Running performance, $\dot{V}O_{2max}$, and running economy: The widespread issue of endogenous selection bias

Nicolai T. Borgen

Department of Sociology and Human Geography, University of Oslo

Abstract

Studies in sport and exercise medicine routinely use samples of highly trained individuals in order to understand what characterizes elite endurance performance, such as running economy and maximal oxygen uptake ($\dot{V}O_{2max}$). However, it is not well understood in the literature that using such samples most certainly leads to biased findings and accordingly potentially erroneous conclusions because of endogenous selection bias. In this paper, I review the current literature on running economy and $\dot{V}O_{2max}$ and discuss the literature in light of endogenous selection bias. I demonstrate that the results in a large part of the literature may be misleading, and provide some practical suggestions as to how future studies may alleviate endogenous selection bias.

Key points

1 Using samples restricted (truncated) to contain only elite athletes or highly trained individuals may result in biased results.

2 The association between running economy and $\dot{V}O_{2max}$ in truncated samples is at least partially spurious.

3 The effect size of running economy and $\dot{V}O_{2max}$ on race performance in truncated samples is attenuated.

Running head

Endogenous selection bias in the literature on running performance

Contact:
Address: Postboks 1096, Blindern, 0317 Oslo, Norway
E-mail: n.t.borgen@sosgeo.uio.no
Telephone: +47-22855248

1 Introduction

Many studies in sport and exercise medicine are conducted to determine what characterizes elite performances, as well as understand how elite athletes can improve further [24, 32, 28, 25, 42]. For instance, studies examine to what extent maximal oxygen uptake ($\dot{V}O_{2max}$) affects race performance [30] or if rearfoot striking is more economical than midfoot striking [37]. There is a widespread belief that in order to gain insights into elite performances, researchers cannot rely on studies of all runners. Instead, these studies typically select subjects based on race performance, either intentionally (e.g., studies of Olympic qualifiers) or as a result of convinience sampling. Having a homogeneous sample of highly trained individuals, or even elite athletes, is assumed to be an advantage in the literature.

However, it is not well-known in the literature that in observational studies (i.e., non-experimental studies), selecting subjects based on prior race performance will likely result in spurious correlations because of *endogenous selection bias* [40, 10]. The problem is twofold. First, when some individuals in the entire population (e.g., all US citizens) have a higher probability of being included in the population of interest (e.g., elite athletes), restricting the analysis to the population of interest amounts to conditioning on whatever increases the probability of being in the group of interest (e.g., prior race performance). Second, conditioning on the outcome variable or an effect of the treatment variable, such as race performance, may substantially bias the correlations and lead to erroneous conclusions. Perhaps because of its paradoxical and counter-intuitive nature [4], this second point is difficult to recognize and not sufficiently acknowledged in the literature.

In this review article, I discuss the widespread issue of endogenous selection bias in the literature on running performance, $\dot{V}O_{2max}$, and running economy (RE), and provide some suggestions on how to solve this issue. After providing a narrative review of the literature, I use simple models and hypothetical data to demonstrate how sample restriction induces bias in the findings in the literature. First, I show how endogenous selection bias induces an inverse relationship between $\dot{V}O_{2max}$ and RE, even though no such relationship exists in the population. Second, I show that having an elite sample most likely results in attenuated estimates of the effects of $\dot{V}O_{2max}$ and RE on race performance.

The primary aim of this article is to review and discuss the literature on running performance, $\dot{V}O_{2max}$, and RE. However, the article has relevance for other studies within sport and exercise medicine. It demonstrates that unless the independent variable of interest is randomized (which it seldom is [16, 7]), the choice of study subjects not only has implications for whom the results could be generalized to (external validity), but may also affect the internal validity. The take-away point of this article is that when defining a population of interest, one should always consider whether individuals with certain characteristics are more likely to be included in the population of interest, and if these characteristics in some way are related to the outcome or independent variables in the analyses.

2 Literature review

2.1 Determinants of running performance

The classic model of endurance running was initially established more than a century ago [32, 25], and suggests that the physiological 'concepts' $\dot{V}O_{2max}$, RE, lactate threshold, and fractional utilization of $\dot{V}O_{2max}$ affect race performance [32]. Still, we have yet to fully understand the determinants of elite race performance, and developments in our understanding of the role of the brain [36] and mental fatigue [27] have recently added to our understanding of race performance.

All of the physiological 'concepts' have been subject to extensive research, but I will mainly focus on RE and $\dot{V}O_{2max}$ in this narrative review. $\dot{V}O_{2max}$ is one of the main studied measures of athletic competence, while RE has become increasingly studied as a response to the complete dominance of Kenya and Ethiopia in distance running over recent decades [28]. However, endogenous selection bias is also relevant in the literature on other physiological factors.

2.2 Maximal oxygen uptake ($\dot{V}O_{2max}$)

$\dot{V}O_{2max}$ reflects an individual's maximal rate of energy expenditure [18], and has for a century been linked to running performance [25]. In studies that include relatively heterogeneous pools of runners, $\dot{V}O_{2max}$ has repeatedly been shown to be highly correlated with race performance [11, 26]. In one study of well-trained but not elite runners (i.e., $\dot{V}O_{2max}$ at about 60 ml kg^{-1} min^{-1} for males and 50 ml kg^{-1} min^{-1} for females), $\dot{V}O_{2max}$ explained 90.2% of the variance in a 10 mile run and was the single best predictor of running performance [32]. Legaz-Arrese et al. [19] provide a graphical overview of the literature, showing the relationship between $\dot{V}O_{2max}$ and International Association of Athletics Federations (IAAF) scores reported in the literature.

The correlation between $\dot{V}O_{2max}$ and race performance is also evident by the fact that elite athletes typically have very high $\dot{V}O_{2max}$, for men often between 70 and 85 ml kg^{-1} min^{-1} and for women about 10% lower [25, 12, 30, 51, 22, 18]. This is about 50 to 100% *higher* $\dot{V}O_{2max}$ than the normal active population [25, 55]. Among elite athletes, $\dot{V}O_{2max}$ is shown to be similar for runners competing in the 3000m to the marathon distance [19].

Although elites have a higher $\dot{V}O_{2max}$, some longitudinal studies of elite athletes suggest that $\dot{V}O_{2max}$ changes very little in well-trained or elite athletes [22, 18]. However, this may also be because elite athletes tend to do very little high-intensity training [34]. While low intensity training may lead to rapid increase in $\dot{V}O_{2max}$ for individuals who initially have low $\dot{V}O_{2max}$, much higher intensity may be needed for well-trained athletes [34]. Studies that include training at or near $\dot{V}O_{2max}$ for well-trained runners indicate that $\dot{V}O_{2max}$ may increase also in well-trained athletes [34].

Studies of the effect of $\dot{V}O_{2max}$ on race performance in well-trained or elite athletes have reported mixed findings. Some studies have found fairly strong

correlations between race performance and $\dot{V}O_{2max}$ (-.5 to -.87) in well-trained runners [15, 29, 9, 35]. However, several studies failed to find a correlation between race performance and $\dot{V}O_{2max}$ in homogeneous samples consisting of only elites [30, 28, 18]. Of particular interest is a longitudinal study of 32 athletes who were followed over three years, in which Legaz-Arrese et al. [18] demonstrated only small changes in $\dot{V}O_{2max}$ and no relationship between changes in $\dot{V}O_{2max}$ and race performance. One interpretation of this finding has been that a high $\dot{V}O_{2max}$ is necessary to gain membership in the elite performance cluster, but that within this elite cluster, $\dot{V}O_{2max}$ does not discriminate further [19].

However, it is possible that the entire relationship between $\dot{V}O_{2max}$ and race performance is spurious. Evaluating longitudinal data on sedentary individuals, Vollard et al. [52] found that $\dot{V}O_{2max}$ and race performance were related in a cross-sectional case, but that not even in this group of sedentary individuals did improvements in $\dot{V}O_{2max}$ lead to any improvements in race performance [52]. Other longitudinal studies have also failed to find a positive correlation between changes in $\dot{V}O_{2max}$ and changes in race performance. In Ramsbottom et al. [44] improvements in a 5 km trial were correlated with RE but not $\dot{V}O_{2max}$, while in Paavolainen et al. [39] 5 km performance actually declined with improvements in $\dot{V}O_{2max}$.

2.3 Running economy (RE)

Over the last few decades, East-African runners, particularly from Kenya and Ethiopia, have dominated middle- and long-distance running events. Several possible explanations have been proposed [56], but $\dot{V}O_{2max}$ is probably not the explanation, as Kenyan and Ethopian runners do not have superior $\dot{V}O_{2max}$ compared to for instance European runners [46]. However, elite East-African runners are typically small, even compared to other elite runners, and studies have shown that smaller runners and runners with thin lower legs have better RE [12]. For instance, a recent study of competitive Kenyan distance runners demonstrated that having low body mass index (BMI), lower mid-thigh and ankle circumference, as well as a short Achilles moment arm, all had a positive influence on RE [28].

RE is defined as the oxygen costs of endurance running at a given speed [25], meaning that efficient or economical runners have a low RE value. RE has been shown to vary about 30-40% among individuals, and about 20 to 30% among elite athletes [28, 30]. We know little about whether RE can be improved [12], but some studies suggest that an increase in high intensity interval training, plyometric training, altitude training, and heat exposure may improve RE [47].

RE has been researched extensively over the last few decades [47], and many studies have found a strong association between RE and race performance. Some studies indicate that RE is an even better predictor of race performance among elite runners than $\dot{V}O_{2max}$ [47]. However, a recent study questions whether RE is indeed the explanation for the East-African running dominance [28]. In a sample of 32 competitive Kenyan distance runners, Mooses et al [28] found that RE was not associated with running performance among elites. Similar findings

were reported by Grant et al. [15], who studied a sample of well-trained runners.

2.4 Association between $\dot{V}O_{2max}$ and RE

An interesting and seemingly non-intuitive finding in the literature is a moderate positive correlation between $\dot{V}O_{2max}$ and RE in samples of highly trained or elite athletes [30, 28, 50, 51], and a weak positive correlation in diverse samples of recreational runners [38]. In the literature, this positive correlation is typically described as an inverse association between $\dot{V}O_{2max}$ and RE, meaning that individuals with high $\dot{V}O_{2max}$ have on average poorer RE. Some researchers, for instance Joyner [23], have suggested that high $\dot{V}O_{2max}$ may be incompatible with excellent RE (or lactate threshold).

An inverse association between $\dot{V}O_{2max}$ and economy is also found among world-class cyclists. Lucia et al. [20] found that cyclists with higher $\dot{V}O_{2max}$ had lower cycling economy (CE) and gross mechanical efficiency (GE), which led to a fruitful discussion about how expressing both $\dot{V}O_{2max}$ and "economy" relative to body mass may lead to spurious findings [1, 21, 2].[1] However, after accounting for body mass, the inverse association among cyclists [21] and runners [51] still exists.

An inverse relationship between $\dot{V}O_{2max}$ and RE is counter-intuitive, because elite athletes have on average both *higher* $\dot{V}O_{2max}$ and *better* RE [31, 43]. At this point there is no clear understanding of why there is an inverse relationship, only speculations [30, 38, 51]. One suggested explanation of this non-intuitive finding is that runners with higher $\dot{V}O_{2max}$ rely more on fat utilization [38]. Another explanation is that greater lower limb mass will result in higher $\dot{V}O_{2max}$ but also poorer RE, which could explain the inverse relationship [30]. Finally, overstriding leads to excessive vertical oscillation and breaking forces (i.e., poor RE) but possibly also recruitment of larger muscle mass, which may result in higher $\dot{V}O_{2max}$ [51].

However, in cross-sectional observational studies, the relationship between RE and $\dot{V}O_{2max}$ is most definitely *biased* by endogenous selection bias. In fact, as the next section explains in detail, considering the strong relationship between RE and $\dot{V}O_{2max}$ on the one hand and being an elite athlete on the other hand, it is likely that the entire inverse association is spurious.

3 Endogenous selection bias

3.1 Inverse association between $\dot{V}O_{2max}$ and RE

Let us consider a hypothetical study (for simplicity, costs and availability are no issue here). Say that we defined the population of interest as male runners

[1] Note that although most studies express economy and efficiency directly as the oxygen cost, some studies define economy and efficiency so that a high value indicate higher efficiency [21]. In the study of Lucia et al. [20], the correlations between $\dot{V}O_{2max}$ and CE/GE are indeed negative.

with a marathon time below 02:15:00, and that we have drawn a random sample from this population of 50 male athletes where we have (perfectly) measured RE and $\dot{V}O_{2max}$. For these 50 hypothetical athletes, we find a correlation between $\dot{V}O_{2max}$ and RE of 0.25. The correlation is significant and the confidence interval is fairly narrow, and we can accordingly generalize the effect to the population.

From a merely descriptive point of view, this correlation is valid in the sense that the correlation would have been very similar had we used data on the entire population of sub-02:15:00 male marathoners. Our hypothetical study lets us conclude that athletes having high $\dot{V}O_{2max}$ on average are less efficient (i.e., higher oxygen cost) and those having exceptional RE on average have lower $\dot{V}O_{2max}$, which is similar to the conclusions that could be drawn based on the studies in the literature [30, 28, 50, 51, 38, 20].[2]

However, observed associations consist of both causal and various non-causal (i.e., spurious) components [10].[3] Although the observed inverse association can be generalized to the population, it is nevertheless at least partly spurious. Meaning that the elite marathoners with the lowest level of $\dot{V}O_{2max}$ typically have, on average, better RE than the elite marathoners with the highest level of $\dot{V}O_{2max}$, but that we should not expect to see a detoriation of an individual's RE if he/she increased his/her $\dot{V}O_{2max}$ (through for instance interval training).

To see why, we have to consider what factors influence the probability of being included in the population of interest (sub-02:15:00 marathon). Consider the empirically based but simplified example in Figure 1a where RE and $\dot{V}O_{2max}$ are determined independently, both RE and $\dot{V}O_{2max}$ affect race performance (RP), and race performance affects the probability of being an elite athlete. In this example, $\dot{V}O_{2max}$ and RE are marginally independent, meaning that knowing an individual's level of $\dot{V}O_{2max}$ in the full population (elites and non-elites) does not provide any information about the individual's level of RE (no correlation between $\dot{V}O_{2max}$ and RE). However, conditioning on elite status (or restricting the population of interest to elites) will induce a *spurious* inverse relationship between $\dot{V}O_{2max}$ and RE, assuming that the path coefficient π is negative and the path coefficient γ is positive [4].

Linear path modeling is a useful tool to see why two marginally independent variables (RE and $\dot{V}O_{2max}$ in Figure 1a) may become dependent if we condition on a common outcome of these variables (elite status in Figure 1a) [41]. If we

[2]Statistical generalization from a sample to a population (of interest) depends on assumptions such as random sampling. When using convenience samples, typical in the literature, the statistical significance of the correlations may be misleading [5]. With convenience samples, not all elites or highly trained individuals are equally likely to be included in the sample, and the study participants are likely to be more alike with regard to for instance training principles (e.g., amount of high-intensity interval training) than what the participants would have been had they been selected through a probability sample. I suspect that this will lead to P-values being too small and that the uncertainty of the results is underestimated. However, the literature routinely report P-values without any discussion. To explain their sampling procedure, and discuss any potential bias, researchers should consider using guidelines for reporting observational studies, for instance the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement [53].

[3]Following the counterfactual model of causality, causal effects are defined as contrasts between potential outcomes [33].

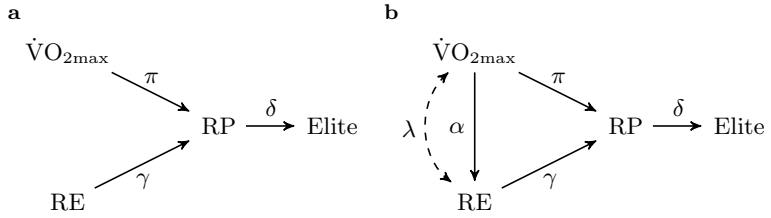

Figure 1: Hypothesized associations between maximal oxygen uptake ($\dot{V}O_{2max}$), running economy (RE), race performance (RP), and being an elite athlete (Elite) in the entire population. The single-headed arrows represent direct effects from causes to effects (e.g., $\dot{V}O_{2max}$ affects RP), while bidirectional arrows indicate that two variables have one or more causes in common. The Greek letters are path coefficients between pairs of variables, and are in this article interpreted as correlations (e.g., π is the correlation between $\dot{V}O_{2max}$ and RP). **a** A simplified example where both $\dot{V}O_{2max}$ and RE are assumed to affect RP, but are marginally independent in the population (no arrow between $\dot{V}O_{2max}$ and RE). Controlling for RP or its descendant Elite induces a spurious association between $\dot{V}O_{2max}$ and RE, because RP is a common outcome of $\dot{V}O_{2max}$ and RE (i.e., collider variable). **b** An example where (1) an increase in $\dot{V}O_{2max}$ is hypothesized to impair RE (α) and (2) the association between $\dot{V}O_{2max}$ and RE is confounded (λ). Controlling for RP or its descendant Elite still induces a spurious association between $\dot{V}O_{2max}$ and RE.

in a full population sample estimate the standardized regression coefficient ($\hat{\beta}$) of RE on $\dot{V}O_{2max}$ conditional on elite status, we identify

$$\hat{\beta}_{RE,VO_{2max}.elite} = \frac{-\gamma\pi\delta^2}{1-\pi^2\delta^2} \tag{1}$$

Since the true causal effect is 0 (i.e., no correlation), equation (1) reveals that conditioning on elite status induces bias. This is generally known as Berkson's paradox [4], endogenous selection bias [10], or collider bias [40]. Restricting the analysis to a subsample of elite athletes amounts to *conditioning* on elite status, and hence results in bias for the same reason [10].[4] This is often known as sample selection bias [41] or sample truncation bias [10].

It may seem counter-intuitive at first that restricting the sample to elite athletes induces a spurious inverse association between RE and $\dot{V}O_{2max}$, but it is actually straightforward. Consider knowing that an individual is an elite marathoner (sub-02:15:00 marathon) and that this individual has only a mediocre RE compared to other elite marathoners. What could then be inferred about his/her $\dot{V}O_{2max}$? It is likely exceptional, because he/she is unlikely to run a sub 02:15:00 marathon with a mediocre RE and a mediocre $\dot{V}O_{2max}$. It is not that a high $\dot{V}O_{2max}$ *leads* to poor RE (causal component), but that individuals that become elite athletes despite being inefficient must necessarily have other traits that compensate, such as a very high $\dot{V}O_{2max}$ (non-causal component).

[4]Linear path models allow us to calculate coefficients under the assumption of linearity and homogeneous effects (no interactions). Restricting the analysis to elite athletes is the same as adding a control for elite athletes and interaction between elite athletes and all independent variables.

Note that this endogenous selection bias occurs not simply because the sample is restricted to a subgroup of the population in itself, but because it is restricted in a specific way. For instance, restricting the sample to individuals with a $\dot{V}O_{2max}$ above 70 ml kg^{-1} min^{-1} will not lead to endogenous selection bias, as long as all individuals in the population of interest (individuals with $\dot{V}O_{2max}$ above 70 ml kg^{-1} min^{-1}) have an equal chance of being sampled regardless of whether they are elites or not. In that case, and assuming that $\dot{V}O_{2max}$ and RE are truly independently determined, there would be no reason to expect that those with a relatively low $\dot{V}O_{2max}$ have a higher RE than those with high $\dot{V}O_{2max}$. Thus, the problem occurs when conditioning on the collider variable elite status or, similarly, restricting the sample to elites only.

Endogenous selection bias is not about whether the effects are different for elites and non-elites. Endogenous selection bias is about how spurious associations are introduced in the data because of for instance sample restriction, resulting in erroneous correlations in the subgroup studied. Thus, the results are not valid even for the subgroup studied, but are rather *statistical artifacts*.

3.2 The size of the spurious inverse relationship

The simple formula in equation (1) demonstrates that the association between $\dot{V}O_{2max}$ and RE becomes inverse after conditioning on race performance when working with population data, and, intuitively, the same holds when restricting the sample to elite athletes only. Figure 2a illustrates the amount of bias we may expect in settings where (1) the amount of variation in race performance that $\dot{V}O_{2max}$ and RE explains is varied and (2) the sample selectivities differ. However, to illustrate the amount of bias, we need some assumptions. In Figure 2a, the correlation between RE and $\dot{V}O_{2max}$ in the full population is constrained to be zero. Additionally, the size of the correlations between $\dot{V}O_{2max}$ and race performance and RE and race performance is constrained to be of equality. The implication of this latter assumption is that Figure 2a illustrates an upper bound of the bias, as further discussed in Electronic Supplementary Material Appendix S2. Electronic Supplementary Material Appendix S3 provides a supplementary data simulation.

Figure 2a illustrates that the amount of bias is a function of two aspects. First, the bias is greatest in cases where $\dot{V}O_{2max}$ and RE explain most of the variation in race performance (i.e., a stronger correlation between $\dot{V}O_{2max}$ and race performance and RE and race performance). Second, a more elite sample will lead to more bias, which is intuitive. In a sample of the best 100 marathoners in the world, those with (relatively) low $\dot{V}O_{2max}$ must have exceptionally RE. However, in a sample that consists of all but the slowest 10% of the population, those with low $\dot{V}O_{2max}$ could very well have a poor RE. That said, the bias may be substantial even if we include all but the slowest 25% of runners (see the top 75% line in Figure 2a).

After accounting for body mass, the correlation between RE and $\dot{V}O_{2max}$ in samples of highly trained individuals ranges from about 0.25 to 0.30 [51]. In Figure 2a, we see that among the top 1 and 25% of the runners, the estimated

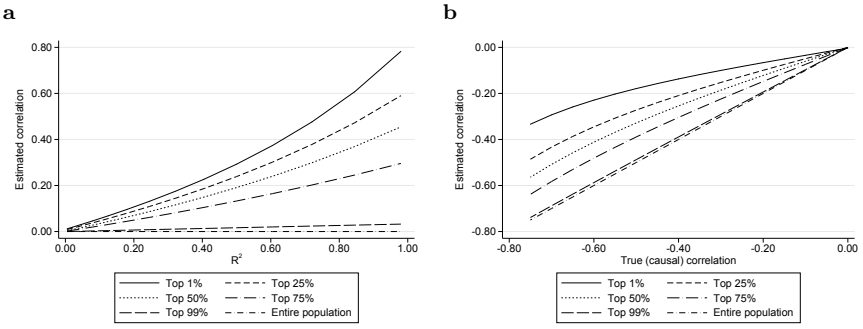

Figure 2: Illustration of the size of endogenous selection bias (see Electronic Supplementary Material Appendix S1 for the Stata code). **a** The figure is constructed by generating a set of hypothetical full populations where the correlation between RE and $\dot{V}O_{2max}$ is constrained to be zero, while the amount of variation in race performance (R^2) that RE and $\dot{V}O_{2max}$ explain varies between 0 and 0.98 (x-axis). The size (but not sign) of the correlations between RE and race performance and $\dot{V}O_{2max}$ and race performance are constrained to be of equality. From each of these populations, the estimated correlation between RE and $\dot{V}O_{2max}$ is estimated (Y-axis) for individuals in the subpopulations with the top 1, 25, 50, 75, and 99% race performances, as well as for the entire popoulation. **b** The figure is constructed by generating a set of hypothethical full populations where the correlation between $\dot{V}O_{2max}$ and race performance varies between 0 and 0.75 (x-axis). From each of these populations, the estimated correlation between $\dot{V}O_{2max}$ and race performance is estimated (Y-axis) for individuals in the subpopulations with the top 1, 25, 50, 75, and 99% race performances, as well as for the entire popoulation. Replacing $\dot{V}O_{2max}$ with RE produces the same results, only with different signs.

correlation is about 0.25-0.30 (upper bound) when $\dot{V}O_{2max}$ and RE explains about 50% of the variation in race performance in the population.[5] Thus, for the entire correlation between RE and $\dot{V}O_{2max}$ to be spurious, RE and $\dot{V}O_{2max}$ may only need to account for approximately 50% of the variation in race performance, which, based on the literature, is plausible [32, 11, 26, 47]. This demonstrates that restricting the sample to elite athletes has the potential to substantially bias the findings in the literature.

The discussion thus far has relied on the simplified model where RE and $\dot{V}O_{2max}$ are assumed to be marginally independent (Figure 1a). However, we could also expect RE and $\dot{V}O_{2max}$ to be marginally dependent (Figure 1b). First, elite runners have both higher $\dot{V}O_{2max}$ *and* higher RE than good runners [43]. This may indicate that some unobserved background factors (e.g., genetics) affects both $\dot{V}O_{2max}$ and RE, as suggested by the curved dotted line in Figure 1b. Second, we have also seen several explanations for why a high $\dot{V}O_{2max}$ may actually impair RE [30, 38, 51]. Thus, Figure 1b includes a direct effect of $\dot{V}O_{2max}$ on RE.

Given this more complex example, the path coefficients become more complicated (for a general example, see Pearl [41]) but the bias is no less present. The correlation between $\dot{V}O_{2max}$ and RE in an elite sample would be equal to the causal effect α and some bias caused by (1) unobserved confounding (λ) and (2) endogenous selection bias.

Section 3.1 explains why restricting the sample to contain only elite athletes may result in a spurious inverse association between $\dot{V}O_{2max}$ and RE, while section 3.2 has illustrated the amount of bias we may expect given some simplified assumptions. In the next section, I show that the effects of RE on race performance and $\dot{V}O_{2max}$ on race performance may be biased for the same reason.

3.3 The effects of RE and $\dot{V}O_{2max}$ on race performance

Many studies that have used elite samples have failed to find a significant effect of RE and $\dot{V}O_{2max}$ on race performance [30, 28, 18]. Consider the study of 32 competitive Kenyan runners, in which Mooses et al. [28] found neither a significant effect of RE on IAAF score ($r = -0.01$) nor of $\dot{V}O_{2max}$ on IAAF score ($r = 0.29$) [28].[6] A study of Olympic trials qualifiers also found a non significant correlation between $\dot{V}O_{2max}$ and race performance ($r = -0.21$) [30].

However, restricting the sample to elite athletes not only induces an inverse spurious association between RE and $\dot{V}O_{2max}$, it also biases the effects of RE and $\dot{V}O_{2max}$ on race performance. Conditioning on elite status, which is a descendant of the outcome variable race performance, induces a spurious

[5] The coefficient of determination (R^2) can be calculated by summing the squared semi-partial correlations, which in this case is identical to the pairwise correlations (Figure 1a): $R^2 = \pi^2 + \gamma^2$. Since π and γ is constrained to equality, π and γ could be calculated from the figure using $(R^2 * \frac{1}{2})^{\frac{1}{2}}$.

[6] This study also found an inverse relationship between RE and $\dot{V}O_{2max}$ ($r = 0.42$), which, as discussed in section 3.1, is at least partially spurious.

association between the predictor variable and all unmeasured causes of elite status.

To keep things simple, let us first consider what happens if we condition on elite status in a regression of race performance on $\dot{V}O_{2max}$ in a population sample where $\dot{V}O_{2max}$ is hypothesized to be exogenous. Based on Figure 1a, we identify [41]:

$$\hat{\beta}_{RP,RE.Elite} = \frac{\pi(1-\delta^2)}{1-\pi^2\delta^2} \tag{2}$$

, which in this particular example means that the estimate is biased towards zero , as $\delta > 0$. To see the estimated effect of RE, π simply needs to be replaced by γ. Restricting the sample to a subsample of elite athletes amounts to conditioning on race performance, and hence results in bias for the same reason.

Figure 2b illustrates the size of the endogenous selection bias in some settings, based on the model in Figure 1a (see Electronic Supplementary Material Appendix S3 for a supplementary data simulation). If there is no bias, we should expect the value on the x-axis (the true correlation between $\dot{V}O_{2max}$ and race performance) to perfectly match the value on the y-axis (the estimated correlation), as is the case when the correlation is estimated in the entire population. If the value of the y-axis is smaller than the value on the x-axis, as is the case in all of the subpopulations, then the estimated effect of $\dot{V}O_{2max}$ is underestimated.

There are two important take-away points from Figure 2b. First, we see that the amount of attenuation bias depends on sample selectivities. When comparing results from studies of elite athletes, highly-trained runners, recreational runners, and untrained runners, it may be tempting to, for example, conclude that $\dot{V}O_{2max}$ matters more for untrained than elites. In fact, cross-sectional studies often suggest that using a homogeneous sample of elite athletes in itself could explain the failure to find a significant relationship between RE and/or $\dot{V}O_{2max}$ and race performance [28, 20, 30]. However, Figure 2b illustrates that the amount of attenuation is a function of sample selectivities, and that even if the (causal) correlation is *identical* for all runners, one may find quite different results in different samples simply because of endogenous selection bias.

Second, the amount of attenuation bias depends on the true (causal) correlation, and the bias is zero when the true correlation is zero. Thus, unlike the case of the inverse relationship between RE and $\dot{V}O_{2max}$, adjusting for a descendant of the outcome variable would only generate bias *if* there is a marginal association (confounding or causal effect) between $\dot{V}O_{2max}$ and race performance. This could also be seen by replacing π with 0 in equation (2), in which the numerator reduces to 0, the denominator reduces to 1, and the estimated association is 0.

The amount of attenuation in studies by Mooses et al. [28] and Morgan and Daniels [30] is difficult to predict, as we do not know the true correlation between race performance and elite status (δ) nor do we know if RE and $\dot{V}O_{2max}$ are exogenous (i.e., no-confounding assumption). Sample sizes of 32 [28] and 22 [30] also mean that the point estimates are imprecise. However, assuming that RE and $\dot{V}O_{2max}$ are exogenous, correlations of -0.01, 0.29, and -0.21 in an elite sample (top 1%) would be expected if the true correlations were about -0.05,

0.7, and −0.6. Although this is only speculation, based on assumptions such as the no-confounding assumption and taking the point estimates at face value, it suggests that we could expect large bias when regressing race performance on predictors in a sample that consists of elite athletes.

Some studies have used longitudinal data to investigate how change in $\dot{V}O_{2max}$ relates to change in race performance [19, 52]. The next section discusses how the within-estimator removes endogenous selection bias, but also has some drawbacks.

4 Why longitudinal data may not be the solution

4.1 Within-subject variance as a solution

Let us consider the question of what level of RE individual i_1 would have, had he/she a $\dot{V}O_{2max}$ of 70 ml kg^{-1} min^{-1} rather than 65 ml kg^{-1} min^{-1}. Since we only could observe the actual $\dot{V}O_{2max}$ (65 ml kg^{-1} min^{-1}) and not the counterfactual one (70 ml kg^{-1} min^{-1}), we have a missing data problem. In the cross-sectional case, we solve this missing data problem by comparing individual i_1 with another individual i_2 that has a $\dot{V}O_{2max}$ equal to 70 ml kg^{-1} min^{-1}, under the assumption that these two individuals are equal in all other *relevant* aspects. However, when using a sample consisting of elites, the individual i_2 is not otherwise identical to i_1, as he/she likely has a lower RE. This is why the cross-sectional comparison breaks down, as explained in section 3.

However, if we have repeated observations on each individual (i.e., longitudinal data), then we can compare individual i_1 at time t_1 with the same individual at time t_2. Under the assumption that the bias is invariant over time and there is no selective attrition, the fixed effects model takes account of bias caused by both confounding and endogenous selection [54, 3, 6], and we can identify causal effects. Although not motivated by endogenous selection bias, some studies have used this methodology to investigate the association between $\dot{V}O_{2max}$ and race performance, finding no association [18, 52].

4.2 Small sample size and noisy measures

Despite the fact that the fixed effects estimator solves endogenous selection bias, use of within-subject variation has major drawbacks that renders it basically ineffective in the literature. For instance, studies of elite athletes have demonstrated that heavy training may not change $\dot{V}O_{2max}$ at all or change it only marginally [18, 22]. Thus, by discarding all between-subject variation, we utilize only a fraction of the variance in the data set. The implication is likely large standard errors and imprecise results.

Imprecise results are especially problematic given the small sample sizes in the literature [51, 18, 52], and accordingly low statistical power. Adding individual fixed effects will likely not increase statistical power, which means

that the correlations need to be very large for this design to be able to detect any significant effects [14, 3].[7] This casts some doubt on the studies that find no effect of RE and $\dot{V}O_{2max}$ using longitudinal data [18, 52].

Additionally, RE and $\dot{V}O_{2max}$ would likely differ depending on factors such as altitude [28], time of year [22], and running surface [49, 48], as well as being measured with some error [47, 17]. For instance, the typical measurement error of RE is shown to be about 2.4% [48]. Although this amount of measurement error is of little concern in a cross-sectional case, it will likely substantially attenuate the estimates when using a within-subject estimator. For example, because $\dot{V}O_{2max}$ changes very little in elite athletes, the within-subject variation we observe may to a large extent be caused by random measurement error (i.e., noise) [17], and because the measurement error is random, it is accordingly not related to race performance. In sum, null findings may not be that surprising in studies that evaluate longitudinal data [18, 52].

5 Conclusions

In this article, I have provided a critical review of the literature that investigates the associations between RE, $\dot{V}O_{2max}$, and elite running performance. Studies in this literature routinely use samples of highly trained individuals, which inevitably results in endogenous selection bias. The crux of the problem is that restricting the analysis sample to a population of interest amounts to conditioning on whatever increases the probability of being in the group of interest, such as prior race performance. If, for instance, (prior) race performance is either (1) the common outcome of two variables ($\dot{V}O_{2max}$ and RE) or (2) the outcome variable or a descendant of the outcome variable (race performance), then the sample restriction induces bias in the analysis.

The main conclusions of this review can be summarized as follows. First, I have demonstrated that the inverse relationship between RE and $\dot{V}O_{2max}$ that many studies find [30, 28, 51, 50] is likely spurious. Second, I have demonstrated that endogenous selection bias may substantially attenuate the effects of predictors on race performance, which may explain why some studies that use elite samples fail to find significant effects of $\dot{V}O_{2max}$ [30, 28, 18] and RE [28, 15]. Third, I have shown that a more elite sample will lead to more bias, but that the bias may be substantial even in samples of recreational runners. Fourth, I have argued that using within-subject variation is problematic. Given the small sample sizes in the literature, the fact that $\dot{V}O_{2max}$ changes only marginally in elite athletes [18, 22], and the problem of measurement error [48, 17], null findings [18] are not surprising.

Studies in the literature provide many interesting findings, for example, the relationship between anthropometric variables and RE in Mooses et al [28].[8]

[7] This means that studies that estimate within-subject correlations with small sample sizes and do find significant effects would most likely exaggerate the correlation.

[8] This association is also biased by endogenous selection, but the bias is likely small and the findings accordingly informative. Given the following model: anthropometric variable →

Like many other studies, Mooses et al. [28] also contribute by describing key characteristics of elite distance runners. However, caution should be exercised when estimating correlations between RE, $\dot{V}O_{2max}$, and race performance in observational samples of elite runners.

The challenge of estimating causal effects using a random sample of observational data from a population is in itself formidable [33], and having a (convenience) sample of elite or well-trained athletes magnifies the challenge. Although eliminating all sources of bias, such as measurement error, is difficult to achieve, the bias caused by endogenous selection in this particular literature is so fundamental that it cannot be ignored. Finding a good solution to the problem of endogenous selection bias is difficult, mainly because elite runners by definition are rare, but also because they are unlikely to participate in randomized trials (which would eliminate endogenous selection bias).

Nevertheless, perhaps the best solution is to conduct experiments to investigate the effects of changes in $\dot{V}O_{2max}$ and RE and how these relate to changes in race performance [44, 39]. Another solution is to use case-control studies, commonly used in the epidemiological literature, assuming that the population of interest is restricted by either the treatment or the outcome, and not both [45, 10]. A third solution is to gather a population-based sample of runners, possibly with oversampling of highly trained individuals, and use an unconditional quantile regression model to investigate the effects in different parts of the race distribution [13], possibly with subject fixed effects [8]. If none of these solutions are possible, perhaps the best alternative is to use qualitative single-case studies to suggest likely causes of success (e.g., Jones [22]).

This review points to a need for greater attention to causal inference in the literature, as well as more careful thought about the implications of the sampling procedures. However, the review also has several limitations. First, the assumptions underlying the models have been very simple, such as assuming a linear relationship between variables, no interactions, and no unobserved confounding. Relaxing these assumptions will not make the problem of endogenous selection bias less problematic, as shown in the literature on directed acyclic graphs [40]. However, it means that the suggested size of the bias is somewhat speculative.

Second, I have, for ease of presentation, treated race performance, $\dot{V}O_{2max}$, and RE as time-invariant in the main discussion (section 3). This choice is to some extent justified by the fact that these factors are relatively stable over time, such as $\dot{V}O_{2max}$ in elite athletes [18, 22]. Additionally, many studies do not measure race performance directly, especially studies of elite or near-elite athletes [28, 30, 29, 20], but rather rely on for example the best performance in the ongoing season [28]. Nevertheless, if race performance, RE and/or $\dot{V}O_{2max}$ change over time, then the bias induced by selection of research subjects based on prior race performance would not be identical to the bias induced by se-

RE $\rightarrow$ RP $\rightarrow$ elite, then the elite variable is a descendant of the outcome variable RE, and restricting the sample to elites amounts to conditioning on the outcome variable. However, since the effects of RE on elite are less than the effects of RP on elite, the bias would most likely be small.

lection of subjects based on current race performance. As briefly discussed in Electronic Supplementary Material Appendix S4, endogenous selection bias may both increase or decrease. More research and thinking about the causes and consequences of endogenous selection bias in the literature on elite running performance is therefore needed.

Acknowledgements: The author thank Solveig T. Borgen and two anonymous reviewers for useful comments and suggestions.

Compliance with Ethical Standards

Funding: No sources of funding were used to assist in the preparation of this article.

Conflict of interest: Nicolai T. Borgen declares that he has no conflict of interest relevant to the content of this review.

References

[1] Atkinson G, Davidson R, Passfield L, et al. Could the correlation between maximal oxygen uptake and "ECONOMY" be spurious? Med Sports Exerc. 2003;35(7):1242-3.

[2] Atkinson G, Davison R, Nevill AM. Response: inverse relationship between $\dot{V}O_{2max}$ and economy in world class cyclists. Med Sports Exerc. 2004;36(6):1085-6.

[3] Allison PD. Fixed effects regression models. Vol. 160. SAGE Publishing; 2009.

[4] Berkson J. Limitations of the application of fourfold table analysis to hospital data. Biometrics. 1946;2(3):47-53.

[5] Berk RA, Freedman DA. Statistical assumptions as empirical commitments. In: Blomberg TG, Cohen S, editors. Law, Punishment, and Social Control: Essays in Honor of Sheldon Messinger. 2nd ed. New York, NY: Aldine de Gruyter; 2003. pp.235-54.

[6] Bland JM, Altman DG. Statistics notes: calculating correlation coefficients with repeated observations: Part 1—correlation within subjects. BMJ. 1995;310(6977):446.

[7] Bleakley C, MacAuley D. The quality of research in sports journals. Br J Sports Med. 2002;36(2):124-5.

[8] Borgen NT. Fixed effects in unconditional quantile regression. Stata J. 2016;16(2):403-15.

[9] Brandon LJ. Physiological factors associated with middle distance running performance. Sports Med. 1995;19(4):268-277.

[10] Elwert F, Winship C. Endogenous selection bias: the problem of conditioning on a collider variable. Annu Rev Sociol. 2014;40:31-53.

[11] Fay L, Londeree BR, LaFontaine TP, et al. Physiological parameters related to distance running performance in female athletes. Med Sci Sports Exerc. 1989;21(3):319-24.

[12] Foster C, Lucia A. Running economy. Sports Med. 2007;37(4-5):316-9.

[13] Firpo S, Fortin NM, Lemieux T. Unconditional quantile regressions. Econometrica. 2009;77(3):953-73.

[14] Gelman A, Carlin J. Beyond power calculations: assessing type s (sign) and type m (magnitude) errors. Perspect Psychol Sci. 2014;9(6):641-51.

[15] Grant S, Craig I, Wilson J, et al. The relationship between 3 km running performance and selected physiological variables. J Sports Sci. 1997;15(4):403-10.

[16] Harris JD, Cvetanovich G, Erickson BJ, et al. Current status of evidence-based sports medicine. Arthroscopy 2014;30(3):362-71.

[17] Hopkins WG. Measures of reliability in sports medicine and science. Sports Med. 2000;30(1):1-15.

[18] Legaz-Arrese AL, Ostáriz ES, Mallen JC, et al. The changes in running performance and maximal oxygen uptake after long-term training in elite athletes. J Sports Med Phys Fitness. 2005;45(4):435.

[19] Legaz-Arrese A, Munguía-Izquierdo D, Nuviala AN, et al. Average $\dot{V}O_{2max}$ as a function of running performances on different distances. Sci Sports. 2007;22(1):43-9.

[20] Lucia A, Hoyos J, Perez M, et al. Inverse relationship between $\dot{V}O_{2max}$ and economy/efficiency in world-class cyclists. Med Sci Sports Exerc. 2002;34(12):2079-84.

[21] Lucia A, Hoyos J, Perez M, et al. Could the correlation between maximal oxygen uptake and "ECONOMY" be spurious?. Med Sci Sports Exerc. 2003;35(7):1244.

[22] Jones A. The physiology of the world record holder for the women's marathon. Int J Sports Sci Coach. 2006;1(2):101-16.

[23] Joyner MJ. Modeling: Optimal marathon performance on the basis of physiological factors. J Appl Physiol. 1991;70(2):683-87.

[24] Joyner MJ, Ruiz JR, Lucia A. The two-hour marathon: who and when? J Appl Physiol. 2011;110(1):275-7.

[25] Joyner MJ, Coyle EF. Endurance exercise performance: the physiology of champions. J Physiol. 2008;586(1):35-44.

[26] Maughan RJ, Leiper JB. Aerobic capacity and fractional utilisation of aerobic capacity in elite and non-elite male and female marathon runners. Eur J Appl Physiol Occup Physiol. 1983;52(1):80-7.

[27] Marcora SM, Staiano W, Manning V. Mental fatigue impairs physical performance in humans. J Appl Physiol. 2009;106(3):857-64.

[28] Mooses M, Mooses K, Haile DW, et al. Dissociation between running economy and running performance in elite Kenyan distance runners. J Sports Sci. 2015;33(2)136-44.

[29] Morgan DW, Baldini FD, Martin PE, et al. Ten kilometer performance and predicted velocity at $\dot{V}O_{2max}$ among well-trained male runners. Med Sci Sports Exerc. 1989;21(1):78-83.

[30] Morgan DW, Daniels JT. Relationship between $\dot{V}O_{2max}$ and the aerobic demand of running in elite distance runners. Int J Sports Med. 1994;15(07):426-9.

[31] Morgan DW, Pate R. Could the correlation between maximal oxygen uptake and economy be spurious?. Med Sci Sports Exerc. 2004;36(2):345.

[32] McLaughlin JE, Howley ET, Bassett Jr. DR, et al. Test of the classic model for predicting endurance running performance. Med Sci Sports Exerc. 2010;42(5):991-7.

[33] Morgan SL, Winship C. Counterfactuals and causal inference: methods and principles for social research. 2nd ed. Oxford: Oxford University Press; 2015.

[34] Midgley AW, McNaughton LR, Wilkinson M. Is there an optimal training intensity for enhancing the maximal oxygen uptake of distance runners? Sports Med. 2006;36(2):117-32.

[35] Noakes TD, Myburgh KH, Schall R. Peak treadmill running velocity during the $\dot{V}O_{2max}$ test predicts running performance. J Sports Sci. 1990;8(1):35-45.

[36] Noakes TD. The central governor model of exercise regulation applied to the marathon. Sports Med. 2007;37(4):374-77.

[37] Ogueta-Alday A, Rodríguez-Marroyo JO, García-López J. Rearfoot striking runners are more economical than midfoot strikers. Med Sci Sports Exerc. 2014;46.3:580-5.

[38] Pate RR, Macera CA, Bailey SP, et al. Physiological, anthropometric, and training correlates of running economy. Med Sci Sports Exerc. 1992;24(10):1128-33.

[39] Paavolainen L, Häkkinen K, Hämäläinen I, et al. Explosive-strength training improves 5-km running time by improving running economy and muscle power. J Appl Physiol. 1999;86(5):1527-33.

[40] Pearl J. Causality: models, reasoning, and inference. 2nd ed. Cambridge: Cambridge university press; 2009.

[41] Pearl J. Linear models: a useful "microscope" for causal analysis. J Causal Inference. 2013;1(1):155-70.

[42] Pitsiladis Y, Wang G, Wolfarth B, et al. Genomics of elite sporting performance: what little we know and necessary advances. Br J Sports Med. 2013;47(9):550-5.

[43] Pollock ML. Submaximal and maximal working capacity of elite distance runners. Part I: Cardiorespiratory aspects. Ann N Y Acad Sci. 1977;301(1):310-22.

[44] Ramsbottom R, Williams C, Fleming N, et al. Training induced physiological and metabolic changes associated with improvements in running performance. Br J Sports Med. 1989;23(3):171-6.

[45] Rothman KJ, Greenland S, Lash TL. Case-Control Studies. In: Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. Philadelphia, PA: Wolters Kluwer; 2008. pp. 111-27.

[46] Saltin B, Larsen H, Terrados N, et al. Aerobic exercise capacity at sea level and at altitude in Kenyan boys, junior and senior runners compared with Scandinavian runners. Scand J Med Sci Sports. 1995;5(4):209-21.

[47] Saunders PU, Pyne DB, Telford RD, et al. Factors affecting running economy in trained distance runners. Sports Med. 2004;34(7):465-85.

[48] Saunders PU, Pyne DB, Telford RD, et al. Reliability and variability of running economy in elite distance runners. Med Sci Sports Exerc. 2004;36(11):1972-6.

[49] Smoliga JM. What is running economy? A clinician's guide to key concepts, applications and myths. Br J Sports Med. 2017;51(10):831-2.

[50] Tartaruga MP, Brisswalter J, Peyré-Tartaruga LA, et al. The relationship between running economy and biomechanical variables in distance runners. Res Q Exerc Sport. 2012;83(3):367-75.

[51] Shaw AJ, Ingham SA, Atkinson G, et al. The correlation between running economy and maximal oxygen uptake: cross-sectional and longitudinal relationships in highly trained distance runners. PloS one. 2015; doi:e0123101.

[52] Vollaard NB, Constantin-Teodosiu D, Fredriksson K, et al. Systematic analysis of adaptations in aerobic capacity and submaximal energy metabolism provides a unique insight into determinants of human aerobic performance. J Appl Physiol. 2009;106(5):1479-86.

[53] Von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Int J Surg. 2014;12(12):1495-9.

[54] Winship C, Mare RD. Models for sample selection bias. Annu Rev Sociol. 1992;18:327-50.

[55] Wang CY, Haskell WL, Farrell SW, et al. Cardiorespiratory fitness levels among US adults 20–49 years of age: findings from the 1999–2004 National Health and Nutrition Examination Survey. Am J Epidemiol. 2010;171(4):426-35.

[56] Wilber RL, Pitsiladis YP. Kenyan and Ethiopian distance runners: what makes them so good. Int J Sports Physiol Perform. 2012;7(2):92-102.

Electronic Supplementary Material Appendix S1

Stata code for Figure 2a

```
local c=0
forvalues i1=.05(.05).71 {
local ++c

matrix b‘c’=J(1,7,-99)

clear
matrix  m = (1, ‘i1’, -‘i1’ \ ‘i1’, 1,0\ -‘i1’,0,1)
qui corr2data race re vo2max, corr(m) mean(0 0 0)  ///
sds(1 1 1) n(300000) seed(10000) clear

regress race re vo2max
matrix b‘c’[1,1]=e(r2)

pctile prosentil=race, nq(100)

local r=1
foreach i2 in 1 25 50 75 99 {
local ++r

cap drop elite
gen elite=race>prosentil[‘i2’]

qui cor re vo2max if elite==1
matrix b‘c’[1,‘r’]=r(rho)
}

qui cor re vo2max
matrix b‘c’[1,7]=r(rho)

matrix b=nullmat(b)\b‘c’

}

mat l b

svmat b

twoway  (line b6 b1) (line b5 b1) (line b4 b1)  ///
(line b3 b1) (line b2 b1) (line b7 b1)  ///
, ytitle(Estimated correlation) xtitle(R{sup:2})  ///
legend(order(1 "Top 1%" 2 "Top 25%" 3 "Top 50%" 4 "Top 75%" ///
5 "Top 99%" 6 "Entire population")) name(b, replace)  ///
graphr(color(white)) ylabel(,angle(horizontal)) ///
xlabel(,format(%3.2f)) ylabel(,format(%3.2f))
```

Stata code for Figure 2b

```
local c=0
forvalues i1=0(.05).8 {
local ++c

matrix b‘c’=J(1,7,-99)
matrix b‘c’[1,1]=-‘i1’

clear
matrix  m = (1, -‘i1’ \ -‘i1’, 1)
qui corr2data race vo2max , corr(m) mean(0 0)  ///
sds(1 1) n(3000000) seed(10000) clear

pctile prosentil=race, nq(100)

local r=1
foreach i2 in 1 25 50 75 99 {
local ++r

cap drop elite
gen elite=race>prosentil[‘i2’]

qui cor race vo2max if elite==1
matrix b‘c’[1,‘r’]=r(rho)
}

qui cor race vo2max
matrix b‘c’[1,7]=r(rho)

matrix b=nullmat(b)\b‘c’

}

mat l b

svmat b

twoway (line b6 b1) (line b5 b1) (line b4 b1)  ///
(line b3 b1) (line b2 b1) (line b7 b1)  ///
, ytitle(Estimated correlation) xtitle(True (causal) correlation)  ///
legend(order(1 "Top 1%" 2 "Top 25%" 3 "Top 50%" 4 "Top 75%" ///
5 "Top 99%" 6 "Entire population")) name(r3, replace)  ///
graphr(color(white)) ylabel(,angle(horizontal)) ///
xlabel(,format(%3.2f)) ylabel(,format(%3.2f))
```

Electronic Supplementary Material Appendix S2

The relative importance of RE and $\dot{V}O_{2max}$ for race performance

In Figure 2a in the main text, the correlation between RE and race performance (γ) is assumed to equal the correlation between $\dot{V}O_{2max}$ and race performance (π), and the x-axis shows the coefficient of determination (R^2). However, different combinations of RE and $\dot{V}O_{2max}$ could generate the same R^2. In cases where RE and $\dot{V}O_{2max}$ explain the same amount of variation in race performance in the full population, the estimated correlation between RE and $\dot{V}O_{2max}$ in a sample restricted to elites will differ depending on the relative importance of RE and $\dot{V}O_{2max}$.

In this appendix, instead of constraining γ and π to be of equal size, I keep the amount of variation in race performance that RE and $\dot{V}O_{2max}$ explain constant, and vary the relative size of γ and π. Figure S2.1 shows the estimated correlation between RE and $\dot{V}O_{2max}$ in hypothetical samples consisting of the top 1% fastest runners (y-axis), and where RE and $\dot{V}O_{2max}$ explains 50, 70, and 90% of the variation in race performance in the full population. The x-axis shows the importance of RE (γ), and we can calculate the importance of $\dot{V}O_{2max}$ (π) based on the following formula: $\pi = \sqrt{R^2 - \gamma^2}$.[1] For instance, when $\gamma = 0.20$ and $R^2 = 0.50$, π is equal to 0.68.

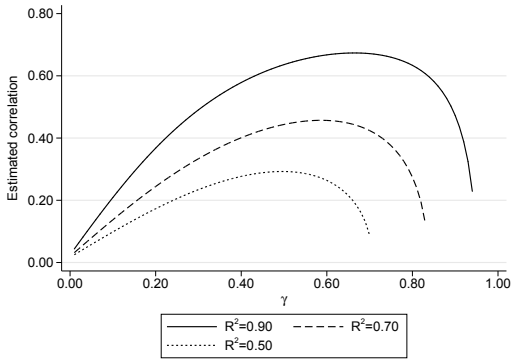

Figure S2.1: The consequences of the relative importance of RE and $\dot{V}O_{2max}$ for the estimated correlation between RE and $\dot{V}O_{2max}$ in a sample consisting of elite runners. The figure is constructed by generating a set of hypothetical populations where the correlation between RE and $\dot{V}O_{2max}$ is constrained to be zero, while the correlation between RE and race performance and $\dot{V}O_{2max}$ and race performance is varied. The x-axis show the correlation between RE and race performance in the full population, while the y-axis shows the estimated correlation between RE and $\dot{V}O_{2max}$ in a sample consisting of the top 1% runners.

[1] R^2 can be calculated by summing the squared semipartial correlations, which in this case is identical to the pairwise correlations (Figure 1a): $R^2 = \pi^2 + \gamma^2$. Thus, if we know γ and R^2, we can calculate π by using the formula: $\pi = \sqrt{R^2 - \gamma^2}$.

Figure S2.1 reveals that the estimated correlation between RE and $\dot{V}O_{2max}$ in samples restricted to elite athletes is highest when γ and π are of equal size (0.50 when $R^2 = 0.50$, 0.59 when $R^2 = 0.70$, and 0.67 when $R^2 = 0.90$), which shows that Figure 2a in the main manuscript provides an upper bound of the bias. When either γ or π is weak, then the estimated bias between RE and $\dot{V}O_{2max}$ is also small. Nevertheless, in most cases, the bias is substantial.

Stata code for Figure S2.1

```
version 14.2

local c=0
forvalues i1=.01(.01).98 {
local ++c

matrix b‘c’=J(1,4,.)

local r=1
foreach r2 in 50 70 90 {
local ++r

local i2=sqrt(0.‘r2’-‘i1’^2)

capture {
matrix  m = (1, ‘i1’, -‘i2’ \ ‘i1’, 1,0\ -‘i2’,0,1)
qui corr2data race re vo2max, corr(m) mean(0 0 0)  ///
sds(1 1 1) n(300000) seed(10000) clear

pctile prosentil=race, nq(100)
cap drop elite
gen elite=race>prosentil[99]

matrix b‘c’[1,1]=‘i1’

qui cor re vo2max if elite==1
matrix b‘c’[1,‘r’]=r(rho)
}
}
matrix b=nullmat(b)\b‘c’
}

mat l b
svmat b
twoway  (line b4 b1)  (line b3 b1) (line b2 b1)   ///
, ytitle(Estimated correlation) xtitle({&gamma}) ///
legend(order(1 "R{sup:2}=0.90" 2 "R{sup:2}=0.70" 3 "R{sup:2}=0.50")) ///
graphr(color(white)) ylabel(,angle(horizontal)) ///
xlabel(,format(%3.2f)) ylabel(,format(%3.2f))
```

Electronic Supplementary Material Appendix S3

Data simulations

The Monte Carlo data simulation in this appendix supplements Figure 2 in the main text by explicitly drawing repeated samples from the hypothetical population of interest and estimating the correlation in each draw.

Full population and population of interest

I begin the simulation by defining the full population, of which the population of interest is selected from. In my hypothetical data, the full population consists of 5,000,000 males, with average marathon race time of 3 hours and 40 minutes (220 minutes), average $\dot{V}O_{2max}$ of 43 ml kg^{-1} min^{-1}, and average RE of 260 ml O_2/kg/km (Table S3.1). The correlation between RE and $\dot{V}O_{2max}$ is constrained to be zero, while the correlation between RE and marathon time is 0.4 and the correlation between $\dot{V}O_{2max}$ and marathon time is -0.7 (Table S3.2). All associations in the hypothetical data are constrained to be linear and there are no interaction effects. That is, the correlation between $\dot{V}O_{2max}$ and marathon time is equal for all, regardless of whether the individual is an elite or not.

Table S3.1: Characteristics of the full population.

	N	Mean	SD	Min	Max
Marathon time	5000000	220	23	105.922	337.162
RE	5000000	260	19	162.262	358.908
$\dot{V}O_{2max}$	5000000	43	8	2.924	84.646

Table S3.2: Correlation matrix in the full population.

	Marathon time	RE	$\dot{V}O_{2max}$
Marathon time	1		
RE	0.400	1	
$\dot{V}O_{2max}$	-0.700	0.000	1

I define three subpopulations of interest: (1) Individuals who have run a sub-03:00:00 marathon (<180 minutes), (2) individuals who have run a sub-02:30:00 marathon (<150 minuts), and (3) individuals who have run a sub 02:15:00-marathon (<135 minutes). From each of these subpopulations (of interest), I draw 10,000 random samples of size 20 with replacement. In each sample, I estimate the correlation between RE and $\dot{V}O_{2max}$, the correlation between RE and marathon time, and the correlation between $\dot{V}O_{2max}$ and marathon time.

Correlation between RE and $\dot{V}O_{2max}$ (supplement to Figure 2a)

Table S3.3 and Figure S3.1 show the correlation between RE and $\dot{V}O_{2max}$ in the three subpopulations of interest. The simulations demonstrate, as Figure 2a in

the manuscript, that when restricting the sample to elite athletes, the results are biased by endogenous selection bias. We see that despite no correlation between RE and $\dot{V}O_{2max}$ in the population, the correlations in samples consisting of sub 03:00:00-, 02:30:00-, and 02:15:00-marathoners are on average 0.35, 0.38, and 0.40 respectively. The reason for this bias is that race performance is a common outcome of RE and $\dot{V}O_{2max}$, and restricting the sample to elite athletes or well-trained recreational runners amounts to conditioning on race performance.

Table S3.3: The average correlation between RE and $\dot{V}O_{2max}$ (Mean), standard deviation of the correlations (SD), the lowest correlation (Min), and the highest correlation (Max) from 10,000 random draws of size 20 in three subpopulations of interest.

Marathon time	N	Mean	SD	Min	Max
<135	10000	0.398	0.196	-0.455	0.890
<150	10000	0.384	0.204	-0.452	0.905
<180	10000	0.346	0.204	-0.618	0.847

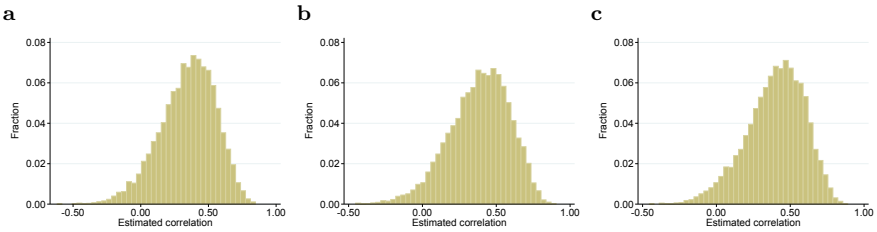

Figure S3.1: The correlation between RE and $\dot{V}O_{2max}$ in 10,000 random samples of size 20 in three subpopulations of interest. **a** The population of interest is all males who have run a sub-03:00:00 marathon. **b** The population of interest is all males who have run a sub-02:30:00 marathon. **c** The population of interest is all males who have run a sub-02:15:00 marathon.

Effects of RE and $\dot{V}O_{2max}$ on race performance (supplement to Figure 2b)

Table S3.4 and Figure S3.2 show that the correlation between RE and marathon time is attenuated when restricting the population of interest to elite athletes or well-trained recreational runners. While the true causal correlation in the full population as well as in the subpopulations is set to be 0.4, the correlations in samples consisting of sub 03:00:00-, 02:30:00-, and 02:15:00-marathoners are on average 0.15, 0.12, and 0.08 respectively.

We see the same pattern regarding the correlation between $\dot{V}O_{2max}$ and marathon time, in table S3.5 and Figure S3.3. While the true causal correlation is -0.7, the correlations in samples consisting of sub 03:00:00-, 02:30:00-, and 02:15:00-marathoners are on average -0.32, -0.23 , and -0.20 respectively.

Table S3.4: The average correlation between RE and marathon time (Mean), standard deviation of the correlations (SD), the lowest correlation (Min), and the highest correlation (Max) from 10,000 random draws of size 20 in three subpopulations of interest.

Marathon time	N	Mean	SD	Min	Max
<135	10000	0.084	0.240	-0.710	0.805
<150	10000	0.119	0.229	-0.716	0.781
<180	10000	0.149	0.226	-0.643	0.812

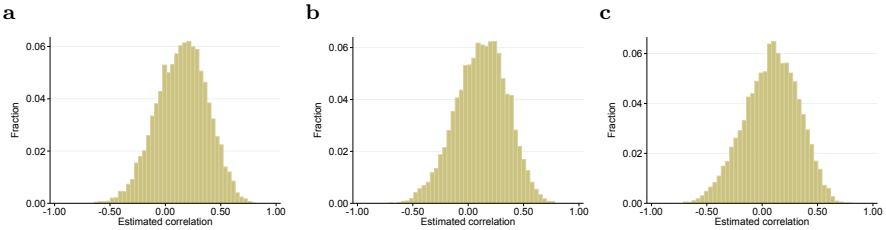

Figure S3.2: The correlation between RE and marathon time in 10,000 random samples of size 20 in three subpopulations of interest. **a** The population of interest is all males who have run a sub-03:00:00 marathon. **b** The population of interest is all males who have run a sub-02:30:00 marathon. **c** The population of interest is all males who have run a sub-02:15:00 marathon.

Table S3.5: The average correlation between $\dot{V}O_{2max}$ and marathon time (Mean), standard deviation of the correlations (SD), the lowest correlation (Min), and the highest correlation (Max) from 10,000 random draws of size 20 in three subpopulations of interest.

Marathon time	N	Mean	SD	Min	Max
<135	10000	-0.197	0.219	-0.823	0.752
<150	10000	-0.227	0.219	-0.869	0.605
<180	10000	-0.320	0.214	-0.885	0.652

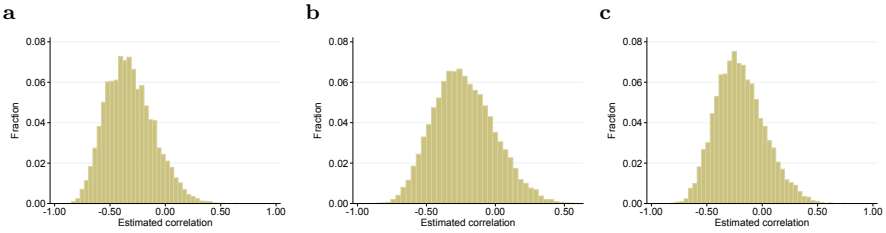

Figure S3.3: The correlation between $\dot{V}O_{2max}$ and marathon time in 10,000 random samples of size 20 in three subpopulations of interest. **a** The population of interest is all males who have run a sub-03:00:00 marathon. **b** The population of interest is all males who have run a sub-02:30:00 marathon. **c** The population of interest is all males who have run a sub-02:15:00 marathon.

Stata code for the data simulations in Appendix S3

```
version 14.2

*[1]* Generate the full population
matrix  m = (1, .4, -.7 \ .4, 1,0\ -.7,0,1)
corr2data race re vo2max, corr(m) mean(220 260 43)  ///
sds(23 19 8) n(5000000) clear seed(5649)
save fullpop, replace

*[2]* Identify elite athletes (based on race times), and save subpopulations
foreach e in 180 150 135 {
preserve
gen race‘e’=race<‘e’
keep if race‘e’==1
save race‘e’, replace
restore
}

postfile buffer racetime cor1 cor2 cor3 using results, replace

*[3]* Repeat simulation for following race times (in minutes)
foreach e in 180 150 135 {

*[4]* Run the simulation (results saved in results.dta)
set seed 12345
forvalues i=1/10000 {
qui {
use race‘e’, clear
sample 20, count
cor re vo2max
local c1=r(rho)
cor re race
local c2=r(rho)
cor vo2max race
local c3=r(rho)
post buffer (‘e’) (‘c1’) (‘c2’) (‘c3’)
}
}
}

postclose buffer
```

Electronic Supplementary Material Appendix S4

Allowing $\dot{V}O_{2max}$, RE, and RP to change across time

Allowing maximal oxygen uptake ($\dot{V}O_{2max}$), running economy (RE), and race performance (RP) to change over time affects the amount of endogenous selection bias. To get an idea of the variation in bias, we need some assumptions. In Figure S4.1, changes in $\dot{V}O_{2max}$, RE, and RP from time 1 ($t1$) to time 2 ($t2$) is assumed to be random, in which case we can think in terms of random measurement error in the variables.

With regard to the inverse spurious association between $\dot{V}O_{2max}$ and RE, it would likely be smaller if RE and $\dot{V}O_{2max}$ are measured at $t2$ compared to RE and $\dot{V}O_{2max}$ measured at $t1$ (Figure S4.1a). With regard to the effects of RE (or $\dot{V}O_{2max}$) on RP, the endogenous selection bias may be larger or smaller (Figure S4.1b). The association between RE at $t2$ and RP at $t2$ would likely be less attenuated, while the association between RE at $t2$ and RP at $t1$ would likely be more attenuated.

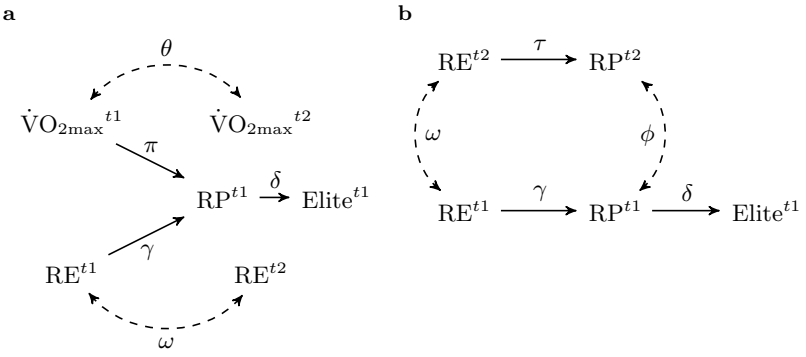

Figure S4.1: Hypothesized associations between maximal oxygen uptake ($\dot{V}O_{2max}$), running economy (RE), race performance (RP), and being an elite athlete (Elite) in the entire population measured at time 1 ($t1$) and time 2 ($t2$).