Anti-doping Policies as an End-game-effect Trigger*

Jakub Šindelář† Liam J. A. Lenten‡

February 20, 2020

Abstract

Previous literature on the economics of doping has argued that bans as a form of deterrence might become ineffective as athletes approach the end of their careers. The model presented here shows that the incentives to dope (by a rational athlete) are rising over time, reaching the maximum threshold shortly before retirement. Our theoretical prediction was tested on two main data sources: (i) publicly available doping data in track and field; and (ii) a leaked-blood-samples database by the IAAF. Estimates generally show a significant relation between age and doping propensity. Since the current punitive bans do not produce the desired outcome, alternatives should be considered, such as conditional superannuation. As a rare empirical application within a field setting, our findings contribute to the more general literature on cheating in contests, by forming a useful complement to the otherwise ubiquitous presence of experimental studies.

Keywords: Cheating, Policy, Punishments, Doping, Sports
JEL Classifications: C21, L83, Z28

*This paper is based on research undertaken when Šindelář was a PhD candidate in the international program in Institutions, Economics and Law. We are grateful especially to Kjetil Haugen and the participants of: (i) the 1st International Conference on Sport Economics & Sport Management (Berlin, 2016); and (ii) the 9th ESEA European Conference on Sport Economics (Paderborn, 2017), for their helpful comments.

†University of Turin, ITALY (Email: xsinj16@gmail.com).

‡Corresponding author: Department of Economics and Finance, La Trobe University, AUSTRALIA (Email: l.lenten@latrobe.edu.au).

1 Introduction

Viewed traditionally as a practitioner problem, doping in professional sport is now an emerging academic topic within economics, given its intuitive interest as a form of cheating within a competitive state. The perennial infamy of doping and corruption in sport continue to occupy the front pages of all major newspapers and other media. Scandals, such as the seminal case involving US cyclist Lance Armstrong, undermine public opinion in the current anti-doping enforcement system. They also show that the problem is not confined merely to a few recalcitrant athletes themselves, but is rather sometimes part of a broader orchestrated regime, in which the sanctioned athletes were coerced into cheating. This creates a challenge for economists to rethink the entire policy suite to make it more effective and efficient.

Within economics, much of the longer-standing research on doping encompasses two-player, one-shot games (Prisoner's-dilemma type) in which competitors choose whether or not to dope. Prominent examples from this stream of research include—but are certainly not limited to—Breivik (1992), Berentsen (2002) and Haugen (2004). More recently, scholars have started to apply alternative theoretical approaches. A study by Emrich and Pierdzioch (2015) highlighted the problem of international coordination of anti-doping policies; while Berentsen et al. (2008) showed the importance of encouraging whistle-blowing to increase the likelihood of detection, thereby disincentivizing cheating. Eber (2002) sees a problem in the insufficient credibility of the current anti-doping system, whereas Buechel et al. (2016) stresses the need for transparent doping tests. Even though the scope of doping literature is now extending, many related policy issues still have not hitherto been discussed and analyzed robustly.

This paper takes a closer look at one specific and important question that stems from the current system of penalties, which is based almost solely on the reliance of bans (as well as fines, selectively). As previously mentioned in the literature (e.g. Maennig, 2002), such a system can distort the level-playing field of policy effectiveness in age-profile terms—it may not sufficiently deter older athletes (specifically) who are in the final stages of their careers, creating numerous policy implications for effective mitigation mechanism design.

The basic intuition follows that towards the end of her career, a rational athlete will no longer carefully consider the threat of punishment (being banned from competition) in her doping choice. That is to say, she will bear no costs of doping other than the direct monetary (obtaining substances) costs and indirect

costs, such as future health consequences (all of which are deemed by many athletes to be negligible).[1] This paper is motivated by shedding light on the end-game effect, both theoretically and empirically, especially so given the ***striking lack of empirical evidence*** (termed as "rare" by Dilger et al., 2007) on doping underpinning this phenomenon.

The remainder of this paper will proceed in the following manner. Section 2 will comprehensively outline the advantages and limitations of the empirical approach against the backdrop of other disciplinary doping research. Section 3 will then present a relatively parsimonious model, allowing us to make some inferences on what the theory would predict. Then, section 4 lays out the hypotheses to be tested and provides precise detail about the unique athlete-testing databases we have constructed. Following that, section 5 presents our regression results; and finally, both a brief discussion and conclusions are presented in section 6.

2 Empirical Background and Approach

While "striking" as noted above, in a certain sense the dearth of empirical evidence on doping is nevertheless unsurprising to economists in particular, given that obtaining representative data from the field has always been fundamentally difficult. More generally, this is also the case with respect to numerous other forms of "cheating" that are of great interest to the discipline. To this end, a common approach in modern times to overcome this challenge is the widespread use of economic experiments to observe cheating accurately in an (albeit purely) internally valid manner.

Such applications of cheating in the literature range in behavioral forms from deception and lying (e.g. Gneezy, 2005) to tax evasion (e.g. Bayer and Cowell, 2009) to sabotage of peers (e.g. Harbring and Irlenbusch, 2008), but this approach has more recently since been extended similarly to the revelation of doping. The experimental design of Wu et al. (2016) included four different enforcement treatments: (i) the no-punishment benchmark; (ii) fine; (iii) ban; and (iv) a conditional superannuation scheme. As predicted, the highest prevalence of dopers was observed in the final (third) round of the ban treatment,

[1] We note that, though it is actually time until retirement that strictly characterizes the end-game effect, this cannot be measured (in fact, not even estimated) reliably in practice. For this reason, athlete age is used instead as a proxy (inversely so) for this.

strongly indicative of a clear end-game effect in the lab. The most effective deterrent was the conditional superannuation scheme.

Despite its methodological advantages, experimental economics nonetheless inherently lacks external validity, being unable to replicate the precise conditions under which the real-world doping phenomenon transpires. Therefore, the use of actual doping test data unambiguously offers a lot of additional value in addressing our research questions. However, given the limitations of these data, as discussed henceforth, we do not necessarily claim our results to be a categorical improvement on the experimental economic approach. Rather, they serve as a highly useful complement to the (far more substantial) existing volume of experimental studies of cheating.

At a basic level, doping test data historically were not generally available publicly, largely because of the array of athlete privacy provisions involved. Nevertheless, data on *positive* tests would effectively become transparent, since the dealing of sanctions is impossible to conceal.[2] Therefore, it would be possible to compare the mean age of known offenders within a competitor cohort (e.g. from a specific competition or event) to that of the overall athlete sample, via uncontrolled t-tests. However, this method fails to account for the possibility that the age-distribution of athletes chosen for testing may be different (in reality, most likely older) to that of the overall cohort.

Therefore, a modeling strategy that fully controls for exogeneity (and possible endogeneity) is required. Dilger and Tolsdorf (2004) looked at elite-level 100-meter runners, and concluded that athletes who had tested positive are significantly older (by about 3 years). However, they used a by-athlete sample centring on a positive doping test at any time during the athlete's career, meaning it lacked contemporaneity with the sample period selection. Coupé and Gergaud (2013) overcome this deficiency by taking a single snapshot in time, sampling the full 2010 Tour de France field. They find qualified evidence of a positive doping relation with age; however, unfortunately their dependent variable was not based on positive tests, but rather merely an "index of suspicion" as assessed by cycling authorities. Moreover, neither study could leverage large enough a sample (fewer than 200) to draw any strong policy-relevant conclusions.

[2]By way of analogy, with tax evasion detection, prosecution can be easily unveiled; however, many tax authorities may strike plea deals involving full (or even partial) repayment of taxation revenue shortfalls, coupled with similar information suppression privileges.

In this light, the most important unique contribution of the current paper to the academic research on doping is to test the theoretical predictions using two alternative large and comprehensive databases. The first one consists of doping tests (positives and negatives) in US track and field, while the other is a famous leaked database of World Athletics (IAAF) blood samples (more details follow in section 4).

The latter of these in particular affords us a methodological advantage over the results from experiments. Due to complexity elsewhere, Wu et al. (2016) are forced to treat doping as dichotomous for the sake of model tractability, whereas the IAAF data can unravel varying degrees of extremity in test results. This is a most useful characteristic, not only because of the extra dimension it provides on the end-game effect via extreme doping levels from athletes simply taking the gamble of not being tested, but also given the wealth of anecdotal evidence on widespread systematic "micro-doping", so as still to return a negative test. This feature reinforces our earlier point on external validity. In fact, the pros and cons of the mere regulation of such a proposed arrangement (as opposed to the omnipresent zero-tolerance approach) are debated by Savulescu et al. (2013).

More generally, both of these databases overcome the multiple data selection issues discussed before, leaving only a few minor issues with our data, which are addressed in section 4. In addition, the possibility of reverse causality could be thought of as an issue (at least in the first part of the empirical analysis). More specifically, than an increase in the rate of positives in any particular cohort might be followed by an increase in testing frequency within that cohort. Despite this, a decomposition of our empirical results (presented in the Appendix) demonstrates that our estimates of the shares of positives over the range of older age groups were somewhat underestimated.

In a not unrelated point, the precise accuracy of the doping tests themselves in revealing underlying doping activity (i.e. the perceived prevalence of Type-I testing errors), might also seemingly appear to create a validity issue. However, as considered by Lenten et al. (2017, p. 598), it is likely that older, more-experienced athletes are better at: "*...averting suspicion or concealing use*". Under this state of nature, should the results demonstrate a positive relation between athlete age and doping propensity, then the estimates of the sensitivity parameter(s) relating doping to age could actually be considered as the lower bound of the actual effect. Thus, the true strength of the relation would be

underestimated (not overestimated), assuaging the concern of test reliability potentially nullifying such a qualitative finding.

And as a preview, it will indeed be shown in section 5, that the estimates support this intuition of the theoretical model—in two complementary ways: (i) older athletes return a positive test result more often (on average) than their younger counterparts; and (ii) they also tend to register more extreme blood values. Even though the blood-tests database allowed us to estimate a fixed-effects model, one still has to be cautious with a causal interpretation of the results. Nevertheless, these results are still important and interesting for policy makers in the industry, since it can help them to optimize testing procedures.

3 Theoretical Model

In order to make theoretical predictions on the age-profile of doping, we now present a theoretical model of a single subject (professional athlete), who is a rational representative agent—optimizing her level of doping throughout her career with respect to expected "profits". While the stand-alone non-pecuniary motivations for sporting success are certainly non-trivial, the utility of these can be encapsulated into a monetary-equivalent value, and simply included in the profit function—doing so will not change the thrust of any of the results.

The model is a simple one of an intertemporal-choice nature, which is designed with the intention of capturing some salient "real-world" aspects of a typical doping-decision scenario. Time has two effects on the optimal level of doping. The first is the aforementioned end-game effect. An athlete who is in the final round (year) of her career purely objectifies the present period and dopes maximally with respect to the probability of being caught; whereas in all other periods preceding the final one, she also has to consider possible future losses.

The second effect relates to the current trajectory of the athlete's performance in the time domain. Critically, in a similar fashion to Wu et al. (2016), an increasing current trajectory of natural performance (ostensibly when the athlete is still relatively young) is associated with lower incentives for athletes to cheat in the present period, since it raises the future costs of doping.[3] By contrast, a decreasing natural-performance time-trajectory, in the latter years

[3] In this model, "future" refers purely to $t + 1$. In practice, however, these costs could extend into multiple future periods in the event that the absence from competition during the ban itself diminishes the athlete's subsequent performance level below the baseline.

of her career, will signify lower (and declining) possible future losses of being banned.[4] The model predicts that doping effort will be rising throughout her career at an increasing pace (with exceptions in certain model configurations).

Let us assume the athlete is risk-neutral, and in a non-strategic situation—she can induce precisely the effects of doping on her profits (that is, facing a 1-year ban in the event of a positive test). In each round t (year) she can choose to dope $d_t \in [0, 1]$ to enhance her natural performance, P_n, such that the final performance is given by $P_n(1 + d_t \cdot \varphi)$, where $\varphi \in (1, \infty)$ stands for the effect of doping on her final performance.

For a given level of natural performance, there is a level of profit $\widetilde{\pi} = f(P_n)$, which would be earned without taking performance enhancing drugs (PEDs). Now, without loss of generality, we assume that $P_n = \tilde{\pi}$. The athlete's possible earnings are therefore given by $\widetilde{\pi}_t(1+d_t \cdot \varphi)$, where the coefficient φ now denotes directly the effect of doping on earnings.

Now, let us model the costs of doping. The athlete has no direct costs of obtaining doping substances (these are commonly supplied freely to the athlete) or other costs (mainly health deterioration).[5] She may only lose current and future profits when tested positive. The choice to dope depends not only on the costs of doping, but also on the probability of being caught $r \in [0, 1]$, which is defined as $\Pr(caught|d = 1)$—that is, the chance of being caught when the athlete dopes maximally (i.e. when $d_t = 1$).

Doping creates possible costs in two periods. The athlete risks losing her profits from the period in which she is competing (henceforth, we term this as a "fine"), and also in the following period while serving the ban.[6] Her aim is to choose a level of doping that takes all of the current and future expected losses into consideration. As is regularly assumed, she also faces a discount factor, δ.[7] The costs of doping are given by a product of the probability of being caught,

[4] Despite not being modeled here, additionally at this stage, one can easily imagine that psychologically, athletes may become increasingly desperate to arrest this decline, and as such experience a heightened propensity to resort to doping.

[5] The Goldman Dilemma results discussed by Lenten et al. (2017, see pp. 592-593) suggest that athletes place a negligible cost on adverse future health outcomes for themselves—at least, relative to the prospect of sporting success.

[6] By general definition, a "fine" need not necessarily be connected in any way to the level of profit. However, we assume a specific form of fine herein, which equates strictly to $(1 + d_t \cdot \varphi)\widetilde{\pi}_t$, after being determined (since it is an *ex-post* type of punishment, coupled with disqualification).

[7] Conventional wisdom on the mindset of professional athletes suggests, just as the results of Carlson et al. (2015), that we might expect δ to be unusually large in comparison to most other standard intertemporal-choice-model scenarios.

r, and the level of doping, d. It is also assumed there are no Type-II testing errors (i.e. false positives).[8]

The decision she is facing takes the following form:

$$\max_{d_t} E\pi_t = \overbrace{(1 + d_t \cdot \varphi)\widetilde{\pi}_t(1 - d_t \cdot r)}^{\text{Time of decision } t} + \underbrace{(1 - d_t \cdot r)\delta \overbrace{[(1 + d_{t+1} \cdot \varphi)\widetilde{\pi}_{t+1}(1 - d_{t+1} \cdot r)]}^{\text{Time of decision } t+1}}_{\text{Expected future costs } (t+1), \text{ given } d_t} \tag{1}$$

Alternatively, in a form stressing the negative effect of doping on both current and future costs:

$$\max_{d_t} E\pi_t = (1 - d_t \cdot r)[(1 + d_t \cdot \varphi)\widetilde{\pi}_t + \delta[(1 + d_{t+1} \cdot \varphi)\widetilde{\pi}_{t+1}(1 - d_{t+1} \cdot r)]] \tag{2}$$

The first right-hand term in equation (1) is the expected profit from competition at t, and the second term represents the possible loss of profit in the following period (if the athlete is banned). The choice variable is d_t, which affects her performance in period t, therefore enhancing her "natural profit" $\widetilde{\pi}_t$ by increasing her performance relative to all other competitors. On the other hand, a higher d_t comes also with higher costs. Doping in time t plays a crucial role in determining the expected profit from the next round $(t+1)$—specifically, the higher is d_t, the lower is the expected return.

Deriving the first-order condition and solving for the optimal level of doping at time t yields as follows:

$$d_t^* = \frac{\overbrace{\delta({d_{t+1}}^2 \cdot r\varphi + d_{t+1} \cdot r - d_{t+1} \cdot \varphi - 1)}^{\beta}}{2\varphi} \frac{\widetilde{\pi}_{t+1}}{\widetilde{\pi}_t} + \overbrace{\frac{\varphi - r}{2r\varphi}}^{d_T^*} \tag{3}$$

The first observation from equation (3) is that if an athlete finds herself in the final period of her career (i.e. $\widetilde{\pi}_{t+1} = 0$), the level of doping d_T^* simply collapses to $\frac{\varphi - r}{2r\varphi}$, which implies that for non-negative levels of doping, the condition $\varphi \geq r$ must hold. This is always the case, since it is assumed that doping has a positive effect on performance, and therefore, on income ($\varphi > 1$).

If either $d_{t+1} < 1$ or $r < 1$, it holds that the rate of doping is highest in the final round, and in all previous rounds the optimal level of doping is below that

[8] The probability of a false positive is close to zero in practice, because for each doping test two samples are stored, "A" and "B". If sample A is positive and the athlete exercises her option to dispute the result, she can ask for re-test with sample B. If sample B is negative, the first test is automatically considered to be a false positive.

level, because the first right-hand term in equation (3), collectively renamed β, is negative and represents a deviation from the final period's doping level.[9]

The quotient of future-to-current natural profits is an important determinant of d_t. All else being equal, the higher is $\widetilde{\pi}_{t+1}$, the lower is the optimal level of doping at time t. This stems from the fact that $\beta < 0$. Intuitively, if an athlete knows that she can be more successful in the next year than in the present year, doping now is quite risky—not so much because she can lose profits in this period, but rather because she can lose greater profits in the following one. A higher future level of doping (d_{t+1}) leads, *ceteris paribus*, to lower current PED consumption if $d_{t+1} < \frac{\varphi - r}{2r\varphi}$ which, as shown before, always holds given that either condition $d_{t+1} < 1$ or $r < 1$ is satisfied.[10]

The model produces one counter-intuitive result—in certain situations, a higher probability of being caught r leads to higher levels of doping. Since a higher r means lower doping in $t+1$, it also means lower costs of doping in time t. This is obvious when looking at the final period, in which the effect of growth in r always translates into higher deterrence (i.e. a lower d_{t+1} therefore lowers the "future" costs of doping in the previous period).[11]

To recap, the model has demonstrated that there are two effects leading to higher levels of doping in the latter stages of athlete's career. The first is the end-game effect, given by d_T^*; and the second is the time-trajectory-of-performance effect, given by $\frac{\widetilde{\pi}_{t+1}}{\widetilde{\pi}_t}$. The faster the rate at which one's natural performance is declining, the higher the levels of predicted doping are.

4 Research Questions and Data Particulars

Following the thrust of section 2, our empirical approach is one that, importantly, accounts for any age-profile selection biases in the sample. It also allows us to reveal additional information about *extreme* doping levels in the blood samples. To this end, we raise two complementary empirical questions in testing the end-game effect:

1. Is it more probable that a doping test will be positive for older athletes?

[9] Proof: In order for the first right-hand term of equation (3) to be negative, $({d_{t+1}}^2 \cdot r\varphi + d_{t+1} \cdot r - d_{t+1} \cdot \varphi - 1) < 0$ has to hold. This inequality holds if $d_{t+1} < 1$ or $r < 1$.

[10] Proof: In order for this statement to hold, the first derivative of d_t with respect to d_{t+1} has to be negative: $\frac{\partial d_t}{\partial d_{t+1}} = \frac{\delta}{2\varphi}(2d_{t+1} \cdot r\varphi + r - \varphi)\frac{\widetilde{\pi}_{t+1}}{\widetilde{\pi}_t} < 0 \implies d_{t+1} < \frac{\varphi - r}{2r\varphi}$

[11] Since $r \in [0, 1]$ and $d_t \in [0, 1]$, in certain situations, a lower r does not change the outcome since doping is already at its maximum level.

2. Are extreme blood values more prevalent among older athletes?

Answering the second empirical question is possible using the leaked IAAF blood-samples database. To answer the first question, one would ideally have an individualized database consisting of all doping tests for the entire population of athletes in a given sport discipline over a certain time period. For an initial uncontrolled comparison, one could then simply compute the prevalence (denoted henceforth as $PREV_i$) of positive doping tests across each age group i. More formally:

$$PREV_i = \frac{POST_i}{TOTT_i}$$

where $POST_i$ denotes the number of positive tests in i, and $TOTT_i$ the corresponding total number of tests carried out. Unfortunately, the nature of the publicly available data does not allow us to compute $PREV$ precisely, because no comprehensive database of all doping tests exists.

The best public data currently available consist of two types: (i) a list of tested athletes by both IAAF and USADA (United States Anti Doping Agency) over the period 2009-2012; and (ii) a list of positive doping findings in athletics over the period between 2004 and 2012. Further details on the structure of the data will be discussed in the following section.

Next, we provide some basic age-profile comparisons to shed light on the structure of our data, then we will estimate the prevalence of positives among age groups via a common standardization technique known as Standardized Mortality Ratio (SMR), and in the final part we will discuss the estimation of a range of regression models.

This paper exploits both alternative types of data. At first an analysis of positive doping cases is made and subsequently also the IAAF leaked blood-values database is explored. The following paragraphs and this section will deal with the first type of data, positive doping tests.

For the purpose of analyzing the correlation between age and positive doping tests, four datasets were employed. The URL: www.dopinglist.com is currently the most comprehensive database of doping cases. For the purpose of this research, we collected 1,164 cases in track and field throughout the period from 2004-2014. Those are cases recorded throughout all national (NADAs) and other

anti-doping agencies.[12] For the sake of standardization of the distribution of positives, we used two sources. The main dataset for this purpose is comprised of 3,276 out-of-competition (OOC) tests by the IAAF anti-doping program in years 2008-2012. The secondary dataset contains a list of 2,029 OOC and in-competition tested (ICT) athletes by USADA.[13] Since the publicly provided data do not contain information on age of athletes, these had to be compiled separately with the use of the official IAAF athlete database, which contains information on approximately 110,000 athletes.[14]

The number of athletes in all datasets was reduced in the process of assigning information about the age of athletes, as not all positively tested athletes have been found in the official IAAF database. Ultimately, 417 positives in track and field over the years 2004-2012 were taken into consideration, as were 2,635 tested athletes by IAAF throughout the years 2009-2012, as well as 1,598 by the USADA testing program.

There are several minor issues with the data used. The main issue is that the subsample of all tested athletes is not random, therefore a potential bias has to be considered. Throughout my analysis I take the IAAF testing sample as the main one and check the validity of my results with the control sample by USADA.

Another issue is that the information on the number of tests per athlete is available only for the USADA database (for each athlete a number of tests in given year). The IAAF source assigns tested athletes into one of two groups as such: (i) 1-3 tests a year; and (ii) ≥ 4 tests. A bias could emerge if athletes of certain age groups are tested more often. Data by USADA and partly also data by IAAF can help us to identify this bias.

Table 1 summarizes the available data, it shows the number of observations, its share in the sample and a Standardized Mortality Ratio (SMR), which is commonly used to express mortality of a study cohort with respect to the general population. In this case, SMR is an indicator of the prevalence of positive doping tests in a study cohort (age_i). Formally, we have:

$$SMR_i = \frac{OBS_i}{Expected_i}$$

[12]Not all positive doping cases are included in the database, as even in the modern day, some NADAs and their underlying countries have strict privacy laws that do not allow the revelation of doping cases publicly.

[13]Sources: http://www.usada.org, http://www.iaaf.org/about-iaaf/documents/anti-doping

[14]Source: http://www.iaaf.org/athletes

where OBS_i denotes the number of observed doping cases (from the official data), while EXP_i is the number of positives that would have been realized if the prevalence of doping is spread evenly between age groups. This was computed by taking testing numbers (by IAAF and USADA) into consideration.

Looking at table 1, one can observe that both younger and older age-groups (above 30) tend have a higher SMR than the mid-career athletes. However, the small number of positive doping cases in some of these individual age-groups makes it difficult to draw any strong conclusions on the basis of this simple uncontrolled comparison.

Table 1: Distribution of Positives and Samples by Age-group

Age	Positives	%total	IAAF	%total	SMR	USADA	%total	SMR
18	8	1.92	63	2.40	0.80	71	4.57	0.42
19	20	4.80	113	4.31	1.11	198	12.73	0.38
20	12	2.88	70	2.67	1.08	31	1.99	1.44
21	20	4.80	109	4.15	1.15	44	2.83	1.70
22	25	6.00	140	5.34	1.12	78	5.02	1.20
23	26	6.24	163	6.21	1.00	101	6.50	0.96
24	28	6.71	186	7.09	0.95	116	7.46	0.90
25	30	7.19	220	8.38	0.86	134	8.62	0.83
26	25	6.00	237	9.03	0.66	122	7.85	0.76
27	33	7.91	225	8.57	0.92	122	7.85	1.01
28	30	7.19	197	7.51	0.96	102	6.56	1.10
29	19	4.56	171	6.52	0.70	86	5.53	0.82
30	31	7.43	152	5.79	1.28	71	4.57	1.63
31	17	4.08	116	4.42	0.92	68	4.37	0.93
32	20	4.80	104	3.96	1.21	47	3.02	1.59
33	12	2.88	91	3.47	0.83	36	2.32	1.24
34	11	2.64	52	1.98	1.33	33	2.12	1.24
35	7	1.68	47	1.79	0.94	22	1.41	1.19
36	10	2.40	43	1.64	1.46	16	1.03	2.33
37	7	1.68	35	1.33	1.26	14	0.90	1.86
38	8	1.92	37	1.41	1.36	10	0.64	2.98
$\geq$ 39	18	4.32	53	2.02	2.14	33	2.12	2.03
Totals	417	100.00	2,624	100.00		1,555	100.00	

Notes: totals are each only the sum of the 18-49 age range, not all observations.

Now we proceed to look at the entire sample to see if there is any difference in means (i.e. if those who tested positive are, on average, older). The mean age of positive doping tests in our sample is 27.56 years. This is compared to the IAAF-tested-athlete mean of 26.98 years and the USADA-sample mean of 25.91 years (i.e. there is more than a 1-year difference in the mean age between testing samples). Results in table 2 show that a Studentized t-test of the difference in

means shows a highly significant difference between ages of Positives and those in the USADA sample (p-value $= 1.104 \times 10^{-7}$). There is also a significant (albeit less profound) corresponding difference between Positives and the IAAF sample (p-value $= 0.02269$), as well as a highly strongly significant difference between the USADA and IAAF samples (p-value $= 1.108 \times 10^{-9}$).

The final comparison is crucial to understand if both samples are significantly different from each other on one hand, or interchangeable on the other hand. The latter case implies add credibility to the model estimates presented later. A closer look at the data structure reveals that there is an over-representation of the youngest age groups, relative to the expected idealized distribution. More specifically, as can be seen in table 1, the 19 year-old age group in both the IAAF and USADA samples poses an anomaly. In the latter case, 19 year olds are tested nearly three times more often than 18 year olds and over six times more often than 20 year olds. This is explained by the one-off effects of intensive testing in the lead-up to and during the (junior) World Athletics U20 Championships.

Truncating the sample to include only observations above age 20 reveals a comparatively different picture. Table 2 summarizes all correlation coefficients and t-tests for both the full sample and restricted sample (20+). A t-test clearly shows that without the outliers in the 19 year-old group, there is no difference in terms of mean age ($\mu_{iaaf} = 27.5884$; $\mu_{usada} = 27.5876$). In addition, the difference-in-means-test p-values of both samples with respect to positives are now qualitatively identical (and quantitatively approximate).

Table 2: Analysis of Age Distributions within Samples

Correlation Matrix				
	All		Age $\geq$ 20 (Only)	
	USADA	IAAF	USADA	IAAF
IAAF	0.9021	-	0.8563	-
Positive	0.9300	0.9076	0.9020	0.8978
Difference-in-means t-test (p-values)				
	All		Age $\geq$ 20 (Only)	
	USADA	IAAF	USADA	IAAF
IAAF	1.11×10^{-9}	-	0.9961	-
Positive	1.10×10^{-7}	0.2261	0.0364	0.0259

From this, we can conclude that there is a significant difference between age means of those who are tested and those who fail doping tests.

The second main source of data is the leaked-blood-samples database from the IAAF that was exposed by the Sunday Times (UK newspaper) and a documentary aired by German broadcaster ARD/WRD. The validity of the leaked data was confirmed by WADA's *Independent Commission Report* (#2), which also reveals that the database was compiled by Giuseppe Fischetto, who worked as a delegate for IAAF and European Athletics (EA) during the years 2008 to 2012. The dataset contains mainly tests made by IAAF but also by NADAs, event organizers and WADA.[15]

The database covers the period between 2001 and 2012, with 12,360 blood samples from athletes of both genders: 6,626 for men and 5,740 for women. Pre-competition tests are predominant (N=9,336), followed by out-of-competition tests (N=1,068) and in-competition tests (N=447). A total of 6,438 tests were conducted in endurance disciplines, and 3,275 in non-endurance disciplines. For the remainder, the information about discipline was not specified. The original database contained the name of each athlete; however, during the handover of the database, the names were deleted. Afterwards, an ID for each anonymous athlete was created, hence a panel regression can be estimated.

An independent commission has compared the database with data from the Anti-Doping Administration and Management System (ADAMS), to which the IAAF commenced entering data from early 2009. Of all samples, 41.5% were matched within ADAMS. However, several samples from early 2009 were not matched, since the blood analysis at these events was not performed in compliance with Athlete Biological Passport (ABP) guidelines. From the World Championships in Berlin (starting 15 August 2009) onwards, 97.3% of samples were matched in ADAMS.

The number of unmatched samples since 2009 is fewer than 141. These cases are typically caused by data entry errors, such as the wrong discipline; therefore, the actual number cannot be measured precisely. WADA experts also found 35 samples in ADAMS whereby the IAAF was the testing authority and which were not found in the database.[16]

[15] Authenticity of the database can be proved in person to anyone who is interested, upon request.

[16] Those samples were probably lost when the data were manually attached to the spreadsheet.

There are several important blood measures stored for each blood sample: (i) haematocrit level (HCT); (ii) hemoglobin (HGB); and (iii) a percentage of reticulocytes (RET). To determine suspicious blood values, the (so-called) "OFF-score" was developed. It incorporates the information from both HGB and RET into a single metric.[17] A very low or very high value is an indicator of blood doping. The "normal" value of an OFF-score lies between approximately 80 and 110. Values higher than 110 generally indicate recent doping in the blood (in preparation for competition). A low OFF-score is a sign that the body is reverting back to its natural equilibrium by lowering its production of red blood cells. Since most of the observations in the database are pre-competition tests, we will place our emphasis on the latter and check the correlation between age and OFF-score.

The average age for athletes with an OFF-score higher or equal to 110 is 26.66 years (N=917), while for the rest it is 25.67 years (N=8,419, p-value $= 2.228 \times 10^{-8}$). For OFF-scores higher or equal to 120, the average age is even higher, at 27.26 years (N=359, p-value $= 5.467 \times 10^{-9}$). Table 3 provides the relevant information for each age cohort. The following fields are reported: the number of samples in the respective cohort N, and the share of samples with respect to the whole population. The next two fields repeat these, but specifically for samples with an OFF-score higher than 110 (OFF$\geq$ 110), and the following field is the corresponding SMR. The final three fields repeat the previous three, though with the more selective sample with an OFF-score higher than 120 (OFF$\geq$ 120). It is easily recognizable from the table that higher OFF-scores are positively correlated with the older age cohorts. This correlation is particularly evident in the bottom part of the table, where the cohorts are grouped into a small number of wider age ranges. The only group in which OFF-scores are below average is the youngest (18-24). In the oldest age group (36-40), the prevalence of extreme blood profiles is 77% ($\geq$ 110) and even 130% ($\geq$ 120) higher compared to the population average.

The first conclusion based on the blood-samples database is that athletes with high OFF-scores are significantly older on average, by approximately 1 year, and the prevalence of extreme high blood values is higher in older age groups. Both sources of data point to the conclusion that doping in older age-groups is more common, however the validity of this result must be confirmed via a robust regression analysis that also considers relevant control variables.

[17] Critical-count values for determining a positive test are based on a range of scientific studies. For example, refer to Tarallo et al. (1994) for biological RET "reference limits".

Table 3: Distribution of OFF-scores and Sample by Age

Age	N	%total	OFF$\geq$ 110	%total	SMR	OFF$\geq$ 120	%total	SMR
18	159	2.00	7	0.92	0.46	2	0.62	0.31
19	607	7.63	32	4.21	0.55	11	3.42	0.45
20	284	3.57	26	3.42	0.96	9	2.80	0.78
21	451	5.67	36	4.74	0.84	15	4.66	0.82
22	648	8.15	57	7.50	0.92	17	5.28	0.65
23	507	6.37	47	6.18	0.97	15	4.66	0.73
24	563	7.08	56	7.37	1.04	15	4.66	0.66
25	599	7.53	66	8.68	1.15	31	9.63	1.28
26	577	7.25	50	6.58	0.91	26	8.07	1.11
27	591	7.43	44	5.79	0.78	12	3.73	0.50
28	516	6.49	59	7.76	1.20	30	9.32	1.44
29	473	5.95	57	7.50	1.26	27	8.39	1.41
30	405	5.09	39	5.13	1.01	19	5.90	1.16
31	369	4.64	43	5.66	1.22	18	5.59	1.20
32	279	3.51	23	3.03	0.86	11	3.42	0.97
33	222	2.79	26	3.42	1.23	13	4.04	1.45
34	184	2.31	17	2.24	0.97	11	3.42	1.48
35	148	1.86	20	2.63	1.41	10	3.11	1.67
36	110	1.38	23	3.03	2.19	14	4.35	3.14
37	87	1.09	14	1.84	1.68	7	2.17	1.99
38	47	0.59	7	0.92	1.56	3	0.93	1.58
39	41	0.52	5	0.66	1.28	3	0.93	1.81
40	16	0.20	2	0.26	1.31	1	0.31	1.54
41	19	0.24	4	0.53	2.20	2	0.62	2.60
Range								
18-24	3,219	40.47	261	34.34	0.85	84	26.09	0.64
25-30	3,161	39.74	315	41.45	1.04	145	45.03	1.13
31-35	1,202	15.11	129	16.97	1.12	63	19.57	1.29
36-40	301	3.78	51	6.71	1.77	28	8.70	2.30

5 Regression Analysis

So far, we have demonstrated that athletes who have tested positive or register abnormal blood values are generally older. Formal regression modeling will allow us to look deeper into the age-doping relationship and control for other variables that might affect doping decisions. In the first part of this section, we estimate a model using the public data on tests and positives. Regression models based on the leaked-blood database will follow.

5.1 Publicly available data

In order to run a regression, we take the testing data from the IAAF and combine them with the data on positives, forming a single one database. This strategy has two limitations. Firstly, the IAAF sample of tested athletes is not a population; therefore, we have to assume that the testing sample is representative

of the whole population. Secondly, as the share of positives is higher than in reality, this can affect the size of the estimated coefficients.[18]

The sample is restricted to age groups of 20+ years old, in order to circumvent the bias created by excessive testing of junior athletes. The period from 2009-2013 was chosen so that the data coincide in terms of time. This sample reduction leaves 245 positives and 3,218 tested athletes.

The numbers of tests per athlete were not added since they are known only for the tested athletes. A possible effect of adding number of tests into the analysis will be discussed later. Both the IAAF sample and positives contain information on the gender of athletes and their nationality.

The variables included in the following regression are $Positive_i$ (1-positive, 0-negative), AGE_i is the age of an athlete in years, $Male_i$ is 1 for a male athlete and 0 female, and YXX is a dummy for each year. Dummies for each nation represented by an athlete were also added.[19]

A simple (unreported) linear probability (LP) or Logit estimation of equation 4 produces no significant result for the Age parameter estimate. However, given the SMR age profile shown in table 1, one can observe that the relationship rather might be nonlinear, with peaks at the bookends of the age-distribution. Therefore, the model represented by equation 5 specifies a quadratic specification for Age, as well as control variables.

$$Positive_i = \alpha + \beta Age_i + \epsilon_i \tag{4}$$

$$Positive_i = \alpha + \beta_1 Age_i + \beta_2 Age_i^2 + \gamma Male_i + \sum_{i=1}^{i=4} \delta_i Year_i + \sum_{i=1}^{i=128} \varphi_i Country_i + \epsilon_i \tag{5}$$

Table 4 reports the results for equation 5 using three alternative estimators. Both the linear and quadratic terms for Age are significant, and the results are predicting that both younger and older athletes are more often caught in the anti-doping net. This effect is stronger for older athletes—for example, the probability of a 37-year old testing positive is 80% higher than for the corresponding 27 year-old age group. A high prevalence of positives in younger

[18] The proportion of positives, at 7.61% (according to the sample numbers below), is comfortably in excess of the 2.0% upper bound of the typical estimate range of positive-test shares (Lenten and Smith, 2020, p.1)

[19] Other control variables might extend the analysis, the data specifying Out-of-competition (OOC) and In-competition tests could also play a role. Unfortunately our dataset of tested athletes by the IAAF is based solely on OOC tests. The relative performance of an athlete or her relative ranking would be another interesting extension.

athletes is surprising, however, younger athletes may be less skilled at concealing their PED use (see again Lenten et al., 2017, p. 598)

Table 4: Model Estimation Results from Equation (5)

	(i) LP	(ii) Logit	(iii) Probit
Constant	0.3909**	1.0780	0.5870
	(0.2018)	(1.5940)	(0.8401)
Age	-0.0250***	-0.2396**	-0.1344**
	(0.0075)	(0.1038)	(0.0552)
Age^2	0.0004***	0.0043**	0.0024***
	(0.0001)	(0.0017)	(0.0009)
$Male$	0.0161*	0.3768**	0.1974***
	(0.0090)	(0.1487)	(0.0747)
$Y09$	-0.0452***	-0.7484***	-0.3976***
	(0.0151)	(0.2345)	(0.1188)
$Y10$	-0.0526***	-0.7736***	-0.4233***
	(0.0136)	(0.2073)	(0.1071)
$Y11$	-0.0821***	-1.2850***	-0.6470***
	(0.0137)	(0.2500)	(0.1207)
$Y12$	-0.0552***	-0.7812***	-0.4115***
	(0.0125)	(0.1860)	(0.0963)
n	3,218	3,218	3,218
$\bar{R}^2$	0.2250	0.1127	0.1134
ℓ	182.9	-768.7	-768.1

Notes: standard errors in parentheses. For models (i) and (ii),

The results obtained here will be decomposed in the appendix.

5.2 The leaked IAAF data

The additional empirical strategy is to run a regression on the IAAF leaked blood data. To check the correlation between age and doping behavior, we will use several approaches. Firstly, we treat the data as cross-sectional; and secondly, we create an unbalanced panel dataset from all the individuals tested more than once.

The OFF-score is the main measure used in the ABP and combines information about HGB and RET into one number. A normal OFF-score lies approximately between 80 and 115.[20] Athletes with values outside this range can be considered, with high level of certainty, as having doped. For the cutoff points of 116.7 for men and 104.4 for women (1% probability of a false positive,

[20]The range can vary slightly for each individual and situation.

Gore et al. (2003)) the database contains 940 samples with OFF-scores higher than the cut-off point, a striking number.

To run a cross-section regression, we restrict the database to normal or high OFF-scores (higher than or equal to 80) and take into consideration only pre-competition tests. A doper is expected to have a high OFF-score; a non-doper is expected to have a normal value and an ex-doper (who doped in the past and now her normal blood values are recovering) is expected to exhibit a low OFF-score. Therefore, if all of the observations are taken into account, it would lead to an underestimation of the true effects. This restriction leaves 4,490 observations.

$$OFFscore_i = \alpha + \beta_1 AGE_i + \gamma Male_i + \sum_{i=1}^{i=6} \delta_i Year_i + \sum_{i=1}^{i=218} \varphi_i Country_i + \epsilon_i \quad (6)$$

The following table (5) contains four variations of the model represented by equation (6), models (i) and (ii) are regressions without controlling for year of test ($Year_i$) and country, whereas models (iii) and (iv) add those variables (which themselves are not reported in the table).

Table 5: Least-Squares Model Estimation Results from Equation (6)

	(i)	(ii)	(iii)	(iv)
Constant	89.90^{***}	87.43^{***}	78.38^{***}	76.36^{***}
	(1.028)	(1.888)	(5.987)	(6.175)
Age	0.1394^{***}	0.3185^{***}	0.1854^{***}	0.3295^{***}
	(0.4131)	(0.4131)	(0.4119)	(0.4119)
Age^2	???	-0.0031	???	-0.0025
	(0.0373)	(0.1209)	(0.0367)	(0.1145)
$Endurance$	1.544^{***}	1.563^{***}	1.817^{***}	1.830^{***}
	(0.3964)	(0.3965)	(0.4344)	(0.4345)
$Male$	3.851^{***}	3.862^{***}	6.696^{***}	6.706^{***}
	(???)	(???)	(8.254)	(8.254)
n	4,490	4,490	4,490	4,490
$\bar{R}^2$	0.0263	0.0266	0.1739	0.1741
ℓ	-1.777×10^4	-1.777×10^4	-1.730×10^4	-1.729×10^4

Standard errors in parentheses.

As in the previous table, the coefficient estimate for Age is highly significant, increasing Age by one year is associated with an increase in the OFF-score by

approximately 0.2 points (a 20-year difference between 2 athletes translates into 4 points difference in an OFF-score). Athletes in endurance disciplines are predicted to have higher OFF-scores by almost 2 points. Men have generally higher OFF-scores; however, this difference cannot necessarily be attributed to doping behavior. Rather, in reality it is much more likely to be due to physiological differences.

Using athletes' IDs, we can create a panel of those athletes who were tested at least twice. This leaves us with 4,819 observations for 1,332 unique athletes. The following table reports estimates on Pooled OLS (i-iii) and fixed effects (iv-vi). Columns (vii-viii) show model results estimated only for pre-competition tests (3,698 observations) and the final two columns (ix-x) take into account only pre-competition tests and OFF-scores higher than 80 (2,352 observations).

Every single estimate exhibits a positive relationship between age and OFF-score. The linear coefficient (Age) varies between 0.14 for the Pooled OLS model and 0.84 for the fixed-effects model including only pre-competition tests. The estimates in columns i-viii are all expected to have a small bias. The nature of the OFF-score makes it a bit tricky to read the results, since not only very high OFF-scores are a sign of doping, but also very low ones, too. Therefore, columns ix and x report fixed-effects model results with observations limited to pre-competition tests and OFF-scores above 80. Within this limited range, an increase in the value of OFF-score increases the probability that the athlete was taking PEDs to improve her performance, a 1-year Age increase is associated with a higher OFF-score by roughly 0.5 points. This implies that a 10-year age difference between athletes is associated on average with a higher OFF-score by five points for the older one. The drawback of looking only at pre-competition tests and OFF-scores higher than 80 is the reduction in sample size.

On the basis of these results, we infer that a relation between age and doping is established with a high level of confidence. However, self-selection bias can undermine the belief in a true causal relationship here. Since only the athletes in the right-tail of the talent distribution typically survive in competition into their thirties, it can be argued that there is not a random selection, but rather a self-selection by those for whom it is a profitable choice. Nonetheless, an athlete who dopes during her twenties might be more successful and thus be more willing to continue her professional career longer than she would have otherwise without having taken PEDs at all. Despite the issue with causality, the estimates are still highly relevant for policy makers.

Table 6: ???Further Robustness Check Estimation Results???

	(i) Pooled LS	(ii) Pooled LS	(iii) Pooled LS	(iv) FE	(v) FE	(vi) FE	(vii) FE-PC	(viii) FE-PC	(ix) FE-PC80	(x) FE-PC80
Constant	84.73**	85.81**	63.58**	79.98**	105.4**	81.59**	66.99**	111.3**	85.03**	110.0**
	(1.570)	(7.263)	(6.723)	(4.141)	(12.95)	(15.26)	(4.058)	(12.68)	(4.327)	(13.33)
Age	0.1345**	0.0540	0.8254*	0.3117**	−1.628*	−1.293	0.8392**	−2.586**	0.5348**	−1.397
	(0.0577)	(0.5300)	(0.4853)	(0.1544)	(0.9508)	(0.9348)	(0.1537)	(0.9420)	(0.1635)	(0.9876)
Age^2		0.0015	−0.0108		0.0359**	0.8392**		0.0641**		0.0361**
		(0.0095)	(0.0087)		(0.0174)	(0.0171)		(0.0174)		(0.0182)
PCT			3.242**			7.868**				
			(0.6689)			(0.7307)				
ICT			−11.41			−10.37				
			(8.925)			(7.886)				
OOC			0.5312			0.4916				
			(1.799)			(1.630)				
Male			15.51**			14.22				
			(0.5148)			(15.62)				
n	4,819	4,819	4,819	4,819	4,819	4,819	3,698	3,698	2,352	2,352
$\bar{R}^2$	0.0009	0.0007	0.1653	0.0012	0.0024	0.0387	0.0117	0.0170	0.0069	0.0094
ℓ	-2.115×10^4	-2.115×10^4	-2.071×10^4	-1.870×10^4	-1.870×10^4	-1.861×10^4	-1.366×10^4	-1.365×10^4	−8,204	−8,201

Standard errors in parentheses; significance at the 10, 5 and 1 percent levels are indicated, respectively, by ***, ** and *.

6 Discussion and conclusions

This paper has included a theoretical treatment of an athlete's doping choice with respect to her age. Even more significantly, it has also provided an extensive empirical analysis on that matter.

A rational professional athlete approaching the end of her career does not face much risk of future economic costs of her doping behavior. Hence, she will be expected to dope as much as possible with respect to both the probability of being caught and the magnitude of effect that doping has on her income. In rounds preceding the final one (year), her decision to dope is affected heavily by her prospects—higher expected future income will make current doping more costly. The level of doping is expected to grow with age, especially approaching the end of the athlete's professional career.

All of our empirical findings show that older athletes are relatively more often doped; however, one has to be cautious when making inference regarding causality. Even though a fixed-effects estimate can reduce the omitted-variable bias, a self-selection bias is still present. Since only certain athletes continue to compete into their 30s, it could be caused by those athletes involved in doping being more willing to prolong their career because of higher expected returns. Another selection bias stems from the fact that athletes are not chosen randomly for anti-doping sample collection. To account for this bias properly, more personal data would be necessary to estimate a choice model.

Even though causality cannot be identified with high a strong likelihood, it is still worth estimating the sign of correlation and its magnitude. The sign of the association between doping and age is identical among all estimates; however, its estimated magnitude and qualitative significance varies. In 3, there was a 29% higher-than-average level of suspicious blood samples within the 31-35 age group, and even a 230% higher-than-average level in the oldest age group (36-40). A weaker, but still positive, relationship can be found in the public data on positive doping tests in athletics. The regression estimates, controlling for gender, type of discipline and country, produce similar results. The age of athletes seems to be a good predictor of their doping behavior, thus it is reasonable to use this information to improve anti-doping policies in a cost-effective manner.

Maennig (2002), who pointed to the end-game problem, also discussed possible solutions to it. As an alternative (or even supplement) to bans, implementation of sufficient financial penalties—preferably if optimally structured in

an intertemporal sense, could better discourage athletes from doping regardless of their age. One specific proposal to this end is a conditional superannuation scheme, also discussed by Maennig (2002)), and tested by Wu et al. (2016) using economic experiments. In this system, athletes would be obliged to forego immediate access to a nominal share of their prize money, which is allocated to a special fund. The accumulated funds (with capital gains) would be released some elapsed time after the end of her career if she never tested positive.

Superannuation could potentially have two positive effects. Firstly, it mitigates the end-game problem; and secondly, it reinforces the current ban system and increases the costs of doping for all athletes—young and old. The superannuation scheme proposal has the potential to be a complementary tool in the arsenal of anti-doping measures. As Lenten et al. (2017) noted, no single policy intervention seems plausible in fully eliminating doping in sport. Regardless of the empirical results of this paper, the superannuation policy can help to overcome the scientific delay of anti-doping by making post-career punishment possible (within a defined withholding period).

In 2015, WADA changed its recommended ban length from 2 to 4 years, even for a first rule violation.[21] In the case of a second-time violation, the recommended punishment is still a lifetime ban, as before. The idea behind extending ban to 4 years is to strengthen deterrence further; however, it could conceivably be counterproductive by exacerbating the end-game effect problem identified here. The longer the doping ban is, the higher is the inter-generational doping inequality that can be expected. Another contention against increasing the length of bans was made by Hirschmann (2017), demonstrating that under certain circumstances, increasing doping sanctions can decrease the number of participants in the competition. This is because an increase in doping sanctions, not accompanied by a significant decrease of doping, will reduce payoffs for all athletes.

To understand the effect that different anti-doping rules might have, much more work has to be done. All of experimental economics, more sophisticated game-theoretical models and multi-agent systems could provide more insights. More empirical work on doping in sports is also needed, but for significant further extension on this study, the cooperation of sport organizations and anti-doping agencies is necessary to provide a large, comprehensive, individualized and identified database of all PED tests.

[21] Source: https://www.wada-ama.org/en/resources/the-code/world-anti-doping-code

Appendix: Decomposition of Empirical Results

To understand better the meaning of the estimates (those based on the public data) and their possible biases, we hereby decompose the results. Firstly, we proceed to express a decomposition of the share of positive doping tests for a given age group i.

$$s_i = \frac{A_i \cdot T_i \cdot d_i \cdot p_i}{A_i} = T_i \cdot d_i \cdot p_i \tag{A.1}$$

Equation (A.1) expresses the share of positive doping tests from cohort i, s_i; as a function of: (i) the number of athletes, A_i; (ii) the number of tests per athlete, T_i; (iii) the share of doped athletes within the i age-group cohort, d_i; and (iv) the probability of registering a (true) positive doping test when doped, p_i. The known variables are s_i, which is predicted by the empirical model, and T_i is also known (or at least a raw estimate) from the USADA data. Both d_i and p_i are unknown.

Figure 1 shows the distribution of tests per athlete by age cohort from the USADA database. The horizontal axis contains information on both the number of tests (top) and cohort (bottom). At a casual glance, the curve appears to be increasing (concavely) until an athlete age of approximately 30 years. The beginning-of-career starting value is approximately 1.5 tests per athlete, rising to about 3.5 to 4.0 by the career stage at which their natural performance should be diminishing (30 to 38 years of age). This surprisingly non-decreasing age range within the figure can be explained by sample selection bias—only athletes that are sufficiently talented (or doped) to compete, continue to survive in the elite athletics pool throughout that age range. Age groups above 39 were not included, given the scarcity of observations.

Figure 1: *Distribution of tests per athlete per cohort from USADA database.*

A mere eyeball of the figure clearly shows that the difference in testing numbers for older groups and middle-aged groups is not significantly different (i.e. it cannot be an explanation for why older or younger athletes are estimated to have more positive doping tests). If the probability that a doped athlete registers a positive test (when doped) does not depend on age, then the curve in figure 1 would have to be U-shaped to render our estimates irrelevant. We will show this formally using equation (A.1).

Let us introduce another age-group, m, such that $\frac{s_i}{s_m} = \alpha$. We can now decompose the structure of the determinants of the relative difference between age groups. Formally, we have:

$$\alpha = \frac{T_i}{T_m} \cdot \frac{d_i}{d_m} \cdot \frac{p_i}{p_m} \tag{A.2}$$

Now assume that both $\frac{s_i}{s_m} > 1$ (i.e. $\alpha > 1$) and $\frac{T_i}{T_m} < 1$. Two possibilities emerge in order to satisfy all of the conditions:

1. if $\frac{p_i}{p_m} = 1 \Rightarrow \frac{T_i}{T_m} \cdot \frac{d_i}{d_m} > 1$, hence $\frac{d_i}{d_m} > \frac{T_m}{T_i}$
   If an age group i has a higher estimated prevalence of positive tests than group m, and there is no difference in probability of being caught while doped, then the ratio of dopers between groups i and m must be higher than the ratio of tests between i and m. Stated alternatively, if 28 year olds are tested twice as often as 38 year olds, the prevalence of doping must be more than double of that in the older age group.
   Since older athletes are, according to the data, not tested less frequently than younger ones (i.e. $\frac{T_m}{T_i} \approx 1$ and $\frac{s_i}{s_m} > 1$); then assuming $\frac{p_i}{p_m} = 1$, it must hold that $\frac{d_i}{d_m} > 1$. For 38 year-old athletes comparative to 28 year-old athletes, $\frac{s_i}{s_m} \approx 2$; hence $\frac{d_i}{d_m} \approx 2$. Obviously, if the assumption holds and the estimates are correct, then α also correctly corresponds to the prevalence of doping.
   More striking is the estimates for younger athletes at the beginning of their career (IAAF data). For 18 year olds, $\frac{s_{18}}{s_{28}} \approx 2$ yet $\frac{T_{28}}{T_{18}} \approx 3$. This implies that for equation (A.1) to hold, $\frac{d_{18}}{d_{28}} \approx 6$ (i.e. 18 year olds would have to dope approximately six times more than 28 year olds).

2. if $\frac{d_i}{d_m} = 1 \Rightarrow \frac{T_i}{T_m} \cdot \frac{p_i}{p_m} > 1$, so $\frac{p_i}{p_m} > \frac{T_m}{T_i}$
   The other possibility is that both age groups have the same prevalence of doping and the perceived higher prevalence of positive doping tests is instead attributable to different probabilities of being tested positive. If 38 year-old athletes and 28 year-old athletes do have the same share of dopers, then the probability of being caught has to be higher in group i and has to be higher than the respective proportion of tests in group m over group i.
   As in the previous example, since $\frac{T_m}{T_i} \approx 1$ and the estimated share of positive tests is $\frac{s_i}{s_m} \approx 2$, it must then hold that $\frac{p_i}{p_m} \approx 2$ in order to render the estimated coefficients biased.

This decomposition has shown that when the number of tests per athlete is added to the analysis as a variable, then (assuming that the data are representative) the predicted shape of the age-doping distribution will not change. Nevertheless the estimated α can possibly be a product of variation in $\frac{p_i}{p_m}$ more than in variation of $\frac{d_i}{d_m}$, therefore any inference about the prevalence of doping among different age groups have to be taken with this in mind. It should be noted that differences in p_i would have to be quite large in order to create a non-upward-sloping curve in the age distribution of doped athletes. While our empirical estimates seem to be quite robust, there should nevertheless be no inference made beyond the relation between age and positive doping tests.

References

Bayer, R.-C. and F. Cowell (2009). Tax compliance and firms' strategic interdependence. *Journal of Public Economics 93*(11-12), 1131–1143.

Berentsen, A. (2002). The economics of doping. *European Journal of Political Economy 18*(1), 109–127.

Berentsen, A., E. Bruegger, and S. Loertscher (2008). On cheating, doping and whistleblowing. *European Journal of Political Economy 24*(2), 415–436.

Breivik, G. (1992). Doping games a game theoretical exploration of doping. *International Review for the Sociology of Sport 27*(3), 235–253.

Buechel, B., E. Emrich, and S. Pohlkamp (2016). Nobody's innocent: The role of customers in the doping dilemma. *Journal of Sports Economics 17*(8), 767–789.

Carlson, K., J. Kim, A. Lusardi, and C. F. Camerer (2015). Bankruptcy rates among nfl players with short-lived income spikes. *American Economic Review 105*(5), 381–384.

Coupé, T. and O. Gergaud (2013). Suspicious blood and performance in professional cycling. *Journal of Sports Economics 14*(5), 546–559.

Dilger, A., B. Frick, and F. Tolsdorf (2007). Are athletes doped? some theoretical arguments and empirical evidence. *Contemporary Economic Policy 25*(4), 604–615.

Dilger, A. and F. Tolsdorf (2004). Doping als wettkampfphanomen. In H.-D. Horch, J. Heydel, and A. Sierau (Eds.), *Events im Sport: Marketing, Management, Finanzierung: Beiträge des 3. Deutschen Sportökonomie-Kongresses*, pp. 269–279. Deutsche Sporthochschule Köln / Institut für Sportökonomie und Sportmanagement.

Eber, N. (2002). Credibility and independence of the world anti-doping agency: A barro-gordon-type approach to antidoping policy. *Journal of Sports Economics 3*(1), 90–96.

Emrich, E. and C. Pierdzioch (2015). A note on the international coordination of antidoping policies. *Journal of Sports Economics 16*(3), 312–321.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review 95*(1), 384–394.

Gore, C. J., R. Parisotto, M. J. Ashenden, J. Stray-Gundersen, K. Sharpe, W. Hopkins, K. R. Emslie, C. Howe, G. J. Trout, R. Kazlauskas, and A. G. Hahn (2003). Second-generation blood tests to detect erythropoietin abuse by athletes. *Haematologica 88*(3), 333–344.

Harbring, C. and B. Irlenbusch (2008). How many winners are good to have?: On tournaments with sabotage. *Journal of Economic Behavior and Organization 65*(3-4), 682–702.

Haugen, K. K. (2004). The performance-enhancing drug game. *Journal of Sports Economics 5*(1), 67–86.

Hirschmann, D. (2017). May increasing doping sanctions discourage entry to the competition? *Journal of Sports Economics 18*(7), 720–736.

Lenten, L. J. A. and A. C. T. Smith (2020). Testing conditional superannuation as an anti-doping policy supplement for safeguarding athlete health and welfare. *Performance Enhancement and Health*, forthcoming.

Lenten, L. J. A., A. C. T. Smith, and R.-C. Bayer (2017). Adding conditional superannuation to the antidoping policy mix. *Journal of Sport Management 31*(6), 591–604.

Maennig, W. (2002). On the economics of doping and corruption in international sports. *Journal of Sports Economics 3*(1), 61–89.

Savulescu, J., L. Creaney, and A. Vondy (2013). Should athletes be allowed to use performance enhancing drugs? *British Medical Journal 347*(f6150), 1–3.

Tarallo, C., J.-C. Humbert, P. Mahassen, B. Fournier, and J. Henny (1994). Reticulocytes: Biological variations and reference limits. *European Journal of Haematology 53*(1), 11–15.

Wu, Q., R.-C. Bayer, and L. J. A. Lenten (2016). A comparison of anti-doping measures in sporting contests. Working Paper 2016-11, University of Adelaide, School of Economics.