

## ORIGINAL ARTICLE

# Changes in Thoroughbred speed and stride characteristics over successive race starts and their association with musculoskeletal injury

Adelene S. M. Wong  | Ashleigh V. Morrice-West  | R. Chris Whitton  |  
Peta L. Hitchens 

Equine Centre, Melbourne Veterinary School,  
University of Melbourne, Werribee Victoria,  
Australia

**Correspondence**

Peta L. Hitchens, Equine Centre, Melbourne  
Veterinary School, University of Melbourne,  
250 Princes Hwy, Werribee, Victoria 3030,  
Australia.

Email: [peta.hitchens@unimelb.edu.au](mailto:peta.hitchens@unimelb.edu.au)

**Funding information**

Grayson-Jockey Club Research Foundation;  
Tasracing; Racing Victoria; Victorian Racing  
Industry Fund; University of Melbourne

**Abstract**

**Background:** Certain stride characteristics have been shown to affect changes in biomechanical factors that are associated with injuries in human athletes. Determining the relationship between stride characteristics and musculoskeletal injury (MSI) may be key in limiting injury occurrence in the racehorse.

**Objectives:** This study aimed to determine whether changes in race day speed and stride characteristics over career race starts are associated with an increased risk of MSI in racehorses.

**Study design:** Case-control study.

**Methods:** Speed, stride length, and stride frequency data were obtained from the final 200 m sectional of  $n = 5660$  race starts by  $n = 584$  horses (case  $n = 146$ , control  $n = 438$ ). Multivariable joint models, combining longitudinal and survival (time to injury) analysis, were generated. Hazard ratios and their 95% confidence intervals (CI) are presented.

**Results:** The risk of MSI increased by 1.18 (95% CI 1.09, 1.28;  $P < 0.001$ ) for each 0.1 m/s decrease in speed and by 1.11 (95% CI 1.02, 1.21;  $P = 0.01$ ) for each 10 cm decrease in stride length over time (career race starts). A more marked rate of decline in speed and stride length was observed approximately 6 races prior to injury. Risk of MSI was highest early in the horse's racing career.

**Main limitations:** Only final sectional stride characteristics were assessed in the model. The model did not account for time between race starts.

**Conclusions:** Decreasing speed and stride length over multiple races is associated with MSI in racehorses. Monitoring stride characteristics over time may be beneficial for the early detection of MSI.

**KEYWORDS**

GPS (global positioning system), horse, joint model, musculoskeletal injury, speed, stride length

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Equine Veterinary Journal* published by John Wiley & Sons Ltd on behalf of EVJ Ltd.

## 1 | INTRODUCTION

There has been considerable research effort into the aetiology and epidemiology of musculoskeletal injury (MSI) in racehorses.<sup>1</sup> Substantive evidence based on pre-existing pathology found at post-mortem indicates that catastrophic MSI in racehorses are due to the material fatigue of bone, which accumulates over multiple galloping events.<sup>2-5</sup> This gradual onset provides the opportunity to identify injuries prior to their occurrence. Potential methods to detect an impending injury in the racehorse include advanced diagnostic imaging techniques that can identify prodromal structural changes in bone, and the monitoring of bone and cartilage biomarkers.<sup>6,7</sup> Advanced imaging is not practical for population wide monitoring and as microdamage is not easily observed, damage to bone needs to be relatively advanced to be observable.<sup>6</sup> Despite showing some promise, biomarkers are yet to find widespread application in equine injury prevention.<sup>8-10</sup> Statistical models have also been developed to profile horses at risk of fatality and fracture, but these have a predictive ability of only about 65–67%.<sup>11,12</sup>

Because a high proportion of MSI in racehorses develop over time, it is likely that changes in gait could be useful to detect an impending injury. Gait examination at low speed has not proven reliable for injury prediction. For example, 90% of horses that experienced a fatal injury at the Hong Kong Jockey Club showed no significant clinical signs at pre-race veterinary inspection.<sup>13</sup> Furthermore, inertial sensors failed to detect induced foot lameness in horses galloping at sub maximal exertion.<sup>14</sup> With advancement in inertial sensor technology, objective assessment of faster gaits is possible, but changes in gait in response to developing injury may only be apparent at extreme exertion. We have shown that there was substantial inter-horse variation in racehorse stride characteristics with speed predicting half or less of this variation.<sup>15</sup> Therefore, simply comparing stride characteristics between horses is unlikely to predict injury risk. Longitudinally monitoring changes in stride characteristics at maximal exertion over time in an individual horse may provide more beneficial information for MSI risk prediction. To our knowledge, a direct relationship between changes in stride characteristics over time to injury in racehorses has not been studied.

This study aims to determine whether changes in race day stride characteristics over career race starts are associated with an increased risk of MSI in racehorses. To achieve this, a joint modelling approach was employed that firstly estimates the changes in stride characteristics over time using a longitudinal process (mixed-effects linear regression models), and secondly models time to MSI (survival model). By combining these models, we are able to investigate whether changes in stride characteristics over successive race starts predict injury.

## 2 | MATERIALS AND METHODS

### 2.1 | Data sources

Three datasets were obtained and merged for Thoroughbred racehorses competing in flat races across three racing venues in Tasmania,

Australia: (1) StrideMASTER (Thoroughbred Ratings Pty Ltd) data from Tasmania's principal racing authority Tasracing from 10 July 2011 to 21 August 2016 (study period); (2) race result data from the official repository for all Thoroughbred race results across Australia; and (3) racehorse injury data from the Australian Racing Incident Database (ARID). For the latter two datasets, historical data back to 1 August 2004 was obtained from Racing Australia and Racing Victoria, with permission from the principal racing authority in Tasmania, Tasracing.

Stride characteristics, including speed, stride length, and stride frequency were available as summarised data (per 200 m segment of race; "sectional"). Erroneous stride characteristic values were identified and excluded, including where the horse was pulled up, lost a rider, fell, or was injured or disqualified. Information about the cleaning and processing of the StrideMASTER dataset are detailed in Morrice-West et al.<sup>15</sup>

Table S1 details the variables existing in the merged dataset as well as the variables generated or transformed for analysis (i.e. cumulative and percentage variables) that may act as confounders or effect modifiers. Categorical track condition data as predetermined by stewards were treated as ordinal continuous variables depicting Firm 2, Good 3 and 4, Soft 5, 6 and 7, Heavy 8, 9 and 10 respectively. Synthetic tracks do not receive a track rating, and thus were classified as Good 4 tracks due to similarity in risk and commensurate guidelines for management of synthetic surfaces. The track condition variable was then adjusted for track surface type in the model. Because they are not independent, both variables were retained in the models. Official barrier trials and non-starters (defined as horses that did not start the race, where the race was cancelled, or the horse was scratched from the race) were excluded. Only the final sectional for each race was used in analysis.

Each incident from the injury dataset was assessed individually by a specialist equine orthopaedic surgeon (RCW) and manually classified as an MSI if indicated in the free text fields or condition classification fields as being lame or as showing gait abnormalities, shin soreness, fractures, bone injuries, tendon or ligament injuries. Back muscle soreness was excluded as an MSI. An event was defined as a race-day occurrence of a fatal or non-fatal MSI, and a case as a horse having sustained at least one MSI during the study period. Enforced rest was defined as a period of greater than 22 weeks between races and/or trials based on our group's survey of 66 trainers detailing training practices undertaken by Thoroughbred horses in Victoria, Australia. This was derived from the summation of the 90th percentiles for length of rest periods (10 weeks) and post-rest time-to-trial (12 weeks).<sup>16</sup> Horses were considered retired if their last career race was within the study period.

### 2.2 | Data analysis

To maximise computational efficiency we used a case-control study design, randomly selecting three controls (no recorded raceday MSI during the study period,  $n = 438$ ) for each case ( $n = 146$ ). Case horses that did not have StrideMASTER data prior to their injury as well as

control horses with no recorded StrideMASTER data were not considered for inclusion in the case-control sample. Four horses that had non MSI-related deaths were excluded from selection.

## 2.3 | Model framework

We used a joint modelling approach to (1) estimate the changes in stride characteristics over time (successive race starts) using a mixed-effects linear regression submodel; (2) model survival event outcomes (MSI) as a proportional hazards submodel; and (3) integrate these two submodels to determine whether there are associations between changes in stride characteristics over each successive race start and an MSI event. Time-independent covariates were then considered in both submodels.<sup>17</sup> Horse-level random effects were included to account for multiple measurements over time from the same horse. We modelled the unit of time by career race start number. MSI events were only modelled up to the first-occurring injury during the study period, with recurrent injuries not considered. Time to event reflected the actual date of MSI where stride data was missing ( $n = 42$ ). Complexities within the dataset were dealt with as below:

### i. Missing data

Stride data was missing for some races prior to an event (MSI) or censoring. Missing data is informative because it may not have been recorded, for example if the horse was pulled up due to injury during a race (non-ignorable missing data).<sup>18</sup> Where intermittent missing stride data is present, the joint model calculates time-to-event as long as event information is included in the survival portion of the model.<sup>19,20</sup> This reduces bias as horses with missing data are not excluded.

### ii. Left truncation

Horses that raced prior to 10 July 2011 ( $n = 198$ ) were at-risk prior to entering the study, and were therefore subject to left truncation.<sup>21</sup> By setting the time variable in the survival-time data to be modelled based on career race start number, we account for horses that had raced and were therefore at risk prior to entering the study.

### iii. Right censoring

Observations were right-censored if they were still racing and completed the study free of MSI, or they dropped out before the end of the study period (21 August 2016), moved to a different state, or were exported overseas.

During all modelling stages (including transformation to nonlinear trajectories of time, variable selection, and inclusion of random effects), we selected models that minimised the Bayesian Information Criterion (BIC). The BIC favours the simpler model, avoiding overfitting and is generally the more consistent approach.<sup>22,23</sup>

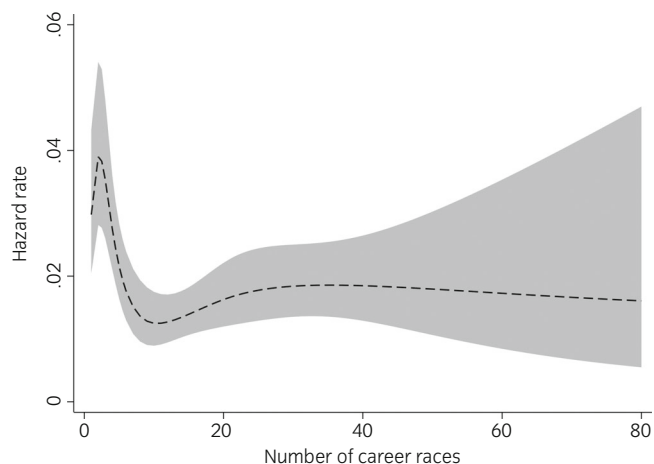
### 2.3.1 | Longitudinal submodel

To determine the relationship of the three different continuous longitudinal stride characteristics (speed, stride length, stride frequency)

with time we fitted univariable mixed-effects linear regression models and allowed for potential nonlinear trajectories of the stride characteristics by fitting both fractional polynomial and restricted cubic splines as a function of time at various degrees of freedom (df). We allowed for random effects to account for variability between and within horses.<sup>24</sup> Speed, stride length, and stride frequency models had the lowest BIC for fixed and random restricted cubic splines of time with no changes in curvature of the model over time (linear function;  $df = 1$ ). Table S2 presents the results from univariable mixed effects linear regression models for changes in speed, stride length and stride frequency over the number of career races. We univariably modelled each stride characteristic outcome against the covariates to identify potential confounders or effect modifiers that may inform the full joint models (Table S1). Continuous covariates were transformed or categorised where required.

### 2.3.2 | Survival (time to MSI) submodel

A horse was considered to have survived to the next point in time (the next race) if it did not sustain an injury in its previous race start. The model was implemented in the Stata package `stgenreg`, allowing for random intercept and slopes to account for variability between and within horses.<sup>25</sup> We investigated baseline hazard function plots for each model fitted with restricted cubic splines with degrees of freedom from one to four (Figure S1), with the best fitting model having four degrees of freedom (three interior knots) (Figure 1).<sup>26</sup> We used the default interior knot locations at the 25th, 50th, and 75th centiles of the distribution of the uncensored log survival times, plus boundary knots placed at the minimum and maximum of the distribution. We also present the Kaplan-Meier curve showing the probability of survival over time (Figure S2). As with the



**FIGURE 1** Predicted baseline hazard function for musculoskeletal injuries sustained by horses in Thoroughbred flat races in Tasmania, Australia, 2011 to 2016 ( $n = 146$ ) modelled with controls ( $n = 438$ ). The model has four degrees of freedom (three internal knots). The shaded portion depicts 95% confidence intervals.

**TABLE 1** Multivariable joint model results applied to the case-control dataset (146 cases; 438 controls) of racehorses racing in Tasmania, from 10 July 2011 to 21 August 2016 presented as regression coefficients and 95% confidence intervals (95% CI).

Covariate	Speed (m/s)		Stride length (m)	
	Coefficient (95% CI)	P-value	Coefficient (95% CI)	P-value
Longitudinal				
Time <sup>a</sup>	0.004 (0.001, 0.008)	0.008	0.002 (0.000, 0.003)	0.02
Race-level variables				
Race distance (km; x)	-2.852 (-3.383, -2.321)	<0.001	-0.196 (-0.223, -0.168)	<0.001
Race distance (x <sup>2</sup> )	0.631 (0.465, 0.797)	<0.001	-	-
Surface type				
Synthetic	Ref		Ref	
Turf	0.262 (0.203, 0.320)	<0.001	0.087 (0.068, 0.106)	<0.001
Track condition <sup>b</sup>	-0.269 (-0.286, -0.252)	<0.001	-0.091 (-0.096, -0.085)	<0.001
Number of starters (field size)	-0.017 (-0.026, -0.009)	<0.001	-	-
Race class				
Listed, Group	Ref		Ref	
Open	-0.256 (-0.388, -0.125)	<0.001	-0.053 (-0.096, -0.010)	0.02
Restricted - Benchmark, Handicap	-0.284 (-0.412, -0.156)	<0.001	-0.058 (-0.101, -0.015)	0.008
Class 2-6	-0.216 (-0.343, -0.089)	0.001	-0.044 (-0.086, -0.001)	0.05
Maiden, Class 1	-0.174 (-0.310, -0.038)	0.01	-0.032 (-0.078, 0.014)	0.2
Season				
Spring	Ref		Ref	
Summer	-0.000 (-0.052, 0.052)	>0.9	-0.018 (-0.034, -0.001)	0.04
Autumn	-0.114 (-0.174, -0.054)	<0.001	-0.044 (-0.063, -0.025)	<0.001
Winter	-0.047 (-0.109, 0.015)	0.1	-0.031 (-0.050, -0.011)	0.002
Day or night meeting				
Day	Ref		Ref	
Night	0.393 (0.343, 0.443)	<0.001	0.103 (0.087, 0.119)	<0.001
Calendar year of race <sup>b</sup>	0.032 (0.013, 0.051)	0.001	-	-
Horse-level variables				
Starting price (AUD) <sup>c</sup>	-0.106 (-0.132, -0.081)	<0.001	-0.016 (-0.024, -0.008)	<0.001
Percentage of career spent in rest (>60 days)	0.000 (-0.001, 0.001)	0.8	0.000 (-0.000, 0.000)	0.2
Percentage places at current race	0.006 (0.004, 0.007)	<0.001	0.002 (0.001, 0.002)	<0.001
Cumulative Australian career winnings 2008-current (AUD)				
Zero	Ref		-	-
More than zero	0.113 (0.014, 0.212)	0.03	-	-
Sex				
Female	-	-	Ref	
Gelding	-	-	0.174 (0.131, 0.216)	<0.001
Intact Male	-	-	0.156 (0.077, 0.236)	<0.001
Age at first race (years)	-	-	-0.039 (-0.068, -0.011)	0.007
Horse country of origin				
Australia	-	-	Ref	
International	-	-	0.263 (0.131, 0.395)	<0.001
Intercept	19.37 (18.94, 19.81)	<0.001	7.417 (7.295, 7.538)	<0.001
Survival				
Association <sup>d</sup>	-1.684 (-2.477, -0.890)	<0.001	-1.050 (-1.888, -0.212)	0.01

(Continues)

TABLE 1 (Continued)

Covariate	Speed (m/s)		Stride length (m)	
	Coefficient (95% CI)	P-value	Coefficient (95% CI)	P-value
Race-level variables				
Race distance (km; x)	-4.366 (-9.202, 0.471)	0.08	0.190 (-0.428, 0.808)	0.5
Race distance (x <sup>2</sup> )	1.107 (-0.327, 2.541)	0.1	-	-
Surface type				
Synthetic	Ref		Ref	
Turf	0.672 (0.096, 1.248)	0.02	0.331 (-0.196, 0.859)	0.2
Track condition <sup>b</sup>	-0.711 (-0.999, -0.422)	<0.001	-0.356 (-0.560, -0.153)	0.001
Number of starters (field size)	0.013 (-0.065, 0.091)	0.7	-	-
Race class				
Listed, Group	Ref		Ref	
Open	-0.928 (-1.981, 0.126)	0.08	-0.503 (-1.511, 0.506)	0.3
Restricted - Benchmark, Handicap	-0.799 (-1.749, 0.151)	0.1	-0.311 (-1.221, 0.600)	0.5
Class 2-6	-0.707 (-1.619, 0.204)	0.1	-0.359 (-1.241, 0.524)	0.4
Maiden, Class 1	-0.631 (-1.614, 0.353)	0.2	-0.291 (-1.247, 0.664)	0.6
Season				
Spring	Ref		Ref	
Summer	0.879 (0.351, 1.407)	0.001	0.874 (0.356, 1.392)	0.001
Autumn	0.745 (0.160, 1.331)	0.01	0.895 (0.339, 1.451)	0.002
Winter	0.585 (-0.089, 1.260)	0.09	0.696 (0.036, 1.357)	0.04
Day or night meeting				
Day	Ref		Ref	
Night	0.803 (0.274, 1.332)	0.003	0.261 (-0.170, 0.693)	0.235
Calendar year of race <sup>b</sup>	0.073 (-0.047, 0.192)	0.2	-	-
Horse-level variables				
Starting price (AUD) <sup>c</sup>	-0.443 (-0.687, -0.200)	<0.001	-0.167 (-0.362, 0.028)	0.1
Percentage of career spent in rest (>60 days)	-0.011 (-0.017, -0.004)	0.001	-0.009 (-0.016, -0.003)	0.005
Percentage places at current race	-0.004 (-0.015, 0.008)	0.5	-0.011 (-0.021, -0.001)	0.04
Cumulative Australian career winnings 2008-current (AUD)				
Zero	Ref		-	-
More than zero	-0.313 (-0.990, 0.365)	0.4	-	-
Sex				
Female	-	-	Ref	
Gelding	-	-	0.159 (-0.224, 0.542)	0.4
Intact Male	-	-	0.719 (0.002, 1.437)	0.05
Age at first race (years)	-	-	0.045 (-0.198, 0.288)	0.7
Horse country of origin				
Australia	-	-	Ref	
International	-	-	0.740 (-0.240, 1.720)	0.1
Spline 1	0.148 (-0.122, 0.419)	0.3	-0.062 (-0.305, 0.181)	0.6
Spline 2	-0.106 (-0.288, 0.076)	0.3	-0.059 (-0.238, 0.119)	0.5
Spline 3	-0.178 (-0.325, -0.030)	0.02	-0.138 (-0.282, 0.007)	0.06
Spline 4	0.092 (-0.075, 0.260)	0.3	0.114 (-0.050, 0.278)	0.2
Intercept	30.52 (15.03, 46.01)	<0.001	5.034 (-1.381, 11.45)	0.1

TABLE 1 (Continued)

Covariate	Speed (m/s)		Stride length (m)	
	Coefficient (95% CI)	P-value	Coefficient (95% CI)	P-value
Random Effects				
sd (Time)	0.010 (0.006, 0.016)	-	0.005 (0.004, 0.007)	-
sd (Intercept)	0.341 (0.294, 0.397)	-	0.250 (0.230, 0.272)	-
corr (Time, Intercept)	-0.478 (-0.702, -0.167)	-	-0.397 (-0.585, -0.168)	-
sd (Residual)	0.613 (0.600, 0.626)	-	0.189 (0.185, 0.193)	-

Note: Regression coefficients in the longitudinal model show the increase or decrease in the outcome variable per one unit change in the covariates. The associations between speed and stride length and time-to-injury are presented under the survival model.

<sup>a</sup>1 unit of time = 1 race start.

<sup>b</sup>Treated as continuous variables with linear associations after checking for appropriateness of transformation.

<sup>c</sup>Natural log-transformed variable.

<sup>d</sup>The association structure is interpreted as the log hazard ratio for a one-unit increase in the value of the longitudinal outcome at time  $t$ .

longitudinal submodel, we modelled each covariate univariably against the injury outcome (Table S1).

### 2.3.3 | Joint longitudinal-MSI model

The association structure in the joint model is interpreted as the log hazard ratio for a one-unit increase in the value of the longitudinal outcome (1 m per second for speed, 1 m for stride length and 1 Hz for stride frequency) at time  $t$ , where  $t$  is the successive race number.<sup>27</sup> Coefficients and their 95% confidence intervals (95% CI) are presented for the multivariable joint model (Table 1), and the association coefficients have been transformed to Hazard Ratios (HR) for interpretation of the relationship between speed and stride length and risk of MSI within text. Confidence intervals that cross 0 (coefficient) or 1 (HR) imply non-significance. Because we observe association coefficients ( $\beta$ ) that are negative, the HR for a one unit decrease in the longitudinal outcome is  $HR = \frac{1}{e^{\beta}}$ . The values of the association coefficients are divided by 10 to calculate the hazard ratios for a 0.1 m/s and a 10 cm increase in the value of speed and stride length respectively. The overall log hazard ratio for the effect of each covariate  $[x]$  is  $\alpha \cdot \delta [x] + \phi [x]$ , where  $\alpha$  is the coefficient of the association value,  $\delta$  is the coefficient of the specified covariate  $[x]$  in the longitudinal submodel, and  $\phi$  is the coefficient of the specified covariate  $[x]$  in the survival submodel of the joint model and was calculated using the Stata postestimation command `nlcom`.<sup>27</sup>

We first generated univariable joint models using covariates that were significant in either the univariable mixed-effects linear regression models or survival models (Tables S3, S4 and S5). Collinearity between variables was assessed using pairwise Pearson correlation coefficients ( $r$ ). Variables that were collinear ( $r \geq 0.6$ ) with career races (time), suggesting time dependency (horse age at start, cumulative race distance, number of career wins and places, career length, cumulative time in rest), were excluded from the multivariable joint model.

The multivariable joint model was generated using the forward stepwise approach, adding variables significant in either component

of the univariable joint model to both sides of the null joint model. The fit of variables to both components of the joint models were informed initially by model diagnostics run on the respective univariable longitudinal (`mixed`) and survival (`stgenreg`) models, including assessing collinearity between predictor variables, AIC/BIC, and the link test (mixed model only) to identify model specification errors. In the multivariable joint model, we further assessed residuals of the longitudinal component, and martingale-like and deviance residuals of the survival component for normality and outliers. Then, using backward stepwise elimination, the covariates with the largest  $P$ -values ( $>0.05$ ) on both sides of the model were removed. Covariates were retained in the joint model if significant in either the longitudinal or survival component of the model.

When generating univariable joint models for stride frequency, we found no significant association between changes in stride frequency over time and injury (Table S5). Therefore, a multivariable stride frequency joint model was not generated.

Joint models were fitted in Stata version 15.1 (StataCorp) using the user written package `stjm`.<sup>27</sup> Significance level was set at  $P \leq 0.05$ .

## 3 | RESULTS

### 3.1 | Description of the study population

Data was obtained from 2814 horses competing in 3107 races in Tasmania during the study period (Figure S3). A total of 25,043 StrideMASTER race starts were available for analysis. During the study period, 179 race-day MSI were reported from 162 horses (5.7%), 10 of which were fatal MSI, and 169 non-fatal. There were 2652 horses that did not record a race-day MSI during the study period, including four horses that died due to non-musculoskeletal related complications. The mean age for entry into the study was 3.77 years (s.d. 1.27; range 1.96, 9.82), and the mean age for exit (including age at exit before the end of the study e.g. retirement, or age at the end of study period) was 5.21 years (s.d. 1.51; range 2.15,

11.44). Mean speed, stride length, and stride frequency from the final 200 m of each race start were  $15.86 \text{ ms}^{-1}$  (s.d. 0.89; range 12.57, 19.70), 6.88 m (s.d. 0.34; range 5.37, 8.63), and  $2.31 \text{ s}^{-1}$  (s.d. 0.09; range 2.04, 2.69) respectively.

A total of 1593 (56.6%) horses experienced enforced rest during the study period, and 2148 (76.3%) horses retired from racing before the end of the study period. A total of 130 of the 2148 (6.1%) retired horses recorded race-day MSI prior to retirement. Only 181 (6.4%) horses did not experience any injury or undergo any enforced rests and were still racing by the end of the study period.

### 3.2 | Hazard of injury in the case-control dataset

Figure 1 presents the baseline survival (hazard) model of the case-control dataset prior to adjusting for confounders or effect modifiers. The hazard of injury was highest at the second race with 0.039 (95% CI 0.028, 0.054;  $n = 371$  horses at risk) injuries predicted to occur per race start; decreasing by the tenth race to 0.012 (95% CI 0.009, 0.017;  $n = 268$ ) injuries per race start; increasing to 0.019 (95% CI 0.014, 0.025;  $n = 50$ ) up to race 35; at which point there is a gradual decrease albeit with much wider confidence intervals. The overall Kaplan-Meier probability of MSI over successive race starts is shown in Figure S2.

### 3.3 | Relationship between covariates of the joint model and speed, stride length and MSI

Speed and stride length increased over career race starts (Table 1; Longitudinal). Speed increased by 0.004 m/s per race start (95% CI 0.001, 0.008;  $P = 0.008$ ), and stride length increased by 0.002 m per race start (95% CI 0.000, 0.003;  $P = 0.02$ ).

Race-level variables such as race distance, surface type, track condition, race class, season, and day or night meeting were associated with speed and stride length (Table 1; Longitudinal). Faster speeds in the final sectional were associated with shorter distance races, with a non-linear reduction to about 2000 m, followed by levelling out of speeds thereafter. Shorter race distances were associated with longer strides. Faster speeds and longer stride lengths were associated with firmer turf surfaces, in Listed and Group races, in spring, and during night races. A smaller field size was associated with faster speeds but had no effect on stride length. Measures of superior performance at the horse-level such as lower starting price (shorter odds) and horses that had a higher percentage of places were associated with faster speeds and longer stride lengths. Horses that had career winnings were faster than horses that had not accumulated any career winnings but did not have different stride lengths. Geldings and stallions, horses that started racing at a younger age, and international horses took longer strides.

Assessing the direct effects of the covariates on MSI risk in the survival component of the multivariable joint model, firmer turf surfaces, lower percentage of career spent resting, as well as races during

summer, autumn and winter compared with spring races were directly associated with an increased risk of MSI for each subsequent race in both speed and stride length models. Additionally, night races and lower starting price (indicative of horses expected to perform better) in the speed model, and horses with a lower percentage of places and stallions in the stride length model were associated with an increased risk of MSI for each subsequent race (Table 1; Survival).

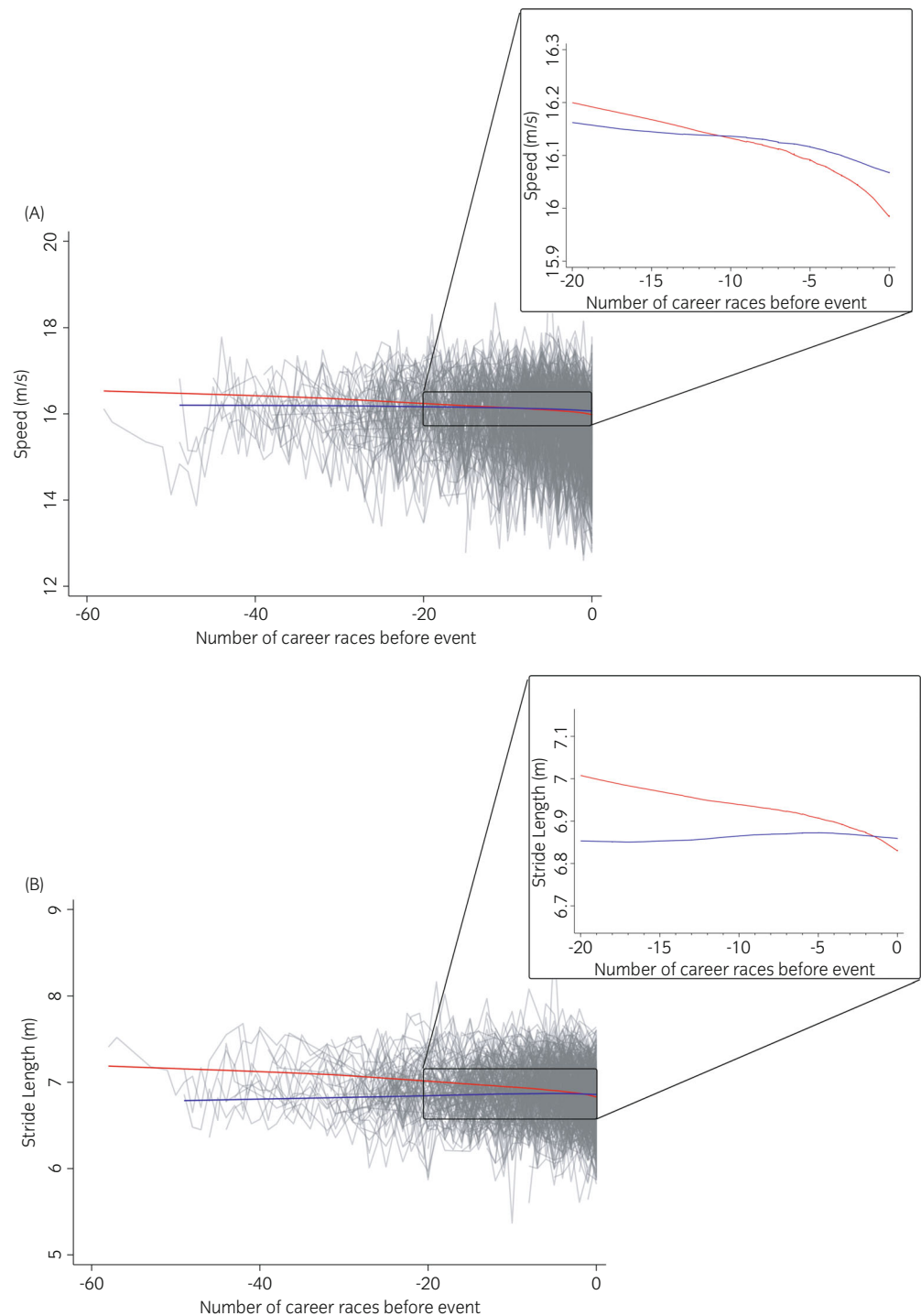
While the main objective of this paper was to determine the relationship between stride characteristics and the risk of injury, an added benefit of joint modelling is that the direct effect of a covariate on MSI can be distinguished from its indirect effect through the longitudinal outcome.<sup>28</sup> The overall effect on MSI is therefore an aggregated effect of time-to-event and the longitudinal process when the covariate is included in both submodels (Figure S4).<sup>29</sup> Taking the indirect effect of speed and stride length changes over time into account, firmer surfaces, races during summer, autumn, and winter compared with spring races, lower percentage of career spent resting and a lower percentage of places contributed to an increased risk of MSI. Horses with a lower starting price contributed to an overall increased risk of MSI in the speed joint model but not the stride length joint model (Table S6).

### 3.4 | Relationship between stride characteristics and MSI

The association coefficient in the speed and stride length joint models were significantly different from zero, providing strong evidence of association between the longitudinal and time-to-event submodels. The risk of race-day MSI increased with decreasing speed and stride length but not stride frequency over career race starts. The risk of MSI increased by 1.18 (95% CI 1.09, 1.28;  $P < 0.001$ ) for each  $0.1 \text{ ms}^{-1}$  decrease in speed and by 1.11 (95% CI 1.02, 1.21;  $P = 0.01$ ) for each 10 cm decrease in stride length over career race starts (Table 1). Speed and stride length for injured horses declined more rapidly approximately six races prior to injury (Figure 2). While Figure 2 shows an apparent finding that injured horses were faster and had longer stride lengths overall, the timeline of these plots depict time preceding injury or censor instead of consecutive career race starts. Therefore, comparing speeds and stride lengths between cases and controls at a particular time in their career cannot be made based on these graphs. Additionally, in the joint model, speed and stride length were observed to increase over racehorses' careers, an observation that cannot be inferred from these graphs.

Models for fatal MSI as the outcome could not converge due to the low incidence of fatality. One of the ten fatally injured horses from the study population that had a non-fatal injury 14 races prior to its fatal injury was not included as part of the fatally injured horses in the case-control dataset as it was modelled only up to its first, non-fatal injury. Only six of the remaining nine fatally injured horses had more than two races with StrideMASTER data prior to their fatal injury. Three of the six horses show an overall decrease in speed and stride length in the six to eight races prior to injury (Figure S5).

**FIGURE 2** Lowess smoothing of longitudinal trajectory predictions for speed and stride length stratified by cases (horses injured during the 2011-2016 racing seasons,  $n = 146$ ) (red line) or controls (horses not injured during 2011-2016 racing seasons,  $n = 438$ ) (blue line) adjusting for significant covariates in the multivariable joint model overlaid on raw trajectory plots of individual horse's (A) speed and (B) stride lengths in successive final 200 m sectionals from  $n = 5660$  race starts in Tasmania, Australia. Inset graphs show zoomed in detail of the model predictions from 20 career races before the event.



## 4 | DISCUSSION

Using a multivariable joint model that allowed for longitudinal monitoring over time, we observed a decrease in speed and stride length, but not stride frequency, over career race starts in association with increased risk of race-day MSI. This was despite a general trend for increasing speed and stride length with greater numbers of career starts for all horses. Adjusted speed and stride length for injured horses were observed to decline more rapidly approximately six races

prior to an MSI being identified. These findings highlight the potential of speed and stride monitoring during races for injury prediction. However, the hazard of injury was highest early in a horse's career, peaking at the second race, suggesting that the additional monitoring of training workloads may be needed for predicting these early career injuries.

Despite the availability of inertial sensors for gait analysis of horses, there are limited data on their use for identification of gait changes in galloping horses. A study by Lopes et al.<sup>14</sup> using inertial

sensors found that induced lameness could not be detected at a gallop. Head and pelvic acceleration and rotation, limb timing, stride duration measurements, and limb lead preference during galloping did not change before and after induction of a single fore- or hindlimb lameness. An increase in stride frequency was observed for horses with induced lameness of the right forelimb when leading with the right limb but this effect was not evident for induced lameness of the left fore when leading with the left limb. Although speed of gait was not measured, the horses were not performing at maximum exertion as in our study.<sup>14</sup> To our knowledge, the direct association between stride characteristics and MSI over time in horses has not been studied.

In human athletes, studies relating stride characteristics to MSI risk have predominantly focused on inherent differences between individuals rather than changes over time. A prospective study of 54 collegiate cross country runners demonstrated that a lower stride frequency was associated with increased bone stress injury risk.<sup>30</sup> In other human studies, indirect relationships between stride characteristics and injury were demonstrated. For example, a reduced stride length along with an increase in stride frequency is associated with key biomechanical factors known to reduce the risk of running injuries.<sup>31-33</sup> And a runner's centre of mass excursion as well as peak vertical ground reaction force were observed to reduce when stride frequency was increased.<sup>34,35</sup> These studies demonstrate that longer strides for a given speed increase the risk of injury. Because injuries develop over time, horses that reduce their speed and stride length could be attempting to circumvent or reduce pain due to an undetected or impending injury.

While our study encompasses a broad range of types of MSI, our findings align with previous studies that have identified that horses in their first year of racing were at increased risk of fatal lateral condylar and distal limb fractures.<sup>36,37</sup> However, a greater number of starts is also associated with an increased risk of MSI as we similarly observed in this present study beyond approximately 15 career starts.<sup>38,39</sup> Increasing horse age is associated with an increase in injury risk likely due to the accumulation of damage with longer racing careers.<sup>40</sup> Current research points towards two differing pathways to injury whereby bone damage could occur when loading is introduced too rapidly to a poorly adapted skeleton or when racing or training duration and intensity exceed the capacity of a well-adapted skeleton.<sup>3,41</sup> Injuries that occur early in a horse's career are hypothesised to occur due to the former, while injuries that occur later in a horse's career would likely have occurred due to the latter pathway. Modifying the horse's training and/or racing schedule when a consistent decrease in speed and stride length across several races are observed could be key to preventing a proportion of musculoskeletal injuries. Detecting changes in race stride characteristics over time requires horses to have a minimum number of starts, however, in the current study 28% developed injuries within the first three races of their career, limiting the use of this technique in this group.

There were insufficient numbers of fatal injuries in this study to draw definitive conclusions about the usefulness of speed and stride

data for predicting fatal MSI with half of the fatally injured horses that had sufficient stride data showing decreasing speed and stride length leading up to injury. Two out of nine horses developed fatal injuries before accumulating enough starts to determine a typical stride pattern, and one horse's stride was not monitored throughout its career prior to its fatal injury.

There are several limitations for this study. First, we did not use all sectionals of the race, but only the final 200 m sectional of each race start. We reasoned that horses are at peak exertional capacity as they approach the end of a race, therefore, the final sectional stride characteristics are a true representation of what the horse can achieve rather than being under the influence of riding tactics or the surrounding horses as may occur earlier in the race. However, by only using the last sectional of the race, the model does not account for the overall race speed and time, and how those might interplay with speed in the final sectional. It is also likely that a proportion of horses were not pushed to full exertion by the jockey if they were perceived to have no chance of finishing competitively, if they encountered interference, or were unable to progress in the race for other reasons. In this study, jockey licence type and weight carried were not associated with any outcome, but other jockey-level factors may require consideration in future studies such as whip use violations or careless riding. Second, we did not have access to stride data for horses in training which may have provided a more complete picture of how a horse's stride characteristics change in response to injury development, particularly in those horses that develop injuries in the first few races of their career. However, horses in training are rarely performing maximally as they are in the final sectional of races so such data may not be as important. Third, we have modelled time as career race starts as this best represents the accumulated race intensity workload the horse has undertaken. However, this approach does not take into account time between each race start or the number of days to event or censor. In order to account for bone reparative processes between preparations, time since the end of the previous spell was adjusted for and found to be univariably significant in the longitudinal components of the joint models (Tables S3, S4 and S5), but not retained in multivariable analysis. Fourth, as this dataset only provides injury events reported on race days, the non-injured controls are likely to include horses with injuries that only became apparent when the horse left the track or occurred during subsequent training, potentially reducing the differences in stride characteristics over time between the injured and uninjured horses. Fifth, we have not assessed recurrent events. Instead, we have modelled injuries up to the first injury during the study period in an attempt to minimise bias from horses that had underlying injuries. Sixth, we have modelled speed, stride length, and stride frequency separately as the univariate joint models were already computationally complex. In future studies, the interdependence between speed, stride length, and stride frequency can be accounted for by modelling them in a multivariate joint model. Seventh, some effects in the joint model may be confounders or effect modifiers. For example, only one racetrack had a synthetic surface, and only one conducted night meetings. Therefore, it could not be determined whether these study factors

affected stride characteristics, or whether these effects were due to other characteristics of those venues. Lastly, the uninjured population of horses in our study included horses that retired or had enforced rest(s) during the study period. Early retirement and enforced rest are competing risks, as their occurrence may preclude the occurrence of MSI, our event of interest.<sup>42–44</sup> Early retirement or enforced rest could be indicators of an underlying injury, hence, including them with the uninjured population of horses may reduce the stride pattern differences observed over time between the injured cases and uninjured controls. The ability for changes in strides over time to predict MSI as well as other competing risk outcomes warrants further investigation.

While more computationally intensive, we chose the joint model framework for this study due to its reduced bias and improved precision over simpler approaches.<sup>19,45</sup> As joint models were developed so that the association and dependence between the longitudinal marker and time to event are considered, the joint model is able to account for measurement error, the intermittent nature of observations, as well as missing data.<sup>29,46,47</sup>

This study demonstrates that a reduction in speed and stride length, but not stride frequency, in the final sectional of a race over multiple starts may predict MSI in racehorses. A wider implementation of inertial sensor technology to monitor horses in training and racing will provide valuable datasets for further investigation. This method is most promising for horses that develop injuries later in their racing careers where multiple race data are available for analysis. It is important that factors that affect stride characteristics are adjusted for when assessing injury risk due to the variations in speed and stride length that result. Accounting for time between race starts and competing risk events will be key to further improving the models.

## AUTHORSHIP

P. Hitchens conceptualised the study. A. Wong, A. Morrice-West, and P. Hitchens contributed to the data analysis. A. Wong drafted the manuscript. All authors contributed to the study design, interpretation of the data, revised the manuscript critically for important intellectual content, and gave final approval of the version to be published.

## ACKNOWLEDGEMENTS

The authors thank Tasracing for provision of data and StrideMASTER for advice on their system. The authors also thank Usha Nattala, Geordie Z. Zhang, and Pratibha Pratibha from the Melbourne Data Analytics Platform of The University of Melbourne for their insight into alternative data analysis methodologies.

## SOURCES OF FUNDING

The Grayson-Jockey Club Research Foundation, Tasracing, Racing Victoria, the Victorian Racing Industry Fund of the Victoria State Government, and The University of Melbourne under the Equine Limb Injury Prevention Program.

## CONFLICT OF INTERESTS

No competing interests have been declared.

## DATA ACCESSIBILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## ETHICAL ANIMAL RESEARCH

The Animal Ethics Committee at the University of Melbourne Faculty of Veterinary and Agricultural Science gave an exemption for formal ethics approval.

## INFORMED CONSENT

Data collection is required by the racing authority. Representatives of Tasracing gave consent for this study.

## ORCID

Adelene S. M. Wong  <https://orcid.org/0000-0002-3231-590X>

Ashleigh V. Morrice-West  <https://orcid.org/0000-0003-2776-4273>

R. Chris Whitton  <https://orcid.org/0000-0003-0012-4065>

Peta L. Hitchens  <https://orcid.org/0000-0002-7528-7056>

## REFERENCES

1. Verheyen KLP. Reducing injuries in racehorses: Mission impossible? *Equine Vet J.* 2013;45:6–7.
2. Riggs CM, Whitehouse GH, Boyde A. Pathology of the distal condyles of the third metacarpal and third metatarsal bones of the horse. *Equine Vet J.* 1999;31:140–8.
3. Martig S, Chen W, Lee PVS, Whitton RC. Bone fatigue and its implications for injuries in racehorses. *Equine Vet J.* 2014;46:408–15.
4. Whitton RC, Trope GD, Ghasem-Zadeh A, Anderson GA, Parkin TDH, Mackie EJ, et al. Third metacarpal condylar fatigue fractures in equine athletes occur within previously modelled subchondral bone. *Bone.* 2010;47:826–31.
5. Stover SM, Johnson BJ, Daft BM, Read DH, Anderson M, Barr BC, et al. An association between complete and incomplete stress fractures of the humerus in racehorses. *Equine Vet J.* 1992;24:260–3.
6. Colgate VA, Marr CM. Science-in-brief: risk assessment for reducing injuries of the fetlock bones in Thoroughbred racehorses. *Equine Vet J.* 2020;52:482–8.
7. Turlo AJ, Cywinska A, Frisbie DD. Revisiting predictive biomarkers of musculoskeletal injury in Thoroughbred racehorses: longitudinal study in polish population. *BMC Vet Res.* 2019;15:1–9.
8. Jackson BF, Lonnell C, Verheyen KLP, Dyson P, Pfeiffer DU, Price JS. Biochemical markers of bone metabolism and risk of dorsal metacarpal disease in 2-year-old Thoroughbreds. *Equine Vet J.* 2005;37:87–91.
9. Jackson BF, Dyson PK, Lonnell C, Verheyen KLP, Pfeiffer DU, Price JS. Bone biomarkers and risk of fracture in two- and three-year-old Thoroughbreds. *Equine Vet J.* 2009;41:410–3.
10. Frisbie DD, McIlwraith CW, Arthur RM, Blea J, Baker VA, Billingham RC. Serum biomarker levels for musculoskeletal disease in two- and three-year-old racing Thoroughbred horses: a prospective study of 130 horses. *Equine Vet J.* 2010;42:643–51.
11. Rosanowski SM, Chang YM, Stirk AJ, Verheyen KLP. Risk factors for race-day fatality, distal limb fracture and epistaxis in Thoroughbreds racing on all-weather surfaces in Great Britain (2000 to 2013). *Prev Vet Med.* 2017;148:58–65.
12. Georgopoulos SP, Parkin TDH. Risk factors for equine fractures in Thoroughbred flat racing in North America. *Prev Vet Med.* 2017;139:99–104.

13. Stewart BD, Lam KH. Racing injury risk assessment at official pre-race veterinary inspections in Hong Kong Jockey Club. *Proceedings of the 18th international conference of racing analysts and veterinarians*; New Zealand: 2010. p. 296–7.
14. Lopes MAF, Dearo ACO, Lee A, Reed SK, Kramer J, Pai PF, et al. An attempt to detect lameness in galloping horses by use of body-mounted inertial sensors. *Am J Vet Res*. 2016;77:1121–31.
15. Morrice-West AV, Hitchens P, Whitton CR. Variation in GPS and accelerometer recorded velocity and stride parameters of galloping Thoroughbred horses. *Equine Vet J*. 2021;53:1063–74.
16. Morrice-West AV, Hitchens PL, Walmsley EA, Stevenson MA, Whitton RC. Training practices, speed and distances undertaken by Thoroughbred racehorses in Victoria. *Australia Equine Vet J*. 2020; 52:273–80.
17. Crowther MJ, Andersson TML, Lambert PC, Abrams KR, Humphreys K. Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. *Stat Med*. 2016;35:1193–209.
18. Dupuy JF, Mesbah M. Joint modeling of event time and nonignorable missing longitudinal data. *Lifetime Data Anal*. 2002;8:99–115.
19. Crowther MJ. *Stjm: Stata module for joint modelling of longitudinal and survival data*. Available at: <https://fmwww.bc.edu/repec/bocode/s/stjm.html>.
20. Li S. Joint modeling of recurrent event processes and intermittently observed time-varying binary covariate processes. *Lifetime Data Anal*. 2016;22:145–60.
21. Cole SR, Hudgens MG. Survival analysis in infectious disease research: describing events in time. *AIDS*. 2010;24:2423–31.
22. Pu W, Niu XF. Selecting mixed-effects models based on a generalized information criterion. *J Multivar Anal*. 2006;97:733–58.
23. Brewer MJ, Butler A, Cooksley SL. The relative performance of AIC, AICC and BIC in the presence of unobserved heterogeneity. *Methods Ecol Evol*. 2016;7:679–92.
24. StataCorp. *Stata multilevel mixed-effects reference manual*. Stata: release 16. Statistical software. College Station, TX: StataCorp LLC. 2019;52:743–9.
25. Crowther MJ, Lambert PC. *Stgenreg: a stata package for general parametric survival analysis*. *J Stat Softw*. 2013;53:1–17.
26. Crowther MJ, Lambert PC. A general framework for parametric survival analysis. *Stat Med*. 2014;33:5280–97.
27. Crowther MJ, Abrams KR, Lambert PC. Joint modeling of longitudinal and survival data. *Stata J*. 2013;13:165–84.
28. van Oudenhoven FM, Swinkels SHN, Ibrahim JG, Rizopoulos D. A marginal estimate for the overall treatment effect on a survival outcome within the joint modeling framework. *Stat Med*. 2020;39: 4120–32.
29. Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. *J Clin Oncol*. 2010;28:2796–801.
30. Kliethermes SA, Stiffler-Joachim MR, Wille CM, Sanfilippo JL, Zavala P, Heiderscheid BC. Lower step rate is associated with a higher risk of bone stress injury: a prospective study of collegiate cross country runners. *Br J Sports Med*. 2021;55(15):851–6.
31. Schubert AG, Kempf J, Heiderscheid BC. Influence of stride frequency and length on running mechanics: a systematic review. *Sports Health*. 2014;6:210–7.
32. Boyer ER, Derrick TR. Select injury-related variables are affected by stride length and foot strike style during running. *Am J Sports Med*. 2015;43:2310–7.
33. Chumanov ES, Wille CM, Michalski MP, Heiderscheid BC. Changes in muscle activation patterns when running step rate is increased. *Gait Posture*. 2012;36:231–5.
34. Morin JB, Samozino P, Zameziati K, Belli A. Effects of altered stride frequency and contact time on leg-spring behavior in human running. *J Biomech*. 2007;40:3341–8.
35. Heiderscheid BC, Chumanov ES, Michalski MP, Wille CM, Ryan MB. Effects of step rate manipulation on joint mechanics during running. *Med Sci Sports Exerc*. 2011;43:296–302.
36. Parkin TDH, Clegg PD, French NP, Proudman CJ, Riggs CM, Singer ER, et al. Risk factors for fatal lateral condylar fracture of the third metacarpus/metatarsus in UK racing. *Equine Vet J*. 2010;37:192–9.
37. Parkin TDH, Clegg PD, French NP, Proudman CJ, Riggs CM, Singer ER, et al. Horse-level risk factors for fatal distal limb fracture in racing Thoroughbreds in the UK. *Equine Vet J*. 2004;36:513–9.
38. Georgopoulos SP, Parkin TDH. Risk factors associated with fatal injuries in Thoroughbred racehorses competing in flat racing in the United States and Canada. *J Am Vet Med Assoc*. 2016;249:931–9.
39. Peloso JG, Vogler JB, Cohen ND, Marquis P, Hilt L. Association of catastrophic biaxial fracture of the proximal sesamoid bones with bony changes of the metacarpophalangeal joint identified by standing magnetic resonance imaging in cadaveric forelimbs of Thoroughbred racehorses. *J Am Vet Med Assoc*. 2015;246:661–73.
40. Bailey CJ, Reid SWJ, Hodgson DR, Bourke JM, Rose RJ. Flat, hurdle and steeple racing: risk factors for musculoskeletal injury. *Equine Vet J*. 1998;30:498–503.
41. Hitchens PL, Morrice-West AV, Stevenson MA, Whitton RC. Meta-analysis of risk factors for racehorse catastrophic musculoskeletal injury in flat racing. *Vet J*. 2019;245:29–40.
42. Rajeswaran J, Blackstone EH, Barnard J. Joint modeling of multivariate longitudinal data and competing risks using multiphase submodels. *Stat Biosci*. 2018;10:651–85.
43. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009;170:244–56.
44. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94:496–509.
45. Sweeting MJ, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biom J*. 2011;53:750–63.
46. Ashra NB, Marriott L, Johnson S, Abrams KR, Crowther MJ. Jointly modelling longitudinally measured urinary human chorionic gonadotropin and early pregnancy outcomes. *Sci Rep*. 2020;10:1–9.
47. Billingham LJ, Abrams KR. Simultaneous analysis of quality of life and survival data. *Stat Methods Med Res*. 2002;11:25–48.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Wong ASM, Morrice-West AV, Whitton RC, Hitchens PL. Changes in Thoroughbred speed and stride characteristics over successive race starts and their association with musculoskeletal injury. *Equine Vet J*. 2023; 55(2):194–204. <https://doi.org/10.1111/evj.13581>