ENDOCRINOLOGY AND WOMEN’S SPORTS: THE DIAGNOSIS MATTERS

RICHARD J. AUCHUS

INTRODUCTION

The participation of women in sports, and in particular elite sports, is a relatively recent phenomenon starting at the turn of the twentieth century. Shortly after the establishment of women’s competitions in the Olympic Games, a few participants who were later found to be biologically male competed in these events.1 These instances led to efforts to police the entry criteria to the women’s events, but sport authorities soon recognized that a single dichotomous criterion for defining a “female” athlete was difficult to identify.

In parallel with the rise of women’s sports, doping with performance-enhancing substances, particularly androgens, began to plague elite sports starting in the 1950s. In an effort to preserve the safety and fairness of sport, extensive programs for anti-doping and testing have become a fixture of professional and now also recreational competitions.

The East Germans quickly recognized that androgen doping was particularly effective in women, in whom androgen action could be raised from the low female status to higher male realms and beyond.2 Today, every elite athlete is subjected to rigorous testing for doping using sophisticated mass spectrometry systems that can cost roughly $1,000,000 with certified authentic standards and vigorous chain-of-custody documentation—testing far more technologically advanced than most non-athletes will ever receive in their medical care.

Using these three points of reference as a framework, this Article will logically approach a discussion of sex in sport from the endocrinologist’s perspective. Part II will explain critical differences between male and female biology. Part III clarifies confusion resulting from the sex versus gender distinction. Part IV concludes that sex, and testosterone in particular, is fundamental to the discrepancies in athletic performance between men and women. Part V proposes solutions to defining the women’s category for the purposes of competitive sport.

Copyright © 2018 by Richard J. Auchus.
This article is also available at http://lcp.law.duke.edu/.

* Departments of Pharmacology and Internal Medicine, Division of Metabolism, Endocrinology and Diabetes, University of Michigan, Ann Arbor, MI 48109.

1. See Joanna Harper, Athletic Gender, 80 LAW & CONTEMP. PROBS., no. 4, 2017, at 140–42 (pointing to early debates about who should compete in the women’s category based on the participation of intersex and transgender athletes).

2. See Doriane Lambelet Coleman, Sex in Sport, 80 LAW & CONTEMP. PROBS., no. 4, 2017, at 72–74 (depicting testosterone reference ranges for genetic males and females).
II
DIFFERENCES IN MALE AND FEMALE PHYSIOLOGY

A. Human Androgen Biology

1. Sexual Differentiation
Early studies, like Alfred Jost’s pioneering work in rabbits, sought to define the mechanisms of external genitalia formation, but subsequent studies illuminated the totality of androgen biology. Jost demonstrated that a secretion from the testis was carried through the circulation as an endocrine hormone to cause fusion of the labia and formation of the penis. Jost reproduced this action with exogenous testosterone (T), confirming that T was the major circulating hormone required to cause male sexual differentiation of the external genitals. What Jost did not know is that an enzyme in the prostate and genital skin, 5α-reductase type 2 (5R2), was required to metabolize T to dihydrotestosterone (DHT), and it was DHT that elicited this action to form the prostate, scrotum, and penis.

Jean Wilson, Jim Griffin, Mike McPhaul, David Russell, and Stefan Andersson in Dallas then went on to define the genetics and molecular biology of 5R2 and the androgen receptor (AR, also known as NR3C4), the latter of which mediates most if not all actions of T, DHT, and synthetic androgens. With colleagues from around the world, they identified genetically male patients with mutations in either the \textit{SRD5A2} gene encoding 5α-reductase type 2 or the \textit{AR} gene, and these seminal studies form the basis of our knowledge about androgen biology.

2. Disorders of Androgen Biology: Androgen Insensitivity and 5R2 Deficiency
AIS and 5R2 deficiency (5R2D) afford very different phenotypes:
AIS patients appear phenotypically female at birth and feminize during puberty. They develop breasts but lack facial and body hair and—critical for our discussion—do not develop male-pattern upper-body muscular development.

4. \textit{See generally} Alfred Jost et al., \textit{Studies on Sex Differentiation in Mammals}, 29 RECENT PROGRESS HORMONE RES. 1 (1973); Coleman, supra note 2, at 73 (discussing the primary (reproductive) and secondary (phenotypic) characteristics caused by testosterone).
8. \textit{See Jean D. Wilson et al., Steroid 5α-Reductase 2 Deficiency, 14 ENDOCRINE REV. 577 (1993).}
during puberty despite a rise of circulating T that reaches the normal male range.\textsuperscript{9} In other words, the casual observer would easily identify these patients as female sex, despite the presence of a 46,XY karyotype and testes capable of normal male T production.

The absence of DHT generation from normal circulating T concentrations in patients with 5R2D renders the external genitalia as phenotypically female at birth and typically results in female sex of rearing. In contrast to AIS, however, these patients respond normally to T production at puberty, with profound physiologic changes including voice deepening, facial and body hair growth, male-pattern upper-body muscle development, and male-pattern effects on the brain.

Furthermore, the presence of the 5α-reductase type 1 isoenzyme in liver and other tissues allows for significant virilization of the external genitalia at puberty. Growth of the penis and descent of the testes into the (often bifid) scrotum occurs, and some individuals with this condition make sperm and have fathered children with assisted reproduction methods. A corollary is that the tissues are exposed to male T concentrations during various windows in fetal and neonatal life, which can induce genetic or epigenetic programming changes on tissues like brain, muscle, bone, and skin. In endocrinology nosology, 5R2D individuals are defined as undervirilized males,\textsuperscript{10} because they share all properties of normal male physiology with the exception of the external genitalia and Wolffian duct development. In fact, roughly half of these individuals who are not identified and treated prior to puberty change their gender to male, a profound testament to the potent actions of T as a hormone in multiple body tissues.

The only conditions with similar virilization and high frequency of gender reversal during a virilizing puberty as 5R2D are 17β-hydroxysteroid dehydrogenase type 3 (17HSD3) deficiency,\textsuperscript{11} discussed below, and partial AIS. The enzyme 17HSD3 converts androstenedione to T in the testicular Leydig cells. The enzyme is only expressed in the Leydig cells of the testis and is the only human enzyme that efficiently converts androstenedione to T. Patients with partial AIS can have a spectrum of phenotypes from normal but infertile male to phenotypic female with some features of androgenization, and partial AIS patients are generally the most difficult to assign into a dichotomous sex category.


\textsuperscript{11} See Wayne M. Geissler et al., \textit{Male Pseudohermaphroditism Caused by Mutations of Testicular 17-Hydroxysteroid Dehydrogenase 3}, 7 NATURE GENETICS 34 (1994) (“The phenotype of 17HSD3-deficient males is similar to that of 5R2D in that both have male Wolffian-duct derived internal genitalia and external genitalia that are predominantly female in character, and in both conditions, masculinization occurs at puberty.”).
3. Disorders of Androgen Biology: 17β-Hydroxysteroid Dehydrogenase Type 3 Deficiency

Because 17HSD3 is only found in the testicular Leydig cells, the ovaries of normal women and nearly all women with ovarian-derived androgen excess states are not capable of producing sufficient T to yield normal male circulating concentrations. Circulating concentrations of T in normal women and men show no overlap: females typically have concentrations of less than 60 ng/dL (~2 nmol/L), while males have concentrations of roughly 300–1200 ng/dL (~10–40 nmol/L). In other words, T concentrations in normal males are five to twenty times higher than those in normal women, and there is no overlap in the normal ranges. Even women with severe polycystic ovary syndrome, which is the most common condition that endocrinologists will agree are “hyperandrogenemic women,” rarely show T concentrations >120 ng/dL (4 nmol/L). Because sex hormone-binding globulin, the major binding protein for T, is typically higher in women than in men, free and bioavailable T concentrations in normal men and women are even more disparate than total T concentrations, for which there is no overlap plus a large gap between normal ranges. For these reasons, the International Association of Athletics Federations (IAAF) established a policy in 2012 to use a serum T concentration <10 nmol/L to define a “female athlete.” This cutoff, based on substantial data (see Section II.A.4), was rather generous, as this concentration is on the low end of T concentrations in average males and far exceeds the T levels found in even hyperandrogenemic women. However, this definition of female athlete is still based on sound endocrinology principles.

4. Androgens in Elite Female Athletes

Given the performance-enhancing properties of T, one might wonder if elite female athletes possess higher circulating T than normal women. This question was addressed at the 2011 IAAF championships in Daegu, and in fact >75% of these elite female athletes had serum T in or close to the normal female range and well below the normal male range. A subsequent study found that women in the highest tertile of T values performed better than those in the lowest tertile

---

15. Coleman, supra note 2, at 72–74.
17. Stéphane Bermon et al., Serum Androgen Levels in Elite Female Athletes, 99 J. Clinical Endocrinology & Metabolism 4328 (2014).
for a few sports like middle-distance running. In both studies, however, a disturbing number of outliers on the high end were observed, which cannot be explained by normal female physiology. Some of these women are doping with T, and others bear the endocrinology disorders that have generated the controversy.

B. Androgens and Athletic Performance

The evidence that T is a performance-enhancing substance is irrefutable. In a landmark study, for example, Bhasin et al. demonstrate that with or without exercise, use of the androgen T-enanthate at 600 mg/week profoundly increases muscle strength and size over a placebo in young healthy male athletes. The dose used in this initial study might be viewed as somewhat between physiologic and the astronomical amounts used in athletes for doping, but this same group of investigators then showed a continuous dose-response for T and muscle size/strength from castrate values through the female range to the male range and beyond. These data strongly support the prominence of T in the genesis of athletic advantage for men over women. Consider the following athletic advantages in men and their origins:

1. Greater muscle size and strength: due to T, as shown in the studies of Bhasin and colleagues.
2. Higher red blood cell mass: due to T, which induces erythropoietin expression and the ~3% higher hematocrit in men versus women.
3. Increased aggression: due to T, as shown in numerous studies of normal and castrate men, transgender men and women, and children.
4. Lack of breasts: due to T, as T counteracts the effects of estrogen to induce breast development, as illustrated in AIS.
5. Narrow hips: actually not due to T but rather due to lack of progesterone, which is derived from the ovaries following ovulation. This is not a male advantage derived from the testes but a female disadvantage derived from the ovaries.

III

SEX VERSUS GENDER

Part of the confusion generated in recent cases derives from an obfuscation of the difference between sex and gender. “Sex” is biologically determined and not the prerogative of the individual, while “gender” is self-determined by each individual person. This Part will address each difference in turn.

19. See id.
21. See Thomas W. Storer et al., *Testosterone Dose-Dependently Increases Maximal Voluntary Strength and Leg Power, but Does Not Affect Fatigability or Specific Tension*, 88 J. CLINICAL ENDOCRINOLOGY & METABOLISM 1478 (2003) (showing a linear correlation of muscle mass and strength gains with the testosterone enanthate does from 50 to 600 mg/week).
A. From Chromosomal Sex to Phenotypic Sex

Our concept of sex dates back to the earliest historical records. Concepts of sex were firmly entrenched in medical science and society at the time when women's sports were first introduced. Sex is an intrinsically fundamental property that we share with other vertebrates and most animal species, and sex forms the basis for reproduction and propagation of our species. Sex has several components:

1. Chromosomal (Genetic) Sex

In most human beings, a 46,XY chromosomal complement (karyotype) equates with male sex and 46,XX with female sex. “Sex reversal” occurs with the chromosomal sex is discordant with one of the following components of sex.

2. Gonadal Sex

A 46,XY chromosome component normally drives the bipotential gonads to develop into testes, while 46,XX affords ovarian development. Testes and ovaries can be completely functional, completely nonfunctional, or partially functional. For example, in Klinefelter's syndrome (46,XXY), the testes can make normal amounts of T during puberty but typically do not produce sperm and lose T synthesis capacity over time. Finally, ovotestes exist in true hermaphrodites with a spectrum of functional qualities.

3. Phenotypic Sex

At birth, only the external genitalia distinguish boys from girls phenotypically. The internal (Müllerian and Wolffian) structures are also sexually dimorphic, but these are not evaluated in most newborns. Because of differences in genetics and in T exposure during fetal life, however, some priming or influence on various organs has already occurred, which manifests later in adolescence and adulthood. At puberty, secondary sexual characteristics progress: breast development and pelvic widening in females versus further growth of the penis, facial and body hair, and upper-body musculature in males. A limited amount of contrasexual characteristics are common, like some sparse facial hair in girls or breast budding in boys, but these changes normally stabilize or regress as puberty progresses and the secondary sexual characteristics expected for the chromosomal sex dominate.

Discordance of genetic and phenotypic sex have been grouped under the terminology “Disorders of Sex Development” or “Differences in Sex Development” (DSD), which replaces the previous “Intersex” or “Pseudohermaphroditism” terms. Although many affected patients and their families dislike these terms, this article will use the term DSD to be consistent with current norms—but with apologies for lack of a better and accepted option for this sensitive topic. Because androgens exert an active effect on multiple

organ systems to drive the constitutive female phenotype to the male phenotype, DSD diagnoses are grouped based on chromosomal sex. Thus, the two largest groups are 46,XY DSD (“undervirilized male”) and 46,XX DSD (“virilized female”). In fact, the aforementioned Daegu study on T levels of elite female athletes at the 2011 IAAF Championships, demonstrated that the prevalence of DSD is seven per 1,000 (consisting of both 46,XY DSD and 46,XX DSD) among athletes in elite women’s competition, a 140-fold increase from the general population. This staggering increase reflects the advantage that intermediate T values provide over the low values found in normal women. The most conspicuous DSD in this group is those athletes with 5R2D,23 who were reared as females and not diagnosed before entering elite competition.

4. Sex of Rearing

Parents are entrusted with the tasks of naming their babies and rearing them in society, and for most societies, this trajectory includes a dichotomous sex of rearing (boy/girl). As expounded below, this dichotomy is not only artificial and inadequate for human biology but also reflects a gender of rearing, although the choice is based primarily on the phenotypic sex and in particular the external genitalia. When these structures are not clearly normal female or male, it is the standard of care to determine the diagnosis leading to this condition. Not only does the diagnosis allow for proper treatment, like corticosteroid replacement for girls with 21-hydroxylase deficiency (21OHD), but also for the opportunity to counsel the family about the natural history of the disorder and the statistics about sex steroid production and sexual maturation, fertility, risk of gonadal malignancy, and other associated conditions. The diagnosis is critical for making treatment recommendations and framing expectations for the family in the best interests of the child.

B. Components of Gender

In contrast to sex, gender is self-determined—only the individual can define their gender, which is reflected in the individual’s interactions with other people in society.24 About 0.5-1% of people in the United States report being transgender, which means that they are living in a gender different from their natal sex and sex of rearing. Constructions of gender and behavioral aspects of gender as different from sex are also found in ancient cultures. In contrast, the biological basis of gender and complexity of a gender spectrum distinct from sex did not appear in the medical literature until well after Jost’s publications on


24. See Wylie C. Hembree et al., Endocrine Treatment of Transsexual Persons: An Endocrine Society Clinical Practice Guideline, 94 J. CLINICAL ENDOCRINOLOGY & METABOLISM 3132 (2009) ("Gender identity is used to describe a person’s fundamental sense of being a man, a woman, or of various indeterminate and flexible genders.").
sexual differentiation. We now recognize that, just as DSD or intersex patients exist, gender is hardly dichotomous and exists as a spectrum, with some individuals living squarely in various intermediate genders or being gender-fluid, not embracing either dichotomous gender, or vacillating between genders based on situation. Sex and gender are concordant and dichotomous in most individuals, which is consistent with a strong influence of androgens and estrogens, however, the causality is far from complete or universal amongst individuals. As an extreme example, there is at least one well-described case of a patient with complete AIS (and therefore female phenotype) but firmly male gender. This individual could not develop male secondary sex (gender-confirming) characteristics despite high-dose T treatment and required surgical construction of a penis equivalent.

C. Care of Transgender Individuals

The care of transgender individuals illustrates the proper medical handling of discordance between sex and gender. The standard of care for a male-to-female trans woman is medical or surgical suppression of T exposure plus estrogen replacement. The institution of gender-concordant hormone therapy with development of gender-affirming secondary sexual characteristics eliminates nearly 99% of the mood disturbances and tendency for personal harm that transgender individuals suffer without proper endocrine care—with or without gender-affirming surgeries to further enhance their phenotype. For a trans woman athlete, the treatment is no different, and World Anti-Doping Agency (WADA) policy specifies lowering T to <10 nmol/L for entering the women’s competition. Not surprisingly, a high fraction of these athletes fail to achieve this goal, despite access to means, given the performance enhancement of higher T. Conversely, female-to-male trans men receive T replacement under a therapeutic use exemption and compete against the men. Here, we find little controversy and accusations for trans men because natal men have comparable T exposure.

---

IV

WHAT CAN WE CONCLUDE?

Based on the principles presented here, an endocrinologist concludes the following:

A. The women’s competition must be defined by sex, not gender

There is no logical reason to separate competitions on the self-determined and potentially fluid parameter of gender, which provides little intrinsic contribution to athletic performance. In contrast, this article illustrates how sex and, in particular, differences in T production is fundamental to the discrepancies in athletic performance typical of men and women.

B. Gender must be respected

We cannot force an athlete to change or present with a different gender just to compete, and all athletes must be allowed to compete as themselves. This principle will require sport and society to be more accepting of the modern concepts of gender and the gender spectrum. We will have to become comfortable with men (gender) competing in women’s (sex) events because that is their biology; with women (gender) competing in the men’s (sex) events because that is their biology; and with genderless or gender-fluid individuals in both competitions based on their biology.

C. The diagnosis matters

We use medical science to train athletes, detect doping, and care for transgender athletes. What would an endocrinologist do for a DSD patient who is not an athlete? We would determine the diagnosis and recommend appropriate therapy—that is the standard of care. The standard of care should be no different if the patient is an elite athlete.

D. Athletes with testes, male T production and normal androgen receptors do not belong in the women’s competition

If we are going to have a “women’s” competition, it is incomprehensible that patients with 5RD2 are allowed to compete against athletes with normal female T values. These athletes are male by every physiologic measure including all those that lead to the male advantages in sport, except atypical external genitalia and impaired fertility, which if anything would provide even greater advantages for athletic performance. While the triaging of athletes with some conditions might be difficult from a scientific perspective, those with 5RD2 are unequivocally male.
E. Athletes with ovaries but higher-than-normal T

The majority of these cases will have some form of virilizing congenital adrenal hyperplasia like classic or nonclassic 21OHD. It is likely that the prevalence of nonclassic 21OHD in elite women’s sports is far higher than the seven per 1,000 figure for DSD. T values for untreated nonclassic 21OHD patients are still well below the male range, whereas classic 21OHD women in very poor control can have T values encroaching into the male range. The flip side for women with classic 21OHD is that the more severe the condition and the higher their T, the lower their cortisol—they have adrenal insufficiency. It is the standard of care to replace the cortisol deficiency in classic 21OHD, and noncompliance carries a risk of adrenal crisis. With at least regular physiologic doses of hydrocortisone, their T values will not be in the male range. Enforcing and assessing compliance will be a difficult challenge, but steroid analysis in hair samples provides a means of long-term assessment. In the end, many of these women will have higher than normal female circulating T concentrations of adrenal origin, but other than instituting the standard of care, there is no compelling and endocrinologically sound reason why they should not be in the women’s competition.

F. Athletes with partial AIS

There is no straightforward answer for these athletes, who in the past were evaluated on a case-by-case basis with liberal criteria favoring the athlete. No alternative to this approach is apparent.

G. Transgender athletes

The current regulations, that transgender athletes should be treated with gender-affirming hormone replacement and/or suppression to achieve T values in the normal range for their trans-sex, are appropriate and consistent with the standard of care. It seems unreasonable that an athlete claims to be a woman (gender) but refuses T suppression and estrogen replacement—this behavior is inconsistent with being transgendered.


29. See Phyllis W. Speiser et al., Congenital Adrenal Hyperplasia Due to Steroid 21-hydroxylase Deficiency: an Endocrine Society Clinical Practice Guideline, 95 J. CLINICAL ENDOCRINOLOGY & METABOLISM 4133 (2010).

V

POTENTIAL SOLUTIONS

A. Eliminate the women’s competition

As an alternative to requiring everyone to compete based on their biological sex regardless of their gender identity or how they present, we could throw up our hands and say that there is no universally acceptable solution to the conundrum and give up. This option would strongly discourage all but the most competitive girls and women from seriously embracing sport and would be a tragedy for the many female athletes who have experienced personal fulfillment from their profession and their successes. This option would also deny countless women competing in high school and college sports, the vast majority of whom do not become professional athletes, an opportunity to develop valuable teamwork and leadership traits in this setting.

B. Let anybody enter the women’s competition

This option might sound absurd. Basing participation on gender, not sex, would open the floodgates for the many male athletes who will go to extreme measures to win medals. Well-publicized studies have demonstrated that many elite athletes would take a substance that would guarantee victory at high risk of premature death.31 Posing to identify as female (temporarily) is a far easier and safer way to gain an advantage.

C. Let the athletes decide

Although this might sound democratic, the vast majority of female athletes will likely want stringent policing of their competition to maintain fairness from their perspective. It is only the minority with DSD that will vote for liberalization, and their voices will not be heard.

D. Develop an algorithm that uses medical science

Some have argued that DSD athletes in the women’s competition come from poor countries without access to medical care and that diagnostic evaluations are intrusive and embarrassing. These arguments are unpersuasive and inconsistent with the care of non-athletes in these same countries. If an adolescent girl has no menses by age sixteen or no menses and no secondary sexual characteristics by age fourteen, a complete medical evaluation is indicated. These criteria are the same the world over, the basic components of medical history and physical exams are available everywhere, and basic laboratory evaluation is available in any developed country. Can we provide athletes with expensive training facilities,

31. See, e.g., Harrison G. Pope, Adverse Health Consequences of Performance-Enhancing Drugs: An Endocrine Society Scientific Statement, 35 ENDOCRINE REV. 341 (2014) (reviewing the high prevalence of performance-enhancing drug use in elite and amateur athletes alike and noting that the athletes consciously accept the increased risks of cardiovascular disease, psychiatric illness, musculoskeletal disorders, other serious chronic health effects, and even death associated with their use).
gear, coaching staff, and doping tests but deny them the most basic medical care? Doing one without the other deprives the athlete of standard basic medical care and exploits minors for the benefit of others. If we are placing the athlete’s health first, they need to have a diagnosis and proper treatment first and foremost. If we only care about winning medals, then we turn a blind eye to an obvious disconnect and allow the other women to compete at an unfair disadvantage.

It is doubtful we will ever arrive at a single criterion for defining athletes who qualify for the women’s competition. Additionally, a single unisex competition is likely not the best solution. Instead, sport should use a basic screen first, and the serum T measurement has performed well because of its physiologic basis. Sports should consider lowering the 10 nmol/L cutoff to 3 nmol/L, which is closer to the upper limit of the normal female range, if we really are serious about providing the best medical care for the athletes because female (sex) athletes with higher (abnormal) T levels might suffer adverse health effects without diagnosis and proper management. Subsequent to this first screen, an algorithm based on sound medical science should distribute athletes fairly into the appropriate competitions. Additional testing will include biochemical and genetic testing of steroid biosynthetic pathways, AR function, and imaging studies, which a panel of experts should design and which will simultaneously provide the standard of care for the athletes. Platforms to sequence of all the genes known that contribute to DSDs on a single small blood sample are available for <$500, a pittance compared to the budget of elite sports. Based on these subsequent results, athletes who may not participate in the women’s competition will be identified. Other athletes who fail the initial T screen will receive a diagnosis, standard-of-care therapy, and clearance for the women’s competition.

These evaluations should be done at home, before any international competition, to avoid the public displays of private information that occur when athletes are denied proper medical care. There is no more embarrassment for an undervirilized male to compete with a female gender in the men’s competition than for a trans man to be in the same competition. No one will know the difference if they start their international competition in this manner. The embarrassment and public exposition is a consequence of accusations that arise from attempting to conceal the obvious.