

# Serum androgen levels are positively correlated with athletic performance and competition results in elite female athletes

Stéphane Bermon,<sup>1,2</sup> Angelica Lindén Hirschberg,<sup>3,4</sup> Jan Kowalski,<sup>5</sup> Emma Eklund<sup>3</sup>

We thank our colleagues for their constructive comments that relate to our two articles in the July 2017 issue of *BJSM*.<sup>1,2</sup>

## BERMON AND GARNIER STUDY

The main criticisms expressed about the data used and our statistical analysis were (1) our 'concentration' on free testosterone (fT) rather than total testosterone (T)<sup>3</sup> (presumably because the T results were only presented in the Internet version of the paper), (2) the fact that 17.3% of the athletes were sampled at both World Championships (Daegu and Moscow),<sup>3</sup> (3) the fact that no correlation analysis was performed other than comparison of fT or T tertiles,<sup>3</sup> (4) the absence of statistical comparison between a group with high T levels and a group with normal T levels<sup>4</sup> and (5) the lack of adjustment for multiple comparisons (suggesting that the significant differences observed in our study could have happened by chance).<sup>5</sup>

Taking this last criticism first, we note that we presented an exploratory study, without no attempt to claim confirmatory results. In fact, the exploratory evidence presented in the study is strong, and correction for multiplicity may be too conservative. But we agree that the results should be put into context. At the type 1 error level of 0.05, we could expect 1 of 20 hypotheses tested to be significant merely by chance, that is,  $p < 0.05$ . In this study, we have observed significant correlations at the 0.05 level in five events, out of 21 events in total. Therefore, it is very unlikely that all these findings are caused by chance. Moreover, the five flagged events were showing similar

findings in both fT and T, which indicates that the evidence and findings are robust.

Franklin *et al* report that, using Fisher's combination test, they were unable to reject the hypothesis that the  $p$  values calculated from the published data were not inconsistent with there being no advantage to high fT women in any of the five events,<sup>5</sup> but they do not explain their analysis in any detail or support their conclusion with any evidence. Their suggestion to conduct one single test of correlation across all 21 events combined is inappropriate because the hypothesis is that elevated testosterone levels enhance performance in certain events but not others. Combining results from the events of long sprint to middle-distance running shows with additional statistical calculation a  $p$  value of 0.003 for the correlation between testosterone levels and competition results, which is very strong statistical evidence for a true correlation not caused by chance.

## ADDITIONAL SENSITIVITY ANALYSES

To address the other criticisms, we have now performed a sensitivity analysis using a modified data set in which (1) observations from athletes who participated in both World Championships were only counted once, that is, the first observed value, and (2) only T concentrations were used (not fT). As for aggregating results, we used running events from 400m up to 1 mile, on the basis that that is where T produces its greatest performance-enhancing effects (by increasing lean body mass and the concentration of circulating haemoglobin). Long sprints that rely on mixed energy pathways (mainly anaerobic) and middle-distance races that rely on mainly aerobic mixed pathways are also considered as restricted events in the International Association of Athletics Federations (IAAF) Eligibility Regulations for the Female Classification that will come into force on 1 November 2018.

Therefore, we have aggregated results from the long sprints (400m events), then

middle-distance runs (800m and 1500m), and finally long sprints and middle-distance runs (400m, 800m and 1500m), into one event group for further statistical analysis. The time results were transformed into an index, that is, percentage of the best performance achieved by each event. We have used the Spearman rank-order correlation coefficient to explore the correlation between competition results and testosterone levels, using a two-sided test at the 0.05 significance level. As was done with the original data set, we analysed the modified analysis population by comparing the data from the highest and the lowest T tertiles, using a Mann-Whitney U test, because it emphasised on the extremes of the magnitude of possible differences in performances. Finally, we used a serum T threshold concentration of 2nmol/L to identify a group of female athletes with 'high T' levels, for comparison against the results of athletes with T levels of less than 2nmol/L ('normal T' levels).<sup>4</sup> We have excluded 230 observations, corrected some data capture errors and performed the modified analysis on a population of 1102 female athletes.

The results showed that T concentration was significantly ( $p < 0.03$ ) correlated to performance for the 400m hurdles, 800m and hammer-throw events. A trend ( $p < 0.09$ ) towards significance was found for the 400m (table 1, online supplementary results). Aggregated performance results obtained in 400m up to 1500m events positively and significantly ( $p = 0.003$ ) correlated with T.

A comparison of the lowest and the highest tertiles in each running event confirmed that the women in the highest T tertiles performed better in 400m, 400m hurdles and 800m, with margins of 2.1%, 2.9% and 2.1%, respectively (table 2, online supplementary results). A similar analysis of long sprints (400m and 400m hurdles) and middle-distance races (800m and 1500m) showed a significant ( $p < 0.01$ ) difference of 2.5% and 1.5%, respectively (table 4, supplementary results). The T threshold of 2nmol/L applied to aggregated results for long sprints and middle-distance races generated 'high T' (median T concentration: 3.51nmol/L) and normal T (median T concentration: 0.64nmol/L) groups. As expected, the 'high T' group outperformed the normal T group by 1.6% (table 5, online supplementary results). In non-running events, the female hammer throwers in the highest T tertile performed 7.35% better than those in the lowest tertile (table 3, online supplementary results).

In conclusion, our complementary statistical analysis and sensitivity analysis using a

<sup>1</sup>LAMHESS, Université Côte d'Azur, Nice, France

<sup>2</sup>Monaco Institute of Sports Medicine and Surgery, Monaco, Monaco

<sup>3</sup>Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

<sup>4</sup>Department of Gynecology and Reproductive Medicine, Karolinska University Hospital, Stockholm, Sweden

<sup>5</sup>JK Biostatistics AB, Stockholm, Sweden

Correspondence to Dr Stéphane Bermon, Monaco Institute of Sports Medicine and Surgery, Monaco 98000, Monaco; bermon@unice.fr

modified analysis population shows consistent and robust results and has strengthened the evidence from this study, where we have shown exploratory evidence that female athletes with the highest T concentration have a significant competitive advantage over those with lower T concentration, in 400m, 400m hurdles, 800m and hammer throw, and that there is a very strong correlation between testosterone levels and best results obtained in the World Championships in those events. A similar trend is also observed for 1500m and pole vault events.

## LIMITATIONS

Our critics are correct that even the large and statistically proven competitive advantages shown in our study do not establish a difference in sport performance of the magnitude typically observed between male and female athletes (10%–12%),<sup>3,6</sup> a point highlighted by the CAS Panel in *Chand v IAAF & AFI*.<sup>7</sup> But it was never expected that our study would show this since it compared normal T concentration groups with medium or medium-high T concentration groups. It did not compare normal T concentration groups with the specific group of androgen-sensitive women with T concentration within the (much higher) normal male range. That will be done in a separate paper.<sup>8</sup>

## EKLUND ET AL STUDY

The Eklund *et al* study<sup>2</sup> was criticised by Sönksen *et al*<sup>3</sup> due to its cross-sectional design of 106 female Olympic athletes and 117 sedentary controls. We agree that an intervention study is preferable to answer the question whether there is a causal effect of testosterone on physical performance in women. However, the aim of our study was to compare the endogenous serum androgen profile (including precursors and metabolites) between female Olympic athletes and sedentary, age-matched and Body Mass Index-matched controls, and to investigate the association between the

androgen profile, body composition and physical performance in the athletes. We believe that a cross-sectional study also can provide valuable information and increase knowledge in the area.

In our study, subjects with disorders of sex development including androgen insensitivity and congenital adrenal hyperplasia were excluded. Furthermore, doping was excluded by doping tests. Sönksen *et al* note from our study that the athletes had comparable serum levels of testosterone with the controls and that there was no correlation between testosterone and physical performance in the athletes.<sup>3</sup> However, they ignore the fact that we found increased serum levels of several precursor androgens including dehydroepiandrosterone (DHEA) in the athletes, compared with the controls. Furthermore, there were clear positive correlations between the androgens DHEA and dihydrotestosterone (DHT) and physical performance in the athletes. This is important because there is increasing evidence that DHEA is the major tissue-specific source of the two bioactive androgens in women (ie, testosterone and DHT) that can bind to the androgen receptor.<sup>9,10</sup> Thus, like the Berman and Garnier study,<sup>1</sup> the Eklund study<sup>2</sup> demonstrated that serum androgen levels within the normal range in elite female athletes are clearly and positively associated with athletic performance.

**Contributors** SB performed the statistical analysis and drafted the text and tables. ALH and EE drafted and reviewed the texts and tables. JK reviewed the statistical aspects.

**Competing interests** None declared.

**Patient consent** Not required.

**Provenance and peer review** Not commissioned; internally peer reviewed.

**Data sharing statement** The data presented here are obtained from statistical analysis of anonymised biological results and their related individual performance. However, raw data cannot be shared because they make identification of individuals (and their personal data) possible.

© Author(s) (or their employer(s)) 2018. No commercial re-use. See rights and permissions. Published by BMJ.

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bjsports-2018-099700>).



**To cite** Berman S, Hirschberg AL, Kowalski J, *et al*. *Br J Sports Med* 2018;**52**:1531–1532.

Accepted 20 June 2018  
Published Online First 7 July 2018



► <http://dx.doi.org/10.1136/bjsports-2017-097792>  
*Br J Sports Med* 2018;**52**:1531–1532.  
doi:10.1136/bjsports-2018-099700

## REFERENCES

- 1 Berman S, Garnier PY. Serum androgen levels and their relation to performance in track and field: mass spectrometry results from 2127 observations in male and female elite athletes. *Br J Sports Med* 2017;**51**:1309–14.
- 2 Eklund E, Berglund B, Labrie F, *et al*. Serum androgen profile and physical performance in women Olympic athletes. *Br J Sports Med* 2017;**51**:1301–8.
- 3 Sönksen PH, Bavington LD, Boehning T, *et al*. Hyperandrogenism controversy in elite women's sport: an examination and critique of recent evidence. *Br J Sports Med* 2018;**52**:1481–2.
- 4 Menier A. Use of event-specific tertiles to analyse the relationship between serum androgens and athletic performance in women. *Br J Sports Med* 2018;**52**:1538–9.
- 5 Franklin S, Ospina Betancurt J, Camporesi S. What statistical data of observational performance can tell us and what they cannot: the case of *Dutee Chand v AFI & IAAF*. *Br J Sports Med* 2018;**52**:420–1.
- 6 Camporesi S. A question of 'fairness': why ethics should factor in the court of arbitration for sport's decision on the IAAF Hyperandrogenism Regulations. *Br J Sports Med* 2018. doi: 10.1136/bjsports-2018-099387. [Epub ahead of print 11 Apr 2018].
- 7 CAS 2014/A/3759, *Dutee Chand v. Athletics Federation of India (AFI) & The International Association of Athletics Federations (IAAF)*, July 2015. [http://www.tas-cas.org/fileadmin/user\\_upload/award\\_internet.pdf](http://www.tas-cas.org/fileadmin/user_upload/award_internet.pdf) (accessed 5 May 2018).
- 8 Handelsman D, Hirschberg A, Berman S. Circulating testosterone as the hormonal basis of sex differences in athletic performance. *Endocr Rev*. In Press. 2018.
- 9 Labrie F, Bélanger A, Bélanger P, *et al*. Androgen glucuronides, instead of testosterone, as the new markers of androgenic activity in women. *J Steroid Biochem Mol Biol* 2006;**99**:182–8.
- 10 Labrie F, Martel C, Bélanger A, *et al*. Androgens in women are essentially made from DHEA in each peripheral tissue according to intracrinology. *J Steroid Biochem Mol Biol* 2017;**168**:9–18.