Including Trans Women Athletes in Competitive Sport: Analyzing the Science, Law, and Principles and Policies of Fairness in Competition

Dr. Rachel McKinnon
Dr. Aryn Conrad

Forthcoming in *Philosophical Topics: Gendered Oppression and its Intersections.*
Bianka Takaoka and Kate Manne, eds..

Abstract:
In this paper, we examine the scientific, legal, and ethical foundations for inclusion of transgender women athletes in competitive sport, drawing on IOC principles and relevant Court of Arbitration for Sport decisions. We argue that the inclusion of trans athletes in competition commensurate with their legal gender is the most consistent position with these principles of fair and equitable sport. Biological restrictions, such as endogenous testosterone limits, are not consistent with IOC and CAS principles. We explore the implications for recognizing that endogenous testosterone values are a 'natural physical trait' and that excluding legally recognized women for high endogenous testosterone values constitutes discrimination on the basis of a natural physical trait. We suggest that the justificatory burden for such prima facie discrimination is unlikely to be met. Thus, in place of a limit on endogenous testosterone for women (whether cisgender, transgender, or intersex), we argue that 'legally recognized gender' is most fully in line with IOC and CAS principles.

1. The Legal and Organizational Structures of International Sport

Rare is the activity and associated body of law that ultimately falls under the jurisdiction of one and only one body the world over. However, most international sporting competitions and the associations and athletes that participate in them fit this singular description: legally, they are under the ultimate jurisdiction of the Court of Arbitration for Sport (CAS), seated in Lausanne, Switzerland.

All Olympic International Federations, as well as many non-Olympic International Sporting Bodies (for example, FIFA), fall under the jurisdiction of CAS. There are a few notable exceptions, such as many North American professional sports leagues--like Baseball, who do not subject themselves to CAS. Those bodies that do recognize CAS as the ultimate authority in the law of their sport thereby exclude their

events from the jurisdictions of other national, regional and international courts.[1] In the case of many countries, including the United States and Canada, this is done by statute. The International Olympic Committee (IOC) requires that federations be subject to CAS if they wish to field teams at the Olympics. Though CAS executes a number of functions, it is most important as the primary appellate body for the decisions of international sporting federations. There are no appeals from CAS's final decisions, effectively rendering CAS the Supreme Court of that sport.

As its name implies, CAS is not a purely adjudicative body, but a quasi-arbitral tribunal. Hence, there is some debate as to whether a common law style doctrine of *stare decisis* exists for CAS rather than a civil law style doctrine of *jurisprudence constante*, and therefore, whether existing precedent is binding or highly persuasive.[2] In common law jurisdictions, under *stare decisis*, existing precedent must be followed by lower courts and higher courts are unlikely to overturn established precedent except in extraordinary circumstances. Under *jurisprudence constante*, civil law courts follow past lines of decision "unless clear error is shown and injustice will arise from continuation of a particular rule of law."[3] Regardless, CAS decisions routinely cite previous CAS decisions, and they are thus at least highly persuasive to subsequent decisions. Therefore, studying previous CAS decisions gives indication of how the court will decide in the future in previous cases.

All eligibility disputes in sport fundamentally concern when discrimination against a particular athlete is justified--whether on the basis of gender, or some other physical characteristic. As in most areas of life, some forms of discrimination are permissible--as Rawls's famous example of the would-be surgeon with the shaky hands, and some are not--as, for example, discrimination on the basis of race in nearly every context.[4] Thus the question of trans inclusion in sport is a question of if and when discrimination on the basis of an athlete's status as transgender is justified. That is, it is a question of when and if trans athletes should be deemed eligible.

The case of Dutee Chand involved a similar question of when an intersex athlete with naturally high endogenous testosterone who identified as female should be

[1] There has been one unique exception where Canadian cyclist Kristen Worley took the UCI and Cycling Canada to the Ontario Human Rights Commission, which survived a challenge that only CAS had jurisdiction. https://www.sportsintegrityinitiative.com/worleys-case-to-proceed-in-human-rights-tribunal/ Last accessed March 3 2019.
[2] See, e.g., Bersagel 2012.
[3] Bersagel 190 quoting Black's 9th.
[4] Exceptions are rare, but would include such things as choosing a black actor over a white actor to play Martin Luther King, Jr. in a Hollywood film.

considered eligible to compete in track events, including those that ultimately determined eligibility for Olympic competition. CAS looked to the International Olympic Charter (hereinafter, "Charter") as well as the applicable regulations promulgated by International Association of Athletics Federation (IAAF), which was the international governing body of the sport recognized by the IOC as the primary governing body as the applicable body of substantive law to decide the case. They also considered the constitution and laws of Monaco.

We assume for this paper that any decision made by CAS on the eligibility of a trans athlete would similarly look to the Charter and the regulations promulgated by the IOC-recognized international governing body of the sport in question. Because of the scope of this paper, we will ignore such picky legalistic questions like how the IAAF's being seated in Monaco rather than Switzerland could affect such decisions.

Another central legal issue to such CAS cases is the concept of burden of proof. The legal concept of the burden of proof and its accompanying standards are often misunderstood and misapplied in popular parlance. Often, one hears that a particular individual who stands accused of a crime but who is not on trial has not been proven guilty beyond a reasonable doubt. However, such statements misunderstand the term "proof beyond a reasonable doubt," which belongs only in courtroom settings, primarily those involving criminal sanctions.

The burden of proof in common law courts refers to which party in an adversarial legal proceeding bears the responsibility, or burden, of making a showing of evidence to the finder of fact, which may be an adjudicator like a judge or a jury, that a given proposition is the case. In the American legal system, the standards range from the forgiving "probable cause" used to justify police searches without a warrant, to the stringent "beyond a reasonable doubt"required for the governmental deprivation of life or liberty in criminal cases with the fifty percent plus a feather "preponderance of the evidence" used notably in many civil cases and the more demanding "clear and convincing evidence," common in patent cases and euthanasia decisions.[5]

In theory, the party bearing the burden of proof may fail to make an adequate showing of evidence even if the other party provides nothing to the finder of fact. In practice, this rarely happens and is rarely likely to be successful. Different questions even within the same legal proceeding may require different standards of proof--for example, whether or not an expert is qualified in a legal proceeding may have one

[5] This is not an exhaustive list of either the standards or their applicability but is provided to illustrate the broad range of standards of evidence.

standard of proof but the verdict may require another. Burdens of proof also may shift within a proceeding from one party to another if certain requirements are met by the party initially bearing the burden. Civil law courts, including CAS, sometimes differ from common law courts in that the burden of proof may require both sides to make a showing of evidence.

2. Chand v AFI/IAAF: The Law and Ethics of Discriminating on the Basis of a Natural Physical Trait

Dutee Chand is an Indian-born athlete who has been competing in junior-level athletics competitions since 2007. (Chand Para. 7.) After complaints from other participants in events she participated in, in 2014 Chand was subjected to a number of gender verification tests by the Sports Authority of India (SAI) in her home country, including a hyperandrogenism test. (Chand Paras. 10-15.) She was subsequently notified that her "male hormone levels" were too high and she was therefore not eligible to compete in the upcoming World Junior Championships. This would affect her ability to qualify for the Commonwealth Games. (Chand Para. 16).

The SAI would later claim that the decision to exclude Chand was based on the IAAF regulations concerning hyperandrogenism, the IAAF Regulations Governing Eligibility of Females with Hyperandrogenism to Compete in Women's Competition (the "Hyperandrogenism Regulations"). (Chand Para. 20). Thereafter ensued a painful and sometimes public battle, wherein Chand was informed she would only be able to compete if she had surgery or pharmaceutical intervention to lower her endogenous testosterone levels, even when she thought that such treatments would be harmful to her health. Chand appealed the decision to exclude her to CAS. (Chand Paras. 15-31).

Chand challenged the Hyperandrogenism Regulations (HRs) as violating the principles against discrimination on the basis of sex and a natural physical characteristic found in the IAAF Constitution, the Charter, international human rights law, and the Monegasque Constitution in her appeal to CAS. (Chand Paras. 112-116). CAS determined that Chand bore the burden of proving, on the balance of probabilities, that the IAAF regulations were impermissibly discriminatory. However, if Chand could make out a case that the regulations were prima facie discriminatory, that is, discriminatory on their face, then the burden would shift to the IAAF to show that the regulations were justifiable necessary and proportionate. The IAAF stipulated to the HRs being *prima facie* discriminatory.

If the IAAF could prove the justification, then the burden would shift back to Chand to disprove the bases of justification. The Panel concluded that the standard of

proof for the IAAF once the burdened shifted needed to be stronger than simply the balance of probabilities. However, they also stated that the strong standard of "beyond a reasonable doubt" was too stringent, as was the intermediate standard of "to the comfortable satisfaction of the panel." (Similar to "clear and convincing evidence" in American jurisprudence.) They ultimately concluded that the balance of the probabilities was the correct standard for the IAAF, but that the IAAF would have to make an affirmative showing--or positively overcome the onus--to meet the burden. In other words, simply making a stronger showing than the other side was not enough--they had to bring the panel to believe, on the strength of their showing, that "the discrimination of a fundamental right, including the right to compete that is recognized in the Hyperandrogenism Regulations themselves," was justified. (Chand Paras. 443-447).

The Charter lays out the Fundamental Principles of Olympism. The Panel stated that Principles 2, 4 and 6 were those relevant to this decision. Principle 2 states that the "goal of Olympism is to place sport at the service of the harmonious development of humankind, with a view to promoting a peaceful society concerned with the preservation of human dignity."

Principle 4 declares that the "practice of sport is a human right." It further states that every individual must have the possibility of practicing sport, "without discrimination of any kind and in the Olympic spirit, which requires mutual understanding with a spirit of friendship, solidarity and fair play." While the Charter does not explicitly state that the practice of sport must be competitive sport, we may infer that this refers to competitive sport as the IOC is concerned only with competition.

Principle 6 states that each of these rights and freedoms must be secured without any discrimination on the basis of race, colour, sex, sexual orientation, language, religion, political or other opinion, national or social origin, property, birth or other status." (Chand 39). We note that while 'gender identity' is not mentioned, the list is explicitly open-ended, which allows 'gender identity' to be considered an additional relevant prohibition of discrimination. Moreover, many US jurisdictions have been recently deciding that 'gender identity' is included under protections against 'sex' discrimination.[6]

[6] For a large list of US case law, see https://transequality.org/federal-case-law-on-transgender-people-and-discrimination

The relevant portion of the IAAF Constitution is Article 3, points 2-4. These state that among the objectives of the IAAF are to promote Athletics and its ethical values as educational, life-affirming and life-enhancing, to encourage participation in the sport regardless of age, gender or race and to "strive to ensure that no gender, race, religious, political or other kind of unfair discrimination exists, continues to exist; or is allowed to develop in Athletics in any form, and that all may participate in Athletics regardless of their gender, race, religious or political views or any other irrelevant factor." (Chand 40).

The Panel discussed the HRs at length, including an exposition of the preface, which lays out the purposes of the HRs. (Chand Paras. 41-68). The Panel was careful to point out that these rules are eligibility rules that flow from the divide between male and female competition. (Chand Paras. 41-43).[7] Thus, the IAAF argued, and the Panel agreed, that the HRs are about distinguishing eligible female athletes from ineligible female athletes, not between female and male athletes.

Generalized references to Monegasque law pepper the decision, with rare instances of specificity. Chand pointed out that the laws of Monaco required the IAAF to take "all appropriate measures to eliminate discrimination against women." (Chand Para. 115) Also, the Panel briefly noted that, as to proportionality, "the detrimental impact of a measure must be proportionate, in that it must not exceed that which is reasonably required in the search of the justifiable aim." (Chand Para. 230)

Common to many human rights legal frameworks, such as the European Commission on Human Rights (ECHR) and the United Nations Universal Declaration of Human Rights, is a structure for determining whether prima facie discrimination can ultimately be justified. A discriminatory policy is permitted when it is in service of a worthy social goal, the the policy is necessary and effective at promoting that worthy social goal, and the benefit to society is proportional to the harm caused to the affected group or individuals.

For example, people have a right to freedom of movement. However, this right is overridden when we imprison people for murder. The policy of imprisoning murderers is in service of the worthy social goal of 'public safety.' Plausibly, imprisoning murderers is both necessary and effective at promoting public safety--particularly from the murderer. And the benefit to society is proportional to the harm experienced by the prisoner. However, a practice such as solitary confinement likely fails to be both

[7] Behrensen (2013) argues that policies such as the HRs are sex-verification policies by another name.

necessary and effective at promoting any worthy social goal, and the harm caused to the victim is not proportional.

The question for our purposes is whether excluding hyperandrogenic women, including trans women (and trans women who aren't hyperandrogenic) meets these criteria for justifying prima facie discrimination. The policy is in service of the worthy social goal of 'fairness in competition.' But the crux of the matter will be whether such policies are necessary and effective at promoting fairness (we argue that they do not), and whether the harm visited upon hyperandrogenic and trans women is proportional to any benefit to society (we argue that it is not).

In Chand, the Panel laid out four issues that needed to be decided in order to dispose of the case. (Chand Para. 32). They stated them thus:

ISSUE 1: Discrimination
Do the Hyperandrogenism Regulations discriminate impermissibly against certain female athletes on the basis of: (i) a natural physical characteristic; and/or (ii) sex? *Id.*

ISSUE 2: Scientific Basis for the Hyperandrogenism Regulations
Should the Hyperandrogenism Regulations be declared invalid on the basis that there is insufficient scientific evidence: (i) that endogenous testosterone improves athletic performance in female athletes; and/or (ii) that IO nmol/L is the scientifically correct threshold at which female athletes are in the "male range" of endogenous testosterone and therefore enjoy the benefits of male levels of androgens? *Id.*

ISSUE 3: Justification
Are the Hyperandrogenism Regulations disproportionate in the context of: (i) the fact they discriminate on the basis of a natural physical characteristic and/or sex; and/or (ii) the harm they cause to female athletes? *Id.*

ISSUE 4: Unauthorized Doping Sanction
Are the Hyperandrogenism Regulations invalid because they are a form of unauthorised anti-doping sanction in violation of Articles 4.3.3, 10 and 23.2.2 of the World Anti-Doping Agency Code? *Id.*

For the sake of logical flow, we shall discuss the issues out of order, beginning with four, then two and then one and three, which merge together. Note the differing

philosophical calls to each question: Issue four is basically a legal question, while issue three is more of a scientific/epistemic question about the warrant for the regulations. Finally, one and three together form a normative question that touches on issues of law, policy and ethics.

Issue four is not particularly important for the purposes of this paper. Moreover, the Panel disposed of it quite quickly. Essentially, the question asks whether an athlete with a high level of endogenous testosterone, who has not in any way done anything affirmative to have this level, is thereby punished by the HAR in the same way that an athlete who had doped with exogenous testosterone as an unauthorized application of the World Anti-Doping Agency ("WADA") Code. The Panel concluded that the HAR was not a sanction under the WADA Code because it was an eligibility rule rather than a sanction created to impose a punishment on an athlete. (Chand Paras. 539-546).

While analyzing issue two, the Panel noted the common ground between the two parties: Both agreed that Lean Body Mass ("LBM") contributed to the differences in strength and athletic performance between males and females. However, the two parties agreed about the effect that testosterone had on LBM, as well as whether or not endogenous testosterone had the same effect on LBM. (Chand Para. 454). For both questions, the onus of proof was on the athlete. After reviewing the expert testimony put forward by both parties, the Panel concluded that while Chand had suggested many testable hypotheses that could eventually establish her position, she had not "established that there is no relationship between testosterone and athletic performance." (Chand Paras. 452-472).

In the case of the effects of endogenous testosterone versus exogenous testosterone, the onus of proof was very important as the Panel ultimately concluded that "no single study has established, to an appropriate level of certainty, a scientific basis to come to a definitive conclusion one way or the other." (Chand Para. 474). Because the onus was on Chand, she was not able to discharge her burden on this matter. (Chand Para. 488). Had the burden been reversed, the IAAF would have lost. Ultimately, because she did not meet her onus on either point, and because the IAAF did put forth sufficient evidence to convince the Panel on the balance of the probabilities to support their case, the Panel concluded that "there is a scientific basis in the use of testosterone as a marker for the purposes of the Hyperandrogenism Regulations." (Chand Paras. 489-499).

As to issue one, the Panel quickly determined that because the HRs required only females and not males to undergo testing, they were discriminatory on their face.

(Chand Para. 448). Because the HAR were prima facie discriminatory, the onus shifted to the IAAF to show that such discrimination was "necessary, reasonable and proportionate for the purposes of establishing a level playing field for female athletes." (Chand Para. 450). The issue of justification just is issue three--whether this discrimination on the basis of gender or a physical trait is justified.

One of the issues that the Panel pointed to in analyzing issue three was the fact that sporting events, with a few exceptions, are currently divided into male and female categories, and yet, sex in humans occurs along a continuum rather than in a binary. (Chand 504). According to the Panel, the divide between male and female athletes is justified by the degree of competitive advantage that one gender has over the other. However, whether or not a person is male or female is, as far as the Panel is concerned, a legal issue. The women affected by the HRs are therefore female, and only eligible to compete as females. Thus, the divide policed by the HRs is a divide between one set of female athletes with a particular physical characteristic and those that lack it, namely, a particular level of androgens and androgen sensitivity. Enforcing the HRs would thus create a group of females who were unable to compete at all. However, both the Charter and the HRs themselves either imply or outright state a right to compete.

For the HRs to be justified, as far as the Panel was concerned, the advantage conferred on the class of female athletes to be excluded would have to be roughly on the same order as that of the advantage conferred on males over females, which is claimed to be 10-12%. While the IAAF had an expert *estimate* that endogenous testosterone could confer some advantage to female athletes (about 3%, Bermon suggested without admissible evidence), there was no evidence that it was anywhere near the same level as that which justifies the male/female split. Indeed, what evidence was adduced indicated the advantage was far lower. This, the Panel concluded, was not sufficient to justify denying female athletes with levels of endogenous testosterone above those allowed by the HRs their right to compete. (Chand Paras. 500-539).

However, what should we make of the CAS decision and reasoning? Endogenous testosterone is the only natural physical characteristic for which women/female athletes are deemed ineligible. There are no such characteristics for which men/male athletes are deemed ineligible. Moreover, there are many natural physical characteristics, we will argue, that confer larger advantages than those being attributed to endogenous testosterone, that are permitted. In what follows, we argue that this places an insurmountable burden on attempts to justify any limit on endogenous testosterone for (cis, trans, or intersex) women athletes. Simply put, such policies are neither necessary nor effective at promoting fairness in sport, and the harm to athletes deemed ineligible

by such policies is not proportional to the supposed social gain. But first, we need to have a firm grasp on the concept of 'fairness' in competition.

3. Fairness in Competition

The concepts of 'justice' and 'fairness' are relatively highly theorized in moral philosophy. However, it's surprising how under-theorized the concept of 'fairness in competition' is, particularly as it relates to the question of whether it's fair for trans athletes to compete in their post-transition sex/gender in single-gender sports. One common pre-theoretic account of fairness in sport is something along the lines of, "Level playing field." So let's start there.

What does one mean by a sporting competition having a level playing field? Its lexical origin is unclear, but it has an attractive literal meaning in sports such as baseball, football, soccer, volleyball, basketball, billiards, and any sport where the field of play itself should be level. There shouldn't be bumps or features of the field of play itself that, in being unlevel, provides an advantage to one team.

Of course, there are many sports where the field itself isn't level in a literal sense. Some sports don't have a 'field.' So in order for this definition of fairness to generalize, we must mean something like, "The conditions of play must be sufficiently equal for all competitors." That is, the conditions of play themselves should not confer a competitive advantage to any competitors.

One way to explicate 'level playing field' is to say that all competitors also *play by the same rules*. So 'level playing field' means something like, "The rules of competition apply to each competitor equally, and the conditions of play must be sufficiently equal for all competitors."[8]

However, this is inadequate for two reasons. First, it doesn't give us a means with which to judge whether the rules or conditions of play are themselves fair. For example, some rules are discriminatory; e.g., racial segregation rules in baseball during the pre-civil rights era of the USA. We need some notion of fairness in order to judge whether rules themselves are fair.

Second, this definition fails as a sufficiency criterion. If there remains a question whether it's fair for trans women to compete in women-only categories, then this definition cannot account for this: if the conditions of play are the same for all

[8] A benefit of this definition is that it also applies to competitive contexts outside of sport such as finance and economics, which is another context where people use 'level playing field.'

competitors, and the trans women are following the rules, then it's fair for them to compete. But this will not satisfy those who wish to say that it is not fair for trans women to compete in women-only categories. So we would need some other notion of fairness. The question here is whether trans women have an unfair competitive advantage over their cisgender counterparts.

So perhaps "Level playing field" requires adding something like, "And no competitor has a competitive advantage." We can't include "unfair" in this definition on pain of circularity. But as Coggon et al (2008, p. 6, emphasis in original) note: "What is clear and must be conceded by all, though, is that while the playing field should be level, the athletes must not be equal. If *all* were equal, there would be little to derive from competition, for everything would result in a draw or a win that was only the result of chance." They're right: one-on-one competitions would become coin-flips, a six-way competition like the track cycling keirin event would become rolling a six-sided die, and so on. Clearly "no competitive advantage" cannot be a requirement of fairness in sport.

Instead, what we tend to mean is that a competitor shouldn't have "too much" of a competitive advantage over their fellow competitors. But this then builds in a degree of arbitrariness to our definition of fairness: how much of a competitive advantage is too much? For now, let's set this aside.

One attractive, common approach to issues of fairness in competition conceives of it and sport as a fundamentally cooperative project. This puts us in the realm of contractarian approaches to ethics. People come together and agree upon rules, agree to abide by those rules, and off they go. Suits (2005/1978, p. 55) offers a pithy definition of games: "playing a game is the voluntary effort to overcome unnecessary obstacles." There's no particular reason why I should want to get this rubber air-filled globe to drop through a metal ring placed 10 feet into the air, let alone restrict the ways in which I'm allowed to do that--such as not using a ladder. But that's how the game of basketball works. Players voluntarily agree to constitutive rules that define the game, but make achieving the lusory goal more difficult than it has to be.

Loland (2002) adopts what Loland and McNamee (2000) refer to as an "eclectic" moral framework. He borrows from many disparate, and often competing, moral frameworks including Aristotelian virtue ethics, deontology, utilitarianism, and contractarianism (i.e., Rawls), among others. However, his notion of *fairness* is distinctly Rawlsian in flavor. He writes:

> I subscribe to the idea of moral norms as based on some kind of unforced, informed agreement that gains authority by the fairness of the procedures by which they arise. However, my ambition is somewhat different from that of Rawls. ... I have presented sport competitions as social practices that are realized through some kind of shared interpretation of the rules among participants, what I called an ethos. ... Consequently, the understanding of the contract procedure I have proposed is not of rational decision-makers seeking full consensus behind a veil of ignorance, but, rather, of persons searching for a set of norms that can be readily accepted, or at least that cannot reasonably be rejected, as a basis for rational agreement (Loland, 2002, p. 31).

He thus proposes the following meta-norm for choosing norms of fair play: "Choose norms that no one can reasonably reject as a basis for unforced, informed general agreement" (Loland, 2002, p. 31).

Following Loland's lead, Coggon et al. (2008) also adopt a contractarian approach to fairness. They write:

> A convincing suggestion for properly understanding fairness in sports asks us to consider it as being a moral requirement to follow a common interpretation of a set of rules....This ideal may be said to rest on a notion of reciprocity: an athlete may benefit from being allowed to participate, and in exchange surrenders liberty to the extent that his or her fellow athletes agree. Fairness so understood has as its core value equal respect for persons, as distinct from equal treatment of persons (Coggon et al, 2008, p. 6).

They continue:

> We may, therefore, conceive of taking part in sport as the consensual participation in a game with agreed, binding rules that are known from the beginning. This means that it can be fair to treat participants differently, so long as it has been agreed to by all the players: their agreement inferred by their willingness to participate according to the

> rules.[9] It is unfair, however, to take some advantage outside of the agreed rules without the consent of other athletes (Coggon et al, 2008, p. 7).

The common thread here is a broadly Rawlsian contractarian account of fairness in competition. Both note that fairness should involve rules that competitors cannot reasonably reject, and both involve some notion of "consensus" in competitors coming to agree upon fair rules of play.[10] Unfortunately, in adopting a broadly Rawlsian framework, they jettison, largely without comment or argument, the most important feature of Rawls's view of justice as fairness: the mutual agreement on rules of fairness must be done behind the *veil of ignorance*.

For Rawls, it's no good to ask what people would reasonably agree to as principles of fairness if we allow their personal biases to influence their answers. He writes:

> One reason why the original position must abstract from the contingencies--the particular features and circumstances of persons--within the basic structure is that the conditions for a fair agreement between free and equal persons on the first principles of justice for that structure must eliminate the bargaining advantages that inevitably arise over time within any society as a result of the cumulative social and historical tendencies" (Rawls, 2001, p. 16).

For example, it's no secret that white people accumulated unfair bargaining advantages in, among many other things, setting up the rules of racially segregated baseball in the US. If we want to answer whether it's fair, for example, for a particular group to compete with another group, we shouldn't simply poll the participants. Their answers will likely be biased by their knowledge of who they are in the society. Our conceptions of 'fairness' tend to unfairly favor people who are like us. Our moral intuitions are often

---

[9] We note that it's a little odd for Loland and Coggon et al to draw on Aristotelian concepts of justice, such as treating unequals unequally, since Aristotle thought democracy was a degenerate version of government, and particularly opposed personal liberty approaches to justice. We also note that Aristotle's concepts of justice, for him, supported slavery and excluding women from politics. Such a conception of justice is incompatible with the Rawlsian framework they (and we) adopt. When we conceive of 'treating unequals as unequal,' the crux is which inequalities are relevant. We no longer think that language or skin color are relevant bases on which to treat people unequally, but Aristotle sure did: he thought it was rational and just to treat 'barbarians' as less than fully human, thus justifying a system of slavery. Aristotle also thought that men cannot be properly friends with women, since women are not equal to--and are merely degenerate forms of--men. This is not a suitable basis for modern conceptions of fairness.

[10] C.f., Loland and McNamee (2000), p. 69 and p. 75.

predictably biased in this way; or, at least, we're inclined to agree to biased policies as 'fair' if they are biased in our favor. Our intuitions are also subject to framing effects.[11]

Instead, Rawls uses the thought experiment of the "original position" to ask what we would reasonably agree to if we didn't know anything about our social identities. The veil of ignorance asks us to imagine that we don't know our race, sex, sexual orientation, gender identity, disability status, socioeconomic status, geographic location, and so on. Only then can the results of our deliberations on what constitute fair rules be, well, fair.

So when Loland (and McNamee), Coggon et al, and others adopt this "fairness by agreement" approach, not including the veil of ignorance is a serious mistake. For Rawls, the veil of ignorance is *everything*: it's doing the heavy lifting in his approach to justice as fairness. As it relates to our project, while one may reasonably agree to trans-exclusionary sports policies (including endogenous testosterone limits for women) if one knows that one is cisgender, one will *not* reasonably agree to such policies of one doesn't know one's gender identity: if it's possible that one would be asked to agree to a policy that could exclude oneself from sport, one wouldn't reasonably agree to such a policy. So while there are legitimate critiques of Rawls's theory of justice, we think that something broadly Rawlsian is a good account of fairness in competition. But such a conception *must* incorporate the veil of ignorance.

The second important aspect of this broadly Rawlsian "morals by agreement"[12] approach to fairness is the role of reason and rationality. Rawls in particular adopts a decision/game-theoretic approach. He argues that no reasonable person, in the original position behind the veil of ignorance, would gamble with basic human rights. He thinks that rational, self-interested people will agree to two fundamental principles of justice: the Liberty Principle and the Difference Principle.[13] The Liberty Principle is the most relevant for our purposes, so we focus only on it. It reads:

> Each person has an equal right to a fully adequate scheme of equal basic liberties which is compatible with a similar scheme of liberties for all.

[11] For a small set of examples, see Petrinovich and O'Neill (1996), Demaree-Cotton (2014), and Andow (2018).
[12] See also Gauthier (1986).
[13] It's certainly remarkable that Rawls argues that everyone would agree to these two and only these two principles of justice. This is a common site of criticism for Rawls's view. However, we think that his *approach* to fairness is a good one.

Crucially, he thinks that rational, self-interested people would adopt maximin reasoning. Maximin reasoning is when we choose options that have the best, worst-case outcomes: we are trying to maximize the minimum we will come away with. In Rawls's original position, we're trying to maximize the situation for the worst-off members of society. We don't want to risk choosing policies that could seriously disadvantage a group of people since, for all we know, we *are* one of those people when the veil of ignorance is lifted.

Applying this to choosing fundamental human rights, Rawls thus thinks that rational, self-interested people would agree to as many as possible. This would likely include such things as a right to meaningful work; adequate food, water, and shelter; play; marriage equality; education; and so on. As noted in Section 2, the IOC has included participation in (competitive) sport as a fundamental human right. Applying a Rawlsian framework, would a reasonable person behind the veil of ignorance agree to a policy that excludes hyperandrogenic (cis, trans, and intersex) women? We suggest not. But it might depend on the policy. We think it's clear that no reasonable person would agree to a blanket ban on trans athletes, which was often the state of sport before the November 2003 IOC Stockholm Consensus on transgender athletes.[14]

> The 2003 policy reads:
>
> The group recommends that individuals undergoing sex reassignment from male to female after puberty (and the converse) be eligible for participation in female or male competitions, respectively, under the following conditions:
>
> - Surgical anatomical changes have been completed, including external genitalia changes and gonadectomy
> - Legal recognition of their assigned sex has been conferred by the appropriate official authorities
> - Hormonal therapy appropriate for the assigned sex has been administered in a verifiable manner and for a sufficient length of time to minimise gender-related advantages in sport competitions.
>
> In the opinion of the group, eligibility should begin no sooner than two years after gonadectomy

[14] https://bit.ly/2RbsZo3. Justification given here: https://bit.ly/2Vx3X1h.

It's also unlikely that a reasonable person would agree to the 2003 policy, since it required both genital surgery and a two-year waiting period. This policy was both without adequate scientific and medical evidence, and it raises issues of bodily autonomy, consent, and socioeconomic inequalities of genital surgery. Not every trans person wants it, nor can most afford it. In most jurisdictions, trans people must pay out of pocket, and the costs are often prohibitive (usually in the $15,000 USD+ range). And for athletes for whom sport is the only means out of poverty, they may feel coerced into surgery or treatment they may otherwise not want.[15] Moreover, genitals do not play sports: genitals are absolutely irrelevant to sport performance.

Perhaps a reasonable person could agree to the 2015 IOC Consensus policy on transgender athletes.[16] It reads:

> In this spirit, the IOC Consensus Meeting agreed the following guidelines to be taken into account by sports organisations when determining eligibility to compete in male and female competition:
>
> 1. Those who transition from female to male are eligible to compete in the male category without restriction.
> 2. Those who transition from male to female are eligible to compete in the female category under the following conditions:
>   - 2.1. The athlete has declared that her gender identity is female. The declaration cannot be changed, for sporting purposes, for a minimum of four years.
>   - 2.2. The athlete must demonstrate that her total testosterone level in serum has been below 10 nmol/L for at least 12 months prior to her first competition (with the requirement for any longer period to be based on a confidential case-by-case evaluation, considering whether or not 12 months is a sufficient length of time to minimize any advantage in women's competition).
>   - 2.3. The athlete's total testosterone level in serum must remain below 10nmol/L throughout the period of desired eligibility to compete in the female category.
>   - 2.4. Compliance with these conditions may be monitored by testing. In the event of non-compliance, the athlete's eligibility for female competition will be suspended for 12 months.

[15] Chand testified that she was not complaining of any health issues based on her, then, undiagnosed intersex condition. She also testified that she believed any treatment she would be required to undertake as both unnecessary and harmful for her health.
[16] https://bit.ly/29IcVCv

We doubt that a rational person, behind the veil of ignorance, would agree to this policy, since it still requires medical interventions that an athlete may not desire--or that an athlete may not have reasonable access to, particularly in privatized medical systems like the US. But it's hard to say conclusively. This policy removed the surgery requirement and the subsequent two year waiting period, and replaced them with a one year testosterone suppression policy. Since we will argue that this policy fails to meet the justificatory burden for *prima facie* discrimination (thus it is not justified *ultima facie*), we suggest that reasonable people behind the veil of ignorance would also not agree to this policy. But we at least grant that this case is less clear than the others.

To conclude, then, we adopt a broadly Rawlsian approach to fairness, one that incorporates the veil of ignorance. It is a mistake, then, to ask what the current pool of athletes would agree to--by consensus or majority vote--since this opens up opportunities for the tyranny of the majority. These decisions aren't made from behind the veil of ignorance. So if we are asking whether we should adopt more trans-inclusive policies, what current athletes think is irrelevant to the ethical question. So let's turn to whether policies such as the 2015 IOC transgender policy can satisfy the justificatory requirements for *prima facie* discrimination on the basis of a natural physical trait.

4. Transgender Athletes, Endogenous Testosterone, and Performance "Advantages"

As noted in Section 2, the primary reason given for having a transgender (woman) athlete policy at all is that elite cis men, on average, tend to have a ~8-12% performance advantage over (cis) women, with the strong implication that such advantages are unfair.[17] We will make two primary claims in this section: endogenous testosterone limits on women are not necessary for promoting fairness in sport, and they are not effective at promoting fairness in sport. Therefore, we argue, there should be no testosterone limit on women (just as there isn't for men).

As noted in the Chand decision, only women have an endogenous testosterone limit for eligibility. The CAS panel was convinced that endogenous testosterone is a primary cause of the performance gap between men and women. At the outset, though, we should note that this approach already treats trans women as 'really' male/men, who

[17] While it's certainly undeniable that, at present, there is a performance gap between the best men and best women, it's not at all clear that that should be primarily attributed to differences in pubertal circulating testosterone (as Handelsman et al 2018 argue), and the gap is consistently closing over time. It's also unclear how much of that gap is due to sociological factors such as how misogynistic societies discourage athletic development in girls and women. For an early discussion of this, see Young (1980).

only may become women after sufficient medical intervention.[18] Why else marshall the claim that (cis) men are, on average, 8-12% stronger, faster, etc. than (cis) women?

One could argue that the performance data on men vs women should be irrelevant for the discussion of trans and intersex women since they are *women*. And to reiterate, the CAS (and by extension the IOC) makes no distinction between sex and gender: trans and intersex women are women *and* female. So physiological data on (cisgender) males should be irrelevant to the evaluation of performance advantages, if there are any, that trans or intersex women have. In a sense, this boils down to the very simple question: are trans women *women* (and *female*), or not?

In Chand 2015, CAS reiterated that sports organizations are no longer in the business of sex-verification testing. Instead, an athlete's sex/gender is whatever their legally recognized sex/gender is in their country of citizenship. Trans women can be so recognized in many countries, and particularly important Olympic countries such as Canada, USA, Australia, UK, and so on. So if trans women are female, we ask, why would "male" physiological data be relevant to the question of fairness?

We know this won't be convincing. But it is an important question to confront. While we will consider some physiological data on men vs women, we will focus on where the AFI/IAAF focused in Chand 2015: endogenous testosterone. What should matter, then, is only data comparing women to other women: comparing men to women shouldn't be relevant. Statistics like 'men are on average x compared to women' are not as relevant as comparing women to other women. So here is the first question: is a hyperandrogenism policy (including the 2015 IOC transgender policy) necessary for promoting fairness in sport? Let's look at the available data.

It may be surprising that there is relatively little empirical data on the relationship between *endogenous* testosterone and athletic performance.[19] Endogenous testosterone is that which is produced by one's own body. Exogenous testosterone is added to one's body through an external source. When people think of testosterone and sport performance, they're typically thinking of *exogenous* testosterone, which is banned as doping for its well-documented significant performance enhancing effects. We *do* have massive amounts of data on the relationship between exogenous testosterone and performance: we know that it provides a large performance advantage. Unfortunately,

[18] Those who oppose trans women athletes often deploy the language of these athletes being "born a man." People are born babies. This language perpetuates a damaging myth that trans women are all big and strong.
[19] See Sonksen et al (2018) for a review of all available literature.

people make a fallacious inference that having high (endogenous + exogenous) testosterone confers a competitive advantage, so having high endogenous only testosterone does too. But this is something that must be borne out by data, not relied upon as an assumption.[20]

In short, all available scientific evidence suggests that there *is no relationship between endogenous testosterone and sport performance*. It will take the rest of this section to substantiate this. There is also no available scientific evidence that post-transition trans women have an unfair competitive advantage. Instead, what little research we have (e.g., Harper 2015, Harper et al 2018) indicates that post-transition trans women have no competitive advantage over cis women. Crucially, trans women are not cis men, so comparing cis men's performances to cis women's is irrelevant to the question at hand.

Gooren and Brunk (2004) note in their small sample size (19 trans women, 17 trans men), there is large overlap in mean muscle area between pre-transition trans women and pre-transition trans men.[21] After 12 months of hormone therapy (testosterone suppression for trans women, testosterone supplementation for trans men), there was complete overlap between mean muscle area. As Jones et al (2008) note, many studies have found that force is linearly related to muscle cross-sectional area. Muscle fibers are not sexed: on average, similarly trained men and women with the same muscle cross-sectional area will produce the same force. What accounts for men's average greater strength than women is that they have a higher average lean body mass (bone+muscle). It's often a myth that men's muscles are stronger than women: they're not, *ceteris paribus*, if they're of the same cross-sectional area.[22]

Continuing with Gooren and Brunk (2004, p. 428), they write: "Androgen deprivation in [trans women] induces a loss of muscle area, but mean muscle area remained larger in reassigned [trans women] than in untreated [trans men], though with an almost complete overlap. This overlap between [trans women] and [trans men] existed even before cross-sex hormones, although not as extensively as after treatment." The key here is that there is near complete overlap between *untreated* trans people. We don't

[20] Interestingly, Bermon et al (2018) discuss Finkelstein et al (2013) who found that testosterone suppression in men, supplemented by an equalizing exogenous testosterone dose did not provide subjects any performance improvements over their pre-intervention state. This may be seen as good evidence that the body treats endogenous and exogenous testosterone sufficiently the same.
[21] This isn't surprising: there is wide overlap between physiological features of cisgender men and women: we only arrive at a non-overlapping bimodal distribution by ignoring the overlapping cases (which is what Handelsman 2018 do). See Sapienza et al (2009) and Stanton (2011). See Figure 1 below.
[22] See also Kanehisa, Ikegawa, and Fukunaga (1994).

want to paint pre-transition people as 'cis until proven otherwise,' but this isn't a surprising outcome.

Pick a secondary sex characteristic like height, bone density, lean body mass, mean muscle cross-sectional area, and so on, and there is almost *complete overlap* between men and women. A lack of overlap only happens at one side of the extremes: the tallest person is a man, the strongest is a man, and so on. But there are plenty of men below the average 'women,' and plenty of women above the average 'man.' Moreover, and this is key, *the differences within a gender are much larger than the average differences between genders*. The difference between the weakest and strongest woman, the shortest and tallest man, are much larger than the differences between the average woman and man.

The same is true for endogenous testosterone concentrations in 'male' and 'female' people. Even in samples that are too small to show a significant number of hyperandrogenic women, the amount of overlap between the groups is complete.

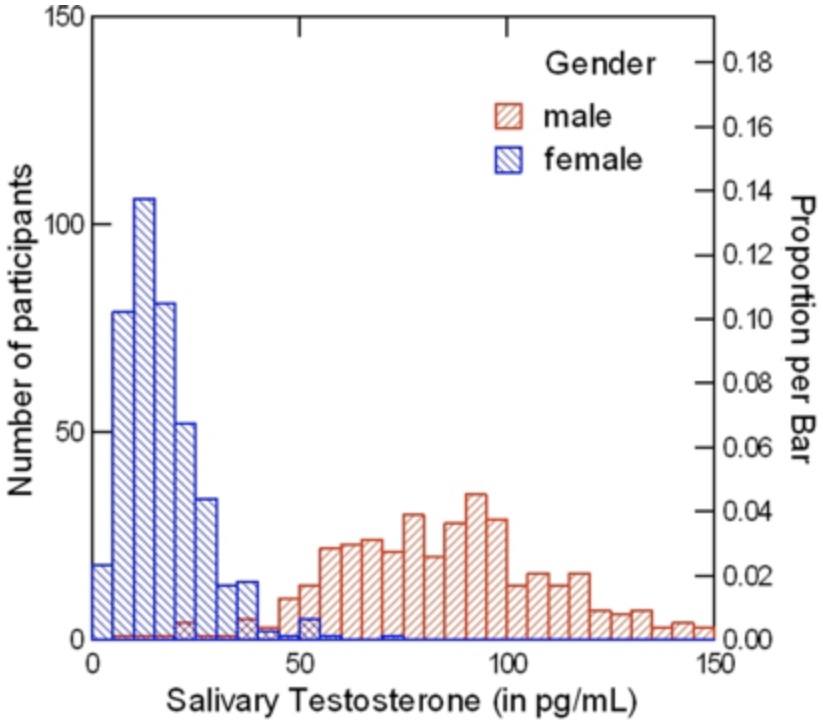

Figure 1. Salivary testosterone distribution in men and women. Sapienza et al (2009).

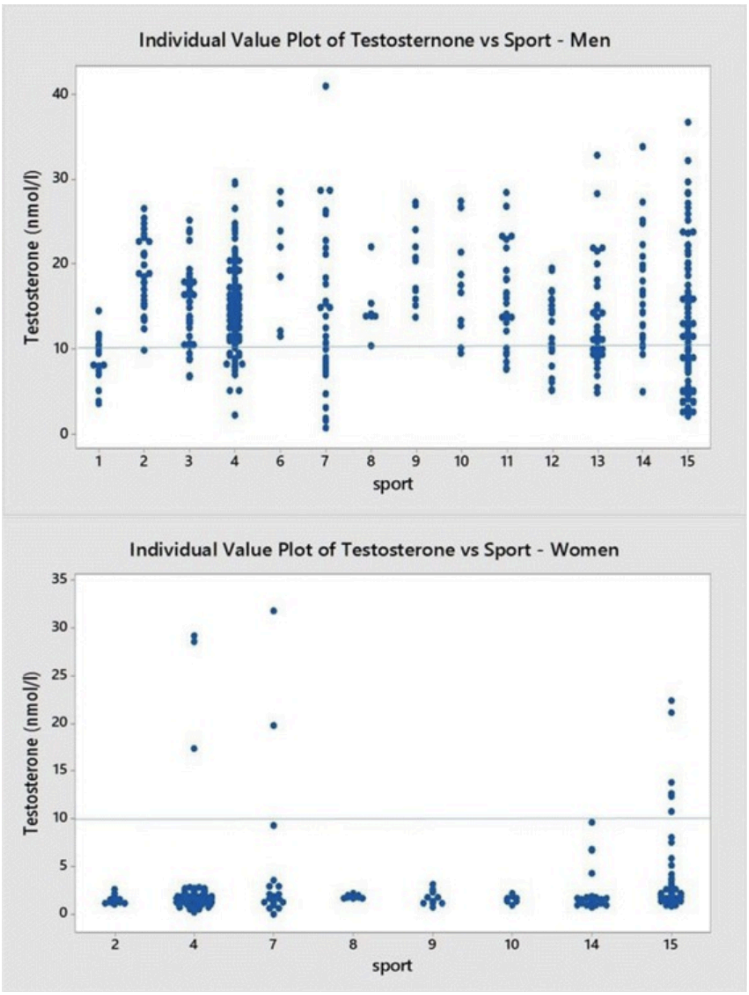

Figure 2. Serum testosterone in 445 elite male athletes and 231 elite female athletes. Sonksen et al (2018). Note how many elite cis men athletes are already naturally below the 10nmol/L limit.[23]

[23] The x axis numbers refer to the following sports: (1) powerlifting, (2) basketball, (3) football/soccer, (4) swimming, (5) marathon, (6) canoeing, (7) rowing, (8) cross country skiing, (9) alpine skiing, (10) weight lifting, (11) judo, (12) bandy, (13) hockey, (14) handball, and (15) track and field.

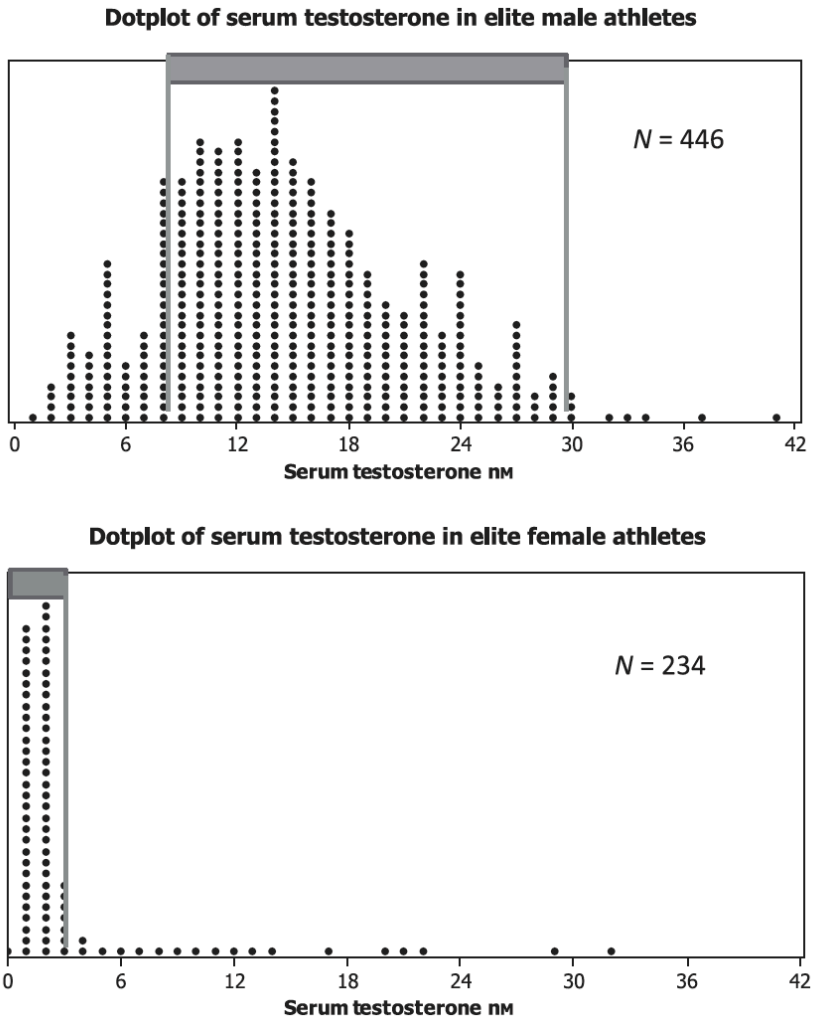

Figure 3. Serum testosterone values in elite male and female athletes. Healy et al (2014).

Note that the male distribution completely covers the entire female range, right down to the lower limit, while the upper limit of the female range does not extend to the upper limit of the male range. This is referred to as "complete overlap" of the male testosterone range with the female range. This is true for every natural physical characteristic we wish to graph: height, lean body mass, power, etc.. Moreover, the difference between the lowest and highest testosterone value for males is much larger than the average difference between males and females. For any given natural physical characteristic, while the geometric means for 'male' and 'female' are significantly different, difference *within* a given sex is much larger.

We permit tall women to compete with large competitive advantages against short women in sports that heavily select for being tall such as volleyball, basketball,

rowing, and so on. We permit small women to compete with large competitive advantages in sports that select against height such as gymnastics. And we call such competition 'fair,' even though height is a natural physical characteristic that can confer large competitive advantages, much larger than the 8-12% being attributed to testosterone.

It's important not to lose sight of this when thinking about physiological data of trans women and asking if there's a competitive advantage for being a trans woman (compared to a cis woman). Endogenous testosterone is the *only* natural physical characteristic for which some women can be deemed ineligible for competition. One of our arguments is that it cannot be necessary to limit endogenous testosterone in women if we permit larger competitive advantages due to other natural physical characteristics such as height or genetic factors.

As we noted in Section 2, the CAS decision focused almost entirely on the connection between endogenous testosterone and performance. So let's analyze the data. In Chand 2015, the IAAF's key witness, Stephane Bermon, could only offer an educated guess that hyperandrogenic women have an approximately 3% performance advantage. We find this a striking admission: the justification for the HRs rests entirely on the claim that higher endogenous testosterone confers an unfair competitive advantage. But *there was no empirical data* to substantiate this. As we note in Section 1, the burden of proof is on the sports organization who seeks to discriminate against the athlete. Discriminatory policies cannot be created prior to adequate justificatory data. Once this admission was made, in our opinion, this is where Chand won her case.

As Chand 2015 was an interim decision, the IAAF was given two years to supply sufficient data to justify the HRs. The results of this study are published in Bermon and Garnier (2017). They argue that their study substantiates the relationship between endogenous testosterone and performance, and substantiates it to a degree that justifies the HRs. In fact, Handelsman, Hirshberg, and Bermon (2018) go further and argue that the data supports an even more restrictive endogenous testosterone limit of 5nmol/L. So let's examine their findings.

Bermon and Garnier (2017) collected endogenous testosterone data on an impressive number of athletes from the 2011 and 2013 IAAF World Championships, 1332 women and 795 men. They broke down the data sets by running and field events: 100m, 100m hurdles, 200m, and so on up to the marathon for running events; and discus, hammer throw, javelin, high jump, and so on for the field events. They compared the bottom tertile (thirds) endogenous testosterone group's performance against the top

tertile group's performance (already a red flag: typically quintiles are used--tertiles are largely unheard of in such analyses).

Table 3 Comparison of the lowest and highest fT tertiles and their associated athletic performances in female running events

Event	n	All athletes	Lowest fT tertile	Highest fT tertile
100 m				
Time s	112	11.88 (0.88)	11.78 (0.83)	12.09 (0.97)
fT pmol/L		(1.4–32.8)	(1.4–8.2)	(12.1–32.8)
T nmol/L		0.90 (0.52)	0.53 (0.18)	1.32 (0.63)
100 m H				
Time s	73	13.15 (0.48)	13.02 (0.46)	13.24 (0.54)
fT pmol/L		(1.0–29.6)	(1.0–6.6)	(9.9–29.6)
T nmol/L		0.75 (0.33)	0.52 (0.20)	0.92 (0.34)
200 m				
Time s	71	23.43 (0.90)	23.28 (0.74)	23.65 (1.11)
fT pmol/L		(1.0–28.2)	(1.0–6.3)	(11.1–28.2)
T nmol/L		0.78 (0.48)	0.47 (0.15)	1.11 (0.62)
400 m				
Time s	67	52.23 (2.56)	52.60 (3.30)	51.16 (1.06)*
fT pmol/L		(1.2–124.4)	(1.2–6.5)	(13.0–124.4)
T nmol/L		2.82 (7.41)	0.45 (0.21)	7.08 (11.64)
400 m H				
Time s	67	56.34 (2.65)	57.38 (3.53)	55.78 (1.60)*
fT pmol/L		(1.4–174.5)	(1.4–7.9)	(12.4–174.5)
T nmol/L		1.03 (1.13)	0.54 (0.26)	1.68 (1.63)
800 m				
Time s	64	121.80 (5.42)	122.68 (3.71)	120.50 (2.90)*
fT pmol/L		(1.1–469.3)	(1.1–7.0)	(12.2–469.3)
T nmol/L		1.47 (4.03)	0.40 (0.14)	3.26 (6.60)
1500 m				
Time s	66	250.16 (6.42)	250.00 (3.23)	250.65 (7.10)
fT pmol/L		(0.6–48.7)	(0.6–7.0)	(10.8–48.7)
T nmol/L		0.79 (0.49)	0.50 (0.27)	1.12 (0.62)
3000 m SC				
Time s	56	581.61 (17.39)	579.81 (17.74)	581.03 (18.70)
fT pmol/L		(2.0–43.2)	(2.0–4.9)	(8.0–43.2)
T nmol/L		0.73 (0.54)	0.39 (0.15)	1.20 (0.66)
5000 m				
Time s	40	932.67 (39.73)	932.80 (32.4)	935.80 (48.30)
fT pmol/L		(2.8–51.8)	(2.8–6.1)	(8.0–51.8)
T nmol/L		0.73 (0.59)	0.41 (0.14)	1.07 (0.84)
10 000 m				
Time s	33	1912.6 (55.6)	1903.6 (61.8)	1913.8 (50.6)
fT pmol/L		(2.7–26.4)	(2.7–4.5)	(9.0–26.4)
T nmol/L		0.59 (0.27)	0.33 (0.13)	0.85 (0.21)
Marathon				
Time s	92	9726.6 (790.9)	9921.3 (1018.9)	9620.1 (736.5)
fT pmol/L		(2.1–363.6)	(2.1–5.6)	(9.0–363.6)
T nmol/L		0.97 (2.55)	0.41 (0.14)	1.85 (4.23)

Data are presented as mean (SD) or (min–max).
Different from the lowest fT tertile: *p<0.05.
fT, free testosterone; H, hurdles; SC, Steeple Chase; T, testosterone.

Table 4 Comparison of the lowest and highest fT tertiles and their associated athletic performances in female non-running events

Event	n	All athletes	Lowest fT tertile	Highest fT tertile
Discus				
Distance m	48	60.18 (4.24)	59.58 (4.97)	60.69 (4.34)
fT pmol/L		(2.1–33.4)	(2.1–7.5)	(12.7–33.4)
T nmol/L		0.82 (0.44)	0.60 (0.24)	1.16 (0.53)
Hammer throw				
Distance m	54	69.31 (4.34)	67.76 (2.75)	70.83 (5.16)*
fT pmol/L		(1.0–50.4)	(1.0–8.5)	(13.1–50.4)
T nmol/L		0.82 (0.40)	0.62 (0.31)	1.13 (0.36)
Shot put				
Distance m	54	18.21 (1.33)	17.94 (1.58)	18.54 (1.23)
fT pmol/L		(1.8–319.9)	(1.8–10.0)	(16.8–319.9)
T nmol/L		1.03 (1.57)	0.57 (0.22)	1.72 (2.45)
Javelin				
Distance m	55	60.40 (4.35)	61.9 (4.64)	60.37 (4.40)
fT pmol/L		(2.2–26.8)	(2.2–7.3)	(9.9–26.8)
T nmol/L		0.73 (0.27)	0.59 (0.21)	0.90 (0.16)
Long jump				
Distance m	62	6.46 (0.28)	6.39 (0.34)	6.47 (0.30)
fT pmol/L		(2.1–24.2)	(2.1–6.9)	(10.4–24.2)
T nmol/L		0.80 (0.43)	0.53 (0.23)	1.00 (0.26)
Triple jump				
Distance m	54	14.00 (0.49)	14.05 (0.52)	13.95 (0.43)
fT pmol/L		(1.4–242.8)	(1.4–6.4)	(9.2–242.8)
T nmol/L		0.92 (1.59)	0.42 (0.16)	1.66 (2.54)
High jump				
Distance m	56	1.91 (0.07)	1.89 (0.06)	1.90 (0.06)
fT pmol/L		(3.6–23.4)	(3.6–6.0)	(8.9–23.4)
T nmol/L		0.74 (0.42)	0.48 (0.21)	1.07 (0.51)
Pole vault				
Distance m	48	4.51 (0.18)	4.41 (0.18)	4.54 (0.17)*
fT pmol/L		(2.0–22.9)	(2.0–6.3)	(8.7–22.9)
T nmol/L		0.78 (0.31)	0.63 (0.28)	0.98 (0.31)
20 km RW				
Time s	97	5600 (617)	5680 (235)	5647 (232)
fT pmol/L		(1.8–31.9)	(1.8–5.9)	(9.7–31.9)
T nmol/L		0.62 (0.29)	0.43 (0.16)	0.87 (0.30)
Heptathlon				
Point	53	6121 (309)	6095 (353)	6096 (263)
fT pmol/L		(1.8–30.4)	(1.8–5.7)	(8.9–30.4)
T nmol/L		0.69 (0.36)	0.47 (0.12)	1.02 (0.43)

Data are presented as mean (SD) or (min–max).
Different from the lowest fT tertile: *p<0.05.
fT, free testosterone; RW, race walking; T, testosterone.

Figure 4. Data tables from Bermon and Garnier (2017).

They show some interesting results. "When compared with the lowest female [fT = free endogenous testosterone] tertile, women with the highest fT tertile performed significantly ($p<0.05$) better in 400m, 400m hurdles, 800m, hammer throw, and pole vault with margins of 2.73%, 2.78%, 1.78%, 4.53%, and 2.94%, respectively" (Bermon and Garnier 2017, p. 1). We should note, though, that after criticism about data duplication, Bermon et al (2018) revised the 400m, 400m hurdles, and 800m to 2.1% (from 2.73%), 2.9% (from 2.78%), and 2.1% (from 1.78%), respectively. Still, these are the only events to be found in the abstract or discussion. This could lead one to think that there is a relationship between endogenous testosterone and performance, which is what would be expected as 'common sense.' However, there are at least two things to note.

First, the paper focuses almost exclusively on women. When discussing men, they note that "[a]mong the 795 male athletes, 101 showed an fT value <0.23 nmol/L" (*ibid.*, 2). There was no observable pattern for event: there were sprinters, middle distance, and long distance runners, jumpers, and throwers with low fT. A surprising proportion of the men studied are overlapping with the top of the reference range for non DSD women. Let's pause: a significant proportion of cisgender men are within the high end of female range for endogenous testosterone.[24] Yet they are competing at the highest levels of international competition. This is consistent with findings in Healy et al (2014) and Sonksen et al (2018). Unfortunately, Bermon and Garnier (2017) don't give us sufficient reporting of the data to see just how many men may be well within the 'normal' female range. There is no table for the men's results analogous to the women's.

Bermon and Garnier (2017) don't dwell on the fact that while they found the aforementioned advantages for higher testosterone in those specific women's events, "[s]uch a pattern was not found in any of the male athletic events" (*ibid.*, 1). That is, this study is some striking evidence that *there is no relationship between performance and endogenous testosterone in men.* Wow! It's a wonder that this hasn't gained greater traction in the social imaginary.

Recall the data in Figure 2 above from Sonksen et al (2018). They found that 25.4% of elite cisgender men athletes, across a large variety of sports, were below 10nmol/L. Healy et al (2014, Figure 3 above) found that 16.5% of the elite male athletes in their sample were below 8.4nmol/L. Further, we can calculate from the men's mean and standard deviations in the samples that approximately 3% of elite cis male athletes are below the geometric mean for women of 1.78nmol/L. All of these values are for total testosterone. Finally, Bermon and Garnier (2017) themselves found that 12.7% of the elite

[24] Bui et al (2015).

cis male athletes in their sample were below 0.23nmol/L fT (they did not report total testosterone, which is normally what is reported--another red flag).

Here's a surprising, unnoticed consequence of this: lowering the endogenous testosterone limit for women from 10nmol/L to 5nmol/L won't do what they think it will if some percentage of elite cisgender men are already well below that. These men could get legal recognition as women from their home country and compete as is. This is, we think, an overwhelming point against such testosterone policies being necessary or effective at promoting fairness (the core criterion of justifying *prima facie* discrimination on the basis of a natural physical trait).

In fact, we argue that this finding alone demonstrates that an endogenous testosterone limit on women cannot be necessary and effective at promoting fairness in sport. Any such policy would allow some elite cisgender men to compete as they are (provided they obtained legal recognition of a 'female' gender, which we imagine unscrupulous nations could do).[25] But let's press on nonetheless. We think the the case against endogenous testosterone limits only gets stronger.

Second, as can be seen in Figure 4, the data is presented in a table that does *not* report the percentage differences in performance between the top and bottom tertiles. One has to break out the calculator or spreadsheet to see the effects of testosterone on performance in other events. A casual reader, reporter, or even relatively interested reader will not see the data and know that there are indeed events where Bermon and Garnier found *disadvantages* for the high fT tertile. And since they don't mention any of these disadvantage events anywhere in the paper, one might come away from reading this paper thinking that there aren't any. We reproduce their data in the following table that includes all of the percentages.

Before you look, humor us. What track and field events would you expect to see an advantage? What events would you expect to see a disadvantage?

[25] In fact, although the 2016 Rio Olympics were the first Olympics after the IOC revised their trans participation policy to the 10nmol/L limit, unscrupulous nations could have simply found elite cisgender men with testosterone below the limit, given them a passport listing a 'F' sex, and allegedly won more medals. But we're happy to see that this hasn't happened in either the 2016 Rio or 2018 Sochi Olympics.

INCLUDING TRANS ATHLETES IN SPORT

BERMON and Garnier 2017

n	Running Sport	%
112	100m	-2.63%
73	100m H	-1.69%
71	200m	-1.59%
67	400m	2.74%
67	400m H	2.79%
64	800m	1.78%
66	1500m	-0.26%
56	3000m SC	-0.21%
40	5000m	-0.32%
33	10km	-0.54%
92	Marathon	3.04%

n	Non-Running Sport	%
48	Discus	1.83%
54	Hammer	4.33%
54	Shot put	3.24%
55	Javelin	-2.53%
62	Long Jump	1.24%
54	Triple Jump	-0.72%
56	High Jump	0.53%
48	Pole vault	2.86%
97	20km Walk	0.58%
53	Heptathlon (pts)	0.02%

Figure 5. Calculated net 'advantage' for highest tertile fT women compared to lowest tertile fT women from Bermon and Garnier (2017).

If you humored us, you probably thought that the pure power/sprint events (100m, 100m hurdles, 200m) would show an advantage for the high fT tertile, but they showed a *dis*advantage to the tune of -2.63%, -1.69%, and -1.59%, respectively. And an event where one might expect a disadvantage (marathon) showed an advantage. As for the field (non-running) events, one might not be surprised to see an advantage for pure power events like hammer throw and shot put, but javelin showed a *disadvantage* of -2.53%.

Here's the kicker: the weighted average "advantage" for running events is a paltry 0.31% and 1.07% for non-running events. And yet, these results are nowhere near the ~8-12% performance gap that is attributed to testosterone (by Handelsman et al 2018, for example). This is a serious mark against using endogenous testosterone as an eligibility criterion.

And while it's certainly true that at the highest level of competition, small advantages can be decisive, sport permits other natural physical characteristics that confer much larger advantages than these the 1-3% being attributed to endogenous testosterone in women.[26] Some such genetic factors related to athletic performance include genes related to skeletal muscle. Ma et al (2013, p. 8), for example, write:

> In conclusion, the present study summarized the associations of sport performance with ACE I/D and ACTN3 R577X polymorphisms. The results consistently provided more solid evidence for associations between ACE II genotype and endurance events, and between ACTN R allele and power events. Our findings provided more solid evident to support that human physical performance might be influenced by genetic profiles.

Yang et al (2003, p. 629) continue: "Although at least 73 genetic loci have been associated with fitness and performance phenotypes (Rankinen et al. 2002), ACTN3 is the first structural skeletal-muscle gene for which such an association has been demonstrated."

This is consistent with Eynon et al (2013, p. 803), who write:

> Several genetic variants (i.e. polymorphisms) have been associated with elite power and sprint performance in the last few years and the current paradigm is that elite performance is a polygenic trait, with minor contributions of each variant to the unique athletic phenotype. … ACTN3 is the only gene that shows a genotype and performance association across multiple cohorts of elite power athletes, and this association is strongly supported by mechanistic data from an Actn3 knockout mouse model.

Elite athletic performance is more complicated than 'more endogenous testosterone means better athlete.' The Bermon and Garnier (2017) show this in the lack of an effect in men, and the sporadic and weak effect (if it's correct to call it a causal effect at all) in women. None of these are used as eligibility criteria, but for some reason endogenous testosterone is even though it confers a smaller competitive advantage (if at all).

As Guth and Roth (2013, p. 656) note, bodies are extremely complicated and it's not as simple as isolating one or a few genetic or hormonal factors as predictive of athletic success:

[26] Epstein (2014) details dozens of genetic mutations that confer very large competitive advantages. And again, even ignoring intersex and trans women, the gap between the bottom of the 'normal' range for female athletes is much farther away from the top of the 'normal' range (pick your attribute) than 8-12%.

> It remains to be seen whether the complex contribution of genetic factors to athletic performance can be used to improve talent selection. It is important to remember that genetic association studies reveal factors that are associated with athletic performance traits at a population level and the relative importance of any given variant for an individual is undoubtedly more variable, thus genetic screening cannot be used to conclusively predict or rule out athletic success. There is neither currently nor is there likely to be a gene variant that is either required or sufficient for superior athletic performance.

However, let's continue to dig deeper into the claims that endogenous testosterone are predictive of athletic performance such that a limit on women may be justified.

Bermon and Garnier (2017) do not mention any of the events that showed a disadvantage for the top fT tertile. Why is that? Moreover, looking at the data, there's no observable, predictable pattern for which events will show which advantages or disadvantages. Even though the average "advantage" in running events is only 0.31%, while the sample size is impressive for a study of this kind, the sample sizes for each individual event are relatively low. We shouldn't draw conclusions for event-specific testosterone limits (as Bermon and Garnier 2017, Bermon et al 2018, and Handelsman et al 2018 do). Bermon et al (2018, p. 1) claims, of the 2017 study, that "the hypothesis is that elevated testosterone levels enhance performance in certain events but not others." They offer no mechanism for why this would be the case.

In fact, it's a mistake to even speak about these results in terms of "advantage" of one group over another. At best we can say that one tertile performed better or worse than another, but it's too hasty to say that they have a competitive *advantage*. Correlation is not causation, after all. Admittedly this is a small, but we think important, linguistic point.

One additional worry with Bermon and Garnier (2017) is that singling out only the events where high fT tertile athletes outperformed low fT tertile athletes is that the study may fall prey to a practice known as "p-hacking" or "data dredging." Very roughly, p-hacking "occurs when researchers collect or select data or statistical analyses until nonsignificant results become significant."[27] There are two worries: first, did Bermon

[27] Head et al (2015, p. 1).

and Garnier (2017) use tertiles, instead of the more standard quintiles, because only tertiles produced significant results? We'd like to see the data analyzed using quintiles.

Second, the claim is that endogenous testosterone confers an unfair competitive advantage. Why would fT only affect women's performances, but not men's, especially when the claim is that what gives men the 8-12% performance advantage over women is testosterone? And why would fT only affect a select few events, including some events that are physiologically similar to events that showed a *disadvantage*?

The 100m, 100m hurdles, and 200m sprints are power events, much like the hammer and shot put: why would the latter show a performance 'advantage,' but not the former? Why would the 200m show a 'disadvantage,' but the 400m an advantage? We think that these are serious worries, particularly for drawing policy conclusions from the 2017 study, as the IAAF has apparently done in attempting to update its HRs only for the 400m to 1500m (and the mile) events for women.[28] Recall that the 1500m showed a -0.26% difference between the high vs low fT tertiles: why was this event included?[29]

Much more importantly, these data outcomes are completely consistent with there being *no advantage* for high fT tertile women, and for the performance differences being what we expect from randomness. Think of it this way. Suppose we ask a group of 50 people to flip a fair coin 10 times. Results will range from 10H:0T to 0H:10T. The mean will be 5H:5T. However, we will see a normal distribution of results spanning this range with a realized mean probably a little above or below the true mean. The realized distribution from the sample size of 50 is *unlikely* to be exactly 5H:5T, with a binomial probability of 8%. That is, we should expect the results of our random trial *not* to indicate the true mean. The upshot is that if there's no effect of endogenous testosterone and performance, we will still expect some samples to "show" an advantage and disadvantage, to be unpredictably distributed (in terms of which events have which dis/advantage), and to average out close to 0%, but not exactly 0% depending on our sample size.

To reiterate, many of the power-based events we'd expect to see an advantage for being in the top fT tertile actually showed a disadvantage. They showed a disadvantage of the magnitude that Bermon and Garnier singled out when it was an advantage. For

[28] https://www.iaaf.org/news/press-release/eligibility-regulations-for-female-classifica We note that this policy has been put on hold, pending a CAS challenge by South African intersex sprinter Caster Semenya.

[29] In Bermon et al (2018) they claim that the revised result is 1.5%. But what about the marathon? Why are the long sprints and middle distance events singled out for the new 5nmol/L policy, but not the marathon?

example, while the 400m showed a 2.74% advantage, the 100m showed a -2.63% disadvantage; shot put showed a 3.24% advantage, the javelin showed a -2.53% disadvantage. However, *this is precisely what we would expect to see if*, just like the data for the men, *there is no relationship between fT and performance*! There's no pattern to the data. There's no reason to think that there's something special about 'middle distance' events like the 400m-800m events, that then disappears for the 1500m, and where the 100m-200m events show a disadvantage. A better hypothesis is that, just like for the men's data, there is no relationship between endogenous testosterone and performance.

Simply put, it is a serious mistake to take the results of a random variable and only single out those that are greater than the mean, which is what Bermon and Garnier (2017) have done, singling out the highest positive results, while completely ignoring the corresponding negative results in physiologically similar events. This would be like taking anyone who flipped more than 7H in the coin experiment, and banning them from competition, claiming that what explains their achieving 7H is some advantage. But the results of the coin flipping are random around the mean of 5. We *expect* some people to get 7H some of the time, just like we should expect some events to show a phantom 'effect.' But this is just a result of noise in the data, not an actual physiological effect of fT on performance. Remember, for men, there is *no* relationship between fT and performance.

Two final notes on Bermon and Garnier (2017): first, they don't report the performance results for the men, broken down by event and fT tertiles like they did for the women. When they report that "such a pattern was not found in any of the male athletic events," are they reporting that every event resulted in identical performances for the bottom and top fT tertiles? That would be remarkable. It's also unlikely. It crucial to see the men's data represented in a table just like the women. We suspect that, just like the women, some events would show a slight disadvantage, and others would show a slight advantage. Again, this is what we would expect if there's no relationship between fT and performance: the results will be randomly distributed close to the mean.

Second Bermon and Garnier (2017) split the tertiles based on fT, or free testosterone in serum. However, the IAAF's HRs control *total* testosterone (T) in serum. It's true that fT is the biologically active level of testosterone in one's body (in addition to albumin-bound testosterone). However, the relationship between fT and T values varies wildly from person to person, depending on the binding proteins in a given person's body. For example, one could instead use the free androgen index (FAI), which is the ratio of total testosterone to sex hormone binding globulin (SHBG), which can

range from nearly 0 to 1,000 (ng/dL).[30] According to Bui et al (2015), the reference interval for women without PCOS (polycystic ovarian syndrome) or any DSDs is an fT of 5.2-26pmol/L. That's a five-fold difference between the high and low end of the 'healthy' interval.[31] The range is even larger when we include women with PCOS and DSDs. This is all to say that given the IAAF's HRs is based on total testosterone, it's odd that Bermon and Garnier (2017) focused on fT.

There are other methodological issues with using fT. It is notoriously difficult to measure accurately.[32] Moreover, other factors such as pregnancy, testosterone-secreting tumors, PCOS, and much more can contribute wildly different free testosterone values. There have been rumors that some women engage in "abortion doping," though there seems to be little empirical evidence of this.[33] And given that the relationship between total testosterone and free testosterone varies wildly by the person, using fT to establish a connection to performance, to be enforced on the basis of total testosterone is methodologically odd.

Moving on, in support of either the 10nmol/L or 5nmol/L limits, Handelsman et al (2018) focus heavily on the claim that performance is dose-responsive to testosterone (whether endogenous or exogenous). We do not dispute the studies they cite. However, we dispute their conclusions from those studies. The write (Handelsman et al 2018, p. 803):

> from puberty onward a clear sex difference in athletic performance emerges as circulating testosterone concentrations rise in men because testes produce 30 times more testosterone than before puberty with circulating testosterone exceeding 15-fold that of women at any age.[34] There is a wide sex difference in circulating testosterone concentrations and a reproducible dose-response relationship between circulating testosterone and muscle mass and strength as well as circulating hemoglobin in both men and women. These dichotomies largely account for the sex differences in muscle mass and strength and circulating

[30] Bui et al (2015).
[31] And there are issues in defining what constitutes "healthy" such that one doesn't beg the question against women with naturally higher testosterone!
[32] Rosner et al (2007).
[33] Sorensen (2009).
[34] This claim is only true making a few additional assumptions: they don't mean the *average* male post-pubertal testosterone is 15-fold the post-pubertal testosterone in all women (even excluding women with PCOS or a DSD). They likely mean that the upper limit of men's post-pubertal testosterone is 15-fold some value for women, excluding PCOS and DSD women.

> hemoglobin levels that result in at least an 8% to 12% ergogenic advantage in men.

This elides the wide range of endogenous testosterone values for cisgender men. As found in Bermon and Garnier (2017), 12.7% of elite cisgender men had fT testosterone values <0.23nmol/L. The men on the low end of the range do not have ~15 times as much endogenous testosterone as the women on the high end of the women's range. There is complete overlap with the women's range for men on the low end of the men's range.

Moreover, when reporting "healthy" ranges, Bermon and Garnier (2017) and Handelsman et al (2018) potentially beg the question by excluding DSD women. DSD women are not an aberration: about 1% of the population is intersex.[35] While not every intersex condition involves elevated testosterone, when considering the reference ranges for endogenous testosterone for women, it's important not to arbitrarily exclude high testosterone women, since that will bias the conclusions we want to draw from the data, and it obscures how much overlap there is between populations, and complete overlap if we include hypogonadal men who produce next to no endogenous testosterone.

Additionally, recall that *there is no relationship between endogenous testosterone and performance* for men. Even if we were to grant that what entirely explains men's performance gap, on average, over women is pubertal testosterone (which we dispute as far from uncontroversial!), why wouldn't higher peri- and post-pubertal testosterone men tend to outperform low peri- and post-pubertal testosterone men? Handelsman (2018) note that the lower 95% confidence limit for men's endogenous testosterone is 7.7nmol/L, while the upper 95% confidence limit is 29.4nmol/L. But the upper limit men are not any bigger, faster, or stronger than the lower limit men. Bermon and Garnier (2017) showed us that.

If athletic performance were so clearly dose-responsive to pubertal testosterone, why is there no performance gap *within* the men's range, especially with a 4-fold difference? After all, the difference between the upper 95% confidence limit for women (excluding DSD women) is 1.68nmol/L, which is 4.58-fold below the bottom 95% confidence limit value for men. If we include women with PCOS, the upper 95% limit is 3.1nmol/L, putting the lower 95% limit men a mere 2.48-fold higher. That is, if men who have approximately a 4-fold more testosterone than other men aren't any bigger, faster, and stronger, why would men with approximately a 4-fold testosterone than women be

[35] http://www.isna.org/faq/frequency

bigger, faster, and stronger? The problem here is that isolating endogenous testosterone as the primary source of physiological differences in men and women is fundamentally flawed.[36]

So the irony is that we think that Bermon and Garnier (2017) establish our argument against testosterone limits for women's sport. One eighth of cisgender men are naturally already below the upper 'normal' range for cisgender women. There's no relationship between endogenous testosterone and performance in men. There is a highly dubious relationship, at best, in women. Testosterone is a hopeless unreliable predictor of performance in post-puberty athletes. It cannot serve the function the IOC, IAAF, and other sports organizations want it to.[37]

For these reasons, no endogenous testosterone limit policy will be effective at promoting fairness in competition. This is a necessary condition for justifying *prima facie* discrimination. But we have also argued that such policies are not necessary for promoting fairness in sport, since we permit much larger competitive advantages (than even the 8-12% people want to attribute to endogenous testosterone) between women based on natural physical traits such as height, myostatin mutations, and dozens of other physiological sources of competitive advantages.

Therefore, endogenous testosterone is clearly neither necessary nor effective for promoting fairness in sport.

We close this section with one final thought. In Chand 2015, the IAAF claimed that the HRs are not a sex-verification policy, but rather an eligibility criterion for female athletes.[38] In various places, Bermon has argued for reducing the endogenous testosterone limit for women down to 5nmol/L, while maintaining that such a policy is merely an eligibility criterion, *not* a sex-verification criterion. But Handelsman et al (2018, p. 803) write the following: "the appropriate eligibility criterion for female athletic events should be a circulating testosterone of <5.0nmol/L. This would include all women other than those with untreated hyperandrogenic disorders of sexual development and noncompliant male-to-female transgender [sic] as well as testosterone-treated female-to-male transgender [sic] or androgen dopers."[39]

---

[36] Karkazis et al (2012) discuss many of the reasons for this.
[37] Unfortunately, for the purposes of length, we cannot discuss Handelsman, Hirschberg, and Bermon (2018) in detail. We save that for a future paper. Suffice it to say, but our position is that their conclusions do not follow from the data.
[38] Behrensen (2013) argues that the HRs are a sex verification policy by another name.
[39] 'Transgender' and 'trans' are adjectives, not nouns. Also, the language of 'male-to-female' and 'MtF' is increasingly dispreferred and outdated.

Why would a female eligibility criterion need to do the work of excluding "testosterone-treated" trans men and androgen doping women? Trans men are *men*. They are excluded from women's sport, under IOC policies, if they are legally recognized as men: trans men do not need to undergo testosterone hormone therapy in order to compete as men. In fact, once they do undergo testosterone hormone therapy, while they can acquire a TUE, they are immediately ineligible for women's competition. For similar reasons, we do not need a gender eligibility criterion to exclude testosterone doping women: that's what anti-doping policies are for.

Bermon (via Handelsman et al 2018), we think, has tipped his hand that the endogenous testosterone policies *are* in fact a kind of sex-verification policy. Moreover, the assumption is that "non-compliant" trans women (that is, trans women who have not undergone testosterone suppression) should be deemed ineligible. But this grossly ignores that a significant proportion of pre-transition trans women already have endogenous testosterone levels below even 5nmol/L. Moreover, remember, Bermon and Garnier (2017) found that 12.7% of cisgender men had endogenous testosterone levels below 0.23nmol/L fT. And as noted above, surprisingly large percentages of elite cis male athletes are naturally below 5nmol/L. Of course, trans women are not cis men. But nonetheless, the newly proposed policy of 5nmol/L wouldn't exclude these athletes from women's competition (provided they received a legal document that they're 'female,' even by an unscrupulous government).

Again, simply put, no endogenous testosterone policy will be necessary or effective at promoting fairness in (women's) sport.

We close with this summary of the available data on the (lack of) relationship between endogenous testosterone and performance from Sonksen et al (2018, pp. 14-15, in text citations omitted, emphasis ours):

> More recently Eklund et al. [2017] related serum androgen levels to performance in a group of female athletes. They found no differences in testosterone levels between sports and no correlation between serum testosterone and performance but were able to show weak correlations between some androgen precursors and performance. In a recent review Bermon [2017] reported that he was able to find a relationship between free testosterone levels and performance such that those with the highest free testosterone levels had between 1.8 and 4.5% advantage; they found no relation between free testosterone and performance in elite men. In the

> full paper [Bermon and Garnier 2017] *it is clear that they showed no relationship between endogenous testosterone concentration and performance in elite women athletes nor, unlike the men, any differences in free testosterone between sports*. Thus neither group showed a significant correlation between serum total testosterone (the endocrine variable used in the 'hyperandrogenism' rule) and performance. They did claim to show a relationship between (estimated) free testosterone and performance in five out of the 21 sub-groups *but in nine of these sub-groups, those with the lower testosterone performed better*.

The current state-of-the-art scientific evidence all shows no relationship between endogenous testosterone and performance. It's time to shed this myth.

5. Fairness Requires That Trans* Athletes Compete in Their Post-Transition Sex/Gender (Conclusion)

Endogenous testosterone is the only natural physical characteristic for which we deem some athletes ineligible for competition, and it's only ever women. There is no limit to endogenous testosterone for men. Moreover, all available scientific evidence (Healy et al 2014, Bermon and Garnier 2017, Eklund et al 2017, and Sonksen et al 2018) shows that there is *no relationship between endogenous testosterone and performance*, even for women. There is no scientific justification for an endogenous testosterone limit on women, whether they're cis, trans, or intersex.

Moreover, even if we granted the premise that there is a relationship between endogenous testosterone and performance, we permit much larger competitive advantages than is being attributed to testosterone, such as height, metabolic mutations, socioeconomic status, coaching, access to facilities, etc.. Therefore, placing a limit on endogenous testosterone for women, including trans women, is neither necessary nor effective at promoting fairness in sport.

We have argued for a Rawlsian conception of 'fairness in competition.' When we consider the question, "Should trans women, who are legally women/female, have the right to compete in women's sport?" from behind the veil of ignorance, we argue that the only rational answer is 'Yes.' Fairness cannot require the elimination of all significant competitive advantages, for the purpose of training, coaching, equipment, and so on is explicitly to *un-level* the playing field in terms of competitive advantage. This is as it should be.

Simply put: trans women are women, and it's fair for trans women to compete in women's sport.

Works Cited

Andow, James. 2018. Are Intuitions About Moral Relevance Susceptible to Framing Effects? *Review of Philosophical Psychology* 9(1), pp. 115-141.

Betancurt, J., Zakynthinaki, M., Martinez-Patiño, M., and Martinez, C.. 2018. Hyperandrogenic athletes: performance differences in elite-standard 200m and 800m finals. *Journal of Sports Sciences*, 36 (21), pp. 2464-2471, DOI: 10.1080/02640414.2018.1464620

Behrensen, M.. 2013. In the halfway house of ill repute: Gender verification under a different name, still no contribution to fair play. *Sport, Ethics and Philosophy* 7(4), pp. 450-466. DOI: 10.1080/17511321.2013.856460

Bermon, Stephane. 2017. Androgens and athletic performance of elite female athletes. *Current Opinion in Endocrinology, Diabetes, and Obesity*, 24(3), pp. 246-251.

Bermon, S. and Garnier, P-Y.. 2017. Serum androgen levels and their relation to performance in track and field: Mass spectrometry results from 2127 observations in male and female athletes. *British Journal of Sports Medicine*, 51, pp. 1309-1314.

Bermon, S., Hirschberg, A., Kowalski, J., and Eklund, E.. 2018. Serum androgen levels are positively correlated with athletic performance and competition results in elite female athletes. *British Journal of Sports Medicine*. Available online June 20, 2018. doi:10.1136/bjsports-2018-099700

Bermon, S., Ritzen, M., Hirschberg, A., and Murray, T.. 2013. Are the new policies on hyperandrogenism in elite female athletes really out of bounds? Response to "Out of bounds? A critique of the new policies on hyperandrogenism in elite female athletes." *The American Journal of Bioethics*, 13(5), pp. 63-65. DOI: 10.1080/15265161.2013.776129

Bialystok, L.. 2016. Transgender inclusion in single-sex competition: The case of beauty pageants. *Social Theory and Practice*, 42(3), pp. 605-635. DOI: 10.5840/soctheorpract201642318

Bui, H.N., Sluss, P.M., Hayes, F.J., Blincko, S., Knol, D.L., Blankenstein, M.A., and Heijboer, A.C. 2015. Testosterone, free testosterone, and free androgen index in women: Reference intervals, biological variation, and diagnostic value in polycystic ovary syndrome. *Clinica Chimica Acta* 23(450), pp. 227-232.

Coggon, J., Hammond, N., and Holm, S.. 2008. Transsexuals in Sport--Fairness and Freedom, Regulation and Law. *Sport, Ethics and Philosophy* 2(1), pp. 4-7. DOI: 10.1080/17511320801896091

Demaree-Cotton, Joanna. 2014. Do framing effects make moral intuitions unreliable? *Philosophical Psychology* 29(1), pp. 1-22.

Eklund, Emma, Bo Berglund, Fernand Labrie, Kjell Carstrom, Lena Ekstrom, and Angelica Linden Hirschberg. 2017. Serum androgen profile and physical performance in women Olympic athletes. *British Journal of Sports Medicine* 51, pp. 1301-1308.

Epstein, David. 2014. *The Sports Gene: Inside the Science of Extraordinary Athletic Performance*. Portfolio/Penguin: New York.

Franklin S., Betancurt J., and Camporesi S.. 2018. What statistical data of observational performance can tell us and what they cannot: the case of Dutee Chand v. AFI & IAAF British Journal of Sports Medicine, 52, pp. 420-421.

Gauthier, David. 1986. *Morals by Agreement*. Oxford: Oxford University Press.

Harper, J.. 2015. Race times for transgender athletes. *Journal of Sporting Cultures and Identities*, 6(1), pp. 1-9.

Harper, Joanna, Jonathan Ospina Betancurt, Maria Jose Martinez-Patino. 2018. Analysis of the performance of transgender athletes. https://www.sportsci.org/2016/WCPASabstracts/ID-1699.pdf Last accessed March 3 2019.

Healy, M.L., J. Gibney, C. Pentecost, M.J. Wheeler, and P.H. Sonksen. Endocrine profiles in 693 elite athletes in the postcompetition setting. *Clinical Endocrinology* 81, pp. 294-305.

Guth, Lisa M and Stephen M Roth. 2013. Genetic influence on athletic performance. *Current opinion in pediatrics*, vol. 25(6), pp. 653-8.

Head M.L., Holman L., Lanfear R., Kahn A.T., and Jennions M.D.. 2015. The Extent and Consequences of P-Hacking in Science. *PLoS Biology* 13(3), pp. 1-15. https://doi.org/10.1371/journal.pbio.1002106

Jones, Eric, Bishop, Phil, Woods, Amanda, and Green, James. 2008. Cross-sectional area and muscular strength. *Sports Medicine* 38(12), pp. 987-994.

Kanehisa, Hiroaki, Ikegawa, Shigeki, and Fukunaga, Tesuo. 1994. Comparison of muscle cross-sectional area and strength between untrained women and men. *European Journal of Applied Physiology and Occupational Physiology* 68(2), pp. 148-154.

Karkazis, K., Jordan-Young, R., Davis, G., and Camporesi, S.. 2012. Out of bounds? A critique of the new policies on hyperandrogenism in elite female athletes. *The American Journal of Bioethics*, 12(7), pp. 3-16. DOI: 10.1080/15265161.2012.680533

Karkazis, Katrina and Jordan-Young, Rebecca. 2015. Debating a testosterone "sex gap." *Science* 22(348), pp. 858-860.

CAS 2014/A/3759 Dutee Chand v. Athletics Federation of India (AFI) & The International Association of Athletics Federations (IAAF)

Loland, Sigmund and McNamee, Mike. 2000. Fair play and the ethos of sports: An eclectic philosophical framework. *Journal of the Philosophic Society for the Study of Sport*, 27, pp. 63-80.

Loland, Sigmund. 2001. *Fair Play in Sport: A Moral Norm System*. London: Routledge.

Ma, F., Yang, Y., Li, X., Zhou, F., Gao, C., Li, M., & Gao, L. (2013). The association of sport performance with ACE and ACTN3 genetic polymorphisms: a systematic review and meta-analysis. *PloS One*, 8(1), e54685.

Patel, Seema. 2015. The IAAF's Hyperandrogenism Regulations Suspended. *World Sports Law Report*, August, pp. 8-11.

Petrinovich, Lewis and O'Neill, Patricia . 1996. Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, 17(3): 145-171

Rankinen T, Perusse L, Raurama R, Rivera MA, Wolfarth B, Bouchard C (2002) The human gene map for performance and health-related fitness phenotypes: the 2001 update. *Med Sci Sports Exerc* 34:1219–1233

Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H.. 2007. Position statement: Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *Journal of Clinical Endocrinology and Metabolism* 92 (2): pp, 405–13. doi:10.1210/jc.2006-1864

Sonksen, Peter, Richard Holt, Walailuck Bohning, Nishan Guha, David Cowan, Chrstiaan Bartlett, and Dankmar Bohning. Why do endocrine profiles in elite athletes differ between sports? *Clinical Diabetes and Endocrinology* 4(3), pp. 1-16.

Sorensen, Elizabeth. 2009. Debunking the myth of pregnancy doping. *Journal of Intercollegiate Sport* 2(2), pp. 269-285.

Suits, Bernard. 2005. *The Grasshopper: Games, Life, and Utopia*. Peterborough, Ontario: Broadview Press. (Original work published 1978).

Yang N, MacArthur DG, Gulbin JP, Hahn AG, Beggs AH, et al. (2003) ACTN3 genotype is associated with human elite athletic performance. Am J Hum Genet 73: 627–631.

Young, Iris Marion. Throwing like a Girl: A Phenomenology of Feminine Body Comportment Motility and Spatiality. *Human Studies*, vol. 3, no. 2, 1980, pp. 137–156. JSTOR, JSTOR, www.jstor.org/stable/20008753.