels after oral glucose load in Thai subjects with normal glucose tolerance. *Metabolism, 59*(2), 172–176.

Regitz-Zagrosek, V., Lehmkuhl, E., & Weickert, M. O. (2006). Gender differences in the metabolic syndrome and their role for cardiovascular disease. *Clinical Research in Cardiology, 95*(3), 136–147.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution, 43*, 223–225.

Richardson, E. A., & Mitchell, R. (2010). Gender differences in relationships between urban green space and health in the United Kingdom. *Social Science & Medicine, 71*(3), 568–575.

Rozanski, A., Blumenthal, J. A., & Kaplan, J. (1999). Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. *Circulation, 99*(16), 2192–2217.

Sapolsky, R. M. (1997). *The trouble with testosterone and other essays on the biology of the human predicament.* New York: Touchstone.

Suarez, E. C. (2006). Sex differences in the relation of depressive symptoms, hostility, and anger expression to indices of glucose metabolism in nondiabetic adults. *Health Psychology, 25*(4), 484–492.

Towfighi, A., Zheng, L., & Ovbiagele, B. (2009). Sex-specific trends in midlife coronary heart disease risk and prevalence. *Archives of Internal Medicine, 169*(19), 1762–1766.

Tuana, N. (2006). The speculum of ignorance: the women's health movement and epistemologies of ignorance. *Hypatia, 21*(3), 1–19.

Whooley, M., de Jonge, P., Vittinghoff, E., Otte, C., Moos, R., Carney, R., et al. (2008). Depressive symptoms, health behaviors, and risk of cardiovascular events in patients with coronary heart disease. *The Journal of the American Medical Association, 300*(20), 2379–2388.

Wizemann, T. M., & Pardue, M.-L. (Eds.). (2001). *Exploring the biological contributions to human health: Does sex matter?* Washington D.C.: National Academy of Sciences.