REVIEW ARTICLE

Spc ts Med 2001, 31 (7` 489-496
C1 2 1642/01/0007-0489 '$22 00/0

Tests of Cycling Performance

Carl D. Paton[1] and *Will G. Hopkins*[2]

1 Centre for Sport and Exercise Science, The Waikato Polytechnic, Hamilton, New Zealand
2 Department of Physiology, School of Medical Sciences, University of Otago, Dunedin, New Zealand

Contents

Abstract

Performance tests are an integral component of assessment for competitive cyclists in practical and research settings. Cycle ergometry is the basis of most of these tests. Most cycle ergometers are stationary devices that measure power while a cyclist pedals against sliding friction (e.g. Monark), electromagnetic braking (e.g. Lode), or air resistance (e.g. Kingcycle). Mobile ergometers (e.g. SRM cranks) allow measurement of power through the drive train of the cyclist's own bike in real or simulated competitions on the road, in a velodrome or in the laboratory. The manufacturers' calibration of all ergometers is questionable; dynamic recalibration with a special rig is therefore desirable for comparison of cyclists tested on different ergometers.

For monitoring changes in performance of a cyclist, an ergometer should introduce negligible random error (variation) in its measurements; in this respect, SRM cranks appear to be the best ergometer, but more comparison studies of ergometers are needed. Random error in the cyclist's performance should also be minimised by choice of an appropriate type of test. Tests based on physiological measures (e.g. maximum oxygen uptake, anaerobic threshold) and tests requiring self-selection of pace (e.g. constant-duration and constant-distance tests) usually produce random error of at least ~2 to 3% in the measure of power output. Random error as low as ~1% is possible for measures of power in 'all-out' sprints, incremental tests, constant-power tests to exhaustion and probably also time trials in an indoor velodrome. Measures with such low error might be suitable for tracking the small changes in competitive performance that matter to elite cyclists.

Cycling coaches and sport scientists use field- and laboratory-based tests of cycling performance to assess competitive cyclists and to investigate nutritional strategies, new cycling equipment and other factors that might affect performance. Time trials and other field-based tests that reproduce the demands of competitive events are generally better than laboratory-based tests for these purposes,[1] but cycling performance outdoors is markedly affected by wind speed. Most cyclists do not have regular access to an indoor velodrome, so tests using stationary cycle ergometers are in widespread use. Mobile ergometers that measure power in the drive train of the cyclist's own bike have also come into use at the elite level in the last few years; these ergometers produce some of the best measures of performance in the laboratory or outdoors. In this review, we describe the various types of ergometer, their errors and the protocols for tests that can be used in the laboratory or field.

1. Types of Cycle Ergometer

All cycle ergometers measure power developed by the cyclist against some form of resistive load. In stationary ergometers, the load is an integral part of the device and is generated by either sliding friction, electromagnetic braking or air resistance. Mobile ergometers measure power developed against the resistance of real cycling.

The Monark is the most widely used make of friction-braked ergometer. The frictional force, which is generated by a belt sliding against a rotating flywheel, is balanced and measured by either a pendulum or a basket holding a range of weights.

In the numerous makes of electromagnetically braked ergometers, an armature rotates through a magnetic field to provide a resistive load. For most of these ergometers, work rate can be set to a constant value independent of pedalling speed, or it can change with changes in pedalling speed, as for friction-braked ergometers.

Several types of air-braked ergometers have been used for testing competitive cyclists. On Repco ergometers vanes are placed equidistantly around a flywheel to create air resistance. On Kingcycle ergometers the rear wheel of the cyclist's own bike drives a roller connected to an impeller fan to create air resistance; the rolling resistance of the rear wheel against the roller also provides resistance similar to the rolling resistance a cyclist would experience on the road.

Mobile ergometers consist of a device fitted to the drive train of the cyclist's own bike for measurement of torque and angular speed. These ergometers can be used during normal cycling or in conjunction with other cycle ergometers in the laboratory. Two makes are available. SRM cranks measure torque in the chain ring (the toothed wheel that drives the chain) with either 2, 4 or 8 (previously 20) strain gauges, depending on the model. Torque and angular velocity data are transmitted by induction from the crank to a unit on the handle-bar that converts the data to power. The latest mobile ergometer is the PowerTap hub, which works on the same principle as the SRM crank but interfaces with the drive cassette and hub of the rear wheel.

2. Errors in Cycle Ergometry

Cycle ergometers contribute two kinds of error to a cyclist's measure of performance: systematic and random. We will explain the meaning and importance of these errors with examples.

Suppose we drive an ergometer with a motor that delivers a true power of 400W. We sample the reading on the ergometer every minute and record 421, 418, 420, 417, 421, 424W ... The average (mean) of a large number of such readings is 420W, and the typical variation about the average (the standard deviation) is 2W. The systematic error is therefore +20W or +5%, and the typical random error is 2W or 0.5%. In statistical terms, systematic error refers to a consistent bias or offset in the reading of power provided by the ergometer, whereas random error refers to fluctuations in readings from measurement to measurement. Random error in an ergometer or other measuring instrument is sometimes referred to as 'noise'. We would correct the bias in any reading of around 420W by subtracting off 20W; the corrected reading would be different from the true reading by typically ±2W.

Systematic error is clearly an issue when the objective of assessment is to compare the performance of an athlete or athletes on different ergometers (e.g. when the test is used as a guide to team selection). Researchers estimate the systematic error by driving the ergometer with a dynamic calibration rig.[2-6] The rig consists of an electric motor connected to the drive shaft (the 'bottom bracket') of the cranks of the ergometer. Torque applied to the drive shaft is measured via a load cell, a weights balance or a torsional strain gauge. The power delivered to the ergometer is the torque multiplied by the angular speed of the drive shaft. By driving the ergometer over a range of powers, the researchers can derive a calibration equation to correct any reading on the ergometer. We summarise the findings of such studies in sections 2.1 to 2.4, under subheadings for each type of ergometer.

Random error reduces our ability to track changes in performance of individual cyclists when they perform repeated tests on the same ergometer (e.g. when monitoring change in fitness of a cyclist or measuring the effectiveness of an ergogenic aid in a controlled trial).[7] Researchers could estimate the random error in an ergometer by performing repeated 'tests' using a given test protocol (e.g. constant power to exhaustion), but with a calibration rig replacing the cyclist. This procedure would give an estimate of the random error that the ergometer contributes when a cyclist performs the test. Unfortunately, no researchers have used this procedure. Instead, they estimate the typical or standard error of measurement (or some other measure of reliability) for a human performing a test repeatedly on a given ergometer, because this error is what practitioners or scientists have to contend with in their assessment of patients, clients or study participants. Although this error is a mixture of ergometer error and biological test-retest variation, comparison of the error for similar participants performing similar tests on different ergometers allows at least a ranking of the random error in the ergometers.

On the basis of available studies, the authors of a recent review of reliability of tests of human power could make no firm conclusions about the random errors in different types of cycle ergometer.[1] They stated that modified Monark ergometers or Monark models with weight baskets seem to be more reliable than the pendulum models. They also pointed out that some of the best values of reliability in any tests of human performance have been observed with SRM, Kingcycle, and Politecnica cycle ergometers, which are used in conjunction with the athlete's own racing bike. These ergometers must have relatively small random errors, and they may also reduce biological variation by giving cyclists the feel of pace and effort of real competitions. We support the authors' call for more studies aimed at comparing the reliability of performance on different ergometers, preferably with the same participants.

2.1 Systematic Errors in Friction-Braked Ergometers

In calibration studies, pendulum models of the Monark ergometer underestimate true power by ~5% at high workloads (~300W).[3,5,6,8] The error is substantially greater at lower workloads, and the error varies between individual ergometers by typically ±2%.[8] These errors arise largely from frictional losses in the drive train linking cranks to the flywheel. There was also a substantial difference (~1 to 5%) between power measurements when the load was gradually increased relative to when the load was gradually decreased.[8] This hysteresis effect, which must be caused by static friction in the bearings of the pendulum and pulley system of the ergometer, might contribute substantially to random error during a test of a cyclist. Other sources of random error with the pendulum model of the Monark include imprecision in reading the load on the pendulum scale and the need to continually reset the load to adjust for changes in friction as the contact surface between belt and flywheel heats up. These random errors probably account for the lower reliability of the pendulum model of the Monark in tests of performance,[1] and, together with the systematic errors, make the pendulum models of the Monark ergometer unsuitable for tests of competitive cyclists. Models that use a weights basket instead of a pendulum may have less error, but frictional losses in the drive train will always be present. Consequently, we do not agree with the claim that the Monark is a 'gold standard' for cycle ergometers.[9]

2.2 Systematic Errors in Electromagnetically Braked Ergometers

An early study of two Elema ergometers (model unspecified) revealed systematic errors of 6 and 12% at high workloads, although it is unclear whether the ergometers were reading too high or too low.[3] The performance of one of the ergometers in constant-power mode at different pedalling rates was also grossly unsatisfactory.

Recently Maxwell et al.[8] performed dynamic calibration of 5 electrically braked ergometers (2 Siemens, 1 Elema-Schonander, 1 Ergoline and 1 Collins; models not specified). At high workloads (250 to 400W for an unspecified time) the average ergometer read slightly high (0.5 to 2.0%), and the worst ergometer read ~4% too low. The authors did not associate specific errors with specific models. Unfortunately, they also chose pedalling speeds of 40 and 60 min^{-1}, which are unrealistic for competitive cyclists, and they did not investigate the accuracy of the ergometers in constant-power mode at different pedalling speeds. These shortcomings apply to two other studies of a single electrically braked ergometer (makes and models unspecified); in one study the ergometer read a little high (by ~2%) at the highest workloads, but immediately after an individual was tested for 15 minutes at low power the ergometer read slightly low (by ~0.5%);[5] in the other study the ergometer read ~3% too high at the highest workloads, and after 1 hour of use at a high workload the error increased to 50%.[2] Such variation in systematic error would contribute substantially to random error in endurance tests.

2.3 Systematic Errors in Air-Braked Ergometers

In a study[8] of five research-grade Repco ergometers, the average reading at a true power of 275W was low by less than 0.5%, although the typical variation (standard deviation) between ergometers was ±2.3%. At the highest power of 1140W the average reading was high by 1.5% and the variation between ergometers was ±1.7%; the pedalling rate at this power was a realistic 150 min^{-1} for competitive sprint cyclists. At pedalling rates of 70 min^{-1}, the average power reading was 3% too low; at 50 min^{-1} the reading was 10% too low, and one ergometer read low by as much as 38%. These authors adjusted the power readings for changes in ambient temperature and pressure, which have considerable effects on air resistance through changes in air density. For example, the power readings increase by 1% for an increase in temperature of 2.7°C or a decrease in pressure of 7.6mm Hg; the readings also increase by 0.1% for each 10% increase in relative humidity.[10]

Researchers have used the Kingcycle in numerous studies, but there are no reports of dynamic calibration for this ergometer. However, there has been a comparison of power recorded simultaneously on a Kingcycle and on either a single 4-gauge or 20-gauge SRM crankset during three types of performance test with competitive cyclists.[11] The SRM power was 0.90 of the Kingcycle power in an incremental test to peak power and in an anaerobic threshold test; in a simulated 16km time trial the factor was 0.92.

Users of the Kingcycle should be aware that the 'run-down calibration' procedure performed before each test is a means of setting the rolling resistance to a standard value;[12] this procedure adjusts the ergometer readings for changes in temperature and barometric pressure for the power at which the run-down is performed (B. Barker, personal communication). The latest version of the computer interface (Mk 3) measures air pressure, temperature and humidity and adjusts the power reading to account for their effects on air density over the full range of power.[13]

2.4 Systematic Errors in Mobile Ergometers

Lawton et al.[14] investigated the systematic errors of 19 4-gauge SRM cranksets at a constant

SCA 001630

pedalling frequency of 100 min^{-1}. They mounted a bike on a stationary wind trainer and changed gears to create 18 loads in the range 50 to 900W. For a given load, the power delivered by the calibration rig was subtracted from the SRM reading; this systematic error was then averaged over the 18 loads for each crankset. The error averaged over the 19 cranksets was 2.5%, and the typical variation between cranksets was ±5.0%. The authors noted that 4 of the 19 cranksets had average errors of ~10%. Even after recalibration, some cranksets had errors of up to 2.5% at some powers. The source of these errors is uncertain, but one possibility is that they arose partly in the calibration rig.

There are no published studies of systematic errors with 20-gauge SRM cranksets using a calibration rig with direct measurement of power delivered to the crankset. However, Jones and Passfield[15] compared the errors in two 20-gauge cranksets and one 4-gauge crankset by driving the cranks from the rear wheel of a bike mounted on a motor-driven treadmill while they measured power with a modified Monark ergometer connected by a second chain to the crankset. They found that all the cranksets read high only by ~1% at high powers. This difference between the cranksets and the Monark probably represents frictional losses in the Monark drive chain, which Jones and Passfield took care to minimise. In the only other study of SRM systematic error, an overall difference of 2.4% in power between one crankset (presumably 4-gauge) and a pendulum Monark powered by a cyclist was also attributed to frictional losses.[9]

No published studies of systematic error in the PowerTap are available. Our own unpublished simultaneous comparison of a single PowerTap, a single 4-gauge SRM crankset and a Kingcycle in 5-minute time trials with 9 competitive cyclists showed that the PowerTap read higher than the crankset by ~8% and lower than the Kingcycle by ~1%.

3. Choosing a Cycling Test

Cycling offers a variety of events ranging in duration from 10 seconds for a track sprint to road races that last over 6 hours. Various types of test are available to measure the power that a cyclist needs to produce in these events:[1]

- Constant work (time trials) – the cyclist completes a set amount of work or a set distance as quickly as possible.
- Constant duration – the cyclist completes as much work or covers the most distance in a set time.
- Constant power – the cyclist maintains a set power to the point of exhaustion.
- Incremental for peak power or oxygen consumption – the cyclist increases power to maximum effort, usually over 8 to 12 minutes.
- Incremental for anaerobic threshold – the cyclist performs a series of constant-power workouts at increasing intensity; the anaerobic threshold is the intensity associated with particular levels of blood lactate or changes in levels of lactate or respiratory variables.
- Critical power – the cyclist performs a series of constant-power, constant-load or constant-duration tests at different intensities; the duration and mean power of each test are combined in a mathematical model to estimate anaerobic work capacity and maximum aerobic power.

The most important consideration when choosing one of these tests is a strong relationship between competitive performance and performance in the test. The usual way to quantify the relationship is to calculate the correlation between performance in a single test and a single competition for a sample of individuals. Comparing correlations from different studies to select the best test can be problematic, because the correlation is sensitive to the inter-individual spread (standard deviation) in performance in the different studies. A better measure of association between competitive and test performance is therefore the typical (random or standard) error of the estimate of competitive performance. We have calculated this statistic for all the studies in which cycling performance was compared between a competition and a test (table I). The small sample sizes in these studies are responsible for considerable uncertainty in the typical errors, but it is reasonable to conclude that peak

SCA 001631

Table I. Typical error of estimate of performance time in a competitive cycling event based on performance in a test. All participants were competitive male cyclists. Studies are sorted by the typical error

Measure[a]	Mean ± SD (n)	Performance test; ergometer[b]	Correlation	Typical error[c] (%)	95% LR (%)	Reference
16.1km power[d]	311 ± 41 (16)	Peak power; SRM (+ Kingcycle)	0.99	0.7[e]	0.5-1.2	Balmer et al.[11]
40km time	61 ± 3 (7)	Simulated 40km time; Kingcycle	0.98	1.0	0.6-2.4	Palmer et al.[12]
40km time	58.0 ± 1.3 (14)	$\dot{V}O_2$ at ventilatory threshold; (Lode)	−0.82	1.3	0.9-2.1	Hoogeveen et al.[16]
40km time	58.0 ± 1.3 (14)	$\dot{V}O_{2max}$; (Lode)	−0.71	1.6	1.1-2.6	Hoogeveen et al.[16]
40km time	61.3 ± 2.3 (8)	Power at ventilatory threshold; Mijnhardt KEM3	−0.81	2.2	1.4-4.8	Hopkins & McKenzie[17]
15km time	23.5 ± 1.5 (22)	$\dot{V}O_2$ at ventilatory threshold; (Monark)	−0.93	2.3	1.8-3.4	Miller & Manfredi[18]
40km time	56.3 ± 3.7 (15)	1h mean power; Monark 819	−0.88	3.1	2.3-5.0	Coyle et al.[19]
20km time	37.3 ± 4.2 (19)	Peak power; Lode	−0.91	4.7	3.5-7.0	Hawley & Noakes[20]
16.1km time[d]	22.5 ± 1.2 (16)	Peak power; SRM	−0.46	4.7	3.5-7.5	Balmer et al.[11]

a Measure of power in Watts; measure of performance time in minutes.

b Ergometers shown in parentheses did not contribute directly to the measure of performance.

c Estimated as $100SD\sqrt{(1-\text{correlation}^2)}/\text{mean}$.

d Each participant performed a time trial on a different day.

e Estimated by dividing the typical error of 16.1km power by 2.5, to convert it to an error in performance time.[1]

95% LR = 95% likely range of true value; **SD** = standard deviation; $\dot{V}O_2$ = oxygen uptake.

power measured with SRM cranks in an incremental test and time or mean power in a simulated 40km time-trial on a Kingcycle are currently the two best measures for predicting competitive time-trial performance. The peak-power test also has the advantage of being less stressful and time consuming for the cyclist, although the time needed to attach the SRM cranks makes both tests similar in overall duration.

The small typical error of the estimate in the peak-power test is remarkable, because each cyclist performed the time trial on a different day, sometimes under quite different environmental conditions.[11] We suspect that the typical error estimated from mean power in time trials performed under stable environmental conditions would be even smaller. The typical error of the estimate for the actual performance time is the worst in table I, reflecting the marked effect of wind and possibly temperature on cycling performance.

Prediction of time-trial performance on the basis of a test can be useful for team selection, but longitudinal monitoring of the performance of athletes as individuals or in studies of factors that affect performance is a more frequent and useful application of tests. To determine whether a test is suitable for this application requires a complex analysis of concurrent estimates of the typical (random or standard) error of measurement for performance in the test and the typical variation of performance between competitive events.[21] If the correlation between test and event is high (>0.90), a relatively simple condition for suitability is that the typical error of the estimate should arise only from typical error of measurement of the test and event (combined by taking the square root of the sum of their variances).[21] Until tests are assessed in this way, researchers should assume that the condition is satisfied for most tests. How well the test tracks the event then depends only on the typical error of measurement of the test: the smaller the error, the better the test. Indeed, in a simple test-retest situation, a test can be trusted for tracking small changes in individuals or used for quantifying small changes in controlled trials only if the typical error is less than or equal to the small change (unpublished observations). For example, a test with a typical error of 1% would track a change of 2%, but tracking a change of 0.5% with this test would require multiple pre and post tests to prevent the noise (the typical error) from swamping the signal (the change).

Are any tests good enough to track the smallest worthwhile change in performance? For an elite athlete, a performance enhancement makes a difference to the chance of winning when it is ~0.3 to 0.5 of the typical variation between events.[21] There are no published data on the typical variation of cyclists from event to event, but our own unpublished observations indicate that elite male cyclists have typical variation for time of ~0.5% in the kilo (duration ~60 sec) and probably ~1% for time trials lasting ~1 hour. These values require transformation to equivalent random variation for power to allow comparison with errors in measures of power from performance tests. The transformation is a factor derived from the relationship between speed and power;[1] for most individual races the factor probably has a value of ~2.5. The smallest worthwhile change in power is thus ~(0.3 to 0.5) × (0.5 to 1) × 2.5, or ~0.5 to 1%.

It is clear from a recent review[1] that most measures of cycling performance in laboratory tests have random errors larger than this ~0.5 to 1% change. Some of the noisier measures of performance, including maximum oxygen uptake and ventilatory anaerobic thresholds, are based on respiratory variables. With random errors usually in excess of 2%, these measures are unsuitable for tracking the smallest changes in performance that matter to elite cyclists. Measures of lactate anaerobic threshold have lower errors in some laboratory studies, but our own unpublished experience of lactate thresholds with athletes preparing for competition leaves us in no doubt that these measures are unsuitable for monitoring small changes in performance, even when the threshold is measured as power on a good ergometer.

Random errors of around 2% are a consistent finding with ~1-hour tests requiring self-selection of pace (constant-duration and constant-distance tests) when the tests are performed on a Kingcycle or SRM ergometer. These tests will not track the smallest worthwhile changes in performance of elite riders *without* multiple tests of each cyclist at each time point. The error increases to ~3 to 4% with tests lasting several hours, possibly because cyclists have less experience with appropriate pace for longer tests. These tests are also unlikely to track the smallest worthwhile changes in performance for longer events.

Measures of power with random errors that come close to the smallest worthwhile change in performance include mean power in all-out sprints and peak power in incremental tests. Errors as low as ~1% are possible when these tests are performed on good ergometers. Another promising measure awaiting a confirmatory study is time to exhaustion in a constant-power test, when the time is converted to equivalent mean power in a time trial of similar duration. Anyone using an electromagnetic ergometer to determine time to exhaustion should first ensure that the ergometer works reproducibly in constant-power mode. Finally, simulated time trials in an indoor velodrome are almost certainly as reproducible as any laboratory-based test and may be as good as competitive time trials. Further research on the reliability of these tests is needed.

4. Conclusion

Any commercially available ergometer is likely to have substantial systematic error in its calibration. Dynamic recalibration with a special rig is therefore desirable for comparison of cyclists tested on different ergometers. For monitoring changes in performance of a cyclist, an ergometer should introduce negligible random error in its measurements; in this respect, SRM cranks appear to be the best ergometer. Random error in the cyclist's performance should also be minimised by the choice of an appropriate type of test. Peak power in incremental tests and mean power in all-out sprints have the lowest random error. Measures of performance derived from constant-power tests to exhaustion and from time trials in an indoor velodrome are probably at least as good.

References

1. Hopkins WG, Schabort EJ, Hawley JA. Reliability of power in physical performance tests. Sports Med 2001; 31 (3): 211-234
2. Clark JH, Greenleaf JE. Electronic bicycle ergometer: a simple calibration procedure. J Appl Physiol 1971; 30: 440-2
3. Cumming GR, Alexander WD. The calibration of bicycle ergometers. Can J Appl Physiol Pharmacol 1968; 46: 917-9
4. Russell JC, Dale JB. Dynamic torquemeter calibration of bicycle ergometers. J Appl Physiol 1986; 61: 1217-20
5. Wilmore JH, Constable SH, Stanforth PR, et al. Mechanical and physiological calibration of four cycle ergometers. Med Sci Sports Exerc 1982; 14: 322-5
6. Woods GF, Day L, Withers RT, et al. The dynamic calibration of cycle ergometers. Int J Sports Med 1994; 15: 168-71
7. Hopkins WG. Measures of reliability in sports medicine and science [letter]. Sports Med 2000; 30: 375-81
8. Maxwell BF, Withers RT, Ilsley AH, et al. Dynamic calibration of mechanically, air- and electromagnetically braked cycle ergometers. Eur J Appl Physiol 1998; 78: 346-52
9. Martin JC, Milliken DL, Cobb JE, et al. Validation of a mathematical model for road cycling power. J Appl Biomech 1998; 14: 276-91
10. Finn JP, Maxwell BF, Withers RT. Air-braked cycle ergometers: validity of the correction factor for barometric pressure. Int J Sports Med 2000; 21: 488-91
11. Balmer J, Davison RCR, Bird SR. Peak power predicts performance power during an outdoor 16.1-km cycling time trial. Med Sci Sports Exerc 2000; 32: 1485-90
12. Palmer GS, Dennis SC, Noakes TD, et al. Assessment of the reproducibility of performance testing on an air-braked cycle ergometer. Int J Sports Med 1996; 17: 293-8
13. Kingcycle. Kingcycle tester/trainer [online]. Available from: URL: http://www.portaprompt.co.uk/king [Accessed 2001 Jan 30]
14. Lawton EW, Martin DT, Lee H. Validation of SRM power cranks using dynamic calibration. Fifth IOC World Congress; 1999 Oct 31-Nov 5; Sydney. Sydney: International Olympic Committee, 1999: 199
15. Jones SM, Passfield L. The dynamic calibration of bicycle power measuring cranks. In: Haake SJ, editor. The engineering of sport. Oxford: Blackwell, 1998: 265-74
16. Hoogeveen AR, Schep G, Hoogsteen J. The ventilatory threshold, heart rate, and endurance performance relationships in elite cyclists. Int J Sports Med 1999; 20: 114-7
17. Hopkins SR, McKenzie DC. Laboratory assessment of endurance performance in cyclists. Can J Appl Physiol 1994; 19: 266-74
18. Miller FR, Manfredi TG. Physiological and anthropometrical predictors of 15-kilometer time trial cycling performance time. Res Q Exerc Sport 1987; 58: 250-4
19. Coyle EF, Feltner ME, Kautz SA, et al. Physiological and biomechanical factors associated with elite endurance cycling performance. Med Sci Sports Exerc 1991; 23: 93-107
20. Hawley JA, Noakes TD. Peak power output predicts maximal oxygen uptake and performance time in trained cyclists. Eur J Appl Physiol 1992; 65: 79-83
21. Hopkins WG, Hawley JA, Burke LM. Design and analysis of research on sport performance enhancement. Med Sci Sports Exerc 1999; 31: 472-85

Correspondence and offprints: *Will G. Hopkins*, Department of Physiology, School of Medical Science, University of Otago, Box 913, Dunedin, New Zealand.
E-mail: will.hopkins@otago.ac.nz