Why sports officials may choose not to fight performance-enhancing drugs

Kjetil K. Haugen\textsuperscript{A} and Pavel Popela\textsuperscript{B}

\textsuperscript{A}Faculty of Logistics, Molde University College, Specialized University in Logistics, Molde, Norway
\textsuperscript{B}Faculty of Mechanical Engineering, Brno University of Technology, Brno, Czech Republic

Abstract
In this article, we present further discussion provoking explanation, why the existing fight against doping in sport is not fully successful although widely presented in media. Our paper is based on economical arguments derived from discussions of practitioners due to lack of statistically valid data and their fundamental mathematical modeling. Hence, the maximization of a (two-variable) sports attendance demand function, depending on athletes’ performance and doping prevalence, may result in a positive optimal doping prevalence and explain the existing sport doping related situation. Given reasonable assumptions on relevant functional behavior, this result can be interpreted as an incentive for sports officials to allow (and even welcome) some doping. As a consequence, we conclude that one should not be surprised that doping exists and is widespread under the assumption of aggregated rational behavior that is common in economical research. Therefore, the need for a global coordinated system of testing and sanctions decoupled from sport may be a necessary condition in order to meet these challenges and together with the proposed model should be the subject of further discussions.

Keywords: Performance-enhancing drugs, Optimal level of doping, Sports performance, Sports demand

1. Introduction
Demand for sport is well treated within Sports Economic research literature, as presented in Borland's [6] excellent review. Most authors seem to agree that many different factors like performance, price, substitutes, quality of stadium infrastructure, TV-coverage and so on all may have imperative effects on a sports producer's ability to fill her stadium. However, we deal with a more specific situation when demand is mainly influenced by athlete performance and doping. Doping in
sport is a hot topic and is often tackled in a provocative way and from different even opposite viewpoints, see e.g., [20] and [21]. Therefore, we follow this style of our ancestors and hope that the fruitful discussion on our conclusions will follow. Our choice is even emphasized by the fact that the needed research papers based on the statistical processing of real-world data only recently have appeared, see [9]. Hopefully, such a need will help to enhance and expand necessary statistical research in a near future. Here, we will focus on two explanatory variables; performance and doping prevalence. We will hence (deliberately) investigate a partial demand function, with the aim of analyzing the link between use of performance-enhancing drugs and (attendance) demand.

The main idea of the paper is to stimulate future discussions on the authors’ explanation of the current situation specified by the fact that doping is not suppressed enough although the existing strong and permanent media campaign creates an impression that such expectations should be fulfilled, see also [15]. Therefore, we have found that the discussion about the relationship among attendance, performance and drugs is and will remain challenging, especially under conditions where useful empirical evidence is missing or is specific or rare. So, we try to tackle the challenge and present the possible explanation based on mathematical modeling, and relying on economical arguments leading to thoughtful assumptions derived from discussions among researchers and practitioners due to lack of statistically valid data. As our considerations are based on economical motivations, we deal with the aggregated characteristics avoiding uncertainty of some data. Therefore, we assume that our initial simplified attempt to give an explanation to the stated problem can be further detailed in the future.

The following notation is used; \( A \) is attendance demand, \( P \) is performance, while \( d \) is the doping prevalence (measured absolutely\(^1\)) in some athlete population. Dependence of \( A \) on \( P \) and \( d \) is specified by a bivariate real-valued function \( f \) as follows:

\[
A = f(P,d).
\]  

We assume the following properties of \( f \) under condition on differentiability of \( f \) (where \( f_j = \frac{\partial f}{\partial j} \) for simplicity)

\[
f_P > 0 \text{ and } f_d < 0.
\]  

Before we discuss differentiability-based assumptions, we have to notice that the detailed discussion is contained in Section 2, where the non-differentiable case is also discussed. The first assumption in (2), \( f_P > 0 \), seems uncontroversial to us, as it means that demand is increasing in performance. Most experts (seem to) agree that higher jumps, faster running or longer throwing attracts more audience. More complex discussions can be found, e.g., in [13] and data processed by media can be seriously studied.

The second one is perhaps also uncontroversial, although perhaps not as much discussed, see [9] for the first study in this direction. Let us also remember the well-known Tour de France doping cases as the Lance Armstrong case that resulted in the several years break of delivery of the related TV program in Germany, so the audience decreased significantly. The meaning of \( f_d < 0 \) is that if the total use of performance-enhancing drugs increase in a sport (or a population of athletes), the audience should

---

\(^1\) The term absolutely, indicates that the doping prevalence is not measured as a fraction. In this setting the possibility of both more dopers as well as stronger substances is convenient.
react negatively, given a certain performance level. Most arguments related to doping and its individual effect on demand seems to correspond with such an assumption.

Finally, and importantly, we consider a link between performance and doping. The whole point of applying performance-enhancing drugs is that they should increase individual performance, and hence, also an increased total performance level in the population. Consequently, this relation is described by a univariate real-valued function $g$ as follows:

$$P = g(d) \quad (3)$$

And, we assume that it should exist. Furthermore, as indicated above, $g'(d) > 0$ for a differentiable $g$ within a certain closed interval $[0,D]$ where $D$ is an upper bound (see Section 2) that is taken to anticipate critical comments saying that after a given level of $d$, the performance either saturates or declines. In addition, we may also say that at some point $d$, doping starts getting dangerous, killing people, and hence, it will not improve performance. However, the athletes may apply doping up to a threshold where it stops seeing performance productive, then they may move to another substance or technique, which again improves performance. That is, performance increases in the population if more athletes use doping or a given number of athletes get access to performance enhancing drugs of “higher quality”.

2. Theoretical framework
Consider the following two-variable unconstrained optimization problem:

$$\max_d f(g(d), d), \quad (4)$$

where $d \in [a, D], D < \infty$, so the domain for $d$ is a closed and bounded interval i.e. a so called compact set [2].

Obviously, given knowledge of $g(d)$, the bivariate real-valued function $f$ collapses to the univariate real-valued function $f^*$ where $A = f(P, d) = f(g(d), d) = f^*(d)$. Suppose the following is known about $f$ and $g$ as they are differentiable functions:

$$f_{g(d)}(g(d), d) > 0, f_d(g(d), d) < 0 \quad (5)$$
and $g'(d) > 0$

on the defined domain of variable $d$.

Thinking somewhat practically on the situation, it is always possible to argue that given no doping ($d = 0$), a certain minimal performance level produces no (real$^2$) demand. We all know sporting activity that produces a minimal amount of spectators, and in this setting, we may safely say that this kind of demand is effectively zero. Consequently (given a suitable choice of $P$ - measurement), we can safely say that:

$$A_0 = f(g(0), 0) = 0 \quad (6)$$

That is, (in a doping free world), no performance leads to no demand in (6). Now, assume that the doping prevalence ($d$) is increased from zero and upwards. The consequence, given the assumption of $g'(d) > 0$ is that performance $P$ increases. Logically, at some point it must produce positive demand (given the assumption $fP > 0$). The fact that $f_{g(d)} < 0$ does of course not lead to negative demand, which is impossible in this

$^2$ By real here, we refer to professional sporting activity with reasonable willingness to pay from spectators that not necessarily are either friends or relatives of the athletes.
situation. Hence, for a certain sufficient level of drug abuse say for \( \hat{d} \), demand shifts from zero to a positive amount. Now, two functional values are identified:

\[
f(g(\hat{d}), \hat{d}) > f(g(0), 0)
\]  

(7)

and consequently, the optimal doping prevalence (not necessarily \( \hat{d} \)) must be strictly positive by (7). That is; \( d^* > 0 \). Q. E. D.

The Weierstrass extreme value theorem (see, e.g., [2]) can be introduced for our purpose (to shed some more light on special cases) as follows: if a univariate real-valued function \( f^* \) is continuous on the closed and bounded interval \([0, D]\), then \( f^* \) must attain a global maximum at least once. The satisfaction of assumptions of the theorem is obvious in our case as continuity of \( f^* \) is implied by differentiability of \( f \) and \( g \) functions.

An interesting generalization of our idea to achieve \( d^* > 0 \), such that \( f^*(d^*) \geq f(d), d \in [0, D] \), can be discussed for the following circumstances. We still assume that \( f \) is an increasing function with respect to \( P \), \( f \) is decreasing with respect to \( d \) and \( g \) is increasing with respect to \( d \). However, we assume that differentiability of \( f \) and \( g \), and hence, continuity of \( f^* \) is no more guaranteed and \( f^* \) is a piece-wise continuous function with existing left-sided or right-sided derivatives in each point \( d \in [0, D] \). Such a feature represents a situation when the change of \( d \) may imply a finite jump in a change of \( A \). Especially, it may appear for the change of \( d \) from zero to a non-zero value with a similar effect as a set-up cost in operations research models. Therefore, discontinuity may occur in 0 and \( \lim_{d \to 0+} f^*(d) > 0 \) may cause that even no global maximum exists; in a situation when the limiting value is greater than any function value of \( f^* \). In such a case, we may redefine function values of \( f^* \) in discontinuity points in such a way that \( f^* \) becomes an upper semi continuous function (see, e.g., [2] for details), and therefore, the assumptions of the generalized Weierstrass theorem, where continuity can be replaced by upper semi continuity, are satisfied again. Thus, a global maximum of a bit redefined upper semi continuous \( f^* \) on \([0, D]\) exists. However, for the redefined function \( f^* \), we have \( f^*(0) > 0 \) and the previous proof idea cannot be applied. Still, with an additional practically interpretable assumption, we can guarantee that \( d^* > 0 \). Let us check the existing right-sided derivative of \( f^* \) in 0 i.e. \( f_d^*(0) \). It can be computed as:

\[
f_d^*(0) = f_{p+}(g(0), 0)g_{d+}(0) + f_{d+}(g(0), 0).
\]

We may conclude that a very small amount of doping \( d \) has initially almost no negative impact on demand of audience \( A \) through \( f \), and it is obviously dominated by the positive impact of increased performance on the demand, see (8). So, \( f_d^*(0) > 0 \) and \( f^* \) is increasing in 0. Thus, there is a \( \hat{d} \) value as in the original proof above, and hence, a global maximum \( d^* > 0 \).

3. Consequences of the theoretical framework

3.1. Incentives of sport officials and league owners

In section 2, it is shown that given reasonable assumptions (either traditional or more general) on functions, \( f \), and \( g(d) \), the optimal doping prevalence exists and is positive. This result is easily interpreted as existence of incentives for sport officials (or league owners in the US) for existence of some doping in sport. Such a result may be
considered important if fighting doping is considered important. If those who manage sports have economic incentives to keep some performance-enhancing drugs alive and present, one obviously could ask if leaving decisions related to minimizing doping to the sport itself is a good idea. As of today, WADA (World Anti-Doping Agency) plays an important part in global anti-doping work, and WADA might be considered more a part of sports than efficiently an “outside agent”. WADA is for instance 50% financed by the Olympic movement and contains for instance an athlete committee [1]. As a consequence, strong forces dealing with sports itself play decisive parts in WADA decision making. A recent case, involving a German proposition of criminalizing doping use was for instance strongly opposed by the present WADA president, Sir Craig Reedle [18]. A seemingly paradoxical statement if the main objective is to fight doping with maximal force. Therefore, the critical opinion of IAAF chairman candidate Sebastian Coe expressed in newspapers in July 2015 does not surprise even in the situation when doubts are about many results of doping tests just before World Championship of Athletics in 2015.

3.2. The economics of doping paradox
Economics of doping, normally meaning game theory applications to investigate various questions related to use and prevention of performance-enhancing drugs, has grown considerably over the past years. Breivik [7], [8], perhaps the first to treat the doping problem through game theory, has later been followed by many researchers [14], [16], [10], [11], [3], [5] and [4].

Common to all these works is a clear notion of the difficulties involved in anti-doping work.

Following Haugen [14], such difficulties can perhaps be summed up through the following inequality:

\[
\frac{1}{2} a > r c
\] (9)

In his extremely simplified doping game, \( a \) is the reward related to winning a competition between two equally good athletes, \( r \) is the probability of being exposed as a drug abuser, while \( c \) is the cost associated with such an exposure. Given acceptance of the direction of the inequality sign in (9), the solution predicts maximal doping; everybody take drugs.

What is of interest here, is the link between anti-doping work up to now, and the inequality (9). Largely, It seems safe to say that most anti-doping efforts have been focusing on \( r \). That is, WADA’s main objective seems to be focusing on increasing this \( r \) through more, as well as better doping tests. This is especially weird, as the costs involved in such a strategy, obviously are much higher than alternative strategies. In order to turn the inequality sign around in inequality (9), both decreasing \( a \), or increasing \( c \) are equally interesting strategies. The fact that decreasing \( a \) might have adverse effects due to necessity of maximal athlete effort (see for instance [22]) is both understandable and acceptable. However, why sanctions still are kept ridiculously low is much harder to understand. Especially, as the normal punishment related to doping exposure is competition withdrawal over a certain time period, obviously constructed in a

---

3 Although several other authors also (as we do) discuss the possibility of optimal doping ([14], [12], [17]), our modeling approach is different and provides (as we see it) added insight.

4 Which according to Haugen [14] would remove doping.
different and more amateurish time period. This is what might be named The Economics of Doping Paradox. Surely, if incentives are such (as we claim here) that sport “owners” gain from certain doping levels, such a paradox is perhaps not paradoxical at all. The case briefly mentioned above; related to WADA's reaction to the German proposition of harsher sanctions, is then perhaps easier to understand.

4 Conclusions
At first, we have identified a problem that is mostly discussed by fans, athletes, newspapers, less by researchers till now, and it can be defined as: "Why the existing fight against doping in sport is not fully successful although widely presented and criticized in media?"

An additional problem appears due to the sensitivity of doping problems. There is quite a lack of data and related statistical studies. (An interesting application by Petroczi and Haugen [19] gives good arguments why one should expect such a lack of reliable data.) Still, the stated problem to be discussed has remained in front of us – very visible in media. So, we have felt challenged and even without data support, we have asked ourselves whether the ratio-based explanation can be found to help answering the before mentioned question.

Our major point in this article is very simple. If sport officials or league owners are involved in major decision making related to anti-doping work, and have incentives for a positive doping prevalence, one should not expect a victorious fight against doping. As such, globalization and sport decoupling of anti-doping work seems a clear necessity.

One simple suggestion could perhaps be to exclude the Olympic movement from WADA financing, as well as taking great care of the structure of athlete participation in relevant decision-making.

Our point is of course not to suggest that athletes' opinions related operational and strategic decisions in anti-doping work should not be heard, but merely to indicate that when it comes to actual decisions, great care should be taken related athlete participation.

Many researchers, including Maennig [15], believe that abuse of performance-enhancing drugs is among the most critical and threatening problems in sport. Failing to handle these problems may undoubtedly have critical implications for the future of professional sport. As a consequence, if decoupling sport itself from anti-doping work can lead to the obvious solution - stricter sanctions - this is what we should do.

Acknowledgement: We address very serious thanks to two anonymous referees whose critical comments and challenging words led us to significant improvement of our ideas, and hence, the presented paper. The present work has been supported by the specific research project no. FSI-S-14-2290, id. code 25053 and by NETME CENTRE PLUS (LO1202) with support from the Ministry of Education, Youth and Sports under the National Sustainability Programme I.
References


[20] J. Savulescu, B. Foddy, and M Clayton. Why we should allow performance
