[redacted]

Department of Economics

University of [redacted]

Measuring Performance in Cycling

Intro

Cycling, as in many other sports, has quantitative metrics that can be used to assess, predict, and compare performance across time and between riders. The best metric to use is VO2max which I will derive and show that it is *consistent across time* and *comparable between riders*. A problem in any activity that people engage in – whether that be academics, sports, or the working world – is the measurement of performance. The use of metrics to measure performance is a matter of debate in any culture that attempts to be a meritocracy. An example in our culture is the SAT. Reformers want to abolish the SAT, since they claim it has a negative effect on socioeconomic diversity in colleges. Other than political criticisms, I argue that the SAT does not measure what it proports to measure. When a measure becomes a target, it ceases to be a good measure, as Goodhart's Law says. The SAT only measures how well a student can take the SAT, whether it measures anything else is uncertain. What we want to achieve in any field, to measure performance, is a metric or combination of metrics that are fair and incentivize improvement. Cycling has many metrics that quantify performance: speed, perceived effort, heart rate, and power. Most of these metrics are either subjective or are prone to bias by confounding factors. Many riders latch onto one metric to compare themselves to other riders and to compare themselves over time. In my opinion, it is foolish to compare performances with most of the previously mentioned metrics. Like the SAT, cycling metrics may not indicate what riders think they do. Comparing SAT scores is like comparing biased cycling metrics, the signal is drowned out by too much noise and this comparison becomes a battle of egos.

Ineffective Methods

There are many ineffective metrics in cycling that all have the same problem: they are biased by confounding factors. There are many more of these factors in cycling that make metrics like speed inconsistent where they are perfectly acceptable in other sports like running. It is important to mention these methods because they are commonly used by riders to compare each other, but without a proper interpretation they are useless. Cyclists will argue with each other over who is better by comparing ineffective metrics like speed or their average heart rate. After riding, some cyclists will self-report that they are not tired to show off to their riding buddies. Disagreement is only possible because the metrics they use are not objective; with an objective metric, there would be no disagreement. Another important note, cyclists will use these metrics to compare themselves over time, but biases make it difficult to determine whether the cyclist has truly improved in fitness. To properly train over time, metrics need to be like a *growth chart,* consistent over time, meaning the chart does not say you are 5’8 one year and 4’11 another year.

Speed

Speed, measured in miles per hour, is a metric that many riders use to compare each other with, since in a road race or time trial that is the metric riders are compared by. Speed is obviously the correct metric for a race when conditions are held constant. But there are two confounding factors that make speed biased for comparing across time and between riders. Equipment and wind conditions affect aerodynamics which make speed vary depending on these variables. Drive train efficiency, bike weight, bike aerodynamics, and clothing all affect speed –

which poses a problem of equity because superior gear can be bought. Wind conditions vary across time, so it is deceiving to compare past performances or compare performance to other riders when a head or tail wind has a massive effect on average speed.

Perceived Effort

Perceived effort is a categorical metric where the athlete rates their workout according to how they feel afterwards. They usually rate their effort on a scale of one to five: very easy, easy, normal, hard, and very hard. Athletes use this metric to report their fatigue in a training log to optimize the appropriate intensity for their training. There are two problems with this. First, the rating is subjective and prone to bias. There is no concrete reference or scale for how the effort relates in intensity to past efforts and the fatigue an athlete feels can change over the course of hours after the workout. Second, the categories are not consistent or well defined. Athletes can rate themselves within two categories or more, making the categories fuzzy and unreliable. Perceived effort is objectively the worst of these metrics since it is not quantitative and suffers from human error. These biases make it difficult to train precisely and monitor fatigue.

Heart Rate

Heart rate, measured in beats per minute, is the best out of the three ineffective methods and is frequently used in running and cycling to gauge cardiovascular exertion during the training session. Heart rate can be compared across time periods to measure cardiovascular fitness, but it cannot be compared between athletes. There are many problems with heart rate data that clouds its ability to measure performance accurately and consistently. First, heart rate is influenced by biological factors and substance consumption. Sleep deprivation and stress levels elevate heart rate beyond what is typical, making it inconsistent. Coffee and alcohol consumption can also make heart rate vary. Second, heart rate lags effort (see figure 1) and stays elevated after an effort is over, making it inefficient. This makes it difficult to analyze heart rate data after the fact to precisely determine at what second the most strenuous effort took place. Finally, maximum heart rate differs between athletes based on age, training level, and genetics, therefore, it is difficult to compare efforts between athletes. These problems are fixed by comparing heart rate with other metrics, but it does not meet the criteria of being a good metric to measure performance on its own.

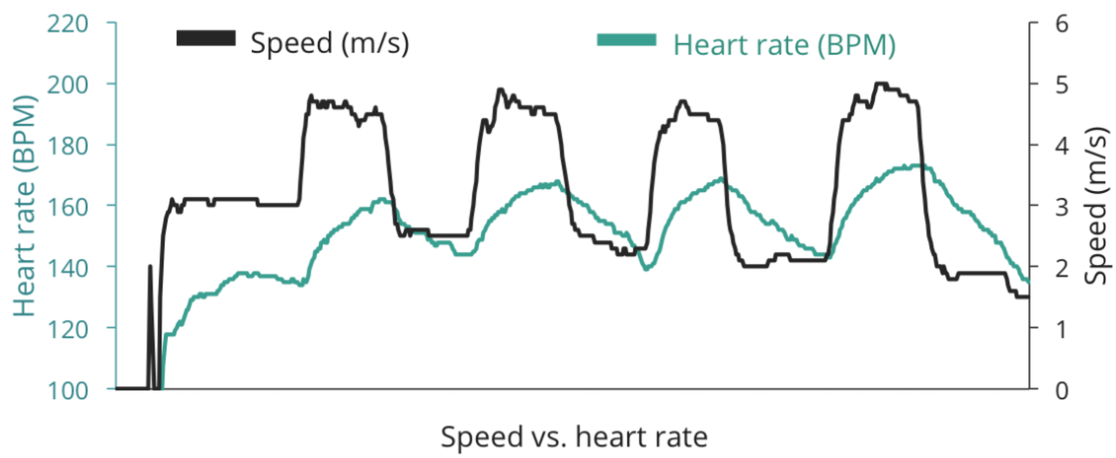

Figure 1: Heart Rate vs. Speed histogram: https://blog.stryd.com/2014/12/28/running-power-and-heart-rate/

Effective Methods

Power

Power, measured in watts, is the best metric to measure cycling performance; it meets all three criteria for being a good indicator of performance. It is consistent across time, assuming an accurate power meter. It is unbiased by gear, weather, or biological factors. It is efficient – correlating perfectly with effort at every time interval. The tradeoff is that power is expensive.

Power is the amount of force produced by the rider onto the pedals. If a person were a machine, power would be the output of the machine, making it a very good metric to represent effort. *Think of power as the output of an effort function*.

How power is measured is critical for it to meet these three criteria. Different power meters yield different results and suffer from confounding variables. There are two power meters that accurately measure power: crank based power meters, and pedal based power meters. There is a third type, the hub-based power meter, but it can be inaccurate due to drive train efficiency. People do use hub-based power meters on an indoor trainer since they may not have an additional power meter. Crank and pedal based power meters measure power at the point of power transfer, the cranks, which eliminates all the downsides of the previously mentioned metrics. When not using an indoor trainer, pedal and crank based power meters are ubiquitous. The tradeoff between the two is that pedal based power meters are heavier than crank based, but pedals can be easily switched between bikes whereas cranks cannot.

Power is not without its downsides. Financially, it is very expensive; a good power meter costs around $1000. There are also three problems with power as a metric for cycling. First, weight is not considered. Lighter riders will produce less power on average, and vice versa. To overcome this problem, power can be standardized by weight to compare riders of different weight by dividing watts by the weight of the rider in kilograms. The strength to weight ratio is measured in watts/kilogram and this metric is used to model speed on virtual cycling apps like Zwift. Second, power does not account for stress on the body. Adding heart rate, in conjunction with power, gives one the best assessment for performance and fatigue. Heart rate per level of wattage is a good way to compare performances across time. *The ratio of heart rate to power is used to infer VO2 Max.* VO2max is normally distributed in the population (See figure 2 and 3) and can be used to objectively compare performance across riders and time. VO2 max is a measure of aerobic capacity measured in milliliters of oxygen consumed per kilogram of body mass per minute (mL/kg/min). Notice, VO2 max is standardized for weight and is an absolute measure of oxygen, which makes it possible to compare riders. Since it measures aerobic capacity, a biological process, it is possible to compare one’s aerobic capacity over time as a proxy for performance.

Age → Rating ↓	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64
Excellent	>62	>59	>56	>54	>51	>48	>46	>43	>40
Very Good	57-62	54-59	52-56	49-54	47-51	44-48	42-46	40-43	37-40
Good	51-56	49-53	46-51	44-48	42-46	40-43	37-41	35-39	33-26
Average	44-50	43-48	41-45	39-43	36-41	35-39	33-36	31-34	29-32
Fair	38-43	36-42	35-40	33-38	32-35	30-34	28-32	27-30	24-28
Poor	32-37	31-35	29-34	28-32	26-31	25-29	24-27	22-26	21-24
Very Poor	<32	<31	<29	<28	<26	<25	<24	<22	<21

Figure 2: VO2max population distribution: https://runninforsweets.com/vo2-max-chart/

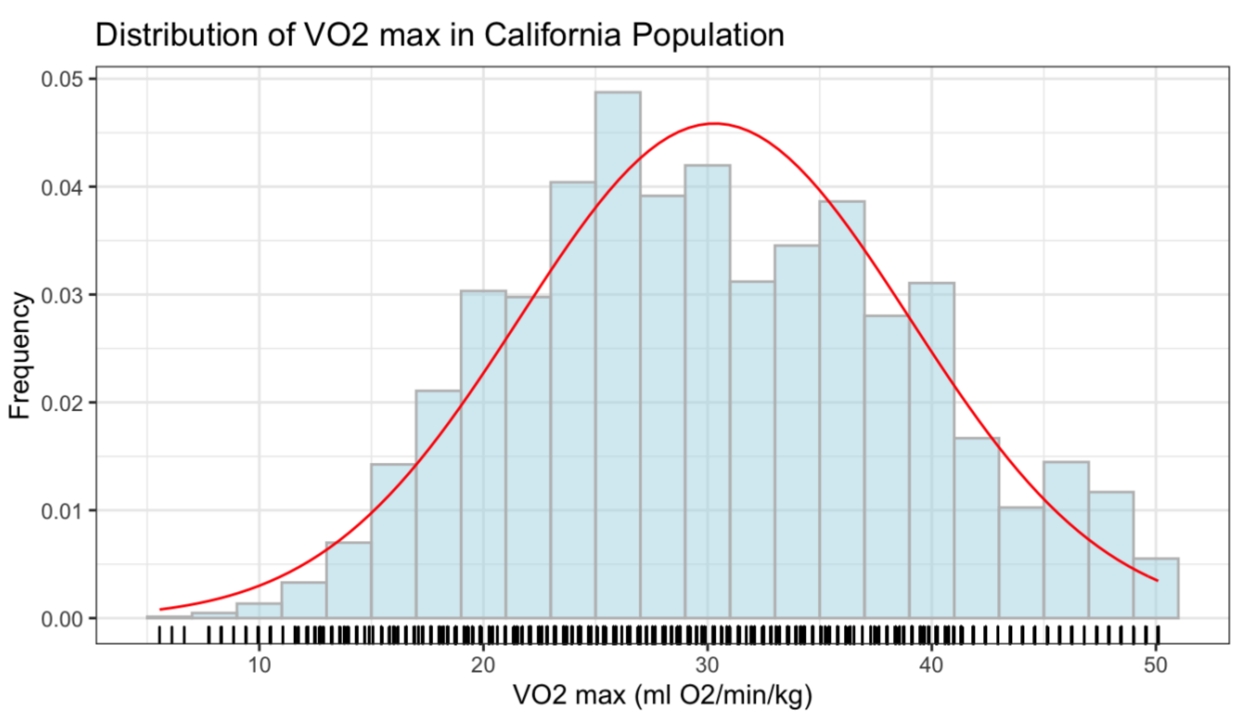

Figure 3:https://rpubs.com/nandiniguntur/632817

Conclusion

The best metric to use is VO2max, since it is normally distributed in the population and represents a biological capacity, making it consistent across time. This metric does accurately assess performance and incentivizes cyclists to improve their overall fitness. The quantitative nature of cycling makes it a fascinating sport to analyze and gain insight into performance. Training with objective metrics makes training both more productive and more satisfying. It is rare in the world to have a game where the variables that need to be optimized are clearly defined and are completely a function of effort.