

Overestimated Effect of Epo Administration on Aerobic Exercise Capacity: A Meta-Analysis

Hein F.M. Lodewijckx^{1,*}, Bram Brouwer¹, Harm Kuipers², Ren é van Hezewijk¹

¹Department of Psychology, Open University, Heerlen, The Netherlands

²Department of Movement Sciences, University of Maastricht, Maastricht, The Netherlands

*Corresponding author: Hein.Lodewijckx@ou.nl, lodex@ziggo.nl

Received March 27, 2013; Revised April 17, 2013; Accepted April 28, 2013

Abstract Recent studies examining the relationship between epo doping and aerobic performance (the EDAP–relationship) yield conflicting results. To resolve this inconclusiveness in an empirical way, we conducted a meta–analysis on 17 laboratory studies and assessed effect sizes (unbiased d , r and r^2) of the epo–induced improvements in aerobic exercise capacity measured by maximal oxygen uptake (VO_{2max}) and maximal aerobic power output (W_{map}). The fixed, pooled EDAP effect size estimates were moderate, $d = 0.41–0.49$, $r = .19–.44$, and $r^2 = .04–.19$, revealing a shift of approximately half SD in performances of the epo–treated compared to the non–treated participants. As to VO_{2max} , we observed the strongest post test performance ($M = 64.39\text{ml kg}^{-1} \text{min}^{-1}$) in double blind, placebo controlled studies on performances assessed at sea level with an increase from pre to post tests of $M = 4.02\text{ml kg}^{-1} \text{min}^{-1}$. Regarding W_{map} , the increase was $M = 26\text{W}$ with the strongest post test performance of $M = 398\text{W}$ observed in similar studies as VO_{2max} . Percents improvement from pre to post tests varied between $M = 6–7\%$ (VO_{2max}), and $M = 7–8\%$ (W_{map}). The largest improvement in VO_{2max} we found equals an increase in velocity of about 1km/h. Consistent with recent studies criticizing the EDAP–relationship our findings indicate that its strength is overestimated. In turn, this entails that the relationship between epo doping and cyclists’ performances at real contests is overrated too.

Keywords: aerobic performance, epo doping, meta-analysis, professional road racing

1. Introduction

As part of our research on the (social) psychology of doping, focusing on professional cyclists [1,2,3,4,5], we reviewed studies investigating the relationship between epo doping and aerobic performance (the EDAP–relationship). With this review we sought to resolve a specific issue which is very relevant for the wider–ranging goals we planned to achieve with our research: “What is the magnitude of the EDAP–relationship observed in the epo laboratory studies?” As we will argue, a literature search did not permit a satisfying answer to this pivotal query. Therefore, we decided to estimate the strength of this relationship in an empirical way by conducting a meta–analysis on the findings of all seventeen experiments that hitherto examined this relationship. VO_{2max} , the maximal oxygen uptake during intense aerobic exercise, and the associated maximal aerobic power output in watts (W_{map}) are conventionally considered to be the most important measures to assess improvements in endurance performance. In their review of findings of epo studies to date, Lundby and Olsen [6] concluded that —if hematocrit (Ht) is artificially increased by epo administration from pre–test baseline values to around $Ht = 50\%$ post test— VO_{2max} is estimated to improve by 8–12%. Over the years, these findings led to the generally shared opinion that the performance–

enhancing (or ergogenic) effects of epo are ‘dramatic’ (cf. [7], see also [8,9,10]).

However, findings of field research [11,12,13,14,15] suggest that VO_{2max} is but one of the many determinants of endurance performance and other variables, such as lactate metabolism and biomechanical efficiency [11], also constitute important factors. What is more, in an extensive and critical review of the physiological processes involved in the EDAP–relationship, Heuberger and co–workers [16, p. 3] state “there is no scientific basis to conclude rHuEPO has performance enhancing properties in elite cyclists.”

1.2. Research Objective

The inconclusiveness of these research findings puzzled us. Note that Kuipers [13] and Heuberger et al. [16] contend that the strength of the EDAP–relationship observed in the epo studies is overestimated. If so, this contention would constrain the predictive validity and generality of the EDAP–relationship, for it would entail that the relationship between the ergogenic effects of epo on cyclists’ performances in real competitions might be overrated too. Effect size indices validly estimate the magnitude of experimental effects obtained in research [17,18]. However, the epo studies did not systematically report such indices. To resolve our uncertainty concerning the status of the EDAP–relationship, we therefore decided to estimate its strength by the well–known indices: d , r ,

and r^2 . Note that our purpose is not to critically appraise the proposed physiological processes involved in the EDAP–relationship. Other research [16] already provided a detailed account of these processes. Rather, our objective is to evaluate the magnitude of the relationship in an alternative way by means of statistical analyses.

2. Method

2.1. Design, Studies and Unit of Analysis

Basically, the design for the analysis constitutes a 2(Treatments: epo vs. placebo / control) x 2(Measurements: pre vs. post test) mixed ANOVA design with repeated measurements on the last factor and with VO_{2max} and W_{map} as the most frequently used dependent variables. We identified seventeen published studies (1991–2010) through database searching and did not include unpublished studies. All seventeen studies were eligible and included in the analysis. The studies are marked by an asterisk in the references [7,19–34]. Table S1 (supporting material and information, see appendix) presents an overview of the characteristics of the studies.

The studies comprised a total of 186 participants in the epo treatments and 107 participants in the placebo / control treatments. Seven (41%) of the studies were double blinded and six (35%) were single blinded. Three (18%) studies did not supply sufficient information concerning their method. In the Rasmussen et al. (2010) study [30], one condition was double blinded, but in the other condition (6%) there was no specific reference to this feature. In one experiment [22] professional athletes took part, whilst in another study [19] only in one condition athletes served as participants. In the remainder of the studies (88%) all participants were non–athletes.

Comparisons of pre vs. post test performances within epo and placebo / control treatments (if present) as well as comparisons of experimental treatments on pre and post tests separately constitute the unit of analysis in the design. Table S1 indicates that the analysis involved a total of 59 separate comparisons (or ‘strata’). Thirty–eight (64%) related to epo vs. placebo treatment comparisons and 21 (36%) involved only epo treatment comparisons without placebo or control conditions. Of these 59 comparisons, 32 (54%) used VO_{2max} as the dependent variable, of which 25 presented this measure in $ml \cdot kg^{-1} \cdot min^{-1}$ and seven in $l \cdot min^{-1}$. Twenty (34%) of the comparisons additionally used W_{map} as a dependent variable, whereas seven (12%) of them expressed aerobic performance in different measurements. Next to VO_{2max} and W_{map} , Ekblom and Berglund [25] and Thomsen et al. [33] assessed participants’ work time in seconds (five comparisons). Rasmussen et al. [30] measured participants’ aerobic capacity in kilojoules (kJ, two comparisons)

2.2. Inclusion and Exclusion Criteria

We expressed VO_{2max} in $ml \cdot kg^{-1} \cdot min^{-1}$. However, two studies [19,21] presented this variable in $l \cdot min^{-1}$, but did not provide sufficient information concerning participants’ body mass, a variable which is germane to convert $l \cdot min^{-1}$ to $ml \cdot kg^{-1} \cdot min^{-1}$. This does not pose a problem for the estimation of effect sizes, but it has consequences for the estimation of the epo–induced increases in VO_{2max} and

W_{map} from pre to post tests. Therefore, we were forced to exclude findings of these two studies from analyses when assessing these improvements. Russell et al. [32] reported incomplete data for six W_{map} –comparisons and they were therefore not used in the analyses. Besides, experiments have shown that the ergogenic effects of epo are less strong at higher altitudes [19,27,31]. Yet, altitudes of 2500–3000 m are not uncommon in professional road racing. At the Tour de France for instance, famous climbs such as the Galibier (2646 m) approximate such altitudes. Therefore, we decided to include six comparisons of studies that involved effects of simulated, moderate hypoxia conditions of < 3000m on performance, but excluded six comparisons of altitudes $\geq 3000m$ [27,31]. Last, we included experiments which examined submaximal exercises, e.g., performing at 80% of VO_{2max} [33] or at workloads of 100 or 200W (e.g., some treatments in [25,28]). Application of these criteria led to varying numbers of strata to estimate effect sizes, which are presented in the tables and supporting materials.

2.3. Assessments

Table S1 reveals that sample sizes in the epo studies are small. We therefore estimated effect sizes by Hedges and Holkin’s [17] unbiased d –index (see appendix Table S2). Conventionally, a $d < 0.20$ is considered a trivial effect, a d of 0.20 to 0.40 a small effect, around $d = 0.50$ a moderate effect, and a $d \geq 0.80$ a large effect. Some studies presented descriptive statistics in SE , which we converted to SD to estimate pooled SDs (see Table S1). To assess the amount of variation explained by the various treatments in aerobic performance, we next converted the resulting unbiased d to r and r^2 , using the formulae provided by Rosenthal [18,35] (see appendix Table S2).

We further specifically assessed what we label the EDAP effect size index, or EDAPES. To this end, we subtracted the d ’s produced by the experimental and placebo / control treatments from pre to post tests from each other. We did the same for the d ’s that resulted from differences produced by the experimental treatments on separate pre and post tests. To appraise the amounts of variation explained by the EDAPES in aerobic performance, we subsequently converted the resulting differential d ’s to r and r^2 .

2.4. Analyses

We conducted four analyses to address our research objective. First, we estimated the unbiased d for all dependent measures relating to all comparisons we distinguished, including performances measured in moderate hypoxia conditions, as well as submaximal and maximal performances demonstrated in studies with or without control / placebo treatments. We next performed similar analyses for VO_{2max} and W_{map} , only. We subsequently refined our analyses to measurements of VO_{2max} and W_{map} that solely related to maximal performances demonstrated in normoxia conditions (sea level) obtained in double blind, placebo controlled studies. In all these analyses, we checked for fixed and random effects, estimated by Cochran’s Q and the associated I^2 –statistic. We further checked for potential bias in the studies using the method developed by Egger and co–workers [36]. Main results of the analyses are graphically presented by

bias assessment and Forest plots. In the third analysis, we evaluated the magnitude of the EDAP_{ES}, estimated by d , r and r^2 . Finally, in the fourth analysis, we assessed the epo-induced improvement in performance from pre to post tests within the epo treatments of the studies. Next to

mean improvements ($M \pm SD$), we calculated the proportional increases (%) obtained on VO_{2max} and W_{map} .

We conducted all analyses using the *StatsDirect* Statistical Software[®] package (version 2.7.9.2013).

Table 1. Effect Size Indices and Associated Parameters of Experimental Treatments on the Post Tests and of Pre vs. Post Test Comparisons within Epo Treatments

	Comparisons / Variables 1	Strata	Fixed Effect: d (CI95%)	Heterogeneity: Q	Inconsistency: I^2 (CI95%)	Random Effect: d (CI95%)	Bias
A	Epo vs. control / placebo conditions, post tests (all measurements)	32	0.43** (0.25–0.62)	77.97**	60% (38–72)	0.44** (0.15 -0.74)	0.63
	Epo conditions, placebo controlled, pre vs. post tests (all measurements)	-	0.54** (0.36–0.71)	28.63	0% (0–38)	-	2.22*
B	Epo conditions, pre vs. post tests (VO_{2max} and W_{map})	40	0.54** (0.39–0.69)	58.55*	33% (0–54)	0.57** (0.38 – 0.76)	3.48**
	Epo conditions, pre vs. post tests (VO_{2max})	30	0.45** (0.28–0.63)	34.10	15% (0–46)	-	2.21*
	Epo conditions, pre vs. post tests (W_{map})	10	0.85** (0.52–1.18)	20.12*	55% (0–76)	0.92** (0.43 -1.42)	10.49**
C	Epo vs. control / placebo conditions, post tests (VO_{2max} and W_{map})	17	0.52** (0.27–0.78)	12.85	0% (0–45)	-	1.58
	Epo conditions, pre vs. post tests (VO_{2max} and W_{map})	-	0.50** (0.27–0.73)	6.61	0% (0–45)	-	1.10
D	Epo vs. control / placebo conditions, post tests (VO_{2max})	14	0.62** (0.34–0.90)	9.09	0% (0–47)	-	1.13
	Epo conditions, pre vs. post tests (VO_{2max})	-	0.50** (0.24–0.75)	5.99	0% (0–47)	-	1.17
E	Epo vs. control / placebo conditions, post tests (W_{map}) ²	3	0.11 (-0.47–0.69)	1.36	0% (0–73)	-	-
	Epo conditions, pre vs. post tests (W_{map})	-	0.53 [‡] (-0.03–1.08)	0.61	0% 0–73)	-	-

Notes:

1. **Panel A** refers to all measurements and all studies, and includes performances measured in hypoxia conditions (< 3000 m) as well as submaximal and maximal performances. **Panel B** presents combined and separate findings from pre to post test within epo treatments only. The findings relate to VO_{2max} and W_{map} but combines findings of single blind and double blind, placebo controlled studies and include performances measured in moderate hypoxia conditions and as well as submaximal and maximal performances. Since the panel combines single and double blind studies, this means that separate pre and post test comparisons cannot be estimated. **Panels C – E** present combined and separate findings of VO_{2max} and W_{map} that relate to maximal performances measured at sea level and are restricted to findings of double blind, placebo controlled studies.

2. Estimations of W_{map} have low power, due to a low number of comparisons, and are therefore unreliable.

[‡] $p \leq .10$; * $p \leq .05$; ** $p \leq .01$

3. Results

3.1. Control / Placebo Treatments

Table S3 in the appendix presents the d -values produced by the control / placebo treatments of the studies from pre to post tests and by the epo vs. control / placebo treatments on the pre tests. Panels A, B, and C indicate trivial effect sizes, $d = -0.02$ – 0.18 . Although all estimates are not significant, positive values indicate minor performance progress from pre to post tests in the control / placebo treatments. The positive values obtained on the pre test estimates further designate that participants in the epo treatments performed slightly better than their

counterparts in the control / placebo treatments. Findings relating to W_{map} in Panel D of Table S3 indicate an effect size for the pre to post test comparisons of $d = 0.20$. The $d = -0.25$ obtained on the pre test comparisons designates that participants in the control / placebo conditions performed better than in the epo treatments. Note, however, that the number of these comparisons is very small ($N = 3$). These analyses thus have low power, resulting in unreliable effect size estimates.

Table S3 further shows that all Q -tests yielded nonsignificant results with 0% inconsistency, indicating fixed effects. The analyses further revealed no significant sign of bias for any of the comparisons.

3.2. Epo Treatments

3.2.1. Heterogeneity and Bias

Table 1 presents *d*-values yielded by the epo treatments of the studies from pre to post tests and by the epo vs. control / placebo treatments on the post tests. Before describing findings relating to effect size estimates, we will first discuss potential heterogeneity of the studies and bias.

Panel A and B in Table 1 show significant results for the *Q*-statistic in three of the five series comparisons. Four of the five series of comparisons also reveal significant bias. Note that Panel A presents findings of all measurements, observed in various experimental treatments that were designed by the researchers, with or without control / placebo conditions. Panel B combines observations on VO_{2max} and W_{map} from pre to post tests observed in studies with or without control / placebo conditions that again relate to a mixture of experimental treatments and performances.

Figure 1 presents the bias assessment plot of the series of pre vs. post test comparisons that consists of the largest number of strata ($N = 40$, Panel B) we examined. The plot reveals lateral asymmetry. The associated Forest plot in Figure 2 indicates that the bias can be traced to studies that generated small negative *d*'s on submaximal performances, produced by four comparisons in the Ekblom and Berglund [25] study. The bias can further be traced to large positive *d*'s ≥ 2.00 for W_{map} , observed in the study of Thomsen et al. [33], which also examined submaximal performances. One condition in the study by Lundby et al. [28], which assessed the influence of epo-induced augmented oxygen capacity on (sub)maximal performances, also yielded a large *d*-value, but only on W_{map} . To scrutinize the effects on W_{map} more closely, we inspected the bias assessment and Forest plots for this variable. It consists of ten comparisons (Table 1, Panel B, last row) and the plots can be seen in Figure 3, Figure 4. Figure 3 reveals a strong asymmetry, due to the same two studies that yielded large *d*'s, which we described above, and to relatively smaller *d*'s for the remaining studies (see Figure 4). Note that this specific series of comparisons again consists of a mixture of studies with or without control / placebo treatments, which either assessed performances in moderate hypoxia or in normoxia conditions, or measured maximal and submaximal performances.

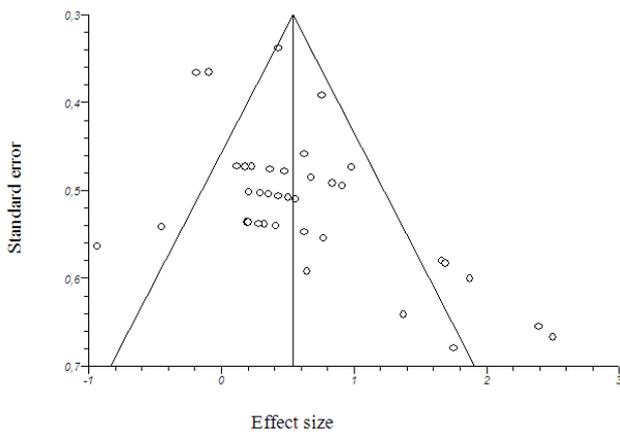


Figure 1. Bias assessment plot of pre vs. post test comparisons on VO_{2max} and W_{map} within epo treatments. ($N = 40$ strata, see Panel B in Table 1)

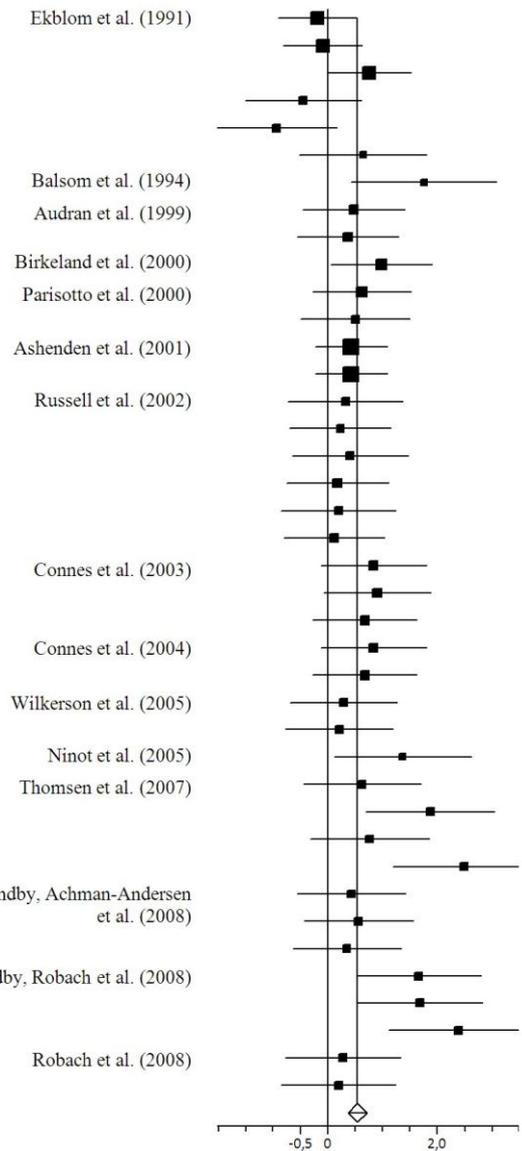


Figure 2. Forest plot of pre vs. post test comparisons on VO_{2max} and W_{map} within epo treatments

The plot includes comparisons of studies with and without control / placebo treatments, which measured submaximal as well as maximal performances in moderate hypoxia and normoxia conditions ($N = 40$ strata, see Panel B in Table 1). The size of the solid black squares represents the weight that the corresponding study exerts in the meta-analysis. Confidence intervals of estimates are displayed as a horizontal line through the black square. The unfilled diamond presents the pooled estimate.

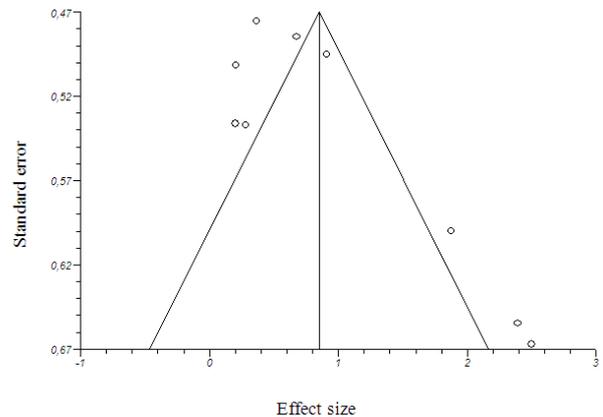


Figure 3. Bias assessment plot of pre vs. post test comparisons on W_{map} within epo treatments. ($N = 10$ strata, see Panel B in Table 1)

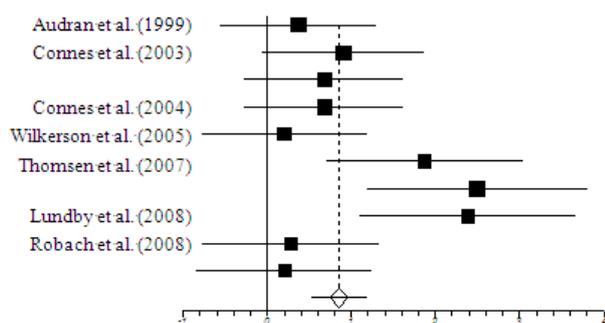


Figure 4. Forest plot of pre vs. post test comparisons on W_{map} within epo treatments. The plot includes comparisons of studies with and without control / placebo treatments, which measured submaximal as well as maximal performances in moderate hypoxia and normoxia conditions ($N = 10$ strata, see Panel B in Table 1). The size of the solid black squares represents the weight that the corresponding study exerts in the meta-analysis. Confidence intervals of estimates are displayed as a horizontal line through the black square. The unfilled diamond presents the pooled estimate

Examination of various other bias assessments plots, all relating to findings presented in Panel A of Table 1, indicated that bias was also observed in studies that examined performance improvements in moderate hypoxia conditions (e.g., [19,31]). These studies revealed that post test performances of participants in the epo treatments were sometimes lower than performances of participants in the control conditions. The studies, however, expected these effects of altitude to occur.

From these findings we can conclude that the heterogeneity and bias of the studies can mainly be attributed to specific performances measured in distinct studies (submaximal performances) or to specific experimental treatments (moderate hypoxia). This means that the epo studies can be considered homogeneous, i.e., they were conducted in similar circumstances and all recruited comparable participants. This conclusion is corroborated by findings presented in Panels C, D and E of Table 1. In these panels we methodically left out all these specific performances and / or experimental treatments. As can be seen in the table, the studies show no significant bias for all of these additional series of comparisons and the nonsignificant values of the Q -statistic and the related I^2 -index all suggest consistent, fixed effects and, thereby, homogeneous studies.

3.2.2. Effect Size Estimates

As regards these estimates, findings in Table 1 reveal a very consistent pattern. Considering fixed effects, Panels A and B in Table 1 show estimates that range from $d = 0.43$ (Panel A, all measurements) to $d = 0.85$ (Panel B, W_{map}). As to random effects, the values vary between $d = 0.44$ (Panel A, all measurements) and $d = 0.92$ (Panel B, W_{map}). As we already outlined, the series of comparisons relating to W_{map} consist of a broad array of measurements and experimental treatments, leading to strong variation in participants' power output. After restricting estimations solely to $VO_{2\text{max}}$ and W_{map} performances realized in normoxia conditions in double blind, placebo controlled studies, the variation in these performances is greatly reduced. This can be seen in Panel C and D of Table 1. They show moderate, yet significant estimates, $d = 0.50$ – 0.62 . As discussed earlier, findings presented in Panel E,

relating to W_{map} , are unreliable, so we will refrain from discussing these results. All in all, we conclude that, across a variety of measurements, comparisons and studies, the estimated pooled, fixed effect size of the EDAP-relationship is moderate and, on average, amounts to $d = 0.55$. This indicates a shift of approximately half SD in performances of the epo-treated compared to the non-treated participants.

3.2.3. EDAP_{ES}

Table 2 presents findings relating to the EDAP_{ES}. The estimates in this table are restricted to double blind, placebo controlled studies. Again we refined our analyses, i.e., we first assessed effects on all measurements and next examined combined and separate effects on $VO_{2\text{max}}$ and W_{map} .

The second column in the table reveals differential d -values for the experimental treatments after subtracting the d -values observed on the pre tests from the d -values obtained on the post tests of the studies. Obviously, since we checked for the effects of the control / placebo conditions, the estimates of these comparisons are lower compared to the estimates presented in the previous section. Again, however, the EDAP_{ES} reveals moderate effect sizes, $d = 0.41$ – 0.45 . The associated correlation coefficients vary between $r = .19$ – $.22$, and the amounts of variation explained in aerobic performance between $r^2 = .04$ – $.05$. Findings relating to pre vs. post test comparisons again yielded modest estimates, $d = 0.46$ – 0.49 . The correlation coefficients vary between $r = .42$ – $.44$, and the amounts of explained variation range from $r^2 = .17$ – $.19$. For reasons we outlined earlier, the findings concerning W_{map} (last row in Table 2) will not be discussed. Note that effects produced by the two EDAP_{ES}-comparisons we discussed, differ in strength of the resulting correlations and associated amounts of explained variation. This can be attributed to the different ways in which the r and r^2 indices are estimated for dependent and independent samples (see appendix Table S2).

3.2.4. Performance Improvement

Table 3 presents findings concerning the average increases and percents improvement in $VO_{2\text{max}}$ and W_{map} from pre to post epo treatment. Panel A refers to all comparisons including submaximal and maximal performances observed in studies with or without control / placebo conditions, demonstrated in normoxia as well as moderate hypoxia conditions (< 3000 m). Panel B presents improvements in maximal performances realized at sea level, obtained in double blind, placebo controlled studies only.

As to post test performances, if we take account of the sequence in exclusion criteria we applied, Table 3 shows that epo administration increases $VO_{2\text{max}}$ going from Panel A to B: $M_s = 54.77$ and 64.39 ml $\text{kg}^{-1} \text{min}^{-1}$, respectively. The improvements from pre to post tests range between $M = 3.19$ – 4.02 ml $\text{kg}^{-1} \text{min}^{-1}$. The percents improvement vary between $M = 6.18$ – 6.65% . The post test performances on W_{map} also improve going from Panel A to B: $M_s = 365$ and 398 W. In both panels, the increase from pre to post tests amounts to $M = 26$ W, whilst the percents improvement range between $M = 6.98$ – 7.67% .

Similar to findings relating to effect sizes estimates, the improvements we obtained are also very regular. This is the case, irrespective of the kind of performances or

experimental treatments which we included in (or excluded from) analyses. In every distinct analysis we conducted, the proportional improvements for VO_{2max} hover around 6–7% and for W_{map} around 7–8%. The

strongest post test performances can be seen in double blind, placebo controlled conditions with performances assessed at sea level.

Table 2. Fixed Effect Size Indices of the EDAP_{ES}

		Pre Tests ^a			Post Tests ^a			Post vs. Pre Tests ^b		
	Variables / Treatments / Comparisons	d	r	r ²	d	r		d	r	r ²
All	Epo vs. control / placebo	-0.02	<.01	< .01	.43	.21	0.4	.54	.48	.23
	Control / placebo	-	-	-	-	-	-	.08	.08	< .01
	EDAP_{ES}				.45	.22	.05	.46	.42	.17
VO _{2max} / W _{map}	Epo vs. control / placebo	.11	.05	< .01	.52	.25	.06	.50	.45	.20
	Control / placebo	-	-	-	-	-	-	.04	.04	< .01
	EDAP_{ES}				.41	.19	.04	.46	.42	.18
VO _{2max}	Epo vs. control / placebo	.18	.09	< .01	.62	.29	.09	.50	.45	.20
	Control / placebo	-	-	-	-	-	-	.006	.006	< .01
	EDAP_{ES}				.44	.21	.04	.49	.44	.19
W _{mapc}	Epo vs. control / placebo	-.25	-.12	.02	.11	.055	< .01	.53	.47	.22
	Control / placebo	-	-	-	-	-	-	.20	.20	.04
	EDAP_{ES}							.33	.31	.10

Notes:

a. For the differential d's, we subtracted estimates of epo vs. control placebo treatments on the pre test from the post test estimates and then converted the values to r and r².

b. For the differential d's, we subtracted estimates of post vs. pre test comparisons in the control / placebo treatments from estimates in the epo treatments.

c. Estimates are unreliable, due to low number of comparisons.

Table 3. Descriptive Statistics of Improvement in Aerobic Performance (VO_{2max} and W_{map}) in the Epo Treatments from Pre to Post Tests

Comparisons	Variable	Epo Treatment		Improvement from Pre Test	
		Pre Test	Post Test	Δ	%
A	VO _{2max} (N = 28)	M (±SD) 51.58 (13.62)	54.77 (15.20)	3.19 (1.83)**	6.18
		95% – CI 46.29 – 56.86	48.87 – 60.66	2.48 – 3.90	5.36 – 6.86
	W _{map} (N = 10)	M (±SD) 339 (43)	365 (46)	26 (11) **	7.67
		95% – CI 308 – 370	333 – 398	18 – 34	5.84 – 9.18
B	VO _{2max} (N = 14)	M (±SD) 60.37 (4.31)	64.39 (4.98)	4.02 (1.14) **	6.65
		95% – CI 57.88 – 62.85	61.51 – 67.26	3.36 – 4.68	5.81 – 7.45
	W _{map} (N = 3)	M (±SD) 372 (53)	398 (57)	26 (5) **	6.98
		95% – CI 241 – 502	256 – 540	15 – 38	6.22 – 7.57

Notes:

Panel A refers to all comparisons including performances observed in studies without control / placebo conditions, performances in hypoxia conditions (< 3000 m) as well as submaximal and maximal performances. Panel B presents maximal performances at sea level obtained in double blind, placebo controlled studies only. Findings of [19] and [21] are excluded from analysis.

** p ≤ .01 by paired t – test performed on the aggregated data

4. Discussion

4.1. Magnitude of the EDAP–Relationship

The studies we included in this meta–analysis were designed for a variety of (theoretical) reasons, some for the development of tests to detect epo abuse in athletes [29], whereas other studies attempted to assess the effects of epo on performance by mimicking anecdotal reports of epo abuse in cycling in their design [32]. Yet, all studies involved the administration of epo in varying units, at varying intervals, during intervention periods that varied in longevity. In many of the studies, these variations were specifically intended to influence (sub)maximal performances. In spite of this variation in studies, our findings consistently show that the effect size of the EDAP–relationship is modest. From the findings in Table 1 we concluded that, across a variety of measurements,

comparisons and studies, the estimated pooled, fixed effect size of the EDAP–relationship amounted to $d = 0.55$ on average. Table 2 indicated that the fixed estimates of the EDAP_{ES} varied between $d = 0.41–0.49$. All estimates indicate a shift of approximately half SD in performances of the epo–treated compared to the non–treated participants. But what do these numbers mean? One way to interpret them is to examine the amounts of variation in performance improvement attributable to epo treatment. Table 2 shows that these amounts varied between $r^2 = .04–.19$. This entails that 81–96% of the differences in performance improvement observed in the epo studies are not explained by epo administration. Another way to illustrate the moderate effect size is to examine the degree of (non–) overlap between maximal performances realized by participants in the control / placebo vs. the epo treatments of the studies. According to Cohen [37] a $d = .40$ indicates a non–overlap in performance of 27.4%, a $d = .50$ of 33%, and a $d = .60$ of 38.2%. These percents imply that in 61.8–72.6% of the observations the epo

studies are not able to discriminate between maximal performances demonstrated by participants that were administered epo or not. Taken together, all these results put pressure on the robustness of the EDAP–relationship.

4.2. Validity of the EDAP–Relationship

A next and much more relevant question relating to the wider–ranging goal we planned to achieve with our research on doping is: “Will these epo–induced improvements obtained on VO_{2max} and W_{map} be decisive to gain an edge and win a competition?” We hesitate to reply this question positively for various reasons. Some arguments relate to the external and ecological validity of the epo studies, while other arguments involve the predictive validity of the EDAP–relationship, which concerns the specific association between VO_{2max} and cycling speed in mean kilometers per hour (km/h). We will start our discussion with the latter relationship.

Nevill, Jobson, Palmer and Olds [38] report that the relationship between VO_{2max} and km/h is curvilinear. Their analysis of data of three time–trial cycling studies yielded a proportional curvilinear association between km/h and energy cost: $Km/h = 36.1 \cdot VO_{2max}^{0.41} m^{-0.13}$. In this equation m denotes cyclists’ body mass in kg. Nevill et al. [38] used this equation to predict cyclists’ mean time trial performances, describing an example of a male cyclist ($m = 72.2kg$) who aspired to increase his speed from 30 to 35km/h. They calculated that he would need to increase his VO_{2max} from 2.36 to 3.44 $l\text{min}^{-1}$, a difference of 1.08 $l\text{min}^{-1}$. However, an improvement of 40 to 45km/h would require an increase in VO_{2max} from 4.77 to 6.36 $l\text{min}^{-1}$, a difference of 1.59 $l\text{min}^{-1}$. Importantly, Nevill et al. [38] note that to reach 45km/h it would take the cyclist an increase in VO_{2max} of more than 50%. And it would require a massive increase of more than 100% in VO_{2max} to attain a speed of 50km/h (8.55 $l\text{min}^{-1}$). These required proportional increases dwarf the 6–7% epo–induced improvements in VO_{2max} found in our meta–analysis. Notably, these relationships also mean that, for highly trained endurance athletes such as cyclists, increases in VO_{2max} will lead to proportionally smaller improvements in performance [16]. Top–level cyclists are known to have high VO_{2max} levels [39]. Consequently, this implies that the epo–induced improvements in VO_{2max} we found, will have very limited effects on their performances [16].

To determine improvements in km/h corresponding to the epo–induced increases in VO_{2max} we obtained, we applied Nevill et al.’s [38] equation to the findings presented in Table 3 (Panel B), which concern the most extreme improvements in performance we found. We assumed cyclists’ body weight to be $m = 72kg$ and converted our VO_{2max} values (which we expressed in $ml\text{kg}^{-1}\text{min}^{-1}$) to $l\text{min}^{-1}$. This conversion yielded an average of 4.35 $l\text{min}^{-1}$ for the pre tests and 4.64 $l\text{min}^{-1}$ for the post tests, an improvement of 0.29 $l\text{min}^{-1}$. The corresponding cycling speeds then become 37.8km/h for the pre tests and 38.8km/h for the post tests, an increment of 1km/h. Nevill et al.’s arguments, however, further entail that increases in VO_{2max} with associated increases in km/h will progressively diminish the assumed epo–induced gains in time when racing. To illustrate this, let us take two riders in a cycling race. They race at 37km/h, and one of the

riders breaks away with a speed of 38km/h. This increase would yield him a 2.6 seconds gain in time per kilometer. If both riders race at 47km/h and one of them breaks away at 48km/h, this would yield him 1.6 seconds per kilometer. However, compared to the 38km/h, the 48km/h–attempt would require a substantial increase of 76.8% in VO_{2max} . In other words, higher speeds require progressively stronger aerobic effort and energy cost, but will result in smaller gains in time. Accordingly, these insights put even greater pressure on the EDAP–relationship. They show that the application of improvements in VO_{2max} observed in the epo studies cannot be linearly extrapolated to cycling races and that the epo–induced increases in speed result in nearly trivial time differences between cyclists.

The EDAP–relationship becomes even more limited if we consider the well–known fact that elite athletes, such as pro 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 [40,41,42]. So, after this period athletes’ exercise capacity will be greatly reduced. From these arguments we can derive that the influence of epo on cyclists’ performances is also strongly constrained by time limits.

Generalization problems also arise when extrapolating laboratory findings to elite cyclists. Several scholars have discussed limitations of laboratory research on the physiological processes involved in endurance performance, particularly where it concerns achievements of top–level athletes [16,43,44,45]. They question the generality of findings of this research, because its main body simply does not involve top–level athletes. This criticism also holds true for the epo studies in which 88% of the participants were non–athletes. Given these generalization problems, Joyner and Coyle [41, p. 42] in their study on the physiology of champions, note that “more work is needed on highly trained athletes performing very intense exercise in real or simulated competitions.”

A final and related problem concerns the ecological validity of the epo studies. Various ex–pro cyclists, commenting on their experiences in races in which they participated, provide anecdotal evidence that weather, terrain and road conditions, between–team cooperative efforts in the bunch, the powerful influence of dominant leaders (Anquetil, Merckx, Hinault, Armstrong), status conflicts between rival team leaders, riders’ assigned roles in the teams, short– and long–term tactical considerations during the races, and individual psychological factors such as hardiness and determination, are far more important than a single physical factor, such as VO_{2max} , to establish who will be the victor, who will be in second place, and even who will finish a race [2]. In other words, VO_{2max} is a necessary, but not a sufficient condition to win a competition and static laboratory circumstances are a far cry from the (group)dynamic, unpredictable, and strategic circumstances that characterize three–week, multi–stage cycling races such as the Tour de France. To illustrate this argument more vividly, we wonder how a randomly selected participant of one of the epo studies would fare, if (s)he was allowed to take part in a brutal, 265km long one–day classic race such as Paris–Roubaix, or in a mountain stage climbing up l’Alpe d’Huez after having mounted three other first–category climbs on the same day. Given the physical demands of these races [46] the pivotal

question arises: “How long would such a participant last, if at all?” What would be the relevance of examining how strong the influence of epo would be in such a case?

5. Conclusion

Heuberger and colleagues [16] contend that the EDAP–relationship lacks scientific evidence. They based their conclusion on a critical appraisal of the intricate physiological processes involved in this relationship. We followed an alternative path and examined the strength of the relationship in a statistical way without paying attention to these processes. Yet, both studies lead to the same conclusion: The magnitude of the EDAP–relationship is overvalued. In turn, this entails that the relationship between epo doping and cyclists’ performances at real contests is overrated too.

Statement of Competing Interests

The authors have no competing interests.

References

- [1] Brouwer, B., Lodewijckx, H. F. M. and Kuipers, H., “Doping confessions of cyclists revisited.” [In Dutch: “Doping-bekentenissen van wielrenners langs de wetenschappelijke meetlat”]. *Sportpsychologie Bulletin*, 20: 24-37, 2009.
- [2] Lodewijckx, H.F.M., Give some and take some: Cycling as double play. [In Dutch: Tussen geven en nemen: Wielrennen als dubbelspel]. 2010 Uitgevers, Rotterdam, 2011.
- [3] Lodewijckx, H. F. M. and Brouwer, B. “Some empirical notes on the ‘epo epidemic’ in professional cycling.” *Research Quarterly for Exercise and Sport*, 82(4): 593-608, 2011.
- [4] Lodewijckx, H. F. M. and Brouwer B., “Tour, Giro Vuelta: Rapid progress in cycling performance starts in the 1980s.” *International Journal of Sports Science*, 2(3): 24-31, 2012.
- [5] Lodewijckx, H.F.M. and Verboon, P., “Lance Armstrong’s era of performance – Part 1: Are his time trial performances much different from other winners?” *Journal of Athletic Enhancement*, 2: 1, 2013.
- [6] Lundby, C. and Olsen, N. V., “Effects of recombinant human erythropoietin in normal humans.” *Journal of Physiology*, 589(6):1265-1271, 2011.
- [7] Ninot, G., Connes, P., and Caillaud, C., “Effects of recombinant human erythropoietin injections on physical self in endurance athletes.” *Journal of Sports Sciences*, 24: 383-391, 2006. (meta-analysis).
- [8] Catlin, D. H., Hatton, C. K., and Lasne, F., “Abuse of recombinant erythropoietins by athletes.” In G. Molineux, M. A. Foote, and S. G. Elliott (Eds.), *Erythropoietins and erythropoiesis*. Birkhäuser Verlag, Basel, 2006, 205-228.
- [9] Lasne, F., “Erythropoietin doping: Detection in urine.” In J. L.Fourcroy (Ed.), *Pharmacology, doping and sport*: Routledge, London, 2009, 107-124.
- [10] Varlet-Marie, E., Gaudard, A., Audran, M., and Bressolle, F., “Pharmacokinetics / pharmacodynamics of recombinant human erythropoietins in doping control.” *Sports Medicine*, 33(4): 301-315, 2003.
- [11] Bassett, D. R. and Howley, E. T., “Limiting factors for maximum oxygen uptake and determinants of endurance performance.” *Medicine and Science in Sports and Exercise*, 32: 70-84, 2000.
- [12] Coyle, E. F., Coggan, A. R., Hopper, M. K., and Walters, T. J., “Determinants of endurance in well-trained cyclists.” *Journal of Applied Physiology*, 64: 2622-2630, 1988.
- [13] Kuipers, H., “Putative effects of doping in cycling.” [In Dutch: “Vermeende effecten van doping in de wielersport”]. *Nederlands Tijdschrift voor Geneeskunde*, 150(48): 2643-2645, 2006.
- [14] Lucia, A., Hoyos, J., Pérez, M., Santalla, A., and Chicharro, J. L., “Inverse relationship between VO₂max and economy/efficiency in world-class cyclists.” *Medicine and Science in Sports and Exercise*, 34: 2079-2084, 2002.
- [15] Lucia, A., Hoyos, J., Pérez, M., Santalla, A., Earnest, C. P., and Chicharro, J. L., “Which laboratory variable is related with time trial performance time in the Tour de France?” *British Journal of Sports Medicine*, 38: 636-640, 2004.
- [16] Heuberger, J. A. A. C., Cohen-Tervaert, J. M., Schepers, F. M. L., Vliegthart, A. D. B., Rotmans, J. I., Daniels, J. M. A., Burggraaf, J. and Cohen, A. F., “Erythropoietin doping in cycling: Lack of evidence for efficacy and a negative risk–benefit.” *British Journal of Clinical Pharmacology*, 2012.
- [17] Hedges, L. V. and Olkin, I., *Statistical methods for meta-analysis*. Academic Press, New York, 1985.
- [18] Rosenthal, R., Rosnow, R. L., and Rubin, D. B., *Contrast and effect sizes in behavioral research: A correlational approach*. Cambridge University Press, Cambridge, UK, 2000.
- [19] Ashenden, M. J., Hahn, A. G., Martin, D. T., Logan, P., Parisotto, R., and Gore, C. J., “A comparison of the physiological response to simulated altitude exposure and r-HuEpo administration.” *Journal of Sports Sciences*, 19: 831-837, 2001. (meta-analysis).
- [20] Audran, M., Gareau, R., Matecki, S., Duran, F., Chenard, C., Sicart, M.T. et al., “Effects of erythropoietin administration in training athletes and possible indirect detection in doping control.” *Medicine and Science in Sports and Exercise*, 31: 639-645, 1999. (meta-analysis).
- [21] Balsom, P. D., Ekblom, B., and Sjodin, B., “Enhanced oxygen availability during high intensity intermittent exercise decreases anaerobic metabolite concentrations in blood.” *Acta Physiologica Scandinavica*, 150: 455-456, 1994. (meta-analysis).
- [22] Birkeland, K. I., Stray-Gundersen, J., Hemmersbach, P., Hallen, J., Haug, E., and Bahr, R., “Effect of rhEPO administration on serum levels of sTfR and cycling performance.” *Medicine and Science in Sports and Exercise*, 32: 1238-1243, 2000. (meta-analysis).
- [23] Connes, P., Caillaud, C., Mercier, J., Bouix, D., and Casties, J. F., “Injections of recombinant human erythropoietin increases lactate influx into erythrocytes.” *Journal of Applied Physiology*, 97(1): 326-332, 2004. (meta-analysis).
- [24] Connes, P., Perry, S., Varray, A., Prédaut, C., and Caillaud, C., “Faster oxygen uptake kinetics at the onset of submaximal cycling exercise following 4 weeks recombinant human erythropoietin (rHuEPO) treatment.” *European Journal of Physiology*, 447: 231-238, 2003. (meta-analysis).
- [25] Ekblom, B. and Berglund, B., “Effect of erythropoietin administration on maximal aerobic power.” *Scandinavian Journal of Medicine and Science in Sports*, 1(2): 88-93, 1991. (meta-analysis).
- [26] Lundby, C., Achman-Andersen, N. J., Thomsen, J. J., Norgaard, A. M., and Robach, P., “Testing for recombinant human erythropoietin in urine: Problems associated with current anti-doping testing.” *Journal of Applied Physiology*, 105: 417-419, 2008. (meta-analysis).
- [27] Lundby, C. and Damsgaard, R., “Exercise performance in hypoxia after novel erythropoiesis stimulating protein treatment.” *Scandinavian Journal of Medicine and Science in Sports*, 16: 35-40, 2006. (meta-analysis).
- [28] Lundby, C., Robach, P., Boushel, R., Thomsen, J. J., Rasmussen, P., Koskolou, M. et al., “Does recombinant human epo increase exercise capacity by means other than augmenting oxygen transport.” *Journal of Applied Physiology*, 105: 581-587, 2008. (meta-analysis).
- [29] Parisotto, R., Gore, C. J., Emslie, K. R., Ashenden, M. J., Brugnara, C., Howe, C. et al., “A novel method utilizing altered erythropoiesis for the detection of recombinant human erythropoietin abuse in athletes.” *Haematologica*, 85: 564-572, 2000. (meta-analysis).
- [30] Rasmussen, P., Foged, E. M., Krogh-Madsen, R., Nielsen, J., Nielsen, T. R., Olsen, N. V. et al., “Effects of erythropoietin administration on cerebral metabolism and exercise capacity in men.” *Journal of Applied Physiology*, 109 (2): 476-483, 2010. (meta-analysis).
- [31] Robach, P., Calbet, J. A. L., Thomsen, J. J., Boushel, R., Mollard, P., Rasmussen, P. et al., “The ergogenic effect of recombinant human erythropoietin on VO₂max depends on the severity of arterial hypoxemia.” *PLoS ONE*, 3: 1-13, 2008. (meta-analysis).
- [32] Russell, G., Gore, C. J., Ashenden, M. J., Parisotto, R., and Hahn, A. G., “Effects of prolonged low doses of recombinant human erythropoietin during submaximal and maximal exercise.” *European Journal of Applied Physiology*, 86: 442-449, 2002. (meta-analysis).

- [33] Thomsen, J. J., Rentsch, R. L., Robach, P., Calbet, J. A. L., Boushel, R., Rasmussen, P. et al., "Prolonged administration of recombinant human erythropoietin increases submaximal performance more than maximal aerobic capacity." *European Journal of Applied Physiology*, 101 (4): 481-486, 2007. (meta-analysis).
- [34] Wilkerson, D. P., Rittweger, J., Berger, N. J. A., Naish, F., and Jones, A. M., "Influence of recombinant human erythropoietin treatment on pulmonary O₂ uptake kinetics during exercise in humans." *Journal of Physiology*, 568: 639-652, 2005. (meta-analysis).
- [35] Rosenthal, R., "Parametric measures of effect size." In H. Cooper and L. V. Hedges (Eds.), *The handbook of research synthesis*. Russell Sage Foundation, New York, 1994, 231-244.
- [36] Egger, M., Smith, G. D., Schneider, M., and Minder, C., "Bias in meta-analysis detected by a simple, graphical test." *British Medical Journal*, 315: 629-634, 1997.
- [37] Cohen, J., *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.
- [38] Nevill, A. M., Jobson, S. A., Palmer, G. S., Olds, T. S., "Scaling maximal oxygen uptake to predict cycling time-trial performance in the field: a non-linear approach". *European Journal of Applied Physiology*, 94: 705-710, 2005.
- [39] Jeukendrup, A. E., Craig, N. P., and Hawley, J. A., "The bioenergetics of World Class Cycling." *Journal of Science and Medicine in Sports*, 3: 414-433, 2, 2000.
- [40] Johnson, A. T., *Biomechanics and exercise physiology: Quantitative modeling*. CRC Press / Taylor and Francis Group, New York, 2007.
- [41] Joyner, M. J., and Coyle, E. F., "Endurance exercise performance: The physiology of champions." *Journal of Physiology*, 586: 35-44, 2008.
- [42] Wilmore, J. H., Costill, D. L., and Kenney, W. L., *Physiology of sport and exercise*. Human Kinetics, Champaign, IL, 2008.
- [43] Atkinson, G., Peacock, O., St Clair Gibson, A., and Tucker, R., "Distribution of power output during cycling: Impact and mechanisms." *Sports Medicine*, 37: 647-667, 2007.
- [44] Myburg, K. H., "What makes an endurance athlete world-class?: Not simply a physiological conundrum. *Comparative Biochemistry and Physiology - part A. Molecular and Integrative Physiology*, 136: 171-190, 2003.
- [45] Hopkins, W. G., Hawley, J. A., and Burke, L. M., "Design and analysis of research on sport performance enhancement." *Medicine and Science in Sports and Exercise*, 31: 472-485, 1999.
- [46] Lucia, A., Earnest, C., and Arribas, C., "The Tour de France: a physiological review." *Scandinavian Journal of Medicine and Science in Sports*, 13: 275-283, 2003.

Appendix: Supporting Material and Information

Table S1. Characteristics of Seventeen Epo Studies Included in the Meta-Analysis

Studies /comparisons	Researchers	SB /			VO _{2max} (V)		
		DB / NR ¹⁸	A / NA ¹⁹	Other (O)	W _{map} (W)	N _{Epo} ²⁰	N _{plac} ²¹
1-1	Ekblom & Berglund (1991) ¹ L+L/H-d. 100W	SB	NA	V		15	np
#-2	200 W	--	--	V		(15)	--
#-3	Max	--	--	V		(14)	--
#-4	Worktime	--	--	O (sec.)		(14)	--
#-5	L/H-d. 100W	--	--	V		7	--
#-6	200 W	--	--	V		(7)	--
#-7	Max	--	--	V		(6)	--
#-8	Worktime	--	--	O (sec.)		(6)	--
2-9	Balsom et al. (1994) ²	SB	NA	V		6	np
3-10	Audran et al. (1999) ³ VO _{2max}	SB	NA	V		9	np
#-11	W _{map}	--	--	W		(9)	--
4-12	Birkeland et al. (2000) ⁴	DB	A	V		10	10
5-13	Parisotto et al. (2000) ⁵ IM vs. control	DB	NA	V		10	9
#-14	OR vs. control	--	--	V		8	(9)
6-15	Ashenden et al. (2001) ⁶ Epo vs. contr.	DB	A/NA	V		18	22
#-16	Epo vs. placebo	--	--	V		(18)	7
7-17	Russell et al. (2002) ⁷ 4 wk IV VO _{2max}	NR	NA	V		7	5
#-18	4 wk IV W _{map} (data incomplete)	--	--	W		(7)	(5)
#-19	4 wk OR VO _{2max}	--	--	V		9	5
#-20	4 wk OR W _{map} (idem)	--	--	W		(9)	(5)
#-21	8 wk IV VO _{2max}	--	--	V		(7)	(5)
#-22	8 wk IV W _{map} (idem)	--	--	W		(7)	(5)
#-23	8 wk OR VO _{2max}	--	--	V		(9)	(5)
#-24	8 wk OR W _{map} (idem)	--	--	W		(9)	(5)
#-25	12 wk IV VO _{2max}	--	--	V		(7)	(5)
#-26	12 wk IV W _{map} (idem)	--	--	W		(7)	(5)
#-27	12 wk OR VO _{2max}	--	--	V		(9)	(5)
#-28	12 wk OR W _{map} (idem)	--	--	W		(9)	(5)
8-29	Connes et al. (2003) ⁸ IM VO _{2max}	DB	NA	V		9	7
#-30	Epo vs. control submax	--	--	W		(9)	(7)
#-31	Epo vs. control max	--	--	W		(9)	(7)
9-32	Connes et al. (2004) ⁹ VO _{2max}	DB	NA	V		9	7
#-33	W _{map}	--	--	W		(9)	(7)
10-34	Wilkerson et al. (2005) ¹⁰ VO _{2max}	DB	NA	V		8	7
#-35	W _{map}	--	--	W		(8)	(7)
11-36	Ninot et al. (2006) ¹¹ VO _{2max}	DB	NA	V		6	5
12-37	Lundby / Damsgaard (2006) ¹² VO _{2max} (hypoxia conditions, 4100 m)	NR	NA	V		10	np
#-38	W _{map} (idem)	--	--	W		(10)	--
#-39	VO _{2max} (idem)	--	--	V		(10)	--
#-40	W _{map} (idem)	--	--	W		(10)	--
13-41	Thomsen et al. (2007) ¹³ 4 wk VO _{2max}	SB	NA	V		7	8
#-42	4 wk W _{map}	--	--	W		(8)	(8)
#-43	11 wk VO _{2max}	--	--	V		(7)	(8)
#-44	11 wk W _{map}	--	--	W		(8)	(8)
#-45	4 weeks Time to Exhaustion	--	--	O (sec.)		(8)	(8)

#-46	11 week Time to Exhaustion	--	--	O (sec.)	(8)	(8)
#-47	11 week Time to Exhaustion (rel. intens. as before epo)	--	--	O (sec.)	(8)	(8)
14-48	Lundby, Achman et al. (2008) ¹⁴ Day 35	NR	NA	V	8	np
#-49	Day 42	--	--	V	(8)	--
#-50	Day 49	--	--	V	(8)	--
15-51	Lundby, Robach et al. (2008) ¹⁵ 100W	SB	NA	V	8	np
#-52	Maximal	--	--	V	(8)	--
#-53	W _{map}	--	--	W	(8)	--
16-54	Robach et al. (2008) ¹⁶ 1500 m vs. control (sea level)	SB	NA	W	7	7
#-55	2500 m vs. control (idem)	--	--	W	(7)	(7)
#-56	3500 m vs. control (hypoxia condition)	--	--	W	(7)	(7)
#-57	4500 m vs. control (idem)	--	--	W	(7)	(7)
17-58	Rasmussen et al. (2010) ¹⁷ 3 days high	NR	NA	O (kJ)	7	7
#-59	3 month low	DB	--	O (kJ)	8	8

Notes:

Where required, control treatments in studies (without placebo) are placed under the category of placebo treatments.

1 Examination of the effect of erythropoietin administration on maximal and submaximal aerobic power

2 Examination of enhanced availability during high intensity intermittent exercise VO_{2max} expressed in l min⁻¹ and not converted to ml kg⁻¹ min⁻¹.

3 Conducted for indirect detection of epo abuse in doping control.

4 Examined using soluble transferrin receptor (sTFR) as an indicator of epo abuse. For placebo condition data are estimated from figures in paper.

5 Examined novel method for detection of epo abuse. SE converted to SD. Differences estimated from reported differences between conditions, which were added to reported pre test values. IV = intravenous iron supplement, OR = oral iron supplement which are compared to placebo treatment.

6 Compared effects of simulated hypoxia with epo administration Recruitment of control group not described. VO_{2max} expressed in l min⁻¹ and not converted to ml kg⁻¹ min⁻¹. Authors note strong interindividual variation in their data. In the hypoxia and control treatments participants were athletes, in the epo and placebo treatments they were non-athletes.

7 Examined effects of prolonged low doses of epo on (sub)maximal exercises. Also manipulated IV and OR iron supplementation at 4, 8, and 12 weeks, which are compared to placebo conditions. VO_{2max} expressed in l min⁻¹ and converted to ml kg⁻¹ min⁻¹. Only maximal performances are included. The study provided incomplete data for W_{map}, forcing us to exclude their six comparisons when analyzing this dependent variable.

8 Examined effects of epo on faster oxygen uptake at the onset of submaximal cycling. SE converted to SD. Data for placebo conditions estimated from pre test.

9 Examined influence of epo on lactate influx into erythrocytes. SE converted to SD

10 Examined effects of epo on pulmonary O₂ uptake during exercise. VO_{2max} expressed in l min⁻¹ and converted to ml kg⁻¹ min⁻¹

11 Examined effects of epo on physical self. Compared are epo vs. placebo treatment, no data on extra (no placebo) control group

12 Examined submaximal and maximal exercises in acute hypoxia (4100 m) in repeated measure design. Only maximal exercises are included in the analysis. Participants were administered NESP. No effect NESP on VO_{2max}. SE converted to SD

13 Effects of prolonged epo administration on submaximal exercises at 4 and 11 weeks. Placebo treatment is DB, epo treatment is SB. VO_{2max} expressed in l min⁻¹ and converted to ml kg⁻¹ min⁻¹

14 Examined laboratories testing urine samples with epo (= epo treatment). There obviously was no placebo treatment. Pre vs. post test comparisons were examined at days 0 (= pre test) vs. days 35, 42, and 49 as repeated measures. VO_{2max} expressed in l min⁻¹ and converted to ml kg⁻¹ min⁻¹

15 Examined the augmenting oxygen capacity of epo. Measurement of W during maximal exercise

16 Control (normoxia) condition compared to 1500 and 2500 m simulated hypoxia conditions in a counterbalanced repeated measure design. SE converted to SD. Study is part of Lundby et al. (2008). Only the data of the non-invasive experiment are analyzed. VO_{2max} expressed in l min⁻¹ and converted to ml kg⁻¹ min⁻¹

17 Examined influence of epo on cerebral metabolism and cognitive functioning. Three months low dose group is considered the epo treatment. Three days high dose group is considered the control treatment. Power output is measured in kJ. No differences on VO_{2max} were observed between high vs. low dose treatments

18 SB = single blind, DB = double blind, NR = no reference to SB/DB in paper

19 A = athletes, NA = non-athletes

20 N in parentheses is the same N used in other conditions of the same study. They are only used for statistical comparisons

21 np = no placebo/control treatment in study

Table S2. Formulae for effect size indices

StatsDirect (2012) estimates g (modified Glass statistic with pooled sample standard deviation) and the unbiased estimator d . M_e and M_c are the sample means of the experimental and control / placebo groups. SD_{pooled} is the pooled standard deviation across both groups. $J(m)$ is the correction factor given m and Γ is the gamma function.

$$g = \frac{M_e - M_c}{SD_{pooled}}$$

$$d = g\Gamma[J(N-2)]$$

$$J(m) = \frac{\Gamma(m/2)}{\Gamma[(m-1)/2] \sqrt{(m/2)}}$$

Conversion from d to r and r^2

For two dependent samples (i.e., pre vs. post test comparisons):

$$r = \sqrt{\frac{d^2}{d^2 + 1}}$$

For two independent samples (i.e., epo vs. placebo treatment comparisons):

$$r = \sqrt{\frac{d^2}{d^2 + \frac{1}{p_1 p_2}}}$$

In the latter formula p_1 and p_2 present the proportion of participants in the different experimental treatments relative to the total N of the experiment.

Table S3. Effect Size Indices and Associated Parameters of Experimental Treatments on the Pre Tests and of Pre vs. Post Test Comparisons within Control / Placebo Treatments

	Comparisons / variables ^{1, 2, 3}	Strata	Fixed effect: <i>d</i> (CI _{95%})	Heterogeneity: <i>Q</i>	Inconsistency: <i>I</i> ² (CI _{95%})	Random effect: <i>d</i> (CI _{95%})	Bias
A	- Control / placebo conditions, pre vs. post tests (all measurements)	32	0.08 (-0.11–0.25)	1.53	0% (0 - 38)	-	-0.33
	- Epo vs. control / placebo conditions, pre tests (all measurements)	-	-0.02 (-0.20–0.16)	31.78	2.5% (0 - 38)	-	-2.06
B	- Control / placebo conditions, pre vs. post tests (VO _{2max} and W _{map})	17	0.04 (-0.22–0.31)	0.65	0% (0 - 45)	-	-0.66
	- Epo vs. control / placebo conditions, pre tests (VO _{2max} and W _{map})	-	0.11 (-0.14–0.35)	6.89	0% (0 - 45)	-	-1.17
C	- Control / placebo conditions, pre vs. post tests (VO _{2max})	14	0.006 (-0.28–0.30)	0.16	0% (0 - 47)	-	-0.58
	- Epo vs. control / placebo conditions, pre tests (VO _{2max})	-	0.18 (-0.09–0.46)	4.15	0% (0 - 47)	-	-1.11
D	- Control / placebo conditions, pre vs. post tests (W _{map}) ⁴	3	0.20 (-0.41–0.81)	0.16	0% (0 - 73)	-	-
	- Epo vs. control / placebo conditions, pre tests (W _{map})	-	-0.25 (-0.83–0.33)	1.00	0% (0 - 73)	-	-

Notes:

1. All estimated indices and parameters are not significant.
2. **Panel A** refers to all measurements and includes performances measured in hypoxia conditions (< 3000 m) as well as submaximal and maximal performances.
3. **Panels B – D** present combined and separate findings of VO_{2max} and W_{map} that relate to maximal performances measured at sea level and are restricted to findings of double blind, placebo controlled studies.
4. Estimations have low power, due to a low number of number of strata (N = 3) and are therefore unreliable.