



PERSPECTIVES



An athlete competes in the women's triple-jump final in the 2012 Olympic games in London, UK.

SCIENCE AND SOCIETY

Debating a testosterone “sex gap”

Policies unfairly exclude some women athletes from competition

By **Katrina Karkazis**¹
and **Rebecca Jordan-Young**²

Sexual 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). Policy-makers infer that men’s higher

POLICY T is the “one factor [that] makes a decisive difference” between men’s and women’s athletic performances (2)—so that women

with naturally high T may unfairly enjoy a “massive androgenic advantage” over other women athletes (2). We report on 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 for women of 10 nmol/liter in blood, which it identifies as “within the normal male range,” whereas the IOC policy targets levels “within the male range” (1, 3). Women with high natural T, according to the IAAF, have an unfair advantage over women with lower

natural levels (1). Unless they are androgen-resistant, women must lower their T in order to continue competing (1), which would require surgery or antiandrogens (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 chro-

mosomes. 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 10 times those 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).

Recently, though, the idea of unequivocal sexual 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). In the first, data are 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 report states, "hormone profiles from elite athletes differ from usual reference ranges" in both men and women (10). 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 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 policy-makers 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); the Daegu study used mass spectrometry (MS). IA overestimates

T at lower values. There is no question that MS yields more accurate T readings at lower values. The use of IA in the GH-2000 study might have resulted in some inaccurately high readings among women, but it cannot explain the fact that a considerable proportion of men had very low T levels (in fact, IA underestimation would have countered the latter pattern). 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 re-

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 antidoping context, because of the need to understand hormone profiles "after competition when anti-doping tests are usually made" (10). Doping tests are often how women with high natural T are flagged, so understanding how natural T responds to competition is 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



Close-up of a baton before women's 4 x 400-meter relay final, 2012 Olympic games, London, UK.

sponse to competition. IAAF-IOC policy-makers suggest that the female-male overlap in T observed in the GH-2000 data may be an artifact of sampling within 2 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 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 the individual's sex (14, 15). Intense resistance exercise and short-duration exercise are associated with increase in T, whereas endurance exercise (especially 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

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 (17). The "correct" time to sample T depends on the purpose of the study, but the timing of blood draws seems unlikely to determine whether a study finds 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

¹Stanford University, Stanford, CA 94305, USA. ²Barnard College, New York, NY 10027, USA. E-mail: karkazis@stanford.edu

known as intersex (DSDI). In simple terms, some of the biological characteristics of women with DSDI would be classified as female and others 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

“What looks like a [scientific] controversy ... is ultimately a social and ethical one concerning how we understand and frame human diversity.”

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 are not 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 is 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 levels no different from nonathlete women—reinforce their values as outsiders and justify 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 is 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 that the problem is that women with naturally high T have unfair advantage, despite having acknowledged 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: 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 policy-makers 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 AND NOTES

1. IAAF, *IAAF Regulations Governing Eligibility of Females with Hyperandrogenism to Compete in Women’s Competitions* (IAAF, Monaco, 2011).
2. S. Bermon, M. Ritzen, A. L. Hirschberg, T. H. Murray, *Am. J. Bioeth.* **13**, 63 (2013).
3. IOC, *IOC Regulations on Female Hyperandrogenism* (Lausanne, Switzerland, 2014).
4. R. M. Jordan-Young, P. H. Sönksen, K. Karkazis, *BMJ* **348**, g2926 (2014).
5. V. B. Padmadeo, *Indian Express*, 12 September 2014.
6. J. L. Simpson et al., *JAMA* **284**, 1568 (2000).
7. J. L. Elsas et al., *Genet. Med.* **2**, 249 (2000).
8. IAAF, *IAAF Regulations: Explanatory Notes* (IAAF, Monaco, 2011), pp. 1–4.
9. J. Macur, *New York Times*, 24 June 2012, p. SP6.
10. M. L. Healy et al., *Clin. Endocrinol. (Oxf.)* **81**, 294 (2014).
11. S. Bermon et al., *J. Clin. Endocrinol. Metab.* **99**, 4328 (2014).
12. M. Ritzen et al., *Clin. Endocrinol. (Oxf.)* **82**, 307 (2015).
13. G. S. Ginsburg et al., *Clin. Chim. Acta* **305**, 131 (2001).
14. J. L. Vingren et al., *Sports Med.* **40**, 1037 (2010).
15. C. Enea et al., *Sports Med.* **41**, 1 (2011).
16. S. M. van Anders, N. V. Watson, *Psychoneuroendocrinology* **31**, 715 (2006).
17. C. J. Cook, B. T. Crewther, *Horm. Behav.* **61**, 17 (2012).
18. P. A. Lee et al., *Pediatrics* **118**, e488 (2006).
19. A. Ljungqvist, J. L. Simpson, *JAMA* **267**, 850 (1992).
20. Women’s Sports Foundation, *Participation of Intersex Athletes in Women’s Sports* (WSF, New York, 2011), pp. 1–5.

ACKNOWLEDGMENTS

Supported by NSF grants 1331115 and 1331123.

10.1126/science.aab1057

NEUROSCIENCE

Reading the mind to move the body

Decoding neural signals of intention and movement should guide the development of neural prosthetics

By J. Andrew Pruszynski¹
and Jörn Diedrichsen²

Imagine a world in which your smartphone can read your mind. Just at the moment that you decide to move your finger to delete a message, it is already gone. This sounds like science fiction, but for one human in California, this fantasy is becoming reality. On page 906 of this issue, Aflalo et al. (1) report the case of a tetraplegic individual (called “EGS”) who volunteered to have his brain implanted with two small silicon chips that allow researchers to read his intentions directly from his brain activity. The chips—initially developed at the University of Utah (2) and now commercially available and approved for human use by the U.S. Food and Drug Administration—consist of a matrix of 96 microscopic electrodes that can record the activity of about 100 nerve cells at the same time.

The main goal of the implantation procedure was to restore EGS’s ability to act in his environment. Paralyzed from the neck down, he currently relies on the help of others to perform almost all the daily actions that the vast majority of us take for granted. Using the signals from his brain and bypassing his damaged spinal cord, researchers hope to help him do these things again by allowing him to steer a robotic arm so that he can, for example, reach out, grasp a glass, and take a drink. Alternatively, the acquired signals can be used to control a cursor on a screen so that he can efficiently interact with a computer.

Previously, researchers have implanted chips into regions of the human brain that are closely related to the production of movements, such as the primary motor cortex (3, 4), with the aim of reanimating a limb or controlling a prosthetic. Aflalo et al. have taken a different approach. They have