

Joanne Moffatt, Phil Scarf*, Louis Passfield, Ian G. McHale and Kui Zhang

To lead or not to lead: analysis of the sprint in track cycling

Abstract: This paper uses a statistical analysis of match sprint outcomes to guide tactical decisions in this highly tactical contest and to provide competitors and coaches with a potential, marginal gain. Logistic regression models are developed to predict the probability of the leading rider winning at different points of the race, based on how the race proceeds up to each point. Key tactics are successfully identified from the models, including how the leading rider might hold the lead and how the following rider might optimize overtaking.

Keywords: logistic regression; strategy; tactics.

***Corresponding author: Phil Scarf**, Salford Business School, University of Salford, Salford, M5 4WT, UK, e-mail: p.a.scarf@salford.ac.uk

Joanne Moffatt and Ian G. McHale: Salford Business School, University of Salford, Salford, M5 4WT, UK

Louis Passfield: School of Sport & Exercise Sciences, University of Kent at Medway Chatham, Kent ME4 4AG, UK

Kui Zhang: Wuhan Polytechnic University, Wuhan 430023, Hubei, China

1 Introduction

In all sports, coaches and players seek competitive advantage through tactical analysis, and quantitative analysis of tactics may provide marginal gains. The sprint (or match sprint) in track cycling is highly tactical – in a balanced contest the decisions made by riders are decisive – and so it is interesting to quantify tactics in this context. In this paper, we aim to do just this, providing insight into winning tactics that might be adopted at different points or stages of the race, using a statistical analysis of race outcomes and characteristics.

The match sprint is a race between two riders over 3 laps of a 250 m banked track (Figure 1). Riders travel in the anti-clockwise direction. The first rider to cross the finish line wins the race. In a major championship the overall contest is organized as a knockout or single elimination tournament with several rounds. At the start of the contest, all riders complete an individual “flying” 200 m test against the clock that decides who qualifies and who competes with whom in the first round, the rider with the

fastest time being paired with the slowest. In early rounds each pairing competes in one race only with the winner proceeding to the next round. From the quarter-final stage, a round comprises 3 races; the first rider to win two races proceeding to the next round (UCI 2009).

The race is highly tactical because broadly speaking riders vie for track position in the first part of the race and sprint for the finish line in the second part. The tactical fascination of the race is its positional asymmetry: the leading rider controls track position; the following rider has better sight of his opponent; the leading rider does not benefit from slipstreaming; the following must ride further and higher on the track to overtake. Riders make decisions about speed and position based upon their opponent’s speed and position, the distance to the finish, and their pre-race tactical plans. Pre-race tactical plans will depend on flying speeds (determined during qualification), how coaches and riders expect opponents to race, and starting position – a specified rider must lead at the start. The effect of slipstreaming is large as overcoming air resistance accounts for up to 40% of the work done by the leading rider (Atkinson et al. 2003). The slipstream is the region of reduced pressure created by a moving object as it travels through air, and slipstreaming is the act of riding in the slipstream of the leading rider. Riders typically cannot sustain an early sprint; in fact, a flat out sprint from the start is a rare occurrence and the rider who initiates such a move will be unsuccessful unless his opponent is sufficiently surprised or is very much weaker.

Research in cycling to date has focused extensively on how physiological (maximal aerobic power, muscle fiber type and lactate threshold) and physical (air resistance and rolling resistance) factors influence cycling performance in terms of speed and power. See for example Craig and Norton (2001) and Atkinson et al. (2003). Optimal pacing strategies have been investigated using such factors (Atkinson, Peacock, and Passfield 2007; de Koning et al. 2011; Thomas et al. 2012). However, to the authors’ knowledge no research has developed probabilistic models to investigate how tactics, such as choice of position and choice of overtaking manoeuvre, affect the probabilities of winning for the riders. Our contribution then is to develop such models and describe their tactical implications.

The layout of the paper is as follows. In the next section, we describe the data collection. Section 3

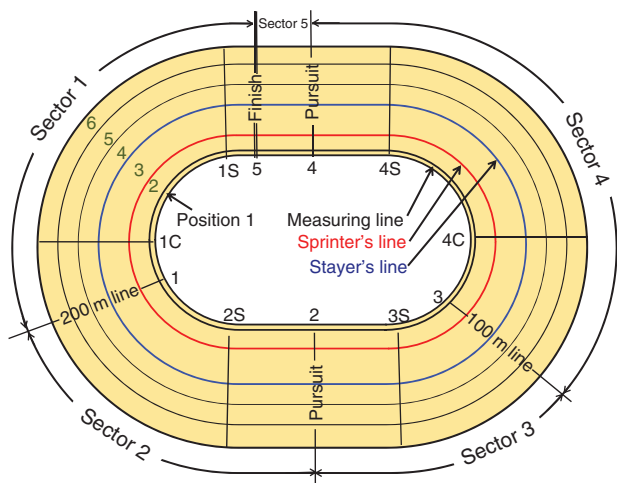


Figure 1 Plan view of a track, showing the track division for determining speed and position and describing covariates and tactics: longitudinal marks (1–5) divide the track into five sectors; latitudinal lines divide the track into six positions. The position of each virtual mark (1S, 1C, 2S, 2C, 3S, 3C, 4S, 4C) is also shown.

discusses the statistical modeling. In Section 4, our results are discussed and interpreted to identify tactics that influence race outcome probabilities. We finish with a summary of our findings and suggestions for further research.

2 Data collection

Video recordings of 231 races at 4 UCI Track Cycling World Cup competitions held at Manchester (February 2007), Sydney (November 2007), Beijing (December 2007) and Copenhagen (February 2008) were used to calculate speed and positional data as described below. The 200 m flying times of riders were obtained from the qualifying round of each competition (Tissot 2009). For the 231 races, 69% of race winners had the shorter flying time, making

this variable a good predictor of race outcome. For our analysis the flying times were converted into flying speeds (average speed over 200 m).

2.1 Calculating sector times and track positions

In order to describe our calculations, a brief description of the track is necessary. Although no two tracks are the same, modern tracks built to Olympic standard have a wooden surface that is banked, the track height being low and horizontal on the inside and high on the outside. The gradient changes around the track, shallow on the straight and steep on the curve. Five longitudinal lines or marks are drawn from the inner edge to the outer edge of the track (see Figure 1). These are the start/finish line, two lines 200 m and 100 m from the finish, and the pursuit lines, half way along each straight. The position of the start/finish line varies between velodromes so that distances between each of the 5 marks vary. There are three latitudinal lines drawn around the track, namely the stayer’s line, the sprinter’s line and the measuring line. The track is 250 m at the measuring line.

We use the five track marks to define five sectors, and the three latitudinal lines plus three others superimposed on the track during the video analysis to define six positions. Sector distances along the measuring line were obtained for each track. Times taken to complete each sector were determined from the video footage using the Dartfish software, providing a total of 15 split times, to an accuracy of 1/50th of a second, over 3 laps. Track position was collected at each track mark and at virtual marks defining the beginnings, ends and apexes of each curve, giving a total of eleven track positions per lap, 33 per rider per race. Figure 2 shows the distribution of position by

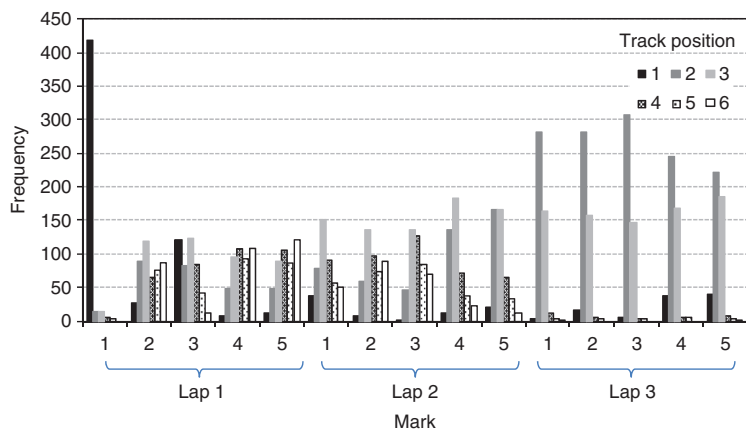


Figure 2 Distribution of track positions of riders at each mark (231 races).

mark and lap: we can see for example that with 700 m to go (mark 1, lap 1) riders are nearly always in track position 1 and with 250 m to go (start of last lap) riders are most frequently in track positions 2 and 3 (between the measuring line and the stayer's line).

2.2 Calculating sector speeds

Calculating the sector speed is not straightforward because riders' paths are three-dimensional. Only when riders travel along the measuring line is sector speed the ratio of sector distance to sector time taken. To provide a better estimate of sector speed, we used a topographical model of a velodrome that defines the height of the track $z=f(x, y)$ over a fine grid. Such a model was only available to us for the Manchester velodrome; for other tracks we use the same function modified slightly to account for differing lengths of the straights and curves. The x -axis is parallel to the straight, the y -axis perpendicular. For a small change in rider coordinate (x, y, z) to $(x+dx, y+dy, z+dz)$ the path length is $dP=\sqrt{dx^2+dy^2+dz^2}$. From the video analysis, positions are recorded at the marks and the virtual marks (curve starts, finishes, and apex). Riders can take one of many paths between two known points; the path they take will depend on factors such as speed, position of opponent and how easy it is to ride the path. It is necessary to make an assumption about the path taken, and for convenience we assume riders take the path with constant rate of change of height. This will not necessarily be the path of least distance and therefore the path of least work, but it will be close to it. Then on the straight the path length is

$$P(p_0, p_1) = \int_{x_0}^{x_1} \sqrt{1 + \left(\frac{dy}{dx}\right)^2 + \left(\frac{dz}{dx}\right)^2} dx$$

where $p_0=(x_0, y_0, z_0)$ and $p_1=(x_1, y_1, z_1)$ are known points (x_0 and x_1 are recorded and (y_0, z_0) and (y_1, z_1) are calculated from the track model $z=f(x, y)$), $dz/dx=(z_1-z_0)/(x_1-x_0)$, and $dy/dx=(dy/dz) \cdot (dz/dx)$ where $dz/dy=\gamma(x)$ is the known track gradient. On the curve we formulate the path length in terms of the angle θ around the curve:

$$P(p_0, p_1) = \int_{\theta_0}^{\theta_1} \sqrt{\left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2 + \left(\frac{dz}{d\theta}\right)^2} d\theta$$

where θ_0 and θ_1 are the angles around the curve at p_0 and p_1 . We then proceed similarly to above, assuming the curve is circular at the base of the track ($z=0$) and $dz/d\theta \approx (z_1-z_0)/(\theta_1-\theta_0)$. The other derivatives can be calculated from the known track model, $z=f(x, y)$. Numerical integration is used to evaluate the path lengths. To determine the distance traveled in a sector, the lengths of the paths in the sector are added. The sector speed is then the ratio of sector path length to sector time taken. Figure 3 shows the distribution of sector speeds of riders for each sector. We can see that sector speeds broadly increase through a race, with the variability decreasing. In the final sector [between the pursuit line and the finish line on the final lap (~15 m)], riders would appear to slow as they tire.

2.3 Calculations for other covariates

Other covariates included in the analysis are described in Table 1; these have been calculated using the information described above. For example, distance between riders at each mark proved difficult to determine directly from video analysis, so it was estimated using the time difference between the riders crossing each mark and the sector speed for the leading rider in

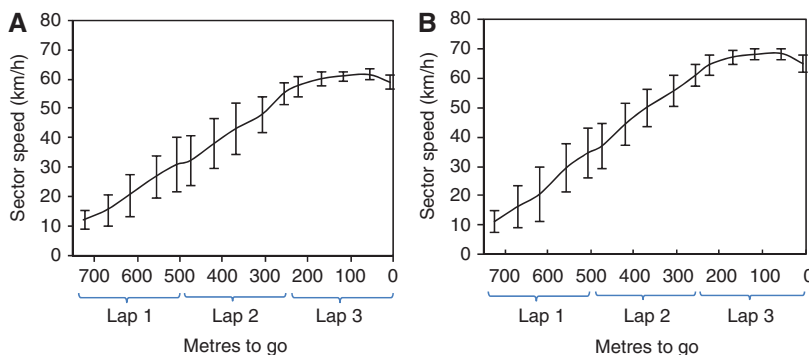


Figure 3 The mean speed of riders in each sector over all races for (A) female and (B) male riders (231 races). The error bars are drawn at ± 1 sample standard deviation.

Table 1 Description of covariates considered for analysis.

Covariate	Description
RF	The ratio of the leading rider's flying speed to the following rider's flying speed
$S(l, s)$	The ratio of the sector speed in sector s , lap l of the leading rider to the sector speed of the average rider in sector s , lap l (by gender and event)
$RS(l, s)$	The ratio of the sector speed in sector s , lap l of the leading rider to the sector speed of the following rider in sector s , lap l
$A(l, s)$	The ratio of the sector speed in sector s , lap l of the leading rider to his (or her) sector speed in previous sector ($s-1$)
$T(l, s)$	The ratio of the average time taken for a rider to complete sector s in lap l (by gender and event) to the time taken for leading rider to complete sector s in lap l
$RT(l, s)$	The ratio of the time taken for following rider to complete sector s in lap l to the time taken for leading rider to complete sector s in lap l
$TA(l, s)$	$T(l, s)$ relative to $T(l, s-1)$
$D(l, m)$	Distance between riders at mark m , lap l
$LC(l, m)^*$	Lead has changed in the race between mark m in lap l and current mark
$PL(l, m)^{**}, PF(l, m)^{**}$	Track positions for leading and following riders respectively at mark m in lap l
$DL(l, m)^*, DF(l, m)^*$	Leading and following riders have decreased track position between mark m in lap l and the current mark respectively
$FH(m)^*$	Following rider is in a higher or in the same track position as the leading rider at mark m in lap l
$FL(m)^*$	Following rider is in a lower or in the same track position as the leading rider at mark m in lap l

NB: () denotes covariate is a function of lap l and either sector s or mark m .

*is a binary variable taking values 0 if no and 1 if yes, ** is a categorical variable.

the preceding sector. Also sectors 2 and 3, and sectors 4 and 5 were merged together when calculating variables relating to the speeds and durations (time taken). Sectors 4 and 5 were merged together because sector 5 was relatively short; sectors 2 and 3 were also merged to make sector lengths consistent across all velodromes; for these variables the merged sectors were either 50 m or 100 m long.

3 Methodology

Our analysis investigates tactics that influence the outcome of a race, classified as either win or lose. In the individual sprint there are potentially a large (or unlimited) number of tactics and states. Therefore a natural approach is to consider the influence on race outcome of variables that summarize the race state and development. Logistic regression has the advantage of being straightforward to implement and readily allows many relationships between the outcome and explanatory variables to be identified and interpreted. It is then a matter of interpreting such relationships in order to determine winning tactics. Logistic regression has been used by many authors to predict match outcomes in many sports: e.g., in basketball (Schwertman, Schenk, and Holbrook 1996), marathon running (Yeung, Yeung, and Wong 2001), cricket

(Allsopp and Clarke 2004; Scarf and Shi 2005; Scarf and Akhtar 2011; Scarf, Shi and Akhtar 2011; McHale and Asif 2013), tennis (Magnus and Klaassen 2008), football (e.g., Leitner, Zeileis and Hornik 2011), golf (e.g., Fearing, Acimovic and Graves 2011), and hockey (Bedford and Baglin 2009; Gramacy, Jensen and Taddy 2013).

Models are fitted at each of 8 intermediate points or stages, namely 700 m, 600 m, 500 m, 450 m, 350 m, 250 m, 200 m and 100 m to go (to the finish). The explanatory variables used are detailed in Table 1 and describe the actions riders can apply, e.g., increase track position. These actions are then interpreted in order to indicate potential winning tactics, with the aim of identifying the following:

- How the leading rider can defend his or her position.
- How the following rider can optimize his or her chance of overtaking.
- When it is best to lead the race and when it is best to follow.
- Which tactics are best for a faster rider (by flying speed) and which are best for a slower rider.

Stages are indexed by $i, i=1, \dots, 8$. The leading rider at point i is considered as the reference rider. The model determines the probability p_{ij} that the leading rider wins the j th race given the race state at point i . The race state at point i is represented by selected explanatory variables, \mathbf{X}_{ij} , that describe the actions of riders up to point i , together with riders' flying speeds. Conditional upon \mathbf{X}_{ij} ,

the outcome of race j at point i , Y_{ij} , is assumed to follow the Bernoulli distribution, $B(p_{ij})$, with $\log\{p_{ij}/(1-p_{ij})\}=\beta^T X_{ij}$ where β^T is a vector of parameters. This model is fitted by maximum likelihood.

At each intermediate point i , to determine the best fitting model, we used a forward step-wise procedure for variable selection, with influential variables and two-way interactions being entered into the model. Thus at each intermediate point i , a “best” model is chosen using the forward stepwise procedure. In the following section results are presented for the “best” model at each intermediate point, although for brevity the qualification “best” is dropped. Three-way interactions were also considered but none proved to be influential. The Akaike Information Criterion (AIC) (Dobson and Barnett 2008) was used in Moffatt et al. (2009) as a selection criterion, but this method allowed too many terms to enter into the model, making interpretation difficult. A finite sample-size correction factor for AIC might be used as the resulting criterion then penalizes additional parameters more strongly. The Bayesian Information Criterion (BIC) also penalizes the number of model parameters more heavily than AIC. However, using BIC only a few terms were selected in the models. Therefore stepwise regression using the likelihood ratio chi-square test (accepting variables with a p -value of <0.05) was used; this allowed an intermediate number of terms to be added.

To restrict the number of variables and 2-way interactions being tested, only flying speed and variables close (within the previous 200 m) to the current point of the race were considered for each model. Variables relating to early parts of the race may interact with variables relating to later parts of the race, making it difficult to interpret the results and identify key tactics.

4 Model results and tactical implications

The overall accuracies of the models are first discussed. Then the overall effect of influential variables is compared at the different points of the race. Finally, influential actions identified by each of the models are discussed in terms of tactics, in race order, taking the overall model accuracies into account.

In order to assess the model accuracy, the probability that the leading rider wins a race was calculated for each intermediate point or stage for all 231 races. If the model predicted the leading rider had over a 50% chance of winning the race he was classified as winning the race, otherwise he was classified as losing the race. The model predictions were compared to actual outcome and Figure 4A shows the percentage of races for each of the 8 models correctly classified. At 700 m and 600 m to go 69% of races were correctly classified. The only variable found to be influential was the ratio of flying speed of riders, and as mentioned in Section 2, 69% of faster riders (by flying speed) win. See Table 2 for the parameter estimates for both these models. Thus, at 600 m to go, covariates from which tactics might be inferred did not influence race outcome, and so tactics in the first 150 m do not appear, on the basis of the model, to be important.

From 600 m to go covariates and hence race tactics become increasingly important. The difference between the percentage of races correctly classified and the percentage of faster riders (by flying speed) who go on to win indicates the percentage of race winners being accounted for by race tactics. This is 8% at 500 m, 450 m and 350 m to go, 13% at 250 m to go, 15% at 200 m to go and 23% at 100 m to go, see Figure 4.

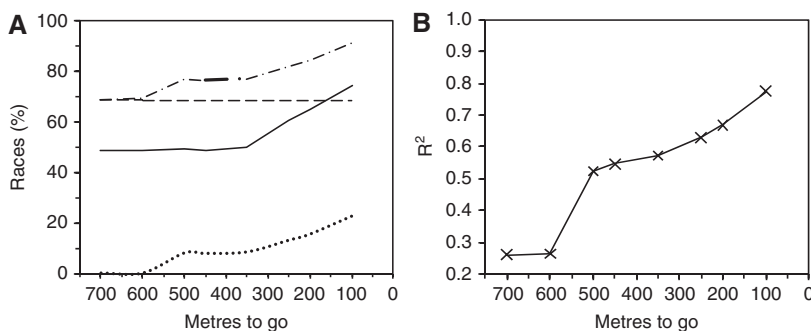


Figure 4 (A) Comparison of percentage of races predicted correctly by the model (---), percentage of leading riders who go on to win (—), and percentage of faster riders (by flying speed) who win the race (- - -), the percentage of race winners being accounted for by race tactics (percentage of races predicted correctly by the model minus the percentage of leading riders who go on to win) (.....); (B) Nagelkerke R^2 for each model.

Table 2 Parameter estimates at 700 m (mark 1, lap 1) and 600 m (mark 3, lap 1) to go.

Model at 700 m to go ⁽¹⁾		Model at 600 m to go	
Variable	Coefficient	Variable	Coefficient
Intercept	-66.47 (11.1)***	Intercept	66.08 (10.88)***
RF	66.54 (11.12)***	RF	66.1 (10.89)***

NB: See Table 1 for covariate definitions, ***significant at 1% level.

⁽¹⁾ 5 observations removed due to missing values.

The models perform considerably better than the simple assumption that the leading rider wins, even towards the end of the race. For the first 2 laps who is leading does not affect the chances of winning, as there is around a 50/50 chance for each rider. In the last lap being in front increases a rider's chance of winning; at 100 m to go only 25% of those behind go on to win (see Figure 4A).

The models fitted at each stage were also compared using Nagelkerke R^2 , a modified form of Cox and Snell pseudo R^2 (Nagelkerke 1991) that is, broadly speaking, the proportion of variability in race outcomes that is explained by the variables in the model. As expected, the R^2 value is higher for models relating to stages closer to the finish (see Figure 4A).

Each model, starting with the 450 m to go model, is discussed in turn in the following sections. We have omitted the discussion for the 500 m to go model as the covariate effects are complex and their explanatory power are quite low. However the covariates for the 500 m to go model are shown in Table 3. Rather than repeatedly use he/she and his/her throughout, we will without prejudice consider

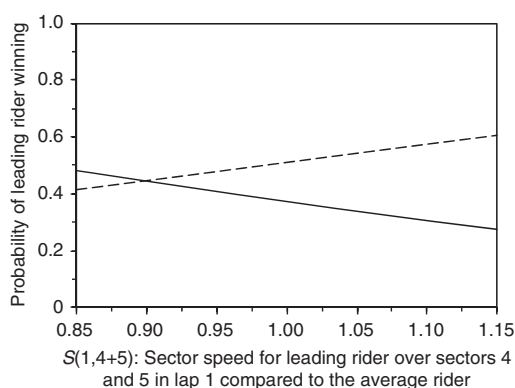


Figure 5 Win probability for leading rider at 450 m to go (mark 1, lap 2) as a function of sector speed over sectors 4 and 5 on lap 1 for the leading rider compared to the average rider ($S(1,4+5)$): following rider in a lower or in the same track position as the leading rider, $FL(1,3)=1$, (—), following rider in a higher track position, $FL(1,3)=0$, (---). The leading rider is in position 3 (modal position) at mark 1, lap 2, and all other model variables were set to average values if continuous and 0 if binary, bar $FH(1,4S)$ which was set to 1.

riders as male. It should be noted that in the remainder of this paper although we will say a rider should carry out certain actions/tactics in order to increase his chance of winning, strictly we will mean that our interpretation of fitted models imply that such actions increase the win probability for the rider.

4.1 Model at 450 metres to go

At 450 m to go, and just prior to this, the actions riders applied in this part of the race as identified by the model only have a small influence (8%) on race outcome, see Figure 4A. Here riders are on the first curve on the second lap, and are typically traveling slowly (see Figure 3).

The model suggests the leading rider should increase speed in sectors 4 and 5 on lap 1 if his opponent is in an attacking position (hence higher track position) at mark 3 (on the 2nd bend in lap 1), see Figure 5 where $(FL(1,3)=0)$. However, the rider should not be too far ahead (<3 m front wheel to front wheel), keeping the follower just on his shoulder [see Table 3, the interaction $D(2,1) \times FH(1,4S)$ is negative]. Being far ahead may indicate that too much energy has been expended or that the following rider has the opportunity to accelerate and overtake quickly. A slow speed (~90% of average speed, or 25 and 26 km/h for women and men respectively) should be adopted to save energy if the opponent is not in an attacking position at mark 3 [see Figure 5 where $(FL(1,3)=1)$].

The best position for the leading rider to be in at mark 1, lap 2 is position 2, as this is where the track gradient is at its highest; being in position 1 allows the opponent to overtake relatively easily. Conversely being in position 5 or 6 is also advantageous; this could be to stop an overtaking move [see Table 3 $PL(2,1)$].

4.2 Model at 350 metres to go

At 350 m to go (on the second bend on the second lap) most riders are trying to get in a good position ready to accelerate to maximum speed within the next 100 m. The 350 m to go model was the only model to contain variables relating to the time taken for riders to complete sectors. These variables had greater predictive power than speed variables. Note, as per the discussion in Section 2.2, a sector time alone does not imply a speed, because riders change track position and may do so to a great extent in this sector (between 450 and 350 m to go). Similarly to the case at 450 m to go, 8% of race winners are accounted for by race tactics. The model indicates that the leading rider

Table 3 Parameter estimates at 500 m (mark 5, lap 1), 450 m (mark 1, lap 2) and 350 m (mark 3, lap 2) to go obtained by stepwise selection.

Model at 500 m to go ⁽¹⁾		Model at 450 m to go ⁽²⁾		Model at 350 m to go	
Covariate	Coefficient	Covariate	Coefficient	Covariate	Coefficient
Intercept	-61.69 (12.93)***	Intercept	-71.45 (17.71)***	Intercept	-8.79 (19.70)
RF	62.89 (13.14)***	RF	56.86 (16.36)***	RF	41.73 (15.83)***
D(1,3)	0.59 (0.16)***	D(2,1)	3.33 (1.43)**	TA(2,2+3)	3.13 (1.41)**
FH(1, 4S)	-3.04 (1.10)***	FH(1,4S)	11.51 (4.84)**	FH(2,2S)	-36.25 (10.94)***
IF(1,4S)	1.99 (0.67)***	IF(1,4C)	-0.90 (0.57)	IF(2,2S)	1.05 (0.41)***
FL(1,5)	-63.38 (31.35)**	FL(1,3)	4.99 (1.83)***	IF(2,2)	-1.56 (0.51)***
LC(1,3)	-6.46 (2.99)**	IL(1,4C)	-2.56 (0.92)***	LC(2,1)	-18.10 (8.56)**
D(1,3)×IF(1,4S)	-0.61 (0.16)***	PL(2,1,1)		PL(2,1,1)	
		PL(2,1,2)	2.17 (0.76)***	PL(2,1,2)	1.79 (0.72)***
D(1,3)×LC(1,3)	-2.14 (0.81)***	PL(2,1,3)	1.05 (0.73)	PL(2,1,3)	0.05 (0.71)
		PL(2,1,4)	1.57 (0.89)*	PL(2,1,4)	0.18 (0.85)
FL(1,5)×RF	62.54 (31.33)**	PL(2,1,5)	2.30 (0.92)***	PL(2,1,5)	0.40 (0.87)
		PL(2,1,6)	2.16 (1.22)*	PL(2,1,6)	-0.39 (1.06)
		S(1,4+5)	2.56 (1.60)	TA(2,1)	-36.59 (11.08)***
		DL(1,4)	-73.33 (27.38)***	IL(2,1)	-81.08 (30.29)***
		D(2,1)×FH(1,4S)	-3.45 (1.44)**	TA(2,1)×FH(2,2S)	36.44 (11.01)***
		FL(1,3)×S(1,4+5)	-5.54 (1.81)***	TA(2.1)×TC(2,2S)	19.64 (8.94)**
		DL(1,4)×RF	74.27 (27.42)***	RF×IL(2,1)	80.91 (30.30)***
		IF(1,4C)×IL(1,4C)	2.65 (1.06)***		

NB: See Table 1 for covariate definitions.

PL(2,1) is a categorical variable, with PL(2,1,1) corresponding to the leading rider at mark 1, lap 2 being in track position 1.

Marks denoted by C occur at the curve apexes, and by S at the beginning and end of the straights e.g., 4S for beginning of the straight before mark 4 in sector 4.

*, **, ***significant at 10%, 5% and 1% level respectively.

⁽¹⁾ 2 observations were removed due to missing values.

⁽²⁾ 3 observations were removed due to missing values.

increases his win probability by slowing the race down in sector 1, lap 2, then increasing speed in sectors 2 and 3, on lap 2 [see Table 3 variables TA(2,1) and TA(2,2+3) which are indicative of acceleration]. The leading rider is thereby forcing his opponent to either slow down or overtake (or undertake). If the leading rider is a faster rider (by flying speed) he can further increase his win probability by moving to a higher track position between marks 1 and 3 on lap 2 [see Table 3 RF, DL(2,1) and RF×DL(2,1)]. A slower rider should not increase track position and should conserve energy for the sprint later on.

The following rider increases his chance of winning by using the track gradient to his advantage to save energy or increase speed. This is shown by Table 4 when IF(2,2S)=0 and IF(2,2)=1, which corresponds to a rider moving to a lower track position between marks 2S and 2 on lap 2 then moving to a higher track position to mark 3.

A successful overtaking (or undertaking) tactic involves first slowing the race down in sector 1, lap 2, and then attaining an attacking position (opponent in a lower track position) at the beginning of the straight, in sector 2 on lap 2. Then an opponent is surprised by a high acceleration to overtake (or undertake), see Figure 6

where FH(2,2S)=0. A different successful overtaking (or undertaking) tactic is for the rider to increase speed in sector 1 (on lap 2) while being in a lower or in the same track position as the opponent at the beginning of the straight in sector 2, lap 2, see Figure 6 where FH(2,2S)=1. Being in a higher track position requires taking a longer route around the bend, and hence considerably more energy to overtake, especially with the race increasing in speed.

Table 4 Win probability for following rider at 350 m to go (mark 3, lap 2) as a function of whether the following rider moves to a higher track position between marks 2S and 3 on lap 2 (IF(2, 2S)), and moves to a higher track position between marks 2 and 3 on lap 2 (IF(2, 2)). We suppose the leading rider is in position 3 (modal position) at mark 1 on lap 2 and all other variables were set to average values if continuous and 0 if binary, bar FH(2, 2S) which was set to 1.

IF(2,2S)	IF(2,2)	
	0	1
0	0.57	0.86
1	0.32	0.69

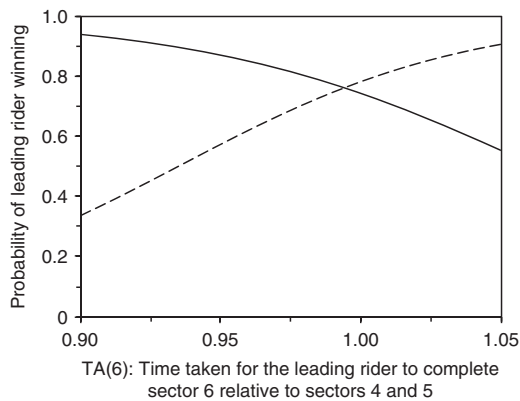


Figure 6 Win probability for the leading rider who has overtaken between marks 1 and 3 on lap 2 as a function of time to complete sector 1 on lap 2 relative to sectors 4 and 5 on lap 1 for the leading rider, as at 350 m to go, $TA(2,1)$: following rider at mark 3 (350 m to go) was in a higher or in the same track position as the leading rider at mark 2S in lap 2, $FH(2,2S)=1$, (---); following rider was in a lower track position, $FH(2,2S)=0$, (—). We set the leading rider at 350 m to go in position 3 (modal position) at mark 1 on lap 2 and all other variables were set to average values if continuous and 0 if binary.

4.3 Model at 250 metres to go

At 250 m to go (finish line with one lap to go) riders are now typically at full speed or making an attacking move to accelerate to maximum speed. Being the leading rider at this point is more advantageous with 61% of leading riders going on to win the race (see Figure 4A). At this stage, 13% of race outcomes are accounted for by tactics-related covariates (see Figure 4A). As detailed in Section 4, 69% of race outcomes are accounted for by flying speed (a measure of rider’s physical ability); the remaining 18% of race outcomes is unexplained.

The model (Table 5) indicates the leading rider should defend his lead by increasing speed and position between 350 m and 250 m to go (sectors 4 and 5, lap 2) with the most important part of this tactic being a move to a higher track position between marks 4 and 5 on lap 2, see Figure 7. The leading rider can also defend his lead and also increase his chance further by being in position 4 at mark 4 on lap 2 [see Table 5, where $PL(2,4)=4$]. Being in a higher track

Table 5 Fitted parameter estimates for the model at 250 m (mark 5, lap 2), 200 m (mark 1, lap 3) and 100 m to go (mark 3, lap 3) obtained by stepwise selection.

Model at 250 m to go ⁽¹⁾		Model at 200 m to go ⁽²⁾		Model at 100 m to go ⁽³⁾	
Covariate	Coefficient	Covariate	Coefficient	Covariate	Coefficient
Intercept	-130.26 (21.19)***	Intercept	-122.27 (20.41)***	Intercept	712.08 (228.64)***
RF	92.56 (15.72)***	RF	111.43 (18.80)***	RF	100.49 (22.80)***
RS(2,2+3)	19.53 (6.95)***	FH(3,1S)	-3.65 (1.33)***	A(3,2+3)	-791.49 (227.29)***
RS(2,4+5)	18.58 (9.58)**	RS(3,1)	13.92 (7.48)*	S(3,2+3)	-855.85 (244.58)***
D(2,3)	4.56 (1.60)***	D(3,1)	0.08 (0.26)	D(3,3)	4.62 (0.87)***
PL(2,4,1+2)		PL(2,4,1+2)		D(3,2)	-2.69 (0.59)***
PL(2,4,3)	-0.24 (0.43)	PL(2,4,3)	0.16 (0.48)	PF(3,1)	0.48 (0.60)
PL(2,4,4)	3.22 (1.01)***	PL(2,4,4)	3.33 (1.08)***	IF(3,2)	2.88 (1.00)**
PL(2,4,5+6)	-2.32 (1.08)**	PL(2,4,5+6)	-1.11 (0.93)		
DF(2,3)	0.29 (0.54)	DF(2,3)	-1.85 (1.18)	S(3,2+3)×A(3,2+3)	833.15 (235.02)***
IL(2,4C)	32.97 (20.03)*	IL(2,4S)	-2.95 (0.98)***		
FL(2,4)	-4.14 (1.20)***	IL(2,5)	2.31 (0.83)***	PF(3,1)×IF(3,2)	-4.29 (1.41)***
FL(2,4S)	-0.15 (0.55)	PF(3,1S,1+2)			
IL(2,4C)×RS(2,2+3)	-103.25 (27.01)***	PF(3,1S,3)	3.08 (1.14)***		
D(2,3)×RS(2,2+3)	-4.48 (1.60)***	PF(3,1S,4)	2.58 (1.56)*		
IL(2,4C)×RS(2,4+5)	73.77 (28.36)***	PF(3,1S,5+6)	8.40 (3.62)**		
FL(2,4)×DF(2,3)	2.89 (0.96)***	D(3,1)×PF(3,1S,1+2)			
FL(2,4)×FL(2,4S)	2.36 (0.98)**	D(3,1)×PF(3,1S,3)	-1.31 (0.36)***		
		D(3,1)×PF(3,1S,4)	-1.05(0.42)***		
		D(3,1)×PF(3,1S,5+6)	-2.74 (0.87)***		
		D(3,1)×DF(2,3)	1.18 (0.38)***		

NB: See Table 1 for covariate definitions.

$PL(2,4)$ is a categorical variable, with $PL(2,4,1+2)$ corresponding to the leading rider at mark 4 in lap 2 being in track positions 1 or 2 etc.

Marks denoted by C occur at curve apexes, and by S at the beginning and end of the straights.

e.g., 4S for beginning of the straight before mark 4 in sector 4.

*, **, *** significant at 10%, 5% and 1% level respectively.

⁽¹⁾ 5 observations, ⁽²⁾ 4 observations were removed due to missing values.

⁽³⁾ 4 observations were removed with 3 due to missing values and a further 1 for being an extreme outlier.

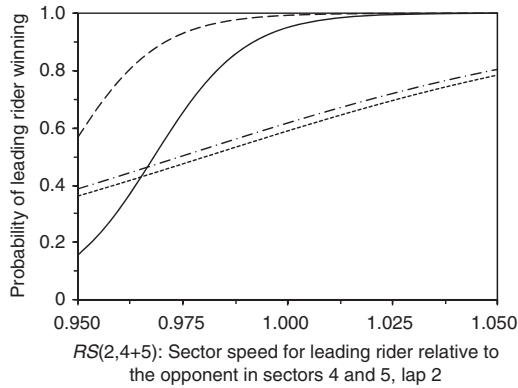


Figure 7 Win probability for leading rider at 250 m to go (mark 5, lap 2) as a function of sector speed for the leading rider relative to that of the following rider in sectors 4 and 5, lap 2, ($RS(2,4+5)$). The graphs correspond to $RS(2,2+3)$ (the sector speed for the leading rider relative to following rider in sectors 2 and 3, lap 2) and $IL(2,4C)$ (the leading rider moves to a higher track position between mark 4C and 5 on lap 2): Taking the following values: $RS(2,2+3)=0.99$ and $IL(2,4C)=0$ (.....), $RS(2,2+3)=1.01$ and $IL(2,4C)=0$ (- · - · -), $RS(2,2+3)=0.99$ and $IL(2,4C)=1$ (- - -), $RS(2,2+3)=1.01$ and $IL(2,4C)=1$ (—). We set leading rider in position 3 (modal position) at mark 4 on lap 2, and all other model variables were set to average values if continuous and 0 if binary.

position here also delays riders in reaching maximum speed. The following rider therefore will spend less time benefiting from the slipstream of the leading rider making it more difficult for the rider to overtake to win.

The following rider can increase his win probability if he does not move to a lower track position between marks 3 and 5 on lap 2 [see Table 5, $DF(2,3)$ is positive] and if he is in the same or in a lower track position than the leading rider at mark 4 on lap 2 [see Table 5: $FL(2,4)$ is highly negative]. If both riders are in a low track position (positions 1–3), the race is likely to be already at maximum (or close to maximum) speed, therefore the following rider can then benefit from the opponents slipstream, saving energy to overtake later. If both riders are in a high track position (4–6), and if the following rider is in a higher track position than the opponent they would be expending more energy as the track gradient is higher. The following rider reduces his win probability if he moves to a lower track position; this is likely to be a response to the leading rider suddenly moving to a lower track position to accelerate to full speed, to gain a lead.

4.4 Model at 200 metres to go

Generally riders at 200 m to go (on first curve on third lap) are at or descending to the bottom of the track and are traveling close to maximum speed (92±5% of their flying

speed). At this point in the race 65% of leading riders go on to win and 15% of race outcomes are accounted for by tactics-related covariates, noting that the model predicts 84% of races correctly with 69% of faster riders (by flying speed) winning – see Figure 4A; 16% of race outcomes remain unexplained.

The model indicates that it is optimal for the leading rider to be in position 4 at mark 4 on lap 2 [see Table 5 $PL(2,4,4)$ is positive]. A rider should also be in a higher track position than his opponent at the end of the straight in sector 1, lap 3 [see Table 5, $FH(3,1S)$ is negative]. These two actions make overtaking more difficult for the following rider. The leading rider can also defend his position by first moving to a lower track position to mark 5 on lap 2 before moving to a higher track position around the bend, where the track gradient is high. This corresponds to $IL(2,4S)=0$ and $IL(2,5)=1$, see Table 5 where $IL(2,4S)<0$ and $IL(2,5)>0$. The following rider should be close to the opponent at mark 1 on lap 3 when in a low track position (1 or 2), see Figure 8A and B where $PF(3,1S,1+2)$, to benefit from the slipstream of the opponent. If the following rider does not move to a lower track position and remains in the higher track position (typically 3 as this would mean the rider will then be in a high position at mark 1 on lap 3, when most races are at maximum speed) the rider will benefit if he is further back from the leading rider [as shown in Figure 8A, where $PF(3,1S,3)$]. The race is likely to be slower in this case and it is more difficult for the leading rider to judge what move the following rider will attempt. When the following rider is in an high position (5 or 6) at the end of the straight in sector 1, lap 3, and he then moves to a lower track position [see Figure 8B, $PF(3,1S,5+6)$], the rider also gains an advantage from being far behind.

4.5 Model at 100 metres to go

Riders are now on the last bend (mark 3, lap 3) with the following rider making his last attempt to overtake; all riders are now at full speed. As generally riders will be positions 1–3 (see Figure 2) and do not change track position significantly, intermediate marks (e.g., Mark 2) were not considered in the modeling process. At this point in the race still only 75% of leading riders go on to win and tactics-related covariates account for 23% of race outcomes (see Figure 4). Only 8% of race outcomes are unexplained by this point in the race.

The model (see Table 5) now indicates winning factors rather than tactics over which riders have some control. Thus the leading rider should be as far ahead as possible at 100 m to go (mark 3, lap 3), he should also be increasing

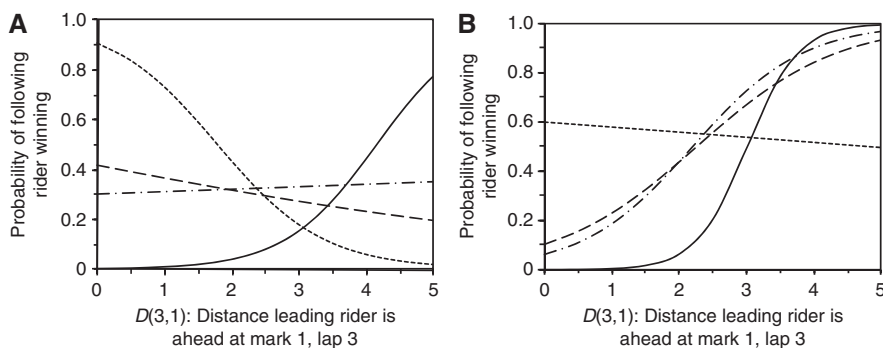


Figure 8 Win probability for following rider at 200 m to go (mark 1, lap 3) as a function of distance behind at mark 1 on lap 3 ($D(3,1)$), and track position at mark 1S, lap 3, for the following rider: $PF(3,1S,1+2)$ (.....), $PF(3,1S,3)$ (- · -), $PF(3,1S,4)$ (- - -), $PF(3,1S,5+6)$ (—); (A) following rider has not moved to a lower track position between marks 3 on lap 2 and mark 1 on lap 3 ($DF(2,3)=0$), and (B) following rider has moved to a lower track position between marks 3 on lap 2 and mark 1 on lap 3 ($DF(2,3)=1$). We set the leading rider in position 3 (modal position) at mark 4, lap 2, the following rider in a higher or the same track position at mark 1S on lap 3 and all other model variables were set to average values if continuous and 0 if binary.

his or her lead between marks 2 and 3 in lap 3 [see Table 5 $D(3,3)$ is positive and $D(3,2)$ is negative]. Riders should also be traveling at a high speed and should be accelerating over sectors 1–3 on lap 3 [see Table 5 the interaction $A(3,2\&3)\times S(3,2\&3)$ is positive, while the main effects are negative]. If a rider is unable to accelerate he is less likely to win, indicating he has gone too fast too early, and is tiring.

The best tactic for the following rider involves being in a high position (3 or above) on the track on the 1st bend of the last lap (mark 1, lap 3), to use the declining track gradient to accelerate, and then move to a higher track position to mark 3, to attain a good position to overtake. See Figure 9 where $PF(3,1)=1$ and $IF(3,2)=1$.

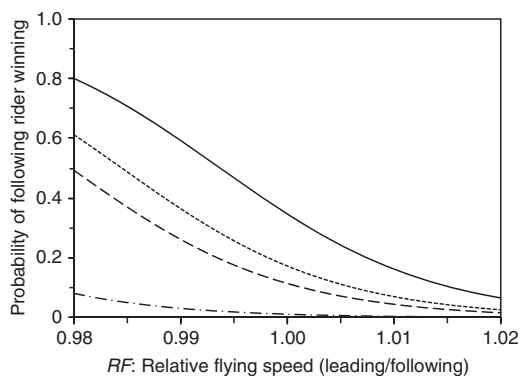


Figure 9 Win probability for the following rider at 100 m to go (mark 3, lap 3) as a function of the ratio of flying speeds (leading rider to the opponent, RF). The graph shows the dependence on the following rider being in a low track position at mark 1 on lap 3 ($PF(3,1)=1$ for 1 or 2, and 0 otherwise) and on the following rider moving to a higher track position between marks 2 and 3 on lap 3 $IF(3,2)$: $PF(3,1)=0$ and $IF(3,2)=0$ (.....), $PF(3,1)=0$ and $IF(3,2)=1$ (- · -), $PF(3,1)=1$ and $IF(3,2)=0$ (- - -), $PF(3,1)=1$ and $IF(3,2)=1$ (—). We set the following rider at 1 m behind at marks 2 and 3 in lap 3 and all other model variables are set to the average.

5 Discussion

We develop statistical models to allow key tactics that influence the win probabilities for riders to be identified. Eight regression models are described for different points of the race. These models consider actions in terms of optimal positions and speeds, allowing tactics to be identified. They show that one clear tactic cannot be defined for a rider throughout the race (unlike for endurance events such as the 4000 m pursuit) because riders cannot sustain maximum speed for the whole race and must react to race circumstances, maintaining a best position given the development of the race. Thus the riders vie for position in the early part of the race before riding flat out in the later part, evidence for which is provided by the position and speed distributions as the race progresses (Figures 2 and 3). The optimum tactics for riders are also interdependent

The models provide insights for answering the key questions asked at the beginning of Section 3:

- *How can a leading rider defend his/her position?* In the early part of the race (450 m to go model) riders should increase speed to stop the opponent from overtaking. However just before riders reach maximum speed (in sector 4, lap 2), riders should also move to a higher track position, forcing their opponent up the track; this action may be about holding a position that allows a rider to accelerate quickly.
- *How can the following rider optimize his or her chances of overtaking?* In the first lap, riders only benefit from overtaking if far behind, and the surprise move catches their opponent off guard. The element of surprise can also be used in the second lap to overtake. However for a fast race, following riders can use the slipstream

of their opponent to conserve energy provided they are close behind and in the same or lower track position on the second lap. The following rider must attain an attacking position regardless of speed by the later part of the race.

- *When it is best to lead the race and when it is best to follow?* Leading the race was found not to be important until the start of the final lap. At a lap to go, 60% of leading riders went on to win.
- *Which tactics are best for a faster rider (by flying speed) and which are best for a slower rider?*
- There were only a few interactions involving flying speed, indicating that the above tactics apply equally to faster or slower riders. The interactions which were significant apply early in the race. For example, the faster rider, if they are leading, benefits from moving to a higher track position between marks 1 and 3 on lap 2 (450–350 m to go), as overtaking is made more difficult or the following rider. Slower leading riders reduce their win probability by adopting this tactic, perhaps because extra effort is required to move to a higher track position.

Nonetheless, the explanatory effects in the models are quite difficult to interpret. Also, when putting tactical interpretations on explanatory effects, we must

distinguish those factors under a rider's control from those not. Another limitation of our approach is that each part of the race is considered in isolation. The joint effect of a rider applying a tactic at for example 450 m to go and another tactic at 250 m to go cannot be measured. Models for later parts of the race could have included variables relating to earlier parts of the race but, as discussed in Section 3, the number of 2-way interaction being considered would be extremely large making it difficult to interpret the models.

Finally, while feedback obtained from sprint coaches prior to the 2012 Olympic Games was positive – the general tactical implications of the models were sensible – the prevailing view was that tactics should be modeled in rider specific contexts. This would require much more data (video analysis) of rider-versus-rider contests and modeling of specific rider-versus-rider effects. Alternatively, rider attributes (other than gender and flying speed) might be quantified.

Acknowledgments: This work has been supported by the Engineering and Physical Sciences Research Council of the UK, under grant number EP/F005792/1. We are grateful for the cooperation of the English Institute for Sport for use of the data and the help of Paul Barrett, Mike Hughes and Duncan Locke, and Jan Van Eijden of British Cycling.

References

- Allsopp, P. E. and Stephen R. Clarke. 2004. "Rating Teams and Analysing Outcomes in One-Day and Test Cricket." *Journal of the Royal Statistical Society: Series A* 167: 657–667.
- Atkinson, G., R. Davison, A. Jeukendrup, and L. Passfield. 2003. "Science and Cycling: Current Knowledge and Future Directions for Research." *Journal of Sports Sciences* 21:767–787.
- Atkinson, G., O. Peacock, and L. Passfield. 2007. "Variable versus Constant Power Strategies During Cycling Time-Trials: Prediction of Time Savings Using an Up-to-Date Mathematical Model." *Journal of Sports Sciences*, 25: 1001–1009.
- Bedford, A. and J. Baglin. 2009. "Evaluating Performance of an Ice Hockey Team Using Interactive Phases of Play." *IMA Journal of Management Mathematics* 20: 159–166.
- Craig, N. P. and K. I. Norton. 2001. "Characteristics of Track Cycling." *Sports Medicine* 31: 457–468.
- de Koning, J. J., C. Foster, A. Bakkum, S. Kloppenburg, C. Thiel, et al. 2011. "Regulation of Pacing Strategy during Athletic Competition." *PLoS One* 6(1):e15863. doi:10.1371/journal.pone.0015863.
- Dobson, A. J. and A. G. Barnett. 2008. *An Introduction to Generalized Linear Models*. London: CRC Press.
- Fearing, D., J. Acimovic, and S. Graves. 2011. "How to Catch a Tiger: Understanding Putting Performance on the PGA TOUR." *Journal of Quantitative Analysis in Sports* 7(1).
- Gramacy, R., S. Jensen, and M. Taddy. 2013. "Estimating Player Contribution in Hockey with Regularized Logistic Regression." *Journal of Quantitative Analysis in Sports* 9: 97–111.
- Leitner, C., A. Zeileis, and K. Hornik. 2011. "Bookmaker Consensus and Agreement for the UEFA Champions League." *IMA Journal of Management Mathematics* 22: 183–194.
- Magnus, J. and F. Klaassen. 2008. Myths in Tennis. *Statistical Thinking in Sports*. edited by J. Albert and R. H. Koning. Chapman & Hall/CRC (Taylor & Francis Group), pp. 217–240.
- McHale, I. and M. Asif. 2013. "Modelling One Day International Cricket Match Outcomes in Play." *Submitted for publication*.
- Moffatt, J., P. Scarf, K. Zhang, L. Passfield, and I. Mchale. 2009. An analysis of strategy in the sprint cycle race in track cycling. Proceedings of the I.M.A. 2nd International Conference on Mathematics in Sport. Groningen, The Netherlands: University of Groningen, pp. 106–112.
- Nagelkerke, N. J. D. 1991. "A Note on a General Definition of the Coefficient of Determination." *Biometrika* 78: 691–692.
- Scarf, P. A. and S. Akhtar. 2011. "An Analysis of Strategy in the First Three Innings in Test Cricket: Declaration and The Follow-On." *Journal of the Operational Research Society* 62: 1931–1940.
- Scarf, P. A. and X. Shi. 2005. "Modelling Match Outcomes and Decision Support for Setting a Final Innings Target in Test Cricket." *IMA Journal Management Mathematics* 16: 161–178.

- Scarf P. A., X. Shi, and S. Akhtar. 2011. "The Distribution of Runs Scored and Batting Strategy in Test Cricket." *Journal of the Royal Statistical Society, Series A* 174: 471–497.
- Schwertman, N. C., K. L. Schenk, and B. C. Holbrook. 1996. "More Probability Models for the NCAA Regional Basketball Tournaments." *The American Statistician*, 50: 34–38.
- Thomas, K., M. R. Stone, K. G. Thompson, A. S. Gibson, and L. Ansley. 2012. "The Effect of Self- Even- and Variable-Pacing Strategies on the Physiological and Perceptual Response to Cycling." *European Journal of Applied Physiology* 112: 3069–3078.
- Tissot. 2009. Flying times. <http://www.tissotiming.com/>.
- UCI. 2009. Cycling. <http://www.uci.ch/>.
- Yeung, S. S., E. W. Yeung, and T. W. Wong. 2001. "Marathon Finishers and Non-finishers Characteristics: A Preamble to Success." *Journal of Sports Medicine and Physical Fitness* 41: 170–176.