

SCIENCE AND SOCIETY

Debating a testosterone “sex gap” in elite athletes

T-based policies unfairly exclude some women from competition.

By Katrina Karkazis^{1*} and Rebecca Jordan-Young²

Sex dimorphism of testosterone (T) in elite athletes was at the center of a recent case at the “Supreme Court of Sport”, the Court of Arbitration for Sport in Switzerland, after teenage Indian sprinter Dutee Chand challenged a sports policy regulating competition eligibility of women with naturally high T. The idea of a “sex gap” in T is a cornerstone of this policy (1). Policymakers infer that men’s higher T is the “one factor [that] makes a decisive difference” between men’s and women’s athletic performances (2)—so they claim that women with naturally high T unfairly enjoy a “massive androgenic advantage” over other women athletes (2). Yet there’s an emerging scientific debate about whether the sex gap in T applies to elite athletes.

In 2011 and 2012, respectively, the International Association of Athletics Federations (IAAF) and the International Olympic Committee (IOC) adopted controversial policies that regulate levels of natural T in women athletes (1, 3). The IAAF policy sets a ceiling of 10nmol/L for women, which it identifies as “within the normal male range”, whereas the IOC policy targets levels “within the male range” rather than setting a threshold (1, 3). Women with high natural T, according to the IAAF, have unfair advantage over women with lower natural levels (1, p. 12). Unless they are androgen resistant, women must lower their T in order to continue competing (1), which would require surgery or anti-androgens (4).

Appealing her exclusion under the IAAF policy, Chand told the *Indian Express* “At every level of my life...I have competed the way I am. I’ve been told the hormonal issue with me is natural so that’s why we have decided [to appeal]” (5). Her March appeal was the first formal challenge to the policy; a decision is forthcoming.

The T policy is the latest attempt to use a biological marker to draw a bright line between women and men for sex-segregated sports. Decades of sex testing of all women athletes relied on biomarkers such as chromosomes. Officials dropped blanket testing in the 1990s, acknowledging that sex is irreducibly complex

and that there is no scientific criterion for separating all men from all women (6). Nevertheless, they retained an ad hoc policy for when a woman’s sex was questioned, which was criticized for continuing the doomed project of sex testing and for being arbitrary (7).

Still determined to find a biological way to regulate who can compete as a woman, policy makers turned to testosterone, arguing that T is both the “performance enhancing hormone” (8) and a sharply differentiated trait between men and women (2,3). In most studies, men’s T levels are about ten times that of women, and the highest levels seen in women are well below the lowest levels seen in healthy men. One policy-maker characterized this as “a huge no man’s land” (9).

Policymakers have repeatedly stressed that: “there is no overlap between testosterone blood levels in healthy men and women” (2). Recently, though, the idea of unequivocal sex dimorphism in T levels has been challenged, at least in elite athletes.

Only two large-scale studies of T in elite athletes exist, and they draw contradictory conclusions regarding a sex gap in T (10, 11). The first published study analyzes data from 446 men and 234 women across 15 highly varied Olympic events (10). These data were collected as part of the GH-2000 study, an IOC- and World Anti-Doping Agency-funded project aimed at developing a test to detect human growth hormone abuse. The GH-2000 data show that “hormone profiles from elite athletes differ from usual reference ranges” in both men and women (10, p. 1). In fact, there was “overlap between men and women, although the mean values differ”. Among women, 13.7% had T above the typical female range, and 4.7% were within the typical male range. In contrast to clinical reference ranges, 16.5% of these elite male athletes had T below the typical male range, with 1.8% falling in the female reference range.

Not long after the GH-2000 report appeared, IAAF researchers published their own study on serum T in 849 elite women athletes in track and field from the 2011 Daegu World Championships (11). That study showed just 1.5% of women athletes with T above the female reference range, a sharp contrast with the 13.7% in the GH-2000 study.

DEBATING THE EVIDENCE. Three critiques of the GH-2000 report raised by IAAF policymakers in the Daegu study and in an IAAF/IOC rebuttal bear on whether there is a sex gap in T (11, 12). The first issue is how sera were analyzed: the GH-2000 study used immunoassay (IA), while the Daegu study used mass spectrometry (MS). IA overestimates T at lower values, which may partly explain the discrepancy between the studies in terms of the proportion of women athletes who show T levels above the reference range. There is no question that MS yields more accurate T readings at lower values. However, IA would overestimate T in *anyone* with low levels—male or female. When the focus is simply on women with high T, the argument over method might seem reasonable because some women’s true values would be lower than were reported. But the overlap in men’s and women’s T levels isn’t simply the result of some women with high levels: a considerable proportion of men have low levels. If anything, the use of IA would understate the proportion of men with very low levels of T. So the use of IA cannot account for the finding of a male-female overlap in the GH-2000 data. The Daegu study did not report men’s values, so it can only shed partial light on the question of a gap.

The second disagreement concerns when to draw serum because T changes in response to competition. Citing evidence that men’s T falls while women’s may stay the same or rise just after competition, IAAF/IOC policymakers suggest that the female-male overlap in T observed in the GH-2000 data may be an artifact of drawing samples within two hours after competition (12).

This criticism requires a selective reading of the evidence on the effect of competition on T levels. The IAAF/IOC critique cites a single report showing male T levels dropped and female levels were steady or modestly rose after an Ironman competition (13). The broader literature, though, shows that T may rise, fall, or remain unchanged after competition, and the main factors determining the response seem to be the type and duration of competition, not sex (14, 15). Intense resistance exercise and exercise of shorter duration are associated with increase in T, whereas endurance exercise (es-

¹Stanford University, Stanford, CA, USA. ²Barnard College, New York, NY, USA. *Email: karkazis@stanford.edu

pecially lasting greater than 3 hours) is associated with decrease in T (14, 15). There are few data on endogenous T in women athletes, but the most recent review again indicates a great variety of responses to exercise are possible—including a large and lasting increase in resting T from long-term resistance training (14).

The timing of serum collection in the GH-2000 study makes sense in the anti-doping context, reflecting: “a need to clarify hormone profiles that may be expected to occur after competition when anti-doping tests are usually made” (10). Doping tests are often how women with high natural T are flagged, so understanding the response of natural T to competition is particularly important.

The tussle over timing may obscure the important point that T is dynamic. Recent research shows that, in both sexes, T dramatically responds to physical situations as well as social cues and contexts, diurnal rhythms, training, and other factors (14–17). For example, positive feedback from a coach can cause a rapid doubling of T level (though there is great variation among individuals) (17). Thus there is no universally correct time to sample T—it depends upon the purpose of the study. As with the choice of IA versus MS, the timing of blood draws is unlikely to account for the presence or absence of an overlap in T between the sexes.

The third issue raised by the IOC/IAAF critiques of the GH-2000 study is the most fundamental: the rules for subject inclusion and exclusion. Both scientific groups agree that subjects who have doped should be excluded. Where they part is whether women with naturally high T should be excluded.

The two camps take opposite views on whether to include these women—a decision that bears directly on whether their findings support or undermine the policies. The GH-2000 study includes all women with high natural T in the sample. The Daegu study included women with high T of unknown etiology, but excluded as “confounding factors” all women whose high natural T can be traced to Diverse Sex Development, also known as Intersex (DSDI). In simplest terms, women with DSDI have some biological characteristics that would be classified as female and others that would be classified as male. This challenges common ideas about sex, but it is widely recognized in medicine, law, and the social sciences that when people are born with mixed markers of sex (e.g., chromosomes, genitals, gonads), the medical standard is that gender identity is the definitive marker of sex—there is no better criterion (18).

What, then, is the logic that classifies women with DSDI as confounders? The Daegu

report consistently pairs clinical language such as “diagnosis” and “disorder” with hyperandrogenism for the women with DSDI, and in their rebuttal to the GH-2000 paper, IAAF/IOC policy makers use the phrase “hyperandrogenic disorders of sex development” (12). This signals their judgment that women with DSDI are not healthy, and therefore should be excluded from reference ranges. But DSDI women aren’t necessarily unhealthy. High T can be associated with health issues, but is not in and of itself a health problem for women (4).

An a priori understanding of women with DSDI as unhealthy and thus outside normal variation creates a rationale for their exclusion both in reference ranges and the policies. But it’s also circular: because women with DSDI are a priori excluded when the reference ranges are created, the findings from the Daegu study—that women athletes have T no different from non-athlete women—reinforces their values as outsiders and justifies the policy.

There is a strong scientific argument for including DSDI women in the sample. These studies aim to establish T reference ranges for elite athletes: i.e., the focus is on physiological ranges, not clinical ranges. This calls for descriptive statistics, and in this case, there’s no valid basis for discarding some values as outliers. In both studies, if the full range of values for women’s endogenous T is included, there is an overlap in T.

CALCULATING FAIRNESS. What looks like a controversy rooted firmly in science is ultimately a social and ethical one concerning how we understand and frame human diversity. These assessments are not trivial: they shape not only the research methods and findings, but also how we understand what is at stake in this policy. And this has very real consequences for people’s lives.

Policy makers, among others, claim the problem is that women with naturally high T have unfair advantage despite acknowledging in their Daegu study that “there is no clear scientific evidence proving that a high level of T is a significant determinant of performance in female sports” (11).

Others see a very different problem: that women who have lived and competed as women their whole lives suddenly find themselves having to undergo medical interventions in order to remain eligible to compete in a category to which everyone agrees they belong.

Calculating what counts as a fair and level playing field for women must take all women athletes into account, including those with naturally high T and/or DSDI. We could return to a consensus reached decades ago, where poli-

cymakers faced these same concerns and concluded that women “who were raised as girls and classify themselves as female should not be excluded from competition as women” (19). In other words, ensuring that women with high endogenous T and/or DSDI “have the same rights to participation in athletics as all women” (20) would be a good place to start.

REFERENCES

1. International Association of Athletics Federations, IAAF regulations governing eligibility of females with hyperandrogenism to compete in women’s competitions (IAAF, Monaco, 2011), vol. 2011, pp. 1–28.
2. S. Bermon, M. Ritzen, A. Hirschberg, T. Murray, *Am. J. Bioethics* **13**, 63 (2013).
3. International Olympic Committee, IOC Regulations on Female Hyperandrogenism (Lausanne, Switzerland, 2014).
4. R. M. Jordan-Young, P. H. Sonksen, K. Karkazis, *Br. Med. J.* **348**, 4 (2014).
5. V. Bhushan Padmadeo, “I am what I am, says sprinter Dutee, refuses hormone therapy,” *Indian Express*, 12 September 2014.
6. J. L. Simpson et al., *JAMA*, **284**, 1568–1569 (2000).
7. L.J. Elsas et al., *Genetics in Medicine*, **2**, 249–54 (2000).
8. International Association of Athletics Federations, HA Regulations: Explanatory Notes (IAAF, Monaco, 2011), vol. 2011, pp. 1–4.
9. J. Macur, “I.O.C. adopts policy for deciding whether an athlete can compete as a woman,” *New York Times*, 24 June 2012.
10. M. Healy, J. Gibney, C. Pentecost, M. Wheeler, P. Sonksen, *Clin. Endocrinology* **81**, 294 (2014).
11. S. Bermon et al., *J. Clin. Endocrinol. Metab.* **99**, 4328 (2014).
12. M. Ritzen et al., *Clin. Endocrinol.* **82**, 307 (2015).
13. G. Ginsburg, M. O’toole, E. Rimm, P. Douglas, N. Rifai, *Clinica Chimica Acta* **305**, 131 (2001).
14. J. L. Vingren et al., *Sports Medicine* **40**, 1037 (2010).
15. C. Enea, N. Boisseau, M. A. Fargeas-Gluck, V. Diaz, B. Dugue, *Sports Medicine* **41**, 1 (2011).
16. S. van Anders, N. Watson, *Psychoneuroendocrinology* **31**, 715 (2006).
17. C. J. Cook, B. T. Crewther, *Hormones & Behavior* **61**, 17 (2012).
18. P. Lee et al., *Pediatrics*, **118**, e488–500 (2006).
19. A. Ljungqvist A, J. Simpson, *JAMA*. 277, 850–852 (1992).
20. Women’s Sports Foundation, Participation of Intersex Athletes in Women’s Sports (WSF, New York, n.d.), pp. 1–5.

10.1126/science.aab1057