Training and racing using a power meter: an introduction

Andrew R. Coggan, Ph.D.

Revised: 25 March 2003

1. Introduction

The ideal training program for any athlete is one that is challenging enough to result in continual improvement, but is not so taxing that it results in illness, injury, or overtraining. Achieving this delicate balance can be difficult in any sport. However, this is especially true in cycling, because the resistance to forward motion varies markedly depending on altitude, weather, terrain, road or trail surface, and/or the effects of drafting. Consequently, current or even average speed is often a poor indicator of training intensity, which can make it difficult to regulate the overall training load (which is also determined by training duration and frequency). Although this tends to be less of a problem in track cycling (especially indoors), there can still be significant day-to-day or track-to-track differences in the physical effort required to achieve a given level of performance, e.g., a certain lap time. Thus, some measure other than simple velocity is required to accurately quantifying training intensity while cycling.

Monitoring heart rate (HR) provides one possible way around the above problem, since at least under carefully standardized conditions there is a close relationship between HR and the actual exercise intensity (i.e., power output or rate of oxygen consumption (VO_2)) (Fig. 1). This method has therefore been widely adopted in cycling and to a lesser degree in other sports (e.g., running). However, while theoretically sound the use of HR to quantify training intensity does have certain practical limitations. One is that although HR is closely correlated with exercise intensity in a laboratory-type setting, this relationship is not nearly as strong while cycling outdoors (Fig. 1). This is due to the wide variety of factors that can influence HR during exercise. For example, altitude, heat, hypohydration/dehydration, recent illness or infection, lack of sleep, and large fluctuations in power output (e.g., in a group ride setting, or in hilly terrain) all tend to increase HR during exercise at a given intensity, whereas acute overreaching has the opposite effect. In addition, the relationship of HR to power can differ between individuals, even if normalized in some manner, e.g., to the HR measured during a time trial (TT), or to maximal HR measured at the end of an incremental exercise test. As a result of such factors, the actual demands imposed by training can differ considerably between workouts or between individuals even if HR or relative HR is kept the same. Moreover, since HR responds relatively slowly (half-life = ~30 s) to changes in exercise intensity, HR monitoring cannot be used to regulate the intensity of shorter efforts, such as brief intervals aimed at enhancing anaerobic capacity or sprints designed to increase neuromuscular power. Finally, it must be kept in mind that HR is not a direct determinant of performance, but is simply a reflection of the strain imposed on the cardiovascular system by the exercise. (This last point is seemingly often overlooked, as demonstrated by the frequency with which coaches and athletes emphasize the need to minimize HR during exercise, when in fact the true goal is to maximize performance regardless of the "cost" in terms of HR.) Thus, while HR monitoring can be useful for detecting training-induced changes in cardiovascular fitness (i.e., maximal oxygen uptake, or VO_2max), it will generally be insensitive to changes in other key determinants of performance, most importantly the rider's metabolic fitness, i.e., their lactate threshold (LT).

The above limitations can be avoided by directly measuring the rider's actual power output, something that can be easily done now that commercial on-bike power meters are widely available. Compared to measuring speed or HR, measuring power has the advantage of providing both a more **direct** and a more **immediate** answer to the question "how hard am I working?" That is, an individual's power output directly determines not only how fast they can pedal down

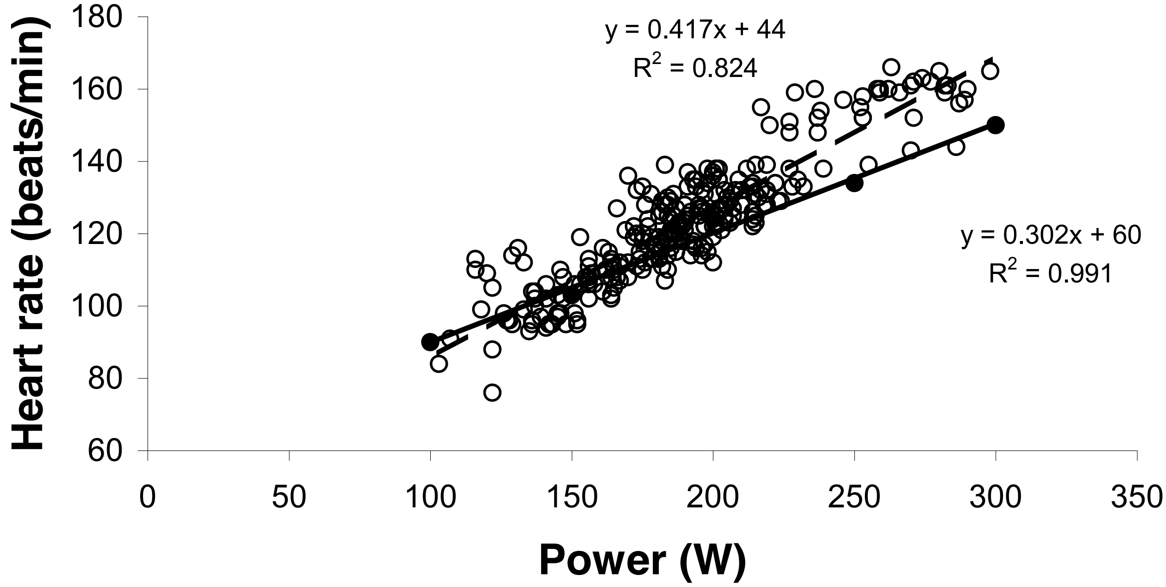

Fig 1. Relationship of heart rate to power in one athlete cycling outdoors (open circles) vs. indoors on an ergometer (closed circles).

the road or up a hill, but also their cardiovascular, metabolic, and perceptual responses to doing so. In other words, it is power output that matters, not only from the perspective of physics, but also from the perspective of physiology. Furthermore, changes in power are detected quite rapidly, without the lag inherent in HR, or even in velocity. Consequently, knowing the rider's power should make it possible to better regulate, or at the very least assess, the overall intensity of training. In addition, regularly measuring power in training and especially during races provides a direct indicator of the efficacy of training, and thus allows the training program to be fine-tuned to achieve maximum results.

Despite such advantages, many coaches and athletes remain uncertain about the actual benefit to "training by power", and/or how to best implement the use of a power meter as a training tool. This is probably because power meters, unlike portable HR monitors, have only recently become widely available – as a result, to date few (if any) training approaches built around the use of such instruments have been described. The purpose of this chapter is therefore to describe the author's approach to training using a power meter, as a means of illustrating some of the possibilities, as well as some of the pitfalls, of power-based training. A series of training levels, or zones, based on power will first be presented, followed by sample workouts meant to serve as examples of how a power meter can be employed to advantage in various situations. Analysis of power meter data will then be discussed, and a means of quantifying the overall training stress based on such measurements will be presented. Finally, other potential uses of a power meter (e.g., as a pacing tool in TTs) will be briefly discussed.

2. Power-based training levels

The training levels presented in Table 1 were developed based upon fundamental principles of exercise physiology, as well as approximately two decades of experience with power-based training, originally in a laboratory and more recently (with the advent of commercial on-bike power meters) in a field setting. The goal was to formulate a logical system for training by power from first principles, rather than attempt to derive training levels secondarily from HR measurements. (The latter approach is fraught with difficulties because of the variability of HR within and between individuals.) Even so, the resulting power-based system has certain parallels with HR-based systems developed previously by others, most notably that put forth by Peter Keen and used by the British Cycling Federation (especially in regards to the verbal descriptions of each training level). This parallelism is largely due to the fact that both the current power-based system and prior HR-based systems are founded on the same underlying phenomena, i.e., the physiological responses to exercise. However, to some extent it also reflects a conscious attempt to build upon previous efforts by incorporating desirable features of these prior systems into the present classification scheme. Some of the thinking that went into the development of this system is described below.

Basis for system/number of levels: Power at LT is the most important physiological determinant of endurance cycling performance, since it integrates VO_2max, the percentage of VO_2max that can be sustained for a given duration, and cycling efficiency (1). As such, it is more logical to define training levels relative to an athlete's threshold power, vs., for example, power at VO_2max (just as it is more logical to define HR-based training levels relative to threshold HR vs. maximal HR). On the other hand, determining the appropriate number of levels is somewhat arbitrary, since the physiological responses to exercise really fall on a continuum, with one intensity domain simply blending into the next. A compromise must therefore be made between defining more levels, thus better reflecting this fact, and defining fewer levels, for the sake of simplicity. In the present system, seven levels were felt to be the minimum needed to represent the full range of physiological responses and to adequately describe the different types of training required/used to meet the demands of competitive cycling. Table 2 lists the primary physiological adaptations expected to result from training at each level, although these will obviously be influenced by factors such as the initial fitness of the individual, the duration of each workout, the time taken between each interval effort, etc.

Determination of LT power: At least in theory, the most precise way of determining an athlete's power at LT would be to rely on laboratory-based testing with invasive blood sampling. Very few individuals, however, have access to such measurements on a routine basis. Furthermore, while LT is often defined by sports scientists as the initial non-linear increase in lactate with increasing exercise intensity (Fig. 2), this intensity tends to be significantly below that which coaches and athletes tend to associate, on the basis of practical experience, with the concept of a "threshold" exercise intensity. The latter corresponds more closely to what the sports science community has termed OBLA (onset of blood lactate accumulation, defined as a blood lactate concentration of 4 mmol/L), but is really conceptually closest to MLSS (maximal lactate steady state) or IAT (individual anaerobic threshold), both of which represent the highest exercise intensity that can be maintained without a continual increase in blood lactate. In terms of understanding the physiology of exercise, it actually makes little difference which of these various definitions is used, since they are all highly interrelated. On the other hand, this plethora of definitions does tend to complicate the use of lactate measurements for the purposes of exercise prescription, especially since determining the precise lactate level that corresponds to a given athlete's sustainable power (or HR) can be problematic.

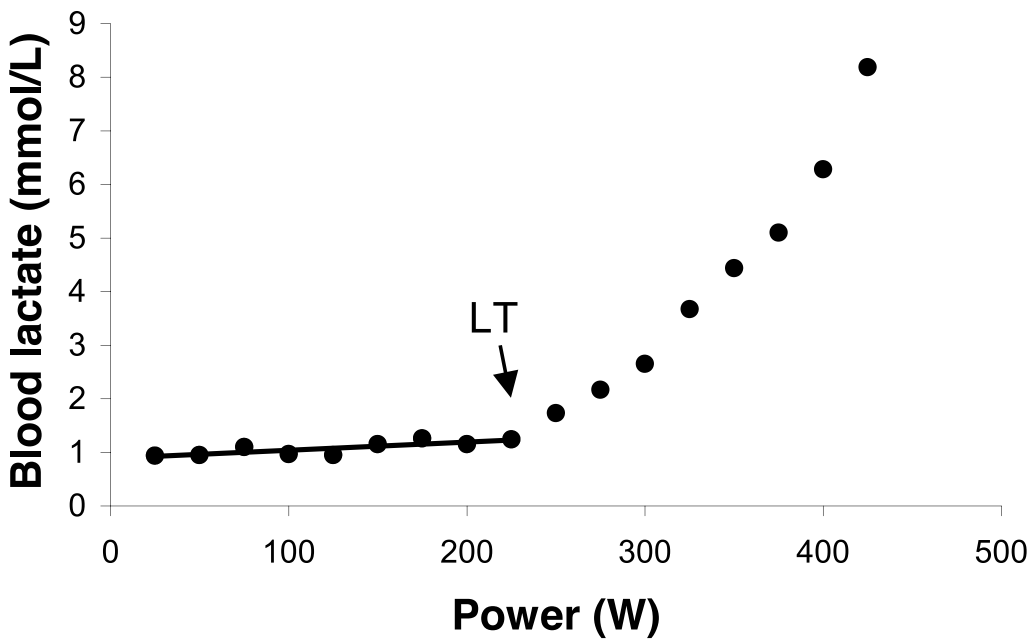

Figure 2. Blood lactate response in a well-trained cyclist during an incremental exercise test.

Given the limitations of laboratory testing as discussed above, probably the easiest and most direct way of estimating a rider's ***functional*** threshold power is therefore to simply measure their average power during a ~40 km (50-70 min) TT. This highly pragmatic approach is justified by laboratory research showing that the power a cyclist can generate for 60 min correlates very highly with, but is slightly greater than, their power at LT (defined as a 1 mmol/L increase in blood lactate over exercise baseline) (2). The precise value obtained for threshold power using this approach may vary slightly depending on the exact distance/duration of the TT, the terrain, the athlete's level of motivation and ability to pace themselves properly, etc. However, such variability is likely to be small relative to the breadth of the defined training levels and the somewhat arbitrary division between them. Furthermore, the simplicity of the approach means that the test (which doubles as a level 4 training session) can readily be repeated if the data obtained are considered suspect, or if there is reason to believe that the athlete's fitness has changed significantly. If for some reason (e.g., phase of training) it is considered undesirable to have the athlete perform a full 40 km TT, data from a shorter TT can be used instead, although this may require slight adjustment of the exact percentages of threshold power for each level and/or application of an appropriate correction factor (e.g., threshold power = average power during a 20 km TT multiplied by 0.93). Again, however, given the breadth of the specified power levels, day-to-day variability in performance, and individual differences in the precise shape of the power-duration curve, the real effect of employing such a correction factor may simply be to convey a false sense of precision.

An even easier way of estimating an athlete's threshold power is to just measure the power that they can routinely produce in training during long intervals or repeats aimed at raising LT (e.g., 2 x 20 min at level 4). Typically, this will be very close (within perhaps 5 percent) to what can be sustained during a 40k TT, with the shorter duration and recovery period(s) between efforts compensating for the generally lower motivational level in training vs. competition. (Average HR during such efforts, on the other hand, will often be significantly below that observed when racing.) The primary advantage to this approach is the ease of measurement, which in some cases may make it preferable to more formal testing.

Yet another, albeit more complicated, way of estimating threshold power is to rely on the critical power paradigm originally described by Scherrer in 1954 (cf. 3). Conceptually, critical power is a power that can be sustained "for a very long time without fatiguing", and is "an inherent characteristic of the aerobic energy supply system". Experimentally, an individual's critical power has been found to be closely related to (although again somewhat higher than) their power at LT as determined via laboratory measurements. A number of mathematically equivalent expressions exist for calculating critical power, but in the present context the most convenient formula is:

$$W = CP * t + AWC$$

where W is the total work (in joules) accomplished during a high intensity exercise task performed to fatigue, CP is critical power (in watts), t is time (in s), and AWC is anaerobic work capacity (in joules). The above equation describes a straight line (i.e., $y = mx + b$), which can be easily fitted to the data using commonly available software (e.g., Microsoft Excel). In this formulation, the slope (CP) reflects the maximum rate at which work can be performed aerobically without fatigue occurring, whereas the intercept (AWC) equals the total amount of work that can be accomplished by relying on non-renewable anaerobic energy sources (i.e., breakdown of ATP and PCr and production and accumulation of lactate) (Fig. 3). This interpretation is supported by experiments showing that CP is influenced by interventions that would be expected to affect aerobic energy production, e.g., hypoxia, whereas AWC is not. Conversely, interventions expected to influence anaerobic capacity, such as creatine loading, have been shown to alter AWC without changing CP. Finally, close correlations have been found between AWC and the total work performed during an all-out 30 s exercise test (i.e., a Wingate test), or between AWC and maximal accumulated oxygen deficit (currently considered the gold standard measurement of anaerobic capacity).

While useful, the CP concept is not without certain limitations. For example, it greatly overestimates the power that can be generated during very short duration exercise, and it incorrectly predicts that there should be a power output below which fatigue will *never* occur. In addition, the precise values obtained for AWC and, to a somewhat lesser extent, CP, depend in part on the testing protocol, especially the exact combination of powers and durations used (specifically, inclusion of progressively longer efforts tends to result in progressively lower estimates of CP). For this reason, it is best to carefully standardize testing conditions and to use data only from efforts that are between 3 min and perhaps 30 min in duration (anaerobic capacity may not be fully utilized during efforts that are shorter than 3 min in length, leading to underestimation of AWC and overestimate of CP). Despite such limitations, however, the CP approach can be useful if carefully applied, and at the very least provides a theoretical framework for understanding two of the most basic factors influencing exercise performance, i.e., anaerobic and aerobic energy production, and how the relative contribution of each varies as

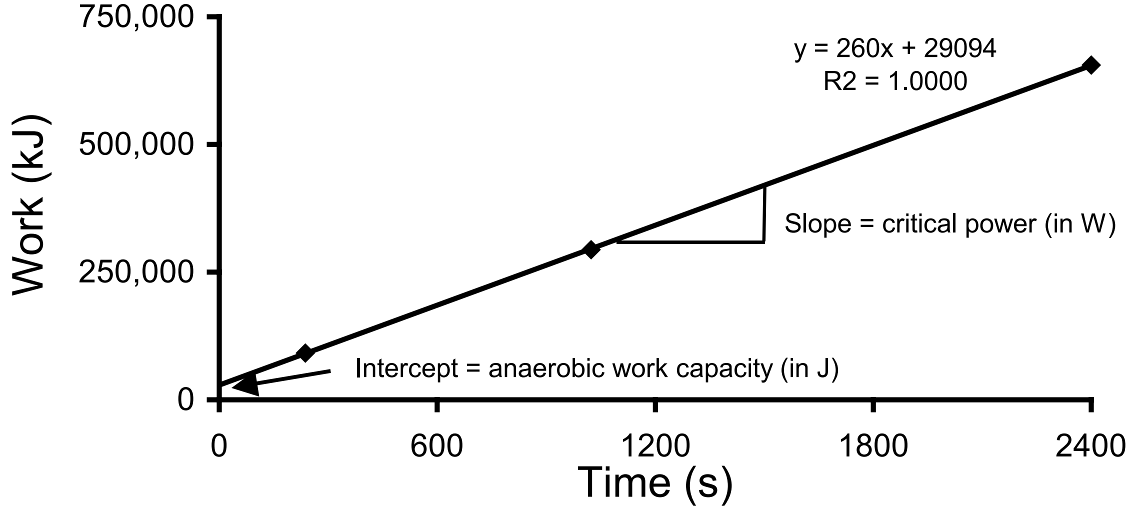

Figure 3. Calculation of critical power and anaerobic work capacity.

a function of time. For example, simply looking at the power-duration curve might lead one to conclude that the factors determining performance in a track pursuit and in a road TT are significantly different, since power falls off very rapidly during the first few minutes of exercise. By applying the CP concept, however, it becomes clear that performance in both events is heavily dependent on the individual's power at LT, since critical power plays a significant role in determining how much work they can perform even during relatively short duration exercise (Fig. 3.). From a training perspective, this makes it easier to understand why elite pursuiters often train 30,000-40,000 km/y. Similarly, application of the CP concept helps explain why even lower category or masters racers whose events might be less than 1 h in duration can often still benefit from multi-hour training sessions.

Lastly, it is also possible to estimate an athlete's threshold power from highly variable power data, such as that recorded during a typical criterium or hilly circuit race, by applying an appropriate mathematical algorithm. This method, which is explained in detail later in this chapter (see *Analysis of Power Meter Data*), has the advantage of not requiring any formal testing or adjustment to a rider's training or racing schedule, and can be used independently of the methods described above, or (better still) in conjunction with one of these other three methods as a means of confirmation. This technique works best, however, when applied to data from races in which the rider was very aggressive, and/or where the level of competition was high – otherwise, threshold power may be somewhat underestimated simply because the rider was not pushing themselves to the limits of their ability.

HR guidelines*:* Relating or translating the specified power levels to corresponding HR ranges or zones is somewhat difficult, due to the inherent variability of HR as well as individual differences in the power-HR relationship (even when referenced to threshold power). Nonetheless, approximate HR guidelines have been provided in Table 1, such that they can be used along with power to help guide training if desired.

Perceived exertion (PE) guidelines*:* The values given are from Borg's 10 point category-ratio scale (reproduced below), not the original 20 point scale that is more commonly used. The category-ratio scale is used because it explicitly recognizes the non-linear response of many physiological variables (e.g., blood and muscle lactate), and thus provides a better indicator of overall effort.

Borg's 10 point category-ratio scale of perceived exertion:

0 = Nothing at all	6
0.5 = Extremely weak (barely noticeable)	7 = Very strong
1 = Very weak	8
2 = Weak (light)	9
3 = Moderate	10 = Extremely strong
4 = Somewhat strong	* = Maximal
5 = Strong (heavy)	

Since perceived exertion increases over time even at a constant exercise intensity (power), the suggested values or ranges refer to perceived effort as determined relatively early in a training session/series of intervals.

***Other issues**:* While the system is based on the average power during a workout or interval effort, consideration must also be given to the distribution of power (this issue is discussed in greater detail under *Analysis of Power Meter Data* and *Limitations of Power-Based Training*). For example, average power during mass start races typically falls within level 3, but races are often more stressful than training at level 3, due to the greater variability (and therefore higher peaks) in power. Similarly, due to soft-pedaling/coasting, the same average power achieved during a hilly ride or group training session will not reflect the same stress as the same average power achieved during a completely flat ride or solo workout. In part, the variability in power is taken into account in defining the various levels, especially levels 2 and 3 (training at the higher levels will tend to be much more structured, thus limiting variations in power). Furthermore, there is obviously an inverse relationship between power output and the duration that power can be sustained. Thus, it is axiomatic that power during shorter training sessions or efforts will fall towards the higher end of a given range, whereas power during longer sessions or efforts will fall towards the lower end of a given range. Nonetheless, a workout consisting of, for example, 30 min of cycling at level 1 (as warm-up), 60 min of cycling at level 3, and another 30 min of cycling at level 1 (as warm down) would best be described as a tempo training session, even though the overall average power might fall within level 2.

Sample workouts: Table 3 illustrates application of the classification scheme for an athlete whose power and HR during a 40k TT averaged 300 W and 162 beats/min, respectively. Sample workouts for this individual are then listed in Table 4. These examples are given primarily to demonstrate how a power meter can be useful in prescribing/monitoring the intensity of training, and should not be viewed as "perfect" workouts necessarily intended to be emulated.

3. Analysis of power meter data

At least in theory, one of the advantages of training and racing with a power meter is that doing so makes it easier to more precisely control the overall training load. By continuously recording power output, the exact demands of each workout can be more accurately quantified, and the intensity or duration (or both) of subsequent training sessions can be modified as necessary to avoid either under- or overtraining. Successful application of this approach, however, requires that the athlete or coach be able to quickly make sense out of the huge amounts of data that are amassed when power output (along with other variables, e.g., HR) is recorded every second or so during multi-hour training rides. This task is made more difficult by the fact that power is highly variable when cycling outdoors, such that the overall average power may give little insight into the actual stress imposed by a given workout. This is especially true for races, since fluctuations in power are further exaggerated by tactical considerations, e.g., by the need to maintain one's position in a large field, or by the need to initiate or respond to attacks. The issue is therefore how to best summarize or condense power meter data while still adequately capturing the actual demands of each race or training session.

One approach that has been used by some is to simply record the total work (in kJ) performed during a race or training session. Expressing the data in this manner can be helpful in understanding the overall energy demands of training and e.g., how this compares to energy intake (useful, for example, when an athlete is trying to alter their body composition). However, like keeping track of miles or hours of training, measuring total work only provides an indication of overall training volume, and says nothing about the actual intensity of that training.

Another means of analyzing power data is to determine the frequency distribution of power output, i.e., the percentage of total ride time when power falls within a certain range (e.g., between 200 and 250 W) or level/zone (e.g., within level 4). Such analyses can be useful, but have two major limitations:

1) a relatively large number of numeric values is still needed to represent a single training session. Such data are therefore best presented graphically (e.g., as a bar chart), and are themselves not readily amenable to further analysis. Furthermore, while large differences in power distribution are readily detectable using this approach, more subtle differences are harder to identify.

2) more importantly, such analyses do not (and in fact readily cannot) take into account how long each "foray" into a given power range or level actually lasts. That is, the frequency distribution histogram will look essentially the same regardless of whether an athlete produced, e.g., 300 W continuously for 30 min, or performed six, 5 minute intervals at that power output. Obviously, however, the physiological responses and adaptations to two such different training sessions would be markedly difference. In theory, this problem can be overcome by preparing a three dimensional histogram, in which each data bin is defined not only by the power output, but also the time spent at that power (Fig. 4). This requires, however, establishing rather arbitrary cut-off criteria to define when a given effort begins and ends. Perhaps more importantly, representing the data in this manner is too complex for routine use.

The limitations of the above methods for analyzing or summarizing power meter data files led to development of an alternative approach, which is described below.

Intensity factor (IF) and training stress score (TSS): Dr. Eric Banister has previously proposed quantifying training load in terms of a HR-based "training impulse", or TRIMPS, score (4):

TRIMPS = exercise duration x average HR x a HR-dependent intensity weighting factor

Since HR is related to oxygen uptake or metabolic rate (Fig. 1), the product of the first two factors in the above equation is proportional to the amount of energy expended, or (since

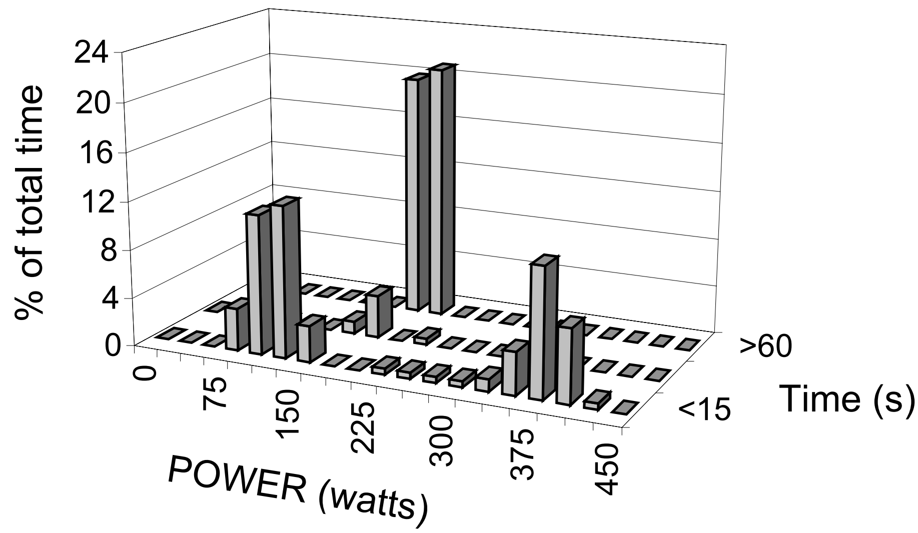

Figure 4. Three dimensional frequency distribution histogram of power output during a "micro interval" workout (15 s "on"/15 s "off " for 2 h).

efficiency is relatively constant), the amount of work performed. The third term then takes into account the intensity of the exercise, since many physiological responses (e.g., glycogen utilization, lactate accumulation) increase non-linearly with increasing intensity.

By analogy, power meter can be used to derive a "training stress score", or TSS:

TSS = exercise duration x average power x a power-dependent intensity weighting factor

Similar to TRIMPS, the product of the first two factors in the above equation is equal to the total work performed, whereas the "intensity factor" (IF) serves to account for the fact that the physiological stress imposed by performing a given amount of work (e.g., 1000 kJ) depends in part on the rate at which that work is performed (i.e., on the power output itself).

To derive an appropriate algorithm for calculating IF, blood lactate data collected from a large number of trained cyclists exercising at intensities both below and above their LT were analyzed. This choice was made because many physiological responses (e.g., muscle glycogen and blood glucose utilization, catecholamine levels, ventilation) tend to parallel changes in blood lactate during exercise – in this context, then, blood lactate levels can be viewed as an overall index of physiological stress. To reduce variability between individuals, the data were normalized by expressing both the power output and the corresponding blood lactate level as a percentage of that measured at LT. The normalized data were then used to derive a best-fit curve. Perhaps not surprisingly, an exponential function provided the best fit, but a power function of the following form proved to be nearly as good:

blood lactate (% of lactate at LT) = power (% of power at LT)$^{3.90}$; R^2=0.806, n=76

Based on these data, a 4th-order function was used in the algorithm for determining the IF (the exponent was rounded from 3.90 to 4.00 for simplicity's sake).

The other physiological information incorporated into the algorithm for calculating IF is the fact that physiological responses to changes in exercise intensity are not instantaneous, but followed a characteristic time course. Because of this, for example, exercise in which the intensity rapidly (e.g., every 15 s) alternates between a high and a low level (e.g., 400 and 0 W) results in physiological, metabolic, and perceptual responses nearly identical to steady-state exercise performed at the average intensity (i.e., 200 W). The specific reasons for this are beyond the scope of this discussion, but the important facts are 1) the half-lives (50% response time) of many physiological responses are directly or indirectly related to metabolic events in exercising muscle, and 2) such half-lives are typically on the order of 30 s. Thus, to account for this fact the power data were smoothed using a 30 second (~1 half-life) rolling average before applying the 4th order weighting as described above.

Finally, to make comparisons across individuals more convenient (e.g., for coaches who must deal with multiple athletes), 1) the IF was expressed as a ratio of the normalized power (obtained by smoothing/weighting as described above) to that individual's threshold power, and 2) the TSS was normalized to the amount of work that could be performed during one hour of cycling at threshold power (=100 TSS points). The steps required to calculate IF and TSS from power meter data therefore are:

1) starting at 30 s, calculate a 30 second rolling average for power
2) raise the values obtained in step 1 to the 4th power
3) take the average of all the values obtained in step 2
4) take the 4th root of the number obtained in step 3
5) divide the normalized power obtained in step 4 by the individual's power at LT – the resulting decimal value is the IF
6) multiply the normalized power by the duration of the effort (in s) to obtain the normalized work performed (in J)
7) multiply the normalized work by the IF (step 5) to derive the "raw" TSS
8) divide the "raw" TSS by the amount of work that could be performed in one hour at threshold power (i.e., threshold power x 3600 s) and multiply by 100 to obtain the final TSS

(These calculations are obviously too cumbersome to routinely perform on every power meter file, or part thereof, even when e.g., using a macro in Excel – however, software is available to automate the process.)

Applications: The most obvious application of the method described above is to quantify the overall training load, in terms of the number of TSS points accumulated during a given training block. For example, by keeping track of the total TSS per week or per month, it may be possible to identify an individual's "breaking point", i.e, the maximum quantity and quality of training that still leads to improvements, rather than overtraining. As well, a very high TSS resulting from a single race or training session may be an indicator that additional recovery on subsequent days is required. The table below gives some rough guidelines for typical TSS scores, and the impact they would be anticipated to have on an athlete's subsequent performance ability:

<150	low (relatively easy to recover by following day)
150-300	medium (some residual fatigue may be present the next day, but gone by 2nd day)

300-450	high (some residual fatigue may be present even after 2 days)
>450	epic (residual fatigue lasting several days likely)

Note that while the TSS score is normalized to the individual's threshold power, such that comparison across athletes is possible, there could still be differences between riders in how they respond to a given "dose" of training. Such difference may be due to natural ability, or may be the result of specific training (i.e., the more you train the more you can train). This is not a major issue, however, since comparisons within a given individual are of primary interest.

While the goal of developing TSS was to provide a way of quantifying the overall training load (duration x intensity) based on power meter data, the IF score and the algorithm used to derive it have other important uses as well. For example, the IF can be used to compare the intensity of markedly dissimilar training sessions or races, either within (most valid/relevant) or across (e.g., to assess tactical or drafting skill in the same race) individuals (see below):

Typical IF values for different events or training sessions:

<0.75	level 1 recovery rides
0.75-0.85	level 2 endurance training sessions
0.85-0.95	level 3 tempo rides, various aerobic and anaerobic interval workouts (work and rest periods combined), longer (>2.5 h) road races
0.95-1.05	level 4 intervals (work period only), shorter (<2.5 h) road races, criteriums, circuit races, 40k TT (by definition)
1.05-1.15	shorter (e.g., 15 km) TT, track points race
>1.15	level 5 intervals (work period only), prologue TT, track pursuit, track miss-and-out

[It should be evident that the IF values given in the table above are actually the fraction or percentage of threshold power that was equivalently maintained. As such, the IF is analogous to the percentages used to define the training levels described in Table 1 – the absolute values differ, however, because the IF score corrects for the effects of variations in power on physiological responses, whereas the training levels have simply been offset to lower power levels to account for this fact. For example, a level 1 training ride would have an IF value of <0.75 (i.e., normalized power was <75% of threshold power), but the average power (uncorrected for variability) would be <55% of threshold power.]

The algorithm used to derive IF also makes it possible for the first time to accurately estimate an individual's threshold power from highly variable power data such as that obtained in criterium. That is, if sustainable power (either constant or non-constant) is essentially "capped" by the athlete's LT, and if the 30 s smoothing/4^{th} order weighting algorithm appropriately adjusts the variable power data, then the normalized power obtained following step 4 in the calculation of IF/TSS (see above) provides an estimate of the equivalent steady power that could be produced for the same physiological stress. Stated another way, the algorithm simply provides a means of expressing highly variable power data in physiologically-relevant "language". Consequently, if an individual pushes themselves just as hard in a ~1 h mass start race (or TT in very hilly terrain) as they might in a flat TT, then the normalized power provides an estimate (generally to w/in 5-10 W) of their threshold power. This observation reduces, and in some cases may even completely eliminate, the need for the rider to perform a TT to determine their threshold power – instead, the results of mass start races can be used for this purpose. This approach may prove useful for beginning power meter users who have never had the opportunity to use such a tool in

TT. Even for riders whose threshold power is well established, the IF score can be used to detect significant changes in fitness – for example, if a rider's IF score for a ~1 h race is greater than 1.05, then their threshold power should be reassessed (ideally using the same means used to establish it originally) to determine whether it has truly changed.

Finally, yet another application of the IF algorithm/score is as a teaching tool, as it helps demonstrate why, even when power is highly variable, it is still an individual's "metabolic fitness" (i.e., LT) that is important in determining performance. That is, by illustrating (via a 4th order relationship – greater even than the 3rd order relationship between power and wind resistance!) how physiologically "costly" every sustained burst above LT truly is, the IF algorithm may 1) help less experienced riders understand why it is important to learn how to modulate their effort during mass start races, so that they don't fatigue themselves unnecessarily, and 2) help even experienced riders understand how appropriate training aimed at raising threshold power can improve performance even in events seemingly much different from a TT (e.g., a criterium).

4. Limitations to power-based training

While power-based training has a number of advantages, it has disadvantages as well. One is that very goal-oriented athletes may become too focused on the power data, expecting or attempting to improve every training session and becoming disappointed or discouraged when they fail to do so. This is not, however, a problem unique to just power-based training, and is best dealt with *a priori*, i.e, by making sure that the athlete has realistic expectations and is provided with appropriate feedback.

A somewhat more important limitation to power-based training is that on a moment-by-moment basis, power during outdoor cycling tends to be extremely variable in nature. This is primarily due to the constantly changing resistances that are encountered outdoors, as most riders are capable of maintaining a relatively constant power on a trainer or on rollers. In any case, this variability can make it difficult to modulate power, at least over brief periods of time, to remain within a prescribed range (as is typically done during HR-based training). As a result, some coaches have advocated relying primarily on HR during low to moderate intensity training (e.g., levels 1-3), using power data only to guide training at higher intensities (when variations in power are likely to be smaller due to the more structured nature of the training). Others have chosen to use even tighter ranges when prescribing training based on power instead of HR, in an attempt to force the athlete's performance to more closely coincide with that envisioned or desired by the coach. In the author's opinion, however, it is counterproductive to excessively limit variations in power during training, regardless of whether one does so by using a power meter *or* a HR monitor. The simple fact is that power during outdoor cycling is highly variable, especially during races, and attempting to "micromanage" the athlete's efforts to minimize such variations merely makes the training less specific. (This is not to say, however, that great lengths must be taken in training to mimic the variations in power that occur during racing – only that at least some spontaneous variation in power output during training is desirable, even necessary, for optimal effectiveness.) The training levels provided in Table 1 are therefore based on the *average* power during an interval effort or training session, and it is the average that the individual should be instructed to primarily focus on regulating, relying on their current power output, their perceived effort, and the verbal descriptions given in Table 1 to help them do so. With constant feedback from a power meter helping to hone their effort sense, athletes generally quickly develop the ability to "dial up" the approximate desired power output when doing intervals, and/or learn to pace themselves appropriately during longer (e.g., level 2) training sessions. The feedback provided by observing changes in power across repeated intervals efforts, or by the overall IF for longer workouts, can also be helpful in teaching an athlete how to properly modulate the exercise intensity.

5. Other uses for a power meter

While the primary use for a power meter is as a training tool, such devices have other applications as well. For example, a power meter can be used in combination with a simple magnetic trainer to conduct simple but informative fitness tests, such as estimating VO_2max from submaximal HR measurements, or to track changes in anaerobic capacity via periodic determination of "critical power". Feedback from a power meter can also be helpful for pacing purposes during TTs – this is true even for experienced athletes, since the tendency to start out too fast is very difficult to overcome, especially in competition. This approach works best in flat TTs held under low-wind conditions, where an isopower effort is optimal, but can also be used to advantage in hilly or windy TTs, where the best performance (lowest cumulative time) is achieved by increased power output on uphill or headwind segments, and recovering on downhill or tailwind sections of the course. While how much an individual athlete should vary their effort under such conditions can only be learned by experience, being able to observe one's actual power output can help speed up this "learning curve".

Data obtained via racing with a power meter can also be used to evaluate performance, and thus to assess the efficacy of training and to determine what changes might need to be made to an athlete's program. Obviously, this is most true for TTs, where variations in power output may be apparent even when time or average speed do not differ, due to differing environmental conditions. However, power data from mass start races can also be used to assess fitness on the basis of the normalized power maintained, i.e., the IF value. Furthermore, such data may provide insight into an athlete's relative strengths and weaknesses, or even to help choose appropriate tactics. Suppose, for example, that an athlete loses contact with the lead group in a road race the 4^{th} time up a steep, 1 mile long hill on the course. Comparison of the power profile from that ascent to what the athlete generated the previous three times up the hill and to what they are capable of producing in training during a single all-out effort of similar duration may indicate whether they were dropped due to cumulative fatigue, or if the pace of the race had simply increased at that point to a level they simply could not match. If the former, then perhaps longer and/or more frequent level 2 and 3 training sessions and/or an increase in total training volume may be necessary. On the other hand, if the latter were true then perhaps a greater emphasis on level 5 and level 6 intervals would help the individual perform better in future similar races. In either case, the post-race analysis of the power data may also help the athlete recognize the need to institute certain strategies, such as making a concerted effort to conserve energy just prior to a major climb or starting such climbs near the front of a group, that may aid their overall performance.

Finally, yet another potential use of a power meter is for estimation of an individual's aerodynamic drag characteristics, i.e., the product of their coefficient of drag (Cd) and frontal area (A). If speed and power data are collected in a relatively well-controlled environment (e.g., on a flat road or velodrome under low wind conditions) and air density is determined, it is possible to rearrange the equation describing the power requirements of outdoor cycling (5) to solve for CdA with a precision (reproducibility) of approximately 2% (Table 5). Unfortunately, even this small amount of variability is enough to prevent this method from being particularly useful – with experience, it is generally possible to position an athlete on their TT bike well enough "by eye" that the potential for further improvements in CdA is similar in magnitude to the error of the measurement. This means that numerous trials must be performed (ideally on the same day, with no change in air density or wind speed or direction) to detect a significant change resulting from, say, altering arm position or aerobar height. Detecting changes in CdA due to differences in equipment is even more difficult, if not impossible. It may be possible to obtain more precise (and thus more useful) estimates of CdA by performing such testing on an indoor velodrome, but few individuals have regular access to such a venue. Moreover, any estimate of CdA obtained under such still-air conditions will not reflect what happens when winds come from one side or the other, i.e., when there is an appreciable yaw angle. Thus, because of its greater accuracy and precision (as well as convenience/time requirement – although not necessarily cost), wind tunnel testing clearly remains the method of choice for determining CdA.

6. Summary

The availability of affordable, reliable on-bike power meters is contributing to changes in the training of cyclists for competition. Although such changes are really far more evolutionary than revolutionary in nature – after all, much knowledge about training (empirically derived or otherwise) already exists and athletes must still perform the necessary "homework" – the increasingly widespread use of such tools is nonetheless likely to contribute to improved performances by many. Coaches and athletes who do not avail themselves of the opportunities provided by such instrumentation therefore risk finding themselves left behind.

7. References

1. Coyle EF. Integration of the physiological factors determining endurance performance ability. Exerc. Sports Sci. Rev. 23:25-63, 1995.
2. Coyle EF, Coggan AR, Hopper MK, Walters TJ. Determinants of endurance in well-trained cyclists. J. Appl. Physiol. 64:2622-2630, 1988.
3. Vanderwalle H, Vautier JF, Kachouri M, Lechevalier JM, Monod H. Work-exhaustion time relationships and the critical power concept: a critical review. J. Sports Med. Phys. Fitness 37:89-102, 1997.
4. Banister EW, Calvert TW, Savage MV, Bach TM. A systems model of training for athletic performance. Aust. J. Sports Med. 7:57-61, 1975.
5. Martin JC, Milliken DL, Cobb JE, McFadden KL, Coggan AR. Validation of a mathematical model for road cycling power. J Appl Biomech 1998; 14:276-291.

Table 1. Power-based training levels

Level	Name/purpose	Average power (% of threshold power)	Average HR (% of threshold HR)	Perceived exertion	Description
1	Active recovery	≤55%	≤68%	<2	"Easy spinning" or "light pedal pressure", i.e., very low level exercise, too low in and of itself to induce significant physiological adaptations. Minimal sensation of leg effort/fatigue. Requires no concentration to maintain pace, and continuous conversation possible. Typically used for active recovery after strenuous training days (or races), between interval efforts, or for socializing.
2	Endurance	56-75%	69-83%	2-3	"All day" pace, or classic long slow distance (LSD) training. Sensation of leg effort/fatigue generally low, but may rise periodically to higher levels (e.g., when climbing). Concentration generally required to maintain effort only at highest end of range and/or during longer training sessions. Breathing is more regular than at level 1, but continuous conversation still possible. Frequent (daily) training sessions of moderate duration (e.g., 2 h) at level 2 possible (provided dietary carbohydrate intake is adequate), but complete recovery from very long workouts may take more than 24 hs.
3	Tempo	76-90%	84-94%	3-4	Typical intensity of fartlek workout, 'spirited' group ride, or briskly moving paceline. More frequent/greater sensation of leg effort/fatigue than at level 2. Requires concentration to maintain alone, especially at upper end of range, to prevent effort from falling back to level 2. Breathing deeper and more rhythmic than level 2, such that any conversation must be somewhat halting, but not as difficult as at level 4. Recovery from level 3 training sessions more difficult than after level 2 workouts, but consecutive days of level 3 training still possible if duration is not excessive and dietary carbohydrate intake is adequate.

Table 1. Power-based training levels (cont.):

Level	Name/purpose	Average power (% of threshold power)	Average HR (% of threshold HR)	Perceived exertion	Description
4	LT	91-105%	95-105% (may not be achieved during initial phases of effort(s))	4-5	Just below to just above TT effort, taking into account duration, current fitness, environmental conditions, etc. Essentially continuous sensation of moderate or even greater leg effort/fatigue. Continuous conversation difficult at best, due to depth/frequency of breathing. Effort sufficiently high that sustained exercise at this level is mentally very taxing – therefore typically performed in training as multiple 'repeats', 'modules', or 'blocks' of 10-30 min duration. Consecutive days of training at level 4 possible, but such workouts generally only performed when sufficiently rested/recovered from prior training so as to be able to maintain intensity.
5	VO_2max	106-120%	>106% (may not be achieved due to slowness of heart rate response and/or ceiling imposed by maximum heart rate)	6-7	Typical intensity of longer (3-8 min) intervals intended to increase VO_2max. Strong to severe sensations of leg effort/fatigue, such that completion of more than 30-40 min total training time is difficult at best. Conversation not possible due to often 'ragged' breathing. Should generally be attempted only when adequately recovered from prior training - consecutive days of level 5 work not necessarily desirable even if possible.
6	Anaerobic capacity	≥121%	N/a	>7	Short (30 s to 3 min), high intensity intervals designed to increase anaerobic capacity. Heart rate generally not useful as guide to intensity due to non-steady-state nature of effort. Severe sensation of leg effort/fatigue, and conversation impossible. Consecutive days of extended level 6 training usually not attempted.
7	Neuromuscular power	N/a	N/a	* (maximal)	Very short, very high intensity efforts (e.g., jumps, standing starts, short sprints) that generally place greater stress on musculoskeletal rather than metabolic systems. Power useful as guide, but only in reference to prior similar efforts, not TT pace.

Table 2. Expected physiological/performance adaptations resulting from training at levels 1-7.

	1	2	3	4	5	6	7
Increased plasma volume		✓	✓✓	✓✓✓	✓✓✓✓	✓	
Increased muscle mitochondrial enzymes		✓✓	✓✓✓	✓✓✓✓	✓✓	✓	
Increased lactate threshold		✓✓	✓✓✓	✓✓✓✓	✓✓	✓	
Increased muscle glycogen storage		✓✓	✓✓✓✓	✓✓✓	✓✓	✓	
Hypertrophy of slow twitch muscle fibers		✓	✓✓	✓✓	✓✓✓	✓	
Increased muscle capillarization		✓	✓✓	✓✓	✓✓✓	✓	
Interconversion of fast twitch muscle fibers (type IIb -> type IIa)		✓✓	✓✓✓	✓✓✓	✓✓	✓	
Increased stroke volume/maximal cardiac output		✓	✓✓	✓✓✓	✓✓✓✓	✓	
Increased VO_2max		✓	✓✓	✓✓✓	✓✓✓✓	✓	
Increased muscle high energy phosphate (ATP/PCr) stores						✓	✓✓
Increased anaerobic capacity ("lactate tolerance")					✓	✓✓✓	✓
Hypertrophy of fast twitch fibers						✓	✓✓
Increased neuromuscular power						✓	✓✓✓

Table 3. Training level guidelines for an athlete who time trials at an average power of 300 W and an average HR of 162 beats/min.

Level	Name/purpose	Average power (W)	Average HR (beats/min)	Perceived exertion
1	Active Recovery	≤165	≤110	<2
2	Endurance	166-225 W	111-134	2-3
3	Tempo	226-270 W	135-152	3-4
4	LT	271-315 W	153*-170	4-5
5	VO_2max	316-360 W	>171†	6-7
6	Anaerobic capacity	≥361 W	N/a	>7
7	Neuromuscular power	N/a	N/a	(maximal)

*May not be achieved during initial phases of effort(s). †May not be achieved due to slowness of heart rate response and/or ceiling imposed by maximum heart rate.

Table 4. Sample workouts for the athlete whose training levels are described in Table 3:

Level	Purpose of training session	Prescribed workout	Notes/details
1	Active recovery	Ride for 1 h maximum @ level 1	Keep power <250 W (below middle of level 3) on all hills - avoid steep climbs, jumping out of turns, or being forced to ride harder than desired by road/wind conditions or by any training companions, especially at beginning of ride. If feeling better/more recovered than expected, power "ceiling" may be increased to level 2 (up to 225 W average) on level terrain and top of level 3 (270 W) on short hills, but only during last 15 min of workout and only if average power for entire session is still kept $\leq$165 W.
2	Basic endurance training	Ride for 3 h @ level 2	Unless feeling exceptionally tired or especially vigorous, power will almost automatically fall into level 2, but nonetheless consult power meter periodically to be certain average is within range. Occasional periods of continuous riding at level 3 (power up to 270 W) acceptable, but if so must be balanced with comparable periods of lower intensity training. Avoid extended periods with power at level 4 and above (i.e., power >270 W) unless necessitated by terrain (e.g., long steep climb).
3	Tempo training (race simulation)	Warm up by riding for 15-30 min @ level 1-2, then ride for 1.5 h @ level 3, followed by an additional 15-30 min @ level 1-2.	Best done on rolling to hilly terrain, alone or perhaps with a training partner of comparable ability. Attempt to maximize variation in power during tempo period by attacking climbs, accelerating hard out of turns, etc., while still keeping intensity high enough at all times to maintain average power within level 3. Concentrate on not letting average power fall during 2nd half of tempo period, as fatigue develops.

Table 4. (cont).

4	Development of LT	Warm up thoroughly as if for a race, by e.g., riding for 15-30 min @ level 1-2, including a few short (1-3 min) efforts at level 3-4. Then, perform 2 x 20 min @ level 4, with 5 min @ level 1 between efforts. Warm-down by riding an additional 15 min @ level 1-2.	Carefully "roll into" 1st interval, making sure that intensity does not exceed targeted power during the first few minutes. Thereafter, try to maximize average power while still keeping perceived effort just below actual race intensity. During 2nd interval, attempt to replicate this effort - inability to maintain average power during 2nd interval within 10 W of that of 1st efforts indicates either A) poor pacing (1st interval too intense), and/or B) inadequate recovery from prior training. Abandon workout if, based on prior experience, perceived effort is excessively high relative to average power.
5	VO_2max training	Warm up thoroughly as above, then complete 6 x 5 min @ level 5, with 2.5-5 min @ level 1 between efforts. Warm-down by riding an additional 15 min @ level 1-2.	Use power data to avoid starting out at an unsustainable intensity, either at the beginning of each interval or during the first few intervals, and as a "carrot" to maintain intensity during later efforts. Stop before completion of all efforts if unable to achieve goal power – if this happens in several consecutive workouts, reduce goal power for next such training session.
6	Anaerobic capacity	Warm up thoroughly as above, then perform 10 x 1 min @ level 6, with 3 min @ level 1 between efforts. Warm-down by riding an additional 15 min @ level 1-2.	Effort during routine training should be very high but not quite "all out" – see above. During peaking phase, increase absolute power, interval duration (e.g., to 2 min), and amount of recovery between efforts, and reduce number of repetitions to as few as 3 or 4. Terminate workout when average power during interval decreases by >10%.
7	Neuromuscular power	After a very thorough warm up, perform 6-10 all-out 10 s sprints, with complete recovery between efforts.	Perform sprints on slight uphill and/or from low velocity so as to maximize fast-twitch fiber recruitment. Terminate workout when maximum power achieved during sprint decreases by >10%

Table 5. Estimation of the product of coefficient of drag (Cd) and frontal area (A) from power meter data.*

Trial No.	Velocity (m/s)	Power (W)	Temperature (°C)	Barometric pressure (mm Hg)	Air density (kg/m^3)	CdA (m^2)
1	12.22	317.0	17.2	29.98	1.200	0.248
2	12.17	318.6	17.7	29.98	1.198	0.254
3	12.26	316.4	18.1	29.98	1.197	0.246
4	12.26	318.1	18.2	29.98	1.196	0.248
5	12.18	301.6	18.4	29.98	1.196	0.238
					Average	0.247
					Std. Dev.	0.006
					CV (%)	2.3%

*An SRM track crank was used the measure velocity and power of a pursuit cyclist performing flying 2 km repeats in the aero position on an outdoor 333.3 m concrete veldodrome. Air density was calculated from temperature and barometric pressure measured at trackside using a portable thermobarometer. CdA was estimated from these data using a previously validated model of the power requirements of outdoor cycling (5), assuming a coefficient of rolling resistance of 0.0035 and an effective incremental wheel frontal area of 0.0027 m^2 (to account for rotational wheel drag). Data were corrected for changes in kinetic energy but not for the effects of wind, which averaged <1 m/s during all trials. The cyclist wore a skinsuit and an unvented time trial helmet, and used an aerodynamically designed bicycle equipped with deep rim front and disk rear wheels.