The Science of Sex Verification and Athletic Performance

Ross Tucker and Malcolm Collins

The division of athletes into male and female categories for competition is a widely accepted practice and is ordinarily straightforward, requiring no intervention from authorities. However, for reasons ranging from deliberate cheating to complex medical conditions resulting in ambiguous development of sex organs, the controversy of sex verification in athletic events has existed for 70 years. Testing procedures, initially implemented to prevent cheating by men masquerading as women, have produced humiliating outcomes for women athletes who were often for the first time confronted with the possibility that they have one of the disorders of sex development. Sporting authorities have and continue to formulate position stands for the management of such cases. An important missing component in this debate is the sound scientific evidence to determine (a) whether a performance advantage actually exists and (b) how large it might be. The division of competition into separate categories and the large difference in sporting performance between male and female necessitate that sport-governing bodies define the boundaries between the sexes in a just and fair manner for all participating athletes. This review will therefore provide the historical context of the debate and aim to discuss relevant physiological and performance aspects of the sex verification process.

Keywords: performance, gender, testing, disorders of sex development, intersex

The division of sports competitors into separate categories for competition is a widely accepted practice. The primary purpose of these classifications is to ensure equality of competition, and to control for factors such as age, weight, and even skill level, as in the hand-to-hand combat sports such as karate at the amateur level.

The classification of athletes into gender categories is ordinarily straightforward, requiring no intervention from authorities, and receiving little attention from those viewing and participating in a sport. However, for reasons ranging from deliberate cheating to complex physiology that results in a discord between genetic and anatomical sex and ambiguity in sex organs, the controversy of males competing in female classifications has existed for 70 years.1–6

Tucker is with the UCT/MRC Research Unit for Exercise Science and Sports Medicine of the Department of Human Biology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa. Collins is with the UCT/MRC Research Unit for Exercise Science and Sports Medicine of the Department of Human Biology, Faculty of Health Sciences, University of Cape Town, and with the South African Medical Research Council, Cape Town, South Africa.
Recently, the testing procedure, performances, and subsequent media speculation involving South African 800-m athlete Caster Semenya have generated renewed debate in this controversial area. Debate has ranged from gender rights to genetics, sports performance, and physiology. Much of this debate lies outside the scope of the present review, which instead aims to describe relevant physiological and performance aspects of the sex verification process.

**Sex Verification and Biological Complexity**

To understand the complexity of current sex verification policies,\(^7\) it is instructive to use the historical development of testing methods as the background to explain the complexity and genesis of some of the conditions that produce ambiguity in sex development.

First, it is important to emphasize that the words *sex* and *gender* are often used interchangeably in the discussion. This is not accurate, for *sex* is a biological term denoting the genetic and anatomical characteristics of an individual, whereas *gender* is a sociological construct that denotes how an individual identifies themselves according to social norms.\(^8\) The term *gender verification*, used widely in the debate and media, thus refers to whether an individual has appropriately identified themselves for the purposes of athletic competition. The term *sex* is, however, the correct one when discussing the methods used previously to differentiate between men and women in sporting competitions.\(^8\)–\(^10\)

The earliest documented controversies around gender in sporting competition come from the 1932 Olympic Games. The 100-m gold medal was won by a Polish athlete resident in the United States—Stella Walsh (original Polish name, Stanisława Walasiewicz), who became dubbed “Stella the Fella” by the media as a result of her masculine appearance. During the 1936 Berlin Olympic Games, Walsh competed against an American athlete, Helen Stephens, in the 100-m final. During the lead-up to the race, they accused one another of being male. Although there was no official testing during the Games, both are reported to have been asked to undergo a physical examination, which they passed. The race was won by Helen Stephens, with Stella Walsh winning the silver medal. After her death in an accidental shooting during a bank robbery in Cleveland in 1980, Stella Walsh underwent an autopsy, which declared the presence of “ambiguous genitalia.” It has been reported that genetic investigations revealed that she had a mixture of cells with either XY or XX chromosomes\(^6\) (http://en.wikipedia.org/wiki/Stella_Walsh and http://en.wikipedia.org/wiki/Helen_Stephens, Accessed 12 April 2010).

The 1936 Berlin Olympic Games were also characterized by another controversy, one involving the German athlete Dora Ratjen. Ratjen finished fourth in the high jump, and then went on to break the world record in 1938. However, while traveling by train in September of 1938, her fellow passengers became suspicious of her, and the police were then notified that a man was dressed as a woman. A doctor’s examination revealed male genitalia, and it has subsequently been alleged that Dora Ratjen was in fact born with ambiguous genitalia, but was registered as a girl. After this incident, Dora changed her name to Heinrich Ratjen and retired from athletics (http://en.wikipedia.org/wiki/Dora_Ratjen, Accessed 15 June 2010).
The Era of Physical Examinations: Crude Sex Assessments

It took until the 1960s for widespread concerns over this form of cheating to see the introduction of the first measures to screen athletes to ensure their eligibility as females. In the absence of more sophisticated tests and understanding of the complexities of sex development, the initial testing procedure consisted of a parade by all female competitors, in the nude, before a panel of judges. This so-called nude parade generated substantial resentment and would be replaced by molecular techniques over the course of the next decade. Apart from the resentment and embarrassment caused to the athletes, the risk of incorrect evaluations was also high because conditions exist that affect the development of the sexual characteristics (discussed subsequently), thus resulting in ambiguous genitalia, making accurate physical examination difficult.

These screening processes did influence participation in female events, however. The Press sisters of the Soviet Union, Irina and Tamara, had combined to win ten medals at Olympic and European Championships between 1960 and 1966. However, the introduction of compulsory physical screening saw their immediate retirement from all athletic competition, amid widespread allegations that they were men masquerading as women.

The Era of Molecular Sophistication and Physiological Complexity

The advent of molecular techniques, which followed Watson and Crick’s proposed molecular structure of DNA in 1953, invited the possibility that testing accuracy could be improved while negating the need for such embarrassing procedures as the “nude parade.” The first method used, introduced in the 1967 European Cup of Track and Field Athletics and the 1968 Grenoble Winter Olympic Games, and then made compulsory at the 1968 Mexico City Olympic Games, was the buccal smear or sex chromatin test to identify the presence of a Barr body. The Barr body is a microscopic body on the inside of the nuclear envelope, and is formed in the nucleus of cells in females when the second X-chromosome is inactivated during early development. In theory, this body would be present only when a second X-chromosome is present. Males, possessing only one X-chromosome, would produce a negative result in the Barr body test. A negative result in this test would mean exclusion from competition on the basis that the sample had to have come from a genetically male competitor, whereas with a positive test, confirming the presence of two X-chromosomes, a certificate of femininity would be issued.

The first Olympic athlete to be disqualified using these new molecular techniques was a Polish sprinter, Ewa Kłobukowska. A multiple Olympic medal winner in 1964, and a world record holder in the women’s 100 m in 1965, she was banned from competing in professional sports after failing the early molecular tests, although her chromosomal anomaly was never revealed (http://en.wikipedia.org/wiki/Ewa_K%C5%82obukowska, Accessed 12 April 2010).

A significant problem with the Barr body test is that of false-negatives and false-positives, which can be produced as a result of any one of a number of physiological conditions that affect sex development. The most common of these
Androgen insensitivity syndrome is one of a number of different conditions that can result in ambiguity between the chromosomal sex and the anatomical sex of an individual. Collectively, these conditions have been termed disorders of sex development.
Sex Verification and Performance

development (DSDs)\textsuperscript{6,15} and they are defined as a spectrum of conditions ranging from those with ambiguous external genitalia to those with normal female external genitalia but varying degrees of internal testes.\textsuperscript{2,15} Included in these DSDs are AIS (described previously) and various other conditions, the description of which lies beyond the scope of this review and have been previously described by Ritchie et al\textsuperscript{6} and Lee et al.\textsuperscript{15}

To understand the impact of these DSDs on the testing process of athletes, it is important to understand the basic process of sex development. The first steps of sexual development or differentiation take place in the embryo at about 5 wk, when the presence of genes on the X- and Y-chromosomes directs the development of the undifferentiated gonads. When two X-chromosomes are present and the Y-chromosome is absent, the gonads develop into ovaries, with the resultant female development. When the Y-chromosome is present, the gonads are signaled to develop into testes, which produce testosterone, and other male hormones, which ultimately result in the development of male reproductive organs, and, at puberty, development of secondary sexual characteristics such as increased musculature and facial hair.\textsuperscript{12}

Testosterone and other androgens such as dihydrotestosterone are responsible for the development of the primary and secondary sexual characteristics. The primary sexual characteristics are those involving the development of the reproductive organs. Secondary sexual characteristics in males include deepening of the voice, hirsutism (excessive hair growth) as well as the pattern of hair growth (face, chest, underarms), and increased skeletal muscle mass, which is the characteristic linked with athletic advantages experienced by male competitors over females. A key hormone in this process is the biologically active metabolite of testosterone, dihydrotestosterone, which is produced by the enzyme $\alpha$-5-reductase and is three times more potent than testosterone. In a condition known as $\alpha$-5-reductase deficiency, this hormone is not produced, and the result is that an individual fails to develop male characteristics, including the penis, which can resemble a clitoris, and are often raised as girls. At puberty, however, their musculature develops owing to normal levels of testosterone.

These are disorders in which a discrepancy exists between chromosomal and anatomical sex as a result of the absence of or insensitivity to key hormones in the sex development process. Another category of DSD involves genetic abnormalities in which the chromosomal sex is ambiguous. These include conditions such as Klinefelter’s syndrome (two X and one Y sex chromosome), Turner’s syndrome (one X sex chromosome), and mosaicism (mixture of cells with XX and XY or X and XY sex chromosomes).\textsuperscript{6} With the possible exception of some forms of mosaicism, where some cells express the XX combination of chromosomes whereas others have an X- and a Y-chromosome, these conditions are not usually thought to confer any performance advantage because of associated health complications. For example, individuals with Klinefelter’s are usually under-masculinized males and thus highly unlikely to possess athletic capabilities that challenge the very best female athletes.

A final common condition among the DSDs identified is congenital adrenal hyperplasia, in which the adrenal glands produce excessive amounts of testosterone in females. These individuals can develop secondary male characteristics but are genetically female (46, XX), lacking testes and male reproductive organs. Individuals with this condition are often referred to as masculinized females, in contrast to
individuals who are genetically male but develop as females—the so-called under-masculinized males, according to the classification of such individuals discussed by the International Consensus Conference on Intersex. A question remains as to whether these conditions would also confer performance advantages on individuals who have them. In all cases, the individual may present as a female on the basis of their sex organs, which are ambiguous and identified at birth as either male or female (and often, regrettably, incorrectly). However, they would also develop male characteristics as a result of the androgen effects of the DSD. Theoretically, these androgen effects could confer large performance advantages to these individuals, since testosterone stimulates muscle development, and resultant speed, strength, power, and endurance. However, no evidence that we are aware of has yet documented whether this theory is borne out, although this is further complicated by the fact that these DSDs are likely to occur to varying degrees, and in a relatively small proportion of the population. The result is that definitive answers as to whether these conditions may provide a performance advantage remain elusive.

The Abolition of Compulsory Testing

The compulsory, laboratory-based testing of all female participants, begun in 1968 at the Mexico City Olympic Games, produced controversy and opposition among endocrinologists, geneticists, and medical practitioners. It was repeatedly pointed out that the procedure, which had been created to prevent cheating by men masquerading as women, had not once managed to detect a cheat. It had, however, produced humiliating outcomes for women athletes who were not deliberately cheating. The International Olympic Committee (IOC) and the International Association of Athletic Federations (IAAF) began to consider the merits of compulsory testing in the 1980s, when it was repeatedly stated that “the aim of gender verification tests is not to differentiate between sexes but to prevent male impostors from participating in female competitions.”

The IAAF reached a decision to ban compulsory gender verification testing in 1992, replacing it with the option to assess an athlete if suspicion arises. The formal process of gender verification was, however, described by Simpson and colleagues as “moot” because the antidoping process requires that a doping official be present in full view when a urine sample is passed.

Eventually, the swell in opposition to compulsory testing led the IOC to pass a resolution “to discontinue the current process of gender verification during the Olympic Games” at the 1996 IOC World Conference on Women and Health. In 1999, the IOC Executive Board approved the abolition of compulsory sex verification testing, meaning that the Sydney Olympic Games were the first since 1964 at which female athletes were not forced to pass a test to obtain a clearance certificate to compete as a woman.

The current IAAF policy on gender verification, prepared in 2006, states that there is “no compulsory, standard or regular gender-verification during IAAF sanctioned championships,” and that “in resolving cases that may arise, determination should not be done solely on laboratory based sex determination,” reflecting the realization that such methods are fraught with difficulty.
The Challenge System

The abolition of compulsory testing by first the IAAF and then the IOC was applauded by many scientists.\(^\text{14}\) However, it did leave unanswered the difficult question of whether athletes with DSDs should compete as females when they may possess a performance advantage over other female athletes as a result of a condition that is characterized by either possessing testes or having excessively high testosterone levels. As mentioned, the theoretical argument for a performance advantage as a result of elevated testosterone levels exists, but we are unaware of any evidence that DSDs either do or do not provide a performance advantage.

Sport-governing bodies thus find themselves in a difficult position. They had rightly identified that compulsory laboratory testing, established to catch cheats, had never done so, and instead humiliated a number of female athletes who were banned from competing, perhaps unjustly. However, the governing bodies also have the responsibility of ensuring equality of competition for all athletes within the categories they have established for competition to take place in—that is, male and female. The presence of two divisions necessitates that some distinction be made between males and females, however controversial that distinction may prove to be in certain instances. This forces authorities to confront the very difficult question of what to do about individuals with a DSD, who blur the simple distinction between males and females, and who may possess a performance advantage.

The IAAF, recognizing this potential issue, leave allowance in their policy for “challenges” to be issued by other athletes or teams.\(^\text{7}\) Such challenges would result in a more comprehensive medical evaluation by a team of experts including a gynecologist, endocrinologist, internal medicine expert, geneticist, and an expert on gender/transgender issues. It is this process that was leaked to the media in the case of Caster Semenya.

Competition Categories and Performance Implications

As previously discussed, gender categories are not the only such categories in sport—athletes regularly compete in weight, age, and even skill categories, as in the case of certain combat sports such as karate. The fair management of these categories is dependent on the accuracy and validity of measurement of the categorical variable, so that an athlete can be placed into the appropriate competition.

For example, the classification of boxers into one of 11 weight categories at the Olympic Games in 2008 was achievable with an accurate scale and a standardized method for measurement. Sports for which weight would provide a substantial advantage to heavier competitors often adopt this method, which is not without controversy, and athletes are known to adopt extreme weight loss tactics, including doping, to fall into a desired category. Similarly, athletes compete in age categories to ensure equality of competition while skill, strength, speed, and other physiological attributes develop with age. These classifications are often somewhat arbitrary, but their purpose is to ensure that individuals who would otherwise be at a disadvantage for reasons related to weight or age (and its associated physiological effects) can compete fairly with peers.
The classification of athletes into male and female categories is a similar separation devised to ensure fair competition, although only infrequent cases such as that of Caster Semenya force it to be questioned. The fundamental problem with this gender classification is that unlike the relatively simple measurement of weight, for example, the classification of athletes into gender categories is enormously complex in those cases where ambiguity exists.

Fundamental to this debate are performance differences between males and females. These differences are attributed primarily to the hormone testosterone, produced in substantially higher concentrations by males at the onset of puberty, leading to increased muscle mass; increased cardiovascular and lung capacity; and increased strength, speed, and power. Indeed, the only Olympic events where males and females compete together are the equestrian events (www.olympic.org).

An analysis of world record performances highlights the performance advantage enjoyed by males in the track events. The best males outperform the best females by between 9 and 14% in events ranging from 100 m to a 90-km ultramarathon. In the marathon, for example, the current women’s world record of 2 h 15 min 25 s does not feature in a listing of the top 400 times for men from 2009 (www.iaaf.org/statistics/toplists/index.html, Accessed on 12 April 2010), and lies outside the top 1,000 performances in the history of the marathon. This is characteristic of all women’s performances—the best female performances in history lie outside the top 400 men’s performances every year, and often fall outside the top 1,000 performances (www.iaaf.org/statistics/toplists/index.html, Accessed on 12 April 2010).

In Figure 1, performance results from selected running events at the 2009 IAAF World Championships in Berlin are compiled into frequency histograms. The figure clearly shows the gap that exists between the best performing men and the best performing women. In all events with the exception of the 10,000 m, there is some overlap between the best female performance and the male performances. This overlap is important because it means that performance alone is insufficient in distinguishing between males and females in a large enough sample.

However, it must be noted that at the World Championships, restrictions on competitors per country means that many male competitors do not compete at this event, which exaggerates the overlap, since the best women are being compared with men who are not necessarily the top global performers. The overlap in performances disappears completely when the world-ranking lists are analyzed and the very best males are compared with the best females (data not shown).

The implications of these findings are twofold. First, a female athlete cannot be excluded solely if her performance overlaps with male performances—the reality is that a large enough sample of each will overlap. However, it is equally important to note that the performance difference between the best performing males and females is so large that any direct comparison is irrelevant—the difference in athletic performance is vast. Thus, a female who benefits from excessive androgen utilization may possess a large advantage over her peers. This is the basis behind doping with anabolic steroids, a practice widely used in the 1980s, and which is deemed responsible for many of the current women’s world records, which have remained unchallenged since the 1980s.

It is thus clear that two separate categories are required if females are to be given the opportunity to compete fairly against one another to achieve success—
were the gender categories to fall away, no female would ever stand a chance of achieving victory in these athletics events.

When an individual with a DSD competes as a female, and may have a performance advantage, the question is whether they should be placed into the female classification, or whether they should cross over into male competition. The separation cannot be based on performance, since there is already overlap between males and female performances, as described above. In addition, the issue should not be whether the female’s performances approach those of males, but rather whether she has an unfair advantage over her female peers—the comparison with male times and physiology is largely irrelevant. The classification also cannot be done based on testosterone levels, because the ranges in males and females are large, and although not overlapping normally, individuals with conditions such as AIS may have high testosterone levels without being able to use the androgen.

**Natural and Unfair Physiological Advantages and Categorical Variables in Sport**

There is also the argument that within a given population (males or females, for example), individuals who achieve sporting success possess a particular combination of physiological characteristics that is at least partly the result of their genetic makeup. Sporting performance is of course multifactorial, and to attribute success to single physiological traits would be an oversimplification. However, to illustrate the concept, it might reasonably be assumed that track sprinters possess a higher proportion of fast-twitch muscle fibers and a high capacity for force production, whereas basketball is a sport in which having a greater stature can reasonably be assumed to be a performance advantage. Similarly, at the elite level, smaller athletes often have an advantage in sports such as cycling or marathon running. It has even been found that in rare cases, mutations in the gene coding for erythropoietin can significantly increase oxygen-carrying capacity by preventing normal feedback control of red blood cell mass. It can be reasonably assumed that this genetic mutation will improve endurance exercise performance, with Finnish cross-country skier Eero Mäntyranta often cited as an example of this phenomenon.

An extension of this position, and an argument that has been made in the gender debate, is that individuals with DSDs who identify themselves as female have a naturally occurring performance advantage over other females. This “advantage,” it is further argued by proponents, is “natural” and does not represent the intent to cheat, and should thus be accepted and allowed in much the same way that other physiological advantages such as size, speed, or strength are accepted.

We certainly do not possess the answers to this controversial issue. However, it is a position that invokes some intriguing debates. Perhaps the fundamental consideration is that in basketball, for example, athletes do not compete in categories of height, and so competition is deemed “open.” Taller players, possessing the physiological advantage, will tend to succeed in this sport, but since there is no criterion for height, there can be no discrimination against these athletes on the basis of this variable. Even in the case of a mutation that increases oxygen-carrying capacity, it can be argued that athletes do not compete in categories that separate athletes according to red blood cell mass or some other oxygen-carrying variable.
Figure 1 — Frequency histograms of male (solid bars) and female (clear bars) competitors who participated in selected running events at the 12th IAAF World Championships in August 2009 in Berlin. All the finishing times for the male and female heat and final events were downloaded from the official games Web page (http://berlin.iaaf.org/results/). Data were sorted according to finishing times. The fastest time of an individual athlete that qualified to the quarter-, semi-, and/or finals was used in the analysis. A race in which an athlete did not finish or was disqualified was excluded from the analysis.
The number of athletes (n) included in the analysis for each event is indicated. The current male (m) and female (f) world records (WR) for each event are also indicated. a—indicates a WR time during the 2009 Berlin IAAF Championships; b—Caster Semenya’s finishing time of 1 min 55.5 s for the women’s 800-m race; c—the four female athletes with finishing times for the 800 m ranging from 2 min 13.4 s to 2 min 31.9 s are not shown on the graph; d—the single female athlete with a finishing time of 3 h 59 s for the standard marathon is not shown on the graph.
Again, competition is “open” according to these criteria, although for safety reasons, it may be feasible to prevent the athlete from participating if their hematocrit levels are deemed a health risk.

However, athletes do compete in male and female categories. Competition is thus not “open,” and so the argument that a DSD (which challenges the categorization) is a naturally occurring advantage analogous to speed or height seems disingenuous. Gender categories exist for the very reason that performance differences between males and females require that two separate categories exist. If, for example, a height category for athletes below a certain height were created to facilitate participation by shorter individuals in the sport of basketball, then fairness of competition would require that athletes taller than this limit be compelled to participate in the higher competition, even though their height is “natural” and represents no intent to cheat. In much the same way, authorities must defend equality of competition in the female category when the equality is questioned as a result of a physiological factor that challenges the basis for the gender categories in the first place.

The extreme extension and application of this argument would ultimately see the abolition of separate male and female categories in sport. Rather, all athletes would compete in open competition, since accurate categorization of males and females is difficult, and since “natural” genetic differences, of which sex is an example, are allowed. Attempting to maintain the separate gender classifications, but failing to appropriately manage where individuals with DSDs and potential performance advantages should compete would be analogous to having a system where weight classifications exist, but then waiving the weight limit for certain individuals who cannot reduce their weight enough to fit into the required category. It is for this reason that sport-governing bodies are compelled to grapple with the issue of DSDs and classification, and simply dismissing them as naturally occurring advantages in performance would compromise the integrity of the sport.

The above analogies are intended to illustrate a key principle, rather than to provide instruction on how DSDs should be managed. This is because a key factor in the management of male and female categories is the ability to accurately and fairly categorize individuals as male and female. As we have made clear in this review, this is not a simple matter, and it is far more complex than enforcing height or weight restrictions, as in our examples above. The principle, however, remains, and if gender categories are to exist, then authorities cannot simply disregard genetic or physiological conditions as natural and similar to characteristics such as speed, size, or height.

Finally, the most important missing component of this debate is the sound scientific evidence to determine (a) whether a performance advantage actually exists and (b) how large it may be. Certainly, the prevalence of complete AIS appears to be substantially higher in athletic populations than in the general population—1 in 421 through five Olympic Games, compared with approximately 1 in 5,000 for disorders of sex development (excluding Turner’s and Klinefelter’s syndromes) in the general population.

However, this does not constitute evidence of an advantage. Presently, forcing athletes with conditions such as AIS, congenital adrenal hyperplasia, or α-5-reductase deficiency to seek medical treatment before being eligible to compete assumes that a large advantage exists, even though individuals with these conditions often experience side effects that prevent them from achieving sporting success.
Conclusion

Gender verification testing protocols have been the source of substantial controversy ever since they were first introduced in the 1930s. Sport-governing bodies are faced with the challenge of ensuring equality of competition within the categories they have identified, but the lack of evidence for how intersex conditions may affect performance makes this challenge difficult to overcome. Numerous philosophical and sociological arguments are generated by a debate on gender categorization and performance, including the debate over whether any advantage an individual may experience as a result of a DSD constitutes a natural advantage or an advantage that should preclude participation in the female category.

The gender verification testing process has evolved from crude to more sophisticated methods, without full resolution. Recent controversies and the resultant discussion may continue to stimulate discussion, but probably without satisfactorily resolving the complex issue for all involved parties.

References

7. IAAF policy on gender verification. 2006. IAAF Medical and Anti-Doping Commission.