



Georgopoulos, S. P. and Parkin, T. D.H. (2017) Risk factors for equine fractures in Thoroughbred flat racing in North America. *Preventive Veterinary Medicine*, 139(Pt. B), pp. 99-104.  
(doi:[10.1016/j.prevetmed.2016.12.006](https://doi.org/10.1016/j.prevetmed.2016.12.006))

This is the author's final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/133551/>

Deposited on: 27 March 2017

Enlighten – Research publications by members of the University of Glasgow  
<http://eprints.gla.ac.uk/33640>

# Risk factors for equine fractures in Thoroughbred flat racing in North America

Stamatis Panagiotis Georgopoulos<sup>\*</sup>, Tim D.H. Parkin

School of Veterinary Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, 464 Bearsden Road, Glasgow G61 1QH, UK

---

## abstract

The aim of this paper is to identify risk factors associated with equine fractures in flat horse racing of Thoroughbreds in North America. Equine fractures were defined as any fracture sustained by a horse during a race. This was a cohort study that made use of all starts from the racecourses reporting injuries. The analysis was based on 2,201,152 racing starts that represent 91% of all official racing starts in the USA and Canada from 1 st January 2009–31 st December 2014. Approximately 3,990,000 workout starts made by the 171,523 Thoroughbreds that raced during that period were also included in the analysis. During this period the incidence of equine fractures was 2 per 1000 starts. The final multivariable logistic regression models identified risk factors significantly associated ( $p < 0.05$ ) with equine fracture. For example, horses were found to have a 32% higher chance of sustaining a fracture when racing on a dirt surface compared to a synthetic surface; a 35% higher chance if they had sustained a previous injury during racing and a 47% higher chance was also found for stallions compared to mares and geldings. Furthermore, logistic regression models based on data available only from the period 2009–2013 were used to predict the probability of a Thoroughbred sustaining a fracture for 2014. The 5% of starts that had the highest score in our predictive models for 2014 were found to have 2.4 times (95% CI: 1.9–2.9) higher fracture prevalence than the mean fracture prevalence of 2014. The results of this study can be used to identify horses at higher risk on entering a race and could help inform the design and implementation of preventive measures aimed at minimising the number of Thoroughbreds sustaining fractures during racing in North America.

---

## 1. Introduction

This study is based on equine fractures in flat horse racing of Thoroughbreds in the USA and Canada during the period 2009–2014. More than 80% of equine fatalities in this period were the result of a fracture. As such, they are a primary focus of epidemiological analyses of existing racing data aimed at maximising the welfare of the racehorse.

Recent studies investigating equine injuries across different countries and jurisdictions have identified associations between them and plausible risk factors. Horse-related risk factors, such as the age, the sex, and the prior racing history of the horse, have been shown to be associated with injuries: age (Estberg et al., 1996a; Estberg et al., 1998a,b; Williams et al., 2001; Parkin et al., 2005; Henley et al., 2006; Lyle et al., 2012) has been shown to be a sig-

nificant risk factor with older horses having a higher risk of injury. Male horses have also been shown to have a higher risk of injury (Estberg et al., 1996a; Estberg et al., 1998a,b; Hernandez et al., 2001; Hernandez et al., 2005). The prior racing history of a horse was also found to be associated with injuries (Estberg et al., 1995; Hernandez et al., 2001; Hernandez et al., 2005; Parkin et al., 2005; Henley et al., 2006; Boden et al., 2007; Lyle et al., 2012). If there was an extended interval since the last race the risk for catastrophic injury was higher (Hernandez et al., 2001; Hernandez et al., 2005). Lyle et al. (2012) found that the more starts a horse had within the last 60 days the less the risk of injury. Henley et al. (2006) also found a decrease in risk of injury the more starts a horse had during the year prior a race but Boden et al. (2007) looking specifically at the starts in the 31–60 days period prior the race, found a higher risk for fatal injuries if the horse had a start. The risk of fracture was also higher for horses that did no gallop work during training (Parkin et al., 2005) but horses that accumulated an excess timed work distance within a 2 month period prior a race were at higher risk as well (Estberg et al., 1996a). Exercise history (Estberg et al., 1996a,b; Estberg et al., 1998a,b) (Cohen et al., 2000) (Hernandez et al., 2005; Parkin et al., 2005) and specifically the distance galloped in training

---

Abbreviations: AIC, Akaike's information criterion; CI, confidence interval; EID, equine injury database.

<sup>\*</sup> Corresponding author.

E-mail address: [s.georgopoulos.1@research.gla.ac.uk](mailto:s.georgopoulos.1@research.gla.ac.uk) (S.P. Georgopoulos).

(Estberg et al., 1995; Estberg et al., 1996a,b; Estberg et al., 1998a,b; Cohen et al., 2000; Parkin et al., 2004a) have also been associated with injuries.

Furthermore, prerace condition of a horse; horses that were reluctant to start a race, (Parkin et al., 2006) inspection by regulatory veterinarians (Cohen et al., 1997) and shoe characteristics have been identified to be associated with equine injuries (Kane, 1996; Kane et al., 1998).

Finally, there seem to be risk factors directly related to the race course. The racing surface and its conditions have been shown to be associated with injuries (Hernandez et al., 2001) (Williams et al., 2001; Parkin et al., 2004a,b; Parkin et al., 2005; Henley et al., 2006), the distance of the race (Peloso et al., 1994; Parkin et al., 2004b; Henley et al., 2006; Boden et al., 2007; Lyle et al., 2012), the field size (Parkin et al., 2004b; Parkin et al., 2005; Lyle et al., 2012), the type of the race (Estberg et al., 1998a,b; Henley et al., 2006) and even the season the race took place in (Boden et al., 2007; Lyle et al., 2012).

These studies provided a starting point for the analysis of our study. We aim to identify the risk factors associated with equine fractures in the USA and Canada for 2009–2014. We also aim to make use of logistic regression models to quantify the probability of a Thoroughbred sustaining a fracture during flat racing and identify a population of horses at higher risk. This could inform the design and implementation of preventive measures aimed at minimising the number of Thoroughbreds sustaining fractures during racing in North America.

## 2. Materials and methods

The analysis reported in this paper is an observational retrospective cohort study based on racecourses reporting injuries to the Equine Injury Database (EID) from 1st January 2009 to 31st December 2014. The injury reports are recorded into the EID by veterinarians at the participating racetrack. The data were supplied by The Jockey Club and covered all tracks that voluntarily contributed to the EID in each year.

An equine fracture, sustained during the race, was specified as the outcome variable of this study. This definition included any possible fracture sustained by the horse during racing, including fatal and non-fatal fractures. Fractures were recorded by veterinarians at the track. A case start was defined as a start in a race in which the horse sustained a fracture. All official starts were used to calculate historical information for each horse but only the starts from the racecourses reporting injuries were used to assess the risk of equine fractures. Furthermore, approximately 3,990,000 workout starts made by the Thoroughbreds that raced during that period were also included in the analysis. Workout starts are timed exercise starts a horse makes during training observed on the track.

Risk factors specific to the North American jurisdiction have also been included such as a horse having previously entered the veterinarian's list, a list used by association and regulatory veterinarians to provide horses with illness, injury or soundness issues a brief respite from racing. More than 75% of horses exit the veterinary list within a month. Moreover, the number of times a horse has been previously scratched (withdrawn from the race on the day of the event) has been included as a risk factor. Furthermore, the betting odds of each horse, the ratio of payoff to stake, has been included as a risk factor. The odds rank of each horse was also included. Within a race we numerically ranked each horse according to their betting odds, the number 1 given to the horse favoured by the odds to win the race.

To find the association between the possible risk factors and fractures we created a univariable logistic regression model for each risk factor. The risk factors that were found to be associated

with a p-value less than 0.20, were chosen to be included in a stepwise selection process. The p-value threshold of 0.20 was chosen to prevent the exclusion of a predictor that only becomes evident when possible confounders have been accounted for (Dohoo et al., 2003). For the stepwise selection we used a forward bidirectional elimination approach with Akaike's Information Criterion (AIC). AIC uses maximum likelihood to compare between models with different risk factors, with an added penalty for complexity which depends upon the number of parameters used to fit the data. The best model, following Occam's razor, is the one offering the highest information gain with least complexity (Bozdogan, 1987). We arrived at the final multivariable logistic regression models by including the risk factors selected through this process. We relied only on the AIC for including risk factors in the models and did not use any other exclusion criteria based on potential biological interaction or effect modification.

To assess the effect of clustering we created mixed-effect models that examined the potential effect of including horse as a random effect.

One model was created using all available starts and another multivariable logistic regression model was created on a sub-sample of the population consisting only of starts from horses that had been in racing for at least six months. This effectively includes every racing start of a horse, excluding the starts in the first six months of its career. This extra model was created to assess the risk factors that summarize historical racing information prior to each race.

To quantify the probability of a Thoroughbred sustaining a fracture and to obtain predictions, we created new multivariable models, including the risk factors already obtained, on data available only from the period 2009 - 2013. The data from 2014 were used to evaluate the predictive performance of each model. We used the Area Under the Receiver Operating Characteristic Curve (AUC), as suggested by Bradley (1997), to evaluate the models. This evaluation takes into account the extremely imbalanced outcome variable. Since the prevalence of a fracture is very low it is trivial for a model to achieve an accuracy of over 99.7% by simply predicting a 0% probability of an equine fracture at every start. A bootstrap with 10,000 iterations was used on the racing starts of 2014 to obtain the 95% confidence intervals on the AUC estimate.

All the statistical analyses and calculations in this study were conducted using RStudio, developed by RStudio Team (2015), and the R programming language by the R Development Core Team (2008).

## 3. Results

The analysis reported in this paper is based on data available for 171,523 Thoroughbreds racing at 89 racecourses in the United States and Canada. It includes 2,201,152 starts from the racecourses reporting injuries to the EID. The EID contains information for most races that took place in North America during that period and is a near census collection of the available data. These starts represent 91% of all official racing starts (2,429,443) in the USA and Canada for that period. The incidence of equine fractures during the period 2009–2014 was 2 per 1000 starts.

The results reported are from the fixed-effect only models as they were near identical to the models that included horse as a random effect. The absolute difference was less than 0.01 in the odds ratio for the fixed effects and no meaningful changes for the P-values.

The results from a final multivariable model for all horses are shown in Table 1. A total of 17 risk factors were included in the model.

**Table 1**

Risk factors for fractures - multivariable model - all racing starts in North America from 2009 to 2014 - 2,201,152 starts - 171,523 horses.

Risk factor	Odds ratio	95% CI	P-value
(Intercept)	0.002	0.002–0.003	<0.001
Age (years)	1.027	1.001–1.053	0.043
Country			
Canada	Ref	Ref	Ref
USA	1.360	1.191–1.554	<0.001
Entered the vet list			
No	Ref	Ref	Ref
Yes	1.788	1.670–1.914	<0.001
First Start			
No	Ref	Ref	Ref
Yes	0.748	0.652–0.858	<0.001
Sex			
Mare/Gelding	Ref	Ref	Ref
Stallion	1.473	1.363–1.592	<0.001
Low purse race (< = \$7500)			
No	Ref	Ref	Ref
Yes	0.765	0.697–0.839	<0.001
Months since last racing or workout start	1.203	1.169–1.239	<0.001
Months in racing	0.987	0.983–0.991	<0.001
No. of previous injuries	1.353	1.182–1.549	<0.001
No. of previous non-vet scratches	1.039	1.020–1.058	<0.001
Odds rank in race	0.941	0.931–0.952	<0.001
Purse change since previous race			
None (between drop of \$1000 and raise of \$1000)	Ref	Ref	Ref
Moderate Drop (between \$1000 and \$15000)	1.070	0.992–1.154	0.081
Moderate Raise (between \$1000 and \$15000)	0.939	0.863–1.022	0.147
Large Drop (more than \$15000)	1.092	0.968–1.233	0.153
Large Raise (more than \$15000)	1.109	0.967–1.271	0.138
Race distance (furlongs)	0.941	0.920–0.964	<0.001
Start with new trainer			
No	Ref	Ref	Ref
Yes	1.087	0.984–1.201	0.099
Surface			
Synthetic	Ref	Ref	Ref
Dirt	1.328	1.196–1.474	<0.001
Turf	1.276	1.121–1.452	<0.001
Time with same trainer (months)	0.989	0.984–0.995	<0.001
Training with first trainer			
Yes	Ref	Ref	Ref
No	1.076	0.990–1.171	0.086

For the population of horses that have raced for at least six months (Table 2), we were able to include, historical information of up to six months prior to a race and 26 risk factors were included in a final multivariable model. This includes all 1,438,992 starts from 114,996 Thoroughbreds that were made at least six months after their first racing start.

When looking at the models based on the entire study population, stallions were found to have a higher risk of sustaining a fracture than mares and geldings (OR: 1.473,  $p < 0.001$ ). Older horses were also found to have a higher risk (OR: 1.027,  $p: 0.043$ ). Horses that have entered the veterinary list were found to be at a significant higher risk (OR: 1.788,  $p < 0.001$ ).

Regarding race related factors, it was found that starts from races in the USA were at higher risk compared to starts from races in Canada (OR: 1.360,  $p < 0.001$ ). Racing surfaces were also shown to be associated with fractures. Synthetic surfaces were found to have the lowest risk compared to turf (OR: 1.276,  $p < 0.001$ ) and dirt surfaces (OR: 1.328,  $p < 0.001$ ).

When looking specifically at the population of horses that have raced for at least six months we were able to include historical information of up to six months prior to a race and we found that the more racing and workout starts horses made up to a month prior to the race the less the risk per start of sustaining a fracture (OR: 0.713,  $p < 0.001$ ). The same association was found when looking at racing and workout starts between 30 and 60 days prior to the race (OR: 0.953,  $p: 0.025$ ). On the other hand, horses with more racing and workout starts between 60 and 90 days prior to a start were at more risk per start (OR: 1.058,  $p: 0.002$ ). Similarly, more starts

between 90 and 180 days prior to a start were associated with an increased risk (OR: 1.069,  $p < 0.001$ ).

Furthermore, we were able to use the information available in the data to create predictive models and identify horses at lower and higher risk than the average. The models were created using the starts from 2009 to 2013 and predicted the risk for all starts in 2014. We evaluated the predictive performance of our models by calculating the AUC which was 65.2% (95% CI: 63.2%–67.3%). The 5% of starts that had the highest score in our models for 2014 were found to have 2.4 times (95% CI: 1.9–2.9) higher fracture prevalence than the fracture prevalence of 2014. Contrary to this, the 5% of starts that had the lowest score were found to have approximately 1/3 the risk (95% CI: 0.1–0.5) of the mean fracture prevalence of 2014.

#### 4. Discussion

This study has made extensive use of the data available from 2009 to 2014 in the Equine Injury Database regarding fracture injuries. Risk factors plausibly linked with fractures have been assessed and those with a statistically significant association have been identified.

The study showed that stallions have a higher chance of sustaining a fracture than geldings and mares. Stallions in California were found, in a previous study, to be at a higher risk of sustaining fracture of the forelimb proximal sesamoid bones (Anthenill et al., 2007). Stallions have also been found to be at higher risk of sustaining a catastrophic musculoskeletal injury (Estberg et al., 1998a,b),

**Table 2**

Risk factors for fractures - multivariable model - all starts after the first 6 months in racing in North America from 2009 to 2014 - 1,438,992 starts - 114,996 horses.

Risk factor	Odds ratio	95% CI	P-value
(Intercept)	0.002	0.001–0.004	<0.001
Age (years)	1.036	1.001–1.072	0.041
Average speed in previous race (m/s)	1.033	0.998–1.069	0.061
Country			
Canada	Ref	Ref	Ref
USA	1.320	1.096–1.589	0.003
Entered the vet list			
No	Ref	Ref	Ref
Yes	1.741	1.595–1.900	<0.001
Sex			
Mare/Gelding	Ref	Ref	Ref
Stallion	1.634	1.471–1.814	<0.001
Low purse race (< = \$7500)			
No	Ref	Ref	Ref
Yes	0.740	0.660–0.830	<0.001
Months in racing	0.988	0.984–0.992	<0.001
Months since last racing or workout start	1.090	1.030–1.153	0.003
No. of previous injuries	1.332	1.156–1.534	<0.001
No. of previous non-vet scratches	1.025	1.005–1.045	0.013
No. of previous vet scratches	0.965	0.920–1.013	0.149
No. of racing and workout starts (Present–30 days prior race)	0.713	0.677–0.750	<0.001
No. of racing starts (Present–30 days prior race)	0.948	0.889–1.010	0.099
No. of racing and workout starts (30–60 days prior race)	0.953	0.915–0.994	0.025
No. of racing starts (30–60 days prior race)	0.943	0.889–1.000	0.048
No. of racing and workout starts (60–90 days prior race)	1.058	1.021–1.096	0.002
No. of racing and workout starts (90–180 days prior race)	1.069	1.053–1.085	<0.001
Odds rank in race	0.946	0.931–0.961	<0.001
Purse change since previous race			
None (between drop of \$1000 and raise of \$1000)	Ref	Ref	Ref
Moderate Drop (between \$1000 and \$15,000)	1.057	0.963–1.161	0.242
Moderate Raise (between \$1000 and \$15,000)	0.891	0.802–0.991	0.033
Large Drop (more than \$15,000)	1.181	1.015–1.375	0.032
Large Raise (more than \$15,000)	1.074	0.896–1.286	0.442
Race distance (furlongs)	0.926	0.898–0.954	<0.001
Surface			
Synthetic	Ref	Ref	Ref
Dirt	1.349	1.166–1.562	<0.001
Turf	1.242	1.044–1.479	0.015
Time with same trainer (months)	0.988	0.983–0.993	<0.001
Wins/starts (Present–30 days prior race)	1.002	1.000–1.003	0.027
Wins/starts (30–60 days prior race)	1.002	1.001–1.003	0.004
Wins/starts (60–90 days prior race)	1.001	1.000–1.002	0.091
Wins/starts (90–180 days prior race)	1.002	1.000–1.004	0.019

a non-fatal superficial digital flexor tendon injury (Takahashi et al., 2004), or any form of fatal injury (Estberg et al., 1996a,b). In contrast, a case-control study of serious musculoskeletal injuries of Thoroughbreds in Australia (Bailey et al., 1997) found no significant association between the sex of the horse and injuries. It is unclear though if geldings were included in the male category, potentially confounding the result.

In our study a higher risk of injury was found for older horses. The age of the horse has been shown to be associated with the risk of injury in a number of studies (Estberg et al., 1996b; Bailey et al., 1997; Cohen et al., 1997; Cohen et al., 2000; Parkin et al., 2004a; Perkins et al., 2005). This consistent association might be because older horses continue to accrue microdamage and injury risk.

There is a strong indication that horses expected to win a race were at higher risk of sustaining a fracture. Those horses that were expected to perform well in a race, as indicated by their “odds rank” risk factor, were found to be at higher risk. Also, races with a lower than \$7500 purse were found to have reduced likelihood of sustaining a fracture. The low purse in this context might be a proxy for the quality of the horses, thus lowering the risk, as well as the competitiveness of the race. A similar finding of higher risk for the more competitive stakes races was shown in a case-control study of serious musculoskeletal injuries of Thoroughbreds in Australia (Bailey et al., 1997).

Our study showed that for the first couple of months immediately prior to the race, the more racing and training starts, the lower the risk for the horse but the association changed for the periods of two months up to six months prior to the race. It was also found that the more months a horse stays away from racing or training, the higher its risk in a new racing start. A study by Anthenill et al. (2007) also found that an increase in the number of workouts increased the risk of proximal sesamoid bone fractures. A case-control study of Thoroughbreds racing in California (Estberg et al., 1996a) found that an increase in cumulative exercise and race distance over the previous two months was associated with an increased risk of fatal skeletal injury. Previous studies have shown an almost eightfold increase in the risk of sustaining a superficial digital flexor tendon injury for Thoroughbreds that took a break of more than two months (Cohen et al., 1997) and Perkins et al. (2005) found that horses with no starts were at a higher risk of sustaining a non-fatal injury to the superficial digital flexor tendon and to the suspensory apparatus.

It was found that the risk of sustaining a fracture was higher if the horse has entered the veterinary list in its career. This is in accordance with previous studies that showed a 5-fold to 14-fold increase in risk for musculoskeletal injury, injury of the suspensory apparatus and of the superficial digital flexor tendon among horses assessed to be at increased risk by regulatory veterinarians conducting pre-race examinations (Cohen et al., 1997).

Horses racing on a dirt surface were at higher risk of sustaining a fracture. This is in accordance with a previous study of equine fatalities (Arthur, 2010) that reported that dirt had the highest incidence rate followed by turf and synthetic surfaces. Another study (Mohammed et al., 1991) also found that dirt courses are associated with higher risk compared to turf. This association might be due to increased forces acting on the limb on dirt surfaces compared to turf and synthetic ones.

It is unclear why starts made in the USA have a higher risk rate compared to starts in Canada. This is an interesting finding that needs to be examined in further studies. Higher risk might be due to different training regimes and racing schedules followed by horses in the two countries.

The study looked at the workout history of each horse and used the number of workout starts prior to a race as a proxy for increased cumulative exercise. The examination in future studies of management practices and type of exercise might yield further insight as to how a horse's training regime is associated with the risk of injury during racing.

We believe the identified risk factors are as unbiased as possible, since we have included in the statistical analysis 91% of racing starts from all official racing in the USA and Canada for that period. A small source of bias could be the roughly 9% of starts which are not included in this study.

Currently, the models developed are not sufficiently predictive to accurately identify a case start. This is most likely due to the fact that data relating to unidentified risk factors such as previous medical and treatment history and detailed training data were not available during these analyses. It is also very likely that other unknown variables exist that will never be identified and that simple random or chance occurrence of a fracture injury is likely to be an ever present.

The predictive performance of the models in this study reaches an AUC of approximately 65%. This figure needs careful interpretation. It is not stating that these models will correctly identify 65% of all horses sustaining a fracture injury. It is stating that, if we were to choose between two horses entered into the same race, one of which we knew was about to end the race, 65% of the time we would identify the correct horse (i.e. the one about to sustain a fracture). This demonstrates how far we have to go and how important it is to be able to utilise other likely critical, before predictions can be safely made. However, this is not to say that these models could not be used to aid decision making by trainers (and owners) over the future race entries and longer term careers of their horses. The identification of a significantly increased risk of fracture injury when a particular horse is entered into a particular race may be used to redirect that horse into a race with a safer risk profile or indeed to reach a point where retirement was the most sensible next step for that horse.

This study looked broadly at fractures as an outcome variable and no attempt was made to differentiate between different causes of injury. As risk factors for different fracture types have been shown to differ (Parkin et al., 2004a,b; Anthenill et al., 2007), we recognise that it is likely that some specific risk factors associated with a specific type of fracture may not have been identified in the current analysis. Over recent years, the different types of injury sustained have become more accurately reported to the EID. This will enable future analyses to use more specific outcome variables that will identify risk factors that are unique to a more specific injury type.

We believe that current strategies and a greater awareness of the importance of equine welfare in the racing industry in North America will result in the development of strategies to reduce risk. Small, positive changes in the prevalence of known risk factors are now being identified which are likely to gradually reduce the risk of Thoroughbred fracture injury.

In conclusion, the results of this study could help to inform the design and implementation of preventive measures aimed at minimising the number of Thoroughbreds sustaining fractures during racing in North America. We are working with The Jockey Club on identifying the most suitable way to translate these results into advice for stakeholders.

## Acknowledgements

This work is funded through an Industry Partnership PhD provided by The Jockey Club and the University of Glasgow.

## References

- Anthenill, L.A., Stover, S.M., Gardner, I.A., Hill, A.E., 2007. Risk factors for proximal sesamoid bone fractures associated with exercise history and horseshoe characteristics in Thoroughbred racehorses. *Am. J. Vet. Res.* 68 (7), 760–771.
- Arthur, R.M., 2010. Comparison of racing fatality rates on dirt, synthetic, and turf at four California racetracks. *Proceedings of the Annual Convention of the AAEP.*
- Bailey, C., Reid, S., Hodgson, D., Suann, C., Rose, R., 1997. Risk factors associated with musculoskeletal injuries in Australian Thoroughbred racehorses. *Prev. Vet. Med.* 32, 47–55.
- Boden, L.A., Anderson, G.A., Charles, J.A., Morgan, K.L., Morton, J.M., Parkin, T.D., Clarke, A.F., Slocombe, R.F., 2007. Risk factors for Thoroughbred racehorse fatality in flat starts in Victoria, Australia (1989–2004). *Equine Vet. J.* 39, 430–437.
- Bozdogan, H., 1987. Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical expressions. *Psychometrica* 52 (3), 345–370.
- Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn.* 30 (7), 1145–1159.
- Cohen, N., Peloso, J., Mundy, G., Fisher, M., Holland, R., Little, T., Misheff, M.M., Watkins, J.P., Honnas, C.M., Moyer, W., 1997. Racing-related factors and results of pre-race physical inspection and their association with musculoskeletal injuries incurred in Thoroughbreds during races. *J. Am. Vet. Med. Assoc.* 211 (4), 454–463.
- Cohen, N.D., Berry, S.M., Peloso, J.G., Mundy, G.D., Howard, I.C., 2000. Association of high-speed exercise with racing injury in Thoroughbreds. *J. Am. Vet. Med. Assoc.* 216 (8), 1273–1278.
- Dohoo, I., Martin, W., Stryhn, H., 2003. *Veterinary Epidemiologic Research.* AVC Inc, Charlottetown.
- Estberg, L., Gardner, I.A., Stover, S.M., Johnson, B., Case, J.T., Ardans, A., 1995. Cumulative racing-speed exercise distance cluster as a risk factor for fatal musculoskeletal injury in Thoroughbred racehorses in California. *Prev. Vet. Med.* 24 (4), 253–263.
- Estberg, L., Stover, S., Gardner, I., Drake, C., Johnson, B., Ardans, A., 1996a. High-speed exercise history and catastrophic racing fracture in thoroughbreds. *J. Am. Vet. Med. Assoc.* 57 (11), 1549–1555.
- Estberg, L., Stover, S., Gardner, I., Johnson, B., Case, J., Ardans, A., Read, D.H., Anderson, M.L., Barr, B.C., Daft, B.M., Kinde, H., Moore, J., Stoltz, J., Woods, L., 1996b. Fatal musculoskeletal injuries incurred during racing and training in thoroughbreds. *J. Am. Vet. Med. Assoc.* 208 (1), 92–96.
- Estberg, L., Gardner, I.A., Stover, S.M., Johnson, B.J., 1998a. A case-crossover study of intensive racing and training schedules and risk of catastrophic musculoskeletal injury and lay-up in California Thoroughbred racehorses. *Prev. Vet. Med.* 33, 159–170.
- Estberg, L., Stover, S., Gardner, I., Johnson, B., Jack, R., Case, J., Ardans, A., Read, D.H., Anderson, M.L., Barr, B.C., Daft, B.M., Kinde, H., Moore, J., Stoltz, J., Woods, L., 1998b. Relationship between race start characteristics and risk of catastrophic injury in thoroughbreds: 78 cases (1992). *J. Am. Vet. Med. Assoc.* 212 (4), 544–549.
- Henley, W., Rogers, K., Harkins, L., Wood, J.L., 2006. A comparison of survival models for assessing risk of racehorse fatality. *Prev. Vet. Med.* 74, 3–20.
- Hernandez, J., Hawkins, D.L., Scollay, M.C., 2001. Race-start characteristics and risk of catastrophic musculoskeletal injury in Thoroughbred racehorses. *J. Am. Vet. Med. Assoc.* 218, 83–86.
- Hernandez, J.A., Scollay, M.C., Hawkins, D.L., Corda, J.A., Krueger, T.M., 2005. Evaluation of horseshoe characteristics and high-speed exercise history as possible risk factors for catastrophic musculoskeletal injury in Thoroughbred racehorses. *Am. J. Vet. Res.* 66, 1314–1320.
- Kane, A.J., Stover, S.M., Gardner, I.A., Bock, K.B., Case, J.T., Johnson, B.J., Anderson, M.L., Barr, B.C., Daft, B.M., Kinde, H., Larochelle, D., Moore, J., Mysore, J., Stoltz, J., Woods, L., Read, D.H., Ardans, A.A., 1998. Hoof size, shape, and balance as possible risk factors for catastrophic musculoskeletal injury of Thoroughbred racehorses. *Am. J. Vet. Res.* 59, 1545–1552.

- Kane, A.J., Stover, S.M., Gardner, I.A., Case, J.T., Johnson, B.J., Read, D.H., Ardans, A.A.1, 1996. [Horseshoe characteristics as possible risk factors for fatal musculoskeletal injury of thoroughbred racehorses](#). *Am. J. Vet. Res.* 57 (8), 1147–1152.
- Lyle, C.H., Blissitt, K.J., Kennedy, R.N., McGorum, B.C., Newton, J.R., Parkin, T.D., Stirk, A., Boden, L.A., 2012. [Risk factors for race-associated sudden death in Thoroughbred racehorses in the UK \(2000–2007\)](#). *Equine Vet. J.* 44 (4), 459–465.
- Mohammed, H.O., Hill, T., Lowe, J., 1991. [Risk factors associated with injuries in Thoroughbred horses](#). *Equine Vet. J.* 23 (6), 445–448.
- Parkin, T.D., Clegg, P.D., French, N.P., Proudman, C.J., Riggs, C.M., Singer, E.R., Webbon, P.M., Morgan, K.L., 2004a. [Horse-level risk factors for fatal distal limb fracture in racing Thoroughbreds in the UK](#). *Equine Vet. J.* 36 (6), 513–519.
- Parkin, T.D., Clegg, P.D., French, N.P., Proudman, C.J., Riggs, C.M., Singer, E.R., Webbon, P.M., Morgan, K.L., 2004b. [Race- and course-level risk factors for fatal distal limb fracture in racing Thoroughbreds](#). *Equine Vet. J.* 36 (6), 521–526.
- Parkin, T.D., Clegg, P.D., French, N.P., Proudman, C.J., Riggs, C.M., Singer, E.R., Webbon, P.M., Morgan, K., 2005. [Risk factors for fatal lateral condylar fracture of the third metacarpus/metatarsus in UK racing](#). *Equine Vet. J.* 37, 192–199.
- Parkin, T.D., Clegg, P.D., French, N.P., Proudman, C.J., Riggs, C.M., Singer, E.R., Webbon, P.M., Morgan, K.L., 2006. [Analysis of horse race videos to identify intra-race risk factors for fatal distal limb fracture](#). *Prev. Vet. Med.* 74, 44–55.
- Peloso, J.G., Mundy, G.D., Cohen, N., 1994. [Prevalence of, and factors associated with: musculoskeletal racing injuries of Thoroughbreds](#). *J. Am. Vet. Med. Assoc.* 204, 620–626.
- Perkins, N.R., Reid, S.W., Morris, R.S., 2005. [Risk factors for injury to the superficial digital flexor tendon and suspensory apparatus in Thoroughbred racehorses in New Zealand?](#) *N. Z. Vet. J.* 53 (3), 184–192.
- Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051-07-0, URL <http://www.R-project.org>.
- RStudio Team, 2015. *RStudio: Integrated Development for R*. RStudio Inc Boston, MA <http://www.rstudio.com>.
- Takahashi, T., Kasashima, Y., Ueno, Y., 2004. [Association between race history and risk of superficial digital flexor tendon injury in Thoroughbred racehorses](#). *J. Am. Vet. Med. Assoc.* 225 (1), 90–93.
- Williams, R.B., Harkins, L.S., Hammond, C.J., Wood, J., 2001. [Racehorse injuries, clinical problems and fatalities recorded on British racecourses from flat racing and National Hunt racing during 1996, 1997 and 1998](#). *Equine Vet. J.* 33 (5), 478–486.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary Table S1. Risk factors for fractures - univariable model - all racing starts in North America from 2009 to 2014 - 2,201,152 starts - 171,523 horses

Risk factor	OR (95% CI)	Starts - Cases	P-value
Age (years)	0.964 (0.945-0.983)	2,201,152 - 4,558	< 0.001
Age at first start (years)	1.004 (0.979-1.029)	2,201,152 - 4,558	0.769
Country			
Canada	Ref	178,683 - 239	Ref
USA	1.598 (1.403-1.820)	2,022,469 - 4,319	< 0.001
Entered the vet list			
No	Ref	1,811,710 - 3,311	Ref
Yes	1.754 (1.643-1.873)	389,442 - 1,247	< 0.001
Field size	0.992 (0.977-1.007)	2,201,152 - 4,558	0.300
First Start			
No	Ref	2,037,651 - 4,279	Ref
Yes	0.812 (0.719-0.917)	163,501 - 279	0.001
Low purse race (<= \$7500)			
No	Ref	1,867,144 - 3,982	Ref
Yes	0.808 (0.741-0.882)	334,008 - 576	< 0.001
Months in racing	0.993 (0.991-0.996)	2,201,152 - 4,558	0.402
Months since last racing start	0.983 (0.967-.999)	2,201,152 - 4,558	< 0.001
Months since last racing or workout start	1.188 (1.156-1.221)	2,201,152 - 4,558	< 0.001
No. of previous injuries	1.494 (1.311-1.702)	2,201,152 - 4,558	< 0.001
No. of previous vet scratches	1.061 (1.023-1.100)	2,201,152 - 4,558	0.001
No. of previous non-vet scratches	1.022 (1.006-1.038)	2,201,152 - 4,558	0.006
Odds at start of race	0.993 (0.991-0.995)	2,201,152 - 4,558	< 0.001
Odds rank in race	0.935 (0.925-0.946)	2,201,152 - 4,558	< 0.001
Post position	0.997 (0.986-1.008)	2,201,152 - 4,558	0.608
Purse change since previous race			
None (between drop of \$1000 and raise of \$1000)	Ref	982,386 - 1,909	Ref
Moderate Drop (between \$1000 and \$15000)	1.192 (1.108-1.281)	514,033 - 1,190	< 0.001



Moderate Raise (between \$1000 and \$15000)	1.010 (0.932-1.094)	445,870 - 875	0.809
Large Drop (more than \$15000)	1.220 (1.086-1.371)	140,030 - 332	0.001
Large Raise (more than \$15000)	1.091 (0.957-1.245)	118,833 - 252	0.192
Purse (\$1000)	1.000 (0.999-1.001)	2,201,152 - 4,558	0.403
Race distance (furlongs)	0.946 (0.925-0.968)	2,201,152 - 4,558	< 0.001
Sex			
Mare/Gelding	Ref	1,924,792 - 3,744	Ref
Stallion	1.516 (1.405-1.635)	276,360 - 814	< 0.001
Start with new jockey			
No	Ref	1,061,213 - 2,122	Ref
Yes	1.069 (1.008-1.133)	1,139,939 - 2,436	0.025
Start with new trainer			
No	Ref	1,991,675 - 4,007	Ref
Yes	1.308 (1.197-1.430)	209,477 - 551	< 0.001
Surface			
Synthetic	Ref	269,523 - 411	Ref
Dirt	1.427 (1.289-1.581)	1,630,597 - 3,547	< 0.001
Turf	1.308 (1.153-1.483)	301,032 - 600	< 0.001
Time with same jockey (months)	0.990 (0.976-1.004)	2,201,152 - 4,558	0.173
Time with same trainer (months)	0.981 (0.977-0.985)	2,201,152 - 4,558	< 0.001
Training with first trainer			
Yes	Ref	1,202,188 - 2,354	Ref
No	1.127 (1.063-1.195)	998,964 - 2,204	< 0.001

Supplementary Table S2. Risk factors for fractures - univariable model - all starts after the first 6 months in racing in North America from 2009 to 2014 - 1,438,992 starts - 114,996 horses

Risk factor	OR (95% CI)	Starts - Cases	P-value
Age (years)	0.963 (0.939-0.989)	1,438,992 - 2,863	0.005
Age at first start (years)	1.014 (0.980-1.048)	1,438,992 - 2,863	0.431
Average speed change on previous race (m/s)	1.007 (0.986-1.028)	1,438,992 - 2,863	0.531
Average speed in previous race (m/s)	1.080 (1.036-1.127)	1,438,992 - 2,863	< 0.001
Country			

Canada	Ref	112,743 - 123	Ref
USA	1.896 (1.582-2.271)	1,326,249 - 2,740	< 0.001
Entered the vet list			
No	Ref	1,123,526 - 1,910	Ref
Yes	1.779 (1.646-1.923)	315,466 - 953	< 0.001
Field size	0.996 (0.977-1.016)	1,438,992 - 2,863	0.702
Low purse race (<= \$7500)			
No	Ref	1,197,168 - 2.482	Ref
Yes	0.756 (0.681-0.846)	241,824 - 381	< 0.001
Months in racing	0.992 (0.989-0.995)	1,438,992 - 2,863	< 0.001
Months since last racing start	0.974 (0.956-0.992)	1,438,992 - 2,863	0.005
Months since last racing or workout start	1.176 (1.139-1.214)	1,438,992 - 2,863	< 0.001
No. of previous injuries	1.516 (1.323-1.737)	1,438,992 - 2,863	< 0.001
No. of previous vet scratches	1.065 (1.024-1.108)	1,438,992 - 2,863	0.002
No. of previous non-vet scratches	1.028(1.011-1.045)	1,438,992 - 2,863	0.001
No. of racing and workout starts (Present – 30 days prior race)	0.702 (0.674-0.731)	1,438,992 - 2,863	< 0.001
No. of racing and workout starts (30 -60 days prior race)	0.911 (0.881-0.942)	1,438,992 - 2,863	< 0.001
No. of racing and workout starts (60 -90 days prior race)	1.046 (1.015-1.078)	1,438,992 - 2,863	0.003
No. of racing and workout starts (90 -180 days prior race)	1.068 (1.055-1.080)	1,438,992 - 2,863	< 0.001
No. of starts (Present – 30 days prior race)	0.827 (0.783-0.873)	1,438,992 - 2,863	< 0.001
No. of starts (30 – 60 days prior race)	1.065 (1.018-1.115)	1,438,992 - 2,863	0.006
No. of starts (60 – 90 days prior race)	1.152 (1.102-2.204)	1,438,992 - 2,863	< 0.001
No. of starts (90 – 180 days prior race)	1.104 (1.082-1.126)	1,438,992 - 2,863	< 0.001
Odds at start of race	0.991 (0.989-0.993)	1,438,992 - 2,863	< 0.001
Odds rank in race	0.927 (0.913-0.940)	1,438,992 - 2,863	< 0.001
Post position	1.000 (0.986-1.014)	1,438,992 - 2,863	0.994
Purse change since previous race			
None (between drop of \$1000 and raise of \$1000)	Ref	588,198 - 1,110	Ref
Moderate Drop (between \$1000 and \$15000)	1.167 (1.066-1.279)	361,786 - 797	< 0.001
Moderate Raise (between \$1000 and \$15000)	0.977 (0.883-1.080)	316,163 - 583	0.651

Large Drop (more than \$15000)	1.230 (1.065-1.421)	95,671 - 222	0.005
Large Raise (more than \$15000)	1.037 (0.875-1.229)	77,174 - 151	0.676
Purse (\$1000)	1.000 (0.999-1.001)	1,438,992 - 2,863	0.890
Race distance (furlongs)	0.924 (0.898-0.950)	1,438,992 - 2,863	< 0.001
Sex			
Mare/Gelding	Ref	1,299,756 - 2,430	Ref
Stallion	1.665 (1.503-1.845)	139,236 - 433	< 0.001
Start with new jockey			
No	Ref	624,041 - 1,211	Ref
Yes	1.045 (0.970-1.125)	814,951 - 1,652	0.248
Start with new trainer			
No	Ref	1,278,862 - 2,453	Ref
Yes	1.336 (1.203-1.483)	160,130 - 410	< 0.001
Surface			
Synthetic	Ref	156,056 - 211	Ref
Dirt	1.567 (1.361-1.805)	1,082,064 - 2,291	< 0.001
Turf	1.330 (1.122-1.576)	200,872 - 361	< 0.001
Time with same jockey (months)	0.987 (0.972-1.003)	1,438,992 - 2,863	0.114
Time with same trainer (months)	0.978 (0.974-0.983)	1,438,992 - 2,863	< 0.001
Training with first trainer			
Yes	Ref	535,212 - 931	Ref
No	1.229 (1.137-1.329)	903,780 - 1,932	< 0.001
Wins/starts (Present – 30 days prior race)	1.002 (1.001-1.003)	1,438,992 - 2,863	0.005
Wins/starts (30 – 60 days prior race)	1.004 (1.003-1.005)	1,438,992 - 2,863	< 0.001
Wins/starts (60 – 90 days prior race)	1.003 (1.002-1.005)	1,438,992 - 2,863	< 0.001
Wins/starts (90 – 180 days prior race)	1.006 (1.004-1.007)	1,438,992 - 2,863	< 0.001

---