

The Epo Fable in Professional Cycling: Facts, Fallacies and Fabrications

Hein FM Lodewijckx*

Faculty of Psychology and Educational Sciences, Open University of The Netherlands, Heerlen, The Netherlands

Abstract

The massive doping schemes that surfaced in professional cycling suggest that riders' performances, realized in the controversial 'epo era' (>1990), are a cut above achievements delivered by their forerunners. We examined this superior performances assumption (SPA) by conducting six historic studies, which all scrutinized archival records of winning riders' stage race and time trial performances demonstrated in the three European Grand Tours (Tour de France, Giro d'Italia, and Vuelta a España; 1903–2013), including Lance Armstrong's wins. Findings revealed that all riders' wins in the epo years are no exception to the variability in speed progress observed in the three races over time and none of their achievements proved to be outliers. This also holds true for Armstrong's performances. These findings agree with results of a meta-analysis of epo studies we conducted, indicating that the ergogenic effects of epo and blood doping on riders' aerobic performances and associated cycling speeds are overrated. In conclusion, we argue that our observations render the SPA doubtful. They also made us realize that arguments used in contemporary discussions about effects of doping in cycling often involve psychological biases, false reasoning and fabrications. They are presented in the closing sections of this contribution.

The Epo Fable in Professional Cycling: Facts, Fallacies and Fabrications

The recent history of doping affairs in professional cycling placed the sport into grim light and left individual riders with a tainted image [1-6]. Reckoning the proposed, powerful performance-enhancing (or ergogenic) effects of epo and blood doping used by riders in the 'epo era' (>1990), it is often argued that riders' sportive achievements in these years are therefore superior to accomplishments delivered by riders in prior years [7,8] (USADA, 2012a, b). The current paper examines the soundness of this superior performances assumption (SPA) in two different ways. First, at a physiological level, the SPA presupposes strong ergogenic effects of, for instance, epo doping on aerobic performances of endurance athletes such as cyclists, resulting in strong increases in cycling speed. However, as will be shown, findings of a meta-analysis of epo studies we carried out indicated that these effects are far less powerful than generally presumed. These results constitute a first, yet indirect, indication that the SPA may be invalid. Second, to directly evaluate the SPA, we conducted six historic studies which all scrutinized annals of the three main European cycling races: Tour de France, Giro d'Italia, and Vuelta a España (1903–2013). With these records we not only assessed riders' winning performances in stage races, but in time trials as well. The studies were guided by the doping hypothesis model (DHM) that permits a critical appraisal of the SPA because it takes account of the variation in riders' sportive achievements in actual races over time. Alternatively, the same historic performance variation also enables us to examine whether riders' feats in the epo years can be explicated by various historic measures from the past which are unrelated to doping, such as the distances of the races or the years in which riders competed.

In the next sections, we will first elaborate on the DHM. We will then present a short theoretical outline of the physiological reasons why cyclists are tempted to manipulate their blood with doping agents that augment the volume of red blood cells (RBCs), i.e., epo and blood doping. We will then summarize main findings of the meta-analysis and the archival studies. As noted, findings of the meta-analysis evoked our reservations against the SPA. All historic studies confirmed these initial reservations, making the SPA doubtful. Little by little, our empirical facts made us aware that arguments put forward in present-day discussions about effects of doping in professional road racing too often involve psycho-logical fallacies and fabrications. They will be discussed at length in the concluding sections of this paper.

The Doping Hypothesis Model

For many fans the doping affairs that plagued the cycling world in recent years put great pressure on their enthusiasm for the sport. The scandal that struck the final blow involved American, ex-professional racer Lance Armstrong. He was the only rider ever to win seven consecutive Tours de France. The doping agents used by Armstrong and eleven of his team mates at the U.S. Postal and Discovery Channel teams predominantly involved epo, blood transfusions (doping with own blood, or with blood harvested from a compatible donor), and anabolic steroids such as testosterone [8] (USADA, 2012b). Worldwide, the Armstrong affair aroused strong punitive, moral outrage and cleansing responses [9-12] that appear to be fueled by popular beliefs about the potent ergogenic effects of aforementioned doping agents. Intriguingly, these beliefs are questioned by Kuipers [13] in a paper with the compelling title: "Putative effects of doping in cycling." He doubts whether said doping agents indeed improve performances of endurance athletes such as cyclists. Given his conclusions, he further plausibly contends that the doping problem in the sport is therefore partly "due to superstition, hearsay, and insufficient knowledge among the athletes' support personnel, which frequently leads to medical malpractice in sport" (ibid., p. 2645).

One cannot imagine a bigger contrast: The downfall of Lance Armstrong versus Kuipers' arguments that the ergogenic effects of many of the modern doping agents used by cyclists are overestimated and conceivably might even rest on mountebankery and deception. However, Kuipers is not the only dissenting voice. Other scholars arrived at similar conclusions and their arguments will be presented in the pages to follow. When taking the findings and arguments of these

*Corresponding author: Hein Lodewijckx, Faculty of Psychology and Educational Sciences, Open University of The Netherlands, Stationsweg 3A, 4811 AX, Breda, The Netherlands, Tel: 31-76-5711608; Fax: 31-76-5715424; E-mail: Hein.Lodewijckx@ou.nl, lodex@ziggo.nl

Received April 26, 2014; Accepted June 26, 2014; Published July 10, 2014

Citation: Lodewijckx HFM (2014) The Epo Fable in Professional Cycling: Facts, Fallacies and Fabrications. J Sports Med Doping Stud 4: 141. doi:10.4172/2161-0673.1000141

Copyright: © 2014 Lodewijckx HFM. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

scholars seriously, it would be better to speak of a *doping hypothesis* which is still in search of empirical support.

Clearly, many riders in recent years were involved in epo and / or blood doping affairs or acknowledged afterward that they used these aids during their active career [2,14]. So, the pivotal question we sought to answer is not about whether cyclists give in to doping use. In all likelihood they do. While its prevalence is estimated in some studies to be 3-7% [15-18], solid figures, however, are still not available. Rather, reckoning the opposing voices, the essential question of the study became: "To what extent did the use of RBC-augmenting doping aids enhance riders' performances in actual races?" As is the case with the unknown pervasiveness of doping use in the cycling world, research also demonstrates an impressive lack of independent, verifiable and conclusive empirical evidence concerning the effects of doping on riders' sportive achievements in real competitions over the years [1,3,19]. To fill this gap and to resolve our uncertainty concerning the status of the doping hypothesis, we conducted a series of empirical studies that were guided by the doping hypothesis model (DHM). As can be seen in Figure 1, it only consists of question marks.

Lack of Research

Some scholars attempted to fill the lacunae in the model [1,3-6] leading to very interesting, yet undependable insights into the customs and traditions pertaining to doping use in the cycling world. Furthermore, some ex-professional cyclists wrote about their own personal experiences with doping use during their sportive career [20-27]. There are also sketchy accounts about these relationships written by members of the support staff of different cycling teams, such as Willy Voet [28], former 'soigneur' of the much disputed French Festina team, and Jef D'Hont [29], who worked with the controversial German Telekom team. Medical doctors also wrote personal accounts about these subjects, such as Eric Ryckaert [30] who was involved in the 1998 Festina doping affair, and Gérard Porte [31] who served as a physician in the Tour de France for nearly forty years. Finally, you have the journalists and writers, who attempted to uncover the 'omerta' or unwritten code of silence concerning doping use in the professional group of riders [32-35].

Although these sources yield an understanding of the mores in professional cycling, they still mainly concern anecdotic attempts to clarify the historic, sociological, psychological, and physiological reasons as to why riders resort to doping use. As Brewer [1] observed, this is due to the lack of systematic research into actual cycling contests,

which is why scholars often built on their own experiences as fans of, and participants in, cycle races. Consequently, much of the writings about the cycling world necessarily had to fall back on the 'grey' literature, i.e., publications that appeared in magazines, on websites, and in popular books, to some of which we referred to previously. This also means that research that methodically attempted to investigate the relationships between the variables presented in the DHM is practically absent. Our own studies leave the relationships presented in path B unrequited, but they do permit to give direct answers to the relationships presented in path A and some indirect answers to the relationships presented in path C. The latter replies can only be indirect, since there are simply no studies available that unequivocally assessed the statistical relationships between these variables in actual cycle races over time and our studies also failed to do so. Then again, we argue that our answers constitute a rather inconvenient truth concerning the SPA. A first, indirect indication of the unsoundness of the SPA relates to findings of a meta-analysis we conducted [36]. It evaluated results of laboratory studies that all assessed effects of epo administration on aerobic performance and extrapolated the estimated epo-stimulated improvements in performance to cycling speeds in actual races. Main findings are summarized in the next section, preceded by a short outline of the physiological reasons why endurance athletes are tempted to manipulate their blood with RBC-augmenting doping aids.

RBC-augmentation: Physiology and Estimations of its Effects on Performance Physiology

Building on the cardiovascular/anaerobic model [37,38], the proposed, causal chain between epo/blood doping and aerobic performance runs as follows:

$$\text{Epo / blood doping} \rightarrow \text{RBCs} \uparrow \rightarrow \text{Ht} \uparrow \rightarrow \text{O}_2\text{-transport capacity} \uparrow \rightarrow \text{VO}_{2\text{max}} \uparrow \rightarrow \text{W}_{\text{map}} \uparrow \rightarrow \text{Cycling speed} \uparrow$$

The key variable in the chain constitutes $\text{VO}_{2\text{max}}$, described as an estimate of cardio-respiratory, circulatory and muscular fitness that measures the fastest rate at which oxygen (O_2) can be delivered and consumed by the body during strenuous exercise. The model assumes that high levels of $\text{VO}_{2\text{max}}$ are a prerequisite for top-level accomplishments in endurance sports such as cycling [39]. Through the process of erythropoiesis, the hormone erythropoietin (or epo) stimulates the bone marrow to produce newly formed red blood cells (RBCs), or reticulocytes. After one to two days, they grow into mature cells, or erythrocytes, which for 99% make up the population of blood cells. Hemoglobin (Hb) is a protein in RBCs which carries oxygen and, hence, it is essential in the chain of oxygen transport from the lungs to O_2 -consumers such as the muscles. A single RBC contains ~250 million Hb-molecules which each can bind four O_2 -molecules. This is the reason why it is proposed that the oxygen-transport capacity of blood can be improved by augmenting the volume of RBCs in the blood (RBCV) and thereby the blood's Hb mass. Hematocrit (Ht) is commonly used to assess the concentration of RBCs in the blood, or the thickness (viscosity) of the blood. It constitutes the ratio of RBCV vs. PV (blood plasma volume, the liquid part of the blood). For instance, a value of $\text{Ht}=45\%$ means that one litre blood contains 450 ml RBCV and 550 ml PV. Further note that only mature, but not immature RBCs can bind oxygen. This means that, in contrast to blood transfusions that immediately supply the blood with mature RBCs, administration of artificial epo does not have this instant effect. Accordingly, the physiological grounds why cyclists are inclined to boost their performances through RBC-augmenting doping aids is because it is proposed that this augmentation elevates the Ht level in the blood, thereby increasing the corresponding oxygen-carrying capacity.

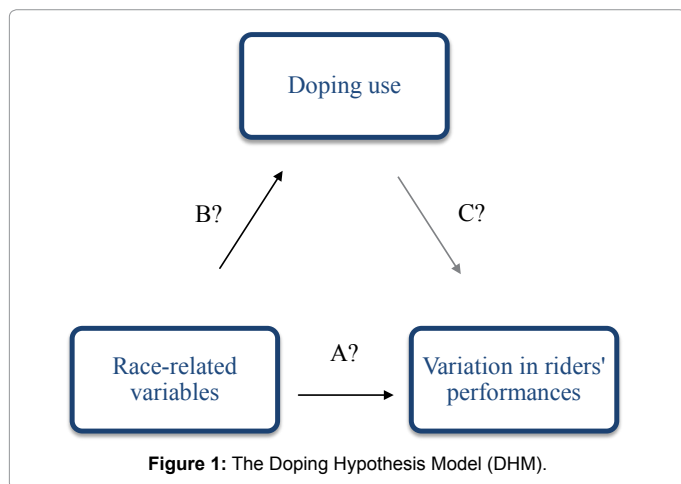


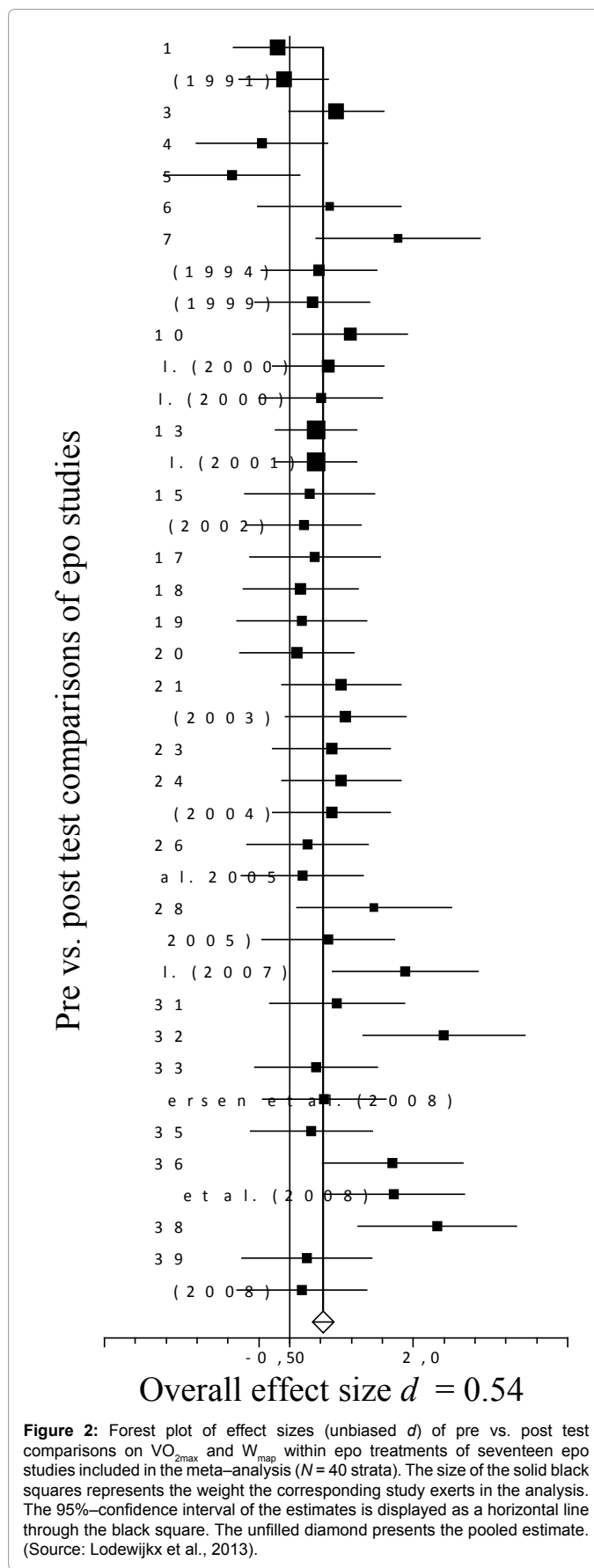
Figure 1: The Doping Hypothesis Model (DHM).

This increase subsequently improves athletes' aerobic performance capacity (VO_{2max}) and the associated maximal aerobic power output, expressed in watts (W_{map}). These improvements are presumed to result in increased speeds in races and, ultimately, even in victories [22,40].

According to Catlin et al. [41] and Joyner [42] the use of artificial epo as an ergogenic aid became rampant in endurance sports such as cycling in the early 1990s. Verbruggen [82], former chairman of the International Cycling Union (UCI), the sport's governing body, labeled the 1990–2000 period in professional cycling as an "epo epidemic", maintaining that epo improved athletes' endurance capacity by as much as 20%. In 2013, Vandeweghe — president of the Flanders Cycling Federation (WBV, Wielerbond Vlaanderen) — arrives at a similar conclusion concerning the estimated, cumulative effects of epo in actual competitions. Lundby and Olsen [43] reviewed findings of laboratory studies which examined the relationship between epo administration and aerobic performance in normal, healthy humans. They concluded that —if Ht is artificially increased by epo treatment from pre-test baseline values to around Ht=50% posttest— VO_{2max} is estimated to improve by 8–12%. For blood doping the estimates are 5–10% improvement [44]. Descriptively, Ninot et al. [45] concluded that the ergogenic effects of, for example, synthetic epo are "dramatic", whilst effects of blood transfusions are labeled "gigantic" in a study by Lundby et al. [46].

Estimations of Effects

Seemingly, the use of descriptive terms such as 'dramatic' and 'gigantic' may lead to the impression that the SPA is indeed valid. However, notice that these are rather suggestive terms that do not have any statistical meaning. To reach some solid, yet preliminary answers concerning the validity of the SPA, we therefore decided to statistically estimate the ergogenic effects of RBC-augmenting doping agents on riders' aerobic performances and the corresponding improvements in cycling speeds by conducting a meta-analysis of epo studies [36]. Kuipers [13] already contended that effects of RBC-doping aids on cyclists' maximal aerobic power output (W_{map}) are overestimated, while Heuberger et al. [58] even maintain that the epo doping – aerobic performance hypothesis is not supported by empirical evidence. In our meta-analysis, we evaluated findings of seventeen laboratory studies and assessed effect sizes (unbiased d , r and r^2) of the epo-stimulated increases in VO_{2max} and W_{map} . The Forest plot in Figure 2 summarizes effect sizes of pre vs. posttest comparisons on these measures, relating to all epo treatments of the studies. The average effect size amounted to $d=0.54$. Figure 2 additionally shows that 31 (77.5%) of the forty d values did not surpass the bandwidth of the 95%-confidence interval. This means that many of the epo-induced improvements in performance yielded by the studies did not exceed chance level. Note, however, that negative and small d 's may have inflated these findings. Additional results revealed that these values could either be attributed to submaximal performances measured in distinct studies, or to specific experimental treatments manipulated in the studies, such as performances assessed at moderate altitudes (hypoxia) or at sea level (normoxia). Of the nine values that did surpass the 95%-bandwidth, five could be traced to two studies, which mainly concern large positive d 's obtained on W_{map} [47,48]. After refining analyses to maximal performances demonstrated by participants at sea level in double-blind, placebo-controlled studies the fixed, pooled effect sizes were moderate: $d=0.41$ – 0.49 . These values slightly deviate from the overall $d=0.54$ described previously. According to Cohen [49] an effect size of roughly half SD ($d=0.50$) indicates that in 67% of the observations the epo studies are not able to discriminate between maximal performances demonstrated by participants that were administered epo or not. The observed amounts of explained



variation in epo-stimulated performance improvement of 4–19% mean that a considerable 81–96% of these improvements cannot be attributed to the effects of experimental conditions. Percentages improvement from pre to post tests ranged between $M=6-7\%$ (VO_{2max}) and $M=7-8\%$ (W_{map}). Important for the SPA, based on Nevill et al. [50], we estimated that the largest improvement in VO_{2max} of $M_{post-pre}=0.29$ l/min yielded by the analysis corresponded to an increase in velocity of about one kilometer per hour (km/h). Perneger [51] reports a similar increase in speed. However, Hopkins et al. [52] strongly warn against directly generalizing this increase to actual races. Additionally, Heuberger et al. [53] and Lodewijckx et al. [36] argue that the epo/blood doping-aerobic performance relationship suffers from external, ecological, and predictive validity problems. For instance, the relationship becomes very limited if we consider the well-known fact that elite athletes, such as professional cyclists, are estimated to be able to exercise at peak VO_{2max} levels for approximately ten minutes before reaching the different stages in the lactate threshold [54]. So, after this period cyclists' exercise capacity will be greatly reduced. Hence, the influence of epo doping on cyclists' performances in actual races is also strongly constrained by time limits.

All these observations suggest that the effects of RBC-augmenting doping aids might indeed be strongly overvalued in actual competitions, including the races that were scheduled during the epo era. Yet, the meta-analysis mainly evaluated achievements of non-athletes who delivered their performances in laboratory situations. It did not examine achievements demonstrated by professional riders in real races. Only a critical appraisal of these performances may provide some conclusive answers concerning the validity of the SPA. In our six archival studies we therefore evaluated these performances.

Historic Studies: Performances in the Three European Grand Tours

Findings of two studies which examined the annals of the three major European stage races appear to agree with the SPA. Perneger [51] investigated mean km/h performances of riders who ranked fifth in the overall standings of the three tours in the period 1990-2009. He reported that between 1990 and 2004, riders' speed increased by 0.16 km/h per year and further observed a decrease in speed of 0.22 km/h per year

since 2004. El Helou et al. [21] analyzed mean km/h performances of riders who reached the first ten places in the final standings of eleven European races from 1892 to 2008, which included all famous one-day classic races as well as the three main stage races. They found that, relative to the pre-epo years (1946–1992), riders' performances in the epo years (1993–2008) showed a significant improvement in km/h of 6.38%.

We also decided to scrutinize archival records [55] of the three main stage races and to assess winning riders' km/h and time performances. However, consistent with Perneger [51] and Vandeweghe [14], we realized that riders' first-ranking achievements, demonstrated after three weeks of competition, could bias conclusions relating to the SPA. In the end, selecting riders' final rankings as the measure to evaluate the SPA may be false, since the very same performances also result from the joint and coordinated labors in the total group of cyclists participating in the races. Hence, these group labors can be considered contaminating variables which may strongly impede a sound evaluation of individual riders' final achievements. Conversely, in time trials riders in person race against the clock and compete for the fastest time. Since they cannot benefit from the efforts of other riders in these races through drafting, time trialing requires the maximum of individual riders' stamina and aerobic exercise capacity [56]. To validly evaluate the SPA, we therefore followed Vandeweghe's [14] recommendation and decided to gather data concerning individual riders' time trial accomplishments as well.

Table 1 presents an overview of the studies and the descriptive statistics of the variables we assessed. Riders' mean km/h and time performances served as the dependent variables. Column five shows that we used seven DHM measures from the past to account for the historic variation in riders' wins. Regarding stage races, we measured the years in which cyclists competed (Y), the distances of the races (D) as well as the number of stages in the races (N_{ST}), which inclined with advancing years. We further developed the brutality rate (B), measuring the harshness of the races. In the early years of the tours sometimes only 30% of the riders managed to finish the race, while in 2011 more than 80% succeeded to do so. Findings revealed that lower distances and brutality rates and increases in the number of stages facilitated riders' speed to a lesser or greater extent over the years. We additionally took

Study	N	Wins	Years	DHM variables ¹	R ² _{adj} (%) explained by DHM variables ²		Main explanatory variable (r)				Outliers (1990–2013)
					Km/h	Time	Km/h		Time		
							Y	D	Y	D	
1 (Lodewijckx & Brouwer, 2011)	181	Stage race	1947–2008	Y, D, B, N _{ST} , SR	.84***	.98***	.79***	-.42***	-.60***	.92***	None
2 (Lodewijckx & Brouwer, 2012)	256	Stage race	1903–2011	Y, D, B, N _{ST} , SR	.94***	.90***	.92***	-.31***	-.68***	.85***	None
3 (Lodewijckx & Verboon, 2013) ³	62	Time trial (50–61 km)	1934–2010	Y, D, A	.58***	.56***	.77***	-.16	-.71***	.41**	None
4 (Lodewijckx & Bos, 2014) ³	100	Time trial (multiple winners)	1949–2013	Y, D, A	.55***	.98***	.73***	-.24*	-.21*	.98***	Indurain (Tour, 1992)
5 (Lodewijckx, 2013) ⁴	19	Time trial (mountain)	1958–2004	Y, CL _i , A	.93***	.88***	.48*	.83***	.14	.94***	Berzin (Tour, 1996)
6 (Lodewijckx & Verboon, 2014) ⁵	324	Time trial (all)	1933–2013	Y, D, SR	.60***	.98***	.71***	-.28***	-.34***	.97***	Plaza (Vuelta, 2005)

Notes:

¹ Y = Competition year. D = Distance of stage race / trial. B = Brutality rate. N_{ST} = Number of stages in the race. SR = Stage race (Tour vs. Giro vs. Vuelta). A = Comparisons of Armstrong vs. other riders. CL_i = Climbing index. Findings concern observations across the Tour, Giro, and Vuelta.

² % Performance differences (R²_{adj}) explained by DHM variables. R²_{adj} are rounded off.

³ These studies compared Armstrong's (A) time trial victories to other riders, demonstrated on flat and rolling terrain, excluded are mountain time trials.

⁴ We only examined mountain trials in the Tour de France.

⁵ Included in this study are all trials on flat and rolling terrain, excluded are prologues, mountain trials, and team time trials.

* p ≤ .05; ** p ≤ .01; *** p ≤ .001

Table 1: Main Findings of Historic Studies.

account of the influence the three different stage races (SR) exerted on riders' achievements over time.

As to time trials, Table 1 shows that the years of competition (Y) and the distances (D) of the trials again served as predictor variables. Study 6 investigated all victories realized by riders on flat and rolling terrain in the three tours over the years (1933–2013). The remaining studies all compared Armstrong's (A) achievements to other riders. Study 3 evaluated his wins against other riders who, from 1934 to 2010, won trials in the three major races and all faced distances comparable to Armstrong's (50–61 km) in the seven Tours he won. Study 4 relates to comparisons of Armstrong's victories with wins of all other famous multiple Grand Tour winners (Coppi, Anquetil, Merckx, Hinault, and Indurain) as well as victories of riders who either were, or were not, involved in doping affairs in the years following Armstrong's domination in professional road racing (2006–2013). Study 5 evaluated Armstrong's 2001 and 2004 wins in mountain time trials (racing uphill). We developed a climbing index (CL_i) to capture the demanding nature of these trials, operationalized as the rise (the corrected altitude of the climbs in km) over the run (the total distance of the trial in km). Higher values of the CL_i designate more demanding trials in terms of riders' instant climbing efforts.

In the sections to follow, we will first discuss the total amounts of variation (R^2_{adj}) the various DHM variables conjointly explained in riders' performances in the different studies as well as the single variables that explained the largest amounts of variation in these achievements. We will then proceed with a comprehensive appraisal of the SPA relating to different findings in different studies: (1) stage race wins; (2) time trial wins; (3) outlying performances; and (4) Armstrong's achievements. The appraisal additionally involves an alternative assessment of riders' performance progress over time, since km/h and time performances *per se* do not permit valid estimates of this progress. The differencing method, used in time-series analysis [57], enables one to appraise the proportional progress (%) in riders' performances per year. We therefore also employed this method to examine the SPA. Last, as can be seen in the figures and tables, we partitioned the years of competition into periods of approximately ten years, using El Helou et al.'s critical year 1993 as the standard to classify the ten-year periods after WW II. Because our main objective is to examine the validity of the SPA, we will restrict the presentation and discussion of our findings to three bordering time periods: the epo era (1993–2002) vs. the immediate preceding years (1983–1992) vs. the following years (2003–2011/2013). A thorough elaboration of developments in the remaining time periods and associated differences between races is beyond the scope of the present thesis. We refer to the separate studies for information concerning these developments.

Main Findings

The sixth column in Table 1 presents the total amounts of variation (R^2_{adj}) the various DHM measures together explained in riders' achievements. As to km/h performances, the amounts range between 55–94%. For time performances they vary between 56–98%. The lowest amounts are found in the first Armstrong time-trial study (Study 3; R^2_{adj} =56–58%). In this study, distance did not significantly contribute to the R^2_{adj} , because it examined a restricted range in the distances of the trials (50–61 km), thereby decreasing the overall R^2_{adj} . The correlations presented in the seventh column show that two variables account for most of the differences in riders' achievements over time. Competition year constitutes the main explanatory variable of riders' km/h performances ($r=0.48-0.92$). As the years proceed, riders race faster. The exception is the mountain time trial study (Study 5) in which the positive influence

of distance ($r=0.83$) was stronger than the influence of competition year ($r=0.48$). Distance turned out to be the main explanatory variable of riders' mean time performances. Larger distances are associated with slower performances. Except for Study 3, which evaluated restricted trial distances ($r=0.41$), the remaining studies yielded robust correlations ($r=0.85-0.98$). Table 1 additionally reveals that, as far as time trials are concerned, competition year has a relatively minor and inconsistent influence on riders' time performances ($r=-0.34-0.14$). We will deal with these relationships in the sections on the psycho-logical fallacies, because they have consequences for the SPA.

Speed Progress in Stage Races

The six panels in Figure 3 graphically present the relationships between competition year and the two performance measures, providing some first answers to the soundness of the SPA. We confined the plots to findings of the study with the largest number of observations (Study 2; $N=256$). The correlations in Panel A and B reveal a linear progress in speed over the years, which is stronger for km/h than for time performances. Notice, first, that the negative competition year-time performance relationship in Panel B also indicates faster time performances over the years. Second, both panels reveal a steady improvement in performance from the epo era up to 2011, designating that riders in the '90s did not outperform riders in succeeding years. This observation constitutes a first indication of the invalidity of the SPA.

A second indication concerns the annual proportional changes in performance we computed. The relationships are illustrated in Panel C and D of Figure 3. Positive numbers on km/h performances denote an incline in performance per year, 0% indicates no progress, and negative numbers indicate a decline. The reverse relationships hold for time performances. The original linear relationships, depicted in Panel A and B, are illustrated by the mean proportional changes in Panel C and D: $M_{km/h}=0.52\%$; $M_{time}=0.42\%$. These percentages indicate, for instance, that riders increased their km/h with approximately 52% within a time span of 100 years, i.e., from the 1900s ($M=26$ km/h) to the 2010s ($M=39.5$ km/h; excluding the years of the two world wars).

Furthermore, in disagreement with the SPA the mean proportional changes in both panels reveal no striking increase in the epo years. As to km/h performances, the analysis yielded $r=-.05$ ($b=-0.007\%$; $p=.41$) and the individual 95%-prediction interval varies between -7.8–8.9%. For time performances, the correlation amounts to $r=-.10$ ($b=-0.04\%$; $p=.10$; 95%-prediction interval: -21.9–22.9%). The same panels additionally designate that the prediction intervals apply to all time periods and, hence, to riders' disputed wins in the epo era as well.

A third indication concerning the unsoundness of the SPA involves the distribution of the mean annual changes that can be seen in Panel E and F of Figure 3. The changes concerning km/h performances are normally distributed, while this is not true for the changes involving time performances, which show kurtosis. This can be explained by the relatively large number of observations that revolve around zero. Importantly, both distributions indicate a symmetrical dispersion of inclines and declines in riders' performances over time. Accordingly, there is no evidence at all for the existence of extraordinary 'superior' developments in riders' first-ranking achievements in the three stage races over the years.

A final indication relates to the evaluation of the predicted, proportional progress in performance realized in the epo era compared to the directly adjacent time periods. Using regression analyses, we assessed the influence of stage races and competition year on the

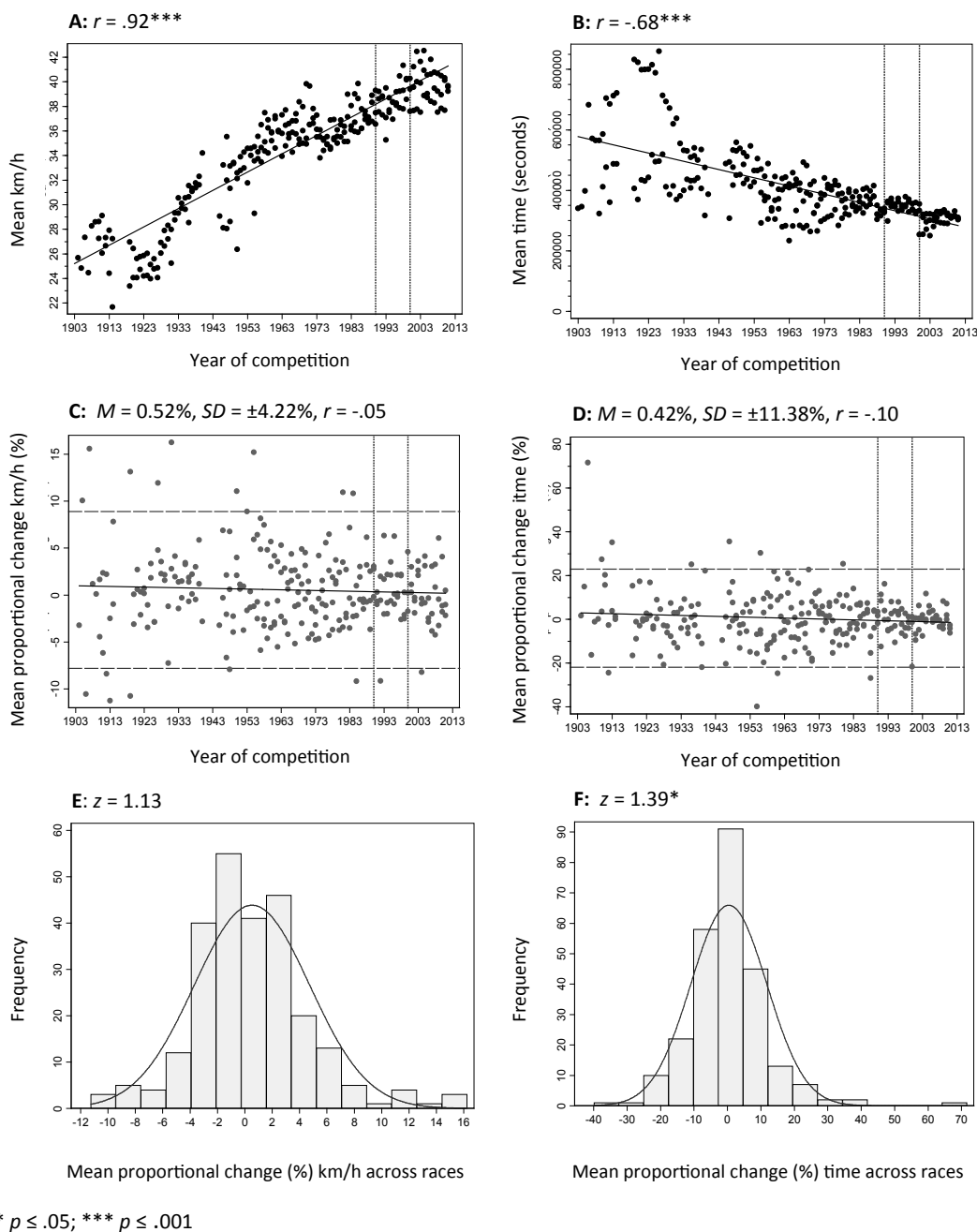


Figure 3: Panel A and B present Pearson correlation coefficients between competition year and winning riders' mean km/h and time performances across stage races. Panel C and D show the Pearson correlation coefficients between competition year and the mean proportional changes (%) per year for both performance measures, averaged across races. Panel E and F present the distribution of the latter performance measures. Vertical dotted lines mark the epo era (1990–2000), horizontal dashed lines present the individual 95%–prediction interval.

percents change per year in which we controlled for the influence of the corresponding annual changes in the three covariates alluded to above (distance, number of stages, and the brutality rate). We will restrict the presentation of the regression findings solely to km/h performances, since time performances yielded virtually identical results. Main findings are summarized in Table 2 and Figure 4. First, relative to Panel C in Figure 3, Figure 4 shows a strong reduction in the variability of the changes that can be attributed to the influence of the variables included in the analyses. Second, Table 2 reveals that, within the epo era, none of the predicted, mean changes differ significantly from zero. Likewise,

the changes in the epo years do no differ significantly from the changes in the immediate bordering years (or from all the other time periods we distinguished). These conclusions hold within and across races. The findings additionally indicate that the relationships between competition year and the mean changes across and within races are small and negative: in speed progress Across races, $r = -0.13$, $b = -0.007\%$ per year, $R^2_{adj} = 1.2\%$, ($p \leq 0.05$); Tour: $r = -0.15$, $b = -0.008\%$ per year, $R^2_{adj} = 1.3\%$, ($p = 0.14$); Giro: $r = -0.14$, $b = -0.008\%$ per year, $R^2_{adj} = 1\%$, ($p = 0.18$); Vuelta: $r = -0.11$, $b = -0.01\%$ per year, $R^2_{adj} = 1\%$, ($p = 0.39$). The relationships reveal a minor, gradual decline over time in the three

	Stage Races ¹				Time Trials ²			
	Year	N	M (SE) ³	95%-CI	Year	N	M (SE) ³	95%-CI
Across Races	1903-1914	16	0.21 (0.47) _a	-0.72–1.14	-	-	-	-
	1919-1929	22	1.48 (0.37) _a	0.75–2.22	-	-	-	-
	1930-1940	22	1.18 (0.64) _a	-0.08–2.43	1933-1942	9	1.46 (0.19) _b	1.07–1.86
	1946-1952	18	0.93 (0.42) _a	0.11–1.75	1946-1952	15	1.10 (0.13) _b	0.85–1.35
	1953-1962	28	0.26 (0.33) _a	-0.40–0.91	1953-1962	24	0.73 (0.10) _b	0.52–0.93
	1963-1972	30	0.49 (0.32) _a	-0.14–1.12	1963-1972	28	0.56 (0.09) _b	0.38–0.74
	1973-1982	30	0.27 (0.32) _a	-0.36–0.90	1973-1982	29	0.34 (0.09) _a	0.17–0.52
	1983-1992	30	0.26 (0.32) _a	-0.37–0.89	1983-1992	30	0.33 (0.09) _a	0.15–0.50
	1993-2002	30	0.48 (0.32)_a	-0.15–1.11	1993-2002	30	0.15 (0.09)_a	-0.02–0.33
2003-2011	27	0.27 (0.34) _a	-0.39–0.94	2003-2013	33	-0.11 (0.08) _a	-0.27–0.06	
Tour	1983-1992	10	-0.14 (0.55) _a	-1.23–0.95	1983-1992	10	0.32 (0.15) _a	0.01–0.62
	1993-2002	10	0.52 (0.55)_a	-0.57–1.61	1993-2002	10	0.20 (0.15)_a	-0.10–0.51
	2003-2011	9	0.19 (0.58) _a	-0.96–1.34	2003-2013	10	-0.01 (0.15) _a	-0.29–0.28
Giro	1983-1992	10	0.40 (0.55) _a	-0.69–1.49	1983-1992	10	0.39 (0.15) _a	0.09–0.70
	1993-2002	10	0.41 (0.55)_a	-0.68–1.50	1993-2002	10	0.13 (0.15)_a	-0.18–0.43
	2003-2011	9	0.19 (0.58) _a	-0.97–1.33	2003-2013	10	-0.16 (0.15) _a	-0.45–0.13
Vuelta	1983-1992	10	0.52 (0.55) _a	-0.57–1.61	1983-1992	11	0.27 (0.15) _a	-0.04–0.57
	1993-2002	10	0.50 (0.55)_a	-0.59–1.59	1993-2002	11	0.13 (0.15)_a	-0.17–0.43
	2003-2011	9	0.46 (0.58) _a	-0.69–1.60	2003-2013	11	-0.16 (0.15) _a	-0.44–0.13

Notes:

¹ Number of stage race wins between time periods vary due to WW I and II and the Spanish civil war. The years of the epo era are in bold type face. Predicted mean proportional changes are based regression analyses in which we used competition year, stage races, distances of the stage races, number of stages in the races, and the brutality rate of the races as predictor variables. N = 253 because proportional changes cannot be computed for the initial races.

² The first time trial ever in professional cycling was scheduled in the 1933 Giro, followed by the Tour in 1934, and in 1941 by the Vuelta. Predicted mean proportional changes are based regression analyses in which we used competition year, stage races, and the distances of the time trials as predictor variables. We aggregated the time trials per year (N = 198).

³ Means without a common subscript differ significantly, p ≤ .05. Across stage races, we contrasted the epo period against all other periods. For separate stage races, contrasts compared the epo years to immediate adjacent time periods. To prevent Type I-errors, we adjusted the 95%-CI using the Bonferroni procedure.

* Within time periods mean proportional changes differ significantly from zero, p ≤ .05.

Table 2. Descriptive Statistics of Predicted Mean Proportional Changes (%) in Winning Riders' Km/h Performances per Time Period in Stage Races and Time Trials.

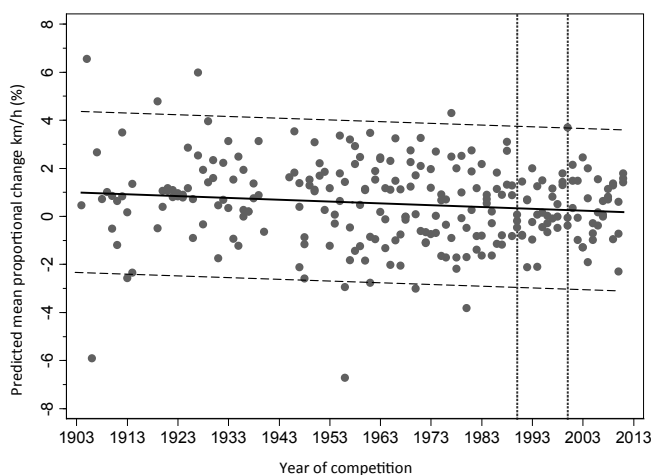


Figure 4: Relationship between predicted mean proportional changes (%) in winning riders' km/h performances and competition year, aggregated across stage races. Changes are based on the regression analyses of the stage race data in Table 2. Vertical dotted lines mark the epo era (1990–2000), dashed lines the individual 95%–prediction interval.

races and competition year explains 1–1.3% of the differences in these developments.

To sum up, all these observations invalidate the SPA as well as

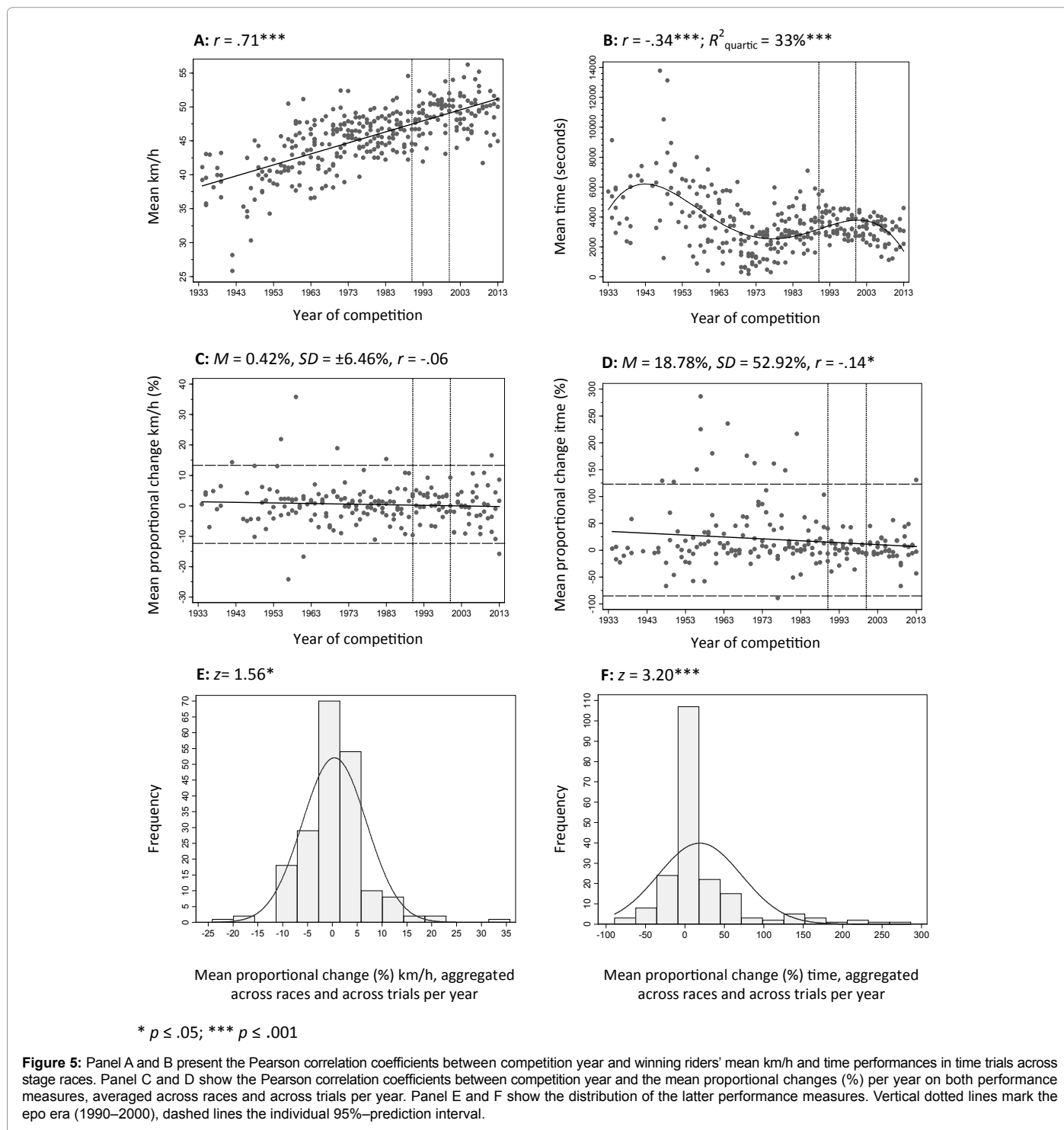
conclusions drawn by El Helou and colleagues [16] and Perneger [51]. They indicate that riders perform progressively faster over time, but this progress gradually levels off with advancing years. However, the progress in the epo period is not superior compared to the progress

observed in immediate bordering periods and, most importantly, does not constitute an exception to the overall, historic variability in speed progress observed in the three main stage races.

Speed Progress in Time Trials

Findings concerning riders' time trial achievements are presented in an identical way as the stage race data. Panel A en B in Figure 5 present the zero-order correlations between competition year and the two performance measures for the study with the largest number of

observations (Study 6; $N=324$). In regard to km/h performances, Panel A shows a linear progress in speed, indicating that riders perform faster with advancing years. Again the panel shows that riders in the epo era did not perform faster than riders in succeeding years. Besides, Panel B reveals that for mean time performances the relationship is not linear, but resembles a significant M-curve, or quartic relationship. Riders raced slower in the '40s, faster in the post-WW II years up till the mid-'70s, slower in the '80s and '90s, and faster again after 2003. By itself, this M-curve again disproves the SPA, because riders in the epo era



delivered comparatively slower not faster time performances compared to riders' achievements in the 2000s. Additional analyses indicated that this M-curve can be explained by the distances of the time trials. This variable shows a robust correlation with time performances ($r=0.97$) and, hence, a development over the years which is similar to the M-curve obtained on the time performance measure. Accordingly, the variation in riders' mean time performance over the years is due to the closely matching variation in trial distances. These observations imply that we should control for the influence of distance on riders' performances.

Panel C and D in Figure 5 illustrate the relationships between the estimated proportional changes and competition year for both dependent variables. To estimate these changes we transformed the data by aggregating the time trials per year. The reason is that the public in the '50s, '60s and '70s wanted to see the performances of the legendary and renowned time trialists Fausto Coppi, Jacques Anquetil, and Eddy Merckx. To satisfy the interest of the public in these riders they therefore faced many trials in their stage races, the distances of which sometimes varied extremely (between 8 and 137 km). These large differences led to huge bandwidths on both performance measures. To circumvent this problem we therefore averaged the time trial performances per year. Consistent with the stage race data, findings of the regression analyses reject the SPA. Regarding the mean proportional changes in km/h, the analyses produced $r=-0.06$ ($b=-0.019\%$; $p=0.39$) and the individual 95%-prediction interval varies between -12.4-13.3%. For time performances, the correlation is $r=-.14$ ($b=-.35\%$; $p=.051$) with the individual 95%-prediction interval ranging between -85.1-123.2%. Notice again that for time performances, positive numbers indicate a decrease and negative numbers an increase in performance. Thus, the negative correlation found for time performances designates a progress in these performances that lessens with proceeding years. As to these performances, we stress that Panel D still reveals a huge variation, despite the data transformation we applied.

Panel E and F in Figure 5 present the distributions of the aggregated, annual changes for the two performance measures, revealing deviations from normality. Similar to the stage race data, the deviations cannot be attributed to a relatively large number of observations that involve extreme forms of fast progress, but rather to a comparatively large number of observations that hover around zero. Once more, both distributions show no evidence for the existence of extraordinary 'superior' developments in riders' time trial performances over the years.

Last, we evaluated the proportional progress in performance in the epo era vs. the directly bordering time periods in which we controlled for the associated yearly changes in trial distances and differences between stage races. Again, the presentation of the regression results will be confined to km/h performances. Findings are presented in Table 2 and Figure 6. Relative to Panel C in Figure 5, Figure 6 shows a strong reduction in the variability of the changes owing to the variables we entered into the regression equation. Additionally, Table 2 reveals that the predicted mean changes in the epo years do not differ significantly from zero. Examination of differences between the three time periods also produced no significant effects across and within races. However, inconsistent with findings relating to the stage races, the regression results revealed strong, significant decreases in the predicted speed progress over time: Across races, $r=-.64$, $b=-.019\%$ per year, $R^2_{adj}=41\%$; Tour: $r=-.62$, $b=-.017\%$ per year, $R^2_{adj}=37.8\%$; Giro: $r=-.82$, $b=-.022\%$ per year, $R^2_{adj}=67.3\%$; Vuelta: $r=-.50$, $b=-.017\%$ per year, $R^2_{adj}=23.5\%$ (all $p \leq .001$). All relationships indicate a strong downturn in predicted speed progress per year within and across races. Inconsistent with the SPA,

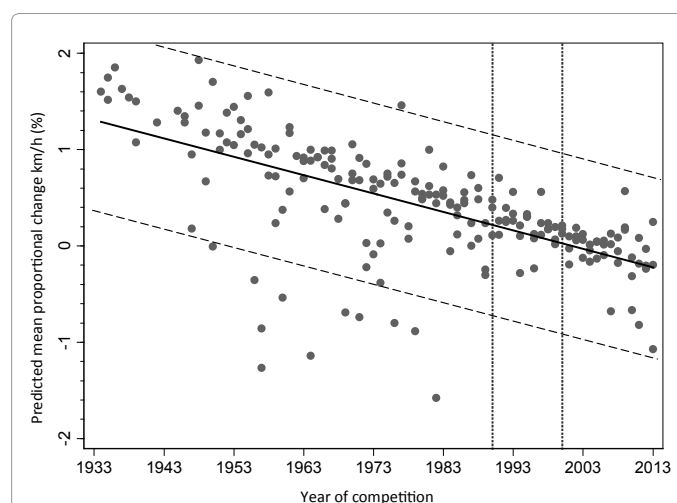


Figure 6: Relationship between predicted mean proportional changes (%) in winning riders' km/h performances in time trials and competition year, aggregated across races. Changes are averaged across trials per year and based on the regression analyses of the time trial data in Table 2. Vertical dotted lines mark the epo era (1990-2000), dashed lines the individual 95%-prediction interval.

however, Table 2 and Figure 6 show that riders' progress in the epo era does not constitute an exception to these developments. The observed decrease in progress can be explained by the fact that, over time, riders deliver faster performances in time trials. However, the more enhanced the speed the more difficult it becomes to make a difference. Hence, the diminishing annual progress in speed demonstrated by riders in these individual races against the clock over the years.

On the whole, we conclude that the time trial data corroborate findings relating to the stage races. Neither the SPA, nor conclusions drawn by El Helou et al. [16] and Perneger [51] are supported.

Outliers

An assessment of outlying (very fast) performances perhaps makes up the most valid way to test the SPA. To determine outliers, we applied the rigorous criterion of $\geq \pm 2SD$ from the sample mean (the individual 95%-confidence and prediction intervals), while conventionally the criterion of $\geq \pm 3SD$ is used (or $z \geq \pm 3.30$ with $N < 1000$; [58]). As to the mean proportional changes, the 95%-prediction intervals and the normal distributions depicted in Figure 3 to 6 already indicate that outliers are very rare indeed. In all analyses, the one and only rider that slightly surpassed the bandwidth in the epo era involved the victory of Spanish rider Roberto Heras in the 2000 Vuelta ($z=2.01$). However, when considering achievements within the Spanish race itself, his performance did not go beyond the criterion anymore.

The last column of Table 1 presents findings of the separate studies. They are based on regression analyses in which riders' observed mean km/h and time performances served as the dependent variables. The table shows that none of the performances of the stage-race winners in the epo era or thereafter fell outside the 95%-bandwidth. This conclusion does not change much after the evaluation of riders' time trial accomplishments. Only three riders surpassed the 95% bandwidth: Spanish rider Miguel Indurain, Russian rider Yevgeny Berzin, and Spanish rider Ruben Plaza. Indurain's performance failed to be an outlier after having taken account of all time trial performances realized by winning riders over the years in the three tours (Study 6). Besides, it can be doubted whether Plaza's achievement in the 2005 Vuelta can be regarded an outlier as well, because the conditions during his race

must have been very good. Closer scrutiny of the overall standings of the trial showed that the rider who reached the 100th position (Alberto Ongarato) already achieved a speed of nearly 50 km/h [55]. Of all riders we investigated, perhaps only Berzin realized a striking performance in his race uphill. Then again, even his achievement did not surpass the bandwidth of $\pm 3SD$.

Once more, these results disprove the SPA. Reckoning the historic variation in riders' achievements in stage races and time trials, all riders' performances in the epo era fell within the range of expected variability. Evidently, many of these performances are outstanding. However, this does not mean to say that, therefore, they are 'abnormal.' Our findings persuasively suggest that this is not the case.

Armstrong's Wins

Study 3 evaluated whether Armstrong's seven time trial wins on flat and rolling courses, demonstrated in the Tour de France from 1999 to 2005, were faster compared to wins of riders who faced similar trial distances (50–61 km) in the three tours from 1934 to 2010. Findings of the study indicated that the American racer initially did realize significantly faster performances relative to the other riders ($M=4.70$ km/h, $R^2=6.4\%$). However, after statistically controlling for the significant influence of competition year, he ultimately raced somewhat slower ($M=-0.43$ km/h, $R^2=0.1\%$). Only one of his wins exceeded the bounds of the 68%–bandwidth (i.e., $\pm 1SD$ from the sample mean), while all his remaining six achievements fell within the bandwidth of this very stringent criterion. Thus, when considering the historic performance variation in these trials of limited distance, all individual accomplishments of the controversial American were not superior.

Study 4 evaluated his time trial–wins against victories of all other multiple Grand Tour winners and against riders who were, or were not, involved in doping affairs in the 2006–2013 periods. It yielded virtually identical results as Study 3. Ultimately, analyses revealed a non-significant difference of $M=142$ seconds ($M=0.33$ km/h) between Armstrong vs. all the aggregated other riders, which explained a trivial 0.1% of the variation in riders' performances. Only two of Armstrong's wins surpassed the bandwidth of the 68%–CI and none went beyond the 95%–CI. The study further revealed that riders who were involved in doping affairs (including Armstrong) raced somewhat slower than riders who were not ($M=-68$ s; $M=-1.17$ km/h). Yet, these differences were far from significant and explained an inconsequential 0.1–3.2% of the performance differences between riders.

Findings of the mountain time trial study (Study 5) confirmed the findings of the two other studies. The climbing index (CL_c) had a robust influence on riders' speed ($r=-.97$; $R^2=94\%$), indicating that riders raced $b=2.302$ km/h slower per unit of the index. The significant mediating influence of the index subsequently reduced riders' yearly progress in speed to a non-significant $b=26$ m per year ($R^2=0.3\%$). Besides, Armstrong's wins did not prove to be outliers, but came out comparatively slow. As to the results of the remaining studies, they all substantiate the foregoing conclusions: None of his performances were extraordinary, including his controversial seven victories in the Tour de France.

As a final conclusion, we maintain that all our empirical facts convincingly invalidate the SPA. Implications of this conclusion for contemporary discussions about the effects of doping in the cycling world will be addressed in the next sections.

Implications: Fallacies and Fabrications

Logical fallacies

As our body of evidence against the SPA accumulated, we came to

realize that arguments used in these discussions often involve logical fallacies and psychological biases. We will first discuss the logical fabrications. An often heard argument refers to the appeal to ignorance (*argumentum ad ignorantiam*), which poses that "something is true only because it has not been proved false, or that something is false only because it has not been proved true" [59]. Evidently, this logic again exemplifies the SPA, because it illustrates the assumed, strong positive association between the year in which riders competed with the concomitant doping use (i.e., in the epo era) and their performances. Our stage race and time trial data reject this logic. Findings relating to the predicted, mean proportional changes in stage races indicated that competition year explained a minor 1–1.3% of the differences in winning riders' progress in speed over time. The small, negative relationships we found indicate that this progress slowly levels off up till the present. We emphasize that riders do perform increasingly faster over the years, but the speed progress in the epo era is consistent with the progress observed in the other time periods we distinguished. The negative relationships we found for riders' predicted progress in time trials were more substantial, explaining 23.5%–67.3% of the performance differences between riders over time. Yet, they do not alter our main conclusion. Again, our findings revealed that the speed progress in the epo period was not superior compared to the progress observed in immediate bordering time periods and did not constitute an exception to the overall, historic variability in speed progress in time trials in the three main stage races. Besides, the assessment of the magnitude of the competition year– performance relationship is also contingent upon whether or not research examines the powerful influence of distance on riders' achievements. For instance, the last three time trial studies in Table 1 indicate that the impact of this variable on riders' time performances ($r=.94-.98$) is far more substantial and consistent than the influence of competition year ($r=-0.34-0.14$). All these observations render the reasoning underlying the appeal to ignorance implausible.

Study 5 illustrates the *post hoc ergo propter hoc fallacy* (translation: after this, therefore because of this). It reflects the erroneous belief that, because there is a temporal sequence in events, one event is the cause of the other. For example, Armstrong's winning achievements in mountain time trials in 2001 and 2004 ('after this') are assumed to be caused by his doping use ('because of this'). However, Study 5 revealed that the main determinant of riders' wins in these trials appeared to be the climbing index rather than the year in which riders won their trial and the doping use associated with it. Thus, when taking race-related variables into consideration that are essential for mountain time trialing, Armstrong's victories did not prove to be superior to achievements realized by famous climbers such as Charly Gaul in 1958 or Frederico Bahamontes in 1959 or 1962. These findings invalidate the logic used in the post hoc fallacy in addition to the reasoning presented in the argument from ignorance. They further entail that we tend to underestimate the athletic achievements demonstrated by the very gifted riders in the early days of the races and to overestimate riders' performances in the modern era.

Study 6 elucidates the *Texas sharpshooter fallacy*. If research is based on selective or limited data or uses an invalid dependent variable, this may result in biased conclusions concerning riders' evolution in speed over time. This fallacy applies to the two archival studies [16,51], described previously, that both supported the SPA. However, the 6.38% performance progress in the epo era, reported by El Helou and co-workers, can be attributed to the selective way this increase was statistically tested [16]. The researchers aggregated all riders' performances demonstrated in the pre-epo years (1946–1992) and compared the resulting mean performances to the aggregated mean performances of riders in the epo years (1993–2008). Because they

also found a strong linear increase in speed over time, the observed difference might thus have been inflated by the relatively slower speeds demonstrated by riders in the years following WW II (see also Panel A in Figure 3). In a similar vein, Perneger based his conclusions on limited data. He restricted his analyses to performances delivered by riders between 1990 and 2009, but did not include performances demonstrated by riders in 1980s in his sample. However, the choice of an alternative way to measure developments in performance (annual proportional changes), the decision to include confounding variables (the control variables) in our studies and to examine all stage race and time trial wins as well as performance differences between the epo era and directly bordering time periods proved them wrong.

All in all, the logical fallacies suggest that discussions about the (reputed) effects of doping in cycling may often involve false reasoning and fabrications. According to Carroll [59], the use of these errors becomes more tempting among 'believers'. Thus, for people who believe in the effects of doping, the lack of opposing empirical evidence may be germane to sustaining their belief.

Psychological fallacies

Psychological research shows that some very powerful social-cognitive biases may reinforce these beliefs, offering partial explanations for the naming, blaming, and shaming of cyclists. Our judgments may be guided by the availability heuristic [60], a rule of thumb whereby people base a judgment on the ease with which they can bring something to mind. Regretfully, winning cyclists who are accused of cheating can be brought to mind all too easily. This heuristic is further associated with some other biases that may well clarify stereotypic reactions towards cyclists. The base rate fallacy leads us to overestimate the number of (winning) riders that used doping substances. We also may engage into biased sampling and illusory correlations [61]. Both involve making sweeping statements based on selective samples of information that are atypical, i.e., on the basis of a few selected observations we come to the conclusion that, because some performances were associated with doping use, all endeavors of cyclists probably have to do with cheating.

These biases may have an influence on people's attribution of blameworthiness. In his culpable control model, Alicke [62] argues that such attributions or often guided by automatic, heuristic processes, not by rational arguments. If we apply his arguments to the doping problem in cycling they mean that, whenever we are confronted with a rider that by hook or by crook is associated with doping use, we implicitly tend to "exaggerate his volitional or causal control, lower evidential standards for blame, or seek information that supports our blame attribution" (ibid., p. 558).

This blaming may be directed by the fundamental attribution error [63], which refers to the tendency to strongly overestimate the extent to which people's behavior is caused by internal, dispositional factors and to underestimate the role of external, situational factors that may plausibly account for the same behavior. So, if we are confronted with an outstanding achievement of a cyclist in a mountain stage, the heuristic tendency will be to attribute this performance to some internal factors (athletic capability, training, doping/cheating?) and to underestimate the role of external, race-related circumstances (favorable wind, competition between teams, distance, relatively easy stage) that may have strongly affected the very same achievement. Three factors facilitate the operation of this bias. It occurs if performances are evaluated as being highly distinctive ("I've never seen a rider race so fast"), if we all agree that this is indeed the case (high consensus), and if there is low consistency in the performance ("He has never climbed like that before"). According to Hilton and Slugoski's [64] abnormal

conditions focus model, the combination of high distinctiveness, high consensus and low consistency will result in the conclusion that riders' performances are 'abnormal'. Combined with the culpable control processes and the other biases we described this will lead us to implicitly construe that these 'abnormal' winning achievements probably involve the use of banned substances and their attributed ergogenic effects. To his regret, Christopher Froome, winner of the 2013 Tour de France, faced this kind of reasoning throughout the entire race.

Conclusions

In the Introduction of this contribution we argued that the findings yielded by our studies would constitute a rather inconvenient truth concerning the validity of the SPA. We additionally put forward that our studies would permit to give direct answers to the relationships presented in path A of the DHM (Figure 1) and some indirect answers to the relationships presented in path C. Our findings indicated that the race-related variables we distinguished in path A explained considerable amounts of 90–94% (Study 2, all stage races) and 60–98% (Study 6, all time trials) of the performance differences between riders over time. These percentages imply that 6–10% (stage races) or 2–40% (time trials) of these differences are not explained by our DHM variables. The rather high 40% of unexplained differences, obtained on km/h performances in Study 6, can be attributed to the weak influence of distance ($r=-0.28$) on these particular performances, while the influence of the same variable on cyclists' time performances is robust ($r=0.97$). Therefore it seems fair to discard this 40%. This means that, ultimately, 2–10% of the differences in riders' achievements over the years in stage races and time trials can be attributed to other, unknown and perhaps confounding variables which we did not include in our studies. Doping use is but one of the variables to account for these differences next to other performance-enhancing variables, many of which relate to the circumstances under which cyclists practiced their sport over the years [1,3-6,14,33,34,56]. They all facilitated riders' achievements with advancing years, such as more favorable road, terrain, and race conditions; less demanding racing programs; growing insights from exercise physiology with associated sophisticated and effective training regimes; improved technology of bikes and racing gear; increased specialization of riders; and improvements in nutrition and hydration, leading to an enhanced maintenance of riders' energy balance during stage events [13]. In addition, in his socio-historical analysis of the cycling sport, Brewer [1] describes other facilitative factors relating to changes in team organization and inter-team dynamics, sponsorship, financial incentives, and a progressively deepening commercialization of the sport, which led to increased speed in races from the mid-1980s onward. We emphasize that this list of expediting variables is by no means exhaustive. They can all be regarded variables which, apart from doping use, may plausibly account for the unexplained performance differences between riders we obtained.

Importantly for the present thesis, the indirect relationships we found for path C render the superior performances assumption doubtful. Our empirical facts yielded no proof that riders in the epo era raced strikingly faster than riders in immediate adjoining years. Moreover, their first-ranking performances in stage races and time trials did not constitute outliers and the progress in performance they demonstrated did not depart from the variation in progress observed in the three major European stage races over time. These null results are consistent with findings of the meta-analysis we carried out, from which we concluded that the ergogenic effects of RBC-augmenting doping agents on riders' aerobic performances and associated cycling speeds are overestimated. However, despite the clear uniformity in our findings, we stress that conclusions from the historic studies can only

be tentative. The studies lack essential base lines and control conditions and the findings may have been influenced by an inestimable, systematic error: We have no idea how riders would have performed over the years, had they abstained from taking banned substances.

Nevertheless, we argue that awareness of our empirical observations and their implications would greatly contribute to the exchange of evidence-based pros and cons in the sometimes frenzied, societal discussions about the effects of doping in the cycling world, set in motion by the Armstrong affair. Moreover, Kuipers [13] and Heuberger et al. [53] might not be mistaken in their final conclusion that the doping problem in professional cycling indeed rests on superstition, hearsay, insufficient knowledge, medical malpractice, and lack of (opposing) empirical evidence. On balance, the sportive feats demonstrated by riders in the years of the 'epo epidemic' were not exceptional at all, despite their doping use.

References

1. Brewer BF (2002) Commercialization in professional cycling 1950–2001: Institutional transformations and the rationalization of doping. *SSJ* 19: 276-301.
2. Dilger A, Frick B, Tolsdorf F (2007) Are athletes doped? Some theoretical arguments and empirical evidence. *Contemporary Economic Policy* 25: 604-615.
3. Dimeo P (2007) A history of drug use in sport 1876–1976. New York: Routledge / Taylor & Francis Group.
4. Hoberman J (1992) *Mortal engines: The science of performance and the dehumanization of sport*. The Free Press, New York.
5. Hoberman JM (2005) *Testosterone dreams: Rejuvenation, aphrodisia, doping*. University of California Press, Berkeley/Los Angeles.
6. Lodewijckx HFM (2011) Reasoned decision of the United States Anti-Doping Agency on disqualification and ineligibility. Rotterdam: 2010 Uitgevers.
7. USADA (2012a) Statement from USADA CEO Travis T. Tygart regarding the U.S. Postal Service Pro Cycling Team doping conspiracy.
8. USADA (2012b) Reasoned decision of the United States Anti-Doping Agency on disqualification and ineligibility.
9. Lodewijckx HFM, Brouwer B (2011) Some empirical notes on the 'epo epidemic' in professional cycling. *Res Q Exerc Sport* 82: 740-754.
10. Lodewijckx HFM, Verboon P (2013) Lance Armstrong's era of performance – Part I: Are his time trial performances much different from other winners? *Journal of Athletic Enhancement* 2:1.
11. Tetlock PE (2003) Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences* 7: 320-324.
12. Tetlock PE (2003) Thinking the unthinkable: sacred values and taboo cognitions. *Trends Cogn Sci* 7: 320-324.
13. Tetlock PE1, Kristel OV, Elson SB, Green MC, Lerner JS (2000) The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *J Pers Soc Psychol* 78: 853-870.
14. Kuipers H (2006) [Putative effects of doping in cycling]. *Ned Tijdschr Geneesk* 150: 2643-2645.
15. Diamanti-Kandarakis E1, Konstantinopoulos PA, Papailiou J, Kandarakis SA, Andreopoulos A, et al. (2005) Erythropoietin abuse and erythropoietin gene doping: detection strategies in the genomic era. *Sports Med* 35: 831-840.
16. El Helou N1, Berthelot G, Thibault V, Tafflet M, Nassif H, et al. (2010) Tour de France, Giro, Vuelta, and classic European races show a unique progression of road cycling speed in the last 20 years. *J Sports Sci* 28: 789-796.
17. Gaudard A1, Varlet-Marie E, Bressolle F, Audran M (2003) Drugs for increasing oxygen and their potential use in doping: a review. *Sports Med* 33: 187-212.
18. Juhn M (2003) Popular sports supplements and ergogenic aids. *Sports Med* 33: 921-939.
19. Joyner MJ1, Coyle EF (2008) Endurance exercise performance: the physiology of champions. *J Physiol* 586: 35-44.
20. Bobet J (2008) *Tomorrow, we ride*. Norwich (UK), Mousehold Press/Hampstead (UK), Sport and Publicity.
21. Fignon L, Ducoin JE (2009) We were young and carefree [In French: Nous étions jeunes et insouciant]. Editions Grasset & Fasquelle, Paris.
22. Hamilton T, Coyle D (2012) *The secret race*. Bantam Books, New York.
23. Kimmage P (2007) *Rough ride: Behind the wheel with a pro cyclist*. Yellow Jersey Press / Random House, London.
24. Krijgsman E, Dekker Th (2011) Clean enough [In Dutch: Schoon genoeg]. Amsterdam, Arbeiderspers.
25. Millar D, Whittle J (2011) *Racing through the dark*. Orion Books, London.
26. Vanwalleghem R, Museeuw J (2009) Museeuw speaks [In Dutch: Museeuw spreekt]. Tielt (Belgium): Lannoo.
27. Winnen P (2000) From Santander to Santander. [In Dutch: Van Santander naar Santander]. Thomas Rap, Amsterdam.
28. Voet W (2001) *Breaking the chain*. Yellow Jersey Press / Random House, London.
29. D'Hont J (2007) Recollections of a 'soigneur' [In Dutch: Mémoires van een wielerverzorger]. Halewyck, Leuven, Belgium.
30. Rijckaert E (2000) The Festina affair [In Dutch: De zaak Festina]. Tielt (Belgium): Lannoo.
31. Porte G (2011) *Physician in the Tour* [In French: Médecin du Tour]. Editions Albin Michel, Paris.
32. Ballester P, Walsh D (2009) L.A. Confidential: Les secrets de Lance Armstrong [In Dutch: Vuil spel: het systeem Armstrong]. No Monkey Books, Westmalle, Belgium.
33. Fotheringham W (2009) *Fallen angel: The passion of Fausto Coppi*. Yellow Jersey Press, London.
34. Maso B (2005) *The sweat of the Gods*. Mousehold Press, Norwich, (UK).
35. Whittle J (2008) *Bad blood. The secret life of the Tour de France*. Yellow Jersey Press / Random House, London.
36. Lodewijckx HFM, Brouwer B, Kuipers H, Van Hezewijk R (2013) Overestimated effect of epo administration on aerobic exercise capacity: A meta-analysis. *American Journal of Sports Science and Medicine* 1: 17-27.
37. Noakes TD (2000) Physiological models to understand exercise fatigue and the adaptations that predict or enhance athletic performances. *Scandinavian Journal of Medicine & Science in Sports* 10:123-145.
38. Noakes TD (2008) Testing for maximum oxygen consumption has produced a brainless model of human exercise performance. *Br J Sports Med* 42: 551-555.
39. Varlet-Marie E, Gaudard E, Audran M, Bressolle F (2003) Pharmacokinetics / pharmacodynamics of recombinant human erythropoietins in doping control. *Sports Medicine* 33: 301-315.
40. Joyner MJ (2003) VO2MAX, blood doping, and erythropoietin. *Br J Sports Med* 37: 190-191.
41. Catlin DH, Hatton CK, Lasne F (2006) Abuse of recombinant erythropoietins by athletes. In Mouligneux G, Foote MA, Elliott SG (Eds.) *Erythropoietins and erythropoiesis*. Basel, Birkhäuser Verlag, 205-228.
42. Verbruggen H (2001) The EPO epidemic in sport. *Bloodline Reviews* 1: 3-4.
43. Lundby C1, Olsen NV (2011) Effects of recombinant human erythropoietin in normal humans. *J Physiol* 589: 1265-1271.
44. Bytomski JR, Moorman III CT, MacAuley D (2010) *Oxford American handbook of sports medicine*. Oxford University Press, Oxford.
45. Ninot G1, Connes P, Caillaud C (2006) Effects of recombinant human erythropoietin injections on physical self in endurance athletes. *J Sports Sci* 24: 383-391.
46. Lundby C1, Robach P, Saltin B (2012) The evolving science of detection of 'blood doping'. *Br J Pharmacol* 165: 1306-1315.
47. Lundby C1, Robach P, Boushel R, Thomsen JJ, Rasmussen P, et al. (2008) Does recombinant human Epo increase exercise capacity by means other than augmenting oxygen transport? *J Appl Physiol* (1985) 105: 581-587.
48. Thomsen JJ, Rentsch RL, Robach P, Calbet JAL, Boushel R, et al. (2007) Prolonged administration of recombinant human erythropoietin increases submaximal performance more than maximal aerobic capacity. *European Journal of Applied Physiology* 101: 481-486.

49. Cohen J (1988) *Statistical power analysis for the behavioral sciences* (2nd Edn.). Lawrence Earlbaum Associates, Hillsdale.
50. Nevill AM1, Jobson SA, Palmer GS, Olds TS (2005) Scaling maximal oxygen uptake to predict cycling time-trial performance in the field: a non-linear approach. *Eur J Appl Physiol* 94: 705-710.
51. Perneger TV (2010) Speed trends of major cycling races: does slower mean cleaner? *Int J Sports Med* 31: 261-264.
52. Hopkins WG1, Hawley JA, Burke LM (1999) Design and analysis of research on sport performance enhancement. *Med Sci Sports Exerc* 31: 472-485.
53. Heuberger JA1, Cohen Tervaert JM, Schepers FM, Vliegenthart AD, Rotmans JI, et al. (2013) Erythropoietin doping in cycling: lack of evidence for efficacy and a negative risk-benefit. *Br J Clin Pharmacol* 75: 1406-1421.
54. Johnson AT (2007) *Biomechanics and exercise physiology: Quantitative modeling*. Taylor and Francis Group, CRC Press, New York.
55. Magnier D, Picq P, Debreilly M, Zingoni P, Haffreingue H , et al. (2014) Cycling remembered. [In French: Mémoires du cyclisme].
56. Lucia A1, Earnest C, Arribas C (2003) The Tour de France: a physiological review. *Scand J Med Sci Sports* 13: 275-283.
57. Koop G (2000) *Analysis of economic data*. John Wiley & Sons, New York.
58. Tabachnick BG, Fidell LS (2001) *Using multivariate statistics* (4th Edn.). Allyn and Bacon / Pearson Education Company, Boston.
59. Carroll RT (2003) *The skeptic's dictionary: A collection of strange beliefs, amusing deceptions, and dangerous delusions*. John Wiley & Sons, Hoboken (NJ).
60. Tversky A, Kahneman D (1973) Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5: 207-232.
61. Hilbert M (2012) Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making. *Psychol Bull* 138: 211-237.
62. Alicke MD (2000) Culpable control and the psychology of blame. *Psychological Bulletin*, 126: 556-574.
63. Ross L (1977) The intuitive psychologist and his shortcomings: Distortions in the attribution process. In Berkowitz L (Ed.), *Advances in experimental social psychology* Academic Press, New York 10: 174-221.
64. Hilton DJ, Slugoski BR (1986) Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review* 93: 75-88.