PREPRINT

A Framework to Guide Decision-Making about Sex and Gender Selection for Sport and Exercise Investigations

Eimear Dolan[1,2], Gabriel P. Esteves[1,2], Paul A. Swinton[3]

[1]Applied Physiology and Nutrition Research Group – School of Physical Education and Sport and Faculdade de Medicina FMUSP, Universidade de Sao Paulo.

[2]Center of Lifestyle Medicine, Faculdade de Medicina FMUSP, Universidade de São Paulo.

[3]School of Health, Robert Gordon University, Aberdeen, UK.

E-mail address for correspondence: eimeardol@gmail.com or eimeardolan@usp.br

Please cite as Dolan et al. (2024) Preprint version: A framework to guide decision-making about sex and gender selection for sport and exercise investigations. *SportRXiv.* DOI: 10.51224/SRXIV.450.

Feedback:

If you would like to provide feedback on this article, please email eimeardol@gmail.com.

ABSTRACT:

Historically, males have been preferentially selected as participants in sport and exercise research, resulting in a large sex and gender bias in almost every aspect of our evidence base. Awareness of the potential implications of this bias is prevalent, and there appears to be a willingness to solve the problem. It can, however, be challenging to make an informed decision on whether to recruit males, females, or mixed cohorts for individual studies. Decisions are frequently made with uncertainty of how biological sex- or gender- specific factors, such as the menstrual cycle, differences in baseline characteristics and response, or societal and cultural perceptions and norms, may influence research findings. Here we propose a framework to guide the decision to recruit males, females or mixed cohorts to sport and exercise studies. The framework comprises a series of conditional branching questions regarding the aims of the research study and the potential influence of sex or gender on outcomes of interest. The questions include: 1) whether the research question centers on a sex or gender specific topic; 2) whether sex or gender specific factors are likely to introduce noise to the outcomes of interest; 3) whether baseline or response differences between sexes or genders are likely to influence the outcomes of interest; and 4) what to do when insufficient data are available to inform answers to questions 2 and 3. We present and discuss examples that may influence the response to each of these branching questions. In many situations, definitive answers may not exist, and the intention of the framework is not to dictate or prescribe the participant group that individual researchers should work with. Instead, the framework is presented to engage with sample recruitment in a structured and systematic way, thereby facilitating informed and evidence-based decision-making, with the ultimate goal of contributing toward a sport and exercise evidence base that is less affected by sex and gender bias.

Introduction:

Historically, males have been preferentially selected as participants in sport and exercise research, resulting in a large sex and gender bias in almost every aspect of our evidence base [1–5]. This is an issue, because important physiological and sociological differences exist between males and females, which may influence exercise participation, performance and responses across interventions [6–9]. As such, data generated primarily on males are not necessarily transferable to females. This bias is not unique to sport and exercise investigations, with females reported to be under-represented in almost every aspect of health and medical research, purportedly leading to common misdiagnoses and inadequate treatment prescription [10–14]. Awareness of potential implications of sex and gender biases in research has increased in recent times, and willingness to solve this imbalance is high. Indeed, many ethics review boards and funding bodies now require that the sex and gender of intended research participants be justified. The difficulty facing researchers, however, is that evidence related to the potential influence of sex and gender on outcomes relevant to sport and exercise studies is limited, rendering it difficult to make informed decisions, or to present evidence-based justifications about when to recruit males, females, or mixed cohorts. To aid this decision-making process, we propose a framework comprising a series of conditional branching questions to help researchers approach this important topic in a more structured and systematic way.

Sex and Gender:

Sex is usually categorized as female or male, and is determined by the biological make-up of the individual and their primary and secondary sex characteristics. The term gender refers to socially constructed roles and perceptions of women, men, girls and boys. The term cis-gender refers to someone whose gender identity matches the biological sex that they were assigned at birth, as opposed to a trans-gender person who does not identify with their assigned biological sex.

Both sex and gender can influence sport and exercise related outcomes and at times these effects may be "entangled" and difficult to differentiate [15]. For example, women tend to experience ACL injuries at far greater rates than men [16]. Historically, this difference was assumed to be due to sex-based anatomical, physiological and hormonal differences [17], with prevention, treatment and rehabilitation programs planned accordingly. More recently, however, the potential role of gendered factors within pre-sport, training, and competition environments have been implicated in ACL injury etiology [15]. Consideration of both gendered (*e.g.,* access to, and attitudes toward, strength and conditioning programs) and sex-specific factors (*e.g.,* hormonal and anatomical differences between males and females), along with their interactions, may be key to understanding, and thus reducing, the gender gap in ACL injury.

The framework presented herein can be used to inform decision making regarding how both biological sex and gender can influence sport and exercise related outcomes. In this commentary, we primarily focus on biological sex related examples, as this is our primary area of research interest. For this reason, we default to the terms males and females throughout, unless specifically referring to gendered factors. We encourage readers to use the examples described herein as a start-point, but to adapt the questions to their own studies and to consider how both sex and gender related concepts may influence study design and research outcomes.

Question 1: *Is this a sex or gender-specific topic?*

Some research questions are sex or gender specific, and as such, participant selection is clear, as the research must be conducted either with the single relevant sex/gender, or with both sexes/genders to provide the contrast. For example, topics such as the influence of the menstrual cycle on exercise performance, or exercising while pregnant are specific to females and as such, must be conducted with female participants. Other topics may be specific to males, such as the exercise hypogonadal male condition [18]. In the context of gender, research questions can also be specific to a single gender or require a contrast between genders. Examples include women's access and attitudes to strength and conditioning programs, or the availability of school physical activity programs for boys and girls.

Many research topics are not exclusively sex or gender specific, but these constructs may still influence outcomes in terms of baseline and response distributions, potentially impacting the location (the mean or median), spread (variation) and shape (symmetric, skewed or multimodal) of the distribution. For example, a researcher may be interested in investigating the influence of a sports supplement on a performance outcome, but be concerned that hormonal fluctuations throughout the menstrual cycle may confound results. Or in a study investigating the influence of resistance training on strength the researcher must consider the possibility that sex dimorphism in muscle mass may influence baseline variability if using a mixed cohort, or that males and females may respond differently, which in both cases could compromise statistical power to detect differences between experimental groups. In these majority cases, researchers should consider the answers to Questions 2 and 3 of this framework.

Question 2: *Are sex-or gender specific factors likely to introduce noise in the outcomes of interest?*

Perhaps the most prevalent example of a sex-specific factor believed to introduce noise in outcomes of interest is the menstrual cycle. Concern that hormonal fluctuations throughout the menstrual cycle may increase variability in female participant groups, potentially obscuring identification of smaller effects, is a common justification for conducting male-only research. Certainly, this concern has merit, given that the female reproductive hormones may influence a range of processes involved in exercise performance or response to training [19]. It is important to critically assess, however, the received wisdom that women are inherently more variable than men. Indeed, within the field of evolutionary biology, the "greater male variability" hypothesis, holds that men are, in fact, the more variable sex [20–22]. Using data from the NHANES database, we recently compared the extent of variation between males and females in fifty morphological and physiological traits. Our analyses indicated that sex differences in variability was trait dependent, with some displaying greater male variability, others displaying greater female variability, while others were equivalent. Furthermore, analysis of a subset of females who reported having a natural menstrual cycle did not influence the extent of greater female variability, further refuting the notion that women are inherently more variable than men across all traits [23]. As such, generalized assumptions about which sex is likely to vary more should not be made, and instead data on traits or outcomes of interest should be considered.

Considering this lack of evidence to support a predominance of either greater male or female variability, researchers must consider whether intra-individual variability for their specific outcome of influence is influenced by the menstrual cycle, or any other sex- or gender-related issue. For example, if estrogen, progesterone, or any other hormone that substantially fluctuates throughout the cycle is an outcome of interest, it stands to reason that the time of testing must be tightly controlled. In contrast, current evidence indicates that acute strength, or responses to resistance interventions, are unlikely to be meaningfully impacted by the menstrual cycle [24]. As such, excluding females, or controlling for menstrual cycle phase in a study investigating the influence of resistance training on acute strength may not be warranted.

In addition to considering whether sex-specific factors such as the menstrual cycle may influence outcomes of interest, it is also important to consider the likely magnitude and resultant effects on inferences made. For example, meta-analytic data provided uncertain evidence that exercise performance may exhibit very small magnitude decrement during the early follicular phase in naturally menstruating females (Hedges' g: ~0 to -0.15), with no evidence for differences between any other phases [25]. How a finding like this should be interpreted will depend on the outcomes of interest and the perspectives of the investigators. Uncertain evidence indicating a very small decrement in exercise performance within the early follicular phase may be considered important in a study evaluating an intervention with a small predicted effect. In contrast, if the proposed intervention is predicted to induce a moderate to large effect, as may occur in many cohort designs or comparisons with non-active conditions, then any potential variability in exercise outcomes induced by the menstrual cycle may be considered negligible. It is also relevant to consider that many outcomes in sport and exercise research exhibit heteroscedasticity where greater variability is present in those with greater absolute values [26] Considering males are usually stronger, more muscular and have a larger aerobic capacity than females, the common observation of positive heteroscedasticity suggests that they may also exhibit greater variability in these outcomes. As such, the potential for a very small increase in variability due to the menstrual cycle may potentially be offset by a more consistent performance in females in outcomes for which they tend to produce lower absolute values.

In the case that a sex- or gender-related factor such as the menstrual cycle is likely to influence outcomes of interest, researchers must then decide how best to account for this within the study design and methods. For example, not all females have a natural menstrual cycle, with 50 to 60% of surveyed British and Dane female athletes reporting using hormonal contraceptives [27,28]. A wide range of contraceptive types exist, but most function via administering small doses of exogenous hormones, which act to suppress the natural fluctuations that occur in the endogenous reproductive hormones throughout the cycle. As such, females who use hormonal contraceptives may have a more stable hormonal profile than their naturally menstruating counterparts, thus reducing the potential for hormonal fluctuations to introduce additional variability to outcomes of interest. Recruiting females who use oral contraceptive may therefore be a useful strategy to reduce menstrual cycle associated noise. Alternatively, data collection could be standardized to take place at a similar time within the participants cycle, thus allowing for a comparable hormone profile between test sessions [29]. Standardization of testing to a specific time in the month may dramatically increase the length of time required for testing and also potentially reduce participant availability, both of which may also increase the extent of variability in test outcomes. As such, researchers must weigh up the pros and cons of different approaches and gauge the feasibility and efficacy of test standardization on outcomes of interest. Variability in test outcomes for both males and females can also be reduced by practices such as performing additional measurements and performing analyses on simple linear combinations of the data obtained. More detailed information on statistical approaches to reduce the effects of measurement error within sport and exercise investigations, along with their underlying assumptions, are described in Swinton et al. [30].

In the case that the menstrual cycle or some other sex- or gender-specific factor is likely to meaningfully interject noise that cannot be accounted for using the aforementioned procedures, then it may be appropriate to recruit the opposing sex or gender. If this is not the case, however, then theoretically male, female or mixed cohorts may be appropriate, and the final decision can be determined based on responses to Questions 3 and 4 of this framework.

Question 3: *Are baseline or response differences between sexes or genders likely to influence the outcomes of interest?*

Prolonged exposure to sex hormones, particularly during the pubertal period, result in markedly different phenotypes between males and females, which can influence exercise performance [8]. For example, males are on average taller, heavier and leaner than females, whereas females are generally more flexible and exhibit greater range of motion. Sex differences in factors such as substrate metabolism, fiber type composition and cardiorespiratory capacity may also influence exercise performance or response to training. These sex differences mean that mixed cohorts may have greater baseline or change variability than an equivalent single-sex sample, which has important implications for sample size determination and study power. Sport and exercise studies often investigate relatively small changes, and it can be challenging to recruit sufficiently large samples to ensure adequately powered and precise studies [31,32]. More variable populations reduce study power, which is an important argument against the use of mixed cohort studies in some situations. Additionally, from an estimation perspective, if males and females respond differently to an intervention, then estimating a single population mean value from a mixed cohort may not be relevant.

In cases where increased variability due to baseline sex differences or response may reduce power or introduce confounding, then researchers should assess whether this can be accounted for using appropriate statistical techniques. For example, analysis of covariance approaches that include sex as a covariate and potentially baseline value may overcome potential confounding and inflated standard errors [33]. Care must be taken, however, as increasing the number of covariates within a model can lead to overfitting, which occurs when the model fits the noise in the data rather than the underlying relationships. In the case of smaller sample sizes and especially where effects are potentially small, mixed cohorts may be inadvisable.

It is important to note the difference between controlling for sex or gender in a research study, versus exploring the potential for these factors to explain variability in an outcome. As outlined in question 1 of this framework, the research topic may be a sex- or gender- specific one where the primary interest is to compare populations and identify whether sex or gender explains variability in an outcome. In contrast, sex or gender may be one of a range of potential explanatory variables that are being investigated in so-called heterogeneous treatment effects [34]. Regardless of whether sex and gender is the primary focus or not, the most appropriate method to investigate the explanatory effect is to recruit a mixed cohort and test for statistical interactions between the intervention and sex/gender [35]. This may be easier said than done, however, given that investigation of potential sex or gender differences through interactions likely requires far greater sample sizes than considering main effects only. For example, when compared with the sample size required to detect the average intervention effect of a particular size, a sample size of roughly four times as much is required to detect a sex or gender difference of the same magnitude in the intervention effect (assuming an even split between groups). In the much more likely case that the interaction effect is smaller than the main effect, the required sample size will be many factors more. Indeed, even when the interaction effect is half the main effect, a sample size approximately sixteen times as large is required [36]. Clearly, investigating potential sex differences in response to an intervention is a far greater undertaking than investigating an assumed consistent main effect across sexes, and in many cases, resource constraints may preclude these types of investigations.

In summary, recruiting mixed cohorts may be appropriate where there are no sex differences in baseline or response, or when these can be controlled for statistically. Where it is feasible to recruit larger samples and a mixed cohort facilitates this, statistical power may be increased, and the risk of overfitting reduced. Even if it is not feasible to recruit sufficiently large samples to detect potential sex differences and the main analysis focuses on main effects, it would be useful to provide individual data segregated by sex or gender as a supplementary file. This would increase the amount of data available that could be included in future meta-analyses and even individual participant data meta-analyses to better estimate the influence of sex or gender on outcomes of interest.

Question 4: *What to do if insufficient information is available to respond to Questions 2 and 3?*

Questions 2 and 3 of the framework assume that sufficient information is available to determine the influence of sex or gender specific factors on noise or systematic effects on the outcomes of interest. For many cases and research questions, however, this may not be the case. This potential lack of information is made more likely by the aforementioned sex and gender bias in sport and exercise research. Where there is uncertainty in how to respond to the questions of the framework, it could be tempting to default to recruitment of single sex or gender samples as the “safer” option. Given the bias that already exists, it is likely that this default would frequently result in recruitment of male participants further exacerbating the existing bias. As such, in the absence of an evidence-based justification to exclude female participants, we suggest that researchers err on the side of inclusion, rather than exclusion. This approach may, at times, increase variability in individual studies, however, as data emerges and is made easily accessible and identifiable, the evidence base will grow and provide a foundation on which better informed decisions about sex and gender selection can subsequently be made.

Summary and Conclusion:

Within this framework, we propose four conditional branching questions, consideration of which may facilitate researchers to make informed decisions about whether to recruit males, females or mixed cohorts to sport and exercise studies. These are: 1) Is this a sex- or gender- specific question? 2) Will sex- or gender- specific factors (e.g., the menstrual cycle) introduce noise to the outcome of interest? 3) Are baseline or response differences between sexes/genders likely to influence results? And 4) What to do if insufficient information is available to answer questions 2 and 3. The intention of this framework is not to dictate, or prescribe, participant selection for individual studies, but instead to provide a structured guide to inform the decision-making process.

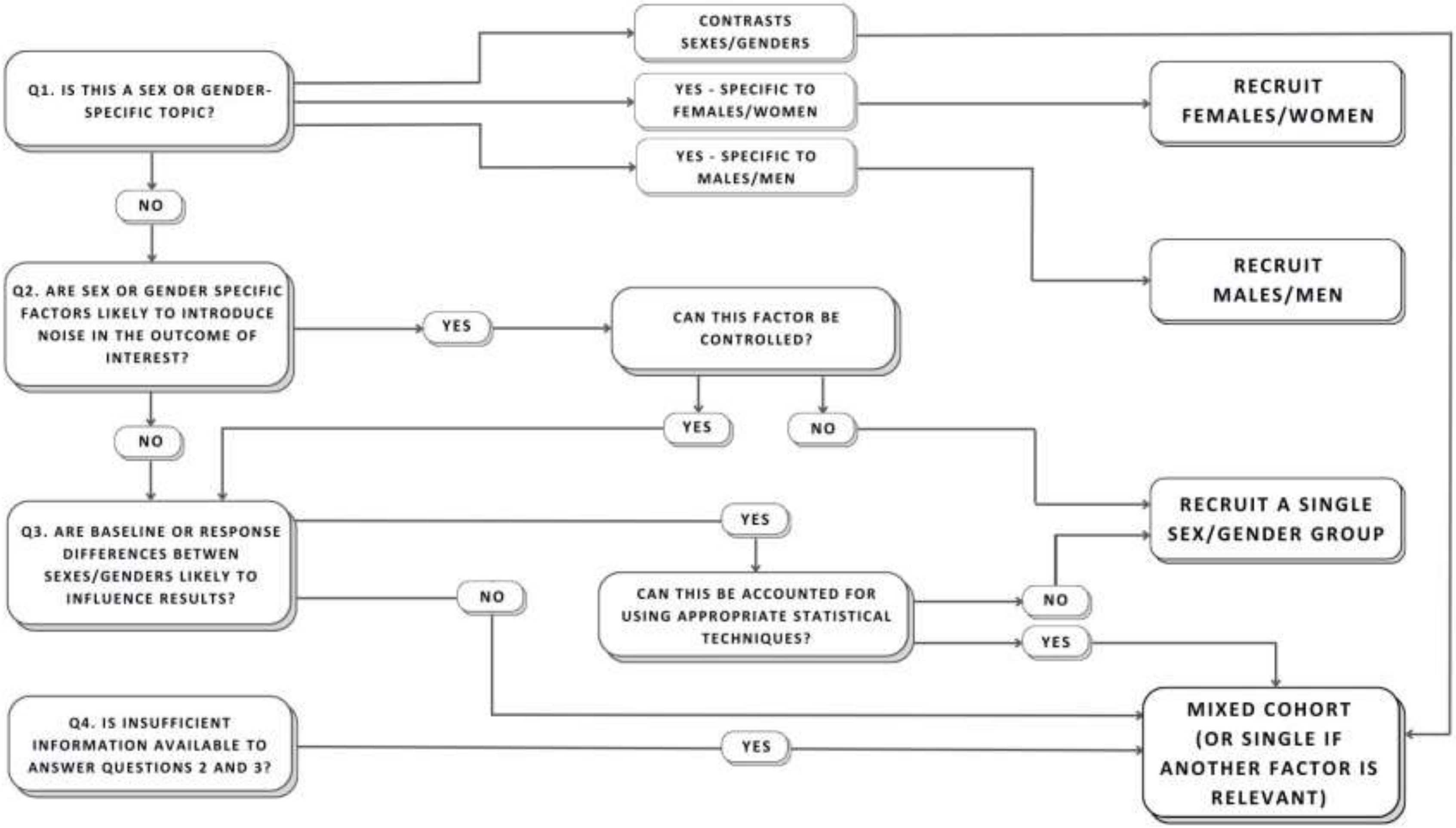

Figure 1: *Framework to guide decision-making about whether to recruit male, female or mixed-cohort groups to sport and exercise studies.*

References:

1. Cowley E, Olenick A, McNulty K, Ross E. “Invisible Sportswomen”: The sex data gap in sport and exercise science research. Women Sport Phys Act. 2021;29:146–51. DOI: 10.1123/wspaj.2021-0028.

2. Hutchins K, Borg D, Bach A, Bon J, Minett G, Stewart I. Female (under) representation in exercise thermoregulation research. Sport Med Open. 2021;22:43. DOI: 10.1186/s40798-021-00334-6.

3. Kuikman M, Smith E, McKay A, Ackerman K, Harris R, Elliott-Sale K, et al. Fueling the Female Athlete: Auditing Her Representation in Studies of Acute Carbohydrate Intake for Exercise. Med Sci Sport Exerc. 2023;55:569–80. DOI: 10.1249/MSS.0000000000003056.

4. Kuikman M, McKay A, Smith E, Ackerman K, Harris R, Elliott-Sale K, et al. Female athlete representation and dietary control methods amons studies assessing chronic carbohydrate approaces to support training. Int J Sport Nutr Exerc Metab. 2023;33:198–208. DOI: 10.1123/ijsnem.2022-0214

5. Smith E, Burke L. Have we considered women in current sports nutrition guidelines? Nutr Today. 2024;59:168–76. DOI: 10.1097/NT.0000000000000692.

6. Ansdell P, Thomas K, Hicks K, Hunter S, Howatson G, Goodall S. Physiological sex differences affect the integrative response to exercise: Acute and chronic implications. Exp Physiol. 2020;105:2007–21. DOI: 10.1113/EP088548.

7. Petrie K, Burbank K, Sizer P, James C, Zumwalt M. Considerations of sex differences in musculoskeletal anatomy between males and females. In: Robert-McComb J, Zumwalt M, Fernandez-del-Vell M, editors. Act Female. Springer Ch, Cham; 2023. DOI: 10.1007/978-3-031-15485-0_1.

8. Hunter S, Senefeld J. Sex differences in human performance. J Physiol. 2024. DOI: 10.1113/JP284198.

9. Thorpe H, Bekker S, Fullager S, Mkumbuzi N, Nimphius S, Pape M, et al. Advancing feminist innovation in sport studies: A transdisciplinary dialogue on gender, health and wellbeing. Front Sport Act Living. 2023;4:1060851. DOI: 10.3389/fspor.2022.1060851.

10. Karp N, Reavey N. Sex bias in preclinical research and an exploration of how to change the status quo. Br J Pharmacol. 2019;176:4107–18. DOI: 10.1111/bph.14539.

11. Yoon D, Mansukhani N, Stubbs V, Helenowski I, Woodruff T, Kinne M. Sex bias exists in basic science and translational surgical research. Surgery. 2014;156:508–16. DOI: 10.1016/j.surg.2014.07.001.

12. Mazure V, Jones D. Twenty years and still counting: including women as participants and studying sex and gender in biomedical research. BMC Womens Heal. 2015;15. DOI: 10.1186/s12905-015-0251-9.

13. Holdcroft A. Gender bias in research: How does it affect evidence based medicine. J R Soc Med. 2007;100:2–3. DOI: 10.1177/014107680710000102.

14. Simon V. Wanted: Women in clinical trials. Science (80-). 2005;308:1517. DOI: 10.1126/science.1115616.

15. Parsons J, Coen S, Bekker S. Anterior cruciate ligament injury: towards a gendered environmental approach. Br J Sports Med. 2021;55:984–90. DOI: 10.1136/bjsports-2020-103173.

16. Agel J, Rockwood T, Klossner D. Collegiate ACL injury rates across 15 sports: National Collegiate Athletic Association Injury Surveillance System Data Update (2004 - 2005 through 2012 - 2013). Clin J Sport Med. 2016;26:518–23. DOI: 10.1097/JSM.0000000000000290.

17. Griffin L, Albohm M, Bahr R, Beynnon B, DeMaio M, Dick R, et al. Understanding and preventing noncontact anterior cruciate ligament injuries: A review of the Hunt Valley II meeting, January 2005. Am J Sports Med. 2006;34:1512–32. DOI: 10.1177/0363546506286866.

18. Hackney A. Hypogonadism in exercising males: Dysfunction or adaptive-regulatory adjustment? Front Endocrinol (Lausanne). 2020;11:11. DOI: 10.3389/fendo.2020.00011.

19. de Jonge X. Effects of the menstrual cycle on exercise performance. Sport Med. 2003;33:833–51. DOI: 10.2165/00007256-200333110-00004.

20. Lehre A, Lehre K, Laake P, Danbolt N. Greater intrasex phenotype variability in males than in females is a fundamental aspect of the gender differences in humans. Dev Psychobiol. 2009;51:198–206. DOI: 10.1002/dev.20358.

21. Halpern D, Benbow C, Geary D, Gur R, Shibley Hyde J, Gernsbacher M. The science of sex differences in science and mathematics. Psychol Sci Public Interes. 2007;8:1–51. DOI: 10.1111/j.1529-1006.2007.00032.x.

22. Thoni C, Volk S. Converging evidence for greater male variability in time, risk, and social preferences. Proc Natl Acad Sci. 2021;118:e2026112118. DOI: 10.1073/pnas.2026112118.

23. Halsey L, Esteves G, Dolan E. Variability in variability: Does variation in morphological and physiological traits differ between men and women? R Soc Open Sci. 2023;10:230713. DOI: 10.1098/rsos.230713.

24. Colenso-Semple L, D'Souza A, Elliott-Sale K, Phillips S. Current evidence shows no influence of women's menstrual cycle phase on acute strength performance or adaptations to resistance exercise training. Front Sport Act Living. 2023;23:1054542. DOI: 10.3389/fspor.2023.1054542.

25. McNulty K, Elliott-Sale K, Dolan E, Swinton P, Ansdell P, Goodall S, et al. The effects of menstrual cycle phase on exercise performance in eumenorrheic women: A systematic review and meta-analysis. Sport Med. 2020;50:1813–27. DOI: 10.1007/s40279-020-01319-3.

26. Nevill A, Atkinson G. Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. Br J Sports Med. 1997;31:314–8. DOI: 10.1136/bjsm.31.4.314.

27. Oxfeldt M, Dalgaard L, Jorgensen A, Hansen M. Hormonal contraceptive use, menstrual dysfunctions, and self-reported side effects in elite athletes in Denmark. Int J Sports Physiol Perform. 2020;15:1377–84. DOI: 10.1123/ijspp.2019-0636.

28. Martin D, Sale C, Cooper S, Elliott-Sale K. Period prevalence and perceived side effects of hormonal contraceptive use and the menstrual cycle in elite athletes. Int J Sports Physiol Perform. 2018;13:926–32. DOI: 10.1123/ijspp.2017-0330.

29. Eliott-Sale K, Minahan C, Janse de Jong X, Ackerman K, Sipila S, COnstantini N, et al. Methodological considerations for studies in sport and exercise science with women as participants: A working guide for standards or practice for research on women. Sport Med.

2021;51:843–61. DOI: 10.1007/s40279-021-01435-8.

30. Swinton P, Stephens Hemingway B, Gallagher I, Dolan E. Preprint: Statistical methods to reduce the effects of measurement error in sport and exercise: A guide for practitioners and applied researchers. SportRXiv. 2023. DOI: 10.5122/SRXIV.247.

31. Skorski S, Hecksteden A. Coping with the “small sample - small relevant effects” dilemma in elite sport research. Int J Sports Physiol Perform. 2021;16:1559–60. DOI: 10.1123/ijspp.2021-0467.

32. Abt G, Boreham C, Davison G, Jackson R, Nevill A, Wallace E, et al. Power, precision, and sample size estimation in sport and exercise science research. J Sports Sci. 2020;38:1933–5. DOI: 10.1080/02640414.2020.1776002.

33. Wang B, Ogburn E, Rosenblum M. Analysis of covariance in randomized trials: More precision and valid confidence intervals, without model assumptions. Biometrics. 2019;75:1391–400. DOI: 10.1111/biom.13062.

34. Kent D, Steyererg E, van Klaveren D. Personalized evidence based medicine: Predictive approaches to heterogeneous treatment effects. Br Med J. 2018;363:k4245. DOI: 10.1136/bmj.k4245.

35. Shapiro J, Klein S, Morgan R. Stop “controlling” for sex and gender in global health research. BMJ Glob Heal. 2021;6:e005714. DOI: 10.1136/bmjgh-2021-005714.

36. Varadhan R, Seeger J. Estimation and reporting of heterogeneity of treatment effects. In: Velentgas P, Dreyer N, Nourjah P, Smith S, Torchia M, editors. Dev a Protoc Obs Comp Eff Res A user’s Guid. Rockville, MD: Agency for Healthcare Research and Quality (US); 2013. Available from: https://www.ncbi.nlm.nih.gov/books/NBK126188/.