An investigation of equine injuries in Thoroughbred flat racing in North America

Stamatis Panagiotis Georgopoulos

MEng, MSc

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

School of Veterinary Medicine

College of Medical, Veterinary and Life Sciences

University of Glasgow

July 2017

Abstract

The aim of this research work was to investigate and quantify the risk of fatal and fracture injury for Thoroughbreds participating in flat racing in the US and Canada so that horses at particular risk can be identified and the risk of fatal injury reduced. Risk factors associated with fatalities and fractures were identified and predictive models for both fatalities and fractures were developed and their performance was evaluated. Our analysis was based on 188,269 Thoroughbreds that raced on 89 racecourses reporting injuries to the Equine Injury Database (EID) in the US and Canada from 1st January 2009 to 31st December 2015. This included 2,493,957 race starts and 4,592,162 exercise starts. The race starts reported to the EID represented the starts for 90.0% of all official Thoroughbred racing events in the United States and Canada during the 7-year observation period.

The annual average risk of fatal and fracture equine injuries for the period 2009 - 2015 was estimated and a description of the different injury types that resulted in fatalities and fractures was given, based on the cases recorded in the EID.

Possible risk factors were pre-screened using univariable logistic regression models; risk factors with an association indicated by $p < 0.20$ were then included in a stepwise logistic regression selection process. A forward bidirectional elimination approach using Akaike's Information Criterion was utilised for the stepwise selection. We identified more than 20 risk factors that were found to be significantly associated with fatal injury ($p < 0.05$) and more than 20 risk factors associated with fracture injury, across the final multi-variable models. The risk factors identified are related to the horse’s previous racing history, the trainer, the race, the horse's expected performance and the horse's racing history.

Five different algorithms were used to develop predictive models based on the data available from the period 2009 - 2014 for both fatal and fracture injuries. Firstly, we used Multivariable Logistic Regression, commonly used in risk factor analysis. Secondly, Improved Balanced Random Forests were developed, a machine learning algorithm based on a modification of the random forests algorithm. Because fatal injuries are extremely rare events, less than 2 instances per 1000 starts on average, balanced samples were used to develop the Random Forest model to deal with the

class-imbalance problem. Furthermore, we trained an Artificial Neural Network with a single layer and two networks with deep architecture, a Deep Belief Network and a Stacked Denoising Autoencoder. As artificial neural networks and deep learning models have been successfully used to solve complex problems in a diverse field of domains we wanted to explore the possibility of using them to successfully predict equine injuries. The performance of each classifier was evaluated by calculating the Area Under the Receiver Operating Characteristic Curve (AUC), using the data available from 2015 for validation. AUC results ranged from 0.62 to 0.64 for the best performing algorithm and similar predictive results were obtained from the wide array of different models created.

This is the first study to make use of the extensive information contained in the EID to identify risk factors associated with equine fatal and fracture injuries in the US and Canada for this period. To our knowledge, this is the largest retrospective observational study investigating the risk of equine fatal and fracture injuries during flat racing in the literature. This is also the first study to train logistic regression and machine learning models to predict equine injuries using such an extensive amount of data and a full year of horse racing events for prediction and evaluation.

We believe the results could help identify horses at high risk of (fatal) injury on entering a race and inform the design and implementation of preventive measures aimed at minimising the number of Thoroughbreds sustaining fatal injuries during racing in North America.

Table of Contents

List of Tables

List of Figures

List of Publications

1. Georgopoulos, S.P. and Parkin, T.D., 2016. Risk factors associated with fatal injuries in Thoroughbred racehorses competing in flat racing in the United States and Canada. Journal of the American Veterinary Medical Association, 249(8), pp.931-939.

2. Georgopoulos, S.P. and Parkin, T.D., 2017. Risk factors for equine fractures in Thoroughbred flat racing in North America. Preventive Veterinary Medicine, 139, pp.99-104.

Acknowledgements

I would like to express my sincere appreciation to Dr. Tim Parkin for his professional assistance and support throughout my studies. His valuable and constructive suggestions have been essential for the development of this research work. I couldn't have asked for a better supervisor.

My grateful thanks are also extended to Dr. Matt Denwood; his insightful approach and assistance provided at the beginning of my research have been much appreciated.

I would also like to thank the US Jockey Club for providing the data and supporting this project through an industry partnership PhD initiative. Without those resources this PhD research would not have been possible. I would also like to thank the College of Medical, Veterinary and Life Sciences and the School of Veterinary Medicine for co-funding my position with the US Jockey Club.

I was also lucky to have the support, advice and critique from my old friends Christos Alevras and Gina Stavropoulou and, of course, my brother Thanasis. They came through when I needed them despite being, in all probability, bored out of their minds with all those stats.

Finally, I wish to thank my parents for their constant support and encouragement.

Author’s Declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Printed Name: Stamatis P. Georgopoulos

Signature: ________________________

Definitions/Abbreviations

AIC	-	Akaike's Information Criterion
AUC	-	Area Under the Curve
CI	-	Confidence Interval
DBN	-	Deep Belief Network
EID	-	Equine Injury Database
IBRF	-	Improved Balanced Random Forests
RBM	-	Restricted Boltzmann Machines
ROC	-	Receiver Operating Characteristic
SDA	-	Stacked Denoising Autoencoders
US	-	United States of America

1. Introduction and Literature Review

1.1. Introduction

The aim of this study was to quantify the risk of fatal and fracture injury for Thoroughbreds participating in flat racing in the United States and Canada from January 1 2009 to December 31 2015. Information on flat races for this period was available in the Equine Injury Database (EID) and was provided by the US Jockey Club.

The Jockey Club was established in New York in 1894. Its mission is to improve Thoroughbred breeding and racing primarily in the United States, Canada, and Puerto Rico.

The Jockey Club launched the EID in July 2008. It is a near census collection of data available for flat races taking place in the US and Canada. Its stated mission is "to identify the frequency, types and outcome of racing injuries using a standardized format that will generate valid statistics, identify markers for horses at increased risk of injury and serve as a data source for research directed at improving safety and preventing injuries."

To this end, this study aimed to:

- provide a description and summary statistics of the data available in the EID on fatal and fracture injuries,
- assess possible risk factors and their association with fatal and fracture injuries from the plethora of variables available in the dataset,
- train predictive models to identify Thoroughbreds at a higher risk of sustaining a fatal injury and fracture injury before entering a flat race.

As is the case for all sports, injuries are a part of the equation. In 1992 (van Mechelen) running injuries were reported in between 37% and 56% of people, while a study on the world athletic championship showed an injury rate amongst athletes of 13.5% (Alonso, et al., 2012). When it comes to horseracing injury rates to the horses participating in the sport are significantly lower but at the same time they can be significantly more severe and in some cases, they may result in the death or euthanasia of the participating horses.

Therefore, extreme care must be taken to prevent situations that might result in the injury of the participating horses. Correctly identifying the risk factors associated with horseracing injuries and accurately measuring their significance could help design future intervention strategies to decrease the risk of sustaining an equine injury in the flat horse racing population. Additionally, being able to identify horses at high risk on entering a race could inform the design and implementation of preventive measures aimed at minimising the number of Thoroughbreds sustaining fatal and fracture injuries during racing in North America.

1.2. Fatal Injuries

The number of fatalities per thousand starts as has currently been reported in the literature, ranges between 0.44 – 1.7 and 4 – 14 per thousand starts in flat and jump racing respectively (Bourke, 1994; Peloso, et al., 1994; Mckee, 1995; Estberg, et al., 1996; Bailey, et al., 1998; Wood, et al., 2000; Stephen, et al., 2003; Boden, et al., 2006; Rosanowski, et al., 2016). Sudden death has been reported to be between 0.08 and 0.29 per 1000 starts in flat and jump races respectively (Boden, et al., 2006) while other studies calculate it to be somewhere in the region of 9% to 12% of the overall fatalities reported (Johnson, et al., 1994; Lyle, et al., 2011). The current fatality rates are an outcome of studies with a varying sample size while some of them did not have the required data to produce confidence intervals, which might explain the different range of fatal injury prevalence reported in those studies. Furthermore, in all of the studies the reported fatalities occurred while the horses were at the racecourse, not taking into account euthanasia cases at a later time, that were a direct result of an injury that was sustained during the race. The factors that affect the decision of euthanizing an injured horse usually are medical or human based. Furthermore, fatality rates are greatly affected by euthanasia as a major proportion of cases are a result of it (Reardon, 2013).

It has frequently been reported that two categories of racehorses have an increased risk of fatality, those two categories being older and male horses. The most plausible reason for the increased risk for those two categories might be the refusal of providing treatment for old racehorses that will not provide a reasonable return on the investment, this is more probable for older male horses not only for the shorter lifespan but also the decreased breeding potential (Reardon, 2013). It is fair to report that although univariable analyses (Boden, et al., 2007a,b) identify age as a significant reason in Australia that seems not to be the case in the final multivariable models and in a US based study (Johnson, et al., 1994) where the risk of 2-year old racehorses training related deaths was higher than for 3-year olds.

The difference between race types is highly associated with different rates of fatalities of racehorses with jump and steeplechases races being those with the higher risk due to the possibility of racehorses acquiring injuries by hitting the obstacles, that are an integral part of those type of races, and falling (Williams, et al., 2001; Pinchbeck 2004; Boden, et al., 2006); the longer distance that the racehorses have to run when competing in those races (Wood, et al., 2000; Hernandez, et al., 2001; Parkin, et al., 2004a; Boden, et al., 2007a); and the difference in age in the population that competes in those races, it is usually older racehorses that undertake jump racing when compared to flat racing competitions (Krook & Maylin, 1988). Out of the studies that undertook the task to evaluate the racing surface with regards to horse fatalities in jump racing competitions only one study from Australia failed to find any relation between “hard” and “fast” racing courses and the risk of fatality (Boden, et al., 2007b). This is intriguing since a study of similar methodology conducted by the same authors, recognised track “going” as a significant risk factor when it comes to flat racing competitions. This variability in the results might be the outcome of another factor being of a greater significance when it comes to jump racing associated risks. Out of the four analyses that included increased race distance as risk factor, two of them concluded that it is a significant risk (Henley, et al., 2006; Boden, et al., 2007a), while the reason for those results are thought to be increased horse fatigue and increased time at risk.

Furthermore, when considered in a study, risk factors that assess previous racing and training histories of the racehorses were always found to be of significantly associated with the risk of fatalities. The categorisation though, and the means of assessment varied significantly between the studies. A study in 1995 (Estberg, et al., 1995) concluded that the relative risk of fatal musculoskeletal injuries during racing was three times greater for racehorses that accumulated racing and training distances that exceeded a particular cut off defined by those authors. This comes in a direct contradiction with two studies that found reduced risk of fatality for racehorses that undertook high speed training before a race (Cohen, et al., 2000), and increased previous distance in jump racing competitions (Boden, et al., 2007a).

Moreover, a UK study in 2006 (Henley, et al., 2006) reports that there is an increased risk of fatality when decreased previous starts are observed, on the other hand an Australian study in 2007 reported the exact opposite (Boden, et al., 2007a).

Another reported variable considered was the time since the previous race. A 2001 study showed an increased risk of fatality if the racehorse's previous race was more than 33 days ago (Hernandez, et al., 2001). Another study showed an increased risk if the horse participated in racing in the period of 31 to 60 days prior the race (Boden, et al., 2007a). For jump racing events, it was found that having run at least once within 14 days prior to a start and having made fewer starts in any type of race in the 60 days prior to a race increased the risk of fatality (Boden, et al., 2007b). Although there is a consensus that racing and training histories have noteworthy associations with racehorse fatalities, due to differences in the categorisation of the time periods the results are not readily comparable.

Furthermore, the surface and condition of the racetrack has been identified to be associated with fatal injuries in California (Arthur, 2010) and in the UK (Henley, et al., 2006; Williams, et al., 2001; Parkin, et al., 2004a,b; Parkin, et al., 2005).

Other less researched variables that have been considered with the increased risk of fatalities include, racehorse's career duration (Boden, et al., 2007b), altering the type of race for the racehorse (Henley, et al., 2006), competing in a city environment rather than a rural one (Boden, et al., 2007a, 2007b), pre-race veterinary checked horses identified as being at a higher risk, and finally an increased Beyer Grade, a numerical representation of the horse's performance, in the previous race (Cohen, et al., 2000). Those variables should be a starting point for future research in racehorse fatalities.

1.3. Fracture Injuries

Different countries, track surface, and race types have been identified as being associated with the occurrence of fracture incidents in racing. Racecourse veterinary reports or post-mortem reports are the two most commonly used methods for collecting data on racing fracture frequencies (Reardon, 2013). The further evaluation of fatal fractures has been facilitated by the introduction of post-mortem schemes and also by the investigation of specific injuries in the US (Johnson et al., 1994), UK (Parkin, et al., 2004c), and Australia (Boden, et al., 2006). Although post-mortem diagnoses are still the most accurate source of data about site and the extent of fractures, they are still not directly comparable with other studies most of the time because they only report on fatal fractures (Reardon, 2013). The risk of catastrophic fracture ranges from 0.33 to 2.3 per 1000 starts, varying with race type and country (Hill, et al., 1986; Peloso, et al., 1994; Mckee, 1995; Estberg, et al., 1996; Rosanowski, et al., 2016). Some studies have reported findings on specific fractures. Sesamoid and fetlock fractures were reported to be 0.53 per 1000 starts in all race types (Williams, et al., 2001), proximal phalangeal fracture were 0.16 per 1000 flat starts on turf, and proximal sesamoid bone fractures were 0.39 per 1000 flat starts on all-weather synthetic surfaces (Parkin, et al., 2004c). Lateral condylar fractures were found to be 0.3 and 0.35 per 1000 hurdle and steeplechase starts, respectively (Parkin, et al., 2004c).

There have also been studies that report fractures sustained during training. Two studies published in 2004 (Verheyen & Wood, 2004) and 2009 (Ely, et al., 2009) report fracture rates of 1.15 and 1.1 per 100 horse months, respectively during flat and jump racing training in the UK. Another study focusing on pelvic and tibial stress fractures uncovered rates of 0.15 pelvic and 0.16 tibial stress fractures per 100 horse months during training periods and even more importantly that solely 12% of the overall fractures that are being reported occurred during racing events (Verheyen, et al., 2006a).

Various risk factors for limb fractures include race type, country, and whether the fracture occurred during training or racing. Another generally recognised variable in regards with risk of fracture is the age of the horse. A study (Carrier et al., 1998) uncovered that the risk of complete humeral fracture for three-year-old horses was considerably high, and that the risk of complete pelvic fracture was high for “older” horses. Horses that started their racing career at the age of three- or four years of age are 2.6 times more likely to suffer a fracture in the future than horses that first started racing at the age of two (Parkin, et al., 2005).

On the other hand studies that focus their research on evaluating risk factors associated with fractures in general, fatal distal limb fractures, pelvic and tibial stress fractures, training fractures in general and forelimb proximal sesamoid bone fractures did not manage to identify any significant age association (Hill, et al., 1986; Parkin, et al., 2005; Verheyen, et al., 2006a; Verheyen, et al., 2006b; Anthenill, et al., 2007; Ely, et al., 2009). Thus, it is safe to assume that the association of age and fracture risk is not a simple one.

The sex of the horse and its possible association with fracture injuries has been considered in the literature. It has been found that male horses were associated with increased risk of complete humeral fracture (Carrier, et al., 1998) and forelimb proximal sesamoid fracture (Anthenill, et al., 2007). There is also a study that found out that female horses were a risk factor for complete pelvic fractures (Carrier, et al., 1998), and other studies from the UK - that included training information - that did not find any significant association between sex and the risk of fracture (Verheyen, et al., 2006a, 2006b; Ely, et al., 2009).

Considering surface and race length, a 1986 study (Hill, et al., 1986) did not show any association with fracture injuries, on the other hand more recent reports in 2004 and 2005 (Parkin, et al., 2004b; Parkin, et al., 2005) identified firm ground surface as a risk factor for distal limb fractures. Furthermore, dirt tracks and firmer turf tracks, have been found to be associated with higher musculoskeletal injury risk, with the prevailing hypothesis being that the poor cushioning from these tracks leads to more injuries (Bailey et al., 1998; Mohammed et al., 1991). Moreover, an association between a specific type of sand gallop in training and likelihood of pelvic and tibial stress fractures was identified in 2006 (Verheyen et al., 2006). A popular hypothesis for this, is that firmer ground surface might result in increased concussive forces on the bones while also increasing the overall race speeds, which might explain the augmented risk of fracture (Reardon, 2013). Also, two 2004 analyses identified longer race length as a significant risk factor for fatal distal limb fractures and fatal lateral condylar fractures as well (Parkin et al., 2004b, Parkin et al., 2005). Potential explanations for this include increased horse fatigue and increased time at risk for horses in longer races.

Racing and training histories have been shown to have a significant association with equine fracture injuries. Regarding complete humeral fractures there is a study that found an association with lay-up time and increased interval between races (Carrier et al., 1998). Another study identified an association between increased time in training and racing after a lay-up period and increased risk of forelimb proximal sesamoid bone fracture (Anthenill et al., 2007). A possible explanation for the contradicting results might have to do with differences in the aetiologies of these fractures. The 2007 study (Anthenill et al.2007) identified that any changes to the training schedules of the horses in order to reduce proximal sesamoid fractures may lead in the increase of humeral fractures, thus highlighting the difficulty of providing advice to trainers and policy makers. Another group of studies have identified associations between the amount of time in training and the overall risk of fracture, high risk of fatal distal limb fractures for horses in their starting year has been reported by two studies in 2004 and 2005 (Parkin et al., 2004a, Parkin et al., 2005), while increased time in training and racing was associated with an increased risk of forelimb proximal sesamoid fracture by another study in 2007 (Anthenill, et al., 2007).

Another identified risk factor for fatal distal limb fracture is the lack of gallop work in training, increasing the risk of fracture (Parkin et al., 2004a, Parkin et al., 2005);

overworking the horses during a short period of time has also been associated with increased risk, high canter distance in the past 30 days during training increases the risk of pelvic and tibial stress fracture, and high intensity exercise over a short period of time leads to an increased risk of all fracture types for horses in training (Verheyen, et al., 2006a, 2006b); high intensity exercise in the previous 12 months also leads to higher risk of forelimb PSB fracture (Anthenill, et al., 2007). A 2006 study (Verheyen, et al., 2006) and a 2007 study (Anthenill, et al., 2007) likewise report an association between increased accumulated exercise and increased risk of fracture. Numerous researchers came to the result that there is a significant association between risk of fracture and time and intensity of training which appears to be linked to the balance between subclinical bone damage and adaptation. When intense training appears to create an imbalance, clinical fractures are the outcome (Poole & Meagher, 1990; Stover, et al., 1992; Loitz & Zernicke, 1992; Riggs, et al., 1993; Riggs, et al., 1999a, 1999b; Kawcak, et al., 2000; Hill, et al., 2001).

To conclude the list of identified risk of factors, the following must be included; higher number of runners participating in the race, and fewer days between races for fatal distal limb fracture (Parkin, et al., 2004b, Parkin, et al., 2005). Finally, the following associations have also been identified; competing without professional jockeys for fatal lateral condylar fractures where this proposition was made on the basis that jockey's experience in identifying horse distress might influence the overall risk (Parkin, et al., 2005). Trainers have also been investigated for fractures sustained during jump training and racing, on the basis that there are differences in training regimens, veterinary input or horse populations between trainers (Ely, et al., 2009). Differences in the aetiologies of various fracture types, as well as alterations relating to factors associated with training and racing are likely to be the major reasons for the variations between studies.

2. Review of the Equine Injury Database and Exploratory Analysis

2.1. Introduction

The Jockey Club was established in New York in 1894. Its mission is to improve Thoroughbred breeding and racing primarily in the United States, Canada, and Puerto Rico. In 2008, The Jockey Club initiated the EID. The purpose of the EID is to identify the frequency, types, and outcomes of racing injuries in Thoroughbred racehorses competing in flat racing in a standardized format so that valid statistics can be generated in the hope that factors associated with specific injuries can be identified and appropriate measures implemented to prevent such injuries and improve the safety of Thoroughbred racing. The EID contains information for most Thoroughbred races that take place in the United States and Canada and serves as a near-census collection of available data.

2.2. Fatal Injuries

2.2.1. Introduction

The aim of this part of the study is to provide a description of the fatal equine injuries, based on the cases recorded in the EID, sustained during flat racing of Thoroughbred racehorses in the US and Canada.

2.2.2. Study Population

The study population comprised all 188,269 Thoroughbred horses that participated in flat racing in the US and Canada from 1st January 2009 to 31st December 2015 at the 89 race tracks reporting injuries to the EID.

2.2.3. Case Definition

The study was conducted with race start as the unit of analysis as fatal injuries are reported at the start level. Cases were defined as starts from horses that died or were euthanized within three days of sustaining an injury during a race. All other race starts from race tracks reporting injuries to the EID were classified as controls.

2.2.4. Description and prevalence of fatal injuries

There were 1.84 fatal injuries per 1000 starts for the 2,493,957 racing starts in the 7-year study period. A breakdown of injuries for each year is shown in table 2-1.

Table 2-1 Fatal injuries per year for the 2009 - 2015 period

Year	No. of Horses	No. of Starts	No. of fatalities per 1000 starts
2009	68,867	411,282	1.90
2010	65,942	390,847	1.84
2011	62,625	376,912	1.88
2012	59,864	365,774	1.92
2013	56,325	346,668	1.90
2014	52,939	309,669	1.80
2015	50,882	292,805	1.60

Out of all fatalities 83.1% were fractures, 16.4% were soft tissue injuries, 15.2% were joint injuries and 7.5% were non-musculoskeletal injuries. Of the recorded fatal injuries 1.7% were for either unknown or other reasons. A full breakdown of fatal injuries for that period is shown on table 2-2 for fractures, table 2-3 for joint injuries, table 2-4 for soft tissue injuries and table 2-5 for non-musculoskeletal injuries. Furthermore, all acute injury details regardless of overlap of different type of injuries in a single case for 2009 - 2014, are shown on table 2-6.

Table 2-2 Fractures mentioned in fatal injuries for the 2009-2015 period

Fracture	Both Fore	Both Hind	Left Fore	Left Hind	Right Fore	Right Hind	Unknown	NA	Total
Proximal sesamoid bone(s)	14	3	829	18	753	38	9	-	1664
MC3/T3	25	3	556	40	327	69	3	-	1024
Carpal bone(s)	28	-	298	-	392	-	4	-	722
P1/P2	4	-	102	47	68	56	2	-	279
Humerus	0	-	31	-	39	-	0	-	70
Pelvis	-	15	-	16	-	16	0	9	56
Scapula	2	-	24	-	26	-	0	-	52
Skull/Spine	-	-	-	-	-	-	-	40	40
Radius/ulnar	0	-	22	-	15	-	1	-	38
Tibia	-	0	-	18	-	11		-	29
Splint bone(s)	1	0	14	0	7	1	0	-	23
Femur	-	0	-	13	-	9	0	-	22
Tarsus	-	0	-	3	-	6		-	9
P3	0	0	2	0	2	0	0	-	4

Table 2-3 Severe joint injuries with no mention of fracture in fatal injuries for the 2009-2015 period

Joint injury	Both Fore	Both Hind	Left Fore	Left Hind	Right Fore	Right Hind	Unknown	All four limbs	NA	Total
Fetlock	9	2	96	1	77	4	1	0	-	189
Carpal	0	-	20	-	29	-	0	0	-	49
Interphalangeal	1	0	1	1	1	1	0	0	-	5
Stifle	-	0	-	2	-	2	1	0	-	5
Shoulder	0	-	0	-	1	-	0	0	-	1
Elbow	0	-	1	-	0	-	0	0	-	1

Table 2-4 Soft tissue injuries with no mention of fracture or severe joint injury in fatal injuries for the 2009-2015 period

Soft Tissue Injury	Both Fore	Both Hind	Left Fore	Left Hind	Right Fore	Right Hind	Unknown	All four limbs	NA	Total
Suspensory apparatus	4	0	75	2	51	3	1	0	-	136
Superficial Digital Flexor Tendon	0	0	29	2	19	3	1	0	-	54

Table 2-5 Non-musculoskeletal injuries with no mention of fracture, severe joint injury or soft tissue injury in fatal injuries for the 2009-2015 period

Non-musculoskeletal injury	Both Fore	Both Hind	Left Fore	Left Hind	Right Fore	Right Hind	Unknown	All four limbs	NA	Total
Neurological	0	7	1	0	0	0	0	1	9	18
Sudden death	-	-	-	-	-	-	-	-	-	258
Pulmonary Haemorrhage	-	-	-	-	-	-	-	-	-	34
Exercise Induced Pulmonary Haemorrhage	-	-	-	-	-	-	-	-	-	25
Post exertional distress/heatstroke	-	-	-	-	-	-	-	-	-	11
Cardiac arrhythmia	-	-	-	-	-	-	-	-	-	4

Table 2-6 All acute injury details regardless of multiple details in the same report in fatal injuries for the 2009-2014 period

Classification	Number (% of reports)
Proximal sesamoid bone fracture	1486 (35.7)
MC/T3 fracture	917 (22.0)
Carpal fracture	679 (16.3)
Suspensory apparatus injury	533 (12.8)
Fetlock joint injury	438 (10.5)
P1/P2 fracture	254 (6.1)
Sudden death	235 (5.6)
Superficial digital flexor tendon injury	158 (3.8)
Fracture with unreported location	81 (1.9)
Distal limb fracture	72 (1.7)
Humerus fracture	63 (1.5)
Pelvic fracture	49 (1.2)
Carpal joint injury	44 (1.1)
Scapula fracture	44 (1.1)
Non-specific soft tissue injury	42 (1.0)
Skull fracture	37 (0.9)
Radius/ulnar fracture	34 (0.8)
No detail at all 'unknown'	30 (0.7)
Pulmonary haemorrhage	28 (0.7)
Exercised induced pulmonary haemorrhage	27 (0.6)
Tibial fracture	26 (0.6)
Splint bone fracture	24 (0.6)
Unspecified joint injury	23 (0.6)
Femoral fracture	21 (0.5)
Neurological injury	21 (0.5)
Palmar ligament injury	18 (0.4)
Proximal limb fracture	15 (0.4)
Interphalangeal joint injury	12 (0.3)
Non-specific non-musculoskeletal injury	10 (0.2)
Post exertional distress/heatstroke	8 (0.2)
Tarsal fracture	8 (0.2)
Stifle joint injury	7 (0.2)
Cardiac arrhythmia	4 (0.1)
P3 fracture	4 (0.1)
Respiratory distress or disease	2 (0.05)
Colic	1 (0.02)
Elbow joint injury	1 (0.02)
Hock joint injury	1 (0.02)
Shoulder joint injury	1 (0.02)

Data to produce the statistics for the different type of injuries were retrieved from the database by searching for the relevant text in the field where the injury descriptions were recorded, within the R programming environment. The Unknown field in the tables is for injuries where the injury location was specifically recorded in the form as unknown whereas the NA field is for injuries where all the injury location fields were left blank.

2.3. Fracture Injuries

2.3.1. Introduction

The aim of this part of the study is to provide a description of the equine fracture injuries, based on the cases recorded in the EID, sustained during flat racing of Thoroughbred racehorses in the US and Canada.

2.3.2. Study Population

The study population comprised all 188,269 Thoroughbred horses that participated in flat racing in the US and Canada from 1st January 2009 to 31st December 2015 in the 89 race tracks reporting injuries to the EID.

2.3.3. Case Definition

The study was conducted with race start as the unit of analysis as equine fracture injuries are reported at the start. Cases were defined as starts from horses that sustained a fracture injury during a race. This definition includes every possible type of fracture, including skull and spinal fractures, that a Thoroughbred might have sustained. Fractures included were both fatal and non-fatal.

2.3.4. Description and prevalence of fracture injuries

Data to produce the statistics for the different type of injuries were retrieved from the database by searching for the relevant text in the field where the injury descriptions were recorded, in the R programming environment. There were 1.99 fracture injuries per 1000 starts for the 2,493,957 racing starts in the 7-year study period. A breakdown of injuries for each year is shown in table 2-7.

Table 2-7 Fracture injuries per year for the 2009 - 2015 period

Year	No. of Horses	No. of Starts	No. of fractures per 1000 starts
2009	68,867	411,282	1.97
2010	65,942	390,847	1.95
2011	62,625	376,912	2.09
2012	59,864	365,774	2.15
2013	56,325	346,668	2.01
2014	52,939	309,669	1.93
2015	50,882	292,805	1.81

Out of all fractures sustained during racing 90.8% were fractures of the distal limb, 4.8% were fractures of a proximal bone, 2.6% were fractures of the axial skeleton and 2.5% were non-specified fractures. Furthermore, 74.9% of fractures resulted in fatality and 9.2% of them also resulted in joint injuries. A full breakdown of fractures that resulted in a fatality is shown on table 2-2 and a full breakdown of fractures that did not result in a fatality is shown on table 2-8.

Table 2-8 Fractures mentioned in non-fatal injuries for the 2009-2015 period

Fracture	Both Fore	Both Hind	Left Fore	Left Hind	Right Fore	Right Hind	Unknown	NA	Total
Proximal sesamoid bone(s)	0	0	178	9	149	8	3	-	347
MC3/T3	2	1	135	17	105	7	1	-	268
Carpal bone(s)	7	-	162	-	216	-	1	-	386
P1/P2	0	-	14	3	10	3	1	-	31
Humerus	0	-	1	-	2	-	0	-	3
Pelvis	-	4	-	8	-	11	1	3	27
Scapula	0	-	1	-	1	-	0	-	2
Skull/Spine	-	-	-	-	-	-	-	7	7
Radius/ulnar	0	-	2	-	2	-	0	-	4
Tibia	-	0	-	3	-	3	0	-	6
Splint bone(s)	0	0	11	0	5	0	0	-	16
Femur	-	0	-	1	-	0	0	-	1
Tarsus	-	0	-	3	-	1	0	-	4
P3	0	0	15	1	10	0	0	-	26

2.4. Exploratory analysis of EID Variables

The data available for each variable in the EID were explored and density plots and dot plots were generated where appropriate. For the smoothing bandwidth of the density plots the standard deviation of the smoothing kernel was used. For each group in the dot plots, exploring the relation between a variable and fatal or fracture injuries, 95% confidence intervals were calculated using standard errors.

2.4.1. Accumulated distance ran in career (Km)

From the racing and exercise starts available for each horse we calculated the accumulated distance they had run in recorded starts throughout their career. The density of starts by km accumulated prior to the race is shown at Figure 2-1 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for different distance groups are shown at figures 2-2 and 2-3 respectively.

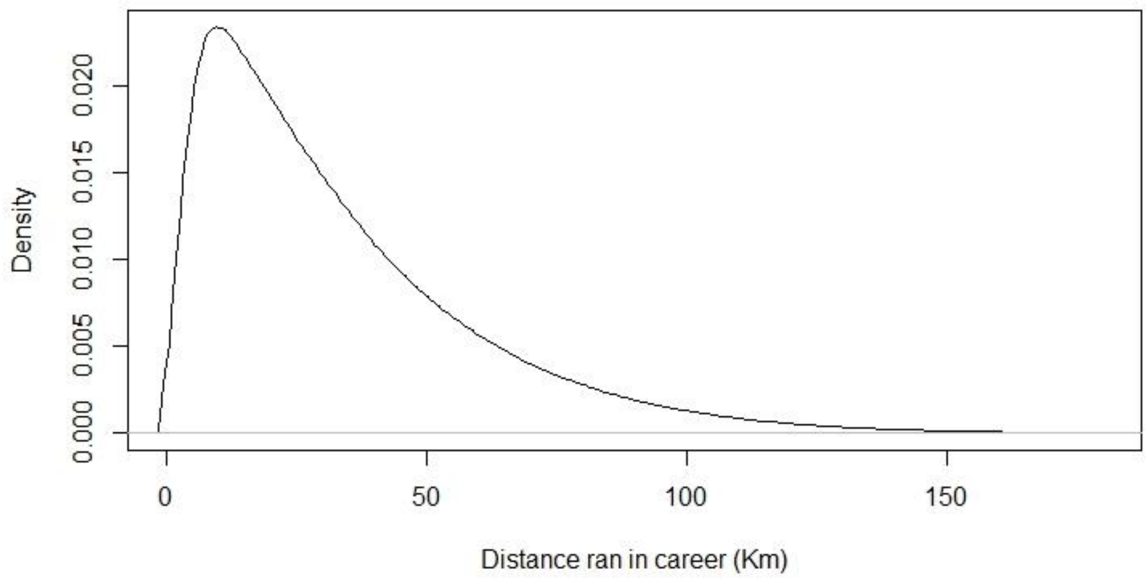

Figure 2-1 Density plot of accumulated distance ran in career (Km)

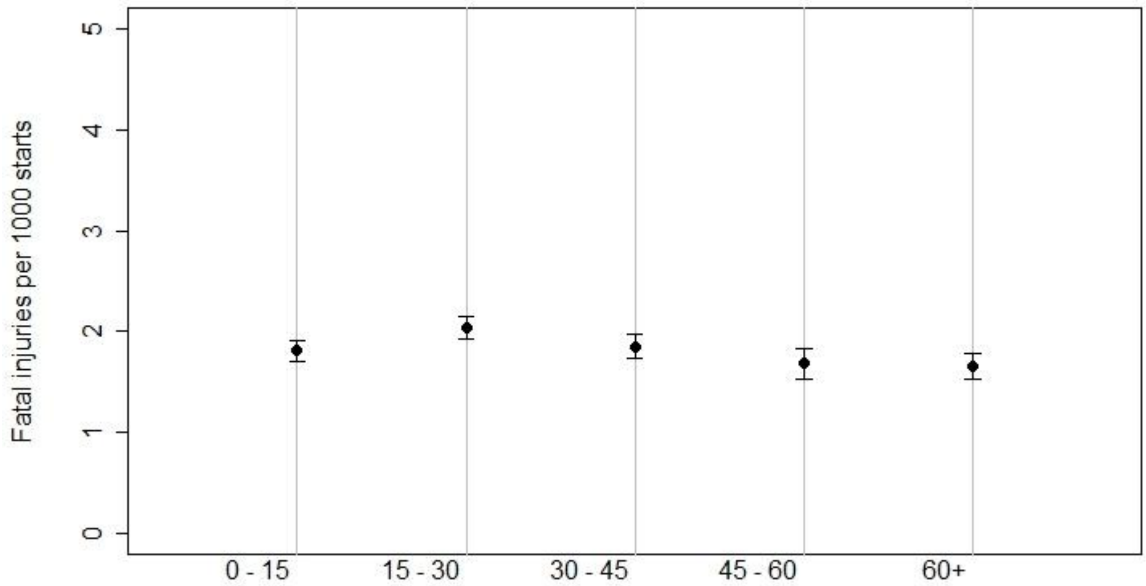

Figure 2-2 Fatal injuries per 1000 racing starts with 95% confidence interval by accumulated distance run in career (Km) group

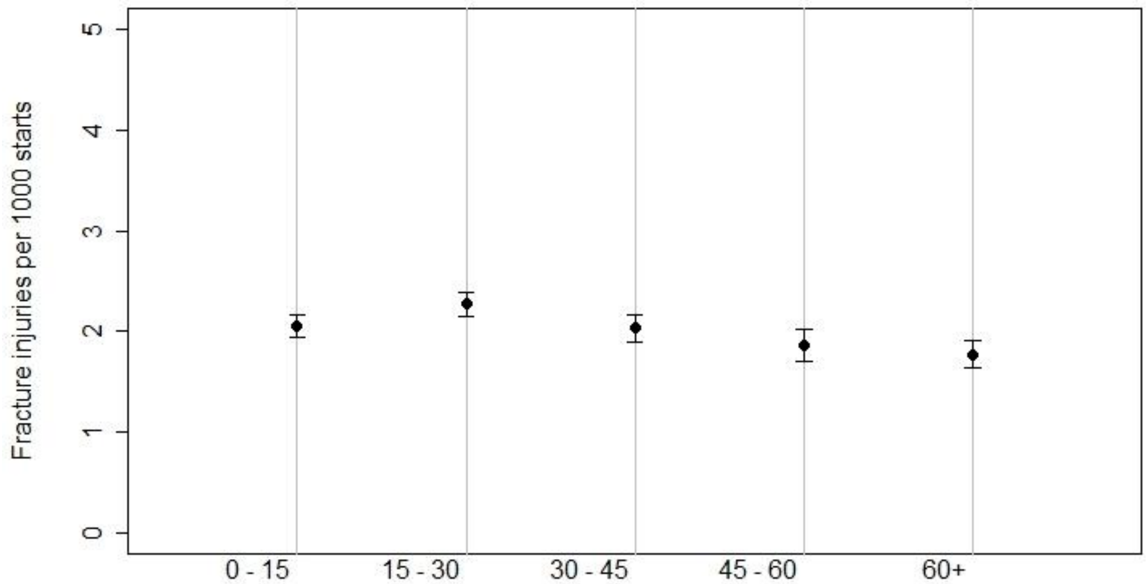

Figure 2-3 Fracture injuries per 1000 racing starts with 95% confidence interval by accumulated distance ran in career (Km) group

2.4.2. Accumulated exercise distance ran in career (Km)

From the exercise starts available for each horse we calculated the accumulated distance they had ran in recorded starts throughout their career. The density of starts by km accumulated prior to the race is shown at Figure 2-4 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for different distance groups are shown at figures 2-5 and 2-6 respectively.

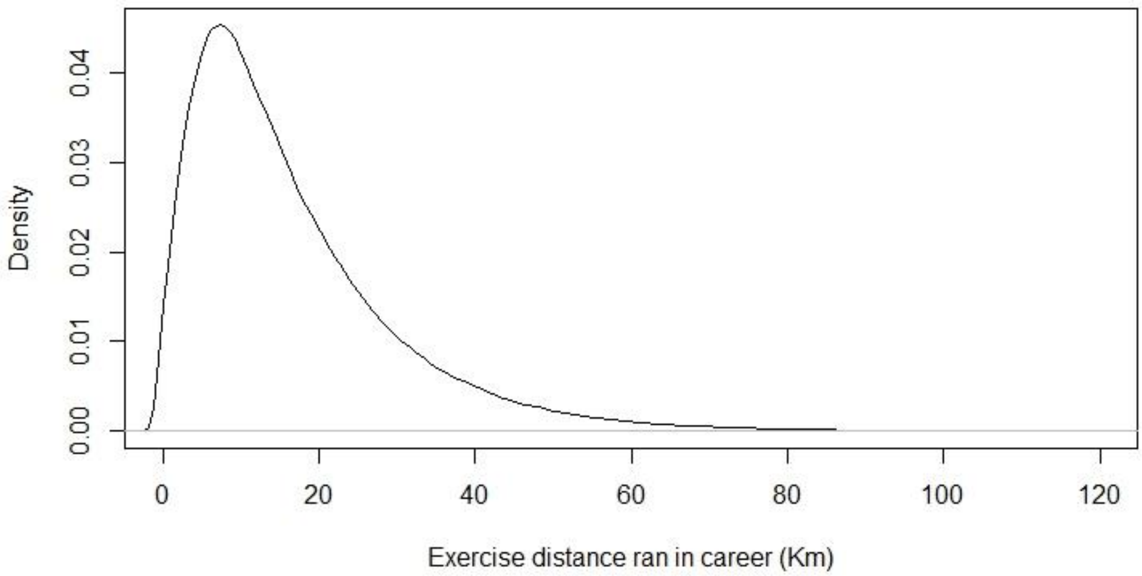

Figure 2-4 Density plot of accumulated exercise distance ran in career (Km)

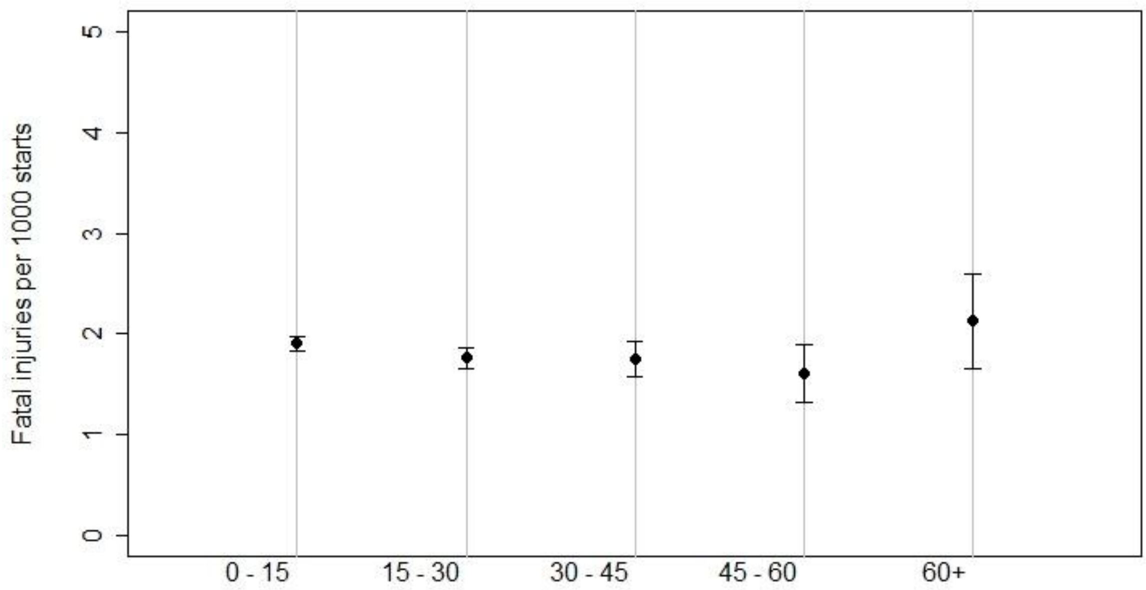

Figure 2-5 Fatal injuries per 1000 racing starts with 95% confidence interval by accumulated exercise distance ran in career (Km) group

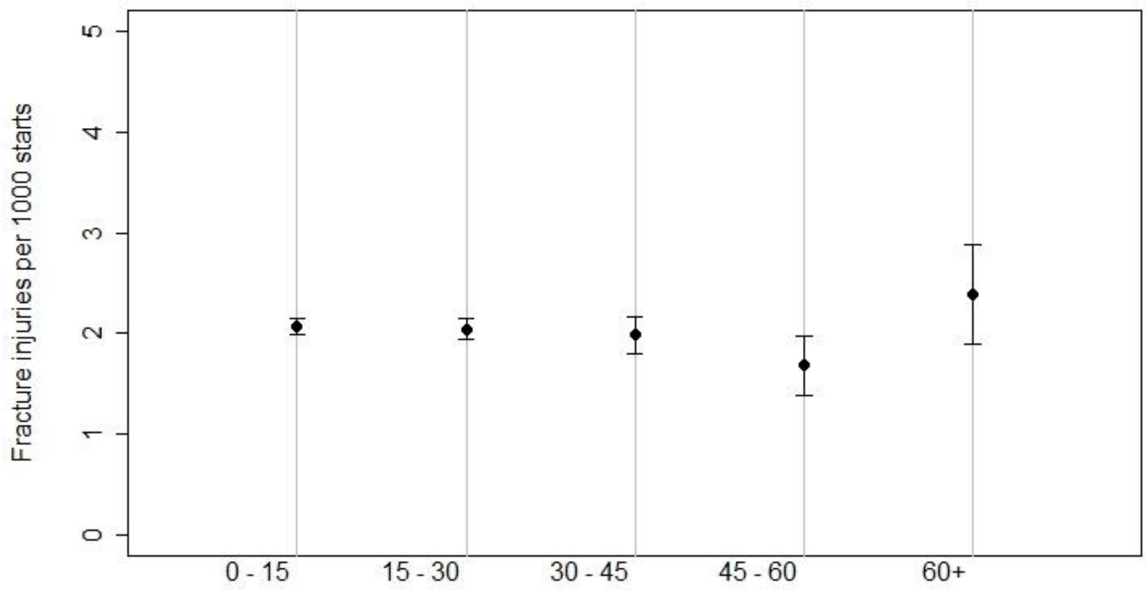

Figure 2-6 Fracture injuries per 1000 racing starts with 95% confidence interval by accumulated exercise distance ran in career (Km) group

2.4.3. Accumulated racing distance ran in career (Km)

From the racing starts available for each horse we calculated the accumulated distance they had run in recorded race starts throughout their career. The density of starts by km accumulated prior to the race is shown at Figure 2-7 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for different distance groups are shown at figures 2-8 and 2-9 respectively.

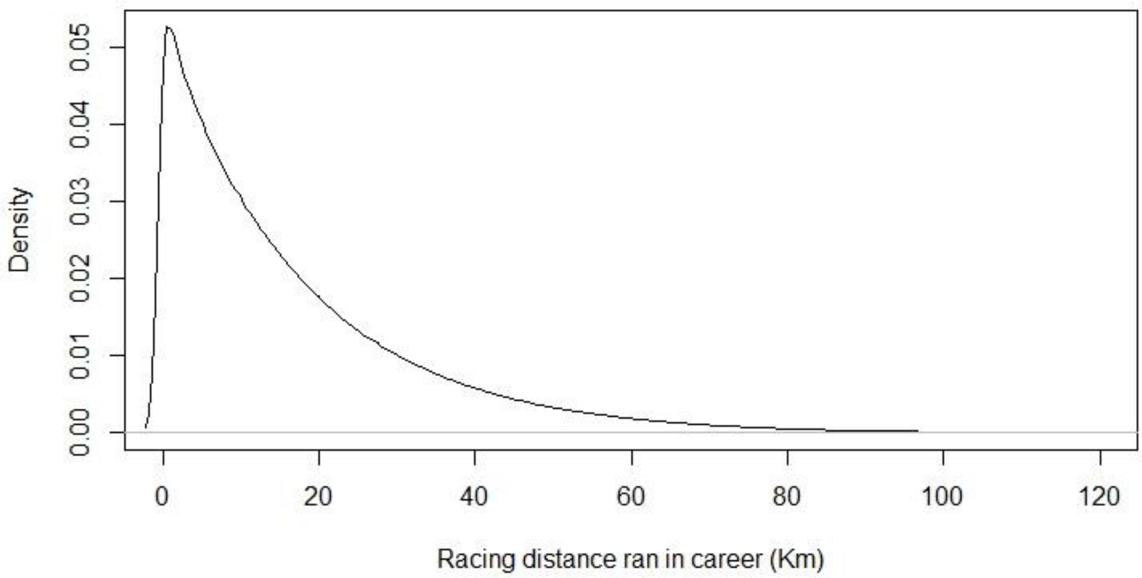

Figure 2-7 Density plot of accumulated racing distance ran in career (Km)

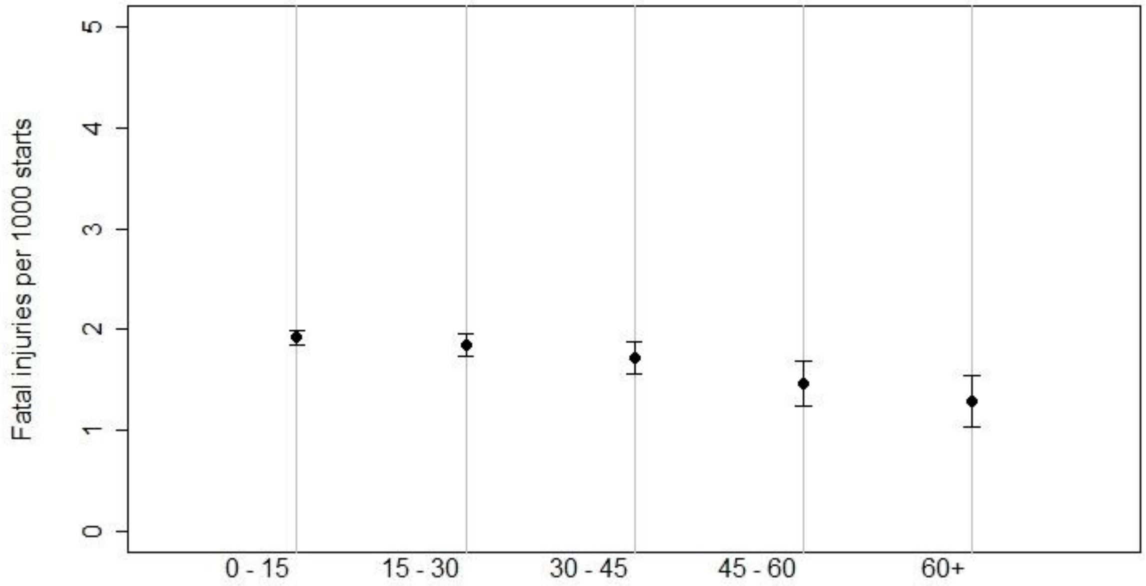

Figure 2-8 Fatal injuries per 1000 racing starts with 95% confidence interval by accumulated racing distance ran in career (Km) group

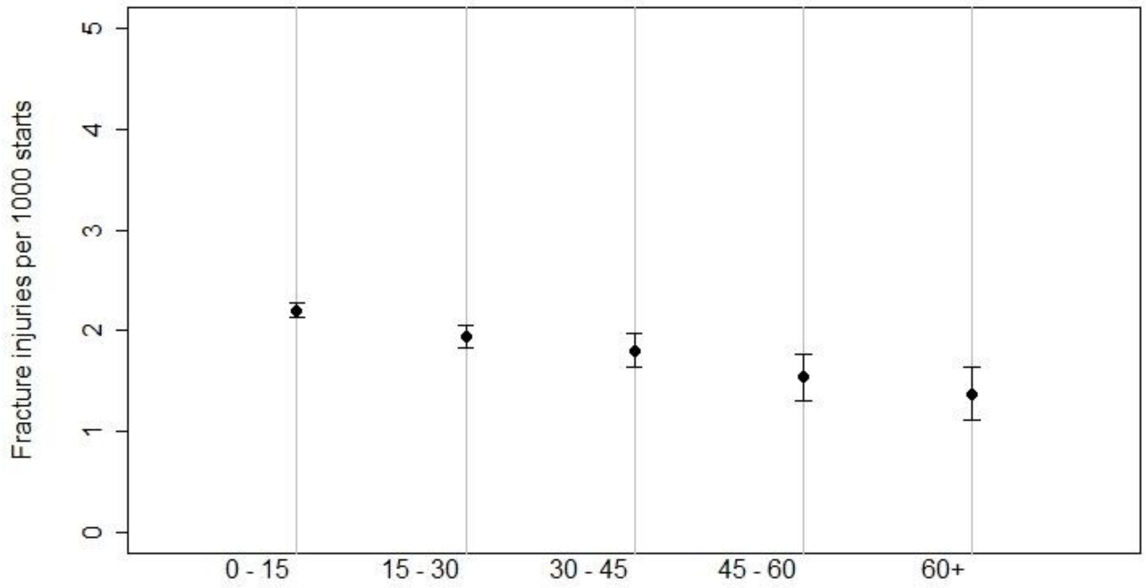

Figure 2-9 Fracture injuries per 1000 racing starts with 95% confidence interval accumulated racing distance ran in career (Km) group

2.4.4. Age (years)

The EID contained the date of birth of each horse and we calculated the biological age at the start of each race. The density of racing starts by age group is shown at Figure 2-10 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for each age group are shown at figures 2-11 and 2-12 respectively.

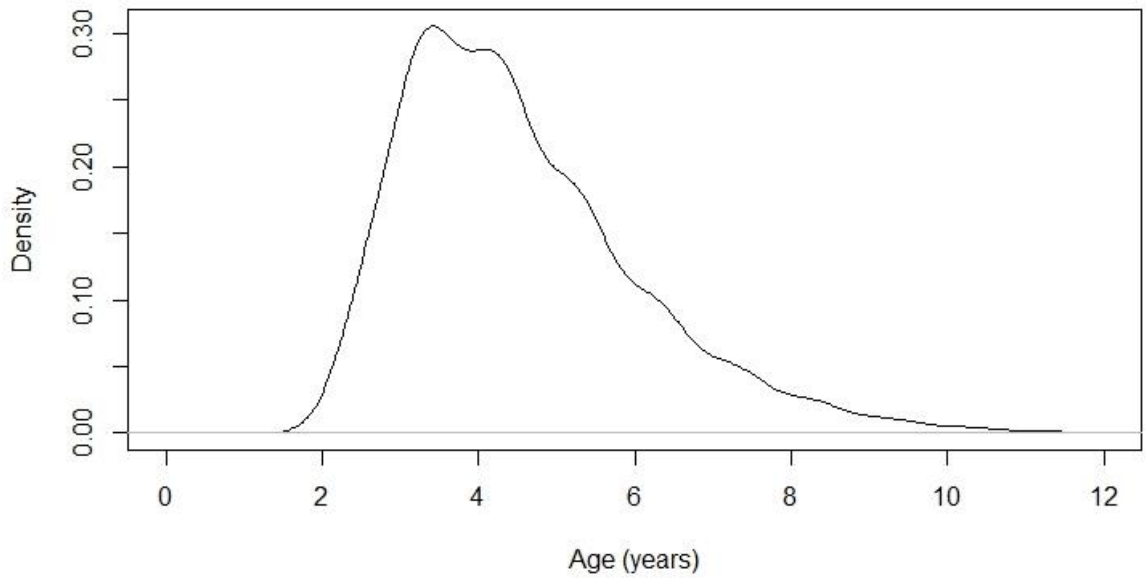

Figure 2-10 Density plot of age (years)

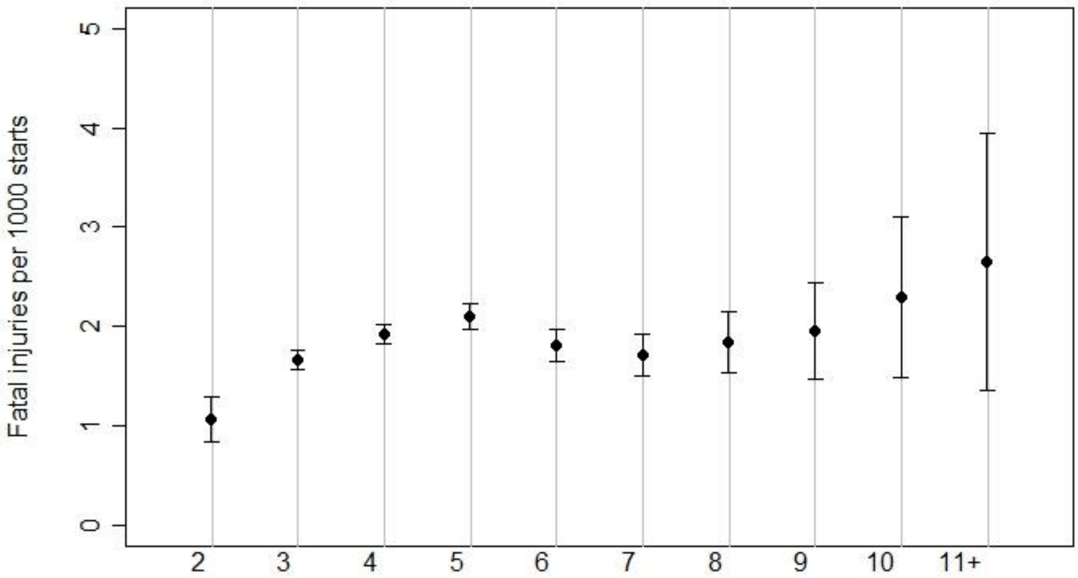

Figure 2-11 Fatal injuries per 1000 racing starts with 95% confidence interval by age group

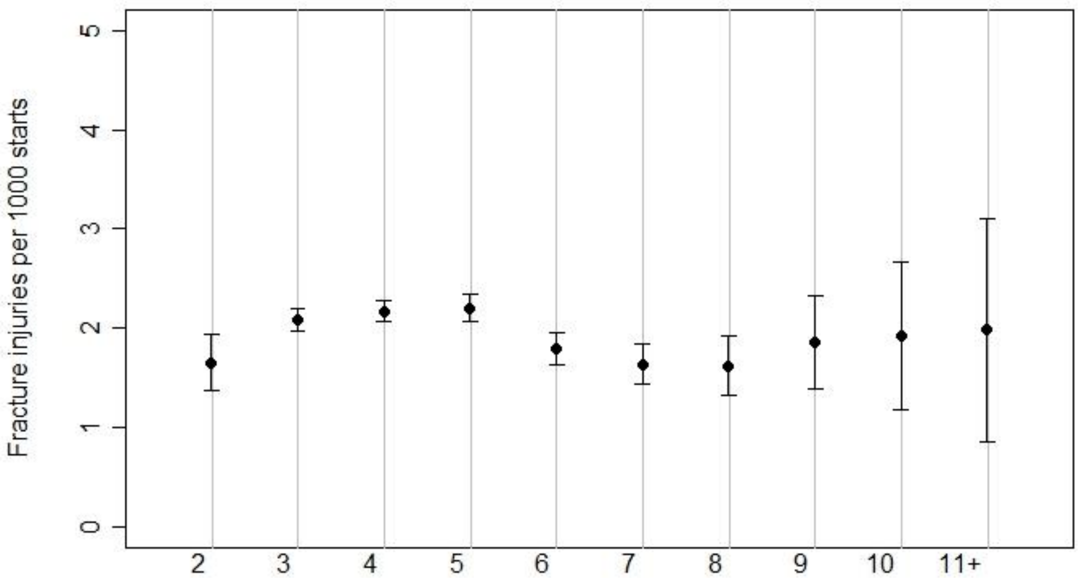

Figure 2-12 Fracture injuries per 1000 racing starts with 95% confidence interval by age group

2.4.5. Age at first start (years)

The EID contained the date of birth of each horse and we calculated the biological age at the start of each race. The density of racing starts by age at first start group is shown at Figure 2-13 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for each age at first start group are shown at figures 2-14 and 2-15 respectively.

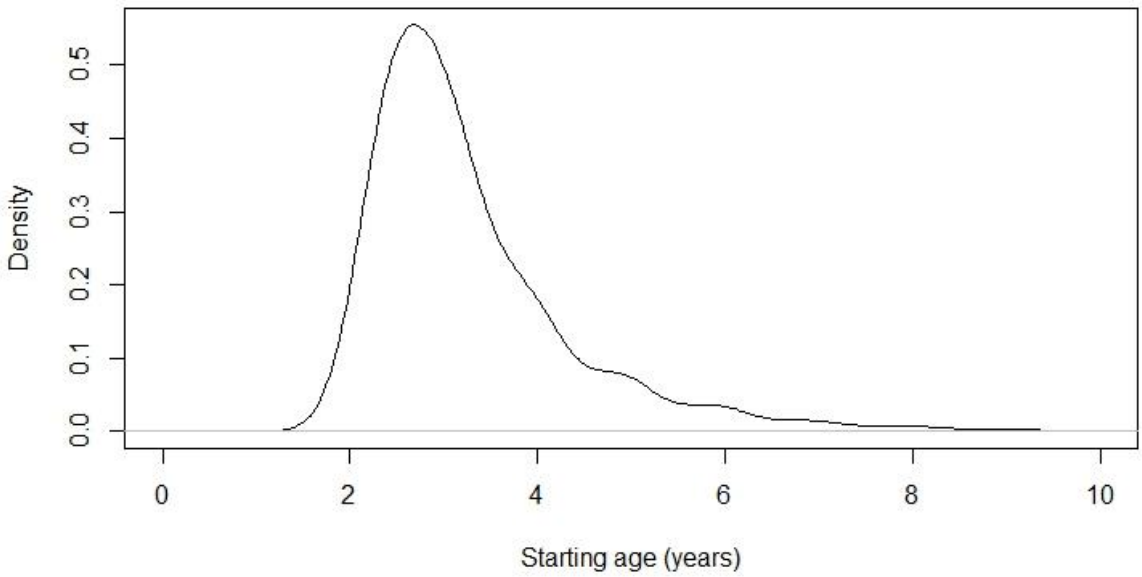

Figure 2-13 Density plot of starting age (years)

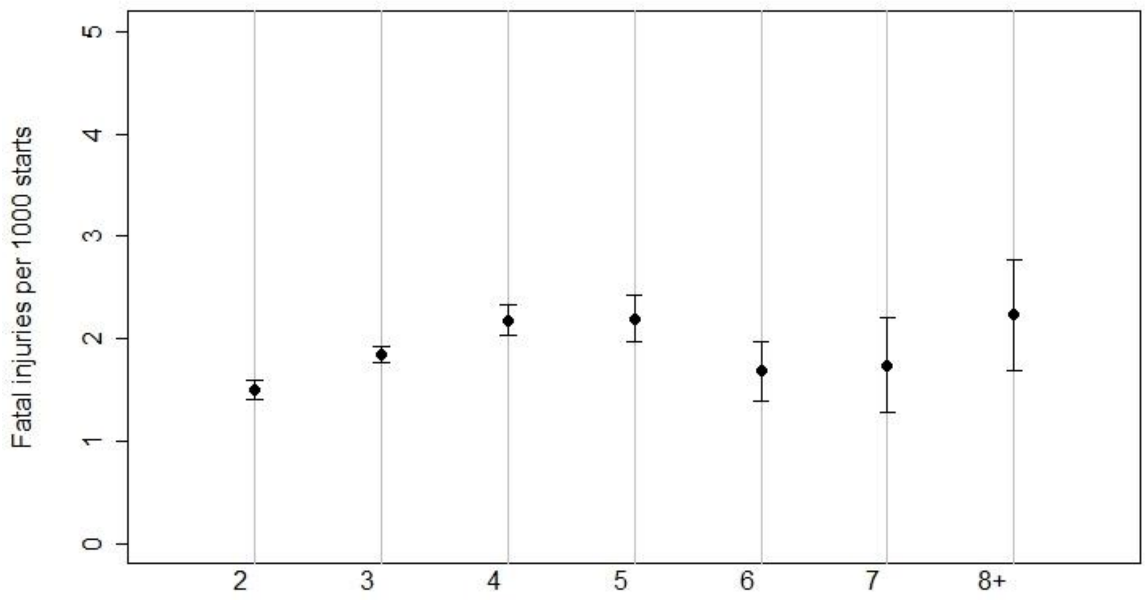

Figure 2-14 Fatal injuries per 1000 racing starts with 95% confidence interval by age at first start group

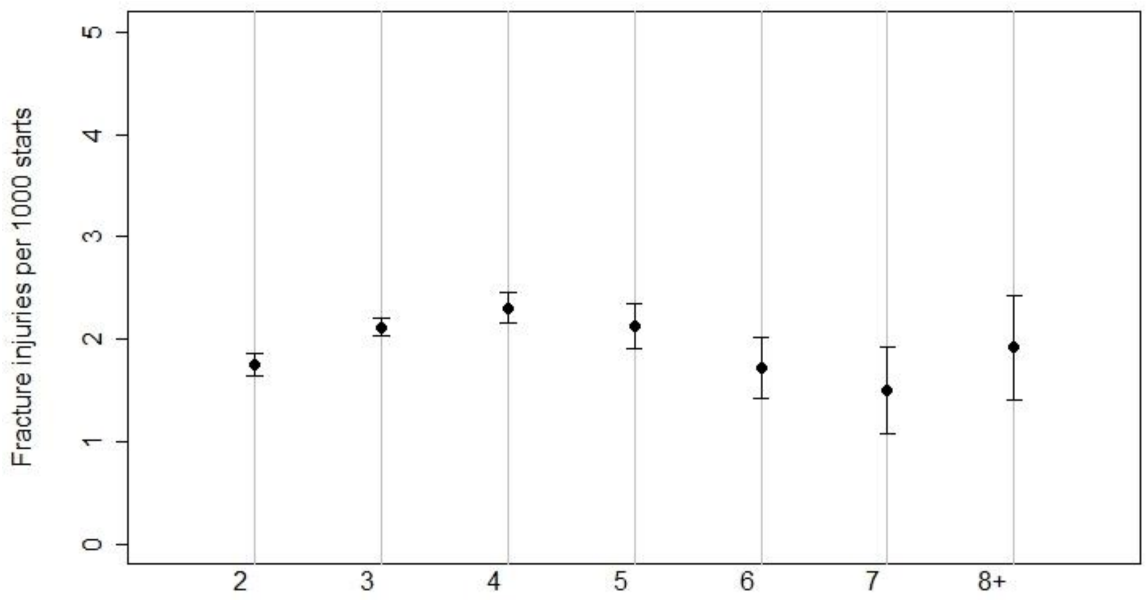

Figure 2-15 Fracture injuries per 1000 racing starts with 95% confidence interval by age first start group

2.4.6. Average speed change on previous race (m/s)

From the starts available for each horse we calculated the difference in each speed of the horse between the two prior races. The density of starts by m/s is shown at Figure 2-16 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for different speed groups are shown at figures 2-17 and 2-18 respectively.

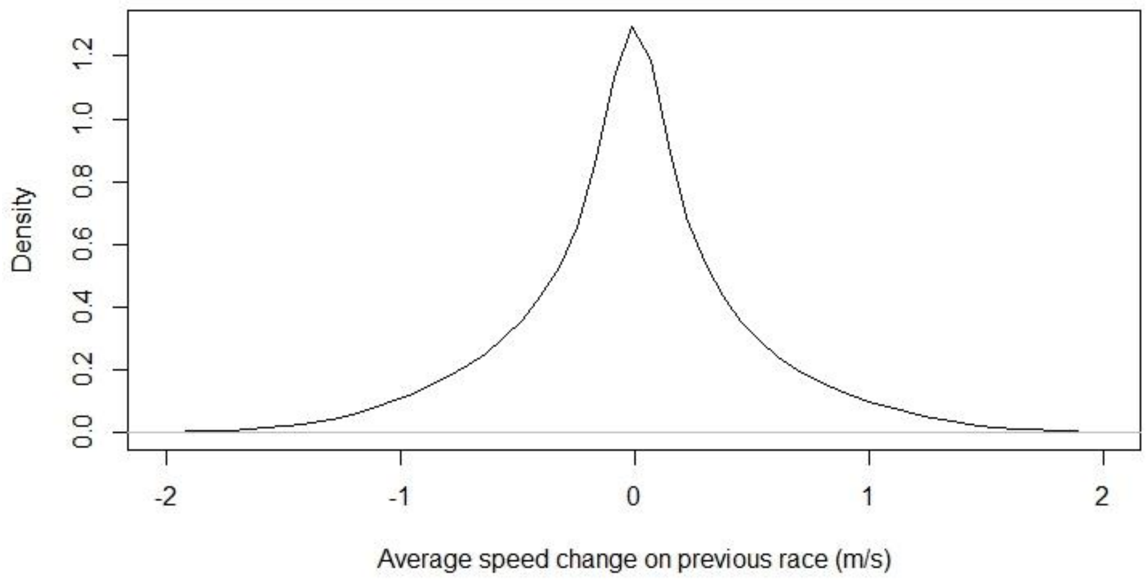

Figure 2-16 Density plot of average speed change on previous race (m/s)

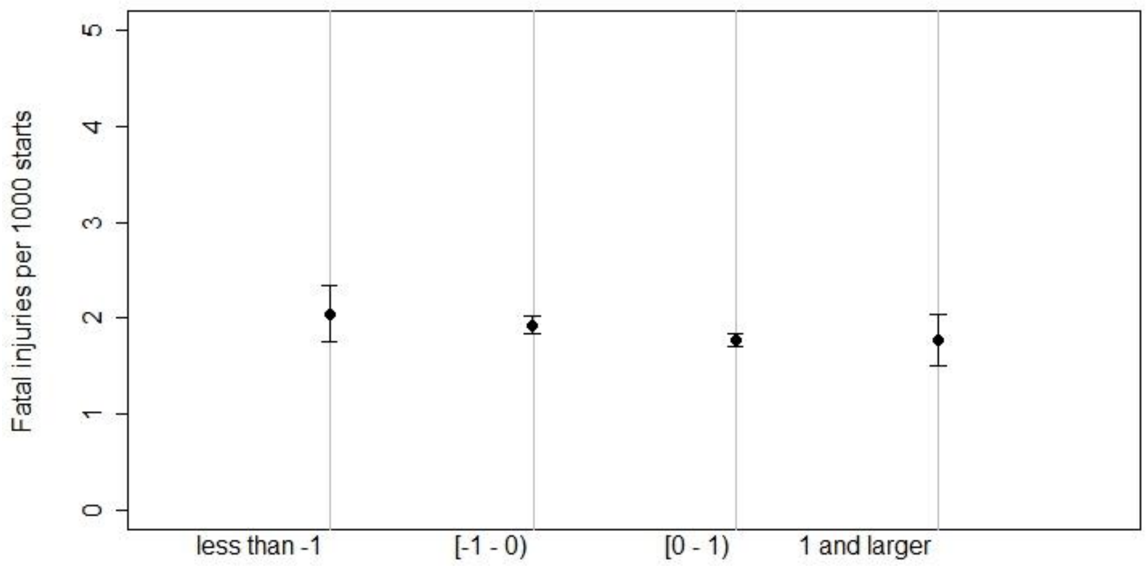

Figure 2-17 Fatal injuries per 1000 racing starts with 95% confidence interval by average speed change on previous race (m/s) group

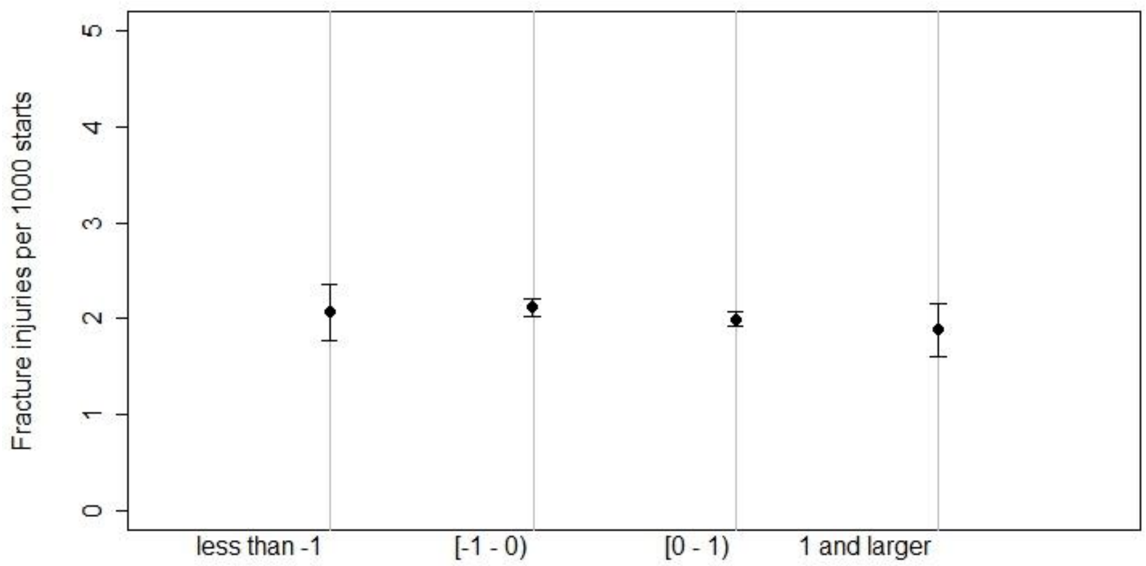

Figure 2-18 Fracture injuries per 1000 racing starts with 95% confidence interval by average speed change on previous race (m/s) group

2.4.7. Average speed in previous race (m/s)

From the information available we calculated the speed of each horse in the previous race. The density of starts by m/s is shown at Figure 2-19 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for different speed groups are shown at figures 2-20 and 2-21 respectively.

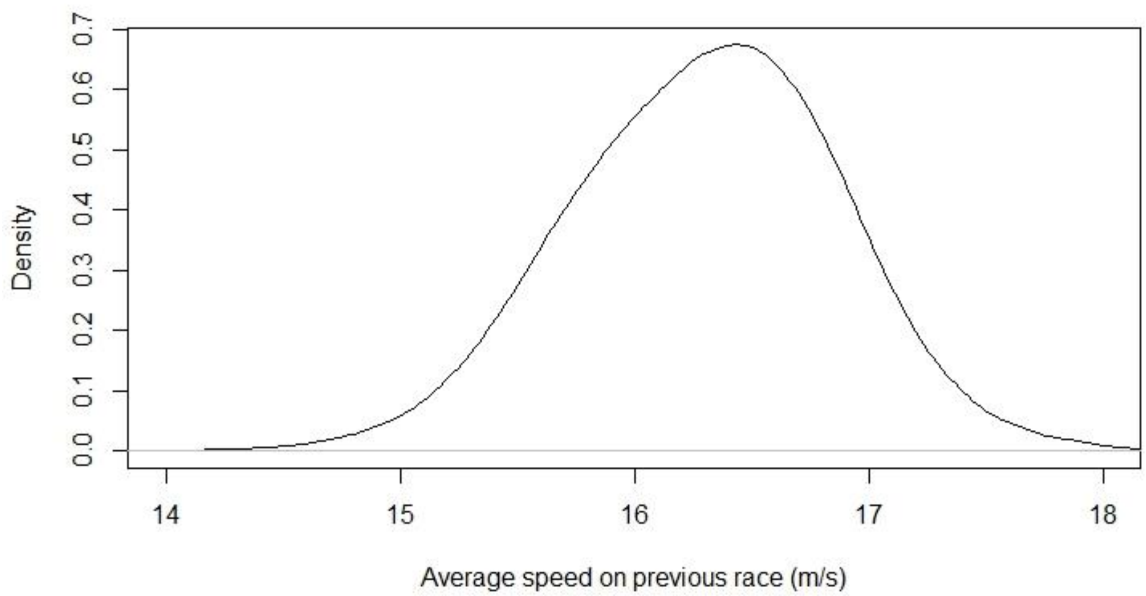

Figure 2-19 Density plot of average speed in previous race (m/s)

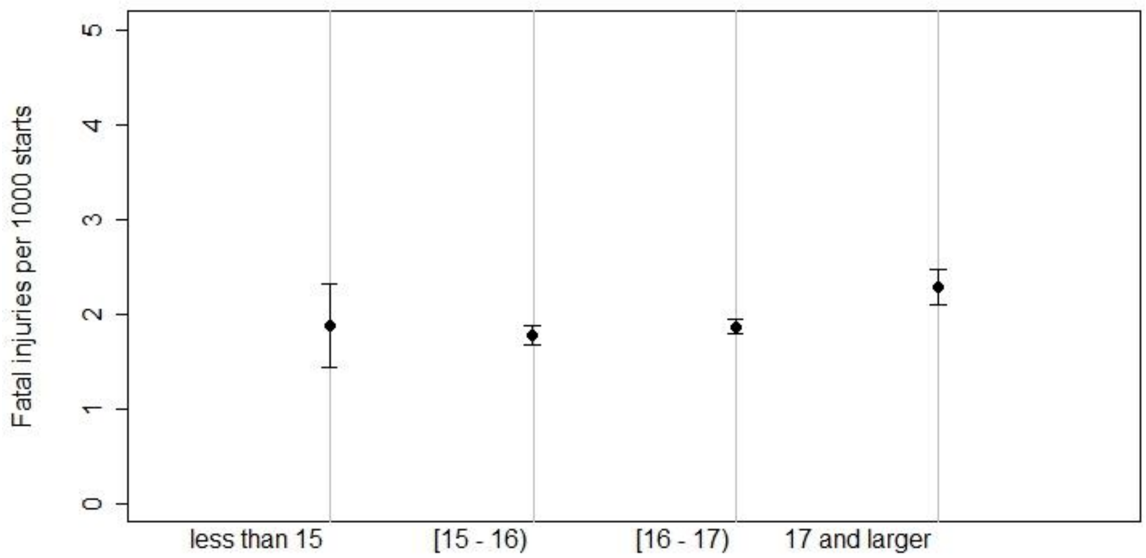

Figure 2-20 Fatal injuries per 1000 racing starts with 95% confidence interval by average speed in previous race (m/s) group

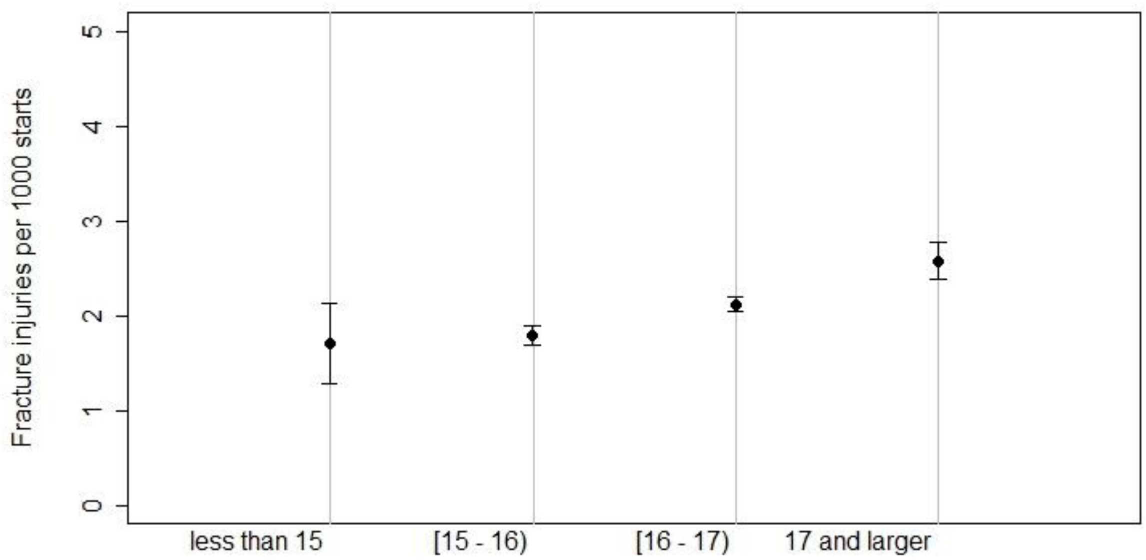

Figure 2-21 Fracture injuries per 1000 racing starts with 95% confidence interval by average speed in previous race (m/s) group

2.4.8. Country

The EID contained information on the country each race took place. The proportion of starts in the USA is 92% and in Canada 8%. The number of fatal and fracture injuries, along with their 95% confidence intervals, for the two countries are shown at figures 2-22 and 2-23 respectively.

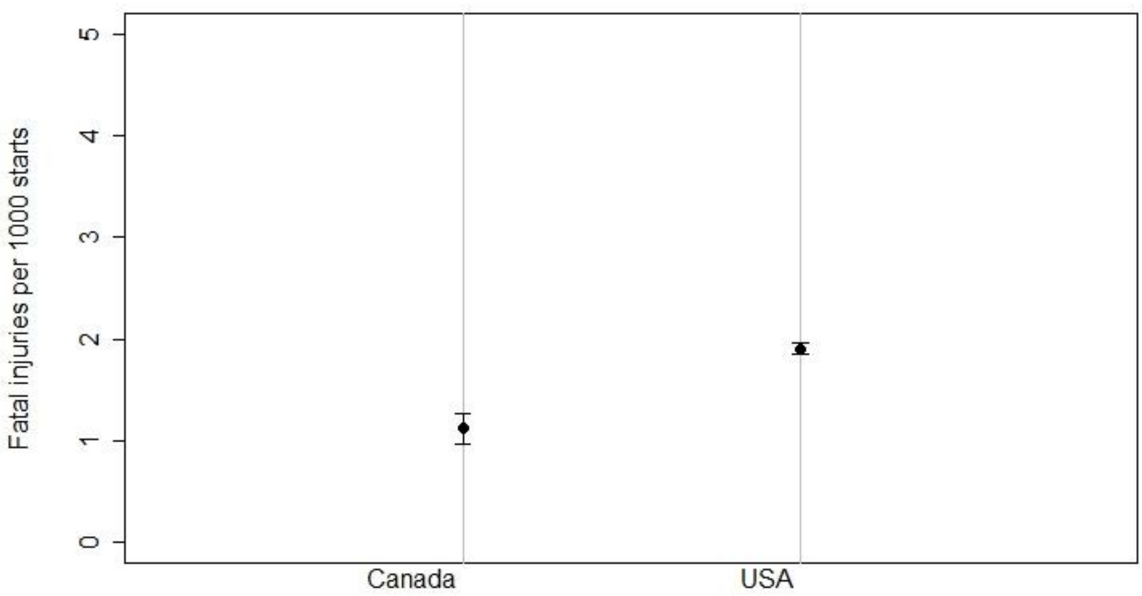

Figure 2-22 Fatal injuries per 1000 racing starts with 95% confidence interval by country

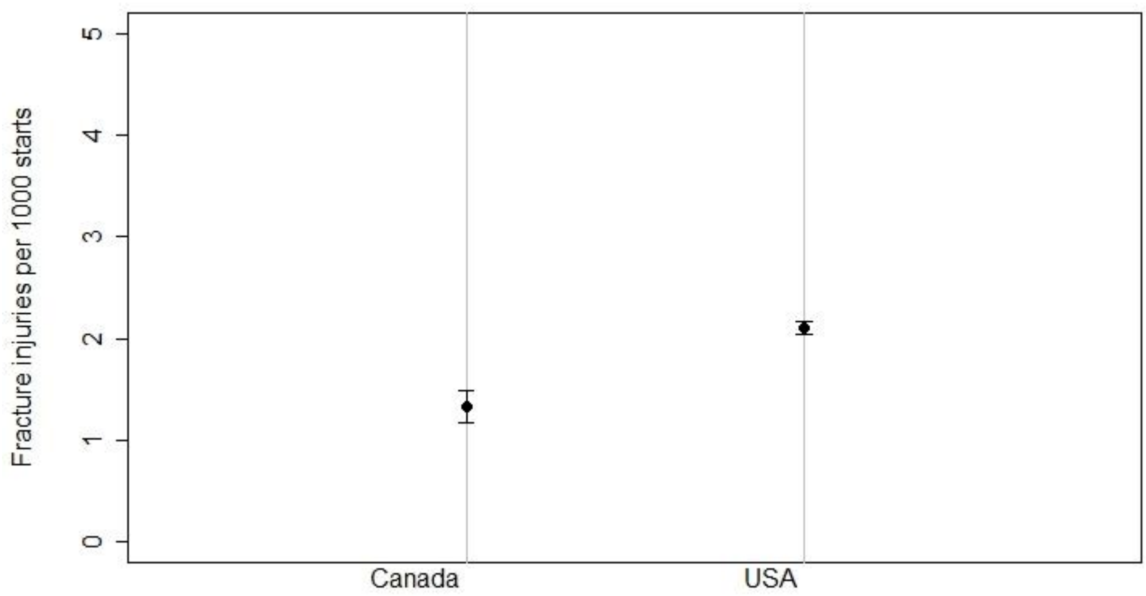

Figure 2-23 Fracture injuries per 1000 racing starts with 95% confidence interval by country

2.4.9. Entered the vet list

A risk factor specific to the North American jurisdiction is horses that have previously entered the veterinarian's list. This is a list used by association and regulatory veterinarians to provide horses with illness, injury or soundness issues a brief respite from racing. The proportion of starts for horses that at some point in their career entered the veterinarian list is 19% and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-24 and 2-25 respectively.

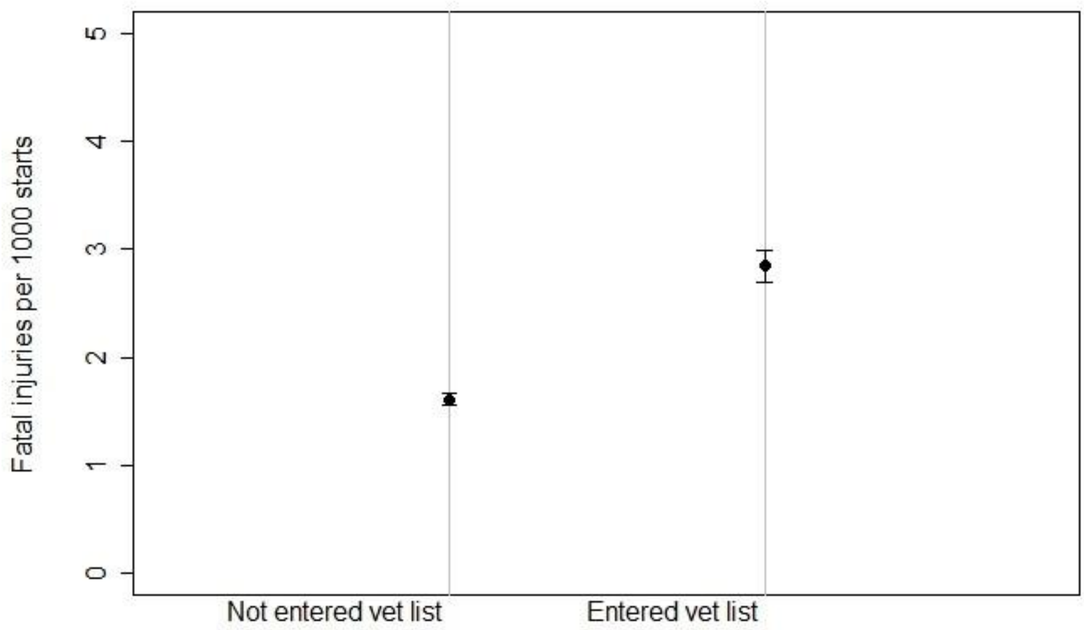

Figure 2-24 Fatal injuries per 1000 racing starts with 95% confidence interval by vet list group

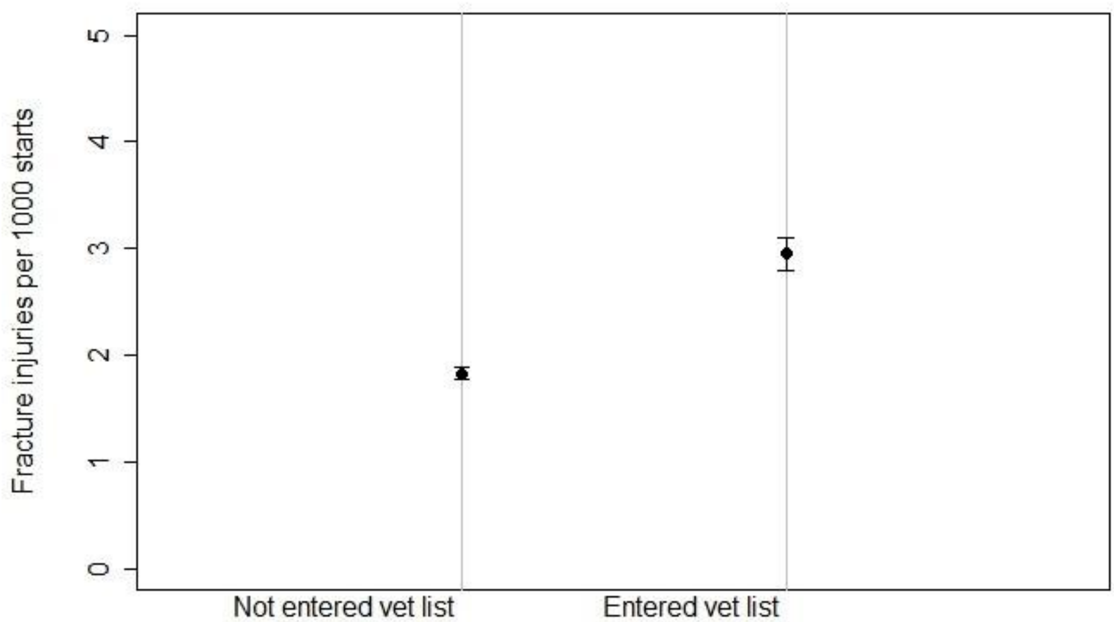

Figure 2-25 Fracture injuries per 1000 racing starts with 95% confidence interval by vet list group

2.4.10. Field size

The EID contained information on the field size of each race. Field size is the number of horses participating in a race. The proportion of starts by different field sizes is shown at Figure 2-26 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for different field size groups are shown at figures 2-27 and 2-28 respectively.

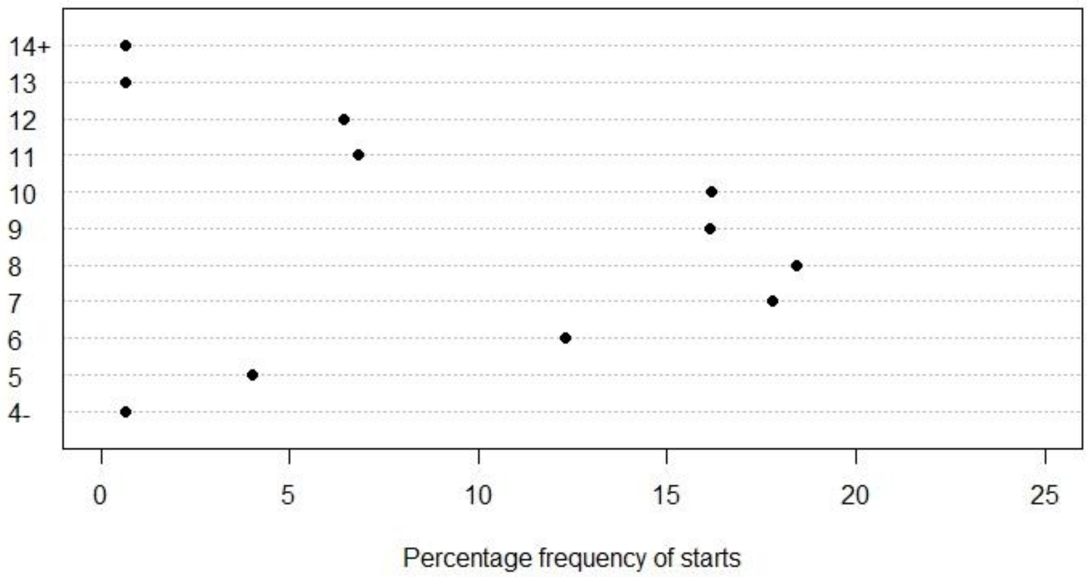

Figure 2-26 Dot plot of field size

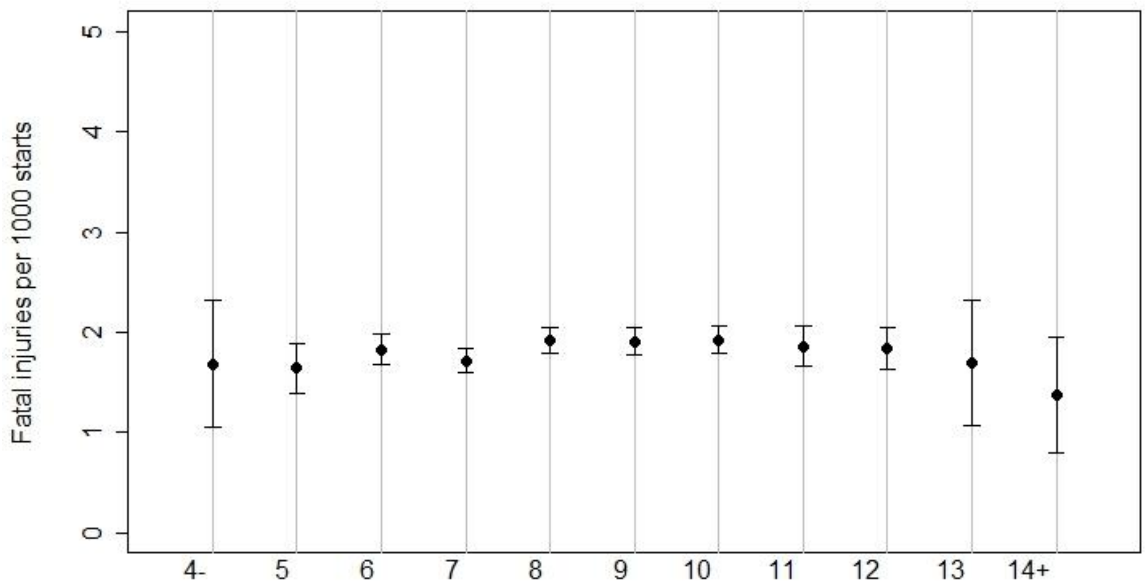

Figure 2-27 Fatal injuries per 1000 racing starts with 95% confidence interval by field size

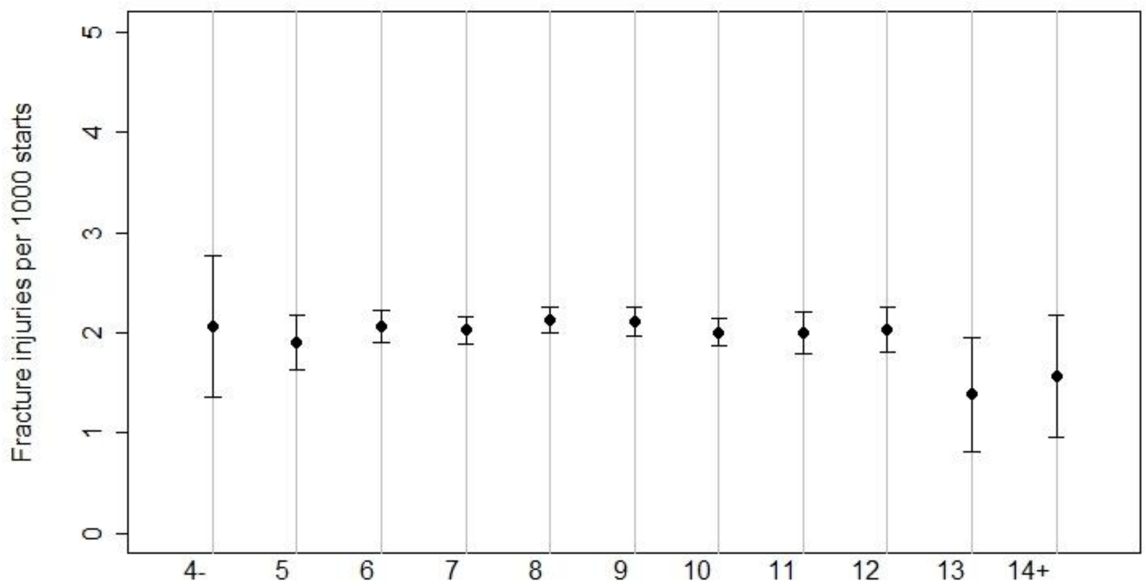

Figure 2-28 Fracture injuries per 1000 racing starts with 95% confidence interval by field size

2.4.11. First start

The EID contained information on the racing starts of each horse and we looked at the first start for each horse. The proportion of starts from horses in their first racing start is 7% and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-29 and 2-30 respectively.

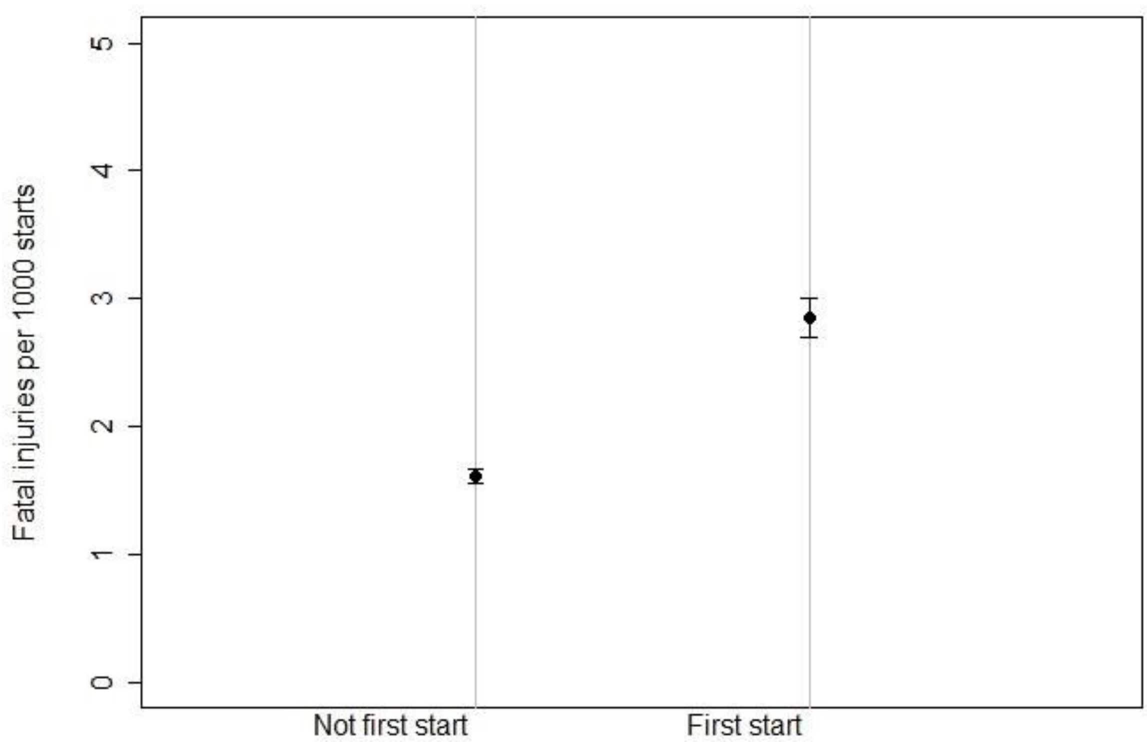

Figure 2-29 Fatal injuries per 1000 racing starts with 95% confidence interval by first start

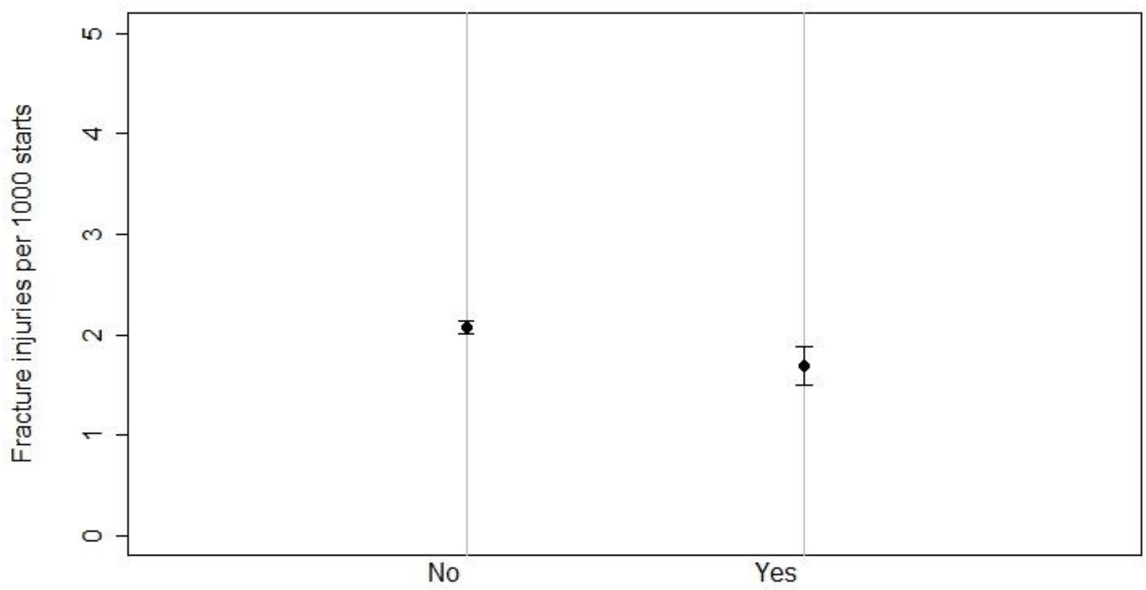

Figure 2-30 Fracture injuries per 1000 racing starts with 95% confidence interval by first start

2.4.12. Low purse race (<= $7500)

The EID contained information on the purse of each race and we looked at races with a low purse of equal or less than $7500. This figure was chosen on an *ad hoc* basis to specifically explore races with the lowest 15% purse. The number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-31 and 2-32 respectively.

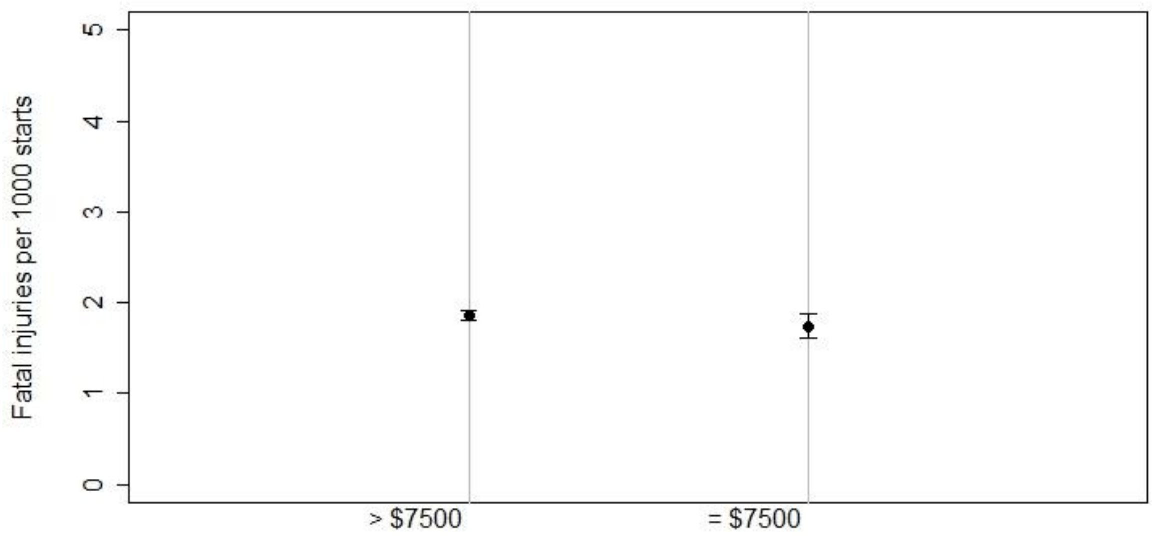

Figure 2-31 Fatal injuries per 1000 racing starts with 95% confidence interval by low purse race (<= $7500) group

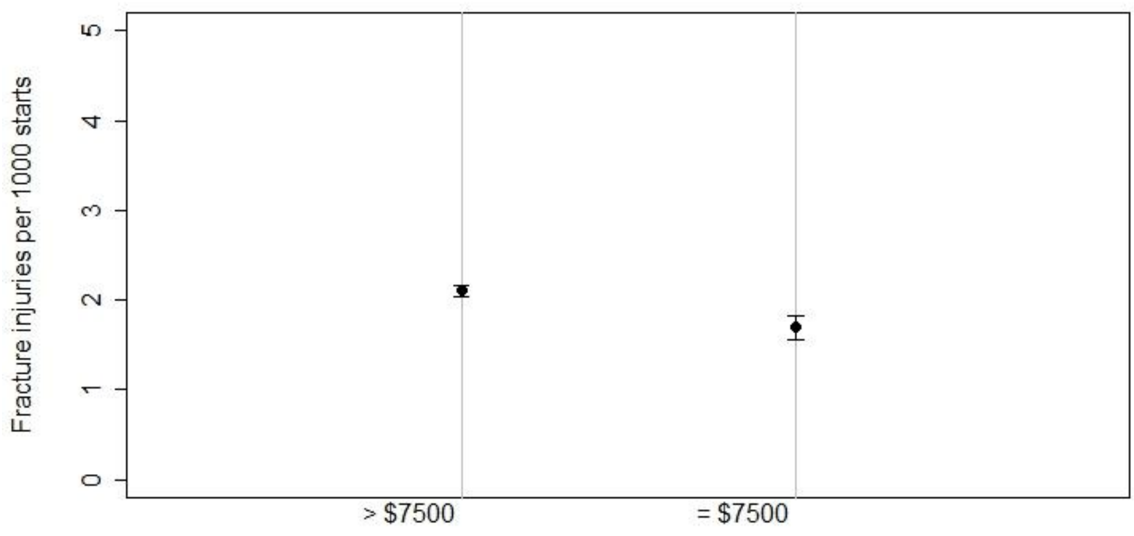

Figure 2-32 Fracture injuries per 1000 racing starts with 95% confidence interval by low purse race (<= $7500) group

2.4.13. Months since last racing start

From the racing starts available for each horse we calculated the time in months since the last racing start for each horse. The density of starts per month since by the time in months since the last racing start is shown at Figure 2-33 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-34 and 2-35 respectively.

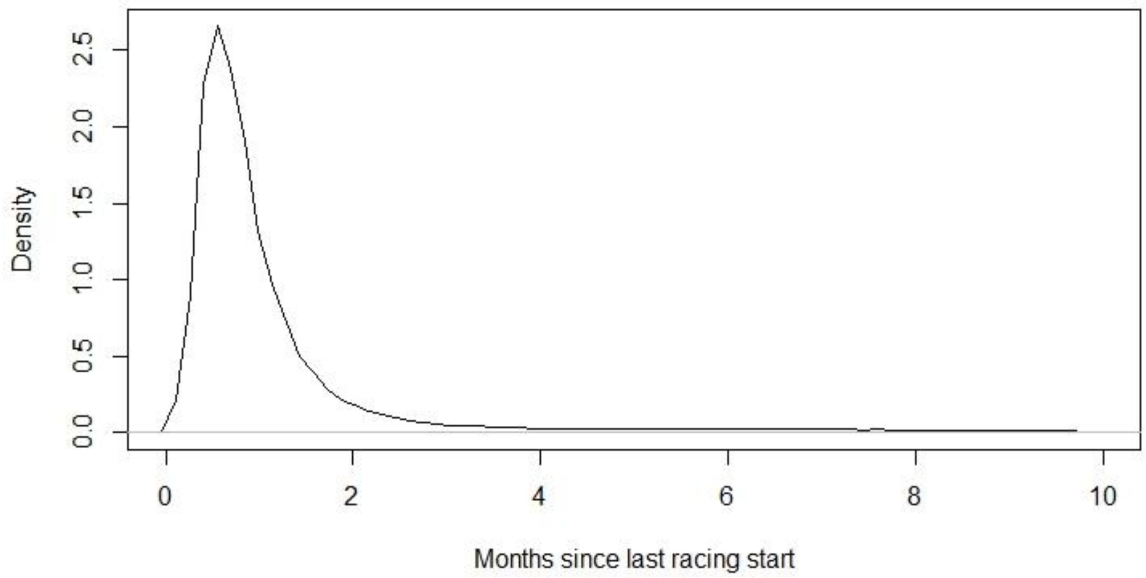

Figure 2-33 Density plot of racing starts by months since last racing start

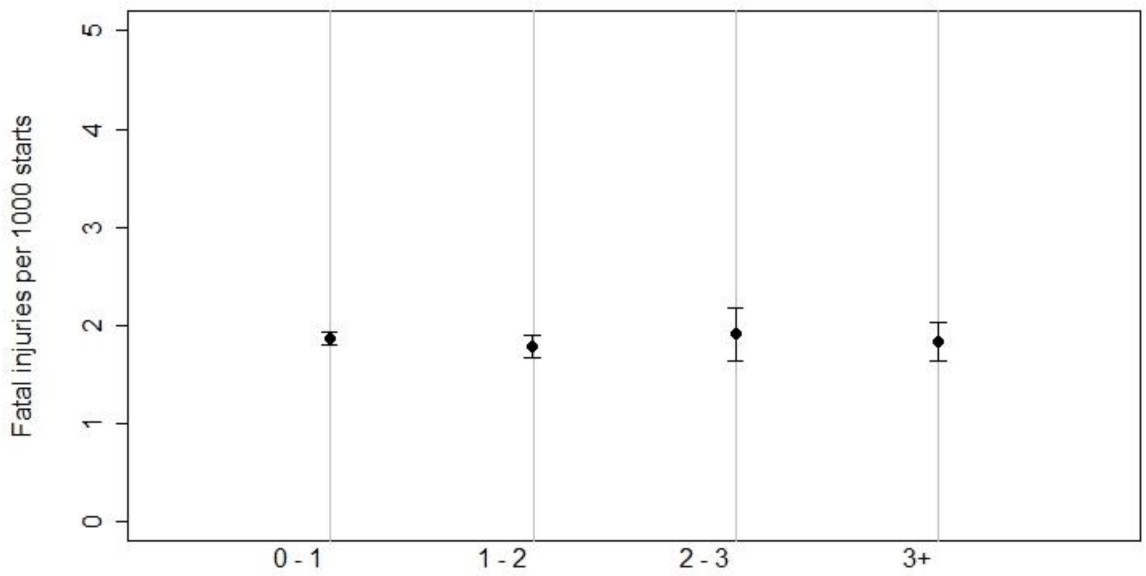

Figure 2-34 Fatal injuries per 1000 racing starts with 95% confidence interval by time (months) since last racing start group

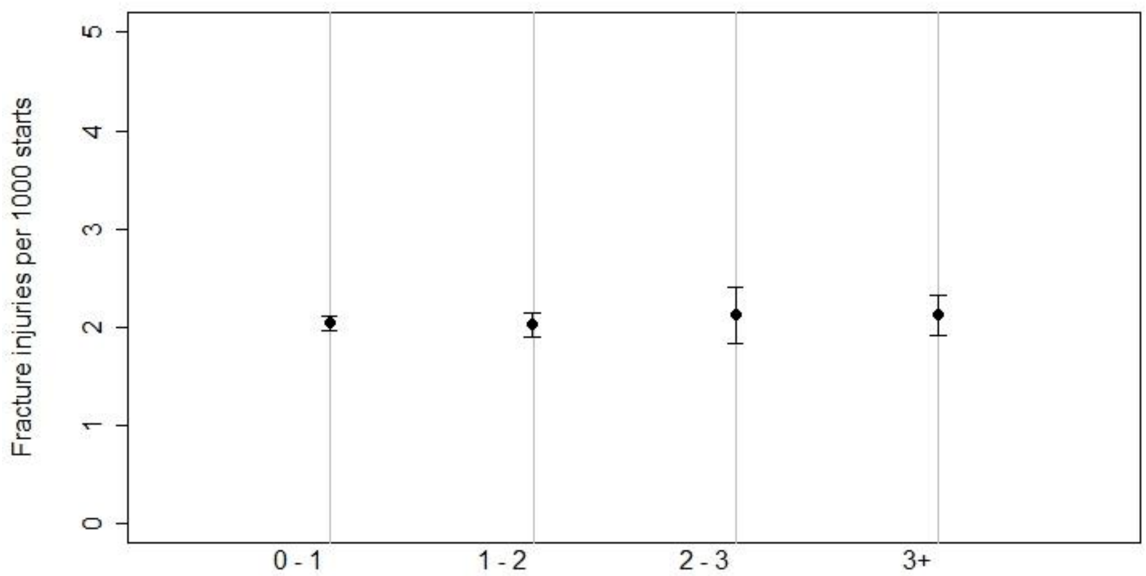

Figure 2-35 Fracture injuries per 1000 racing starts with 95% confidence interval by time (months) since last racing start group

2.4.14. Months since last racing or exercise start

From the racing and exercise starts available for each horse we calculated the time in months since the last start for each horse. The density of starts per month since by the time in months since the last racing or exercise start is shown at Figure 2-36 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-37 and 2-38 respectively.

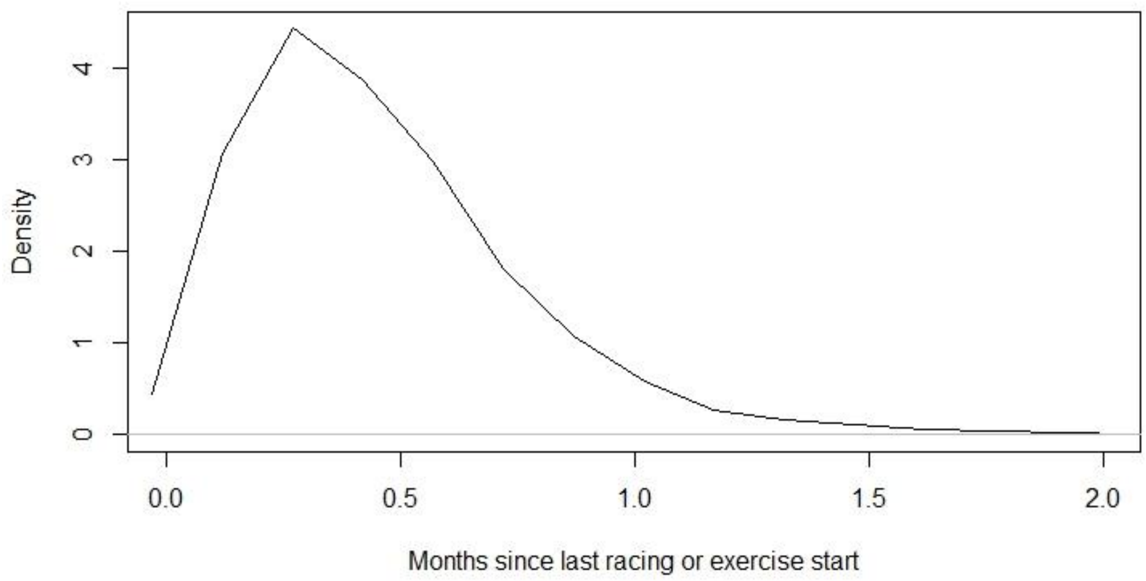

Figure 2-36 Density plot of time (months) since last racing or exercise start

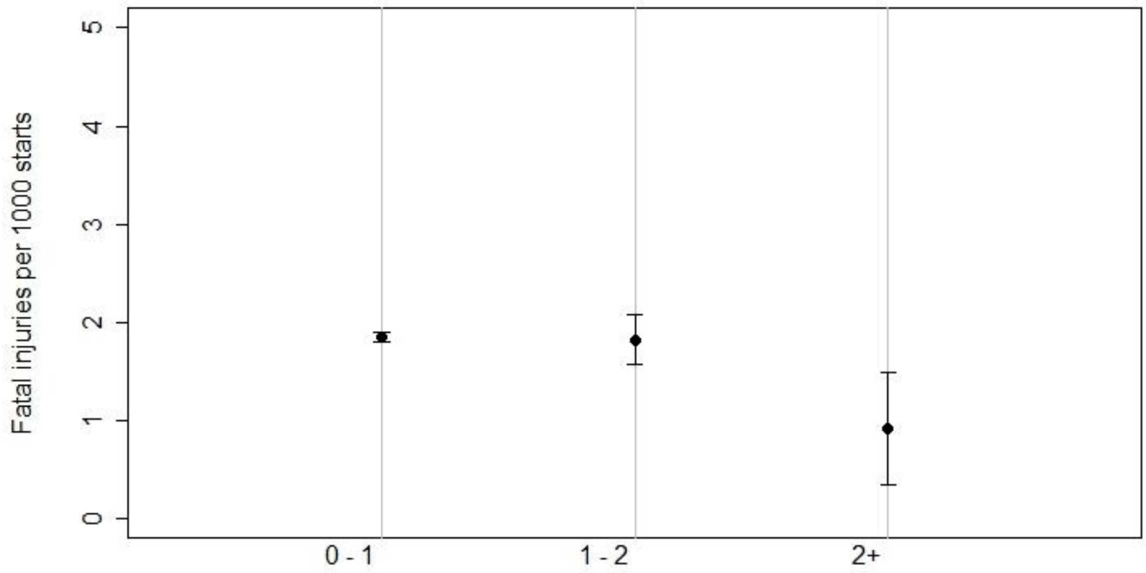

Figure 2-37 Fatal injuries per 1000 racing starts with 95% confidence interval by time (months) since last racing or exercise start group

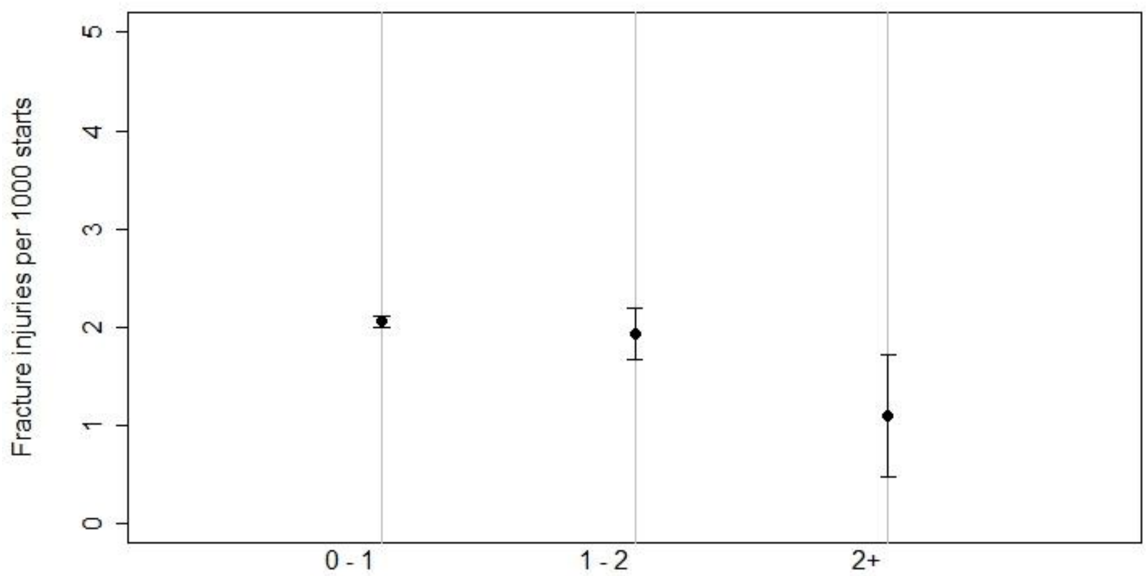

Figure 2-38 Fracture injuries per 1000 racing starts with 95% confidence interval by time (months) since last racing or exercise start group

2.4.15. Number of layups

From the racing and exercise starts available for each horse we calculated the number of layups a horse had throughout its career. A layup was defined as a more than 60-day period without any recorded racing or exercise start. The proportion of starts by number of layups is shown at Figure 2-39 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-40 and 2-41 respectively.

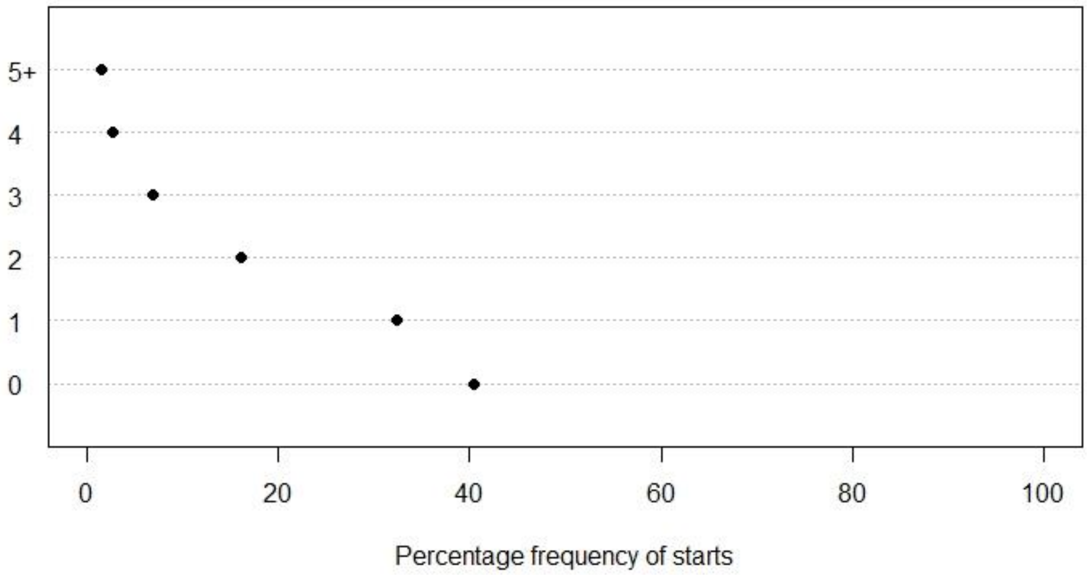

Figure 2-39 Dot plot of No. of layups

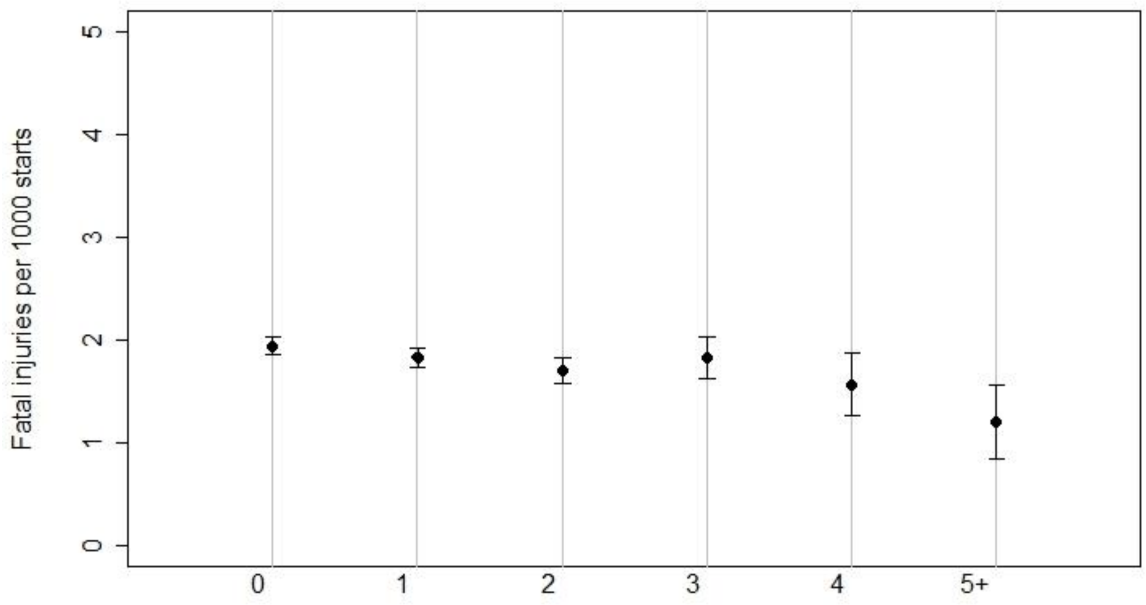

Figure 2-40 Fatal injuries per 1000 racing starts with 95% confidence interval by No. of layups

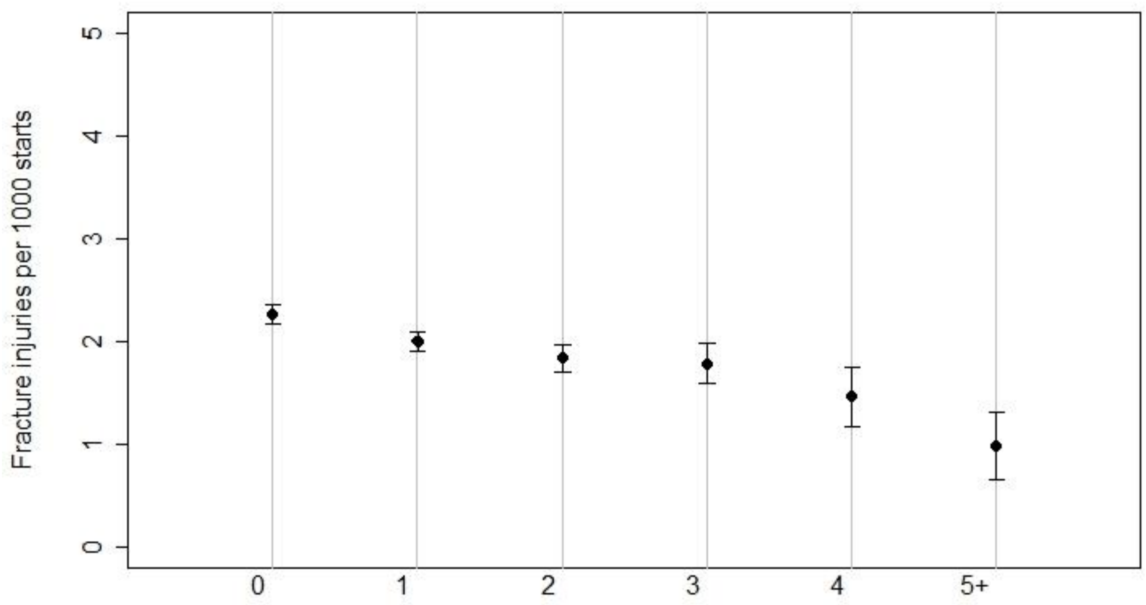

Figure 2-41 Fracture injuries per 1000 racing starts with 95% confidence interval by No. of layups

2.4.16. Number of previous injuries

From the racing and exercise starts available for each horse we calculated the number of EID recorded injuries sustained during a race a horse had throughout its career. The proportion of starts by number of previous injuries is shown at Figure 2-42 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-43 and 2-44 respectively.

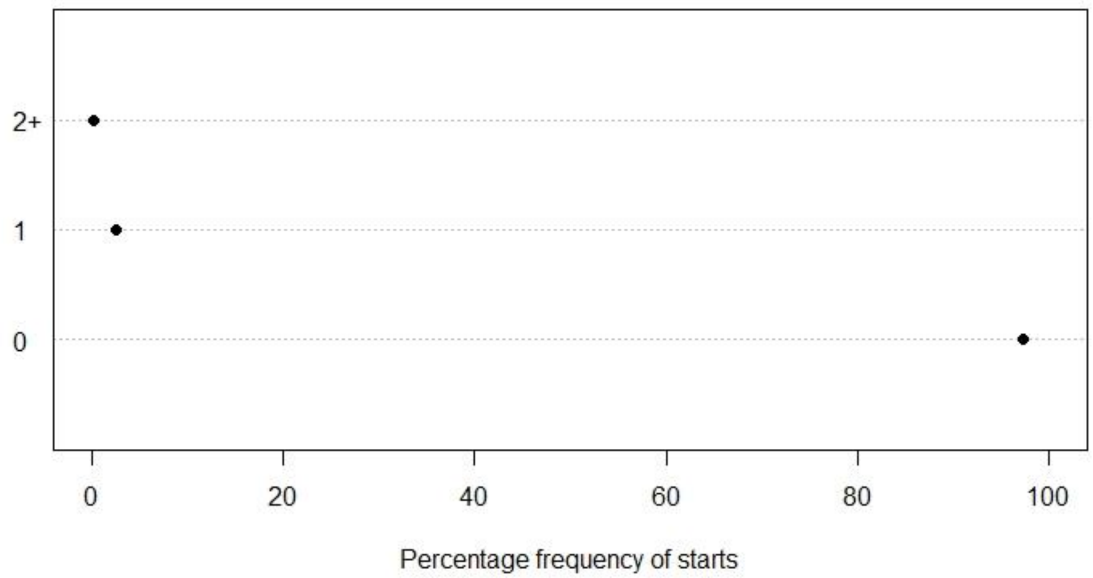

Figure 2-42 Dot plot of No. of previous injuries

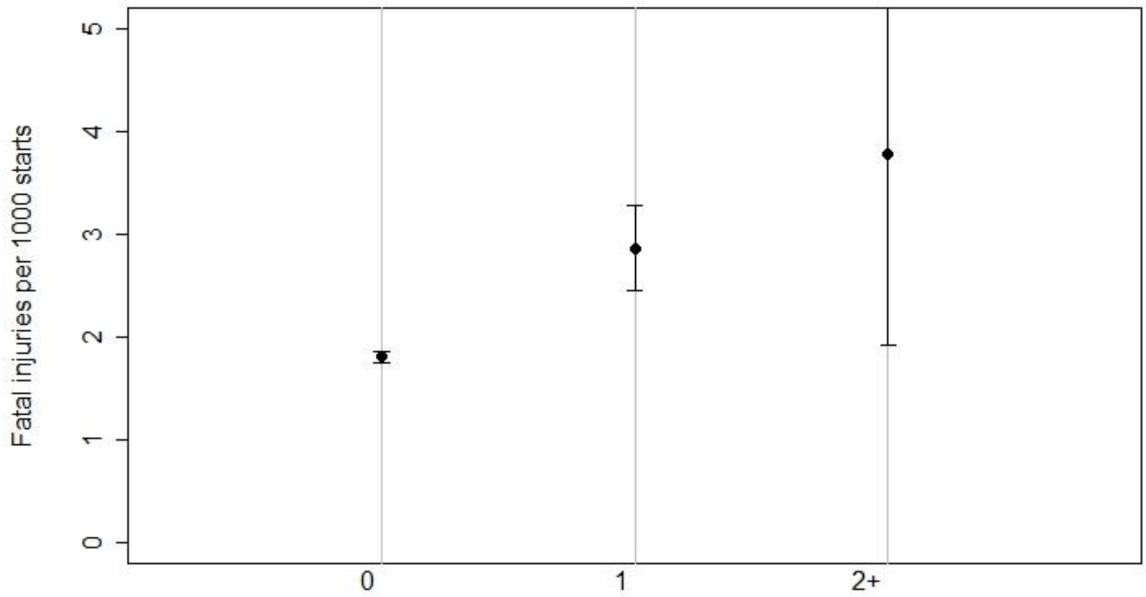

Figure 2-43 Fatal injuries per 1000 racing starts with 95% confidence interval by No. of previous injuries

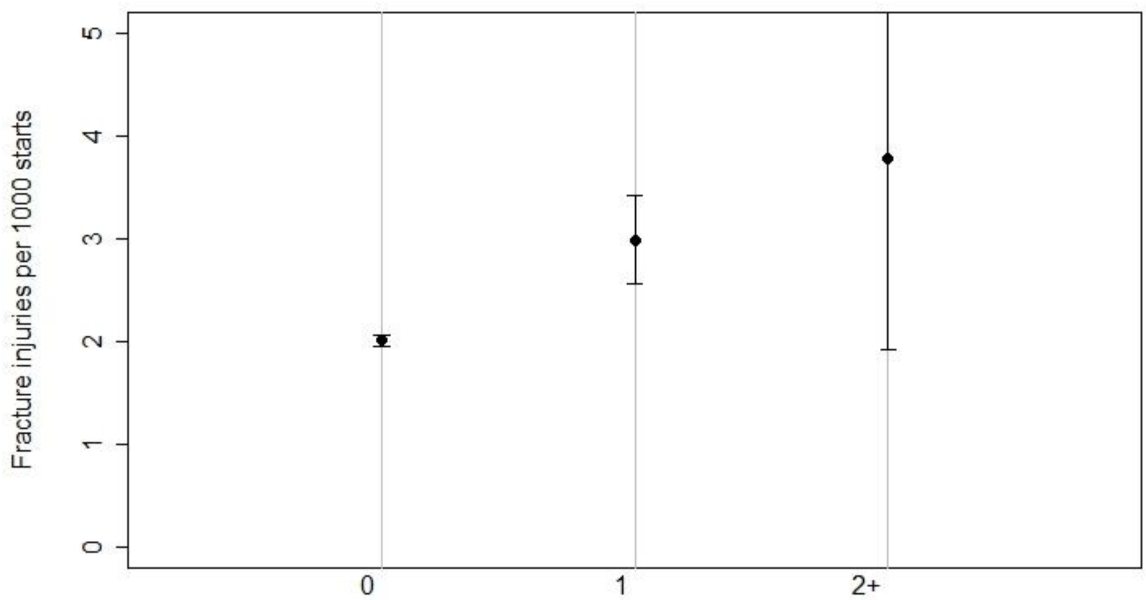

Figure 2-44 Fracture injuries per 1000 racing starts with 95% confidence interval by No. of previous injuries

2.4.17. Number of previous vet scratches

From the racing and exercise starts available for each horse we calculated the number of scratches by a veterinarian a horse had throughout its career. A vet scratch is a withdrawal of the horse from a race by the track veterinarian. The proportion of starts by number of previous veterinarian scratches is shown at Figure 2-45 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-46 and 2-47 respectively.

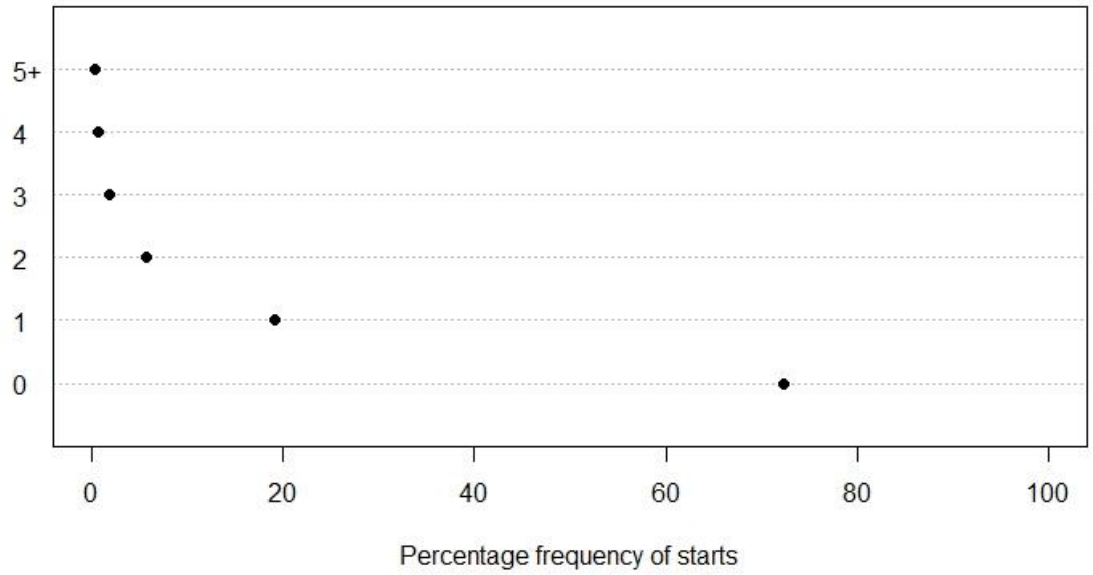

Figure 2-45 Dot plot of No. of previous vet scratches

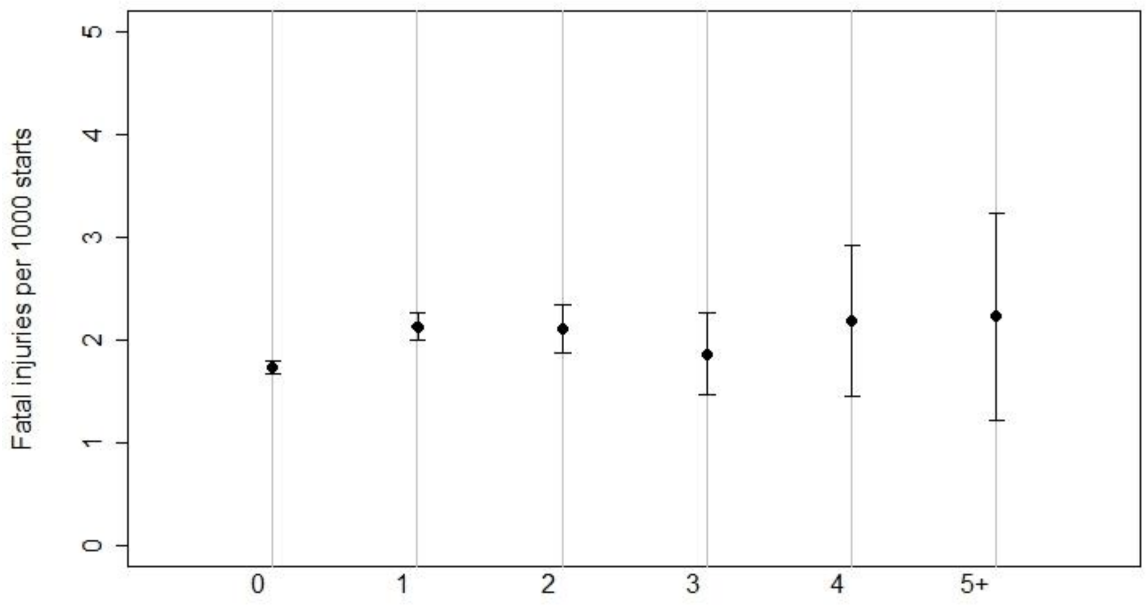

Figure 2-46 Fatal injuries per 1000 racing starts with 95% confidence interval by No. of previous vet scratches

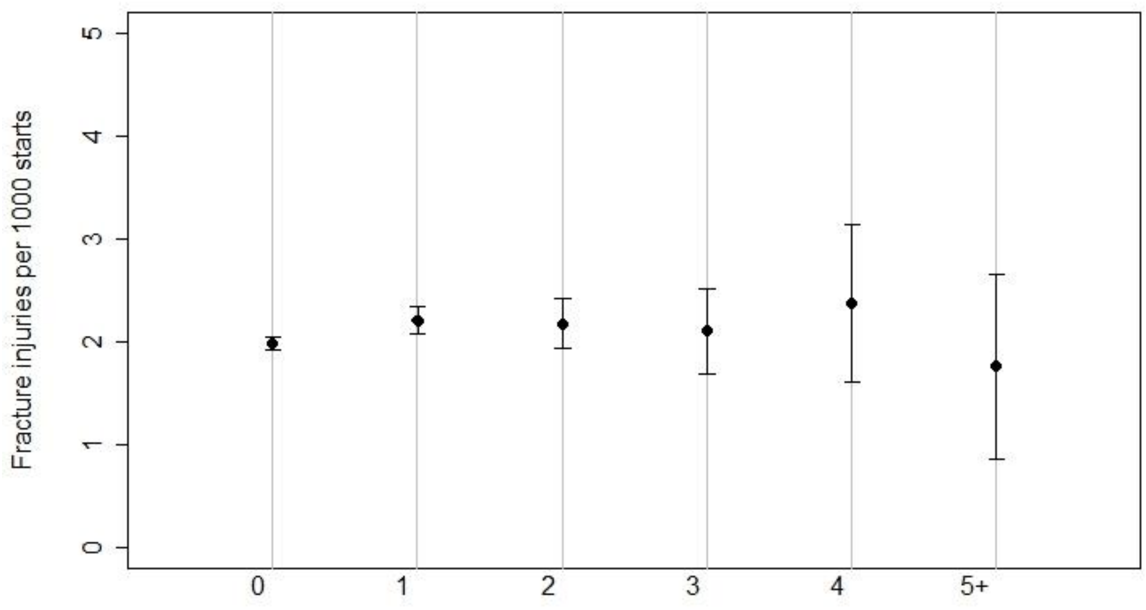

Figure 2-47 Fracture injuries per 1000 racing starts with 95% confidence interval by No. of previous vet scratches

2.4.18. Number of previous non-vet scratches

From the racing and exercise starts available for each horse we calculated the number of scratches not by a veterinarian a horse had throughout its career. A non-vet scratch is a withdrawal of the horse from a race not by the track veterinarian. The proportion of starts by number of previous non-veterinarian scratches is shown at Figure 2-48 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-49 and 2-50 respectively.

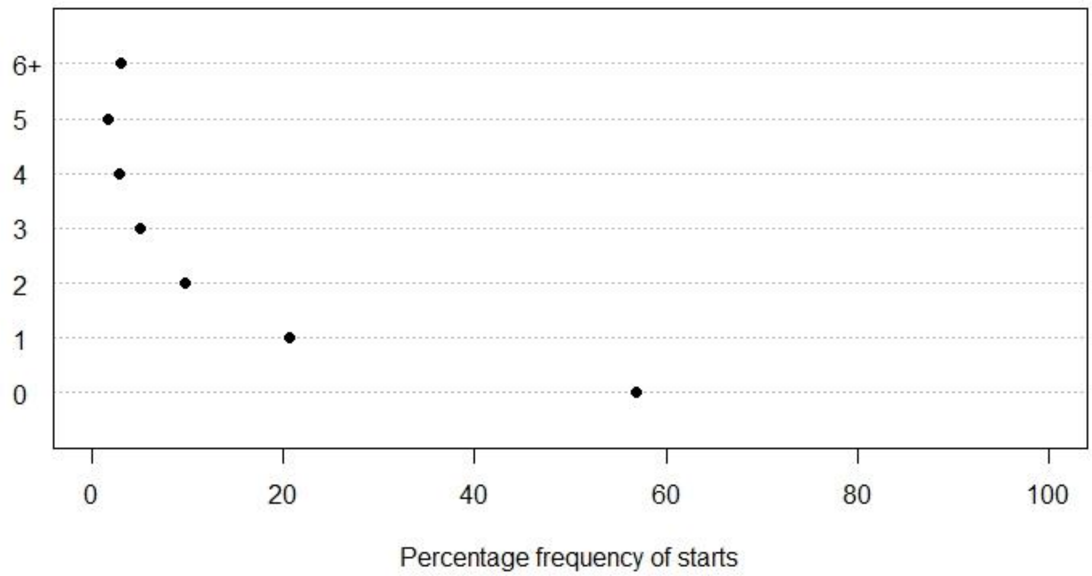

Figure 2-48 Dot plot of No. of previous non-vet scratches

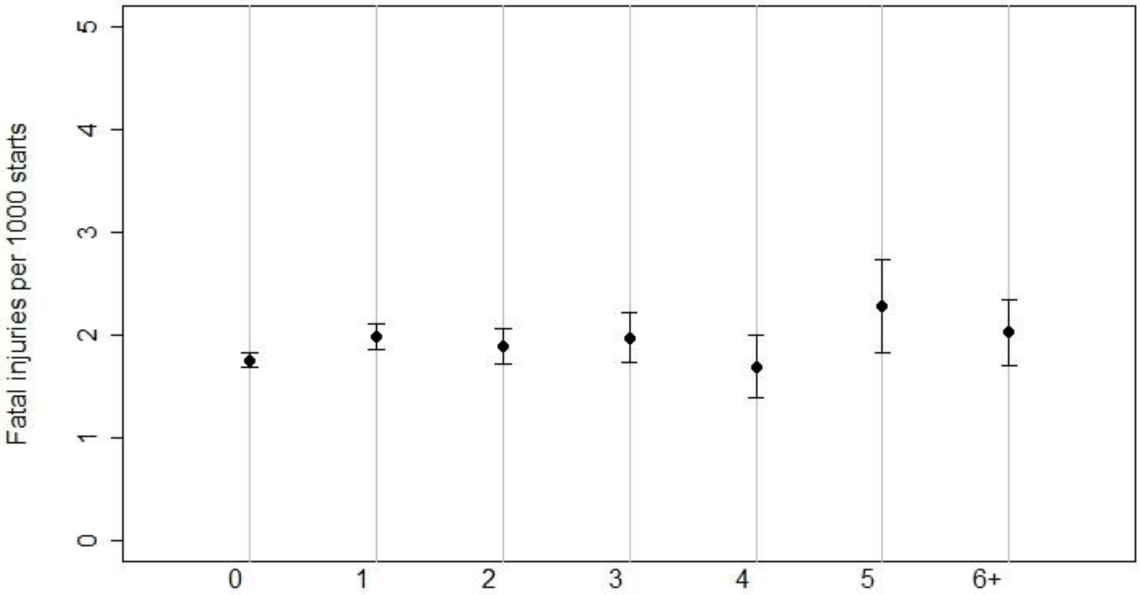

Figure 2-49 Fatal injuries per 1000 racing starts with 95% confidence interval by non-vet scratches

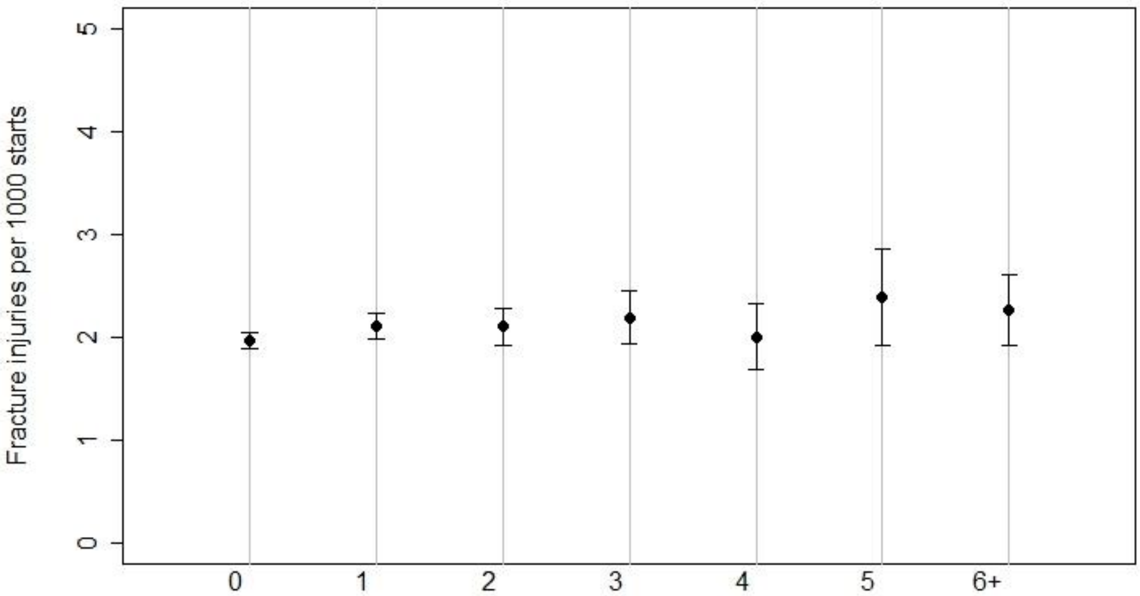

Figure 2-50 Fracture injuries per 1000 racing starts with 95% confidence interval by non-vet scratches

2.4.19. Number of racing and exercise starts (Present - 30 days prior race)

From the racing and exercise starts available for each horse we calculated the number of recorded racing and exercise starts a horse had in the period of 30 days prior the race. The proportion of starts by number of previous starts is shown at Figure 2-51 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-52 and 2-53 respectively.

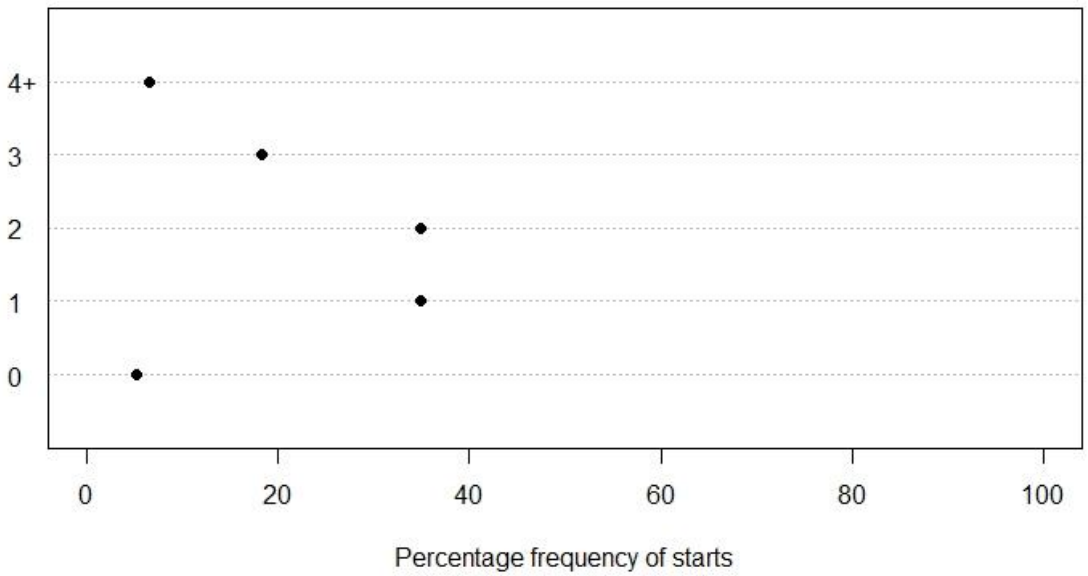

Figure 2-51 Dot plot of No. of racing and exercise starts (30 days prior race)

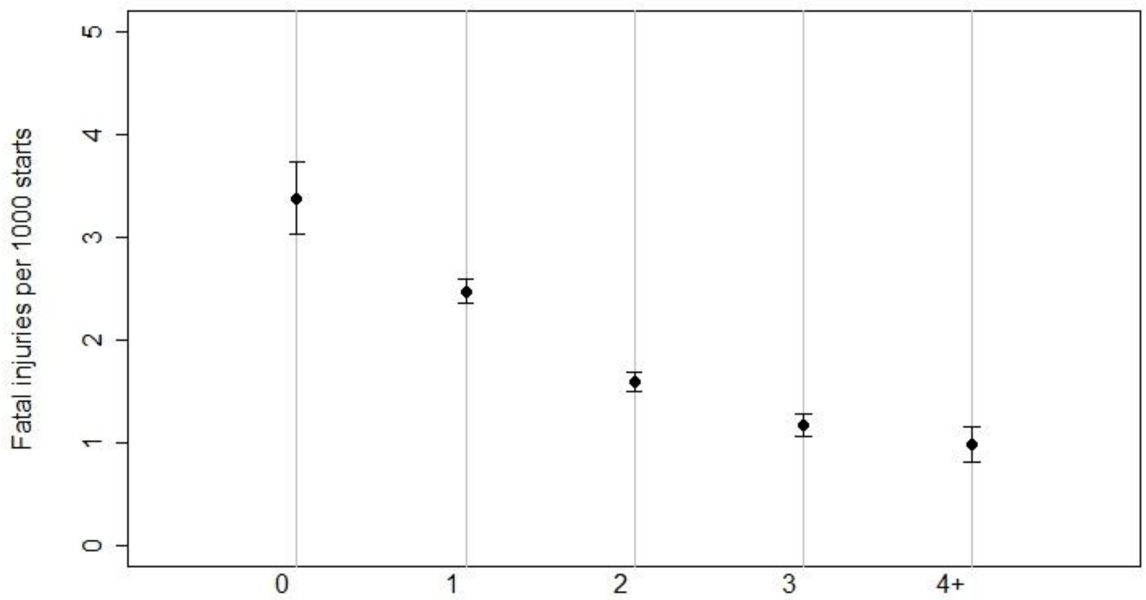

Figure 2-52 Fatal injuries per 1000 racing starts with 95% confidence interval by No. of racing and exercise starts (30 days prior race)

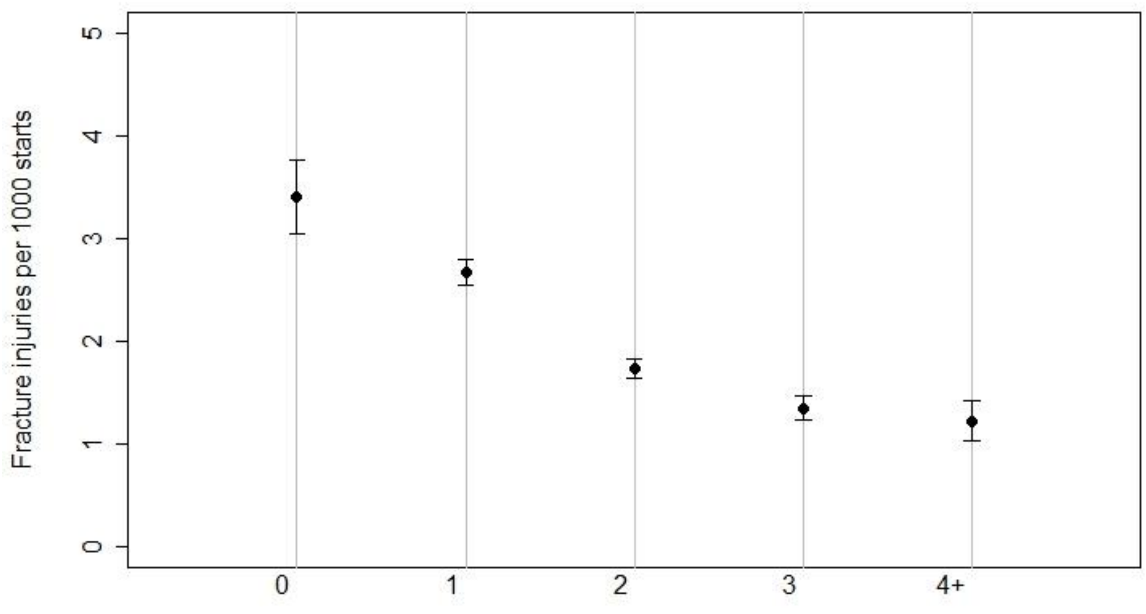

Figure 2-53 Fracture injuries per 1000 racing starts with 95% confidence interval by No. of racing and exercise starts (30 days prior race)

2.4.20. Number of racing and exercise starts (30 - 60 days prior race)

From the racing and exercise starts available for each horse we calculated the number of recorded racing and exercise starts a horse had in the period of 30 to 60 days prior the race. The proportion of starts by number of previous starts is shown at Figure 2-54 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-55 and 2-56 respectively.

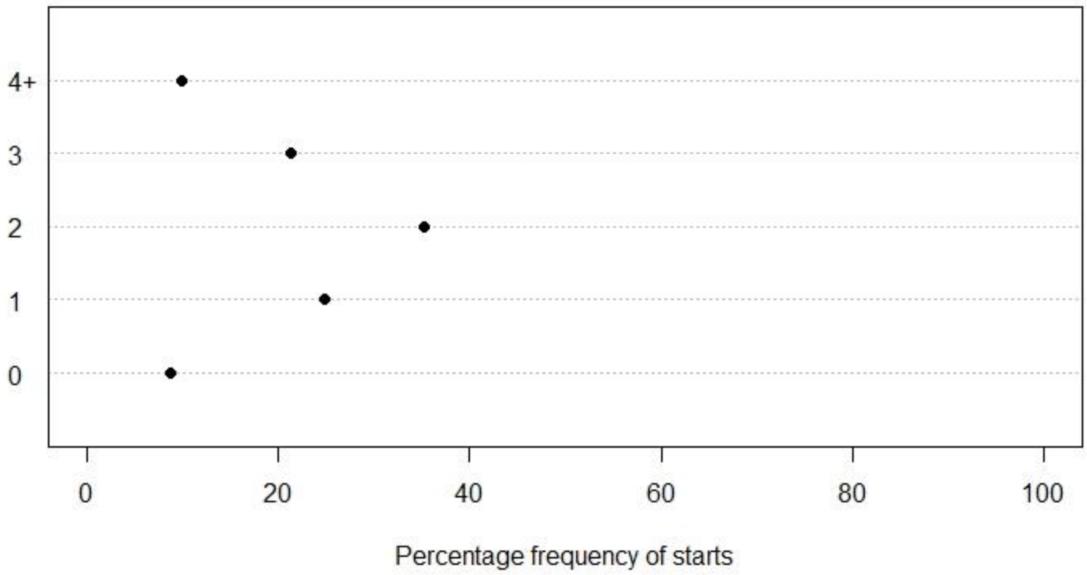

Figure 2-54 Dot plot of No. of racing and exercise starts (30 - 60 days prior race)

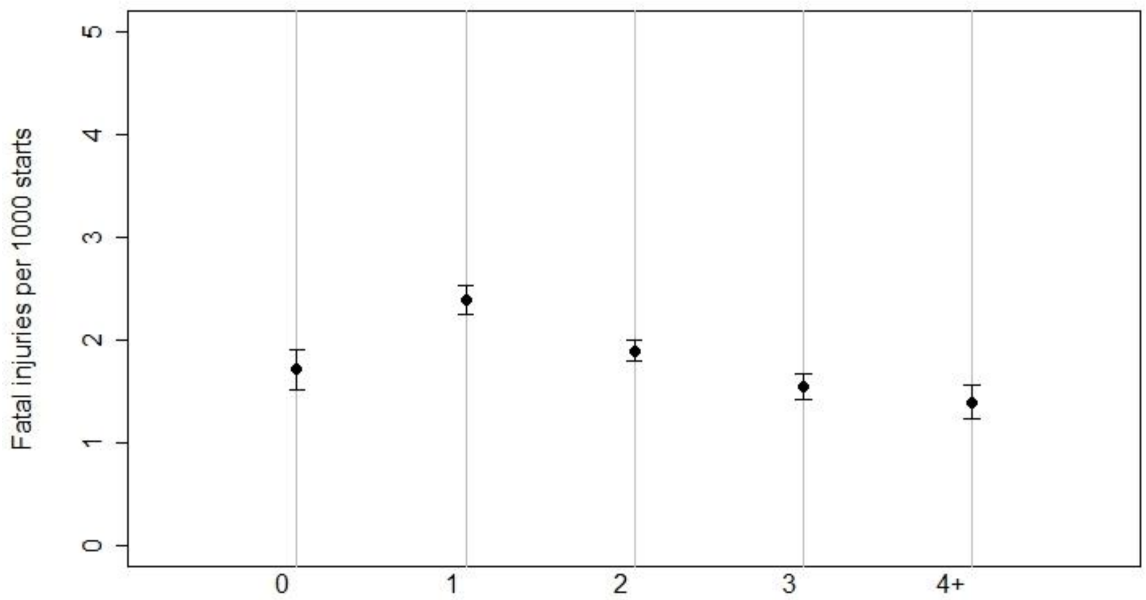

Figure 2-55 Fatal injuries per 1000 racing starts with 95% confidence interval by No. of racing and exercise starts (30 - 60 days prior race)

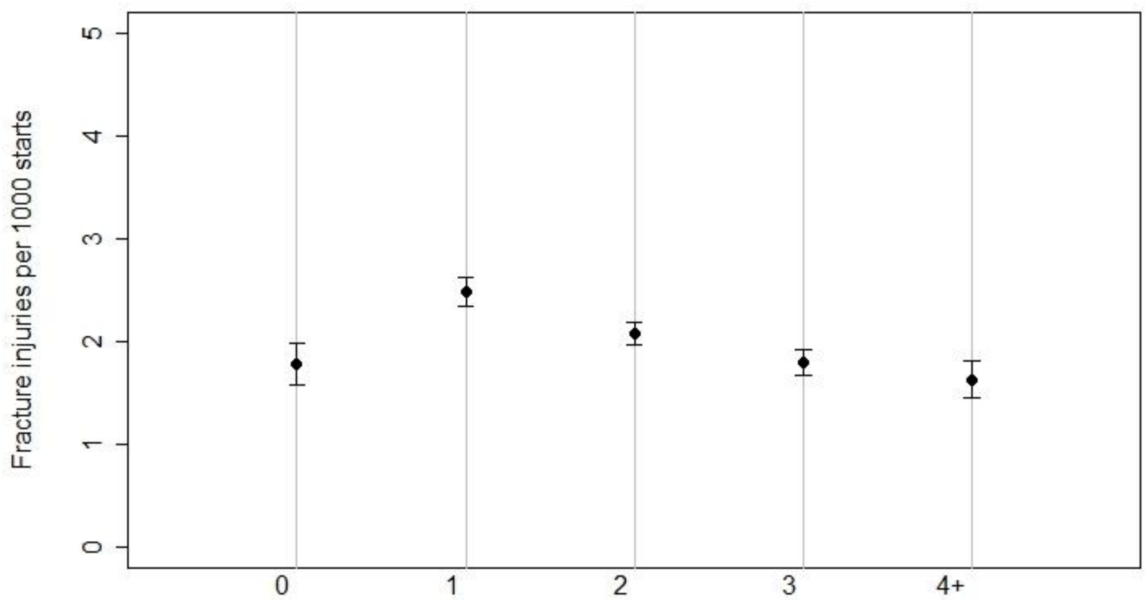

Figure 2-56 Fracture injuries per 1000 racing starts with 95% confidence interval by No. of racing and exercise starts (30 - 60 days prior race)

2.4.21. Number of racing and exercise starts (60 -90 days prior race)

From the racing and exercise starts available for each horse we calculated the number of recorded racing and exercise starts a horse had in the period of 60 to 90 days prior the race. The proportion of starts by number of previous starts is shown at Figure 2-57 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-58 and 2-59 respectively.

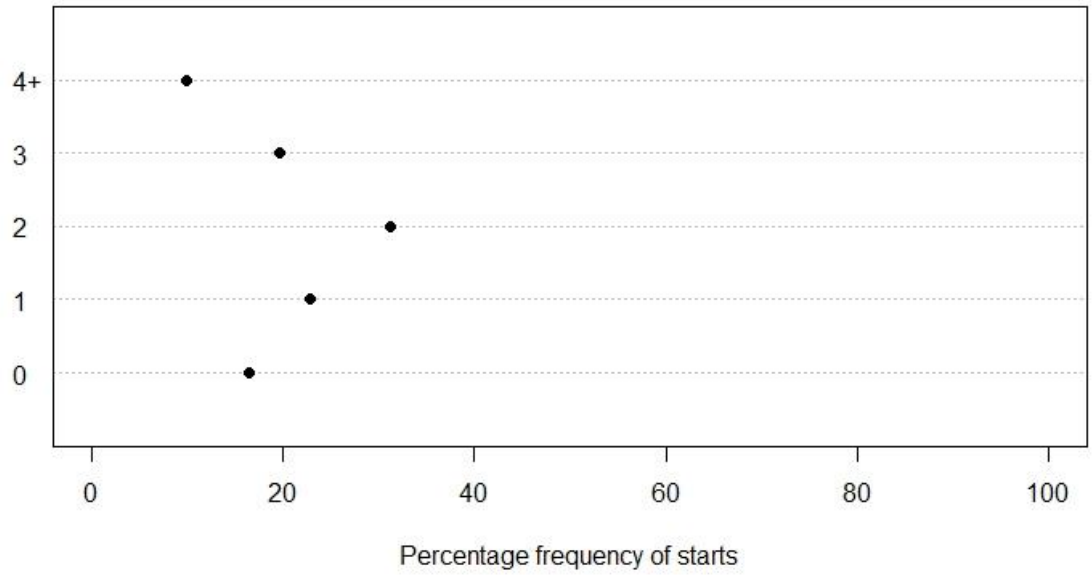

Figure 2-57 Dot plot of No. of racing and exercise starts (60 -90 days prior race)

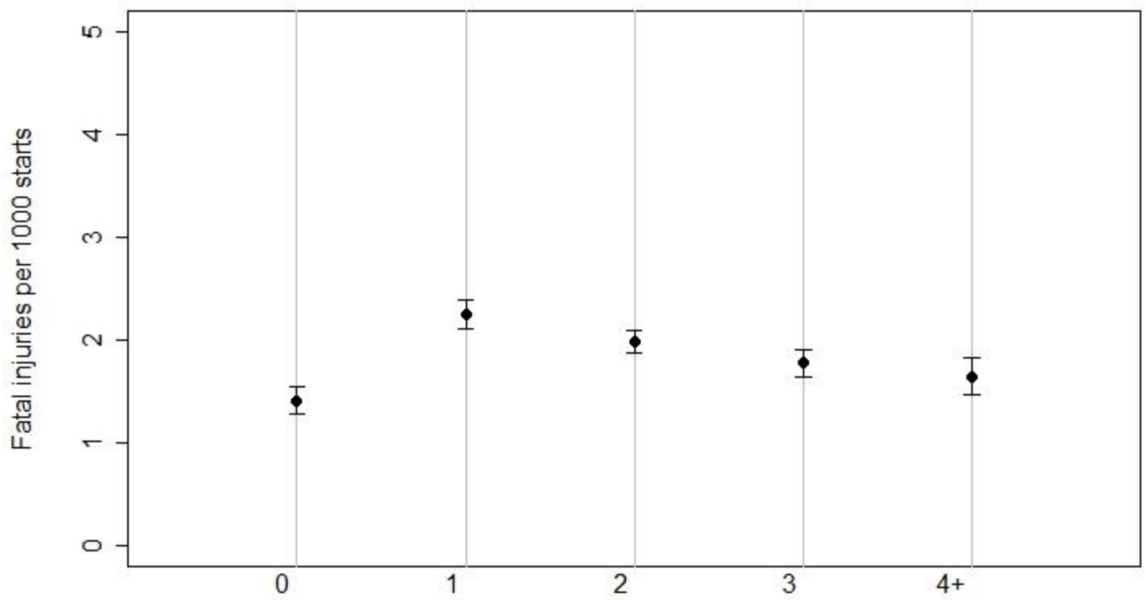

Figure 2-58 Fatal injuries per 1000 racing starts with 95% confidence interval by No. of racing and exercise starts (60 -90 days prior race)

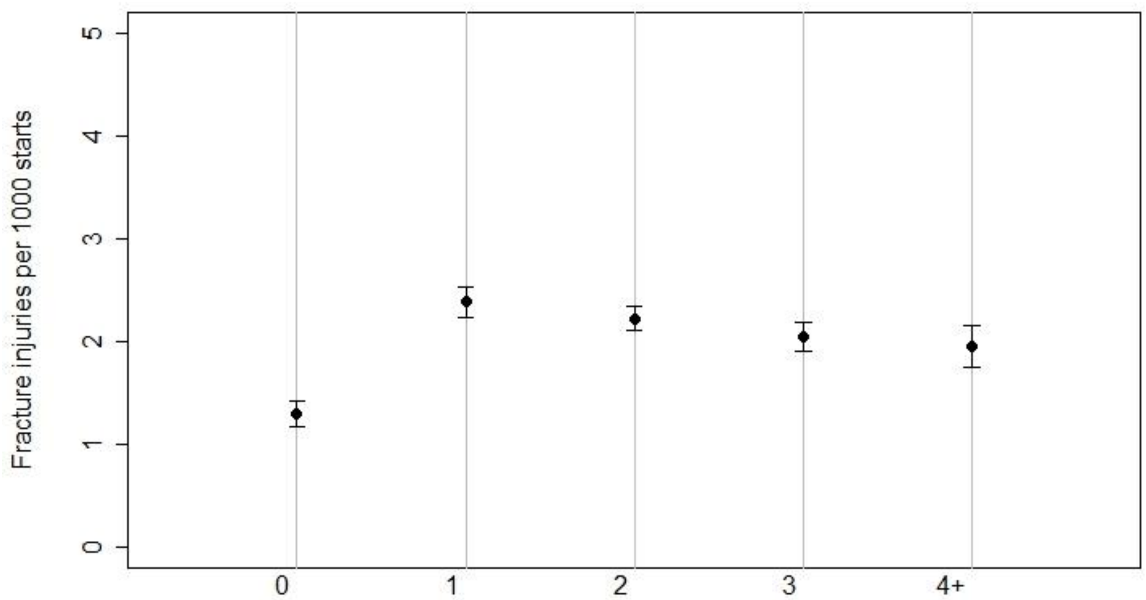

Figure 2-59 Fracture injuries per 1000 racing starts with 95% confidence interval by No. of racing and exercise starts (60 -90 days prior race)

2.4.22. Number of racing and exercise starts (90 -180 days prior race)

From the racing and exercise starts available for each horse we calculated the number of recorded racing and exercise starts a horse had in the period of 90 to 180 days prior the race. The proportion of starts by number of previous starts is shown at Figure 2-60 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-61 and 2-62 respectively.

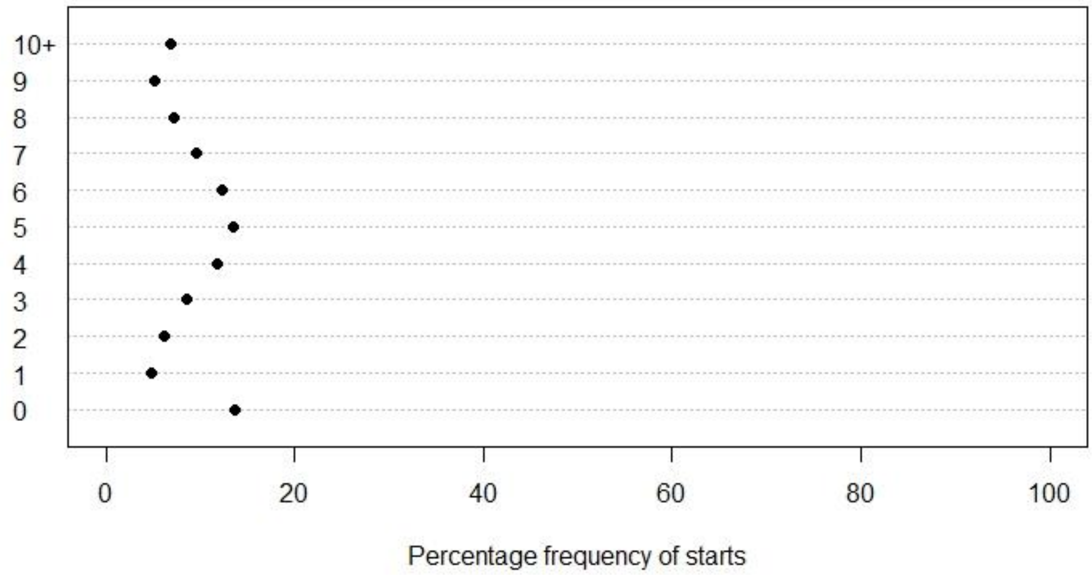

Figure 2-60 Dot plot of No. of racing and exercise starts (90 -180 days prior race)

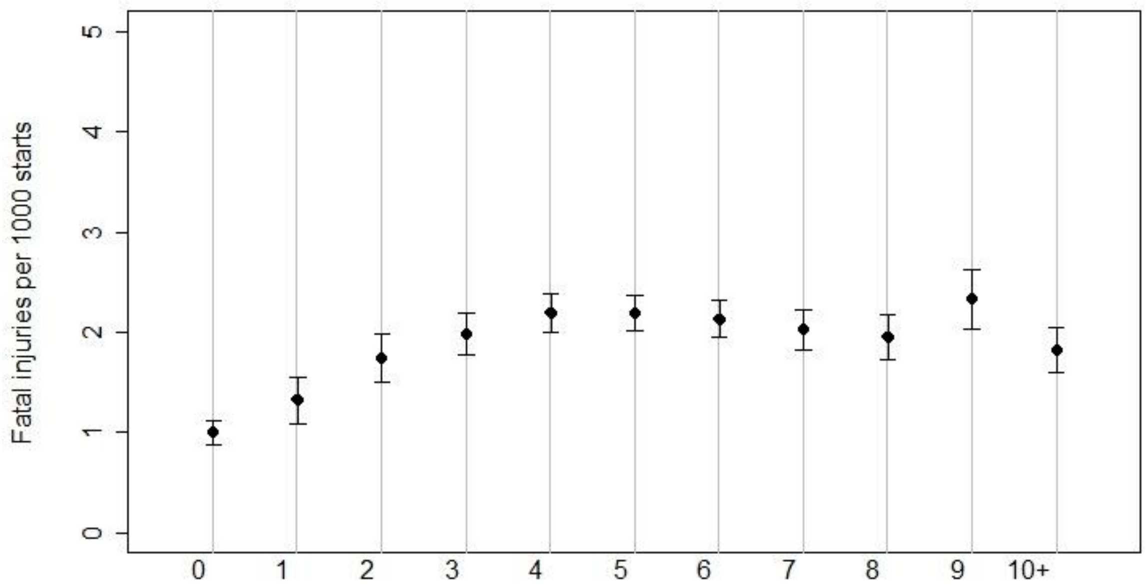

Figure 2-61 Fatal injuries per 1000 racing starts with 95% confidence interval by No. of racing and exercise starts (90 -180 days prior race)

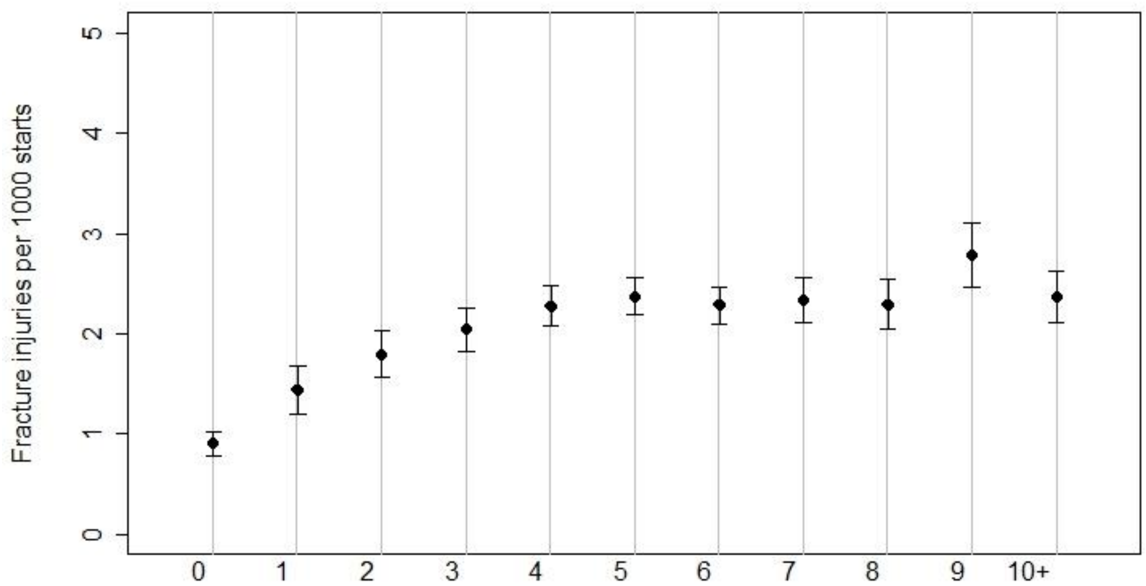

Figure 2-62 Fracture injuries per 1000 racing starts with 95% confidence interval by No. of racing and exercise starts (90 -180 days prior race)

2.4.23. Number of starts (Present – 30 days prior race)

From the racing starts available for each horse we calculated the number of racing starts a horse had in the period of 30 days prior the race. The proportion of starts by number of previous starts is shown at Figure 2-63 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-64 and 2-65 respectively.

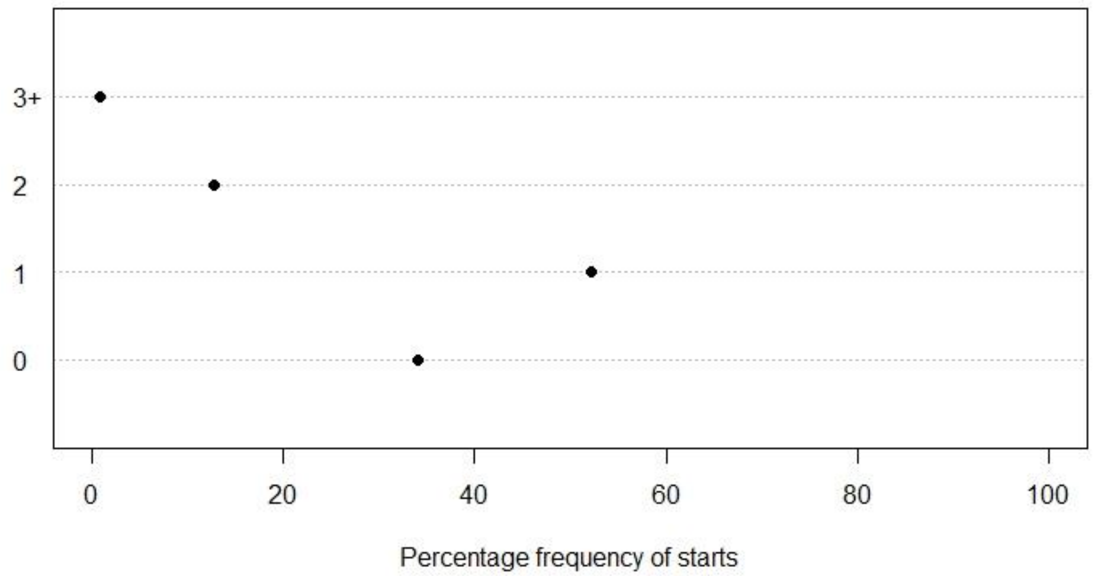

Figure 2-63 Dot plot of No. of starts (Present – 30 days prior race)

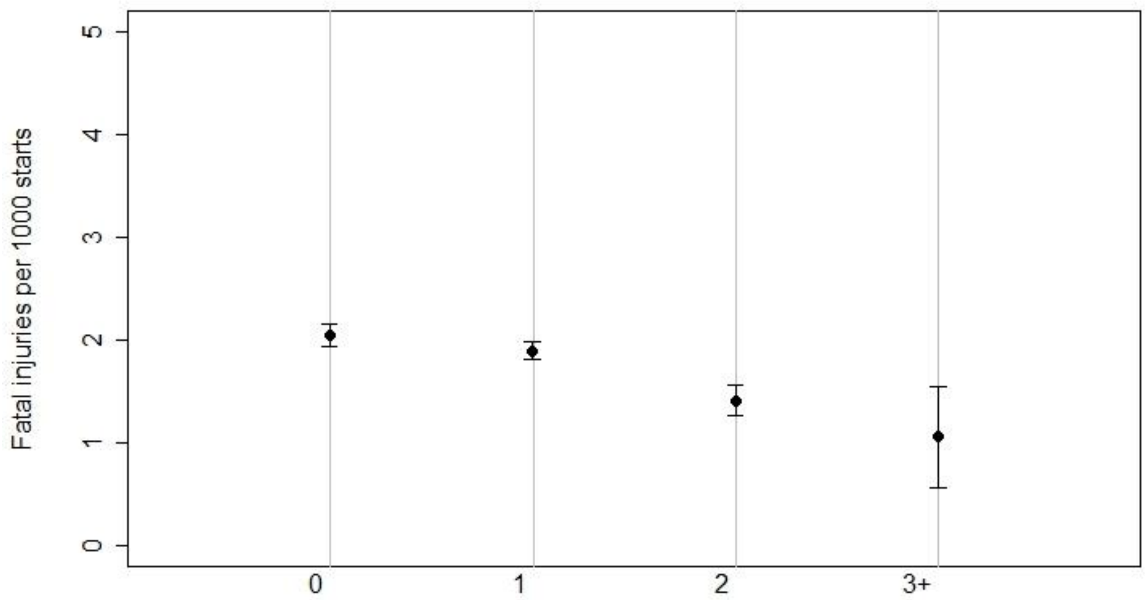

Figure 2-64 Fatal injuries per 1000 racing starts with 95% confidence interval by No. of starts (Present – 30 days prior race)

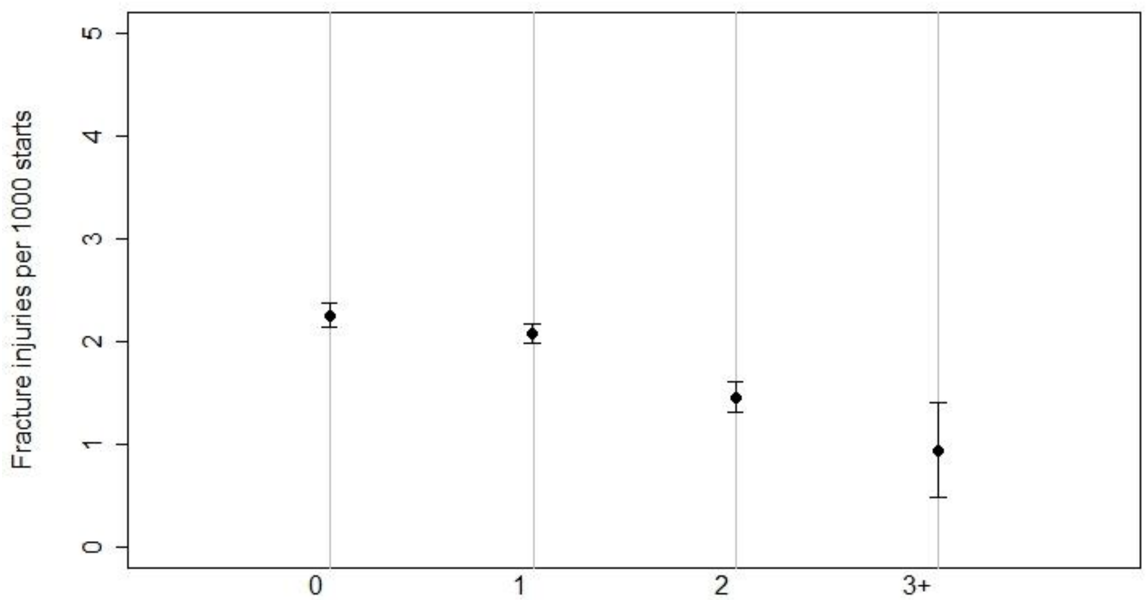

Figure 2-65 Fracture injuries per 1000 racing starts with 95% confidence interval by No. of starts (Present – 30 days prior race)

2.4.24. Number of starts (30 – 60 days prior race)

From the racing starts available for each horse we calculated the number of racing starts a horse had in the period of 30 to 60 days prior the race. The proportion of starts by number of previous starts is shown at Figure 2-66 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-67 and 2-68 respectively.

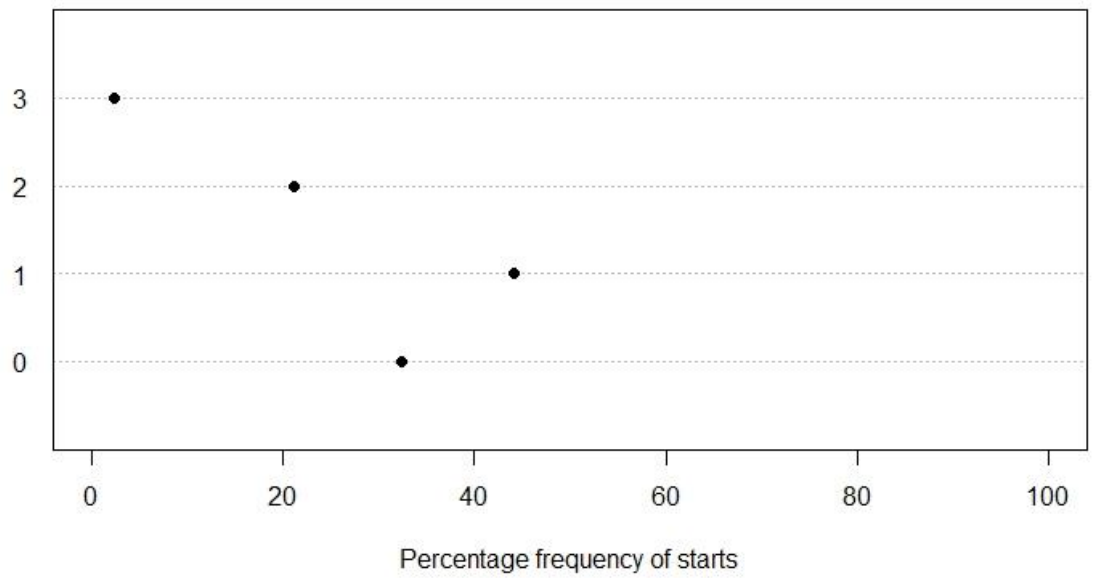

Figure 2-66 Dot plot of No. of starts (30 – 60 days prior race)

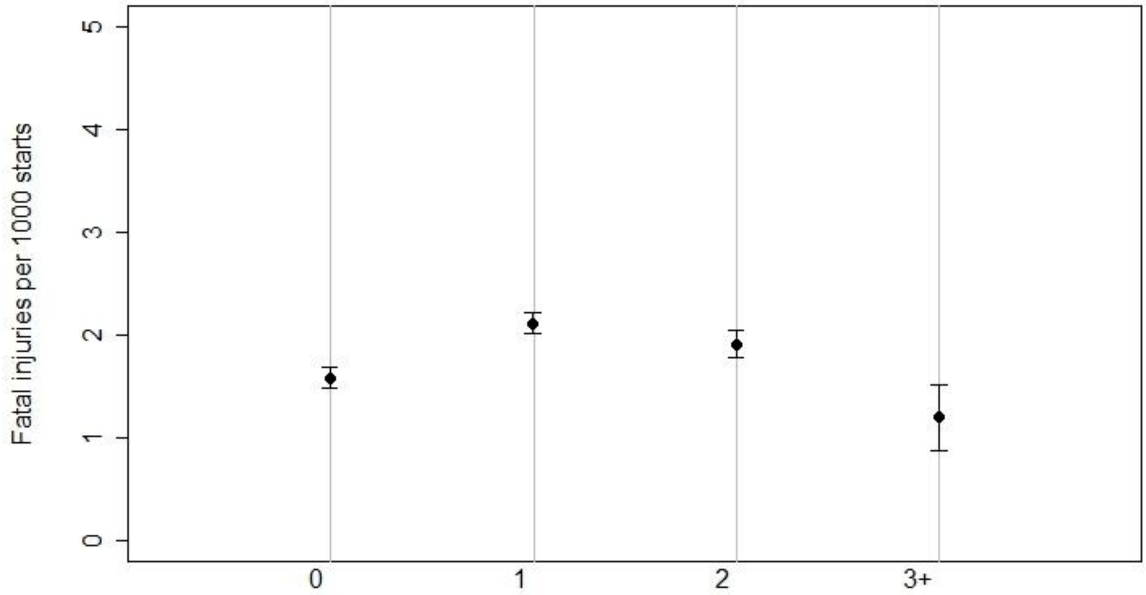

Figure 2-67 Fatal injuries per 1000 racing starts with 95% confidence interval by No. of starts (30 – 60 days prior race)

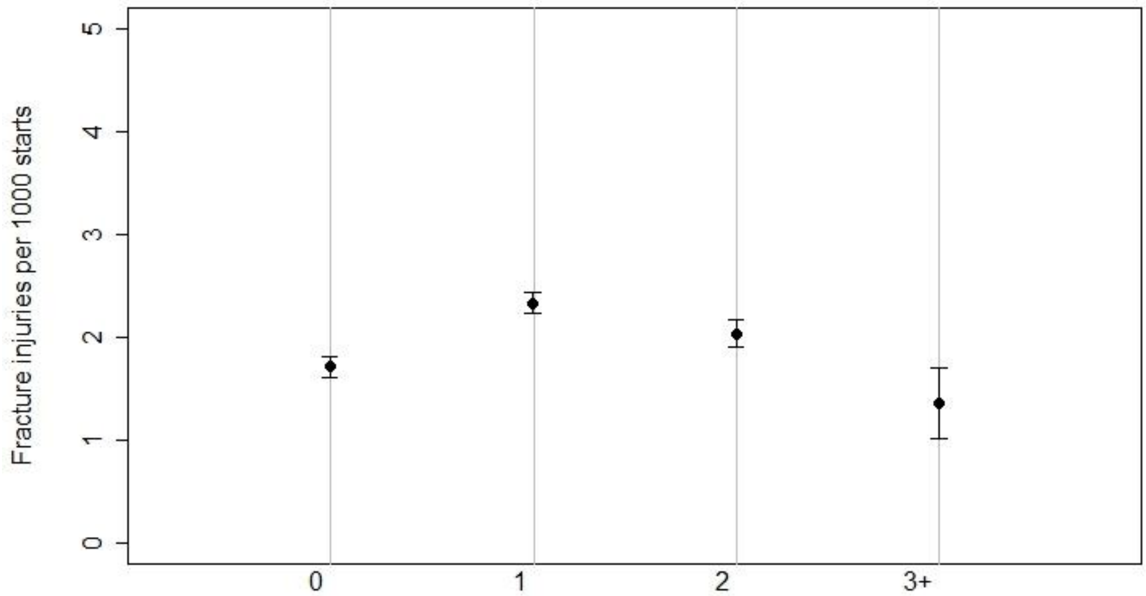

Figure 2-68 Fracture injuries per 1000 racing starts with 95% confidence interval by No. of starts (30 – 60 days prior race)

2.4.25. Number of starts (60 – 90 days prior race)

From the racing starts available for each horse we calculated the number of racing starts a horse had in the period of 60 to 90 days prior the race. The proportion of starts by number of previous starts is shown at Figure 2-69 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-70 and 2-71 respectively.

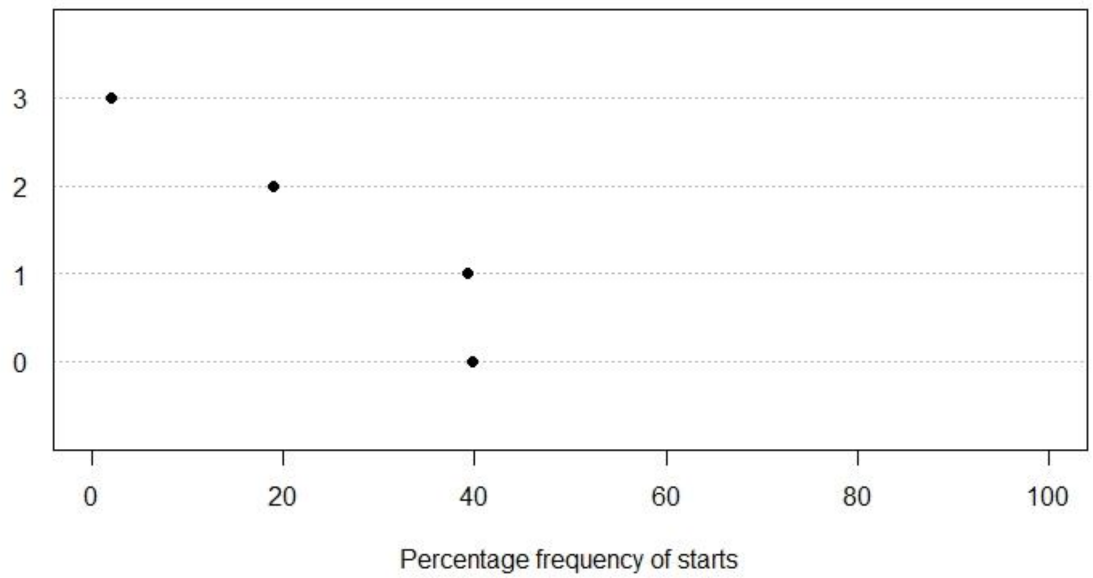

Figure 2-69 Dot plot of No. of starts (60 – 90 days prior race)

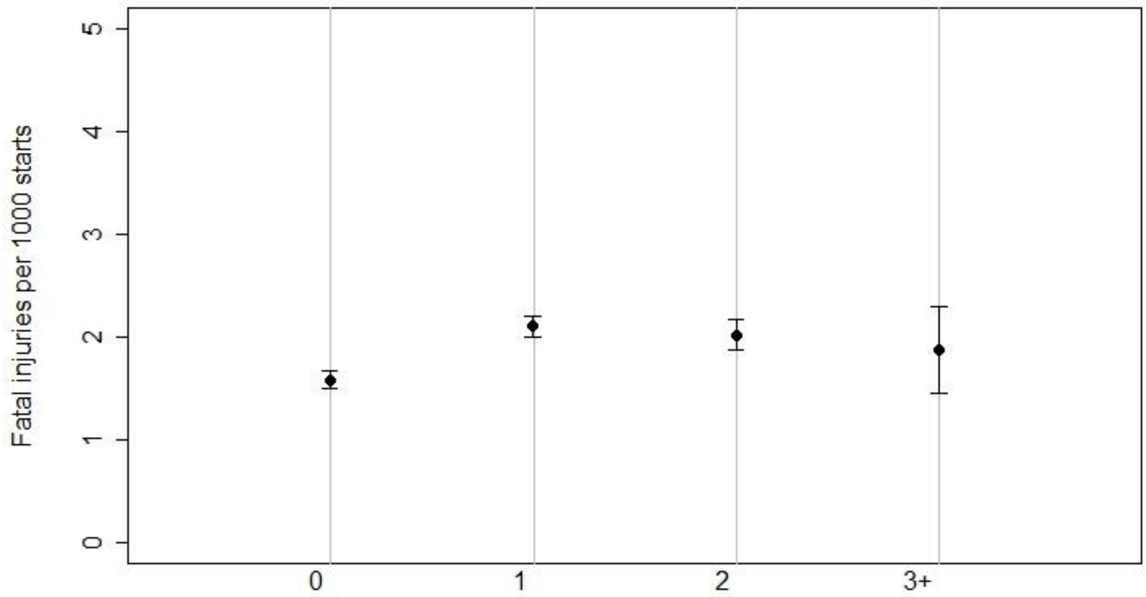

Figure 2-70 Fatal injuries per 1000 racing starts with 95% confidence interval by No. of starts (60 – 90 days prior race)

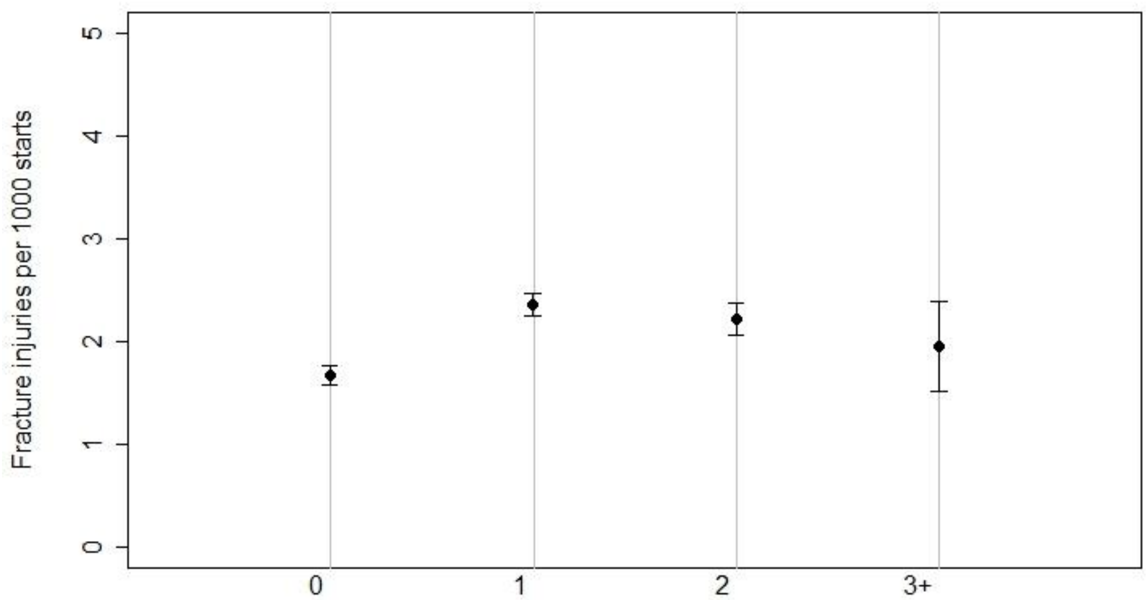

Figure 2-71 Fracture injuries per 1000 racing starts with 95% confidence interval by No. of starts (60 – 90 days prior race)

2.4.26. Number of starts (90 – 180 days prior race)

From the racing starts available for each horse we calculated the number of racing starts a horse had in the period of 90 to 180 days prior the race. The proportion of starts by number of previous starts is shown at Figure 2-72 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-73 and 2-74 respectively.

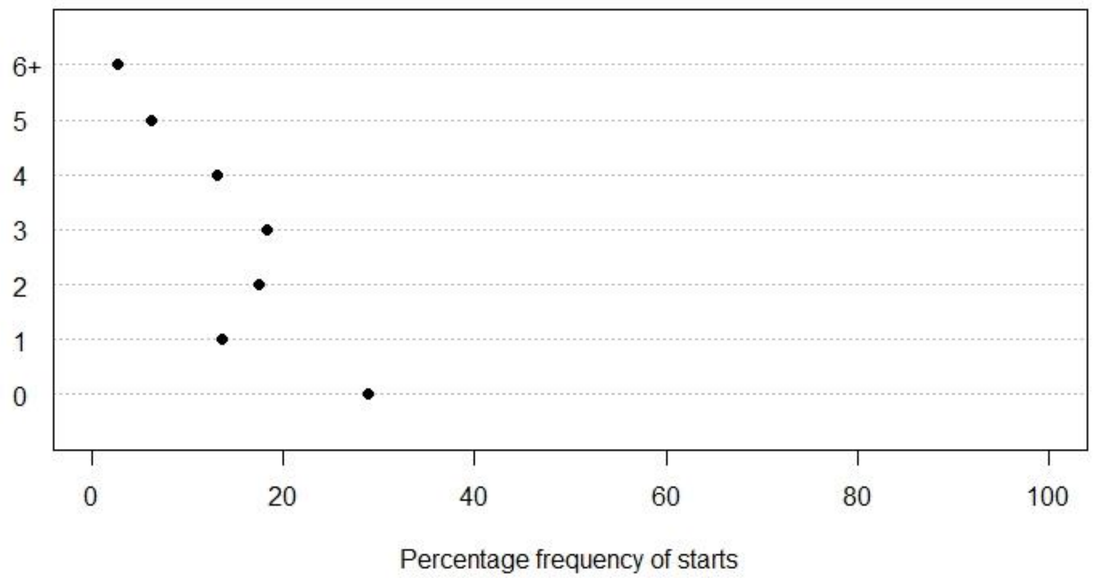

Figure 2-72 Dot plot of No. of starts (90 – 180 days prior race)

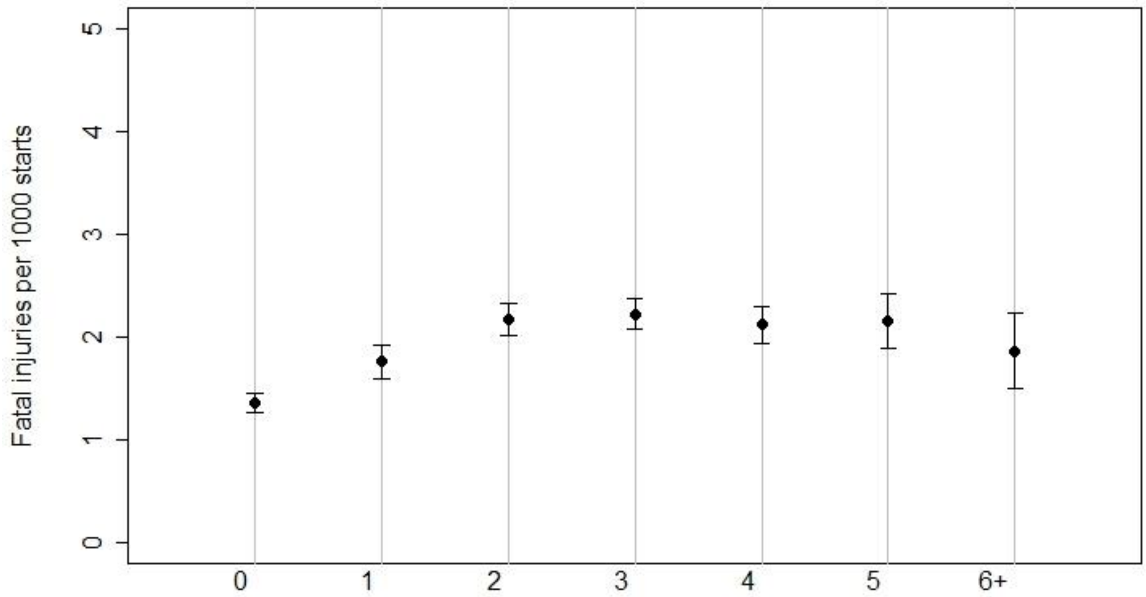

Figure 2-73 Fatal injuries per 1000 racing starts with 95% confidence interval by No. of starts (90 – 180 days prior race)

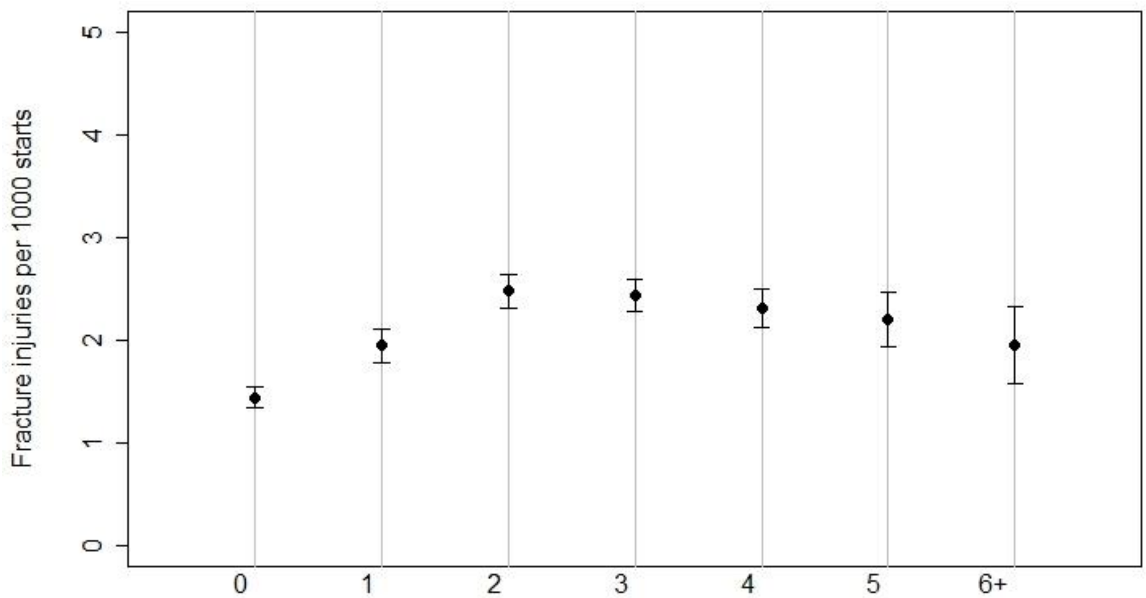

Figure 2-74 Fracture injuries per 1000 racing starts with 95% confidence interval by No. of starts (90 – 180 days prior race)

2.4.27. Odds at start of race

The EID contained information on the betting odds of each horse for each race. The density of starts by the odds is shown at Figure 2-75 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for different speed groups are shown at figures 2-76 and 2-77 respectively.

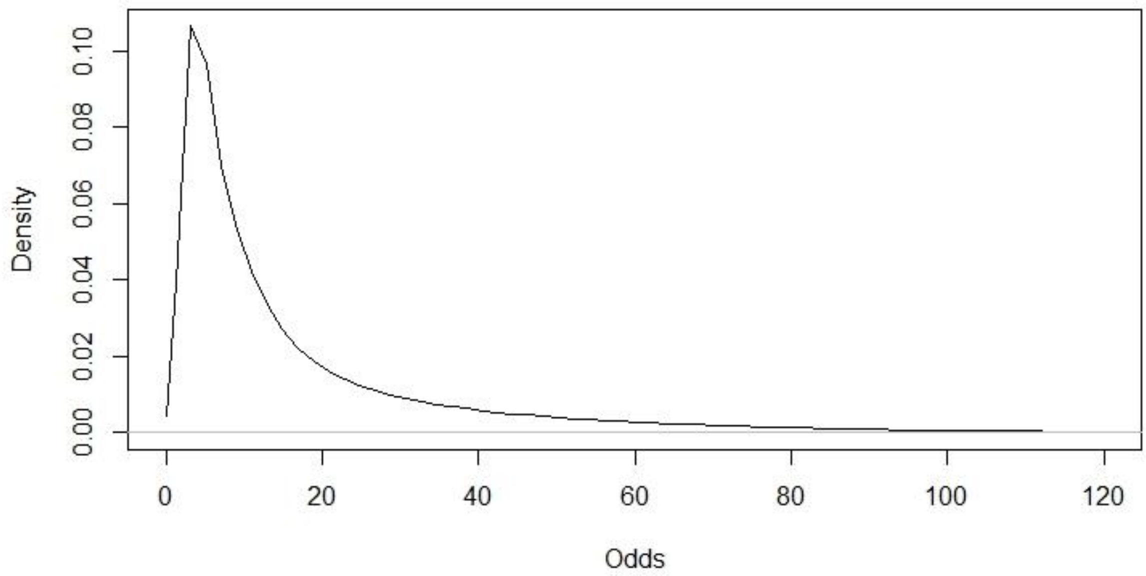

Figure 2-75 Density plot of odds at start of race

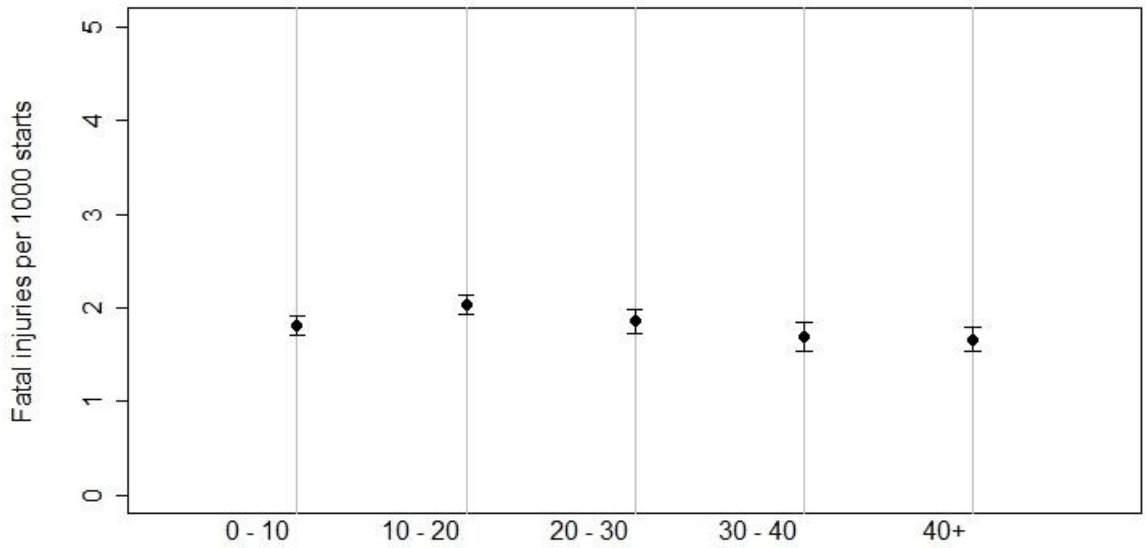

Figure 2-76 Fatal injuries per 1000 racing starts with 95% confidence interval by odds group

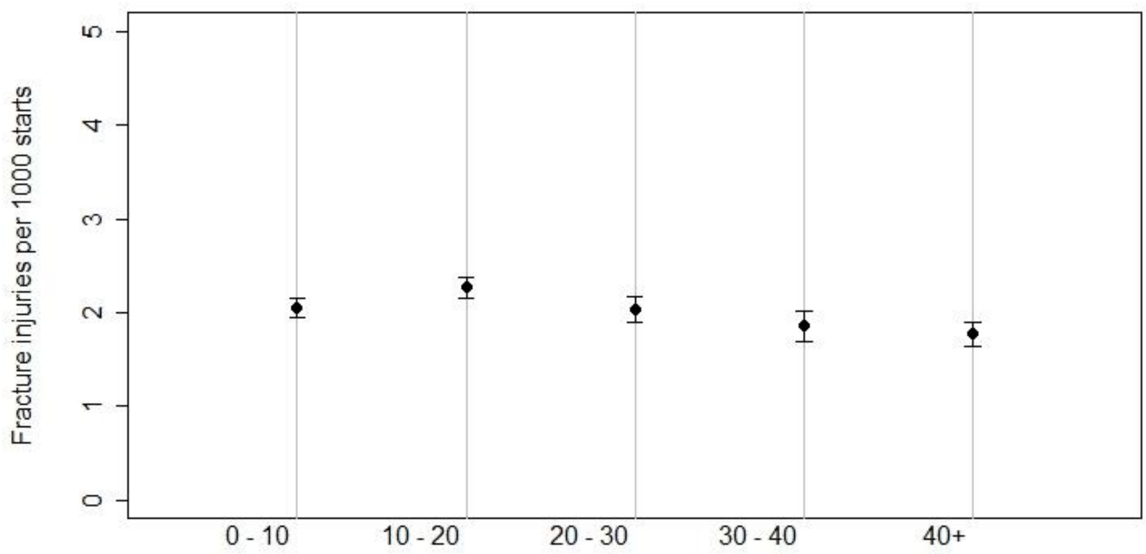

Figure 2-77 Fracture injuries per 1000 racing starts with 95% confidence interval by odds group

2.4.28. Odds rank in race

The EID contained information on the betting odds of each horse for each race. We ranked each horse in each race by its odd, the favored horse to win the race being ranked first. When two horses have the same odds in a race they are both assigned the same lower rank. The proportion of starts by the odds rank is shown at Figure 2-78 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-79 and 2-80 respectively.

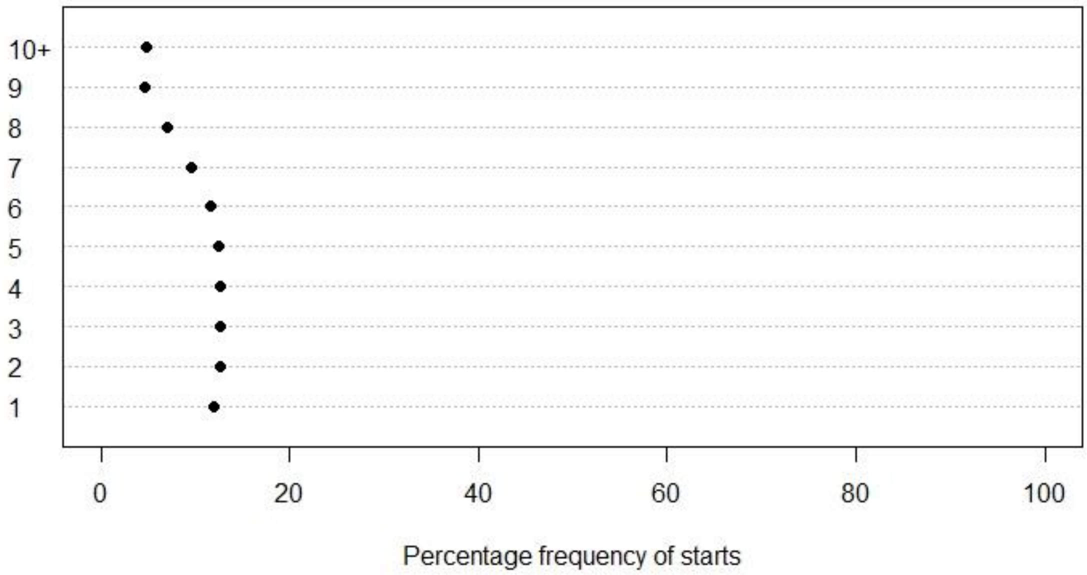

Figure 2-78 Dot plot of odds rank

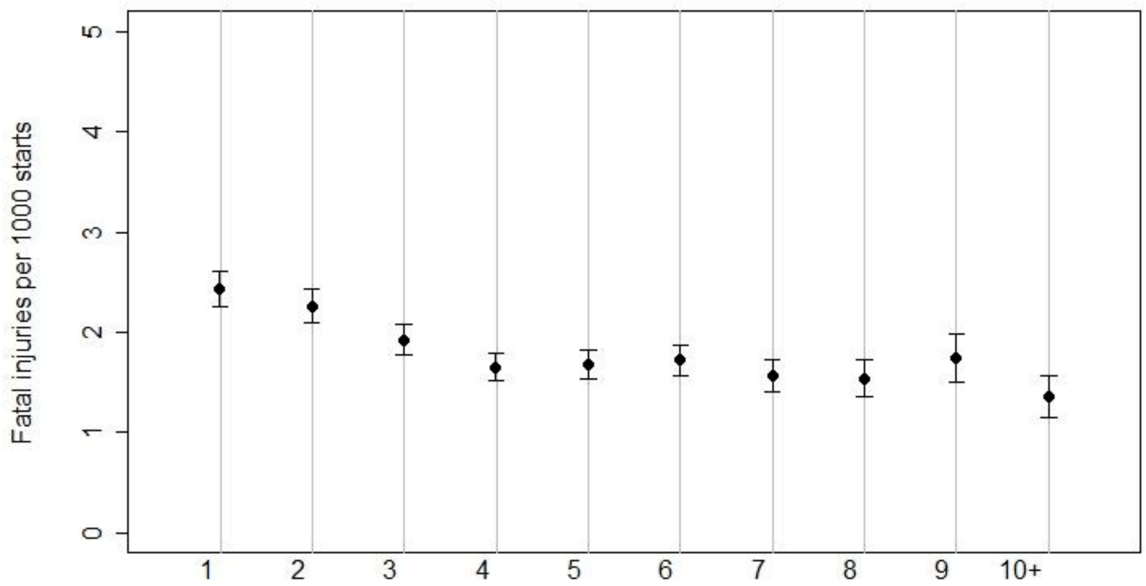

Figure 2-79 Fatal injuries per 1000 racing starts with 95% confidence interval by odds rank

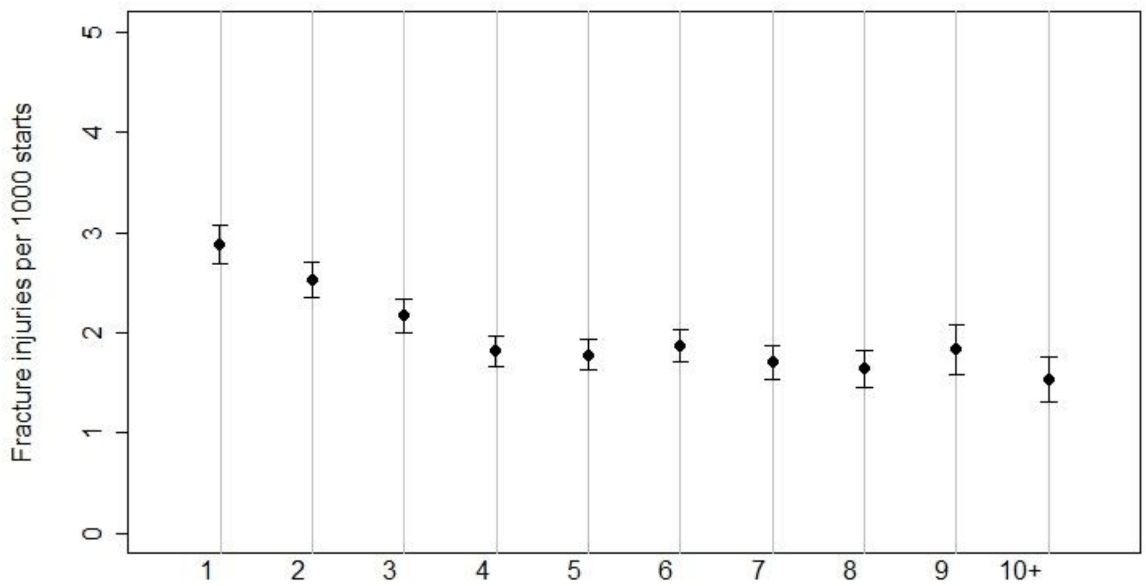

Figure 2-80 Fracture injuries per 1000 racing starts with 95% confidence interval by odds rank

2.4.29. Post position

The EID contained information on the post position of each horse for each race. Post position is the place each horse starts the race. Numbering starts from the horse closer to the inside rail. The proportion of starts by post position is shown at Figure 2-81 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-82 and 2-83 respectively.

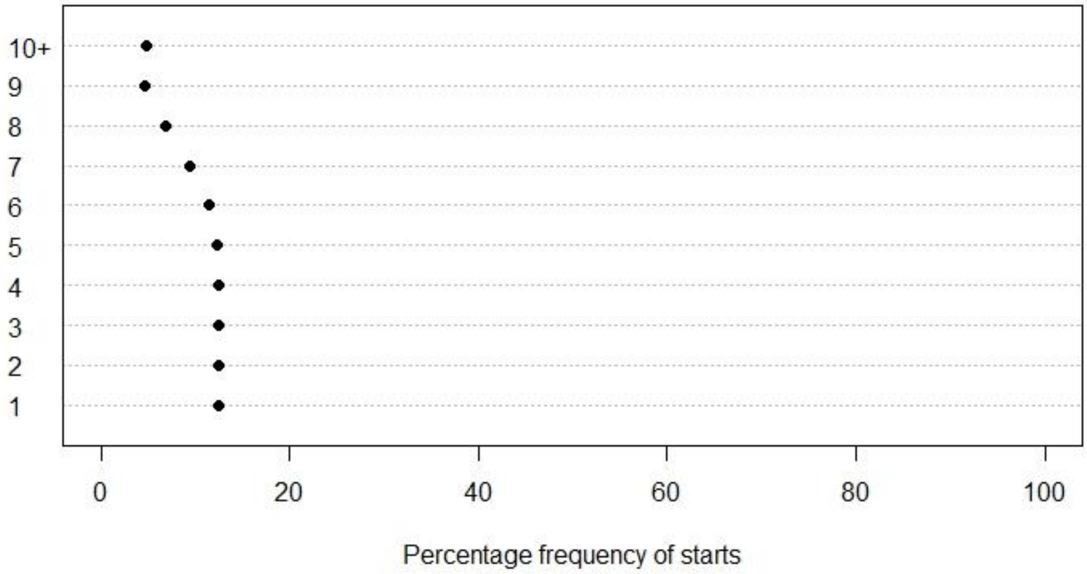

Figure 2-81 Dot plot of post position

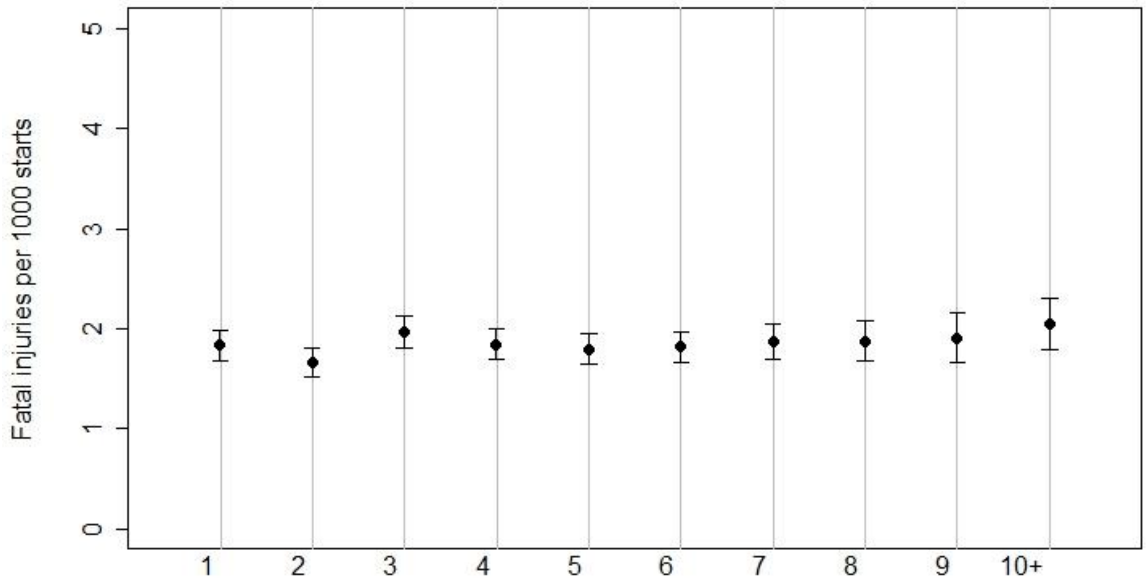

Figure 2-82 Fatal injuries per 1000 racing starts with 95% confidence interval by post position

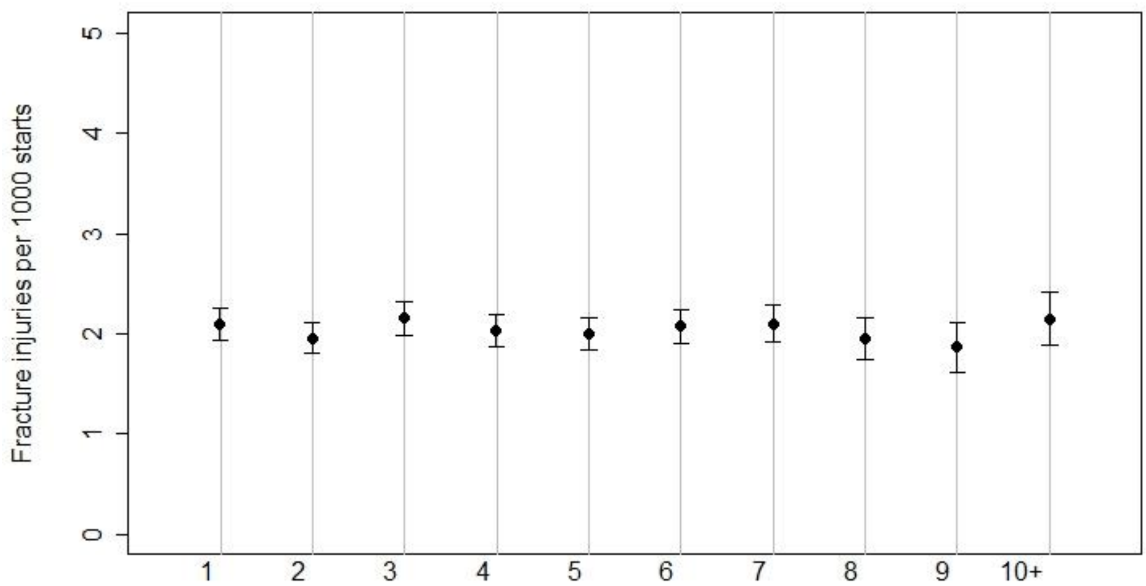

Figure 2-83 Fracture injuries per 1000 racing starts with 95% confidence interval by post position

2.4.30. Purse ($1000)

The EID contained information on the purse of each race. The purse is the total amount of prize money distributed to the winners of the race. The density of starts by the $1000 of purse is shown at Figure 2-84 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-85 and 2-86 respectively.

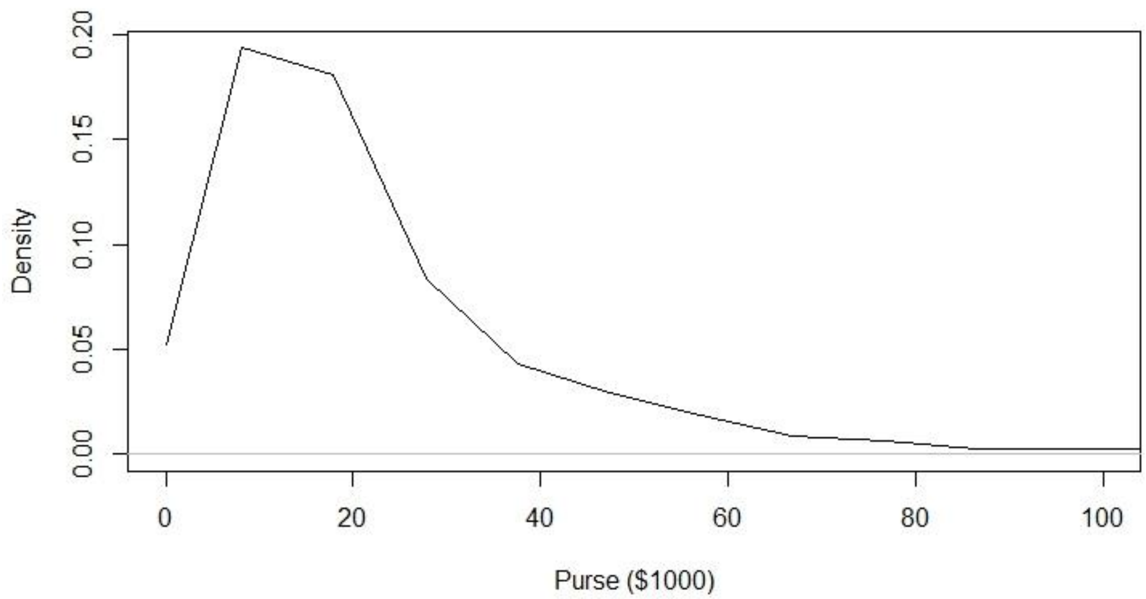

Figure 2-84 Density plot of purse ($1000)

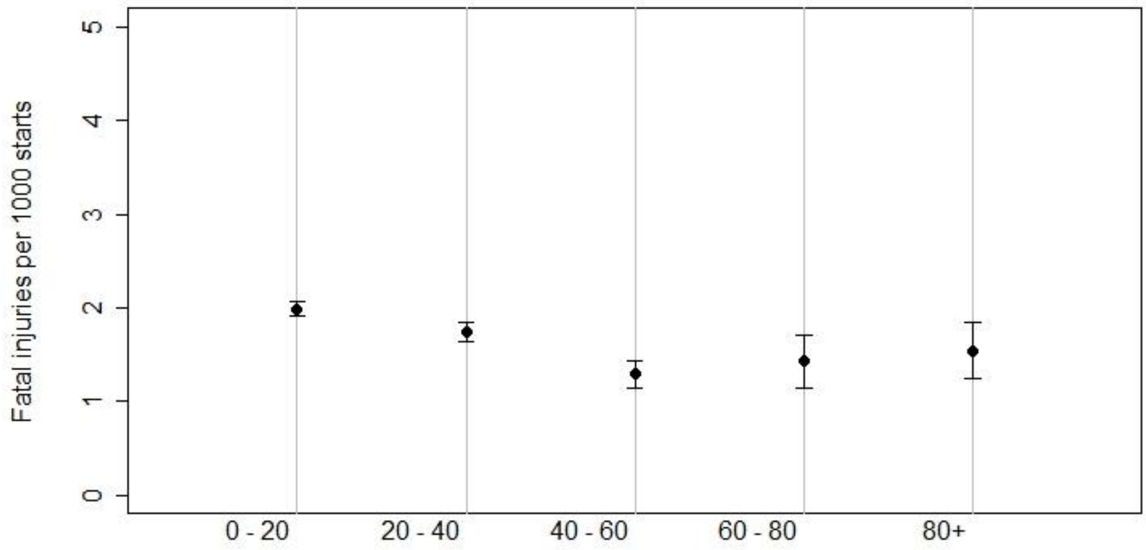

Figure 2-85 Fatal injuries per 1000 racing starts with 95% confidence interval by purse ($1000) group

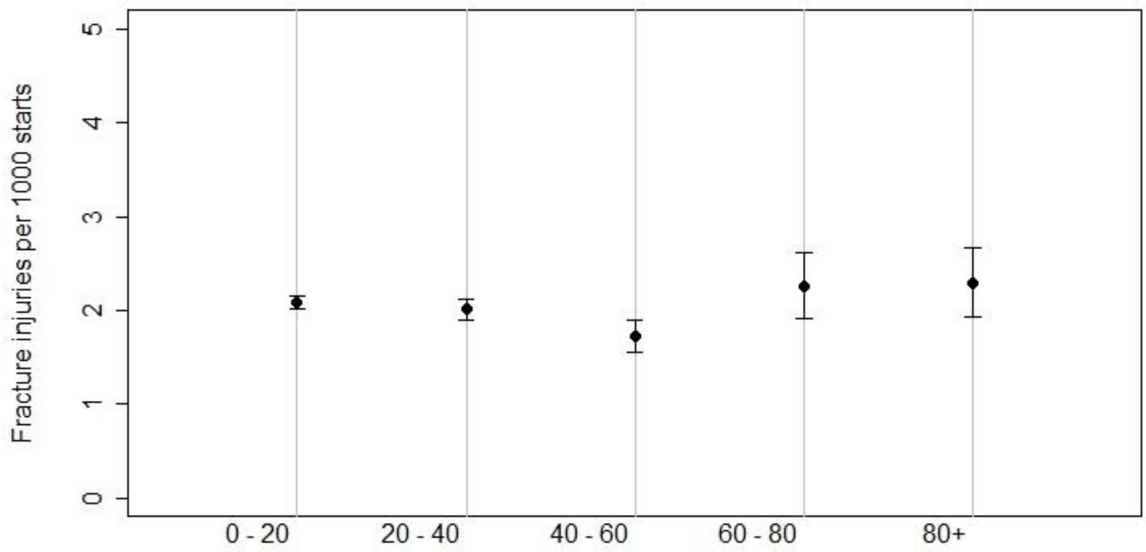

Figure 2-86 Fracture injuries per 1000 racing starts with 95% confidence interval by purse ($1000) group

2.4.31. Race distance (furlongs)

The EID contained information on each race distance. The density of starts by furlong of race is shown at Figure 2-87 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-88 and 2-89 respectively.

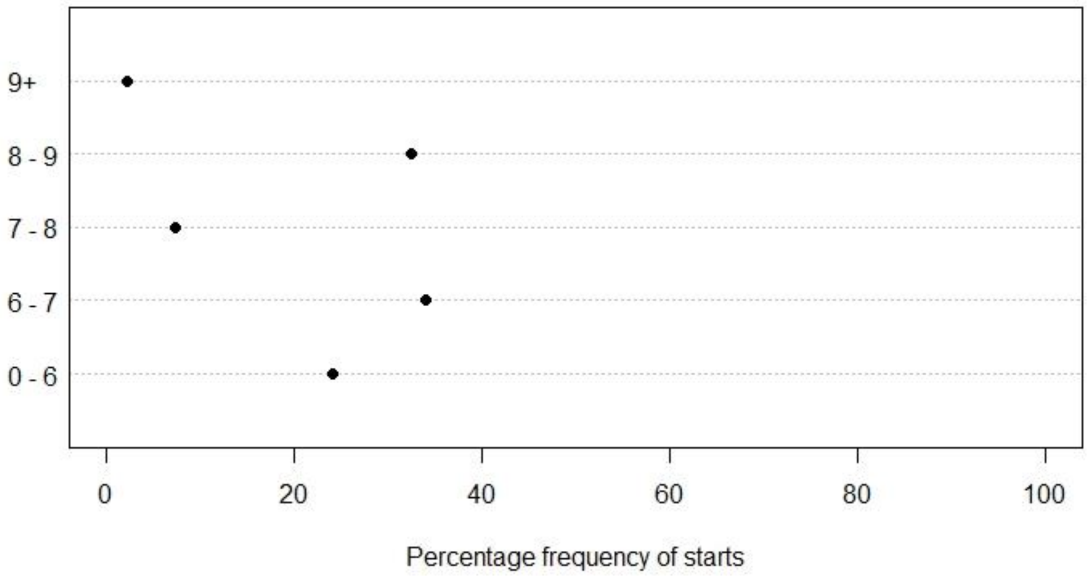

Figure 2-87 Dot plot of race distance (furlongs)

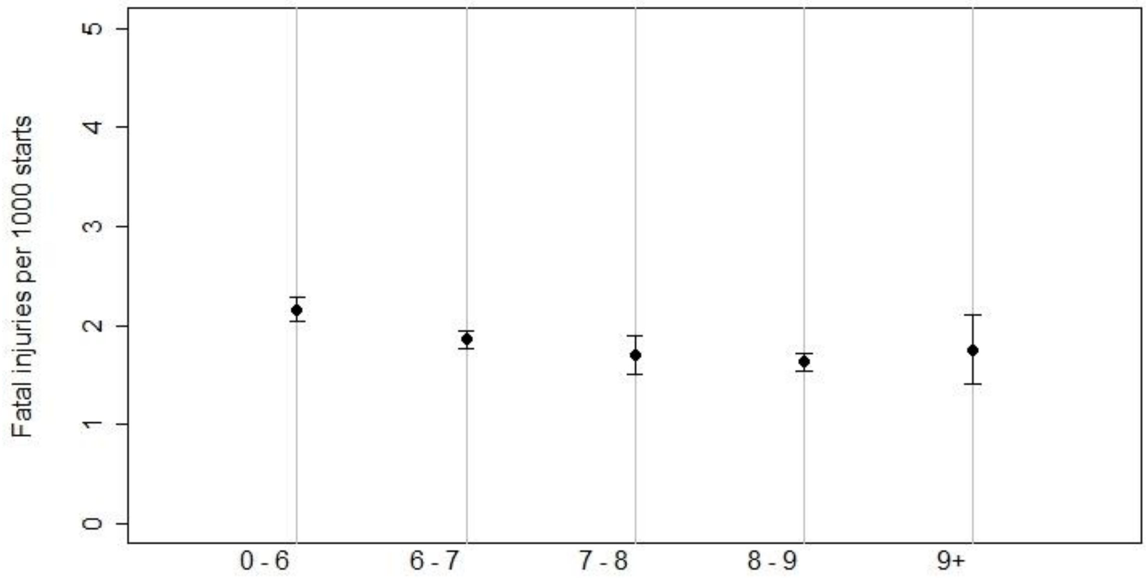

Figure 2-88 Fatal injuries per 1000 racing starts with 95% confidence interval by race distance (furlongs) group

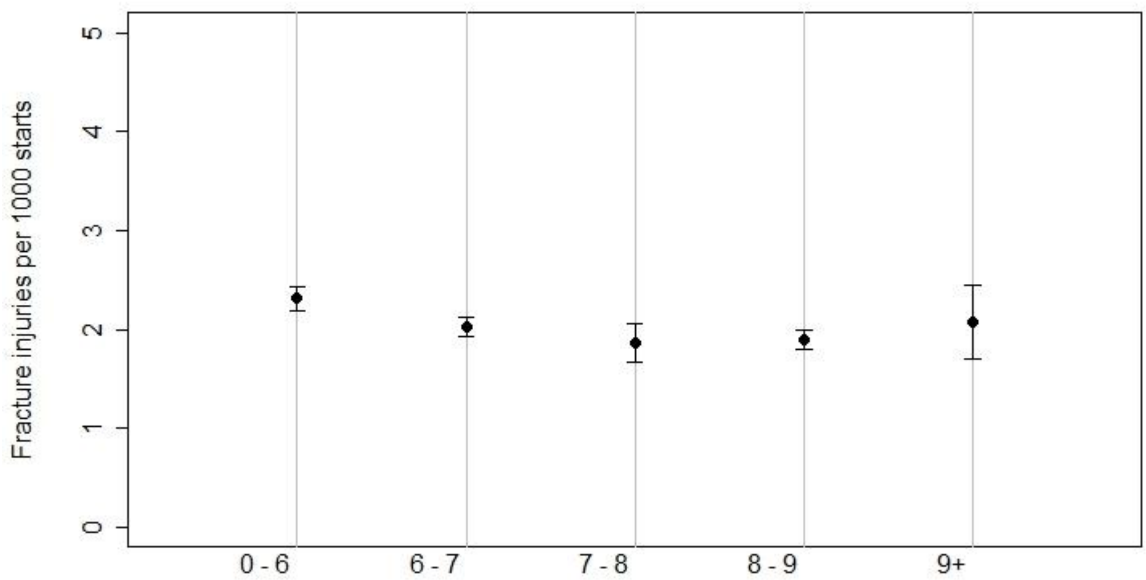

Figure 2-89 Fracture injuries per 1000 racing starts with 95% confidence interval by race distance (furlongs) group

2.4.32. Season

The EID contained information on the season in which each race took place. The proportion of starts by season is shown at Figure 2-90 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different seasons are shown at figures 2-91 and 2-92 respectively.

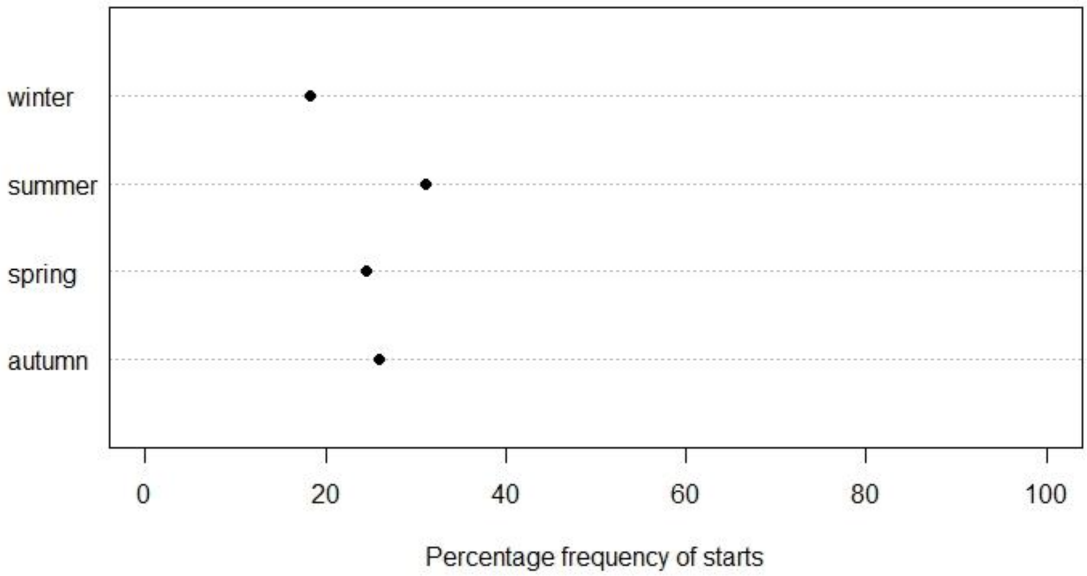

Figure 2-90 Dot plot of season

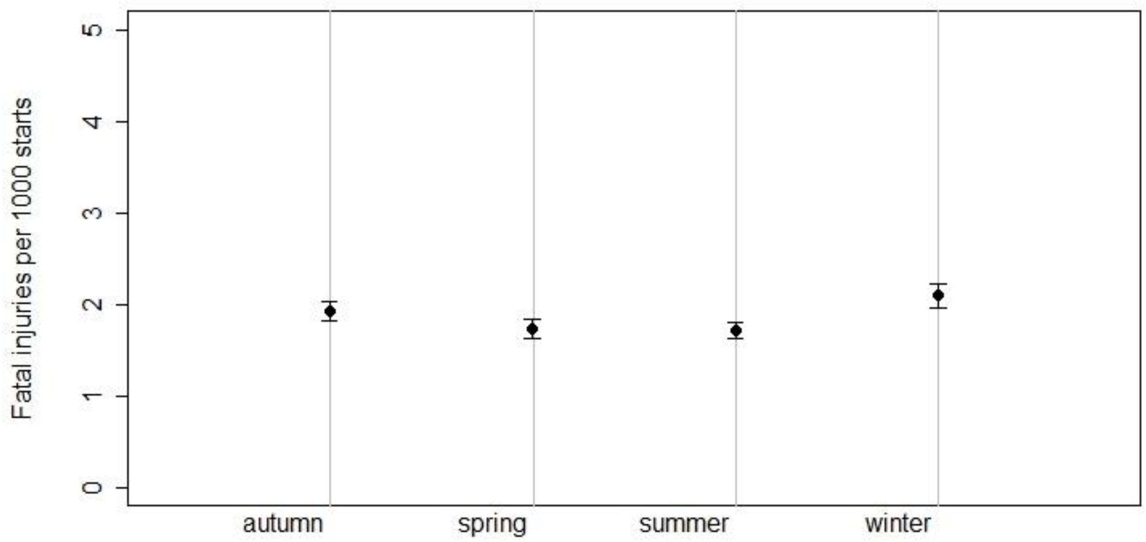

Figure 2-91 Fatal injuries per 1000 racing starts with 95% confidence interval by season

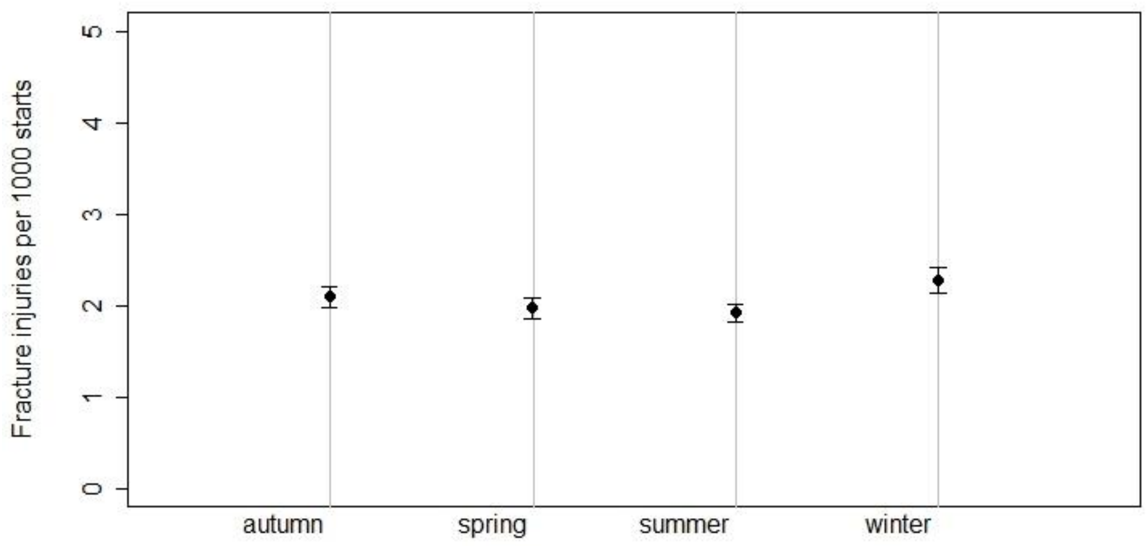

Figure 2-92 Fracture injuries per 1000 racing starts with 95% confidence interval by season

2.4.33. Sex

The EID contained information on the sex of each horse. The proportion of racing starts by stallions is 12% and the number of fatal and fracture injuries, along with their 95% confidence intervals, for each sex group are shown at figures 2-93 and 2-94 respectively.

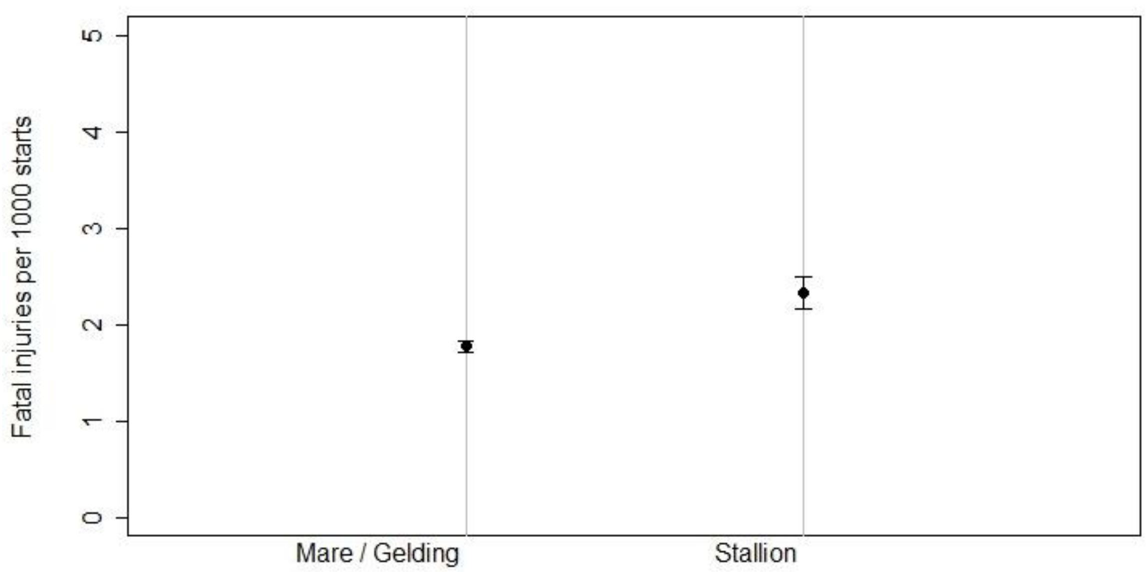

Figure 2-93 Fatal injuries per 1000 racing starts with 95% confidence interval by sex

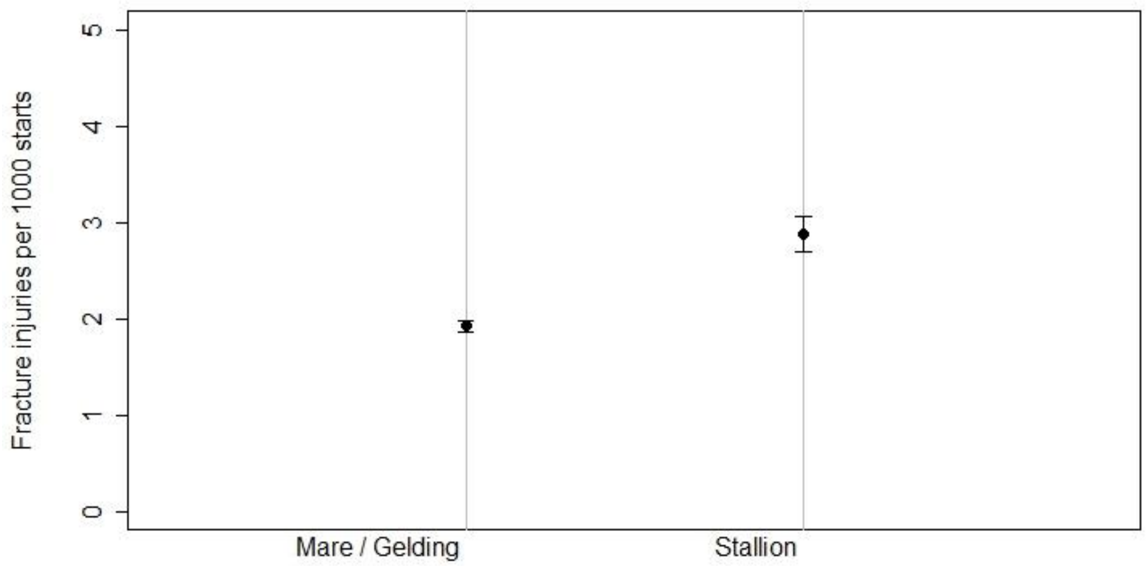

Figure 2-94 Fracture injuries per 1000 racing starts with 95% confidence interval by sex

2.4.34. Start with new jockey

The EID contained information on the jockey of each horse and we looked at the starts for each horse if they changed jockey from the previous race. The proportion of racing starts for horses that changed jockey is 52% and the number of fatal and fracture injuries, along with their 95% confidence intervals, for each group are shown at figures 2-95 and 2-96 respectively.

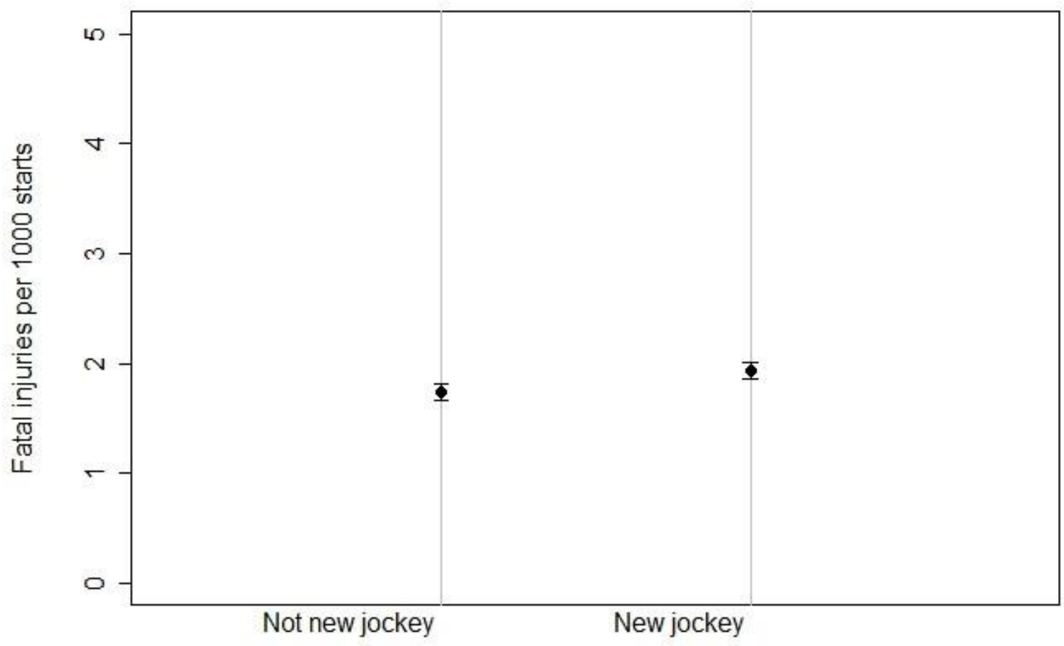

Figure 2-95 Fatal injuries per 1000 racing starts with 95% confidence interval by age starts with new jockey

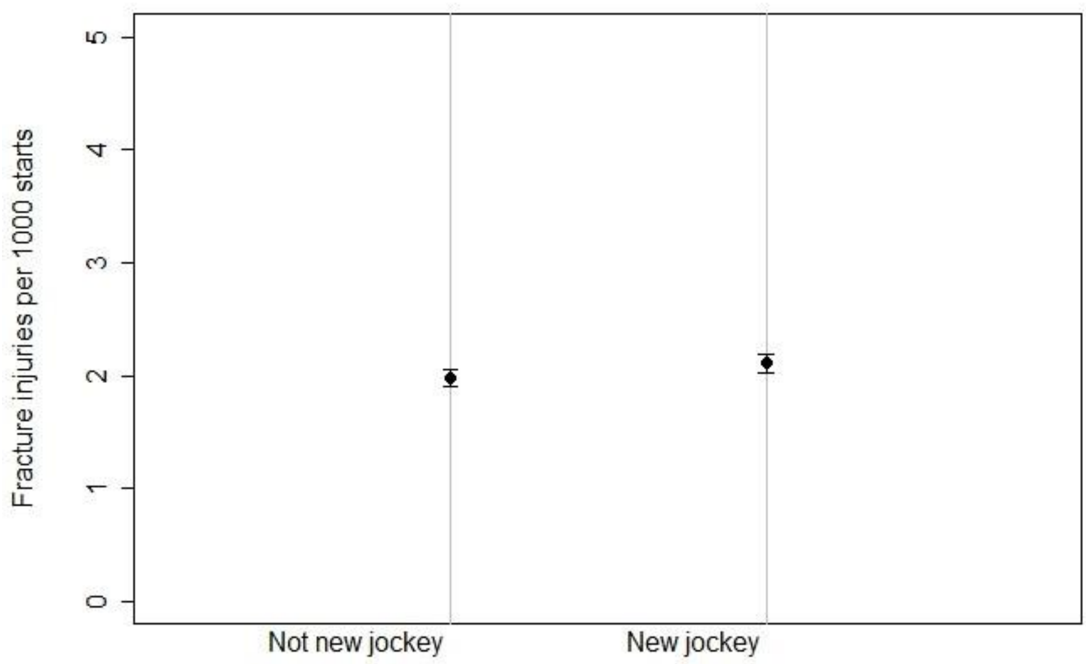

Figure 2-96 Fracture injuries per 1000 racing starts with 95% confidence interval by age starts with new jockey

2.4.35. Start with new trainer

The EID contained information on the trainer of each horse and we looked at the starts for each horse if they changed trainer from the previous race. The proportion of racing starts for horses that changed trainer is 9% and the number of fatal and fracture injuries, along with their 95% confidence intervals, for each group are shown at figures 2-97 and 2-98 respectively.

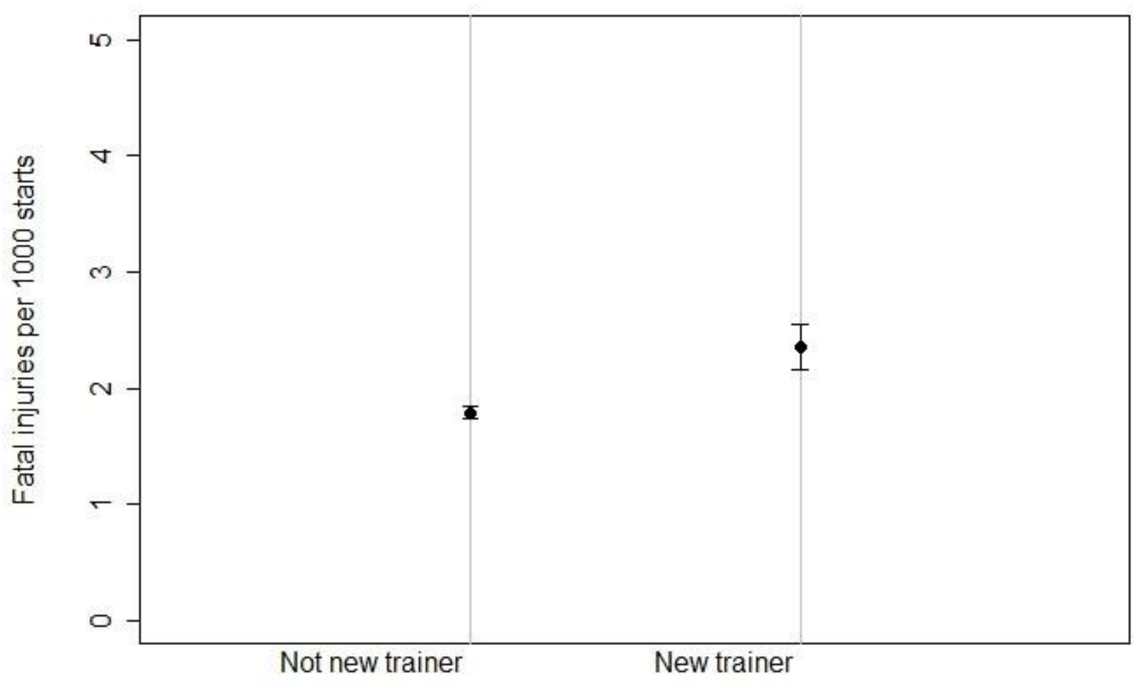

Figure 2-97 Fatal injuries per 1000 racing starts with 95% confidence interval by starts with new trainer

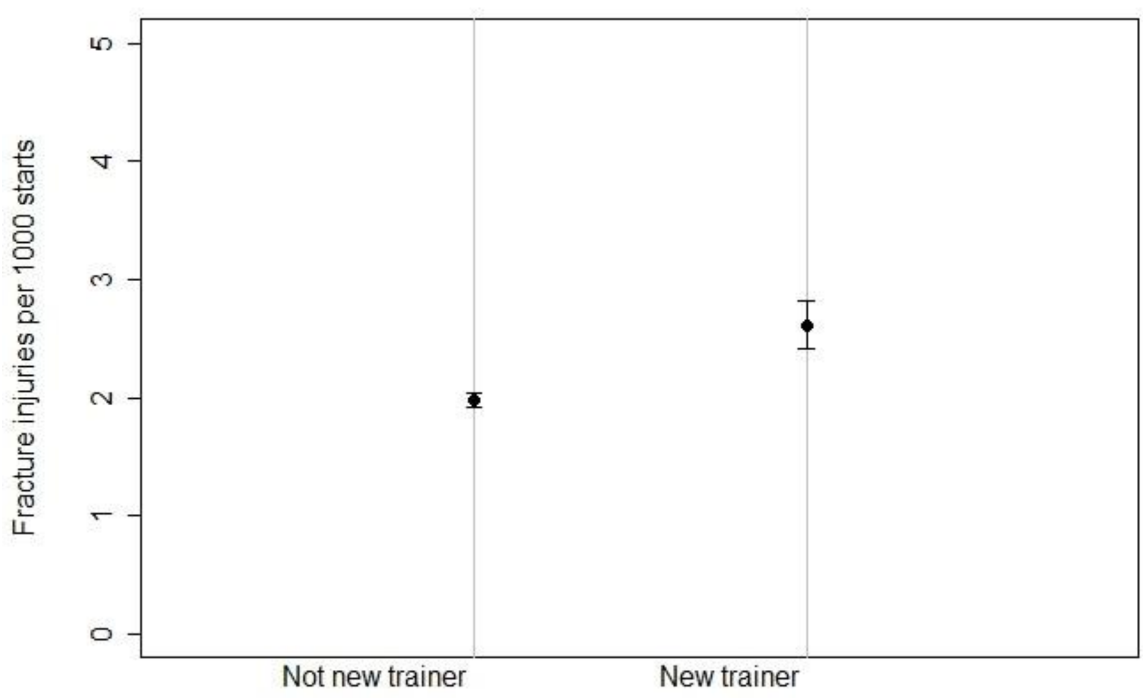

Figure 2-98 Fracture injuries per 1000 racing starts with 95% confidence interval by starts with new trainer

2.4.36. Surface

The EID contained information on the surface each race. The proportion of starts by surface type is shown at Figure 2-99 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different surface types are shown at figures 2-100 and 2-101 respectively.

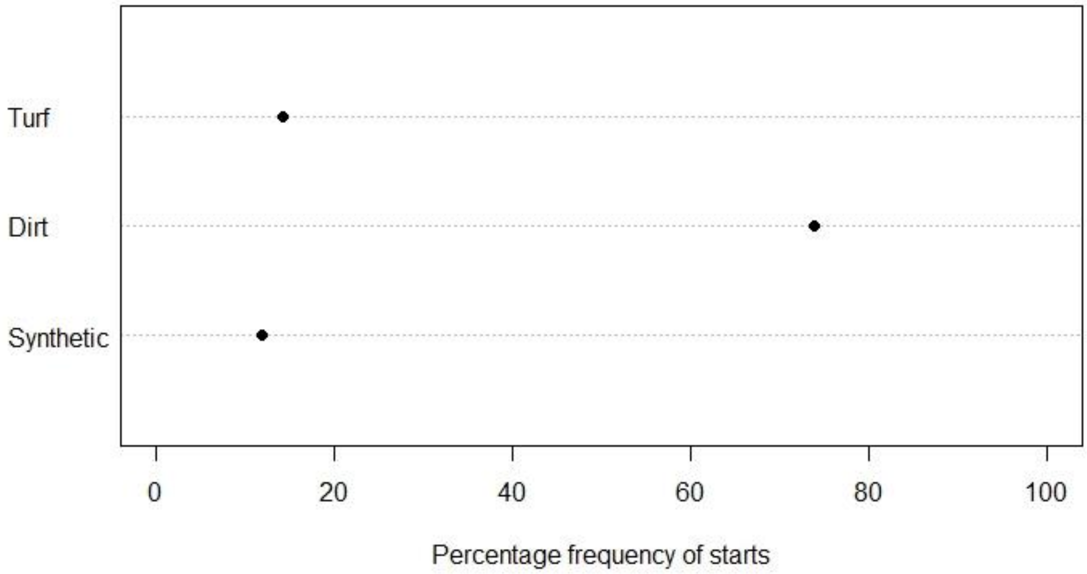

Figure 2-99 Dot plot of surface types

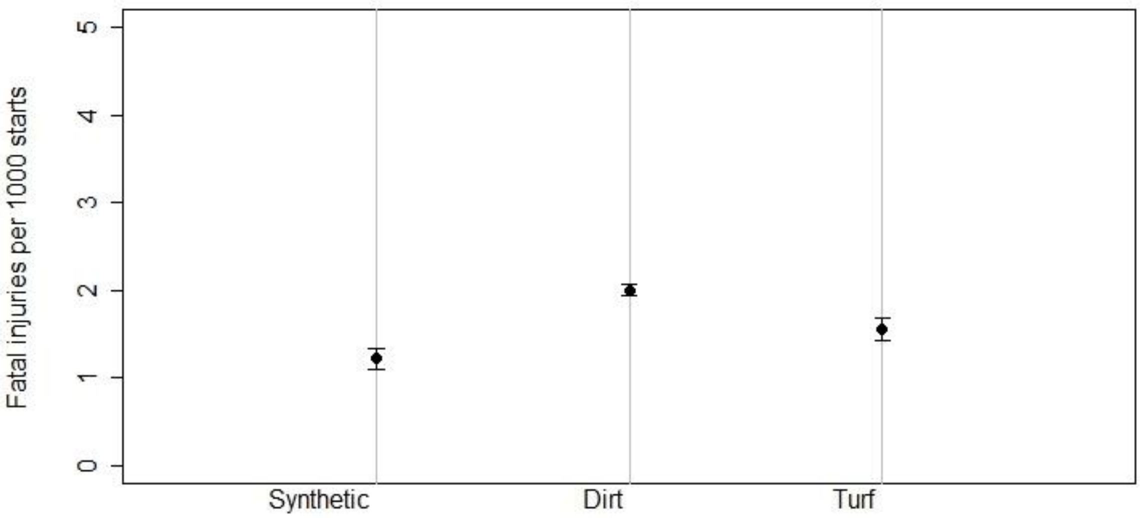

Figure 2-100 Fatal injuries per 1000 racing starts with 95% confidence interval by age surface type

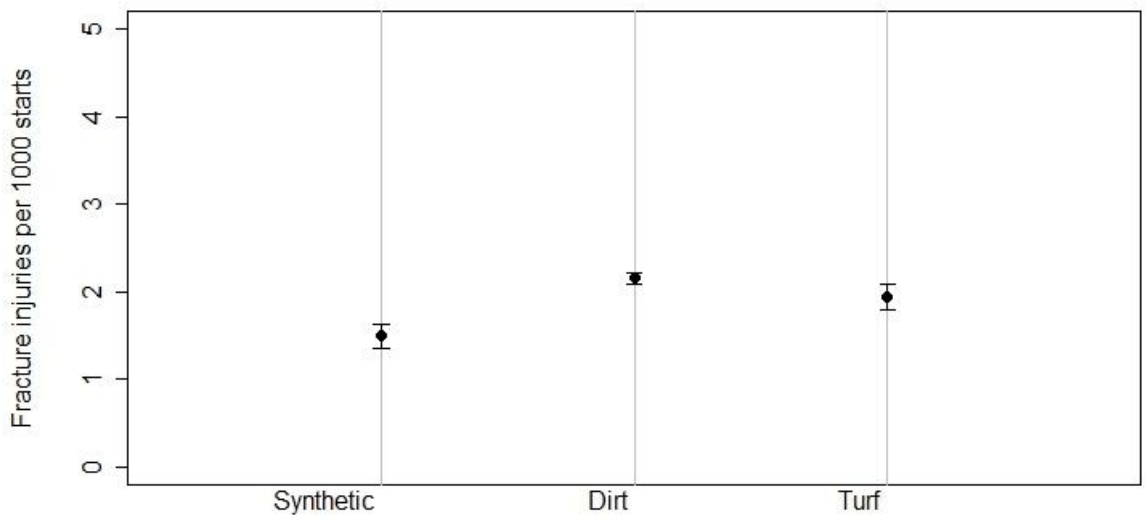

Figure 2-101 Fracture injuries per 1000 racing starts with 95% confidence interval by age surface type

2.4.37. Time between exercise starts – average (months)

From the exercise starts available for each horse we calculated the average time between exercise starts for each Thoroughbred. The density of starts per month is shown at Figure 2-102 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-103 and 2-104 respectively.

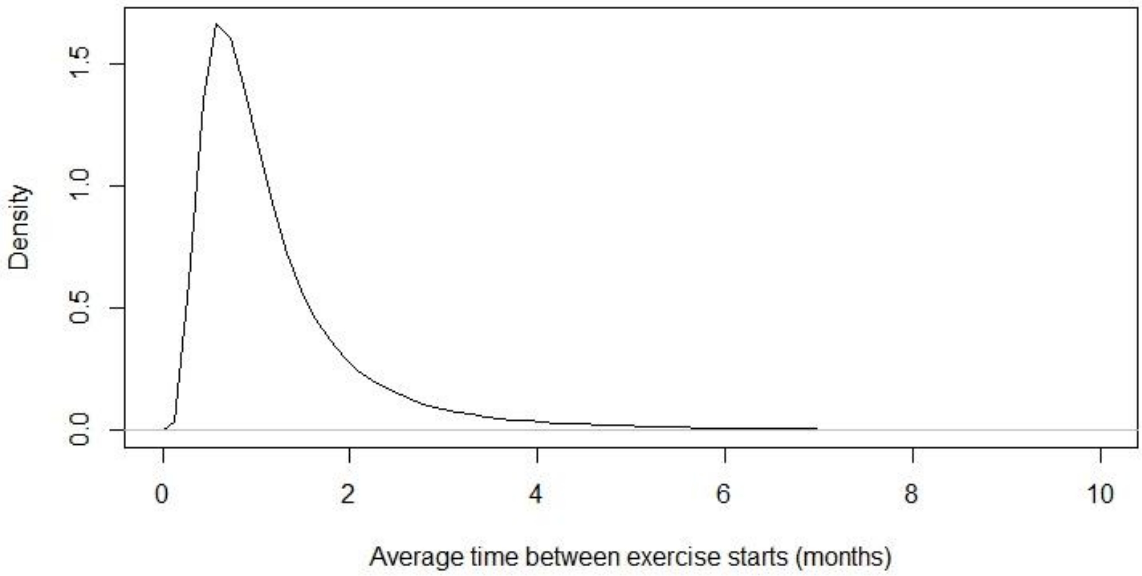

Figure 2-102 Density plot of time between exercise starts – average (months)

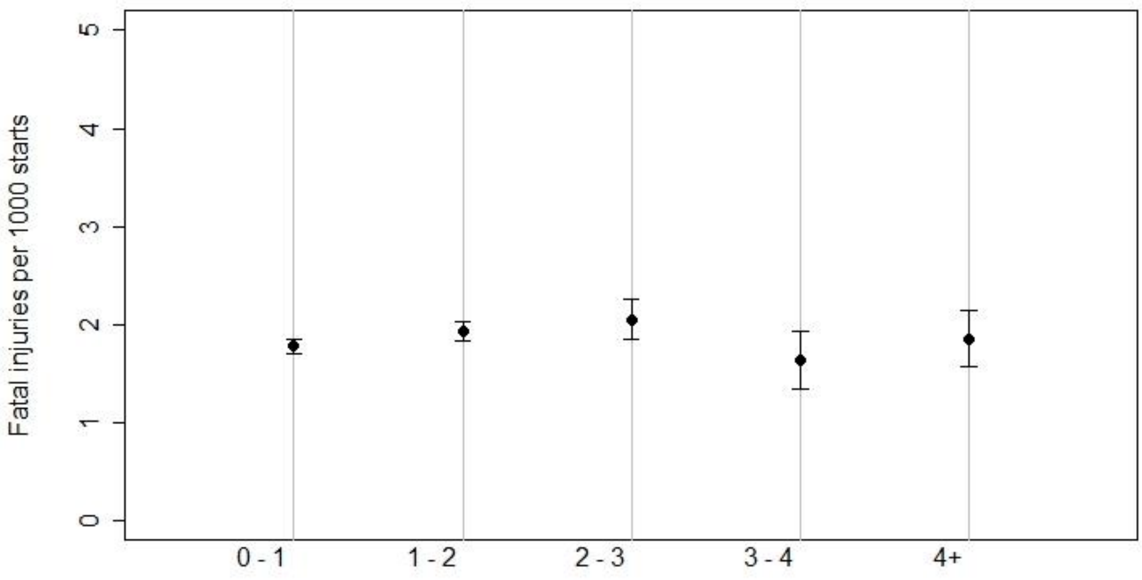

Figure 2-103 Fatal injuries per 1000 racing starts with 95% confidence interval by time between exercise starts – average (months) group

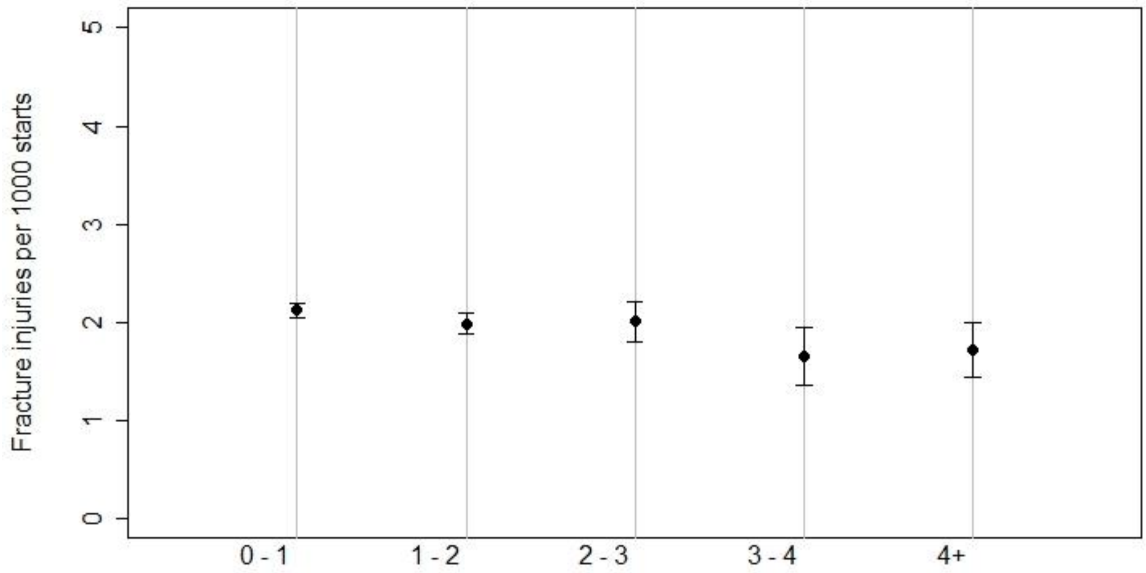

Figure 2-103 Fracture injuries per 1000 racing starts with 95% confidence interval by time between exercise starts – average (months) group

2.4.38. Time between exercise starts - active – average (months)

From the racing and exercise starts available for each horse we calculated the average time between exercise starts for each Thoroughbred excluding our calculations with respect to the time spent on layup for each horse. The density of starts per month is shown at Figure 2-105 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-106 and 2-107 respectively.

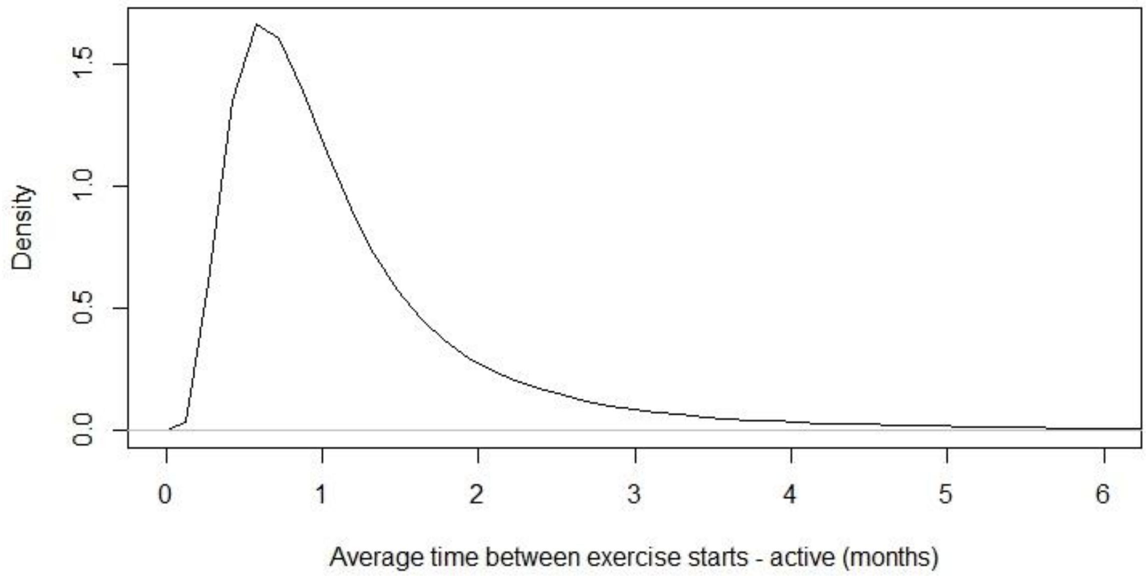

Figure 2-104 Density plot of time between exercise starts - active – average (months)

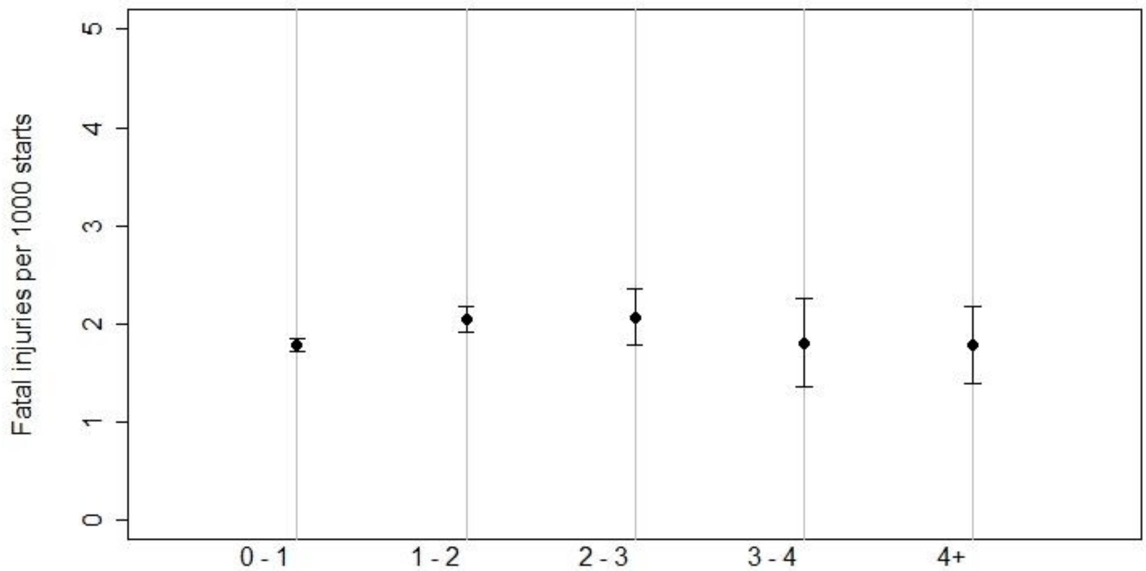

Figure 2-105 Fatal injuries per 1000 racing starts with 95% confidence interval by time between exercise starts - active – average (months) group

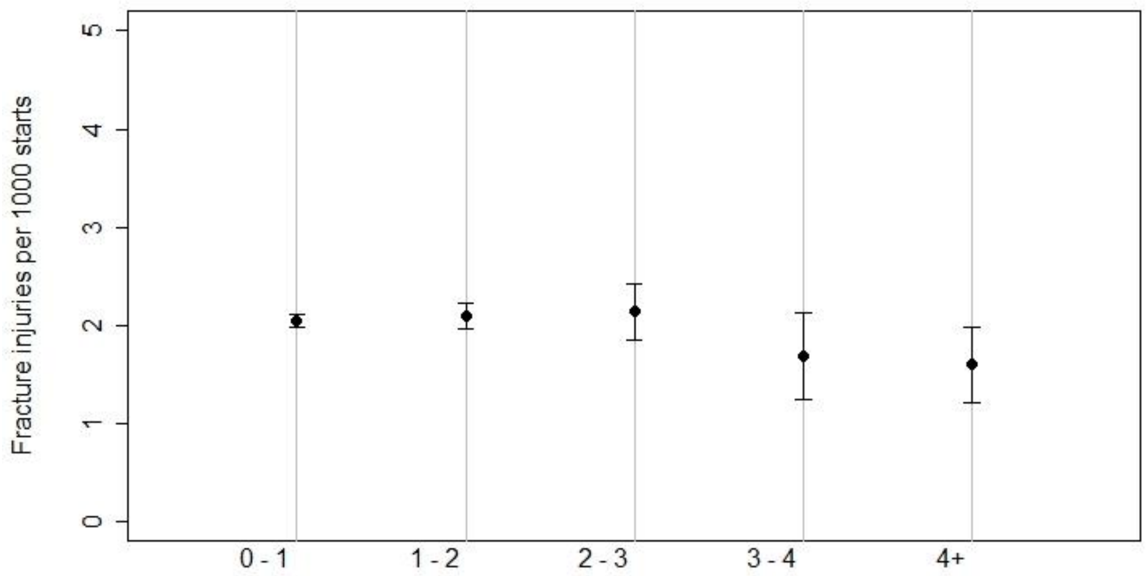

Figure 2-106 Fracture injuries per 1000 racing starts with 95% confidence interval by time between exercise starts - active – average (months) group

2.4.39. Time between racing starts – average (months)

From the racing starts available for each horse we calculated the average time between racing starts for each Thoroughbred. The density of starts by month is shown at Figure 2-108 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-109 and 2-110 respectively.

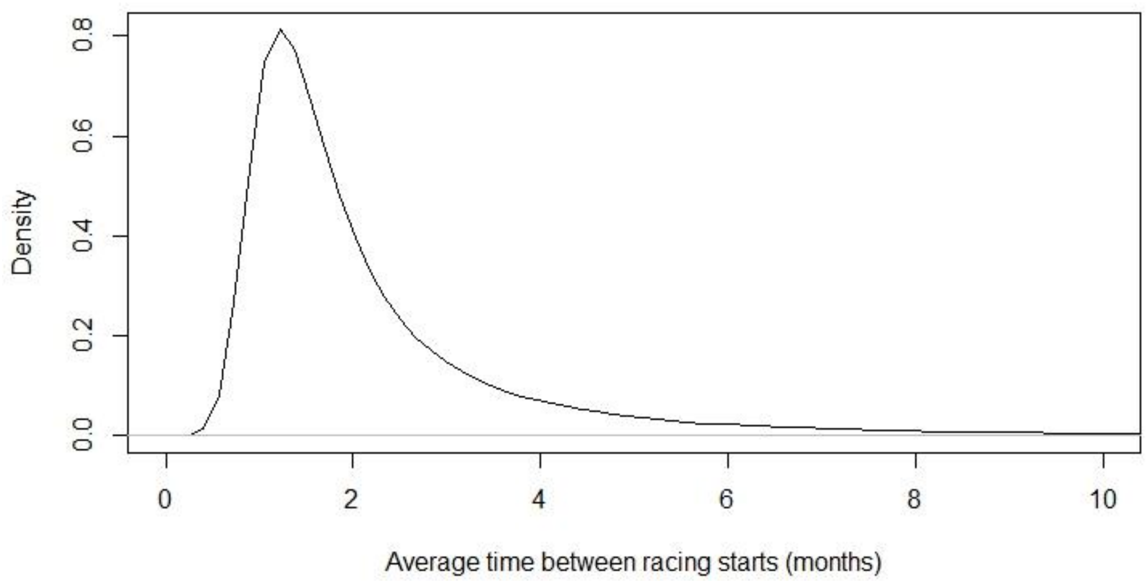

Figure 2-107 Density plot of time between racing starts – average (months)

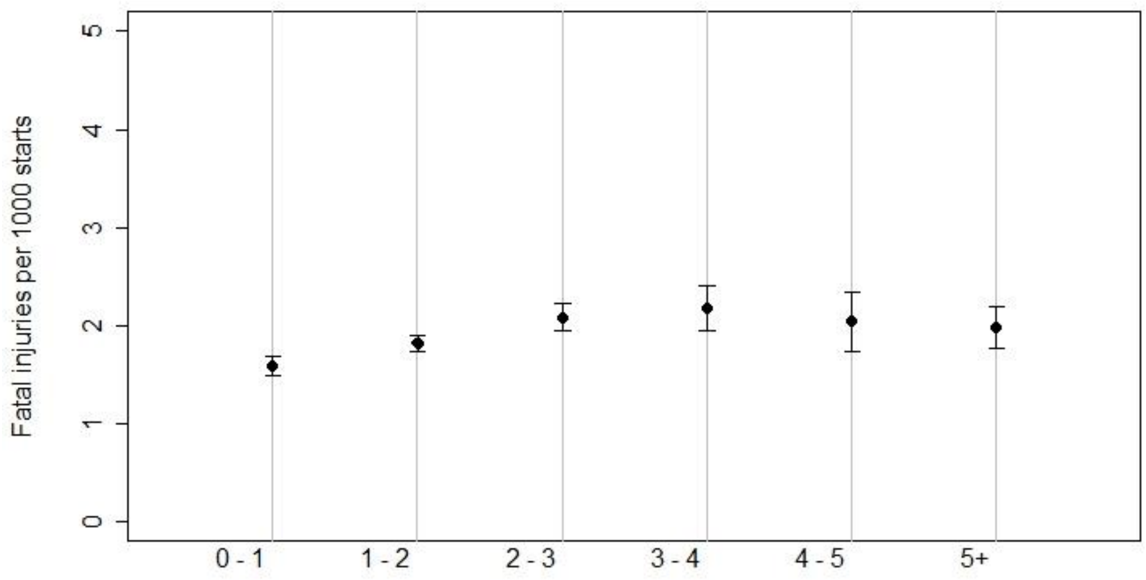

Figure 2-108 Fatal injuries per 1000 racing starts with 95% confidence interval by time between racing starts – average (months) group

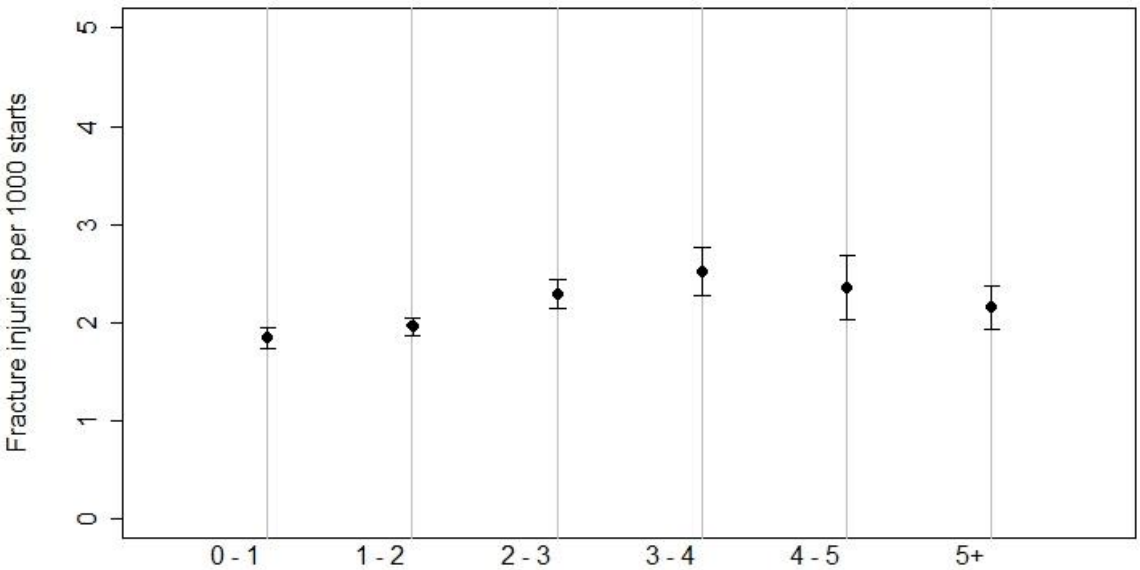

Figure 2-109 Fracture injuries per 1000 racing starts with 95% confidence interval by time between racing starts – average (months) group

2.4.40. Time between racing starts – active - average (months)

From the racing and racing starts available for each horse we calculated the average time between racing starts for each Thoroughbred excluding our calculations with respect to the time spent on layup for each horse. The density of starts by month is shown at Figure 2-111 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-112 and 2-113 respectively.

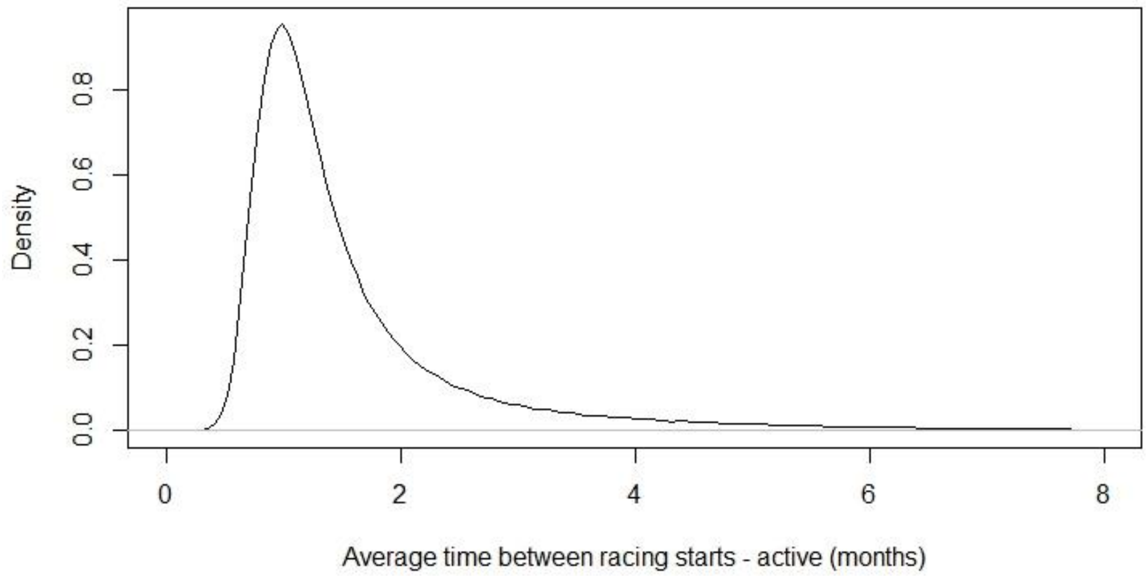

Figure 2-110 Density plot of time between racing starts – active - average (months)

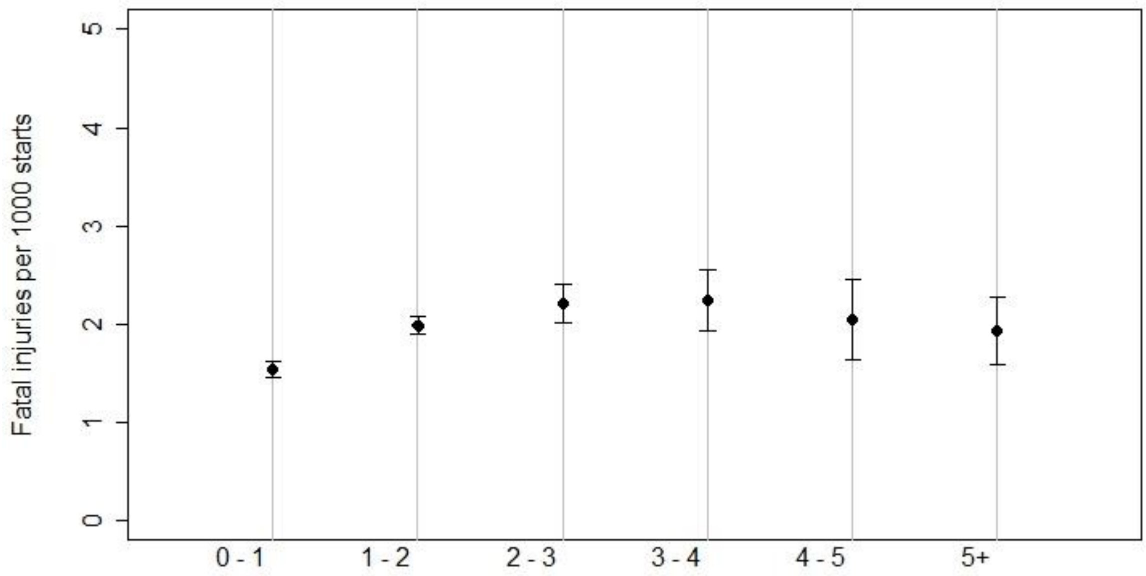

Figure 2-112 Fatal injuries per 1000 racing starts with 95% confidence interval by time between racing starts – active - average (months) group

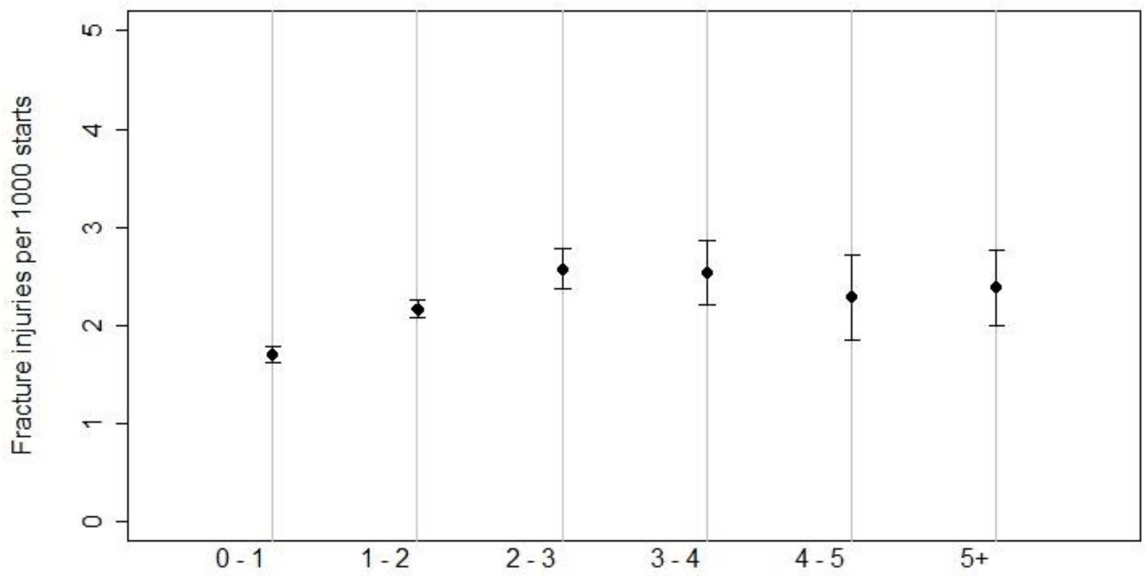

Figure 2-111 Fracture injuries per 1000 racing starts with 95% confidence interval by time between racing starts – active - average (months) group

2.4.41. Time in layup (months)

From the information contained in the EID we calculated the time a horse has spent in layup throughout its career. The density of starts by month in layup is shown at Figure 2-114 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-115 and 2-116 respectively. Most starts from horses with layups in their career are from horses with one short layup and that is the reason for the observed spike in the density plot.

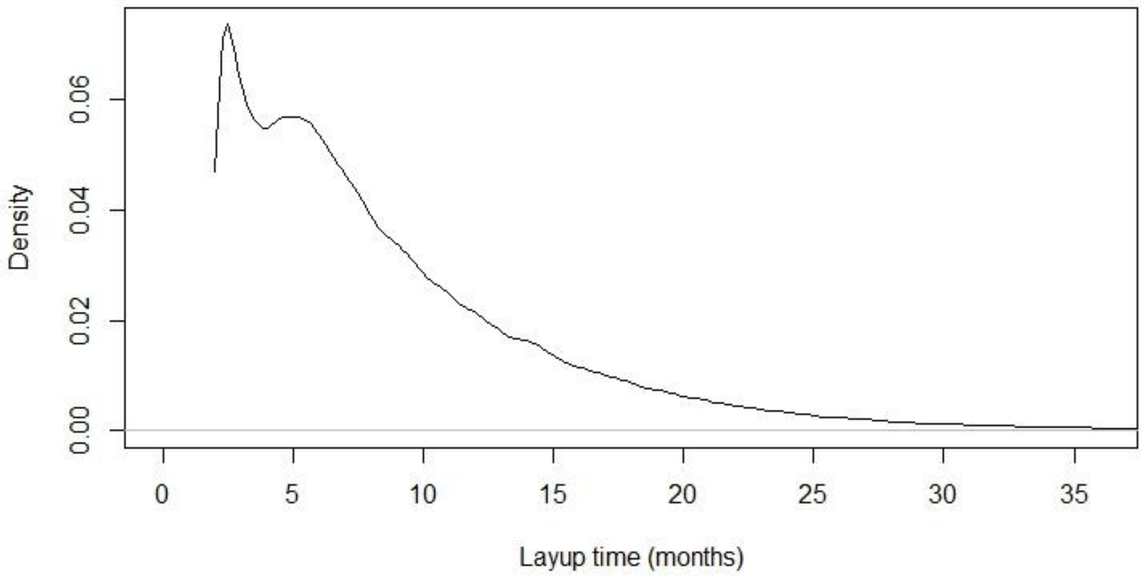

Figure 2-112 Density plot of time in layup (months)

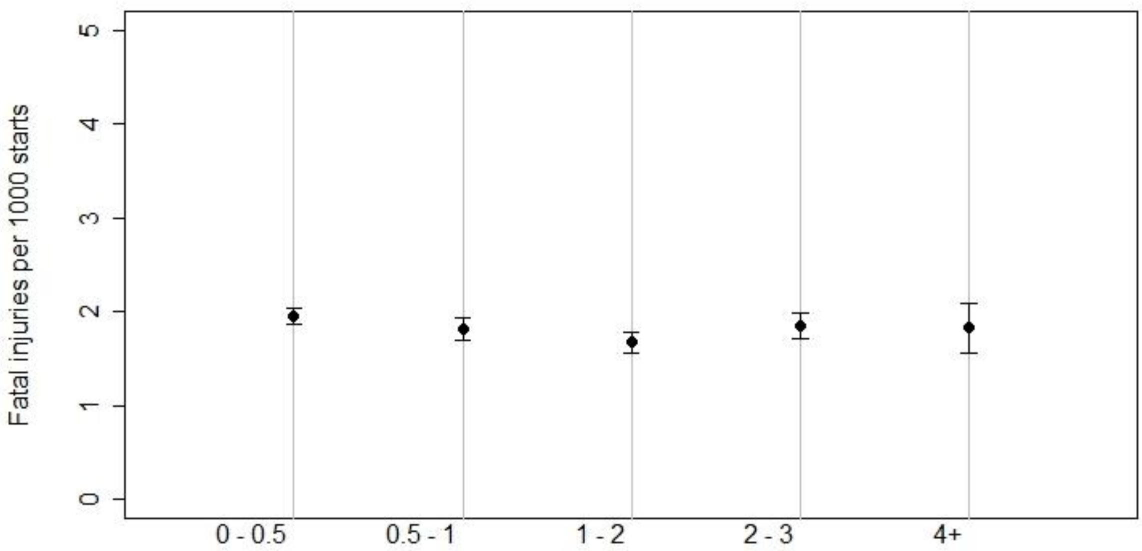

Figure 2-113 Fatal injuries per 1000 racing starts with 95% confidence interval by time in layup (months) group

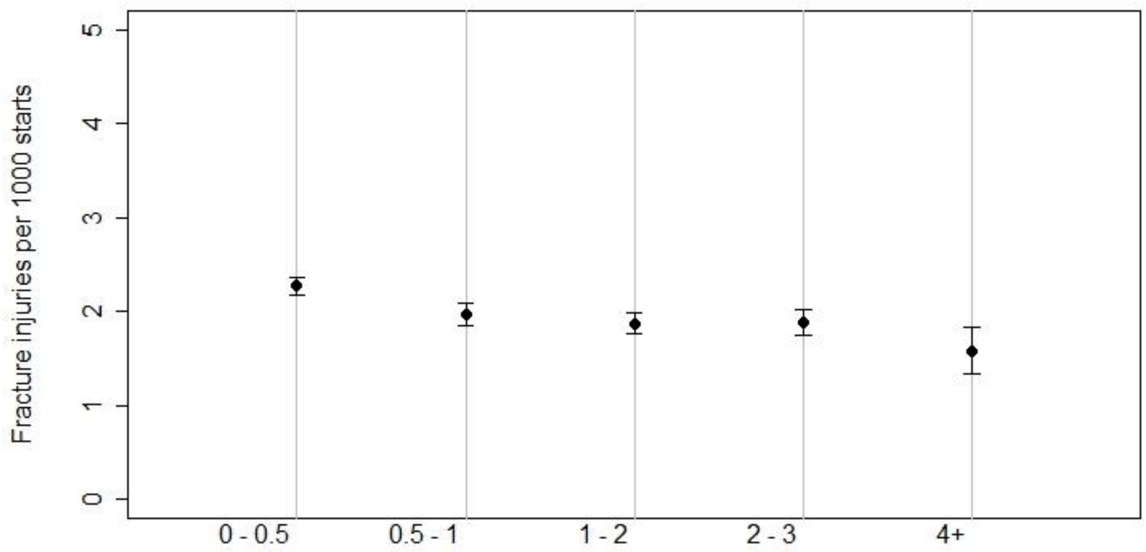

Figure 2-114 Fracture injuries per 1000 racing starts with 95% confidence interval by time in layup (months) group

2.4.42. Time in racing – active (months)

From the racing and exercise starts available for each horse we calculated the time in racing for each Thoroughbred excluding the time it has spent on layup. The density of starts by month is shown at Figure 2-117 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-118 and 2-119 respectively.

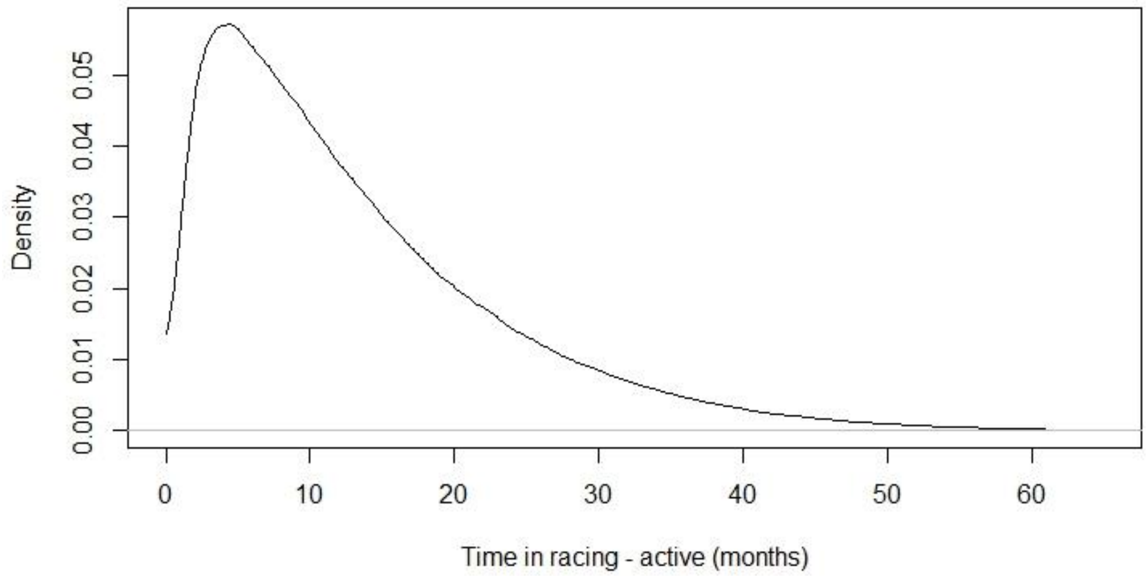

Figure 2-115 Density plot of time in racing – active (months)

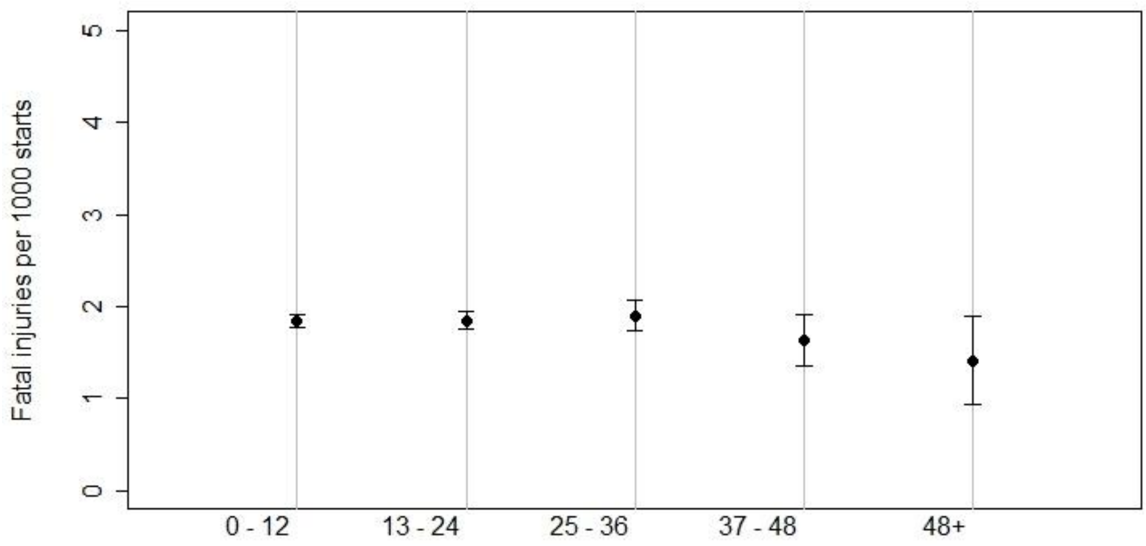

Figure 2-116 Fatal injuries per 1000 racing starts with 95% confidence interval by time in racing – active (months) group

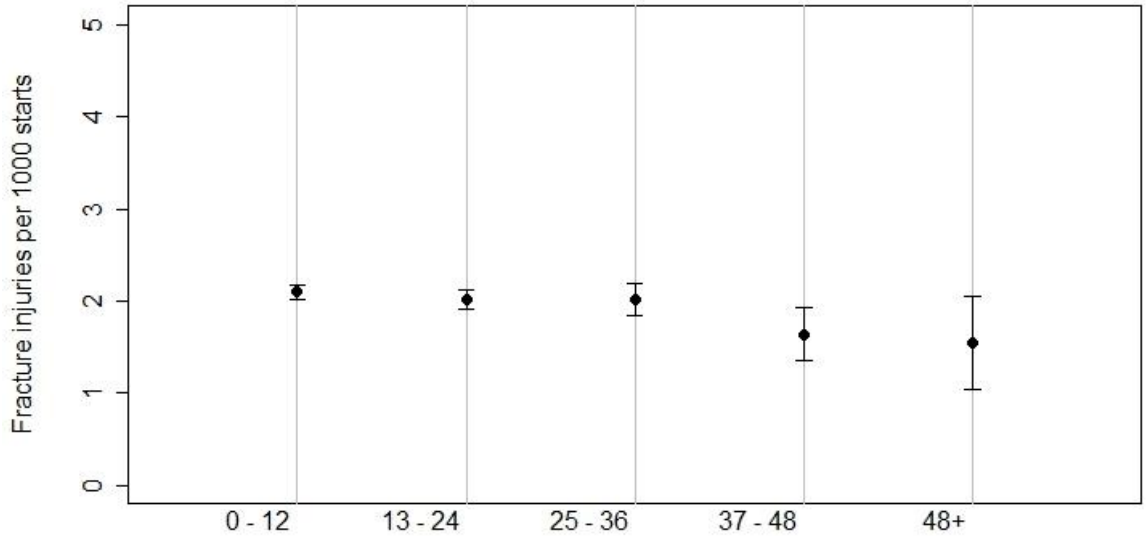

Figure 2-117 Fracture injuries per 1000 racing starts with 95% confidence interval by time in racing – active (months) group

2.4.43. Time in racing (months)

From the racing and exercise starts available for each horse we calculated the time in racing for each Thoroughbred from its first recorded start. The density of starts by month is shown at Figure 2-120 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-121 and 2-122 respectively. We observe different spikes in the density plot that account for the periods of respite between races and for the many horses that start their career with a few starts at an early age but have the majority of their starts later on.

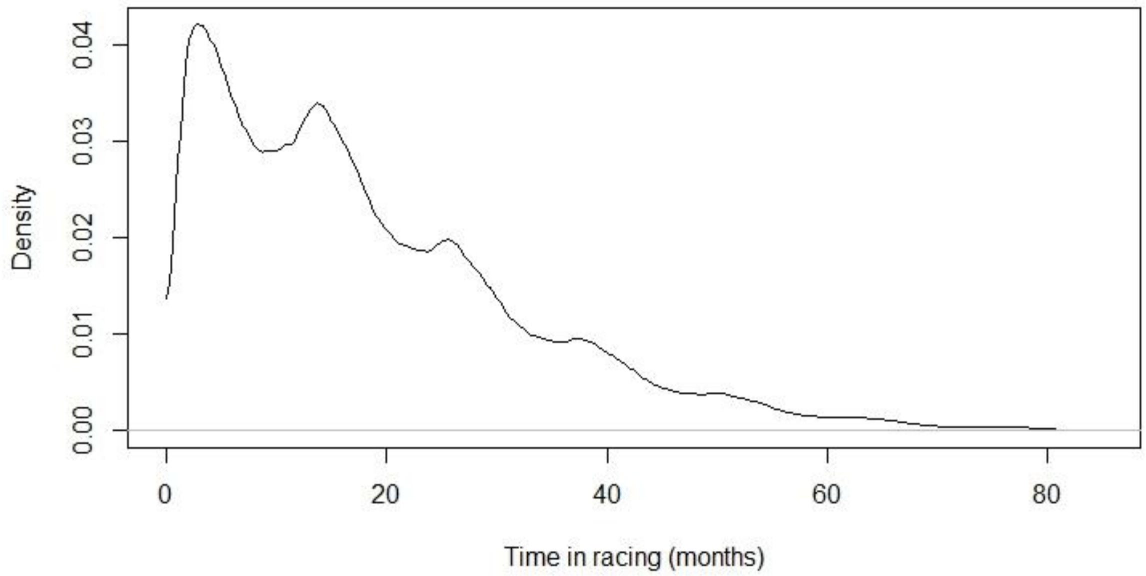

Figure 2-118 Density plot of time in racing (months)

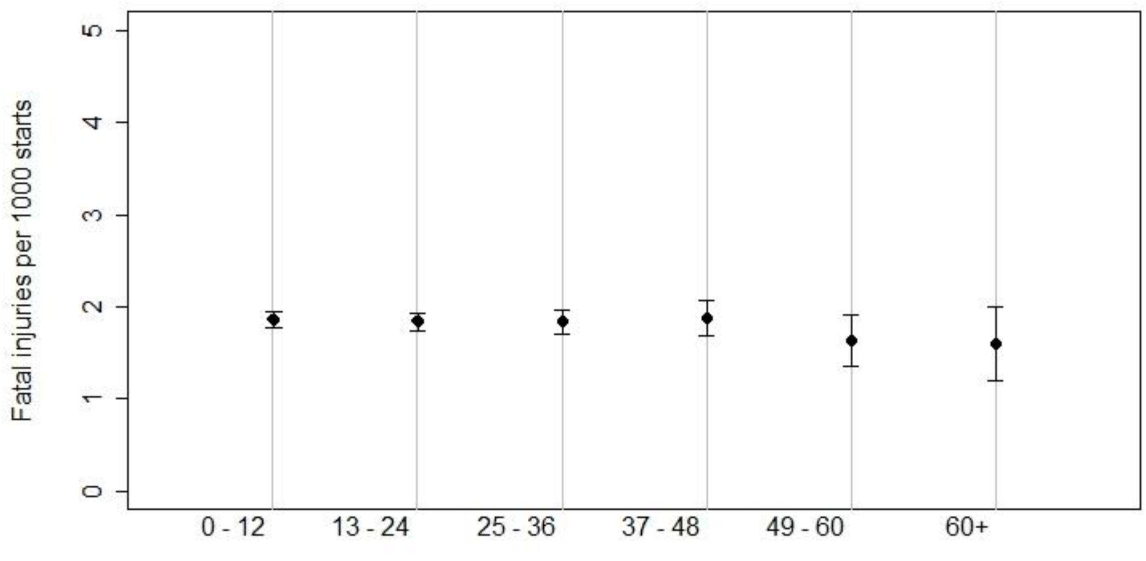

Figure 2-119 Fatal injuries per 1000 racing starts with 95% confidence interval by time in racing (months) group

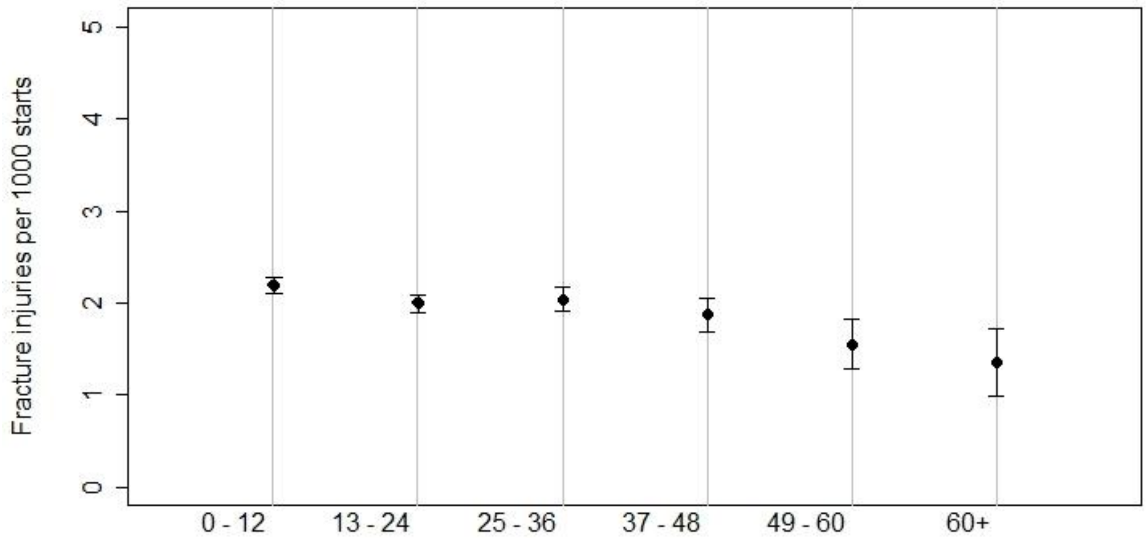

Figure 2-120 Fracture injuries per 1000 racing starts with 95% confidence interval by time in racing (months) group

2.4.44. Time with same jockey (months)

From the racing and exercise starts available for each horse we calculated the time a horse has remained with the same jockey. The density of starts by month is shown at Figure 2-123 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-124 and 2-125 respectively.

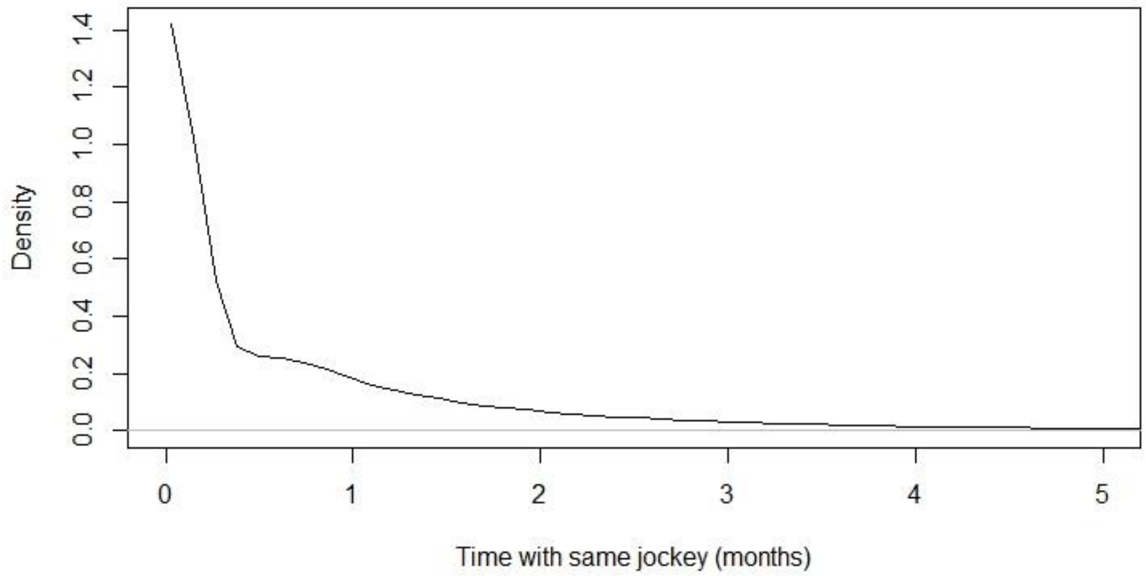

Figure 2-121 Density plot of time with same jockey (months)

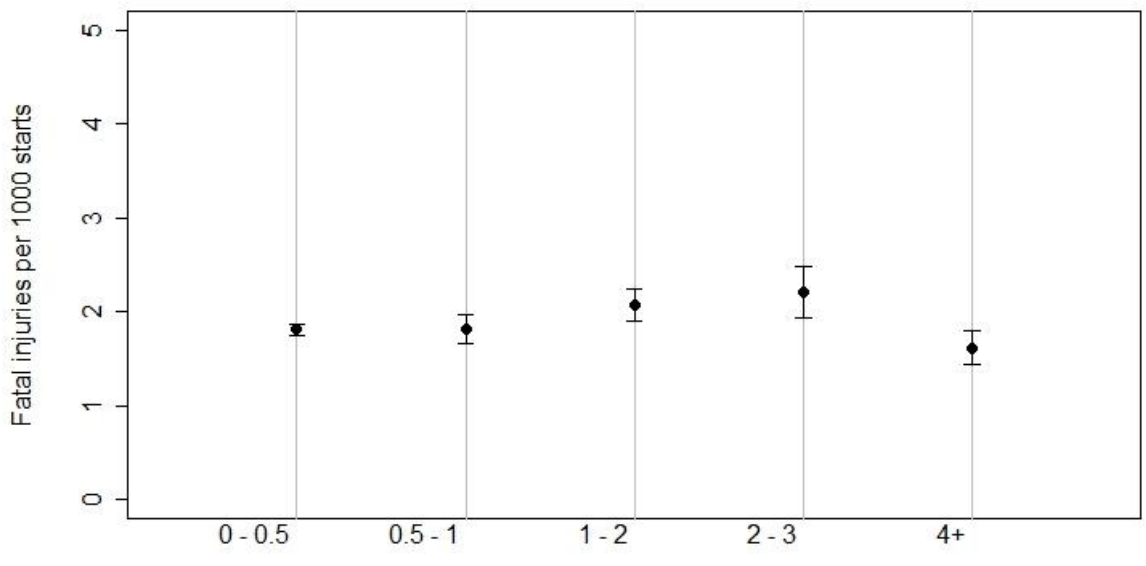

Figure 2-122 Fatal injuries per 1000 racing starts with 95% confidence interval by time with same jockey (months) group

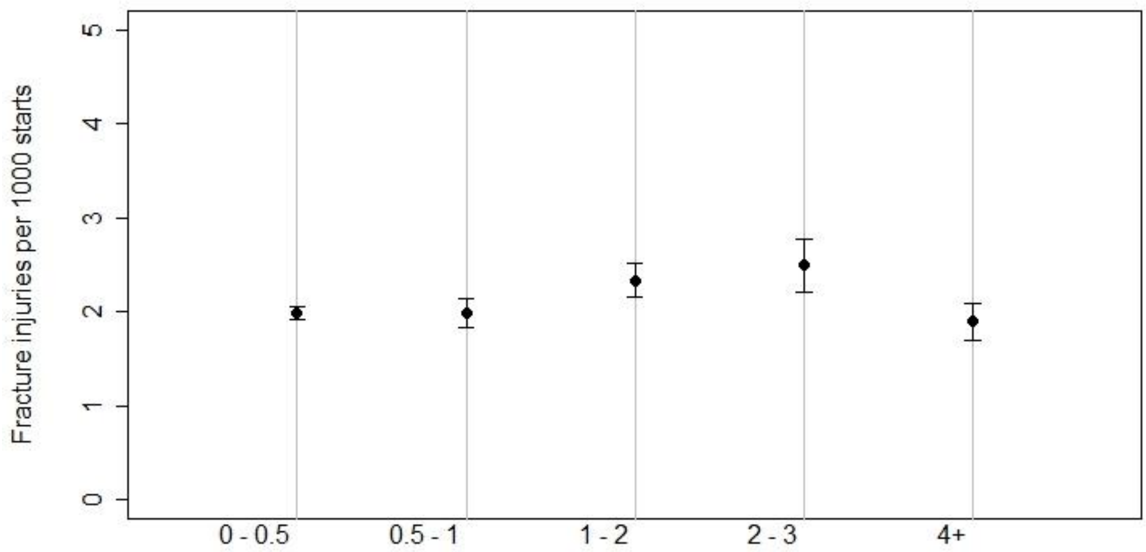

Figure 2-123 Fracture injuries per 1000 racing starts with 95% confidence interval by time with same jockey (months) group

2.4.45. Time with same trainer (months)

From the racing and exercise starts available for each horse we calculated the time a horse has remained with the same trainer. The density of starts by month is shown at Figure 2-126 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-127 and 2-128 respectively.

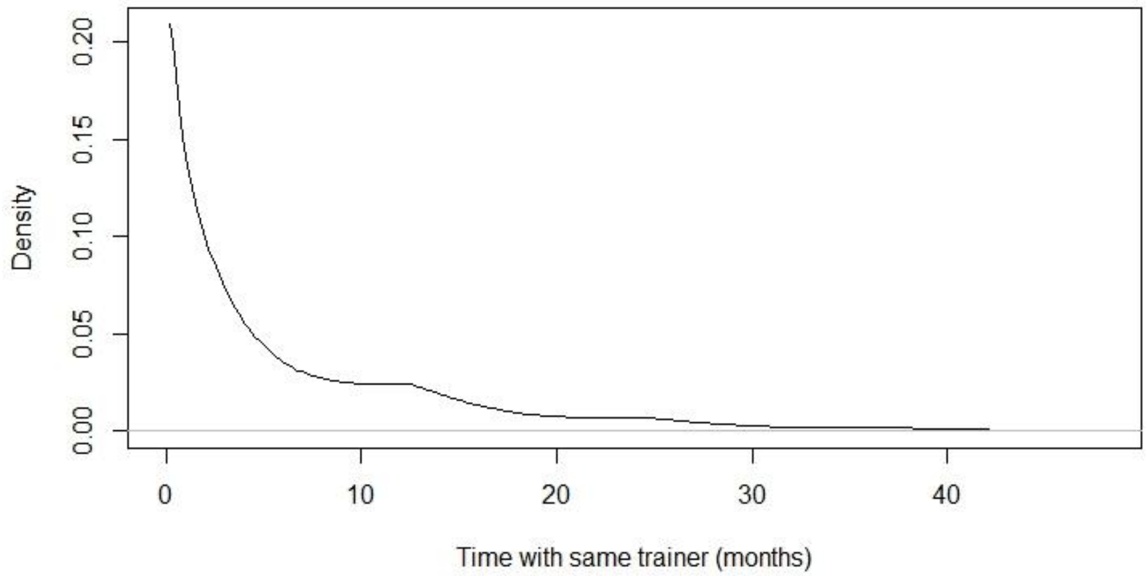

Figure 2-124 Density plot of time with same trainer (months)

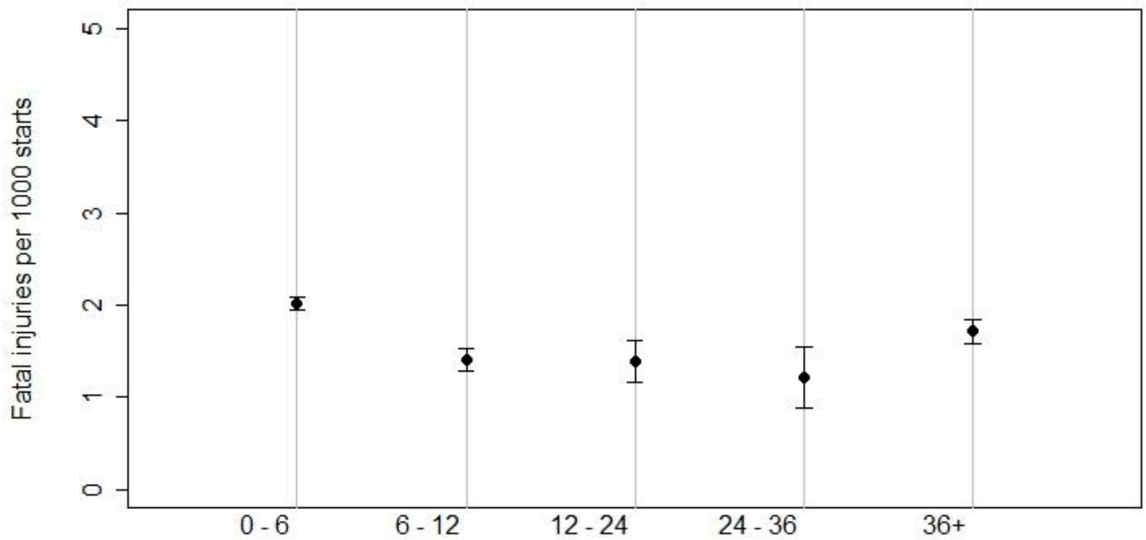

Figure 2-125 Fatal injuries per 1000 racing starts with 95% confidence interval by time with same trainer (months) group

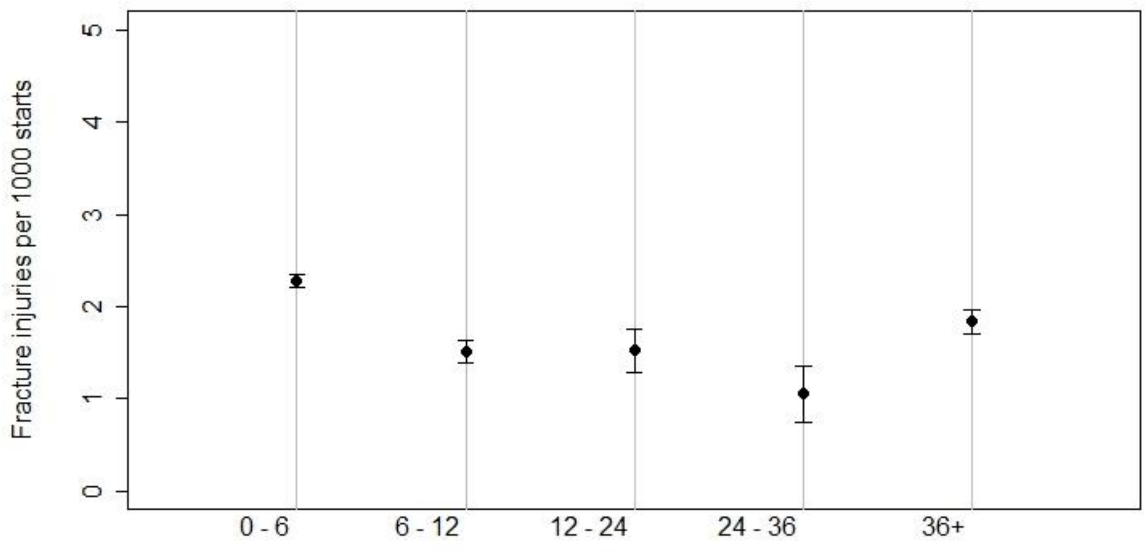

Figure 2-126 Fracture injuries per 1000 racing starts with 95% confidence interval by time with same trainer (months) group

2.4.46. Track size (furlongs)

The EID contained information on the size (i.e. circumference) of each race track. The proportion of starts by furlong is shown at Figure 2-129 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-130 and 2-131 respectively.

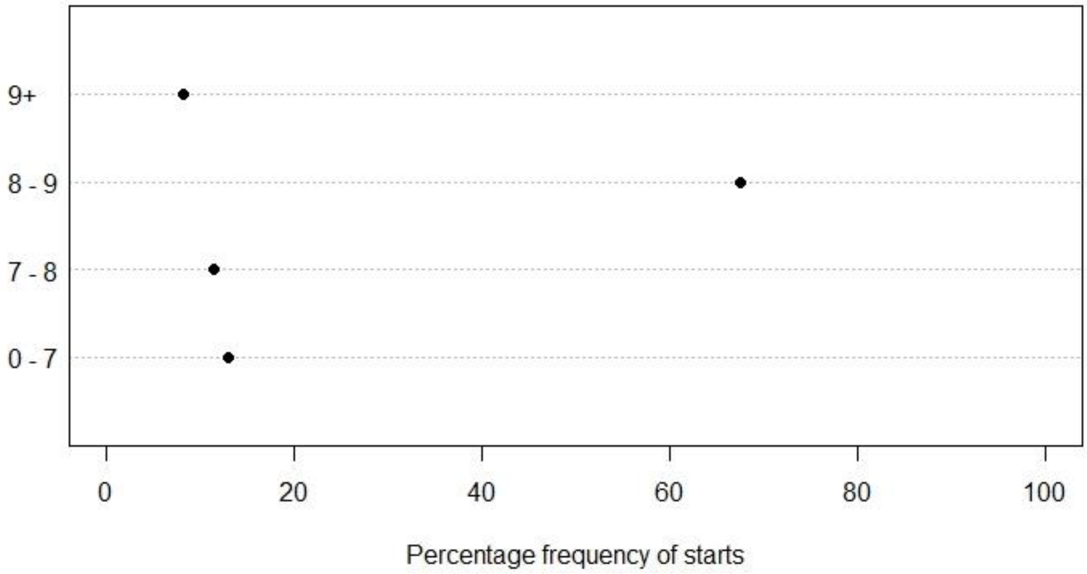

Figure 2-127 Dot plot of track size (furlongs)

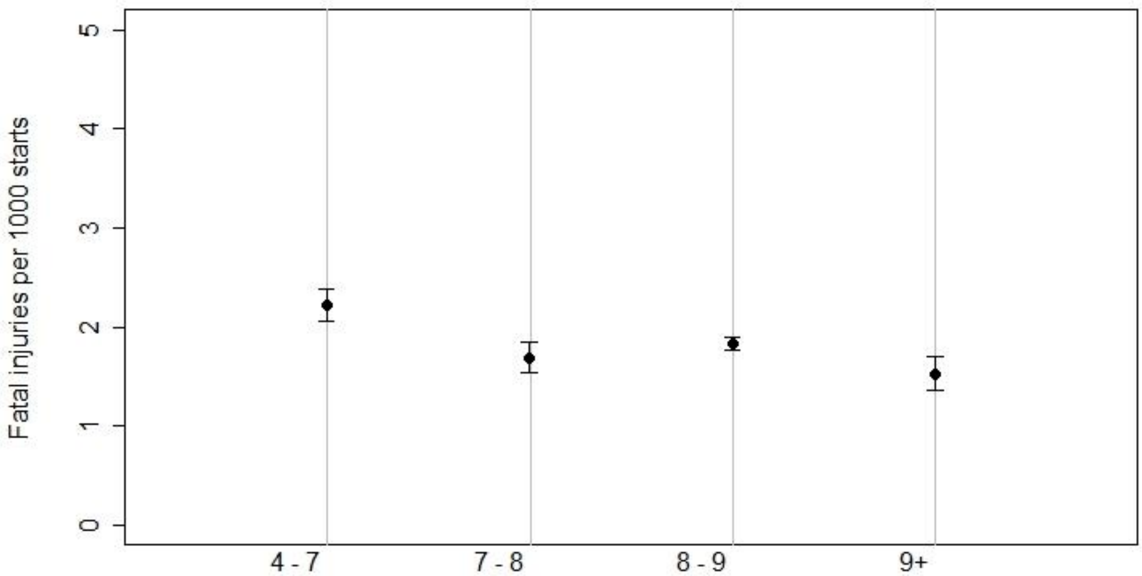

Figure 2-128 Fatal injuries per 1000 racing starts with 95% confidence interval by track size (furlongs) group

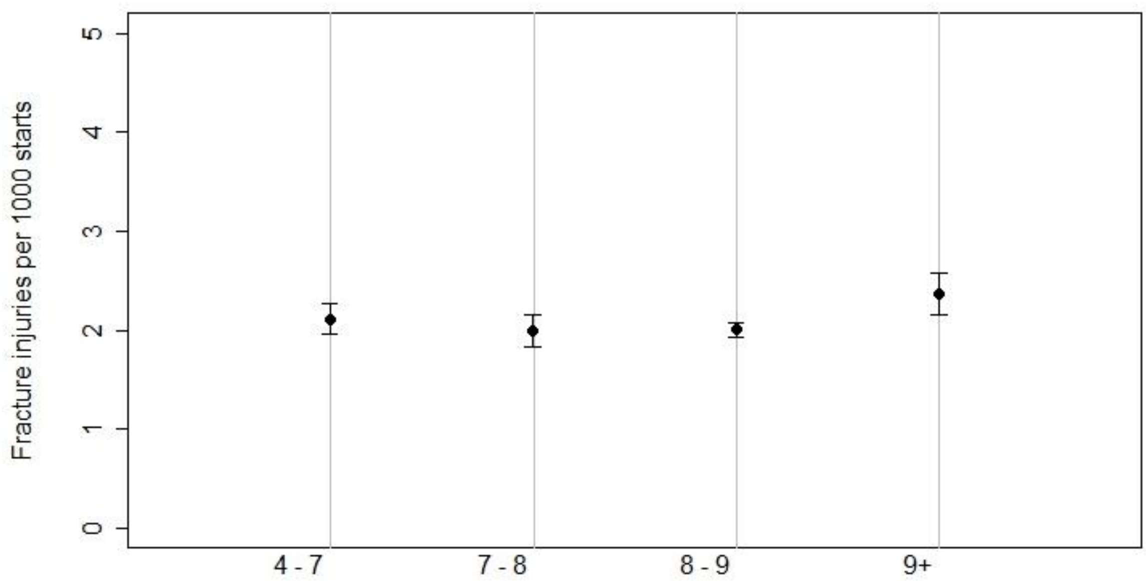

Figure 2-129 Fracture injuries per 1000 racing starts with 95% confidence interval by track size (furlongs) group

2.4.47. Racing with first trainer

The EID contained information on the trainer of each horse and we looked at the start of each horse with its first trainer. The proportion of racing starts for horses racing with their first trainer is 53% and the number of fatal and fracture injuries, along with their 95% confidence intervals, for each group are shown at figures 2-132 and 2-133 respectively.

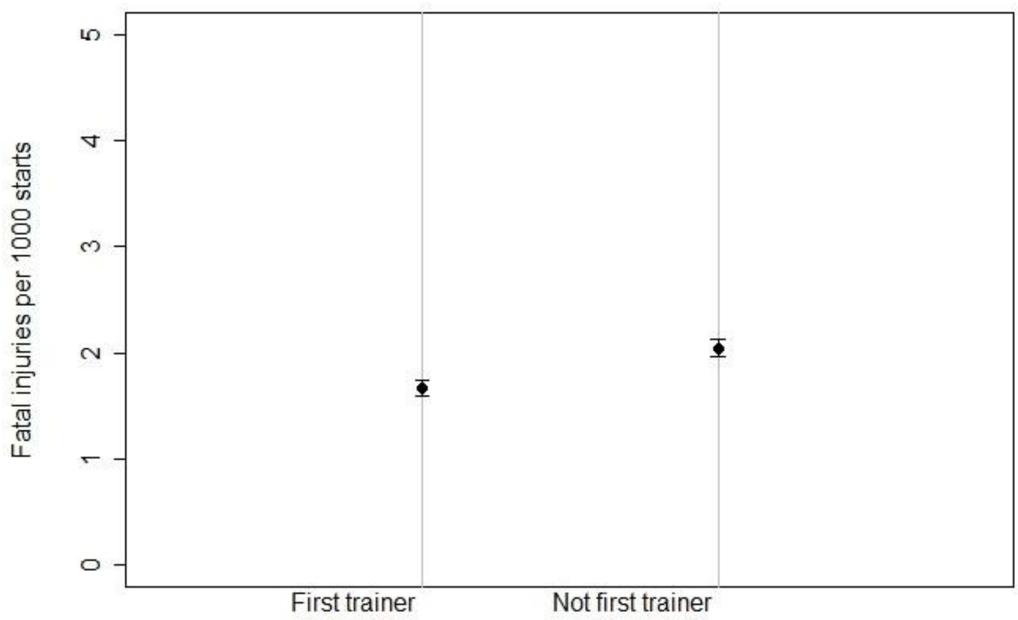

Figure 2-130 Fatal injuries per 1000 racing starts with 95% confidence interval by starts with first trainer

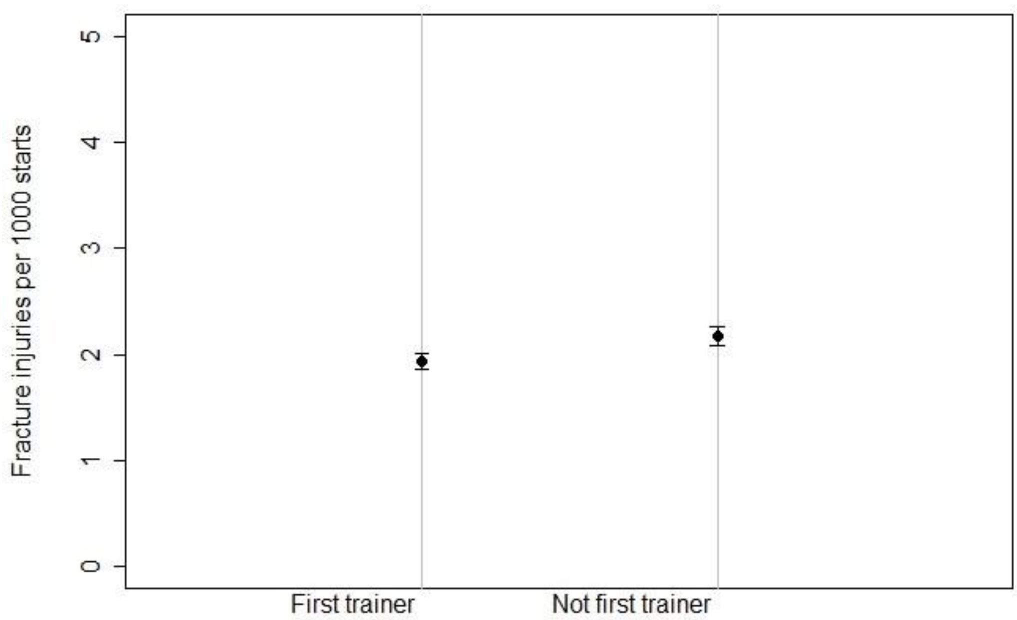

Figure 2-131 Fracture injuries per 1000 racing starts with 95% confidence interval by starts with first trainer

2.4.48. Wins/starts (Present – 30 days prior race)

From the racing starts available for each horse we calculated the number winning races per start a horse had in the period of 30 days prior the race. The proportion of starts by wins per start is shown at Figure 2-134 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-135 and 2-136 respectively.

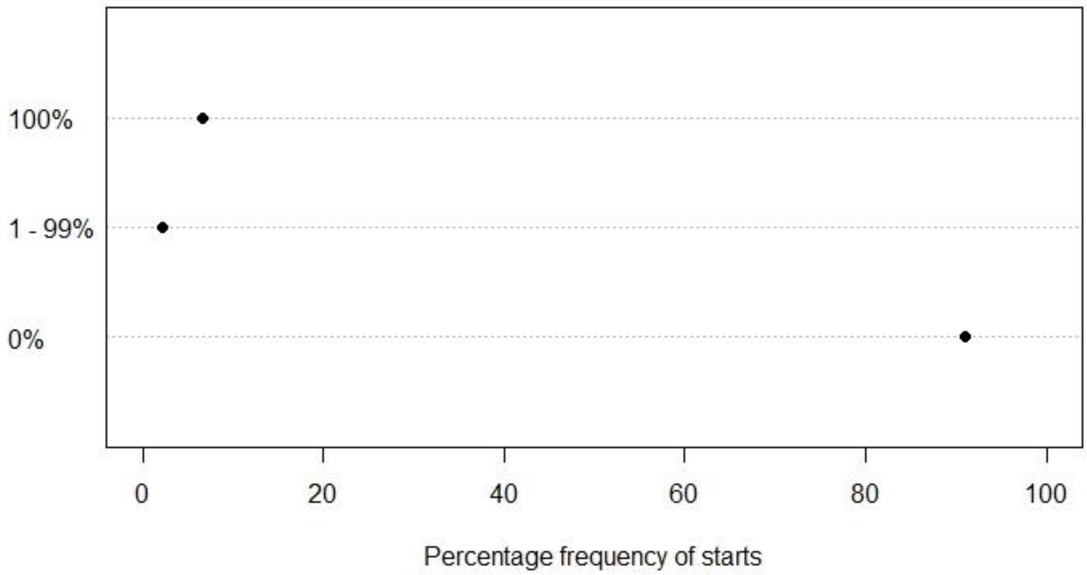

Figure 2-132 Dot plot of wins/starts (Present – 30 days prior race)

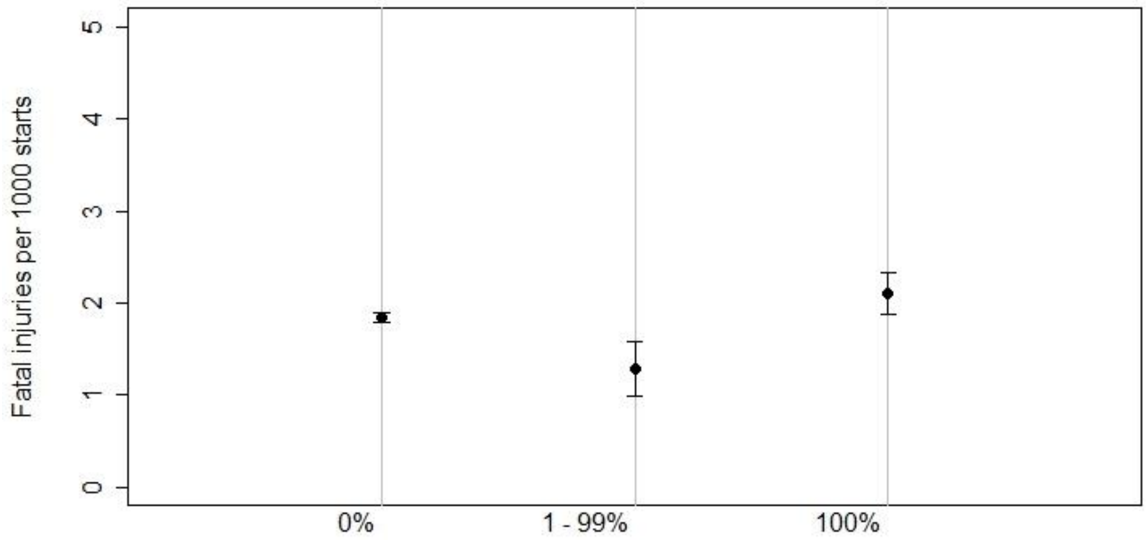

Figure 2-133 Fatal injuries per 1000 racing starts with 95% confidence interval by wins/starts (Present – 30 days prior race) group

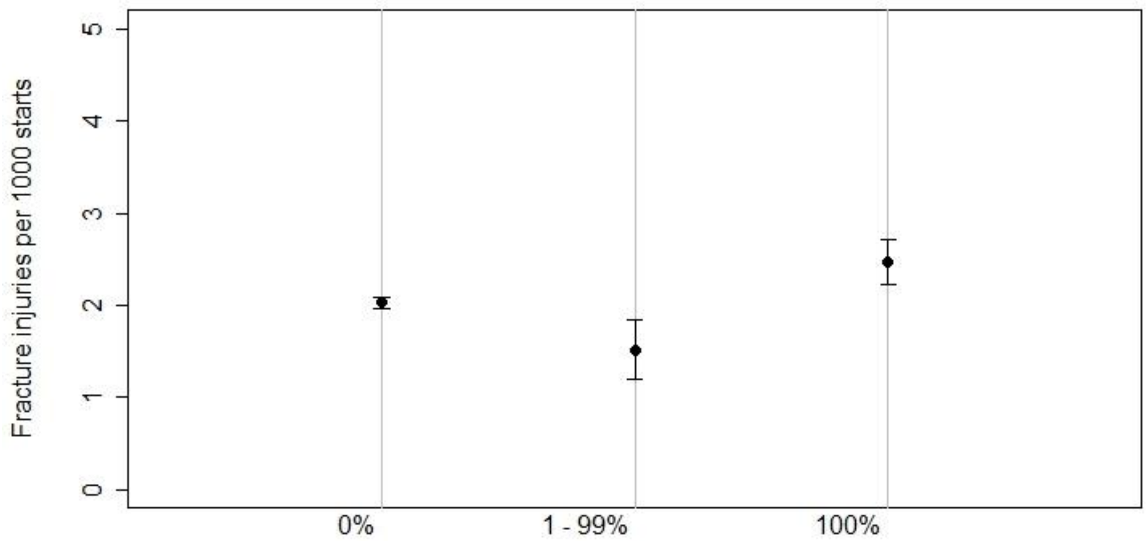

Figure 2-134 Fracture injuries per 1000 racing starts with 95% confidence interval by wins/starts (Present – 30 days prior race) group

2.4.49. Wins/starts (30 – 60 days prior race)

From the racing starts available for each horse we calculated the number winning races per start a horse had in the period of 30 to 60 days prior the race. The proportion of starts by wins per start is shown at Figure 2-137 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-138 and 2-139 respectively.

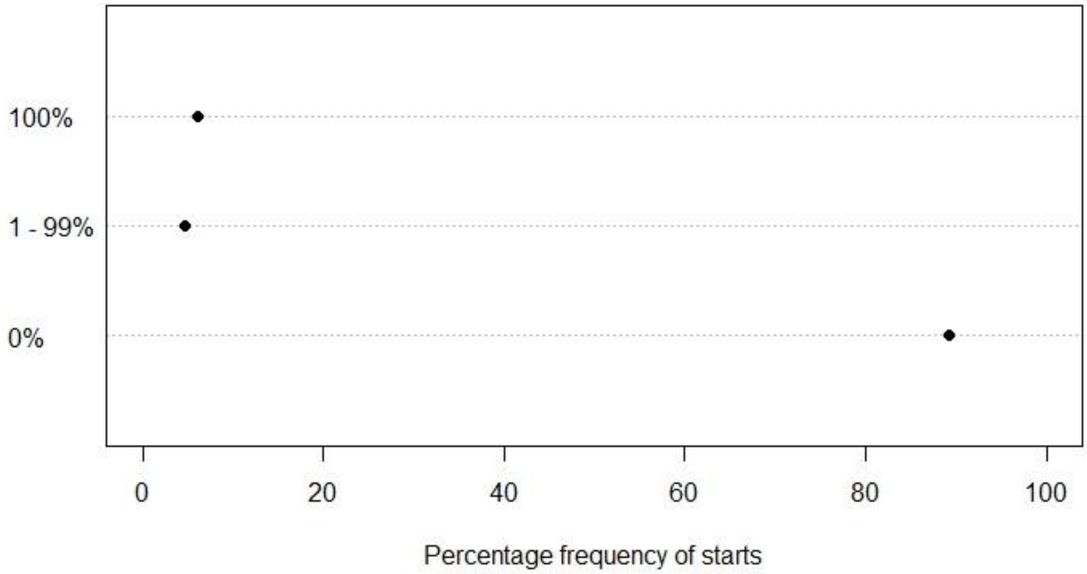

Figure 2-135 Dot plot of wins/starts (30 – 60 days prior race)

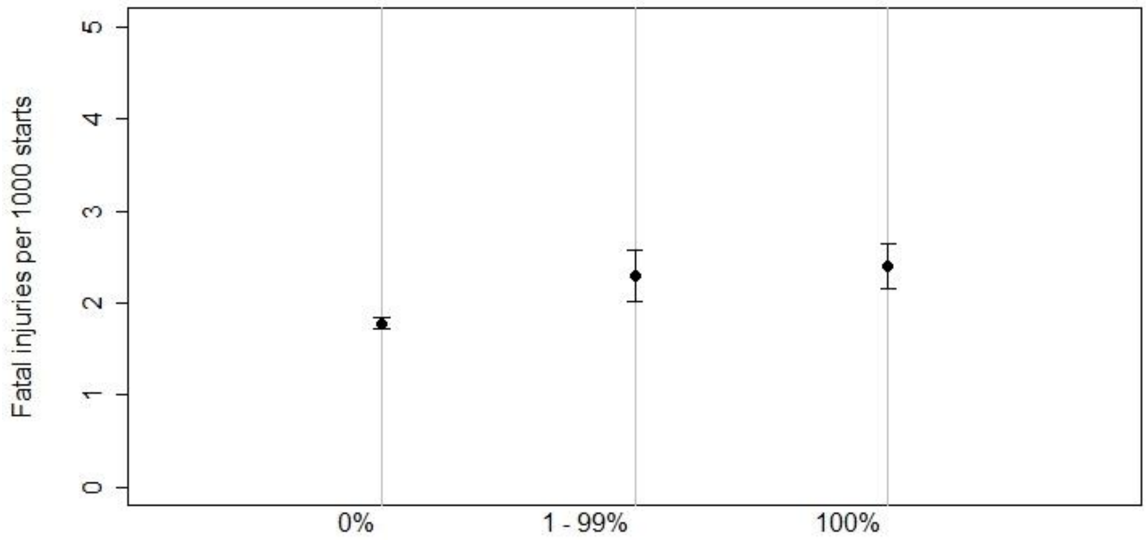

Figure 2-136 Fatal injuries per 1000 racing starts with 95% confidence interval by wins/starts (30 – 60 days prior race) group

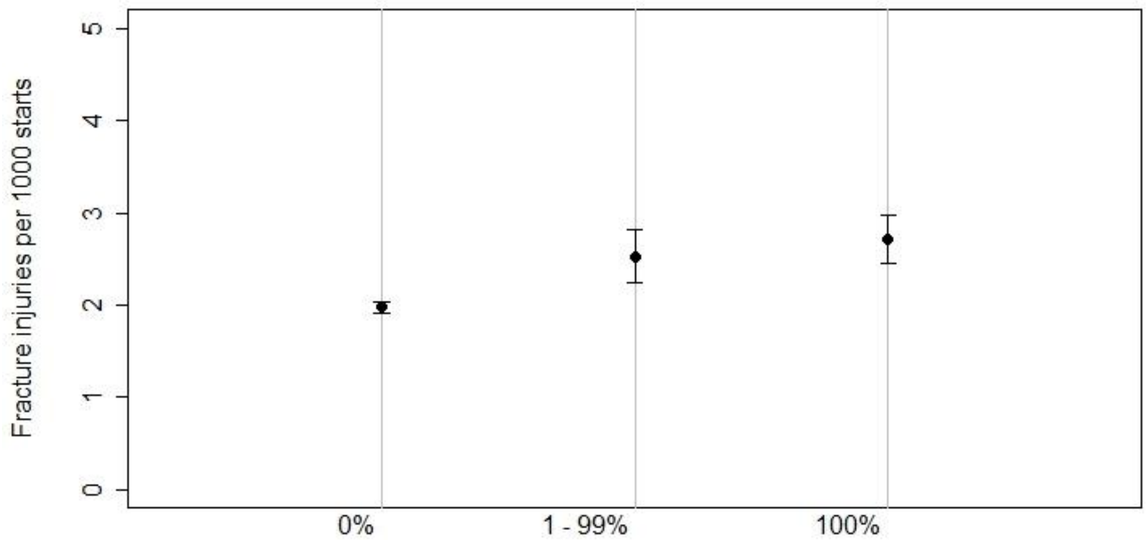

Figure 2-137 Fracture injuries per 1000 racing starts with 95% confidence interval by wins/starts (30 – 60 days prior race) group

2.4.50. Wins/starts (60 – 90 days prior race)

From the racing starts available for each horse we calculated the number winning races per start a horse had in the period of 60 to 90 days prior the race. The proportion of starts by wins per start is shown at Figure 2-140 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-141 and 2-142 respectively.

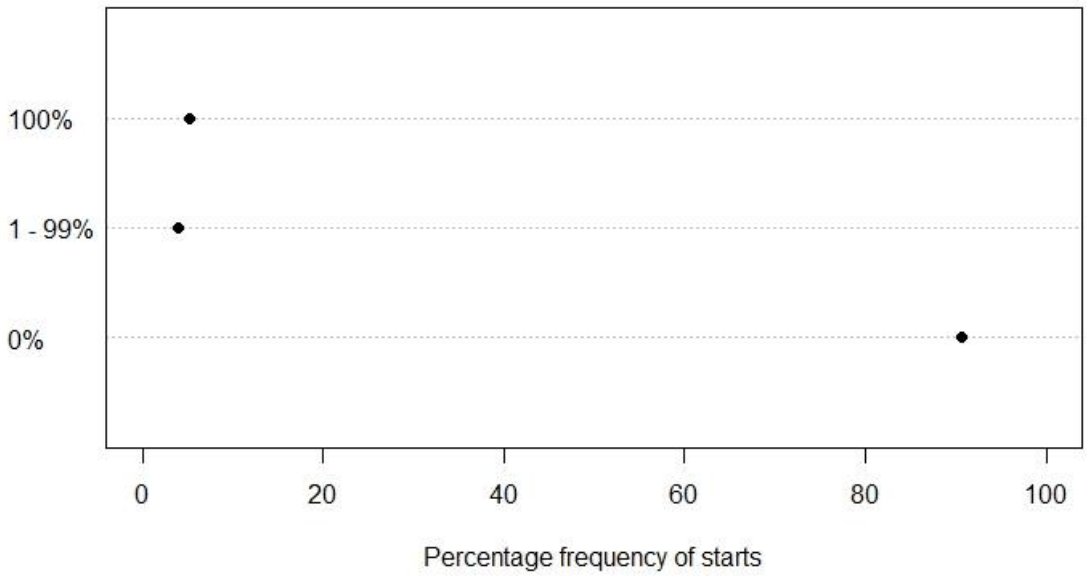

Figure 2-138 Dot plot of wins/starts (60 – 90 days prior race)

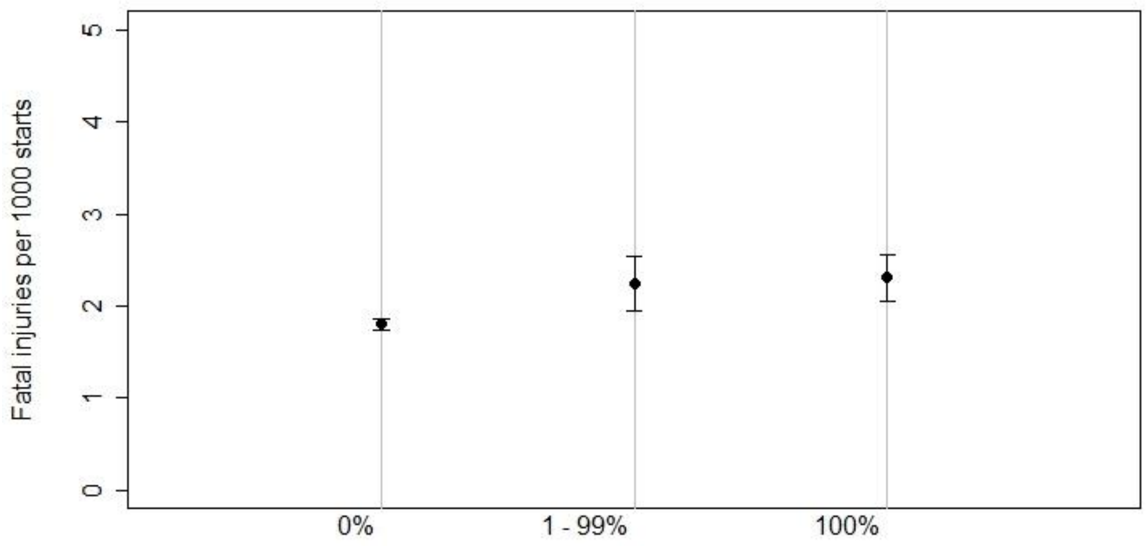

Figure 2-139 Fatal injuries per 1000 racing starts with 95% confidence interval by wins/starts (60 – 90 days prior race) group

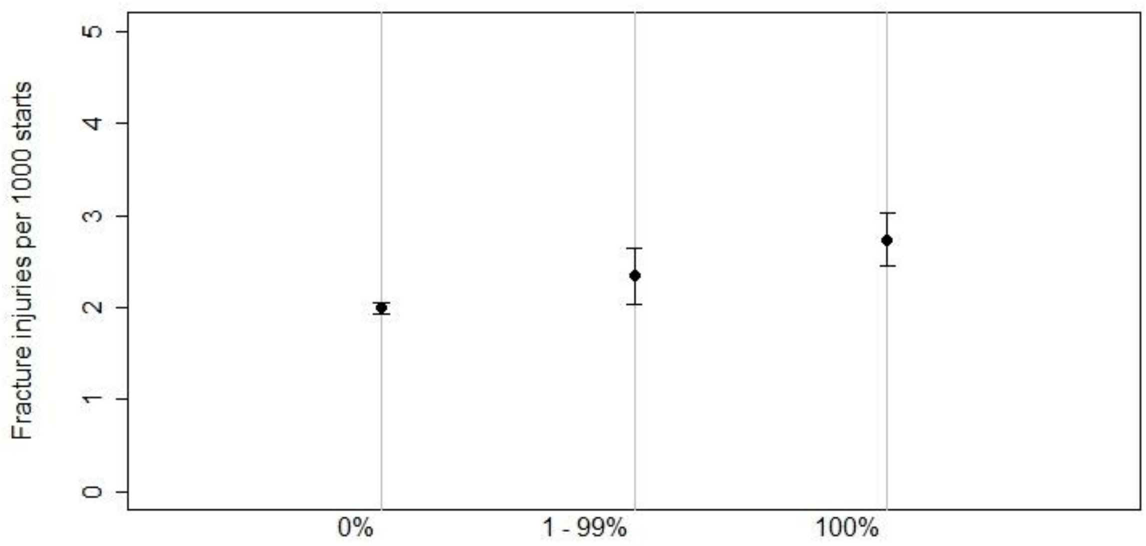

Figure 2-140 Fracture injuries per 1000 racing starts with 95% confidence interval by wins/starts (60 – 90 days prior race) group

2.4.51. Wins/starts (90 – 180 days prior race)

From the racing starts available for each horse we calculated the number winning races per start a horse had in the period of 90 to 180 days prior the race. The proportion of starts by wins per start is shown at Figure 2-143 and the number of fatal and fracture injuries, along with their 95% confidence intervals, for the different groups are shown at figures 2-144 and 2-145 respectively.

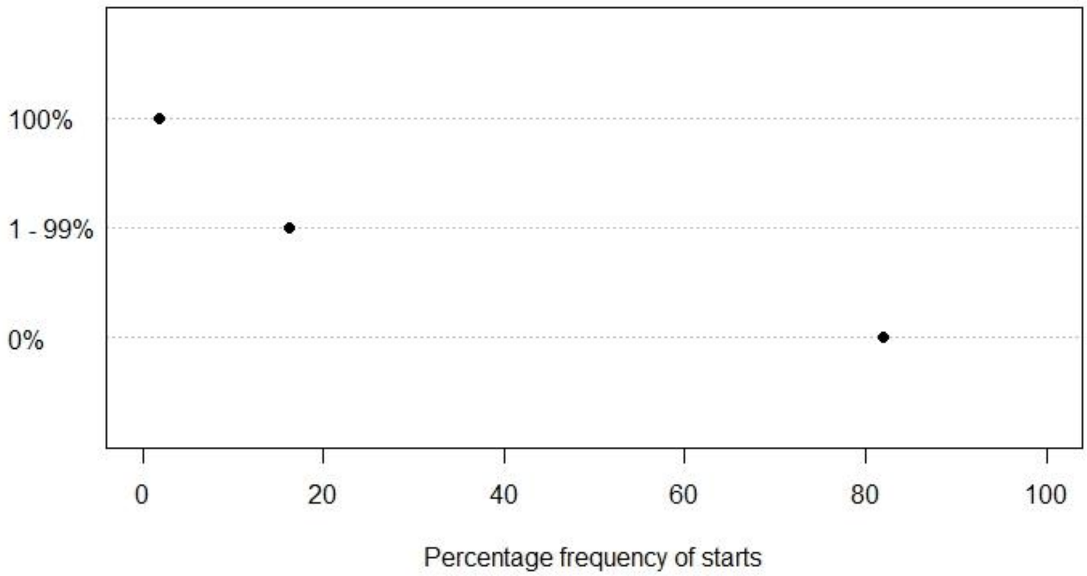

Figure 2-141 Dot plot of wins/starts (90 – 180 days prior race)

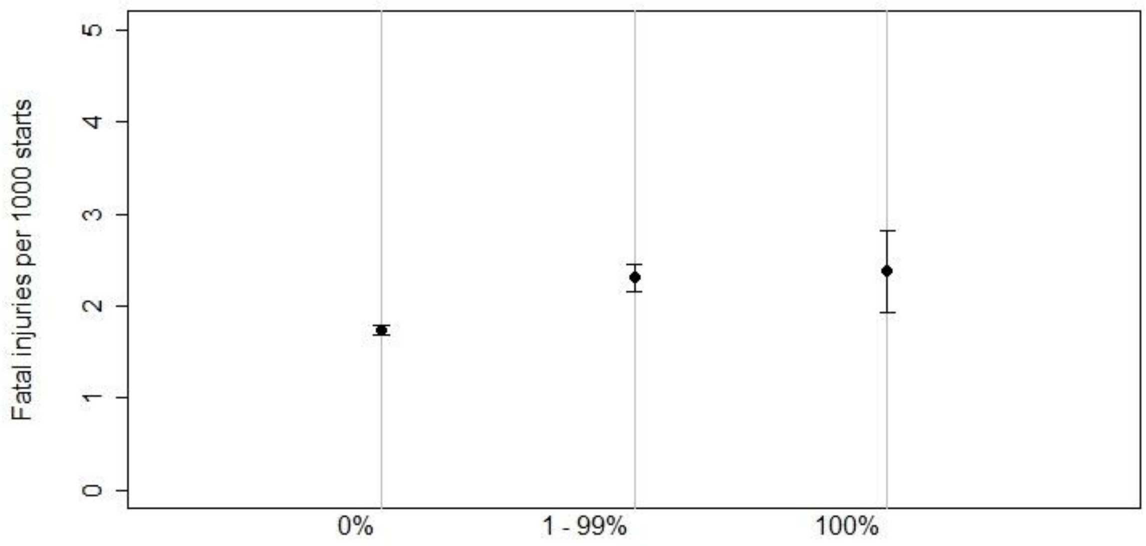

Figure 2-142 Fatal injuries per 1000 racing starts with 95% confidence interval by wins/starts (90 – 180 days prior race) group

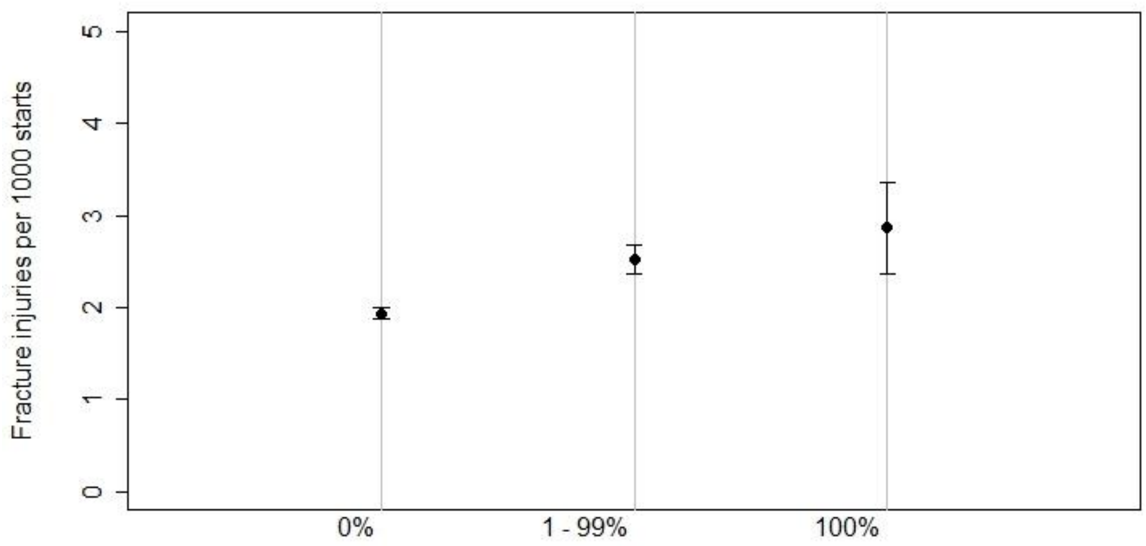

Figure 2-143 Fracture injuries per 1000 racing starts with 95% confidence interval by wins/starts (90 – 180 days prior race) group

3. Risk Factors for Fatal Injuries

3.1. Introduction

This part of the study is focused on equine fatal injuries in flat horse racing of Thoroughbreds in the US and Canada between January 1 2009 and December 31 2015.

Previous studies investigating equine injuries in North America and other jurisdictions have identified risk factors associated with equine injuries in flat racing. Regarding risk factors related to the racehorse, age has been identified as a significant risk factor associated with fatal musculoskeletal injury in California (Estberg, et al., 1996b; Estberg, et al., 1998b), with breakdown in Australia (Bailey, et al., 1997) and with fatal injuries (Williams, et al., 2001; Henley, et al., 2006), sudden death (Lyle, et al., 2012) and fatal lateral condylar fracture (Parkin, et al., 2005) at UK racecourses. The sex of the horse has been identified as a significant risk factor associated with fatal injury in studies in California (Estberg, et al., 1996b; Estberg, et al., 1998b), in Florida (Hernandez, et al., 2001) and in Australia (Boden, et al., 2007a). Prior racing history has been found to be associated with equine fatalities in studies in California (Estberg, et al., 1998a), in the UK (Parkin, et al., 2004b; Parkin, et al., 2005; Henley, et al., 2006; Lyle, et al., 2012) and Australia (Boden, et al., 2007a) with males being at greater risk. Exercise history and distance galloped during training have also been shown to be associated with fatal injuries in California (Estberg, et al., 1995; Estberg, et al., 1996a; Estberg, et al., 1998a), in Florida (Hernandez, et al., 2005) and in the UK (Parkin, et al., 2004a; Parkin, et al., 2005; Lyle, et al., 2012).

Regarding risk factors related to the racecourse, the distance of the race has been found to be associated with fatal injuries in California (Estberg, et al., 1998a) and in Kentucky (Peloso, et al., 1994), in the UK (Parkin, et al., 2004b; Parkin, et al., 2005; Henley, et al., 2006; Lyle, et al., 2012) and in Australia (Boden, et al., 2007a). The surface and condition of the racetrack has been identified to be associated with fatal injuries in California (Arthur, 2010) and in the UK (Williams, et al., 2001; Parkin, et al., 2004a,b; Parkin, et al., 2005; Henley, et al., 2006). The type of the race has been identified to be associated with fatal injuries in California (Estberg, et al., 1998b), in Kentucky (Peloso, et al., 1994) and in the UK (Williams, et al., 2001; Henley, et al., 2006; Parkin, et al., 2006; Lyle, et al., 2012). Finally, field size has been found to be associated with equine fatalities in the UK (Parkin, et al., 2004b; Parkin, et al., 2005).

The aim at this part of the study was to identify risk factors associated with fatal injuries in Thoroughbred flat racing in the US and Canada from 2009 to 2015. Specific factors such as age, sex, race distance, racetrack surface type and conditions, race type, field size, and prior racing history that were associated with fatal injuries in racehorses in previous studies both in in North America and other jurisdictions were evaluated. We also evaluated risk factors specific to North America and made use of the recorded exercise history available to identify risk factors unique to the US and Canada.

3.2. Materials and Methods

3.2.1. Study Design

The analysis reported in this thesis is an observational retrospective cohort study based on racecourses reporting injuries to the EID from 1st January 2009 to 31st December 2015. 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. These data include 2,493,957 race starts from the 89 race tracks reporting injuries to the EID and 4,592,162 exercise starts. The race starts reported to the EID represented the starts for 90.0% of all official Thoroughbred racing events in the United States and Canada during the 7-year observation period.

3.2.2. Case definition

The study was conducted with race start as the unit of analysis as fatal injuries are reported at the start level. This approach allowed analysis of all start level risk factors of interest and allowed comparison of the current results with those of previous research in this field. Cases were defined as starts from horses that died or were euthanized within three days of sustaining an injury during a race. All other race starts from race tracks reporting injuries to the EID were classified as controls. Exercise starts were only used to quantify prior exercise history for each horse entering a race. Horses that died or were euthanised more than three days following a race were not excluded from analysis. Horses that died or were subject to euthanasia following an exercise start were not included in the case population. The range and types of (fatal) injury are described in Chapter 2 of this Thesis.

3.2.3. Risk Factors

A total of 51 potential risk factors were identified from previous studies and from *a priori* hypotheses and were considered in our analysis. These included horse-related potential risk factors and race-related risk factors. The EID database also contained information for approximately 11,000 anonymized trainers and 3,000 anonymized jockeys associated with the recorded races, which enabled us to analyze trainer- and jockey-related risk factors.

3.2.4. Statistical Analysis

Initially, the linear relationship between numerical potential risk factors was assessed by examination of graphical plots of the log odds of the potential risk factors and fatal injuries (Boden, et al., 2007a; Reardon, 2013). If the relationship could be considered non-linear in the log scale, we created categorical alternatives for the risk factors. Binary and polytomous (5-level) categorical terms were considered and the form of the variable that produced the best fit in a univariable model based on the Akaike information criterion (AIC) was retained. For parsimony and to facilitate interpretability of the results we did not consider polynomial terms or interaction terms. Finally, when examining the potential risk factor of the purse of the race a categorical variable was introduced *ad hoc* to specifically explore low purse races.

The association between each potential risk factor and fatal injury was assessed by creating a univariable regression model. Wald P-values were calculated and risk factors with values of $P < 0.20$ in univariable analysis were eligible for inclusion in a multivariable logistic regression model. A threshold of $P < 0.20$ was chosen to prevent exclusion of a potentially significant risk factor that only becomes evident when a confounder has been controlled for in a multivariable analysis (Dohoo, et al., 2003). An automated stepwise selection process was used to build the multivariable model. Potential risk factors were identified by use of a forward bidirectional elimination approach and assessment of their AIC. We preferred a forward stepwise approach compared to a backwards stepwise approach. A backwards approach usually results in more variables retained in the final model. As we have a plethora of available data, and consequently high statistical power, we are confident that a variable that contributes information to the final model would not be excluded. The AICs for competing models were compared, and the model with the lowest AIC was preferred (Bozdogan, 1987). Only risk factors with a statistical significance indicated by a Wald P value of less than 0.05 were retained in the final models.

To assess risk factors that summarize historical racing information prior to each race start, we followed the same analytical procedure to arrive at a second multivariable model on a sub-sample of the population consisting of all the starts from horses at least six months after their first recorded racing or exercise start. Using this sample of the data we were able to assess the relationship between the number of starts a horse had up to six months prior to the race and fatal injuries. This facilitates the interpretation of the relationship as it excludes horses that would have fewer or no starts due to having only recently started racing.

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. However, for the risk factors included in the final models we checked for possible collinearity (Bagley, et al., 2001); correlation coefficients were produced for all pairs with a threshold for inclusion set at 0.7.

Furthermore, we evaluated potential confounders by resubmitting the variables that were excluded from the final model during the stepwise selection process (Boden, et al., 2007a; Reardon, 2013). If the potentially confounding variable altered odds ratios for variables in the final model by >20% (Dohoo, et al., 2003), we retained the confounder in the final model.

Furthermore, the potential effect of horse in the data analyses was evaluated by creating a mixed-effects model that included horse as a random effect (Reardon, 2013; Boden, et al., 2007a; Lyle, et al., 2012). Results were nearly identical (less than 10% change in ORs and no meaningful changes in P values) to results obtained with models that did not include random effects so the single level fixed models were retained. We did not further check mixed models with racecourse, jockey or trainer as random effects. There is little indication in the literature that there are meaningful changes in those models compared to the single level fixed models and we do not think that that would be the case our study especially given the thousand different jockeys and trainers involved.

Model fit was assessed by using the Hosmer-Lemeshow goodness of fit test (Hosmer & Lemeshow, 2000; Dohoo, et al., 2003). Furthermore, we checked for the existence of influential observations in our final models (Cook & Weisberg, 1982; Williams, 1987).

The predictive ability of the models was assessed by calculating the area under the receiver operating characteristic curve (Bozdogan, 1987). Furthermore, the top 5% of fitted scores from the models were used to assess the ability of the models to identify a population of starts with higher prevalence of fatal injury than the average.

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

3.3. Results

3.3.1. Results for all horses

3.3.1.1. Study Population

The study population comprised all 188,269 Thoroughbred horses that participated in flat racing in the US and Canada from 1st January 2009 to 31st December 2015 at the 89 race tracks reporting injuries to the EID. The prevalence of fatal injury was 0.18% for the 2,493,957 racing starts in the 7-year study period.

3.3.1.2. Univariable Models

In total 33 possible risk factors were screened using univariable analysis (table 3-1); 28 of them were found to have a statistically significant association with fatal injuries (P < 0.05). Of the possible risk factors 30 were found to have a P-value of less than 0.20 and were included in the subsequent forward bidirectional elimination to be potentially included in the final multivariable model.

Table 3-1 Results of univariable logistic regression for assessment of risk factors associated with fatal injuries in Thoroughbred racehorses competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Controls - Cases	OR (95% CI)	P value
Age (years)	2,489,358 – 4,599	1.026 (1.007-1.045)	0.007
Age at first start (years)	2,489,358 – 4,599	1.070 (1.045-1.095)	< 0.001
Country			
Canada	200,465 - 225	Ref	Ref
US	2,288,893 - 4,374	1.703 (1.489-1.947)	< 0.001

Table 3-1 (Continued)

Risk factor	Controls - Cases	OR (95% CI)	P value
Entered the vet list			
No	2,012,776 - 3,238	Ref	Ref
Yes	476,582 - 1,361	1.775 (1.666-1.891)	< 0.001
Field size	2,489,358 - 4,599	1.010 (0.995-1.025)	0.200
First Start			
No	2,310,661 - 4,369	Ref	Ref
Yes	178,697 - 230	0.681 (0.596-0.777)	< 0.001
Low purse race (<= $7500)			
No	2,126,510 - 3,965	Ref	Ref
Yes	362,848 - 634	0.937 (0.862-1.019)	0.129
Months since last racing start	2,489,358 - 4,599	0.999 (0.985-1.014)	0.938
Months since last racing or exercise start	2,489,358 - 4,599	1.085 (1.067-1.104)	< 0.001
No. of layups	2,489,358 - 4,599	0.950 (0.926-0.975)	< 0.001
No. of previous injuries	2,489,358 - 4,599	1.510 (1.336-1.706)	< 0.001
No. of previous vet scratches	2,489,358 - 4,599	1.084 (1.048-1.120)	< 0.001
No. of previous non-vet scratches	2,489,358 - 4,599	1.018 (1.003-1.034)	0.021
Odds at start of race	2,489,358 - 4,599	0.994 (0.992-0.996)	< 0.001
Odds rank in race	2,489,358 - 4,599	0.944 (0.934-0.955)	< 0.001
Post position	2,489,358 - 4,599	1.009 (0.998-1.020)	0.112
Purse ($1000)	2,489,358 - 4,599	0.998 (0.997-0.999	< 0.001
Race distance (furlongs)	2,489,358 - 4,599	0.923 (0.902-0.944)	< 0.001
Season			
Autumn	646,489 - 1,247	Ref	Ref
Spring	610,600 - 1,058	0.898 (0.828-0.975)	0. 010
Summer	775,590 - 1,335	0.892 (0.826-0.964)	0.004
Winter	456,679 - 959	1.089 (1.001-1.184)	0.048
Sex			
Mare/Gelding	2,179,652 - 3,874	Ref	Ref
Stallion	309,706 - 725	1.317 (1.217-1.426)	< 0.001

Table 3-1 (Continued)

Risk factor	Controls - Cases	OR (95% CI)	P value
Start with new jockey			
No	1,195,620 - 2,090	Ref	Ref
Yes	1,293,738 - 2,509	1.109 (1.047-1.176)	< 0.001
Start with new trainer			
No	2,250,492 - 4,034	Ref	Ref
Yes	238,866 - 565	1.226 (1.157-1.299)	< 0.001
Surface			
Synthetic	297,211 - 363	Ref	Ref
Dirt	1,837,893 - 3,684	1.641 (1.473-1.828)	< 0.001
Turf	354,254 - 552	1.276 (1.117-1.457)	< 0.001
Time between exercise starts – avg (months)	2,489,358 - 4,599	1.013 (0.994 -1.032)	0.195
Time between exercise starts - active – avg (months)	2,489,358 - 4,599	1.028 (1.003 -1.053)	0.003
Time between racing starts - avg (months)	2,489,358 – 4,599	1.029 (1.018 -1.039)	< 0.001
Time between racing starts - active - avg (months)	2,489,358 – 4,599	1.069 (1.047 -1.090)	< 0.001
Time in racing – active (months)	2,489,358 - 4,599	1.000 (0.998-1.003)	0.757
Time in racing (months)	2,489,358 - 4,599	1.000 (0.998-1.002)	0.746
Time with same jockey (months)	2,489,358 - 4,599	0.986 (0.972-1.000)	0.050
Time with same trainer (months)	2,489,358 – 4,599	0.984 (0.981-0.988)	< 0.001
Track size (furlongs)	2,489,358 – 4,599	0.930 (0.906-0.955)	< 0.001
Training with first trainer			
Yes	1,333,130 - 2,229	Ref	Ref
No	1,156,228 - 2,370	1.127 (1.063-1.195)	< 0.001

3.3.1.3. Multivariable Model

The 30 possible risk factors with a P-value of less than 0.20 in the univariable analysis were included in a forward bidirectional elimination using AIC to assess the models created. Following this procedure, we arrived at the final multivariable model (Table 3-2).

Table 3-2 Results of multivariable logistic regression for assessment of risk factors associated with fatal injuries in Thoroughbred racehorses competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Odds ratio	95% CI	P-value
(Intercept)	0.002	0.002 - 0.003	< 0.001
Age at first start (years)	1.092	1.066 - 1.119	< 0.001
Country			
Canada	Ref	Ref	Ref
US	1.522	1.321 - 1.754	< 0.001
Entered the vet list			
No	Ref	Ref	Ref
Yes	1.747	1.634 - 1.867	< 0.001
First Start			
No	Ref	Ref	Ref
Yes	0.682	0.589 - 0.789	< 0.001
Low purse race (<= $7500)			
No	Ref	Ref	Ref
Yes	0.812	0.744 - 0.886	< 0.001
No. of layups	0.928	0.899 – 0.957	< 0.001
No. of previous injuries	1.287	1.133 - 1.461	< 0.001
Odds rank in race	0.948	0.937 - 0.959	< 0.001
Post position	1.017	1.006 - 1.028	0.003
Race distance (furlongs)	0.927	0.905 - 0.949	< 0.001
Season			
Autumn	Ref	Ref	Ref
Spring	0.863	0.795 - 0.937	< 0.001
Summer	0.908	0.840 - 0.982	0.015
Winter	1.048	0.962 - 1.141	0.285
Sex			
Mare/Gelding	Ref	Ref	Ref
Stallion	1.473	1.363 - 1.592	< 0.001

Table 3-2 (Continued)

Risk factor	Odds ratio	95% CI	P-value
Surface			
Synthetic	Ref	Ref	Ref
Dirt	1.443	1.291 - 1.612	< 0.001
Turf	1.197	1.042 - 1.376	0.011
Time between racing starts – avg (months)	1.031	1.020 - 1.042	< 0.001
Time with same trainer (months)	0.989	0.984 - 0.995	< 0.001
Track size (furlongs)	0.923	0.896 - 0.950	< 0.001
Training with first trainer			
Yes	Ref	Ref	Ref
No	1.093	1.019 - 1.172	0.013

The final multivariable model included 17 risk factors with a statistically significant association with fatal injuries.

From the horse-related risk factors we have identified, the one with the highest potential impact is related to horses that had at some point in their career entered the vet list. Those horses had 74.7% (95% CI: 63.4% - 86.7%) more chance of sustaining a fatal injury than horses that had never been on the vet list. The horse-related risk factor with the second highest potential impact was the sex of the horse. There was 47.3% (95% CI: 36.3% - 59.2%) more chance of sustaining a fatal injury for stallions compared to mares and geldings. Conversely, horses were 31.8% (95% CI: 21.1% - 41.1%) less likely to sustain a fatal injury during their first racing start, compared with all subsequent starts. The number of previous injuries a horse had in its career was also found to be associated with fatal injuries, with horses having a 28.7% (95% CI: 13.3% - 46.1%) higher chance of fatal injury for each previous EID-reported injury they had sustained. The age at which a horse begins its racing career was found to be associated with the risk of equine fatality. We found that for each year older the horse was at their first racing start there was a 9.2% (95% CI: 6.6% - 11.9%) higher chance of fatal injury in all its starts. Horses were also found to have a lower chance of sustaining a fatal injury by 7.2% (95% CI: 4.3% - 10.1%) for each 60-day layup they had in their career though for each extra month time they had on average between racing starts they had 3.1% (95% CI: 2.0% - 4.2%) higher chance of injury. Finally, when ranking the horses of a race according to their betting odds, horses less favoured by the odds were found to be less likely to sustain a fatal injury by 5.2% (95% CI: 4.1% - 6.3%) for each place further down the odds ranking.

Regarding race-related risk factors, the one with the highest potential impact is the country of the race, with horses participating in races in the US having a 52.2% (95% CI: 32.1% - 75.4%) higher chance of fatal injury than those competing in Canada. The risk factor with the second highest potential impact was the surface of the race. Horses were at 44.3% (95% CI: 29.1% - 61.2%) higher chance of fatal injury when running on a track with dirt surface compared to a synthetic one and at 19.7% (95% CI: 4.2% - 37.6%) more risk when running on turf compered to synthetic surfaces. We also found that horses participating in races with a purse of equal or less than $7,500 were 18.8% (95% CI: 11.4% - 25.6%) less likely to sustain a fatal injury compared with horses competing in races with a purse greater than $7,500. Horses were also at less risk when racing in the spring, by 13.7% (95% CI: 6.3% - 20.5%), and in the summer by 9.2% (95% CI: 1.8% - 16.0%), compared to when racing in autumn. The distance of the race was also found to be associated with fatal injuries with horses racing in longer races having 7.3% (95% CI: 5.1% - 9.5%) less risk per extra furlong of the race. Moreover, the track size was also found to be associated with fatal injuries with horses racing at longer tracks having 7.7% (95% CI: 5.0% - 10.4%) less risk per extra furlong of the track. The post position was also found to be associated with fatal injuries with horses having a 1.7% (95% CI: 0.6% - 2.8%) higher chance of sustaining a fatal injury for each post position further from the inside running rail at the start of race.

Finally, we were able to identify risk factors that were related to the trainer of the racing horse. For each extra month a horse spends with the same trainer, it has 1.1% (95% CI: 0.5% - 1.6%) less chance of sustaining a fatal injury. Furthermore, horses that changed trainer in their career and were no longer training with their first trainer had 9.3% (95% CI: 1.9% - 17.2%) higher chance of sustaining a fatal injury.

3.3.1.4. *Model Fit*

The multivariable model had a deviance of 66,188 with 2,493,936 degrees of freedom. The x^2 test statistic of the Hosmer-Lemeshow goodness of fit test was 13.375 with 8 degrees of freedom and a p-value of 0.10 indicating no evidence of a lack of fit.

No numerical variables were assessed to be non-linear in the log odds.

No influential observation was found in the final multivariable model.

The area under the receiving operating characteristic curve was 62.1%.

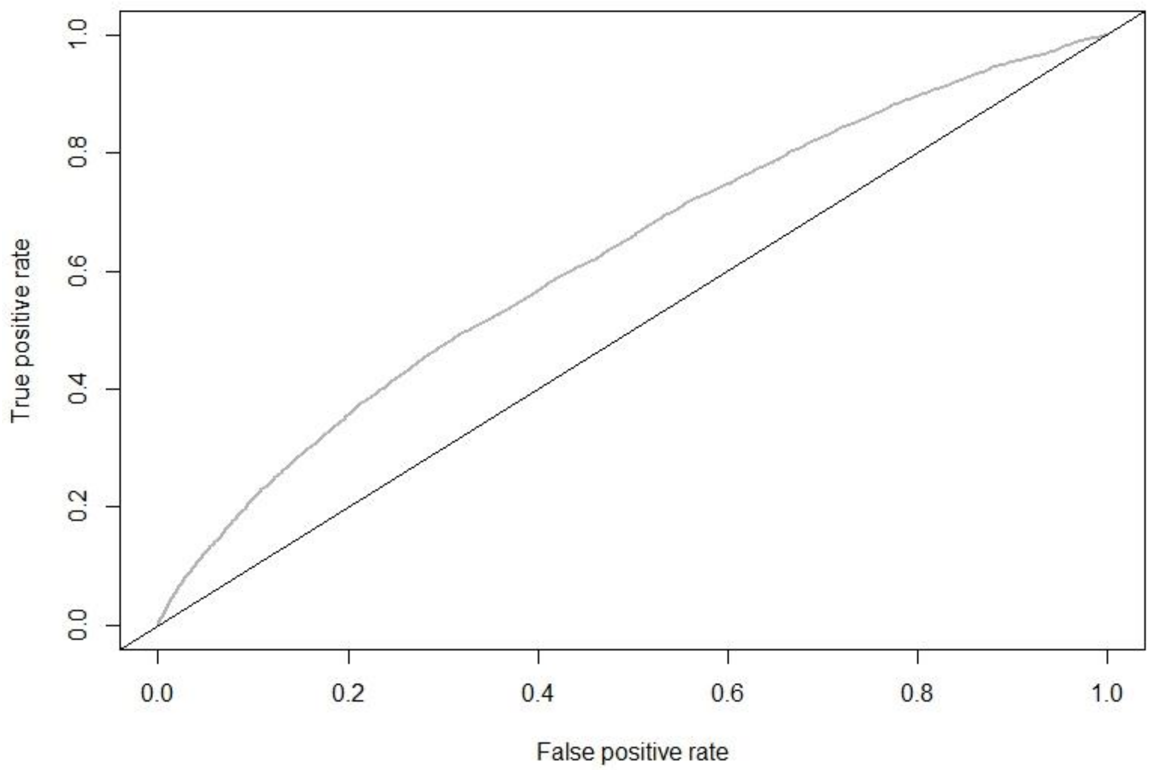

Figure 3-1 ROC curve of the multivariable logistic regression model trained on starts from 2009 to 2015 for fatal injury prediction for the same period

Using the top 5% of fitted scores from our model we were able to identify the starts that resulted in 2.5 times higher risk of fatal injury than the average injury prevalence for that study period.

3.3.2. Results for horses that had been racing for ≥ 6 months

3.3.2.1. Study Population

The study population comprised of 151,820 Thoroughbred horses that participated in flat racing in the US and Canada from 1st January 2009 to 31st December 2015 at the 89 race tracks reporting injuries to the EID and that had already had a racing or exercise start more than six months in the past. The prevalence of fatal injury was 0.19% for the 1,962,418 racing starts in the 7-year study period.

3.3.2.2. Univariable Models

In total 51 possible risk factors were screened using univariable analysis. A total of 40 of them were found to have a statistically significant association with fatal injuries ($P < 0.05$) and 42 of the possible risk factors were found to have a P-value of less than 0.20 and were included in the subsequent forward bidirectional elimination to be potentially included in the final multivariable model.

Table 3-3 Results of univariable logistic regression for assessment of risk factors associated with fatal injuries in Thoroughbred racehorses, six months after their first recorded racing or exercise start, competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Starts - Cases	OR (95% CI)	P-value
Accumulated distance ran in career (Km)	1,958,722 – 3,696	0.996 (0.995-0.997)	< 0.001
Accumulated exercise distance ran in career (Km)	1,958,722 - 3,696	0.997 (0.994-0.999)	< 0.001
Accumulated racing distance ran in career (Km)	1,958,722 - 3,696	0.993 (0.991-0.995)	< 0.001
Age (years)	1,958,722 - 3,696	1.014 (0.989-1.032)	0.351
Age at first start (years)	1,958,722 - 3,696	1.089 (1.058-1.121)	< 0.001
Average speed change on previous race (m/s)	1,958,722 - 3,696	1.009 (0.991-1.028)	0.331
Average speed in previous race (m/s)	1,958,722 - 3,696	1.034 (1.018-1.052)	< 0.001
Country			
Canada	153,176 - 171	Ref	Ref
US	1,805,546 - 3,525	1.749 (1.500-2.039)	< 0.001
Entered the vet list			
No	1,525,066 - 2,471	Ref	Ref
Yes	433,656 - 1,225	1.743 (1.628-1.867)	< 0.001
Field size	1,958,722 - 3,696	1.017 (1.000-1.034)	0.051
Low purse race (<= $7500)			
No	1,661,562 - 3.195	Ref	Ref
Yes	297,160 - 501	0.877 (0.798-0.963)	0.006
Months since last racing start	1,958,722 - 3,696	0.992 (0.977-1.008)	0.317
Months since last racing or exercise start	1,958,722 - 3,696	1.078 (1.058-1.098)	< 0.001
No. of layups	1,958,722 - 3,696	1.089 (1.058-1.121)	< 0.001
No. of previous injuries	1,958,722 - 3,696	1.478 (1.305-1.674)	< 0.001
No. of previous vet scratches	1,958,722 - 3,696	1.072 (1.036-1.110)	< 0.001

Table 3-3 (Continued)

Risk factor	Starts - Cases	OR (95% CI)	P-value
No. of previous non-vet scratches	1,958,722 - 3,696	1.010(0.994-1.026)	0.226
No. of racing and exercise starts (Present – 30 days prior race)	1,958,722 - 3,696	0.704 (0.680-0.729)	< 0.001
No. of racing and exercise starts (30 -60 days prior race)	1,958,722 - 3,696	0.889 (0.864-0.916)	< 0.001
No. of racing and exercise starts (60 -90 days prior race)	1,958,722 - 3,696	1.000 (0.975-1.027)	0.984
No. of racing and exercise starts (90 -180 days prior race)	1,958,722 - 3,696	1.046 (1.035-1.057)	< 0.001
No. of starts (Present – 30 days prior race)	1,958,722 - 3,696	0.856 (0.816-0.898)	< 0.001
No. of starts (30 – 60 days prior race)	1,958,722 - 3,696	1.060 (1.018-1.103)	0.005
No. of starts (60 – 90 days prior race)	1,958,722 - 3,696	1.123 (1.079-1.167)	< 0.001
No. of starts (90 – 180 days prior race)	1,958,722 - 3,696	1.086 (1.067-1.106)	< 0.001
Odds at start of race	1,958,722 - 3,696	0.993 (0.991-0.995)	< 0.001
Odds rank in race	1,958,722 - 3,696	0.940 (0.928-0.952)	< 0.001
Post position	1,958,722 - 3,696	1.010 (0.998-1.022)	0.994
Purse ($1000)	1,958,722 - 3,696	0.998 (0.996-9.999)	< 0.001
Race distance (furlongs)	1,958,722 - 3,696	0.900 (0.877-0.922)	< 0.001
Season			
Autumn	526,421 - 1,039	Ref	Ref
Spring	468,676 - 816	0.882 (0.805-0.967)	0.007
Summer	604,406 - 1,067	0.894 (0.821-0.974)	0.011
Winter	359,219 - 774	1.092 (0.995-1.198)	0.065
Sex			
Mare/Gelding	1,746,643 - 3,168	Ref	Ref
Stallion	212,079 - 528	1.373 (1.252-1.505)	< 0.001

Table 3-3 (Continued)

Risk factor	Starts - Cases	OR (95% CI)	P-value
Start with new jockey			
No	871,865 - 1,571	Ref	Ref
Yes	1,086,857 - 2,125	1.085 (1.017-1.158)	0.014
Start with new trainer			
No	1,748,839 - 3,184	Ref	Ref
Yes	209,883 - 512	1.340 (1.220-1.471)	< 0.001
Surface			
Synthetic	221,406 - 268	Ref	Ref
Dirt	1,447,562 - 2,977	1.699 (1.499-1.925)	< 0.001
Turf	289,754 - 451	1.286 (1.105-1.496)	< 0.001
Time between exercise starts – avg (months)	1,958,722 - 3,696	1.001 (0.980 -1.023)	0.926
Time between exercise starts - active – avg (months)	1,958,722 - 3,696	1.017 (0.990 -1.045)	0.222
Time between racing starts – avg (months)	1,958,722 - 3,696	1.023 (1.011 -1.034)	< 0.001
Time between racing starts – active - avg (months)	1,958,722 - 3,696	1.059 (1.034 -1.083)	< 0.001
Time in layup (months)	1,958,722 - 3,696	0.994 (0.989-0.998)	0.010
Time in racing – active (months)	1,958,722 - 3,696	0.997 (0.994-1.001)	0.129
Time in racing (months)	1,958,722 - 3,696	0.997 (0.995-0.999)	0.015
Time with same jockey (months)	1,958,722 - 3,696	0.982 (0.967-0.997)	0.016
Time with same trainer (months)	1,958,722 - 3,696	0.980 (0.976-0.984)	< 0.001
Track size (furlongs)	1,958,722 - 3,696	0.930 (0.902-0.958)	< 0.001
Training with first trainer			
Yes	848,807 - 1,420	Ref	Ref
No	1,109,915 - 2,276	1.229 (1.137-1.329)	< 0.001
Wins/starts (Present - 30 days prior race)	1,958,722 - 3,696	1.001 (0.999-1.002)	0.377

Table 3-3 (Continued)

Risk factor	Starts - Cases	OR (95% CI)	P-value
Wins/starts (30 – 60 days prior race)	1,958,722 - 3,696	1.003 (1.002-1.004)	< 0.001
Wins/starts (60 – 90 days prior race)	1,958,722 - 3,696	1.003 (1.001-1.004)	< 0.001
Wins/starts (90 – 180 days prior race)	1,958,722 - 3,696	1.005 (1.004-1.006)	< 0.001

3.3.2.3. Multivariable Model

The 42 possible risk factors with a P-value of less than 0.20 in the univariable analysis were included in a forward bidirectional elimination using AIC to assess the models created. Following this procedure, we arrived at the final multivariable model (Table 3-4).

Table 3-4 Results of multivariable logistic regression for assessment of risk factors associated with fatal injuries in Thoroughbred racehorses, six months after their first recorded racing or exercise start, competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Odds ratio	95% CI	P-value
(Intercept)	0.002	0.002 - 0.003	< 0.001
Accumulated racing distance ran in career (Km)	0.989	0.987 – 0.992	< 0.001
Age at first start (years)	1.099	1.067 – 1.132	< 0.001
Country			
Canada	Ref	Ref	Ref
US	1.295	1.102 - 1.522	0.002
Entered the vet list			
No	Ref	Ref	Ref
Yes	1.704	1.587 - 1.830	< 0.001
Field size	1.049	1.030 – 1.068	< 0.001
Low purse race (<= $7500)			
No	Ref	Ref	Ref
Yes	0.787	0.714 - 0.868	< 0.001

Table 3-4 (Continued)

Risk factor	Odds ratio	95% CI	P-value
No. of previous injuries	1.292	1.136 - 1.470	< 0.001
No. of racing and exercise starts (Present – 30 days prior race)	0.717	0.691 - 0.745	< 0.001
No. of racing and exercise starts (30 -60 days prior race)	0.942	0.912 - 0.973	0.001
No. of racing and exercise starts (90 - 180 days prior race)	1.057	1.042 - 1.072	< 0.001
No. of racing starts (90 - 180 days prior race)	1.046	1.021 – 1.073	< 0.001
Odds rank in race	0.938	0.926 - 0.951	< 0.001
Race distance (furlongs)	0.943	0.918 - 0.969	< 0.001
Sex			
Mare/Gelding	Ref	Ref	Ref
Stallion	1.454	1.323 - 1.598	< 0.001
Surface			
Synthetic	Ref	Ref	Ref
Dirt	1.388	1.218 - 1.581	< 0.001
Turf	1.156	0.986 - 1.353	0.074
Time between racing starts - avg (months)	1.025	1.012 - 1.038	< 0.001
Time in layup (months)	1.007	1.002 - 1.013	0.009
Time with same trainer (months)	0.988	0.984 - 0.992	< 0.001
Track size (furlongs)	0.946	0.916 - 0.977	< 0.001
Wins/starts (30 - 60 days prior race)	1.002	1.001 - 1.003	< 0.001
Wins/starts (90 - 180 days prior race)	1.002	1.001 - 1.004	< 0.001

The final multivariable model included 21 risk factors with a statistically significant association with fatal injuries.

From the horse-related risk factors we have identified, the one with the potentially highest impact is related to horses that had at some point in their career entered the vet list. Those horses had 70.4% (95% CI: 58.7% - 83.0%) more chance of sustaining a fatal injury than horses that had never been on the vet list. The horse-related risk factor with the second highest potential impact was the sex of the horse. There was 45.4% (95% CI: 32.3% - 59.8%) more chance of sustaining a fatal injury for stallions compared to mares and geldings. The number of previous injuries a horse had in its career was also found to be associated with fatal injuries, with horses having a 29.2% (95% CI: 13.6% - 47.0%) higher chance of fatal injury for each previous EID-reported injury they had sustained. The age at which a horse begins its career was found to be associated with the risk of equine fatality. We found that for each year older the horse was at their first racing start there was a 9.9% (95% CI: 6.7% - 13.2%) higher chance of fatal injury for each subsequent start. Horses were also found to have a higher chance of sustaining a fatal injury by 0.7% (95% CI: 0.2% - 1.3%) for each extra month they spent in layup and for each extra month they spent, on average, between racing starts they had 2.5% (95% CI: 1.2% - 3.8%) higher chance of injury. Furthermore, when ranking the horses in a race according to their betting odds, horses less favoured by the odds were found to be less likely to sustain a fatal injury by 6.2% (95% CI: 4.9% - 7.4%) for each place further down the odds ranking. Finally, for each extra km horses had accumulated from racing starts in their career there was 1.1% (95% CI: 0.8% - 1.3%) less risk of sustaining a fatal injury.

Regarding risk factors exploring the number of previous races in which a horse has participated we found that for each racing or exercise start a horse had during the month prior the race there was 28.3% (95% CI: 25.5% - 30.9%) less chance of a fatal injury. The same association was also found for the period of 30 to 60 days prior to a race but with less potential impact: The risk of fatal injury was 5.8% (95% CI: 2.7% - 8.8%) less for each racing or exercise start in the period 30 to 60 days prior to a race. Conversely for each racing start in the period 90 to 180 days prior to a race horses had a 4.6% (95% CI: 2.1% - 7.3%) increased risk of fatal injury and additionally for each racing and exercise start an increased risk of 5.7% (95% CI: 4.2% - 7.2%). Furthermore, for the period of 30 to 60 days prior a race, for each extra percentage point in the percent of races won in that period horses were at 0.2% (95% CI: 0.1% - 0.3%) higher risk. The same observation can be made for the period 90 to 180 days prior the race with horses being at 0.2% (95% CI: 0.1% - 0.4%) higher risk for each extra percentage point in percentage of races won.

Regarding race-related risk factors, the one with the highest potential impact was the surface of the race. Horses were at 38.8% (95% CI: 21.8% - 58.1%) higher chance of fatal fatal injury when running on a track with dirt surface compared to a synthetic one and at 15.6% (95% CI: -1.4% - 35.3%) more risk when running on turf compered to synthetic surfaces, though this association was not statistically significant. The risk factor with the second highest potential impact was the country of the race, with horses participating in races in the US having a 29.5% (95% CI: 10.2% - 52.2%) higher chance of fatal injury than those competing in Canada. We also found that horses participating in races with a purse of less than $7,500 were 21.3% (95% CI: 13.2% - 28.6%) less likely to sustain a fatal injury than those competing in races in which the purse was greater than $7500. The distance of the race was also found to be associated with the risk of fatal injury with horses racing in longer races having 5.7% (95% CI: 3.1% - 8.2%) less risk per extra furlong of the race. Moreover, the track size was also found to be associated with fatal injuries with horses racing at longer tracks having 5.4% (95% CI: 2.3% - 8.4%) less risk per extra furlong of the track. The field size of the race was also found to be significant with 4.9% (95% CI: 3.0% - 6.8%) higher risk of fatal injury for each extra runner in the race.

Finally, we were able to identify one risk factor that was related to the trainer of the racing horse. For each extra month a horse spent with the same trainer, it had 1.2% (95% CI: 0.8% - 1.6%) less chance of sustaining a fatal injury.

3.3.2.4. Model Fit

The multivariable model had a deviance of 52,431 with 1,962,395 degrees of freedom. The x^2 test statistic of the Hosmer-Lemeshow goodness of fit test was 6.98

with 8 degrees of freedom and a p-value of 0.54 indicating no evidence of a lack of fit.

No numerical variables were assessed to be non-linear in the log odds.

No influential observation was found in the final multivariable model.

The area under the receiving operating characteristic curve was 66.5%.

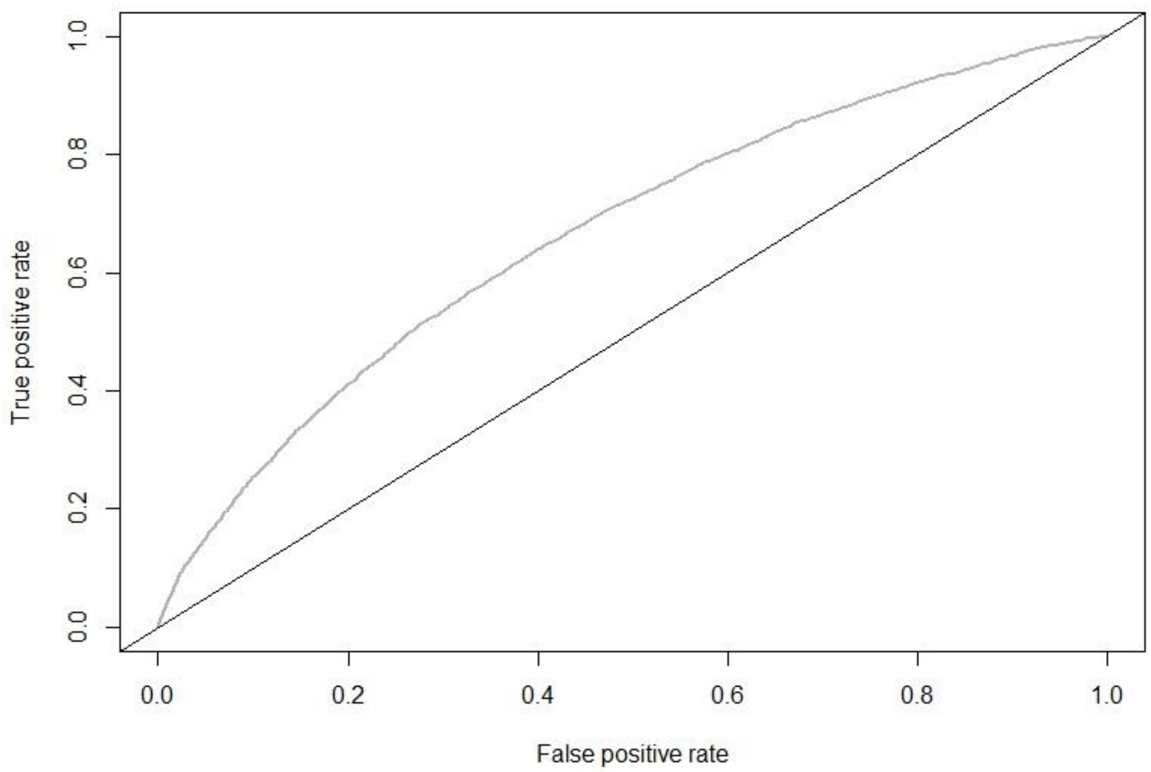

Figure 3-2 ROC curve of the multivariable logistic regression model trained on starts of Thoroughbred racehorses, six months after their first recorded racing or exercise start, from 2009 to 2015 for fatal injury prediction for the same period

Using the top 5% of fitted scores from our model we were able to identify the starts that resulted in 3.0 times higher risk of fatal injury than the average injury prevalence for that study period.

3.4. Discussion

3.4.1. Performance of the Models

The power of both models was above 80% for identifying a statistically significant association with an odds ratio of 1.2 at the 95% confidence level. The plethora of starts available for the study resulted in a high power to identify risk factors significantly associated with fatal injuries and this consequently, means that both models have very low Type II error.

The fit of the models was assessed by using the Hosmer-Lemeshow goodness of fit test and there were no indications that the models did not fit the data well.

Furthermore, the area under the receiving operating characteristic curve was calculated for both models. The model that used all starts had a score of 0.62 while the model for horses that had been racing for more than 6-months, that assessed racing history risk factors, was able to achieve a higher score of 0.65. This indicates that both models are able to identify starts at a higher risk of injury by using the statistically significant risk factors identified.

Another indication that the models are able to identify the starts at higher risk is the population of starts identified by the top 5% of fitted values from the models. The population identified from the model that used all starts had a prevalence of fatal injury that was 2.5 times higher than the average. The population of starts identified from the model for horses that had been racing for more than 6-months had a prevalence of fatal injury that was 3.0 times higher than the average.

3.4.2. Risk Factors

3.4.2.1. Risk Factors for all horses and for horses that had been racing for ≥ 6 months

3.4.2.1.1. Age at first start (years)

The age of the horse at the beginning of its racing career was found to be significantly associated with fatal injury. Horses had a higher risk of injury for each extra year of age at their first start.

Older age has frequently been reported as a risk factor and been associated with an increased risk of injury (Estberg, et al., 1995; Cohen, et al., 2000; Williams, et al., 2001; Henley, et al., 2006; Lyle, et al., 2012). This association might be due to owners being less willing to treat older horses that have reached the end of their racing career once they sustain a treatable injury. This might result in more aged horses being euthanised and therefore associated with higher risk of fatality as a result of an injury sustained during racing. Furthermore, age might be a proxy for higher levels of accumulated strain in a horse's bones and soft tissue resulting in higher risk of a catastrophic fracture during the race. In our study, we identified a higher risk of 3% per year older a Thoroughbred was when entering a race in the univariable level, yet in the multivariable level age was not included as a significant factor in the final model.

However, the age of a Thoroughbred at the beginning of their career was found to be significantly associated with a higher risk of fatal injury at the multivariable level. A plausible explanation for this is that horses that begin their career at a later age are introduced to the exercise regimen racing horses undertake also at later age. Therefore, their bones and soft tissue might be less well adapted to deal with the strain accumulated through the intense galloping during racing compared to horses that have participated in racing at a younger age. Bones model and remodel in order to reduce strain level (Frost, 1983; Frost, 1987; Lanyon & Baggott, 1976; Lanyon, 1982; Rubin & Lanyon, 1985) and most significant changes in bones occur between the ages of 1 year and 2 years old (Nunamaker, et al., 1990). It is also important to note that bone adapts and remodels continuously in older horses in response to exercise, so it is important that appropriate exercise regimens are adhered to throughout a horse's career.

It is also possible that the reason for an increased risk of fatal injury associated with starting racing as an older horse is related to some delay in the commencement of the horse's career, potentially due to (sub-clinical) injury. If so, then for such horses, it may be that it is the injury, or inherent susceptibility to injury, that is causally associated with the future risk of fatal injury. In these cases having a first race as an older horse may simply be a proxy measure of that previous (sub-clinical) injury or susceptibility to injury.

3.4.2.1.2. Country

Thoroughbreds had a higher risk of fatal injury when they entered a race in the US than when racing in Canada. It is unclear why starts made in the US have a higher risk of injury compared to starts in Canada. It is probable that this risk factor captures differences in regulations and practice between the two countries that have not been accounted for by more specific risk factors. This is an interesting finding that may be worth examining in further studies. However, it is also worth noting that in comparison to the number of starts made in the United States those made in Canada were very few, and were staged at a limited number of tracks. It may be that there are certain unique factors related to racing at those particular tracks that account for the difference identified.

3.4.2.1.3. Entered the vet list

A risk factor specific to the North American jurisdiction is horses that have previously entered the veterinarian's list. This is a list used by association and regulatory veterinarians to provide horses with illness, injury or soundness issues a brief respite from racing. Horses that at some point in their career had entered the vet list were found to have a higher risk of sustaining a fatal injury. It is possible that the reason for entering the vet list might be persisting even after the respite from racing therefore increasing the risk of injury (i.e. the underlying pathology remains for some horses).

It is worth commenting that the regulations surrounding the use of the vet list (including how and why horses are placed on the list and how horses get off the list) vary from state to state and have also changed in some states over the period covered by the EID data used in this study. In some states it is still currently possible for a horse to be on the vet list and unable to race in State A, travel to State B and race without being examined and officially taken off the vet list. Further harmonization and sharing of the vet list(s) between states would be of enormous benefit. Future analyses will also focus on regions where racing jurisdictions share their vet lists and where regulations with respect to the use of the vet list are consistent.

3.4.2.1.4. Low purse race (<= $7500)

Thoroughbreds racing in races with a purse lower or equal to $7500 had 19% less risk of sustaining a fatal injury. We hypothesise that the more competitive a race the more stress (i.e. to perform well) the horses were put under, resulting in a higher risk of injury. A study in Australia similarly showed a higher risk of breakdown associated with the more competitive stakes races (Bailey, et al., 1997).

3.4.2.1.5. No. of previous injuries

For each previous EID-reported injury a horse had sustained during a racing start, there was a higher risk of fatal injury indicating that previous pathology remains important. It is known that catastrophic fractures are often the result of pre-existing fractures (Stover, et al., 1992; Riggs, et al., 1999a; Riggs, 2002). It is not unreasonable to hypothesise that a previous injury might predispose horses to having a higher risk of future injuries that might result in a fatality. This result further emphasizes the need to record and share with regulators injury (and veterinary) data. It is understood that confidentiality will always be an issue when sharing such data. However, in order to make further progress in our ability to predict fatal injury, knowledge of the full veterinary history of horses at the time they enter a race or commence a new racing season is imperative. We would encourage greater efforts by all involved in the racing industry to overcome the justified concerns of trainers and owners and find a way to enable inclusion of full veterinary records in future studies.

3.4.2.1.6. Odds rank in race

Horses were at higher risk of fatal injury during racing the more favoured they were by the odds. We hypothesise that horses that are more likely to win are subjected to more stress during the race resulting in a higher risk (i.e. jockeys are more likely to ‘push or encourage’ horses that are favourites, than they are horses that are not expected to win). This may result in jockeys being less likely to pull-up a horse with a potential pre-fatal injury sign in such horses. Additionally, favoured horses are, on average, more likely to be competing in close finishes (i.e. genuinely racing to the end of the race) than less favoured horses who may be eased down, or even pulled-up, in the later part of races when no longer in contention.

3.4.2.1.7. Race distance (furlongs)

Race distance was associated with fatal injury with a decreased risk per extra furlong. Three studies in the United Kingdom; a case-control study (Parkin, et al., 2004b) of fatal fractures in the distal portion of the limbs of Thoroughbred racehorses; a retrospective study of fatal injuries (Henley, et al., 2006); and an observational study of sudden death (Lyle, et al., 2012) also indicated an association between the risk injury and race distance. However, unlike the present study, there was a positive association between risk of fatal injury and race distance. The same positive association was found in a study for Thoroughbred racehorse fatality in flat starts in Victoria, Australia (Boden, et al., 2007a). We hypothesise that the reason for the discrepancy in the nature of the association between risk of fatal injury and race distance between those studies and the present study is most likely due to differences in distance ranges for flat races in the United Kingdom and Australia, compared with the distance ranges for flat races in the United States and Canada. Thoroughbred flat races in the United States and Canada tend to be shorter than those in the United Kingdom and Australia. Consequently, races in the United States and Canada are run at a faster pace than races in the United Kingdom, and that fast pace likely contributed to the negative association between risk of fatal injury and race distance observed in this study. This possibly explains why a study in Kentucky, Peloso, et al., (1994) (Peloso, et al., 1994) found that the distance of the race was significantly shorter for horses with catastrophic injuries than for horses with non-catastrophic injuries in accordance with our findings. Together these findings indicate potentially how useful it would be to have true (sectional) speeds recorded and available for all horses in all races and not just the speed of the winning horse.

3.4.2.1.8. Sex

The odds of sustaining a fatal injury were higher for intact male horses compared to females and geldings. Stallions have been shown to be at a higher risk of injury in previous studies. A study in California showed higher risk of fracture to the forelimb proximal sesamoid bones (Anthenill, et al., 2007). Stallions have also been found to be at a higher risk for catastrophic musculoskeletal injury or any form of fatal injury (Estberg, et al., 1996a,b), as well as, non-fatal superficial digital flexor tendon injury (Takahashi, et al., 2004). Some have hypothesized that this association may be due to a more competitive nature associated with intact male horses due to higher levels of testosterone. As the current study is not the first to identify such an association it is probably worth investigating further. On the other hand no trainer or owner is going to castrate a potentially valuable intact male simply to moderately reduce his risk of fatal injury, so further investigation of modifiable risk factors may be more appropriate.

3.4.2.1.9. Surface

The odds of sustaining a fatal injury were higher for horses racing on turf surfaces and even higher for horses racing on dirt surfaces compared to horses racing on synthetic surfaces. Arthur (2010), in accordance with the findings of the current study, comparing fatalities at four California tracks reported a higher incidence of fatal injury on turf than synthetic surfaces and an even higher risk on dirt surfaces. Dirt courses were also found to be associated with a higher risk of breakdown compared to turf in a New York study (Mohammed, et al., 1991). We believe a possible explanation to be the increased forces that might be acting on the limb of the racing horse on dirt surfaces compared to turf and synthetic surfaces. It is also possible, as suggested by Parkin, et al., (2004) that there are inherently different populations of horses with inherently different levels of risk racing on different surfaces. Certainly, in the UK those horses that race on all weather surfaces are racing in lower class races and therefore are not rated as highly as those horses racing on turf. This may not be the same in North America where there is significantly less turf racing and more dirt racing, but the same potential applies: i.e. for there to be demographic/underlying differences in the racehorses that race on dirt, turf and synthetic surfaces.

3.4.2.1.10. Time between racing starts – average (months)

Increased odds of fatal injury were observed for horses that had longer average time between racing starts. An extended interval since the last race was found to increase the risk of catastrophic injury in two studies in Florida (Hernandez, et al., 2001; Hernandez, et al., 2005). It has also been shown that horses that had a preexisting injury might take longer intervals between races and were more likely to reduce their racing activity and that such horses are also likely to be at a higher risk of bone fracture (Stover, et al., 1992; Carrier, et al., 1998). Pre-existing pathology is associated with fractures (Stover, et al., 1992) and a possible explanation for our findings might be that horses with a pre-existing condition might reduce their racing activity and take a longer average time between racing starts. In other words, healthy horses continue to race (and train) more frequently. Without information about the reason for gaps between racing starts it is not possible to be certain that breaks are not part of a pre-determined racing schedule. As with the associations identified with lay-ups and previous injuries it is critical to future studies and the significant advancement of this area of work that more information relating to veterinary histories, including medications, be made available.

3.4.2.1.11. Time with same trainer (months)

For each month that a horse trains with the same trainer there is 1% less risk of sustaining a fatal injury. We hypothesise that this increased risk might be caused by possible abrupt changes in training regimen, particularly in the situations where trainers acquire horses from claiming races. It is also the case that for the most part when a horse moves to a new trainer significant details pertaining to the horse's veterinary record do not go with that horse. This in itself may put the horse at immediate increased risk as the trainer will be unaware of underlying pathology that could itself increase the risk of fatal injury.

3.4.2.1.12. Track size (furlongs)

The larger the size of the track the horses are racing on the lower the risk of a fatal injury. Boden, et al., (2007a) explored the association between the length of the circuit and equine fatalities and there was a trend indicating higher risk for longer circuits on the univariable level though it was not statistically significant. The finding in the current study may be associated with the fact that smaller tracks will have tighter turns, which without appropriate banking may place extra significant strains on the distal limb, which in turn may predispose to an increased risk of fatal injury.

3.4.2.2. Risk Factors for all horses

3.4.2.2.1. First Start

Thoroughbreds racing for the first time were significantly less likely to sustain a fatal injury. A possible explanation for this might be the reduced fatigue or boney micro-damage a horse has accumulated during the beginning of their career compared to the later stages of its career. This variable may also partly reflect some association with age at the start (which was not retained in this final multivariable model) as horses in their first start are obviously going to be younger (on average) than horses making subsequent starts. There is also some anecdotal evidence from discussion with trainers that horses entering their first race are not 'expected' to win and that jockeys may be encouraged to 'go easy' on such horses. In other words, although not always the case, some trainers treat a horse's first start as a race in which they are familiarizing the horse with the events that surround a race day. If this is the case it may therefore be the fact that horses are not pushed or do not put as much effort into their first race, which, in itself, may reduce the risk of fatal injury.

3.4.2.2.2. No. of layups

Horses had 5% less chance of sustaining a fatal injury for each layup they had in their career. We hypothesise that taking a brief respite from racing might be beneficial to the horse. Furthermore, a possible explanation might be that horses had a layup because they needed rest from a situation threatening to their health that had it not been identified but might have resulted in an increased risk of fatality had they continued to race (Carrier, et al., 1998). It should be noted that the reason for the layup was not recorded in the data available for this study, so it was not possible to differentiate between layups that were the result of injury and layups were simply decided upon by the trainer to be required or part of a normal training regimen. Further details on the exact reasons for layups would obviously be useful and would likely improve the predictive ability of future models.

3.4.2.2.3. Post position

Horses had a 2% greater risk the further outside they were positioned at the start of race. A possible explanation might be the extra effort those horses expend to reach the inside rail during a race. However, given that most races are raced on an oval, one would have expected the racing authorities to have introduced regulations to combat any post position disadvantage. This is assuming of course that one would see a disadvantage in terms of the likelihood of winning the race that was associated with post position, in addition to the association identified here.

3.4.2.2.4. Season

This study showed that the risk for sustaining a fatal injury was less for horses that raced during summer and spring compared to autumn. Lyle, et al., (2012) found a decreased risk of sudden death for horses racing during summer compared to all other seasons with a p value of 0.10 in the univariable model. This association, however, switched direction in the multivariable model showing that horses had a higher risk of sudden death when racing in summer. Furthermore, racing in the spring has been associated with increased risk of epistaxis (Williams, et al., 2008), which on rare occasions has been identified as the primary cause of sudden death (Lyle, et al., 2011). We hypothesise that the identified seasonal associations with injury might be due to training and racing schedules and not due to causal biological reasons, such as a result of a change in weather, brought by the different seasons.

3.4.2.2.5. Training with first trainer

If a horse is not training with its first trainer there is an increased risk of 9% of sustaining a fatal injury. We hypothesise that some horses might change trainers after some minor non-recorded (in the EID at least) injury and this might lead to an increase risk of sustaining a fatal injury. This finding is very closely related to that above for time with the same trainer and could represent similar underlying factors, such as familiarity with a horses sub-clinical pathology and veterinary history, that in this case reduce the risk of fatal injury if a horse is still with its first trainer.

3.4.2.3. Risk Factors for horses that had been racing for ≥ 6 months

3.4.2.3.1. Accumulated racing distance ran in career (Km)

Horses were 2% less likely to sustain a fatal injury per km they had raced in racing starts over their career. This finding might be due to healthy horses participating more in races and therefore having accumulated more racing distance in their career. In other words this association most likely represents the ‘healthy horse’ effect where healthy horses are able to continue to race and therefore do so and healthy horses are also inherently less likely to sustain a fatal injury.

3.4.2.3.2. Field size

Field size, the number of horses participating in a race, was found to be associated with fatal injury with horsed being at 4.9% higher risk for each extra horse participating in the race. Others have hypothesized that larger field size may be associated with more competitive racing which itself may be associated with injury risk (Parkin, et al., 2004b).

3.4.2.3.3. No. of racing and exercise starts (Present to 30 days prior race - 30 to 60 days prior race - 90 to 180 days prior race) - racing starts (90 to 180 days prior race))

Horses were at significantly less risk of sustaining a fatal injury for each racing or exercise start they had in the 30-day period prior to the race. Horses were at less risk of sustaining a fatal injury for each racing or exercise start they had in the period 30 to 60 days prior to the race. Horses were at a 5% higher risk of sustaining a fatal injury for each racing or exercise start they had in the period 90 to 180 days prior to the race and at a 5% higher risk of sustaining a fatal injury for each racing start they had in the period 90 to 180 days prior to the race.

A lot of studies have looked at prior racing history of Thoroughbreds to identify risk factors. Lyle, et al., (2012) found that the more starts a horse had within the last 60 days the lower the risk of sudden death. 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., (2007a), 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 made a start.

We seem to identify a similar trend where horses have a lower risk if they have participated in a race within a period 0 to 30 days and in our study, in the in the period 30 to 60 days prior to the race. However, the association changes direction when looking at 90 to 180 days prior to the race with a higher risk of sustaining a fatal injury for each racing start they had in that period. We believe that Boden, et al., (2007a) were able to identify this switch in direction for an earlier period in their study. Furthermore, the findings of this study are in agreement with two studies that showed a higher risk for catastrophic injury if there was an extended interval since the last race (Hernandez et al., 2001; Hernandez et al., 2005).

3.4.2.3.4. Time in layup (months)

The more time a horse has spent in layup the higher the risk was for sustaining a fatal injury. As an extensive amount of absence from racing might be due to some sort of pathology we hypothesise this is the most likely reason for the identified increased risk. Furthermore, during this layup period, a different kind of exercise regimen might be in place for the horse, that might include no galloping, which was identified as a risk factor for fractures of the distal limb in a previous study (Parkin, et al., 2005).

3.4.2.3.5. Wins/starts (30 – 60 days and 90 – 180 days prior race)

Similarly to the horses favoured by the odds to win the race we found that horses had a slightly higher risk of injury with the more races they had won per racing start in the periods 30-60 and 90–180 days prior the race. Again this may be due to such horses being more competitive toward the end of races, thus increasing their exposure to high risk, high speed parts of races.

3.4.3. Limitations of the Study and Future Analyses

We believe the identified risk factors are as unbiased as possible and representative of racing in North America, since we have included in the statistical analysis 90% of racing starts from all official racing in the US and Canada for that period. A small source of bias could be the roughly 10% of starts which are not included in this study. However, this small percentage of starts would have to be quite radically different to those starts included in the database for them to have any significant effect on the 'national' models produced in this study. We do not know if this is the case as we did not have access to demographic data for this 10% of starts but we do believe this to be highly unlikely. This does not mean that we should not further pursue inclusion of these tracks in future years of the EID or indeed examine characteristics of the racing populations at these tracks to assess whether they are likely to be different from the rest of North American racing in any significant or material way.

It is unclear why starts made in the US have a higher risk rate compared to starts in Canada. This is an interesting finding that requires further examination. The higher risk might be due to different training regimens and racing schedules followed by horses in the two countries.

The study looked at the exercise history of each horse and used the number of exercise 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 horses training regimen is associated with the risk of injury during racing.

It is important to note that we did not make any attempt to differentiate the causes of fatal injury in the present study. Risk factors vary among types of fractures and it is likely that some of those risk factors were not identified in the present study. The types of injuries sustained and the reason for euthanasia have been accurately reported to the EID only recently. Thus, future analyses will be able to use more specific outcome variables to identify risk factors associated with the most common reasons for euthanasia of Thoroughbred racehorses following race-induced injuries.

Furthermore, statistical significance does not necessarily translate to clinical significance. Although we identified several risk factors that were significantly associated with fatal injuries in Thoroughbred horses competing in flat racing, it is important to point out that the vast majority of race starts evaluated in the present study did not result in a fatal injury. Finally, because of the extremely large number of race starts evaluated and the resulting high statistical power of this study, the magnitude of effect for some of the risk factors was very small.

3.4.4. Recommendations

The results of the present study can be used as a guideline for the identification of racehorses at high risk of sustaining a fatal injury during a race. The risk factors identified should be considered in the selection and implementation of measures expected to have the greatest effect on minimising the number of horses that sustain fatal injuries during flat races in the United States and Canada. Priority should be given to the consideration of methods to mitigate the effect of potentially modifiable risk factors with both the highest odds ratios and prevalence in the racing population in North America.

4. Risk Factors for Fractures

4.1. Introduction

This part of the study is based on equine fractures in flat horse racing of Thoroughbreds in the US and Canada during the period 2009–2015. Fractures sustained during racing account for 83% of equine fatalities in this 7-year period, as 75% of them resulted in a fatality. 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., 1996b; Estberg, et al., 1998a,b; Williams, et al., 2001; Parkin, et al., 2005) has been shown to be a significant 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., 1996b; 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). 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). 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 (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 horseshoe characteristics have been identified to be associated with equine injuries (Kane, et al., 1996; Kane, et al., 1998) . Finally, there seem to be risk factors directly related to the racecourse. 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), the field size (Parkin, et al., 2004b; Parkin, et al., 2005; Lyle, et al., 2012) and the type of the race (Estberg, et al., 1998a,b). These studies provided a starting point for the analysis of our study. We aim to identify the risk factors associated with fatal and non-fatal equine fractures in the US 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.

4.2. Materials and Methods

4.2.1. Study Design

The analysis reported in this paper is an observational retrospective cohort study based on racecourses reporting injuries to the EID from 1st January 2009 to 31st December 2015. 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. These data include 2,493,957 race starts from the 89 race tracks reporting injuries to the EID and 4,592,162 exercise starts. The race starts reported to the EID represented the starts for 90.0% of all official Thoroughbred racing events in the United States and Canada during the 7-year observation period.

4.2.2. Case definition

The study was conducted with race start as the unit of analysis as equine fracture injuries are reported at the start level and this approach allowed analysis of all start level risk factors of interest. Cases were defined as starts from horses that sustained a fracture injury during a race. This definition includes every possible type of fracture that a Thoroughbred might have sustained. Fractures included were both fatal and non-fatal. All other race starts from race tracks reporting injuries to the EID were classified as controls. Exercise starts were only used to quantify prior exercise history for each horse entering a race.

4.2.3. Risk Factors

A total of 51 potential risk factors were identified from previous studies and from *a priori* hypotheses and were considered in our analysis. These included horse-related potential risk factors and race-related risk factors. The EID database also contained information for approximately 11,000 anonymized trainers and 3,000 anonymized jockeys associated with the recorded races, which enabled us to analyze trainer- and jockey-related risk factors.

4.2.4. Statistical Analysis

Initially, the linear relationship between numerical potential risk factors was assessed by examination of graphical plots of the log odds of the potential risk factors and fracture injuries (Boden, et al., 2007a; Reardon, 2013). If the relationship could be considered non-linear in the log scale, we created categorical alternatives for the risk factors. Binary and polytomous (5-level) categorical terms were considered and the form of the variable that produced the best fit in a univariable model based on the Akaike information criterion was retained. For parsimony and to facilitate interpretability of the results we did not consider polynomial terms or interaction terms. Finally, when examining the potential risk factor of the purse of the race a categorical variable was introduced *ad hoc* to specifically explore low purse races.

The association between each potential risk factor and fracture injury was assessed by creating a univariable regression model. Wald P-values were calculated and risk factors with values of $P < 0.20$ in univariable analysis were eligible for inclusion in a multivariable logistic regression model. A threshold of $P < 0.20$ was chosen to prevent exclusion of a potentially significant risk factor that only becomes evident when a confounder has been controlled for in a multivariable analysis (Dohoo, et al., 2003). An automated stepwise selection process was used to build the multivariable model. Potential risk factors were identified by use of a forward bidirectional elimination approach and assessment of the AIC. We preferred a forward stepwise approach compared to a backwards stepwise approach that usually results in more variables retained in the final model. As we have a plethora of available data, and consequently high statistical power, we are confident that a variable that contributes information to the final model would not be excluded. The AICs for competing models were compared, and the model with the lowest AIC was preferred (Bozdogan, 1987). Only risk factors with a statistical significance indicated by a Wald P value of less than 0.05 were retained in the final models.

To assess risk factors that summarize historical racing information prior to each race start, we followed the same analytical procedure to arrive at a second multivariable model on a sub-sample of the population consisting of all the starts from horses at least six months after their first recorded racing or exercise start. Using this sample of the data we were able to assess the relationship between the number of starts a horse had up to six months prior to the race and fracture injuries. This facilitates the interpretation of the relationship as it excludes horses that would have fewer or no starts due to having only recently started racing.

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. However, for the risk factors included in the final models we checked for possible collinearity (Bagley et al 2001); correlation coefficients were produced for all pairs with a threshold for inclusion set at 0.7.

Furthermore, we evaluated potential confounders by resubmitting the variables that were excluded from the final model during the stepwise selection process (Reardon, 2013; Boden, et al., 2007a). If the potentially confounding variable altered odds ratios for variables in the final model by >20% (Dohoo, et al., 2003), we retained the confounder in the final model.

Furthermore, the potential effect of horse in the data analyses was evaluated by creating a mixed-effects model that included horse as a random effect (Reardon, 2013; Boden, et al., 2007a; Lyle, et al., 2012). Results were nearly identical (less than 10% change in ORs and no meaningful changes in P values) to results obtained with models that did not include random effects so the single level fixed models were retained. We did not further check mixed models with racecourse, jockey or trainer as random effects. There is little indication in the literature that there are meaningful changes when compared to the single level fixed models and we do not think that that would be the case our study especially given the thousand different jockeys and trainers involved.

Model fit was assessed by using the Hosmer-Lemeshow goodness of fit test (Hosmer & Lemeshow, 2000; Dohoo, et al., 2003). Furthermore, we checked for the existence of influential observations in our final models (Cook & Weisberg 1982; Williams, 1987).

The predictive ability of the models was assessed by calculating the area under the receiver operating characteristic curve (Altman, et al., 2000; Bozdogan, 1987). Furthermore, the top 5% of fitted scores from the models were used to assess the ability of the models to identify a population of starts with higher prevalence of fracture injury than the average.

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

4.3. Results

4.3.1. Results for all horses

4.3.1.1. Study Population

The study population comprised all 188,269 Thoroughbred horses that participated in flat racing in the US and Canada from 1st January 2009 to 31st December 2015 in the 89 race tracks reporting injuries to the EID. The prevalence of fracture injury was 0.20% for the 2,493,957 racing starts in the 7-year study period.

4.3.1.2. Univariable Models

In total 33 possible risk factors were screened using univariable analysis (table 4-1); 26 of them were found to have a statistically significant association with fracture injuries. Of the possible risk factors 27 were found to have a P-value of less than 0.20 and were included in the subsequent forward bidirectional elimination to be potentially included in the final multivariable model.

Table 4-1 Results of univariable logistic regression for assessment of risk factors associated with fracture injuries in Thoroughbred racehorses competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Controls - Cases	OR (95% CI)	P value
Age (years)	2,488,984 - 4,973	0.963 (0.945-0.981)	0.007
Age at first start (years)	2,488,984 - 4,973	1.012 (0.988-1.037)	0.317
Country			
Canada	200,440 - 250	Ref	Ref
US	2,288,544 - 4,723	1.586 (1.402-1.794)	< 0.001

Table 4-1 (Continued)

Risk factor	Controls - Cases	OR (95% CI)	P value
Entered the vet list			
No	2,012,424 - 3,590	Ref	Ref
Yes	476,560 - 1,383	1.627 (1.529-1.731)	< 0.001
Field size	2,488,984 - 4,973	0.995 (0.981-1.009)	0.469
First Start			
No	2,310,330 - 4,700	Ref	Ref
Yes	178,654 - 273	0.751 (0.665-0.849)	< 0.001
Low purse race (<= $7500)			
No	2,126,103 - 4,372	Ref	Ref
Yes	362,881 - 601	0.805 (0.74 - 0.877)	< 0.001
Months since last racing start	2,488,984 - 4,973	0.984 (0.969-.0999)	0.035
Months since last racing or exercise start	2,488,984 - 4,973	1.081 (1.062-1.099)	< 0.001
No. of layups	2,488,984 - 4,973	0.899 (0.876-0.922)	< 0.001
No. of previous injuries	2,488,984 - 4,973	1.427 (1.265-1.610)	< 0.001
No. of previous vet scratches	2,488,984 - 4,973	1.084 (1.048-1.120)	0.018
No. of previous non-vet scratches	2,488,984 - 4,973	1.019 (1.004-1.033)	0.011
Odds at start of race	2,488,984 - 4,973	0.993 (0.992-0.995)	< 0.001
Odds rank in race	2,488,984 - 4,973	0.934 (0.924-0.945)	< 0.001
Post position	2,488,984 - 4,973	0.998 (0.986-1.008)	0.667
Purse ($1000)	2,488,984 - 4,973	1.000 (1.000-1.001)	0.411
Race distance (furlongs)	2,488,984 - 4,973	0.950 (0.931-0.970)	< 0.001
Season			
Autumn	646,405 - 1,331	Ref	Ref
Spring	610,481 - 1,177	0.936 (0.866-1.013)	0.100
Summer	775,483 - 1,442	0.903 (0.838-0.973)	0.007
Winter	456,615 - 1,023	1.088 (1.003-1.181)	0.043
Sex			
Mare/Gelding	2,179,424 - 4,102	Ref	Ref
Stallion	309,560 - 871	1.495 (1.389-1.608)	< 0.001

Table 4-1 (Continued)

Risk factor	Controls - Cases	OR (95% CI)	P value
Start with new jockey			
No	1,195,411 - 2,299	Ref	Ref
Yes	1,293,573 - 2,674	1.075 (1.017 - 1.137)	0.011
Start with new trainer			
No	2,250,162 - 4,364	Ref	Ref
Yes	238,822 - 609	1.315 (1.208 - 1.431)	< 0.001
Surface			
Synthetic	297,145 - 429	Ref	Ref
Dirt	1,837,708 - 3,869	1.458 (1.320 - 1.611)	< 0.001
Turf	354,131 - 675	1.320 (1.170 - 1.490)	< 0.001
Time between exercise starts – avg (months)	2,488,984 – 4,973	0.966 (0.945-0.988)	0.002
Time between exercise starts - active – avg (months)	2,488,984 – 4,973	0.984 (0.957-1.012)	0.254
Time between racing starts - avg (months)	2,488,984 – 4,973	1.025 (1.014 -1.034)	< 0.001
Time between racing starts - active - avg (months)	2,488,984 – 4,973	1.077 (1.057-1.097)	< 0.001
Time in racing – active (months)	2,488,984 – 4,973	0.997 (0.994-1.000)	0.041
Time in racing (months)	2,488,984 – 4,973	0.995 (0.993-0.997)	< 0.001
Time with same jockey (months)	2,488,984 – 4,973	0.991 (0.978-1.004)	0.184
Time with same trainer (months)	2,488,984 – 4,973	0.981 (0.977-0.984)	< 0.001
Track size (furlongs)	2,488,984 – 4,973	1.013 (0.987-1.039)	0.323
Training with first trainer			
Yes	1,332,867 - 2,492	Ref	Ref
No	1,156,117 - 2,481	1.148 (1.086 - 1.213)	< 0.001

4.3.1.3. Multivariable Model

The 27 possible risk factors with a P-value of less than 0.20 in the univariable analysis were included in a forward bidirectional elimination using AIC to assess the models created. Following this procedure, we arrived at the final multivariable model (Table 4-2).

Table 4-2 Results of multivariable logistic regression for assessment of risk factors associated with fracture injuries in Thoroughbred racehorses competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Odds ratio	95% CI	P-value
(Intercept)	0.002	0.002 - 0.003	< 0.001
Country			
Canada	Ref	Ref	Ref
US	1.310	1.154 - 1.788	< 0.001
Entered the vet list			
No	Ref	Ref	Ref
Yes	1.645	1.543 - 1.754	< 0.001
First Start			
No	Ref	Ref	Ref
Yes	0.789	0.694 - 0.897	< 0.001
Low purse race (<= $7500)			
No	Ref	Ref	Ref
Yes	0.790	0.724 - 0.862	< 0.001
No. of layups	0.832	0.790 – 0.876	< 0.001
No. of previous injuries	1.279	1.129 - 1.449	< 0.001
No. of previous non-vet scratches	1.025	1.009 - 1.041	0.002
Odds rank in race	0.942	0.932 - 0.952	< 0.001
Race distance (furlongs)	0.940	0.919 - 0.961	< 0.001
Season			
Autumn	Ref	Ref	Ref
Spring	0.922	0.853 - 0.997	0.041
Summer	0.936	0.869 - 1.008	0.079
Winter	1.063	0.980 - 1.154	0.139

Table 4-2 (Continued)

Risk factor	Odds ratio	95% CI	P-value
Sex			
Mare/Gelding	Ref	Ref	Ref
Stallion	1.449	1.346 - 1.560	< 0.001
Surface			
Synthetic	Ref	Ref	Ref
Dirt	1.373	1.242 - 1.518	< 0.001
Turf	1.293	1.143 - 1.462	< 0.001
Time between racing starts – active - avg (months)	1.050	1.029 - 1.072	< 0.001
Time in layup (months)	1.016	1.008 - 1.025	< 0.001
Time with same trainer (months)	0.985	0.980 - 0.989	< 0.001
Training with first trainer			
Yes	Ref	Ref	Ref
No	1.099	1.005 - 1.203	0.039

The final multivariable model included 16 risk factors with a statistically significant association with fracture injuries.

From the horse-related risk factors we have identified the one with the highest potential impact is related to horses who had at some point in their career entered the vet list. Those horses had 64.5% (95% CI: 54.3% - 75.4%) more chance of sustaining a fracture injury than horses that had never been on the vet list. The horse-related risk factor with the second highest potential impact was the sex of the horse. There was 44.9% (95% CI: 34.6% - 56.0%) more chance of sustaining a fracture injury for stallions compared to mares and geldings. Conversely, horses were 21.1% (95% CI: 30.6% - 20.3%) less likely to sustain a fracture injury during their first racing start, compared with all subsequent starts. The number of previous EID-reported injuries a horse had in its career was also found to be associated with fracture injuries, with horses having a 27.9% (95% CI: 12.9% - 44.9%) higher chance of fracture injury for each previous injury they had sustained. Horses were also found to have a lower chance of sustaining a fracture injury by 16.7% (95% CI: 12.4% - 21.0%) for each 60-

day layup they had in their career though for each extra month in layup they had 1.6% (95% CI: 0.8% - 2.5%) higher chance of injury and for each extra month time they had on average between racing starts they had 5.0% (95% CI: 2.9% - 7.2%) higher chance of injury. Furthermore, when ranking the horses of a race according to their betting odds, horses less favoured by the odds were found to be less likely to sustain a fracture injury by 5.8% (95% CI: 4.8% - 6.8%)%) for each place further down the odds ranking. Finally, for each time a horse was scratched in its career not by a track veterinarian there was a 2.5% (95% CI: 0.9% - 4.1%) increased risk of fracture injury.

Regarding race-related risk factors, the one with the potentially highest impact is the country of the race, with horses participating in races in the US having a 31.0% (95% CI: 15.4% - 78.8%) higher chance of fracture injury compared to horses competing in Canada. The risk factor with the second potential highest impact was the surface of the race. Horses were at 37.3% (95% CI: 24.2% - 51.8%) higher chance of fracture injury when running on a track with dirt surface compared to a synthetic one and at 29.3% (95% CI: 14.3% - 46.2%) more risk when running on turf compered to synthetic surfaces. We, also found that horses participating in races with a purse of less or equal to $7,500 were 21.0% (95% CI: 13.8% - 27.6%) less likely to sustain a fracture injury compared with horses competing in races with a purse greater than $7,500. Horses were also at less risk when racing in the spring, by 7.8% (95% CI: 0.3% - 14.7%) compared to when racing in autumn. The distance of the race was also found to be associated with fracture injuries with horses racing in longer races having 6.0% (95% CI: 3.9% - 8.1%) less risk per extra furlong of the race.

Finally, we were able to identify risk factors that were related to the trainer of the racing horse. For each extra month a horse spends with the same trainer, it has 1.5% (95% CI: 1.1% - 2.0%) less chance of sustaining a fracture injury. Furthermore, horses that were not longer training with their first trainer had 9.9% (95% CI: 0.5% - 20.3%) higher chance of sustaining a fracture injury.

4.3.1.4. Model Fit

The multivariable model had a deviance of 72,484 with 2,493,937 degrees of freedom. The x^2 test statistic of the Hosmer-Lemeshow goodness of fit test was 9.370 with 8 degrees of freedom and an p-value of 0.31 indicating a good fit.

No numerical variables were assessed to be non-linear in the log odds.

No influential observation was found in the final multivariable model.

The area under the receiving operating characteristic curve was 61.7%.

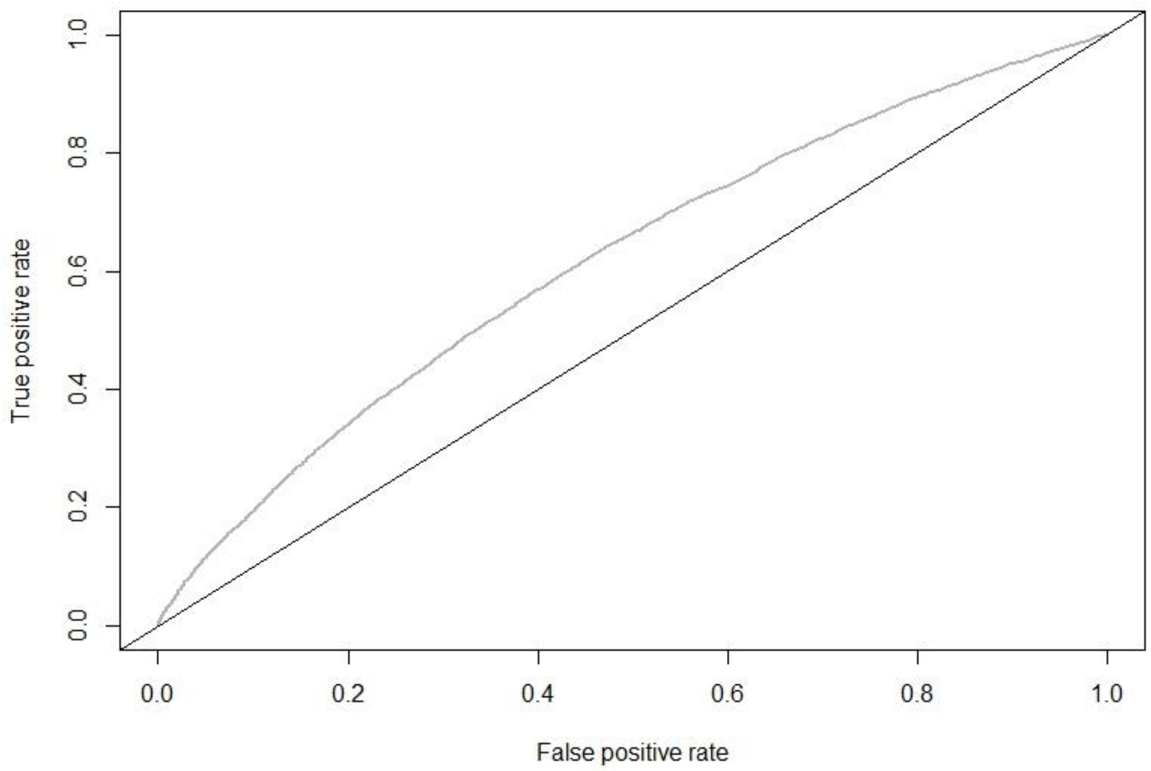

Figure 4-1 ROC curve of the multivariable logistic regression model trained on starts from 2009 to 2015 for fracture injury prediction for the same period

Using the top 5% of fitted scores from our model we were able to identify the starts that resulted in a 2.3 times higher risk of fracture injury than the average injury prevalence for that study period.

4.3.2. Results for horses that had been racing for ≥ 6 months.

4.3.2.1. Study Population

The study population comprised of 151,820 Thoroughbred horses that participated in flat racing in the US and Canada from 1st January 2009 to 31st December 2015 in the 89 race tracks reporting injuries to the EID and that had already had a racing or exercise start more than 6 months in the past. The prevalence of fracture injury was 0.21% for the 1,962,418 racing starts in the 7-year study period.

4.3.2.2. Univariable Models

In total 51 possible risk factors were screened using univariable analysis. 41 of them were found to have a statistically significant association with fracture injuries. 45 of the possible risk factors were found to have a P-value of less than 0.20 and were included in the subsequent forward bidirectional elimination to be potentially included in the final multivariable model.

Table 4-3 Results of univariable logistic regression for assessment of risk factors associated with fracture injuries in Thoroughbred racehorses, six months after their first recorded racing or exercise start, competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Controls - Cases	OR (95% CI)	P-value
Accumulated distance ran in career (Km)	1,958,464 – 3,954	0.996 (0.995-0.997)	< 0.001
Accumulated exercise distance ran in career (Km)	1,958,464 – 3,954	0.998 (0.996-1.001)	0.239
Accumulated racing distance ran in career (Km)	1,958,464 – 3,954	0.992 (0.990-0.994)	< 0.001
Age (years)	1,958,464 – 3,954	0.949 (0.929-0.970)	< 0.001
Age at first start (years)	1,958,464 – 3,954	1.027 (0.997-1.058)	0.074
Average speed change on previous race (m/s)	1,958,464 – 3,954	1.005 (0.988-1.024)	0.529
Average speed in previous race (m/s)	1,958,464 – 3,954	1.030 (1.015-1.046)	< 0.001
Country			
Canada	153,161 - 186	Ref	Ref
US	1,326,249 - 3,767	1.719 (1.483-1.991)	< 0.001
Entered the vet list			
No	1,524,821 - 2,716	Ref	Ref
Yes	433,643 - 1,238	1.603 (1.498-1.714)	< 0.001
Field size	1,958,722 - 3,696	0.999 (0.984-1.016)	0.950
Low purse race (<= $7500)			
No	1,661,272 - 3.485	Ref	Ref
Yes	297,192 - 469	0.752 (0.683 - 0.828)	< 0.001
Months since last racing start	1,958,464 – 3,954	0.977 (0.961-0.993)	0.004
Months since last racing or exercise start	1,958,464 – 3,954	1.074 (1.054-1.095)	< 0.001
No. of layups	1,958,464 – 3,954	0.870 (0.845-0.895)	< 0.001
No. of previous injuries	1,958,464 – 3,954	1.419 (1.255-1.604)	< 0.001
No. of previous vet scratches	1,958,464 – 3,954	1.036 (1.001-1.072)	0.046
No. of previous non-vet scratches	1,958,464 – 3,954	1.016 (1.001-1.032)	0.036

Table 4-3 (Continued)

Risk factor	Controls - Cases	OR (95% CI)	P-value
No. of racing and exercise starts (Present – 30 days prior race)	1,958,464 – 3,954	0.732 (0.708-0.756)	< 0.001
No. of racing and exercise starts (30 -60 days prior race)	1,958,464 – 3,954	0.925 (0.900-0.951)	< 0.001
No. of racing and exercise starts (60 -90 days prior race)	1,958,464 – 3,954	1.004 (1.023-1.075)	< 0.001
No. of racing and exercise starts (90 -180 days prior race)	1,958,464 – 3,954	1.070 (1.059-1.080)	< 0.001
No. of starts (Present – 30 days prior race)	1,958,464 – 3,954	0.837 (0.799-0.876)	< 0.001
No. of starts (30 – 60 days prior race)	1,958,464 – 3,954	1.006 (1.018-1.100)	0.004
No. of starts (60 – 90 days prior race)	1,958,464 – 3,954	1.141 (1.099-1.184)	< 0.001
No. of starts (90 – 180 days prior race)	1,958,464 – 3,954	1.084 (1.065-1.102)	< 0.001
Odds at start of race	1,958,464 – 3,954	0.992 (0.990-0.994)	< 0.001
Odds rank in race	1,958,464 – 3,954	0.928 (0.917-0.940)	< 0.001
Post position	1,958,464 – 3,954	0.997 (0.985-1.008)	0.572
Purse ($1000)	1,958,464 – 3,954	1.000 (0.999-1.001)	0.540
Race distance (furlongs)	1,958,464 – 3,954	0.931 (0.909-0.953)	< 0.001
Season			
Autumn	526,367 - 1,093	Ref	Ref
Spring	468,582 - 910	0.935 (0.856-1.021)	0.136
Summer	604,336 - 1,137	0.906 (0.834-0.985)	0.020
Winter	359,179 - 814	1.091 (0.997-1.195)	0.059
Sex			
Mare/Gelding	1,746,498 - 3,313	Ref	Ref
Stallion	211,966 - 641	1.594 (1.465-1.735)	< 0.001
Start with new jockey			
No	871,735 - 1,701	Ref	Ref
Yes	1,086,729 - 2,253	1.062 (0.998-1.132)	0.059

Table 4-3 (Continued)

Risk factor	Controls - Cases	OR (95% CI)	P-value
Start with new trainer			
No	1,748,612 - 3,411	Ref	Ref
Yes	209,852 - 543	1.326 (1.211-1.452)	< 0.001
Surface			
Synthetic	221,336 - 308	Ref	Ref
Dirt	1,447,429 - 2,291	1.544 (1.374-1.736)	< 0.001
Turf	289,669 - 536	1.330 (1.156-1.530)	< 0.001
Time between exercise starts – avg (months)	1,958,464 – 3,954	0.953 (0.930 -0.977)	0.926
Time between exercise starts - active - avg (months)	1,958,464 – 3,954	0.973 (0.944-1.045)	0.093
Time between racing starts – avg (months)	1,958,464 – 3,954	1.023 (1.012-1.034)	< 0.001
Time between racing starts - active - avg (months)	1,958,464 – 3,954	1.080 (1.058 -1.104)	< 0.001
Time in layup (months)	1,958,464 – 3,954	0.983 (0.978-0.988)	< 0.001
Time in racing – active (months)	1,958,464 – 3,954	0.995 (0.992-0.998)	0.003
Time in racing (months)	1,958,464 – 3,954	0.993 (0.991-0.996)	< 0.001
Time with same jockey (months)	1,958,464 – 3,954	0.989 (0.976-1.003)	0.123
Time with same trainer (months)	1,958,464 – 3,954	0.978 (0.974-0.981)	< 0.001
Track size (furlongs)	1,958,464 – 3,954	1.020 (0.991-1.050)	0.177
Training with first trainer			
Yes	535,212 - 1,578	Ref	Ref
No	1,109,815 - 2,376	1.151 (1.080-1.227)	< 0.001
Wins/starts (Present – 30 days prior race)	1,958,464 – 3,954	1.001 (1.000-1.002)	0.014
Wins/starts (30 – 60 days prior race)	1,958,464 – 3,954	1.004 (1.003-1.005)	< 0.001
Wins/starts (60 – 90 days prior race)	1,958,464 – 3,954	1.003 (1.002-1.004)	< 0.001
Wins/starts (90 – 180 days prior race)	1,958,464 – 3,954	1.006 (1.004-1.007)	< 0.001

4.3.2.3. Multivariable Model

The 45 possible risk factors with a P-value of less than 0.20 in the univariable analysis were included in a forward bidirectional elimination using AIC to assess the models created. Following this procedure, we arrived at the final multivariable model (Table 4-4).

Table 4-4 Results of multivariable logistic regression for assessment of risk factors associated with fracture injuries in Thoroughbred racehorses, six months after their first recorded racing or exercise start, competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Odds ratio	95% CI	P-value
(Intercept)	0.003	0.002 - 0.004	< 0.001
Accumulated distance ran in career (Km)	0.989	0.985 - 0.992	< 0.001
Age at first start (years)	1.060	1.028 - 1.094	< 0.001
Country			
Canada	Ref	Ref	Ref
US	1.195	1.028 - 1.388	0.020
Entered the vet list			
No	Ref	Ref	Ref
Yes	1.575	1.469 - 1.689	< 0.001
Low purse race (<= $7500)			
No	Ref	Ref	Ref
Yes	0.761	0.690 - 0.841	< 0.001
No. of layups	0.892	0.852 - 0.934	< 0.001
No. of previous injuries	1.288	1.134 - 1.463	< 0.001
No. of previous non-vet scratches	1.027	1.008 - 1.045	0.004
No. of racing and exercise starts (Present – 30 days prior race)	0.699	0.674 - 0.726	< 0.001
No. of racing and exercise starts (60 - 90 days prior race)	1.055	1.025 - 1.087	0.001
No. of racing and exercise starts (90 - 180 days prior race)	1.083	1.069 - 1.096	< 0.001

Table 4-4 (Continued)

Risk factor	Odds ratio	95% CI	P-value
No. of racing starts (30 – 60 days prior race)	0.915	0.875 - 0.957	< 0.001
Odds rank in race	0.943	0.931 - 0.955	< 0.001
Race distance (furlongs)	0.949	0.925 - 0.974	0.004
Sex			
Mare/Gelding	Ref	Ref	Ref
Stallion	1.585	1.454 - 1.727	< 0.001
Surface			
Synthetic	Ref	Ref	Ref
Dirt	1.352	1.199 - 1.525	< 0.001
Turf	1.253	1.086 - 1.446	0.002
Time between exercise starts – avg (months)	0.953	0.925 – 0.981	0.001
Time between racing starts – avg (months)	1.024	1.012 - 1.037	< 0.001
Time in racing (months)	1.017	1.010 - 1.025	< 0.001
Time with same trainer (months)	0.986	0.982 - 0.990	< 0.001
Wins/starts (30 – 60 days prior race)	1.002	1.001 - 1.003	< 0.001
Wins/starts (90 – 180 days prior race)	1.003	1.001 - 1.004	< 0.001

The final multivariable model included 22 risk factors with a statistically significant association with fracture injuries.

From the horse-related risk factors we have identified, the one with the potentially highest impact was the sex of the horse. Stallions had 58.5% (95% CI: 45.4% - 72.7%) more chance of sustaining a fracture injury compared to mares and geldings. The horse-related risk factor with the second highest impact was related to horses who had at some point in their career entered the vet list. Those horses were 57.5% (95% CI: 46.9% - 68.9%) more chance of sustaining a fracture injury. The number of previous injuries a horse had in its career was also found to be associated with fracture injuries, with horses having a 28.8% (95% CI: 13.4% - 46.3%) higher chance of fracture injury for each previous injury they had sustained. The age at which a horse begins its career was found to be associated with equine fractures. We found that for each year older the horse was at their first racing start there was a 6.0% (95% CI: 2.8% - 9.4%) higher chance of fracture injury in all its starts. Furthermore, for each month more a horse had been in racing there was an increased risk of 1.7% (95% CI: 1.0% - 2.5%). Horses were also found to have a lower chance of sustaining a fracture injury by 10.8% (95% CI: 6.6% - 14.8%) for each 60-day layup they had in their career. Horses were found to have a higher chance of sustaining a fracture injury by 2.4% (95% CI: 1.2% - 3.7%) for each extra month they spent, on average, between racing starts. Conversely, for each extra month they spent, on average, between exercise starts horses were at 4.7% (95% CI: 2.9% - 7.5%) less risk of fracture. Furthermore, when ranking the horses of a race according to their betting odds, horses less favoured by the odds were found to be less likely to sustain a fracture injury by 5.7% (95% CI: 4.5% - 6.9%) for each place further down the odds ranking. For each time a horse was scratched in its career not by a track veterinarian there was a 2.7% (95% CI: 0.8% - 4.5%) increased risk of fracture injury. Finally, for each extra km horses have accumulated from racing or exercise starts in their career there is 1.1% (95% CI: 0.8% - 1.5%). less risk of sustaining a fracture injury.

Regarding risk factors exploring the number of previous races a horse has participated we found that for each racing or exercise start a horse had during the month prior the race there was 29.1% (95% CI: 27.4% - 32.6%) less chance of a fracture injury. The same association was also found for the period of 30 to 60 days prior a race but for racing starts only. The risk of fracture injury was 8.5% (95% CI: 4.3% - 12.5%) less for each racing start. Conversly, for each racing or exercise start in the period 60 to 90 days prior the race horses had a 5.5% (95% CI: 2.5% - 8.7%) increased risk of fracture injury and additionally, for each racing and exercise start in the period 90 to 180 days prior a race an increased risk of 8.3% (95% CI: 6.9% - 9.6%). Furthermore, for the period of 30 to 60 days prior a race, for each extra percentage point in the percent of races won in that period horses were at 0.2% (95% CI: 0.1% - 0.3%) higher risk. The same observation can be made for the period 90 to 180 days prior the race with horses being at 0.3% (95% CI: 0.1% - 0.4%) higher risk.

Regarding race-related risk factors, the one with the potentially highest impact was the surface of the race. Horses were at 35.2% (95% CI: 19.9% - 52.5%) higher chance of fracture injury when running in a track with dirt surface compared to a synthetic one and at 25.3% (95% CI: 8.6% - 44.6%) more risk when running in turf compered to synthetic surfaces. The risk factor with the second highest impact was the country of the race, with horses participating in races in the US having a 19.5% (95% CI: 2.8% - 38.8%) higher chance of fracture injury. We, also found that horses participating in races with a purse of equal or less than $7,500 were 23.9% (95% CI: 15.9% - 31.0%) less likely to sustain a fracture injury than those competing in races in which the purse was greater than $7500. The distance of the race was also found to be associated with fracture injuries with horses racing on longer races having 5.1% (95% CI: 2.6% - 7.5%) less risk per extra furlong of the race.

Finally, we were able to identify one risk factor that was related to the trainer of the racing horse. For each extra month a horse had spent with the same trainer, it had 1.4% (95% CI: 1.0% - 1.8%) less chance of sustaining a fracture injury.

4.3.2.4. Model Fit

The multivariable model had a deviance of 56,438 with 1,962,393 degrees of freedom. The x^2 test statistic of the Hosmer-Lemeshow goodness of fit test was 3.99 with 8 degrees of freedom and an p-value of 0.86 indicating a good fit.

No numerical variables were assessed to be non-linear in the log odds.

No influential observation was found in the final multivariable model.

The area under the receiving operating characteristic curve was 66.8%.

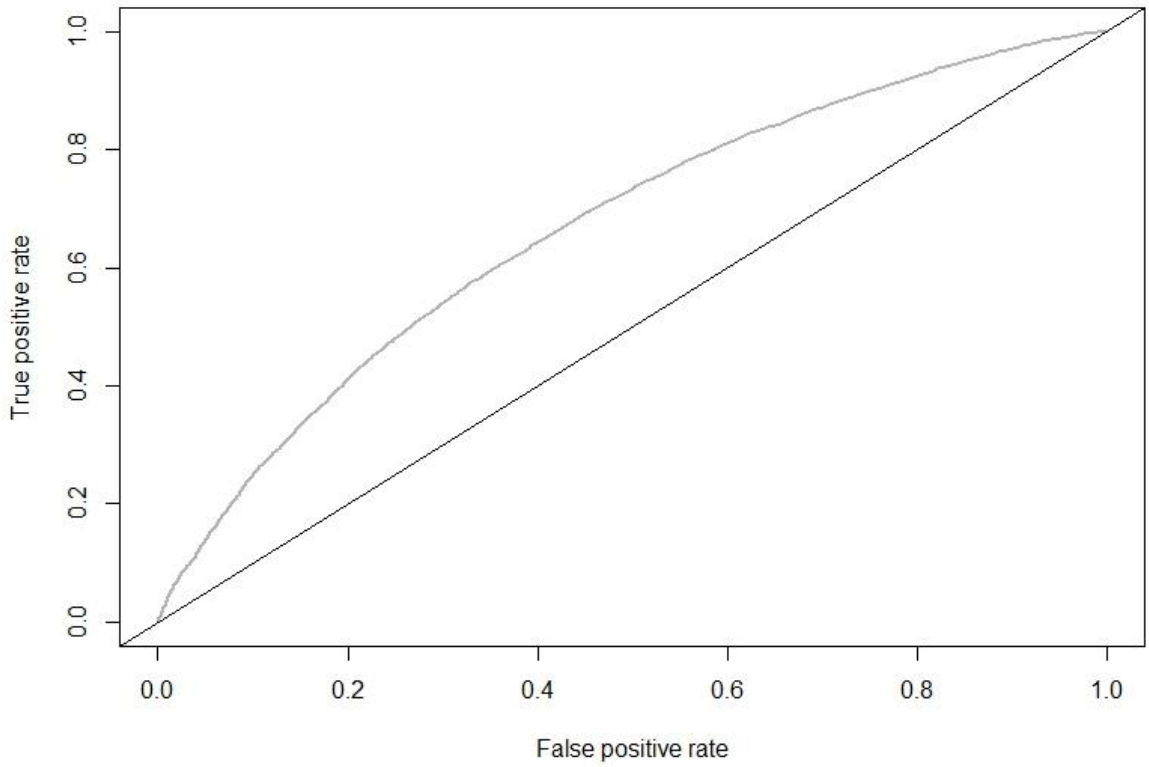

Figure 4-2 ROC curve of the multivariable logistic regression model trained on starts of Thoroughbred racehorses, six months after their first recorded racing or exercise start, from 2009 to 2015 for fracture injury prediction for the same period

Using the top 5% of fitted scores from our model we were able to identify the starts that resulted in a 2.9 times higher risk of fracture injury than the average injury prevalence for that study period.

4.4. Discussion

4.4.1. Performance of the Models

The power of both models was above 80% for identifying a statistically significant association with an odds ratio of 1.2 at the 95% confidence level. The plethora of starts available for the study resulted in a high power to identify risk factors significantly associated with fracture injuries and this consequently, means that both models have very low Type II error.

The fit of the models was assessed by using the Hosmer-Lemeshow goodness of fit test and there were no indications that the models did not fit the data well.

Furthermore, the area under the receiving operating characteristic curve was calculated for both models. The model that used all starts had a score of 0.62 while the model for horses that had been racing for more than 6-months, that assessed racing history risk factors achieved a score of 0.67. This indicates that both models are able to identify starts at a higher risk of injury by using the statistically significant risk factors identified.

Another indication that the models were able to identify the starts at higher risk is the population of starts identified by the top 5% of fitted values from the models. The population identified from the model that used all starts had a prevalence of fracture injury that was 2.3 times higher than the average. The population of starts identified from the model for horses that had been racing for more than 6 months had a prevalence of fracture injury that was 2.9 times higher than the average.

4.4.2. Risk Factors

Many of the risk factors identified to have a statistically significant association with fracture injuries across the final multivariable models are the same as in Chapter 3 of this Thesis. This is not surprising given that approximately 83% of the starts that resulted in a fatal injury were fractures. Therefore, in this part of the discussion we will focus only on new risk factors that were identified for fracture injuries that were not present in Chapter 3.

4.4.2.1. Risk factors for all horses and for horses that had been racing for ≥ 6 months

4.4.2.1.1. No. of layups

Horses had less chance of sustaining a fracture injury for each layup the had in their career. We hypothesise that taking a brief respite from racing might be beneficial to the horse. Furthermore, a possible explanation might be that horses had a layup because they needed rest from a situation threatening to their health that had it not been identified but might have resulted in an increased risk of fracture had they raced (Carrier, et al., 1998). It should be noted that the reason for the layup was not recorded in the data available for this study, so it was not possible to differentiate between lay-ups that were the result of injury and lay-ups were simply decided upon by the trainer to be required or part of a normal training regimen. Further details on the exact reasons for lay-ups would obviously be useful and would likely improve the predictive ability of future models.

4.4.2.1.2. No. of previous non-vet scratches

Horses that have been withdrawn from a race, not from the veterinarian at track, but for other reasons have a higher risk of sustaining a fracture. While inspection by regulatory veterinarians (Cohen, et al., 1997) has been shown to identify horses at higher risk of injury it seems that even horses scratched for other reasons are at a slightly higher risk. A study has also found that horses that were reluctant to start a race (Parkin, et al., 2006) were at higher risk and we hypothesise that there is sound reasoning for withdrawing a horse even if the scratch happens not by a veterinarian.

4.4.2.2. Risk factors for all horses

4.4.2.2.1. Time in layup (months)

The more time horse has spent in layup the higher the risk was for sustaining a fracture injury. As an extensive amount of absence from racing might be due to some sort of pathology we hypothesise this might be the reason for the identified increased risk. Furthermore, during this layup period, a different kind of exercise regimen might be in place for the horse, that might require no galloping which was identified as a risk factor for fractures in a previous study (Parkin, et al., 2005).

4.4.2.3. Risk Factors for horses that had been racing for ≥ 6 months

4.4.2.3.1. Accumulated distance run in career (km)

Horses were less likely to sustain a fracture injury per km they had raced in racing and exercise starts over their career. This finding might be due to healthy horses participating more in races and therefore having accumulated more racing distance in their career.

4.4.2.3.2. No. of racing and exercise starts (Present – 30 days prior race; 60 - 90 days prior race; 90 -180 days prior race) - No. of racing starts (30 - 60 days prior race)

As was the case for fatal injuries in chapter 3, we identified a similar trend when it comes to modelling prior racing starts of a racehorse. Furthermore, for fractures we were able to identify that the more racing and exercise starts a horse had in the period 60 to 90 days prior the race the more the risk of fracture.

4.4.2.3.3. Time between exercise starts – avg (months)

The more time a horse takes between exercise starts the less the risk was for a fracture injury. 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. We hypothesise that an excess time spent in exercise might increase the risk of fracture for the horse as the bones might not have sufficient time to remodel well after a continuous period of stress applied to them during training.

4.4.2.3.4. Time in racing (Months)

The more time a horse has spent in racing the higher the risk of sustaining a fracture. We hypothesise that this might be due to repeated stress applied to the bones as a result of continuous galloping. This is also in accordance to findings from studies identifying a higher risk for older horses (Estberg, et al., 1995; Cohen, et al., 2000; Williams, et al., 2001).

4.4.2. Limitations of the Study

As in chapter 3, we believe the identified risk factors are as unbiased as possible, since we have included in the statistical analysis 90% of racing starts from all official racing in the US and Canada for that period. A small source of bias could be the roughly 10% of starts which are not included in this study.

It is unclear why starts made in the US have a higher risk rate compared to starts in Canada. This is an interesting finding that need 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 exercise history of each horse and used the number of exercise 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 horses training regime is associated with the risk of injury during racing.

It is important to note that we did not make any attempt to differentiate the causes of fracture injury in the present study. Risk factors vary among types of fractures and it is likely that some of those risk factors were not identified in the present study. Thus, future analyses will be able to use more specific outcome variables to identify risk factors associated with the most common reasons for euthanasia of Thoroughbred racehorses following race-induced injuries.

Furthermore, statistical significance does not necessarily translate to clinical significance. Although we identified several risk factors that were significantly associated with fracture injuries in Thoroughbred horses competing in flat racing, it is important to point out that the vast majority of race starts evaluated in the present study did not result in a fracture injury. Finally, because of the extremely large number of race starts evaluated and the resulting high statistical power of this study, the magnitude of effect for some of the risk factors was very small.

4.4.3. Recommendations

The results of the present study can be used as a guideline for the identification of racehorses at high risk of sustaining a fracture injury during a race. The risk factors identified should be considered in the selection and implementation of measures expected to have the greatest effect on minimising the number of horses that sustain fracture injuries during flat races in the United States and Canada. Priority should be given to the consideration of methods to mitigate the effect of potentially modifiable risk factors with both the highest odds ratios and prevalence in the racing population in North America.

5. Predictive Models for Fatal Injuries and Fracture Injuries

5.1. Introduction

The aim of this part of the study was to identify the starts for Thoroughbreds at a higher risk of sustaining a fatal injury and fracture injury before the race. To do so we explored and evaluated the use, as predictive models, of logistic regression models, artificial neural networks both with deep and shallow architecture and random forest models. We aimed to use the vast amount of data available from 2009 to 2014 to train our models and separately use a full year of horse racing data, 2015, to acquire predictions and evaluate the performance of our models.

Logistic regression models have mainly been used in the field to identify risk factors for equine injuries. But rarely, if ever, have they been used to produce predictions, identifying starts for horses at a higher risk of sustaining a fatal injury. We decided to take advantage of the opportunity and make use of the robust technique of arriving at a risk factor model used in chapter 3 and chapter 4, and use these models to predict the risk of future starts.

Artificial neural networks are machine learning models that consist of a large number of processing units called nodes. These nodes are usually arranged in layers and nodes from a layer communicate with nodes from other layers through weighted connections (Figure 5-1).

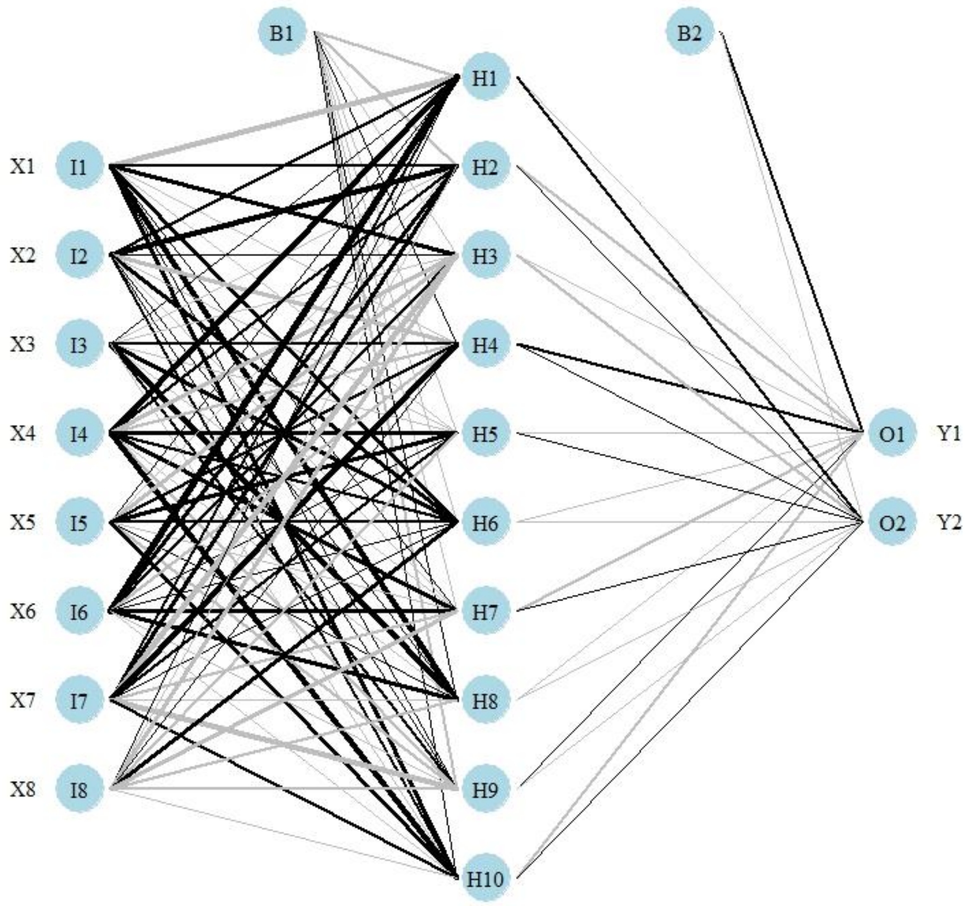

Figure 5-1 Representation of a neural network architecture showing the first 8 input nodes and the first 10 hidden layer nodes and the connection between them

Artificial neural networks were inspired from the way the human brain functions, learns and processes information (Alpaydin, 2014). The human brain consists of billions of neurons that transmit and collect information with each other through a branching network. When neurons are excited they send a signal to the network through a structure called the synapse that regulates the input and output activity of the neuron. Each neuron has hundreds of synapses. It is by changing the effectiveness of the synapses, making it easier or harder for a neuron to send or receive a signal to and from other neurons that learning is achieved.

In much the same way that the human brain uses neurons to process and transmit information; an artificial neural network uses nodes to receive input from other nodes, process it through the use of an activation function and produces output to be send to other nodes. The output of each node is sent through weighted connections to other nodes and it is by changing these weights that learning is achieved in an artificial neural network. Those weighted connections can start out randomly and adapt to the data provided, to learn the underlying relationship between the input data and the outcome (Alpaydin, 2014). One very useful feature of artificial neural networks is that the underlying relationship learned can be unknown (Priddy & Keller, 2005). It is not, therefore, required for the relationship between the possible predictors and the outcome to be explicitly specified for the models to be trained, allowing the models to be less constrained.

Classifying a start as an injury or not, is a complex problem where the underlying relationship between the possible predictors and the outcome is essentially unknown and that is why we decided to make use of the neural networks' powerful learning capabilities. Furthermore, artificial neural networks have been successfully used to solve complex problems in a diverse field of domains (Priddy & Keller, 2005) and we wanted to explore the possibility of using them to successfully predict equine injuries.

Deep Learning methods learn feature hierarchies where higher level features are learned via the use of lower level features (Bengio, 2009; Lee, et al., 2009; Glorot & Bengio, 2010). This procedure is realised using a deep architecture, such as a many-layered neural network, consisting of the composition of computational operations on each layer (Larochelle, et al., 2007). This allows the deep learning model to learn a complex function that maps the input to the output, without relying on features specified *ad hoc* from humans. That property is very useful in highly abstract problems, where the connection between the raw input and the desired output is not known (Bengio, 2009; Erhan, et al., 2009). LeCun, et al., (1998) advocate that, with the availability of computers with fast arithmetic units, large datasets and powerful learning techniques, better models can be built by relying on automated learning procedures than on hand-designed heuristics. Even shallow neural networks have been extremely useful when faced with a complex problem and simplification is unacceptable (Bhadesia, 1999), although there has been evidence suggesting that when the problem is complex enough and there enough data available, deep architectures perform better (Bengio & Delalleau, 2009; Larochelle, et al., 2009).

Since deep learning techniques have been used with success for tasks ranging from classification to robotics and natural language processing (Bengio, 2009), we believe they may also be useful for predicting equine fatal or fracture injuries in flat racing. Our classification problem is a highly abstract one where the road from our input (risk factors at race start) to our output (equine injury) is not known. Furthermore, our response, an equine fatal or fracture injury, is an extremely rare result, less than 0.2 % of the starts, producing an imbalanced, hard to predict, output. However, by making use of the data available in the EID over the 7-year period, we find ourselves in a data rich environment in which we believe deep learning methods could be efficiently used to produce a satisfactory classification result.

We therefore decided to train two networks with deep architecture, a deep belief network (DBN) and a stacked denoising autoencoder (SDA).

Random forests are a machine learning technique introduced by Breiman (2001) where an ensemble of classification tree models is trained and they 'vote' on the most popular class. The technique was inspired by earlier work by Amit and Geman (1997) and is an extension of bagging (Breiman, 1996). By using bootstrap replicates of the learning set to train each tree model, random forest improve accuracy and reduce the risk of overfitting to the data, compared to results obtained from a single tree model.

Random forests have many appealing features, they are fast to train and to predict, they can be used directly for high-dimensional problems (Zhang & Ma, 2012) , they are user-friendly and provide robust results handling the noise in the data well (Liaw & Wiener, 2002).

However, random forests do not produce satisfactory results when they deal with imbalanced data classification problems (Chen, et al., 2004). Chen, et al., (2004) addressed this by introducing two new techniques, weighted random forests and balanced random forests. Balanced random forests deal with the imbalance by using bootstrap replicates of the learning set that contain the same number of observations from the majority class as the minority class. Weighted random forests deal with the imbalance by assigning a heavier penalty when misclassifying the minority class. Xie, et al., (2009) proposed a combination of those two techniques called improved balanced random forest that retains the desirable features of being computationally efficient with large imbalanced datasets as well as being noise tolerant.

As we are dealing with an extremely imbalanced dataset where the prevalence of fatal or fracture injuries is approximately 0.2%, we decided to make use of the improved random forest capabilities of dealing with imbalanced problems, being robust to noise, generalising well and avoiding overfitting and being able to efficiently handle a large amount of data.

To our knowledge, this is the first study to train logistic regression and machine learning models to predict equine injuries using such an extensive amount of data and a full year of horse racing events for prediction and evaluation.

The results could help identify horses at high risk on entering a race and inform the design and implementation of preventive measures aimed at minimising the number of Thoroughbreds sustaining fatal injuries or fractures during racing in North America. In addition, understanding the predictive ability of different models will help the racing regulatory authorities in coming to decisions about what to do with the information that comes from risk factor analysis.

5.2. Materials and Methods

5.2.1. Study Population

The study population comprised all 188,269 Thoroughbred horses that participated in flat racing in the US and Canada from 1st January 2009 to 31st December 2015 in the 89 race tracks reporting injuries to the EID. The prevalence of fatal injury was 0.18% for the 2,493,957 racing starts in the 7-year study period and the prevalence of fractures was 0.20% for the same period.

The data available for the first six years were used to train the predictive models. They contained information on all 171,523 Thoroughbred horses that participated in flat racing in the US and Canada from 1st January 2009 to 31st December 2014. The prevalence of fatal injury was 0.19% for the 2,201,152 racing starts in the 6-year period and 0.21% for fractures.

The data available for the last year in the EID, 2015, were used as a test set to obtain predictions from the models and evaluate their performance. They contained information on all 50,882 Thoroughbred horses that participated in flat racing in the US and Canada from 1st January 2015 to 31st December 2015. The prevalence of fatal injury was 0.16% for the 292,805 racing starts in 2015 and 0.18% for fractures.

5.2.2. Model Evaluation

The performance of each predictive model was evaluated by calculating the Area Under the Receiver Operating Characteristic Curve, as suggested by Bradley (1997), who considers the AUC to be one of the best ways to evaluate a classifier's performance. Also, the evaluation method had to take into account both possible outcomes since the response is extremely imbalanced and it is trivial for a classifier to achieve an accuracy of over 99.8% by simply predicting a 0% probability of an equine fracture at every start. The Receiver Operating Characteristic Curve is a plot of sensitivity, the true positive rate, versus 1-specificity, the false positive rate, for an extensive range of cutoff points. The AUC can range from 0.5 for a model with no discrimination ability (i.e. equivalent to a coin toss) to 1 for a model that perfectly discriminates between the two outcomes. The AUC score can be interpreted as the probability that the measure of risk provided by the models is higher for a case than a non-case.

Furthermore, for each model we looked at the population of starts identified by the top and bottom 5% of predicted risk scores. We compared the average risk of sustaining a fatal injury for this population with the average risk for 2015 to get a relative measure of performance for each model.

Lastly, bootstraps with 10,000 iterations were used to calculate the 95% confidence interval for the area under the curve and for the prevalence for the starts identified by the bottom and top 5% of fitted scores (Efron & Tibshirani, 1986; Wolter, 2007). A bootstrap is a simple random sample with replacement selected from the original main sample. Repeated bootstrap sampling from the main sample produces alternative feasible samples that could have been selected as the main sample from the original distribution. Bootstraps can therefore be used to produce unbiased estimators of the variance of even nonparametric statistics and can be used to produce confidence intervals (Wolter, 2007).

Neural networks were trained using Enthought Canopy (2014) and the Python programming language (Van Rossum, 2007) under the Theano framework (Al-Rfou, et al., 2016). All the rest of statistical analyses and calculations in this chapter were conducted using RStudio, developed by RStudio Team (2015), and the R programming language by the R Development Core Team (2008). The “party” package in R was used to create the tree models that are part of the IBRF models (Hothorn, et al., 2006).

5.2.3. Multivariable Logistic Regression

In order to identify horses at particular risk prior to entering a race we developed and validated predictive models utilising the logistic regression models developed for our risk factor analysis in chapters 3 and 4. Models were trained on the 6-year period 2009 – 2014 and validated on the 2015 starts. To obtain predictions from the logistic regression models we used both the model developed for all starts and the model developed on starts from horses that have been racing for at least six months. For each individual start, based on how long the horse has been participating in racing, the appropriate model was used for obtaining a prediction. Predictions were obtained from the model developed for all starts for starts where the horse had not been in racing for at least six months. For starts from horses that had been in racing for at least six months, predictions were obtained from the model developed specifically on those starts. This selection of predictions was used to validate collectively the classification results of the logistic regression models.

5.2.4. Artificial Neural Networks

Initially, we decided to train a neural network with a shallow architecture for our task. Given a big enough dataset even simple architectures (Ciresan, et al., 2010) or architectures without unsupervised pre-training (Ciresan, et al., 2012) can achieve good competitive results. The artificial neural network used was feed-forward, it was trained using the backpropagation algorithm (Rumelhart, et al., 1986) and had one fully connected hidden layer with 500 nodes. A feed-forward neural network is one where the data entered in the network through the input layer is passed through the network layers until it reaches the output layer and layers are only connected to the previous layer. Backpropagation was used to calculate the gradient of the error with respect to the weights and then update the weights of the network appropriately.

For training the deep belief network two hidden layers were used. The first had 50 nodes and the second 500. We initially trained layer-wise Restricted Boltzmann Machines (RBM), using a hyperbolic tangent activation function (Glorot & Bengio, 2010; LeCun, et al., 2012) and then fine-tuned the models using backpropagation (Rumelhart, et al., 1986).

RBMs are able to capture the connection strength between units and find the underlying structure of its environment (Ackley & Hinton, 1985). RBMs were trained layer-wise using contrastive divergence with a chain step of one, using Gibbs sampling as the transition operator for the Markov chain (Hinton, 2005; Hinton, et al., 2006; Bengio & Delalleau, 2009). By using RBM's in the pre-training phase we initialised the weights of the network in such a way that it achieves better performance in the fine-tuning phase (Larochelle, et al., 2007; LeRoux & Bengio, 2008; Erhan, et al., 2009; Hinton, 2013). RBMs initialise the weights in an unsupervised manner between two consecutive layers of the network that have no intra-layer connections. As indicated by Erhan, et al., (2010) the beneficial effects of the unsupervised pre-training do not diminish during the fine-tuning phase.

To train the SDAs we used the same deep architecture as for the DBNs, two layers of 50 and 500 nodes, as for the DBNs. SDAs use the same philosophy as DBNs of using a local unsupervised criterion to pre-train each layer (Vincent, et al., 2010). SDAs use DAs as this pre-training criterion.

DAs are autoencoders with stochastically corrupted input. Autoencoders map the input to the hidden layer and then use the same input as the output layer (Bengio, 2012). They help initialise the weights of the network in an unsupervised manner by effectively using the same input layer as the output layer creating more robust networks to start supervised training with. DAs were introduced to bypass the limitations of the autoencoders, that is, to be able to use larger representations and not to be limited to "bottleneck" reduced-dimension ones (Vincent, et al., 2010; Bengio & Delalleau, 2011; Bengio, 2012). However, adding a corruption level not only solves the problem of our model trivially learning the identity function but enforces stability and robustness in the network. Having to perform a Denoising task as well ensures that the features learned capture useful structural information of our input data (Vincent, et al., 2010). In our case, a corruption level of 10% was introduced on the first layer and a corruption level of 20% on the second, based on optimal predictive results achieved with corruption levels between 10% and 25% in benchmark tests in a study by Vincent, et al (2008).

From the variables, available in the EID, we used in the input layer of all the models those that were found to be associated with equine fatalities, signified by a p-value of less than 0.2 on univariable logistic regression models. The < 0.20 threshold p-value was chosen for this screening process, to prevent the exclusion of a predictor that only becomes evident when we have controlled for a confounder (Dohoo, et al., 2003).

We transformed our input variables as suggested by LeCun, et al., (2012). Categorical variables were given values of either -1 or 1. For those with more than two levels, dummy variables were created. Numerical variables were centred at zero and normalised. Values beyond -3 and 3 were trimmed and then we further divided by 3 so that the range of values was brought between -1 and 1. The final input layer consists of 52 nodes. Our final datasets consisted of a training set containing information on approximately 1,760,000 starts, a validation set of approximately 440,000 and a test set of approximately 300,000 starts. The training and validation set observations were randomly selected from the observations available for the six years 2009-2014. We specifically use a validation set that is separate from the test set to check and stop the networks from overfitting the data. The test set is not used for stopping overfitting because this would have resulted in optimal predictions from the networks, overestimating their actual performance on unseen data. The test set where we assess the performance of our models contains all starts from 2015.

Training of the models was conducted using 176 mini-batches of 10,000 observations each. This stochastic learning method was chosen because it is faster and it usually leads in better solutions as it is easier for the model to avoid local minima (LeCun, et al., 2012).

For all the networks the sigmoid hyperbolic tangent was used as the activation function for the nodes, that converges faster than the logistic sigmoid (Glorot & Bengio, 2010; LeCun, et al., 2012). We decided to use a constant learning rate of 0.3 and following Simard, et al., (2013) we avoided using momentum, weight decay or structure-dependent learning rates for parsimony. Models requiring pre-training were pre-trained for 30 epochs and a pre-training learning rate of 0.001 was used. Learning rate is the percentage by which the weights of the network are being updated in each epoch.

A final fully connected output layer with two nodes and the Softmax activation function was added on each model (Dunne & Campbell, 1997). All models were trained for 1000 epochs. Each time the network weights were fully updated the same methodology was repeated on the next training or pre-training epoch.

5.2.5. Improved Balanced Random Forest

A predictive model was trained using the Improved Balanced Random Forests (IBRF) algorithm proposed by Xie, et al (2009).

The IBRF algorithm is based on Random Forests developed by Breiman (2001; 2004) and specifically on Balanced Random Forests an approach suggested by Chen, et al., (2004) to accommodate for imbalanced outcomes.

For our model, 500 Classification and Regression Trees (Breiman, et al., 1984)) were trained on samples, randomly drawn with replacement, numbering A*N positive outcomes and (2-A)*N negative outcomes, where N is the number of all positive outcomes and A is a number randomly and uniformly selected for each sample, between 0.8 and 1.2. Positive and negative cases were then weighted by 1/A and 1/(2-A) respectively. The resulting prediction is the averaged outcome from our ensemble of classifiers.

The variables used as predictors to train the classification trees were those that were found to be associated with equine fatalities, signified by a p-value of less than 0.2 on univariable logistic regression models.

5.3. Results

5.3.1. Multivariable Logistic Regression

5.3.1.1. Fatal injuries

The AUC for the multivariable logistic regression model was 63.5% (95% CI: 60.9%–66.0%).

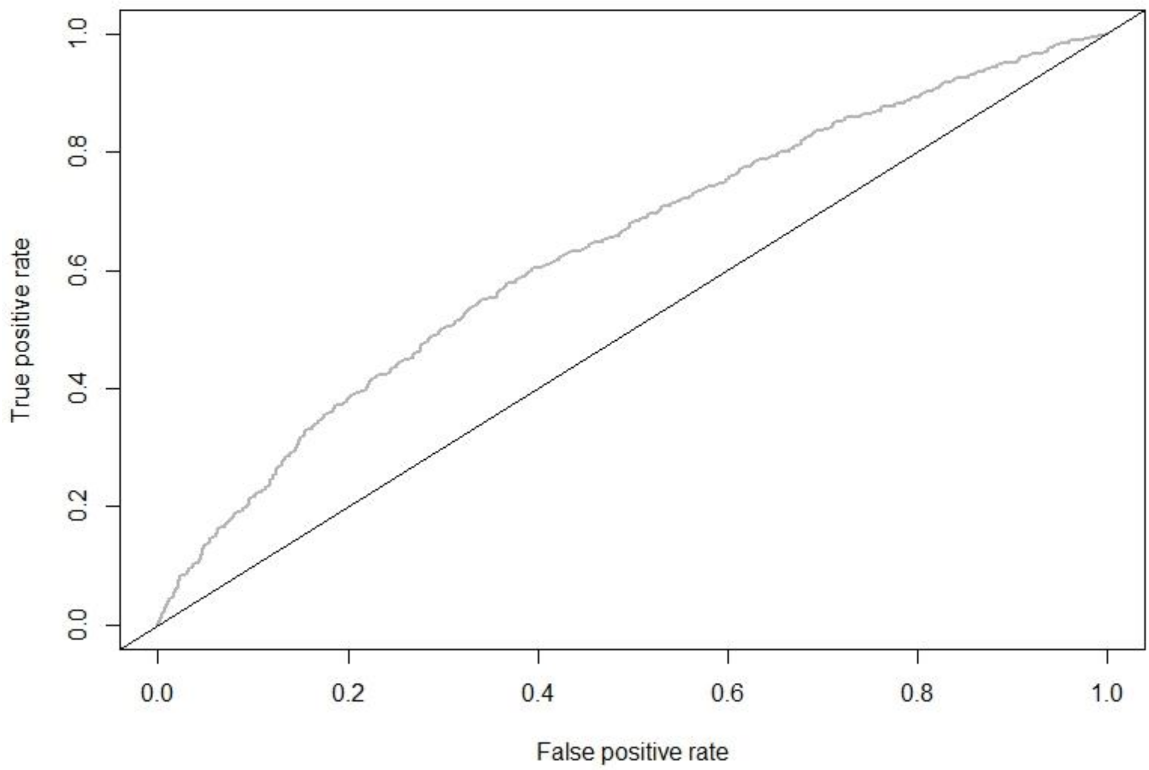

Figure 5-2 ROC curve of the multivariable logistic regression model trained on starts from 2009 to 2014 for fatal injury prediction of 2015 starts

The 5% of starts that had the highest score in the model for 2015 were found to have 2.7 times (95% CI: 2.1–3.4) higher fatality prevalence than the fatality prevalence of 2015. Contrary to this, the 5% of starts that had the lowest score were found to have approximately 0.3 the risk (95% CI: 0.1–0.5) of the mean fatality prevalence of 2015.

5.3.1.2. Fracture injuries

The AUC for the multivariable logistic regression model was 63.2% (95% CI: 60.9%–65.5%).

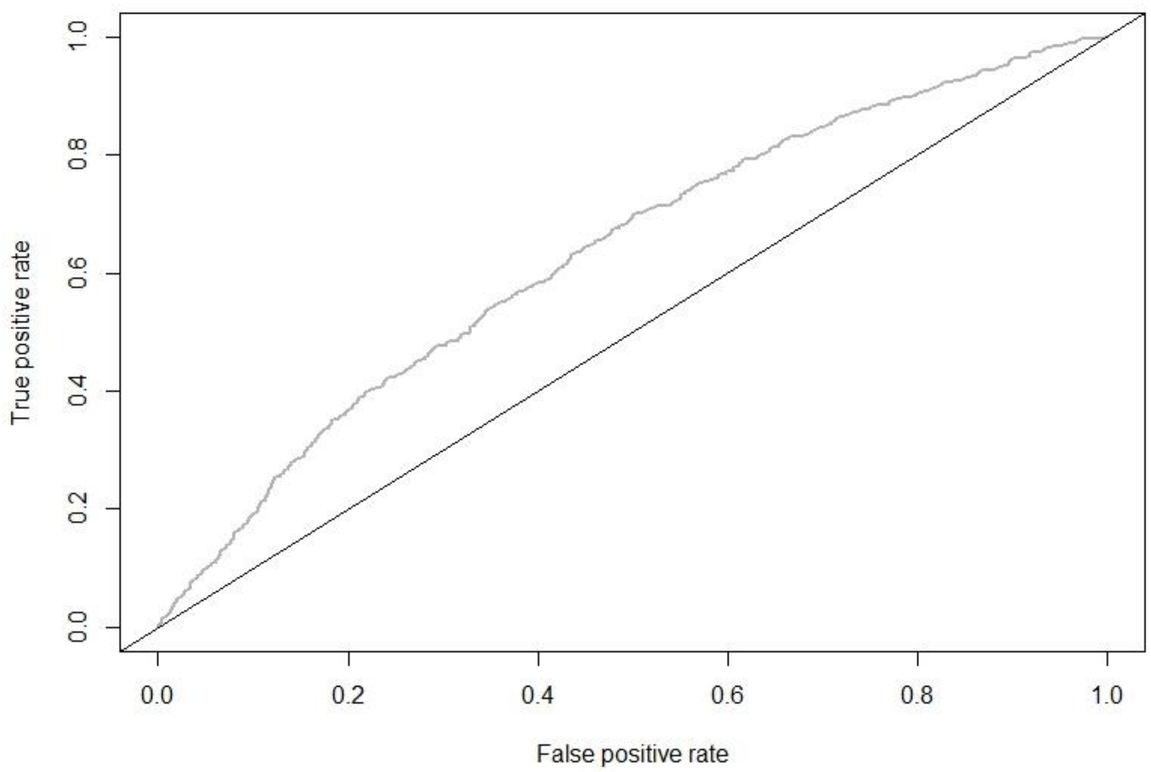

Figure 5-3 ROC curve of the multivariable logistic regression model trained on starts from 2009 to 2014 for fracture injury prediction of 2015 starts

The 5% of starts that had the highest score in the model for 2015 were found to have 2.0 times (95% CI: 1.5–2.5) higher fracture prevalence than the fracture prevalence of 2015. Contrary to this, the 5% of starts that had the lowest score were found to have approximately 0.3 the risk (95% CI: 0.1–0.6) of the mean fracture prevalence of 2015.

5.3.2. Artificial Neural Networks

5.3.2.1. Fatal injuries

The AUC for the artificial neural network model was 62.9% (95% CI: 60.3%-65.5%).

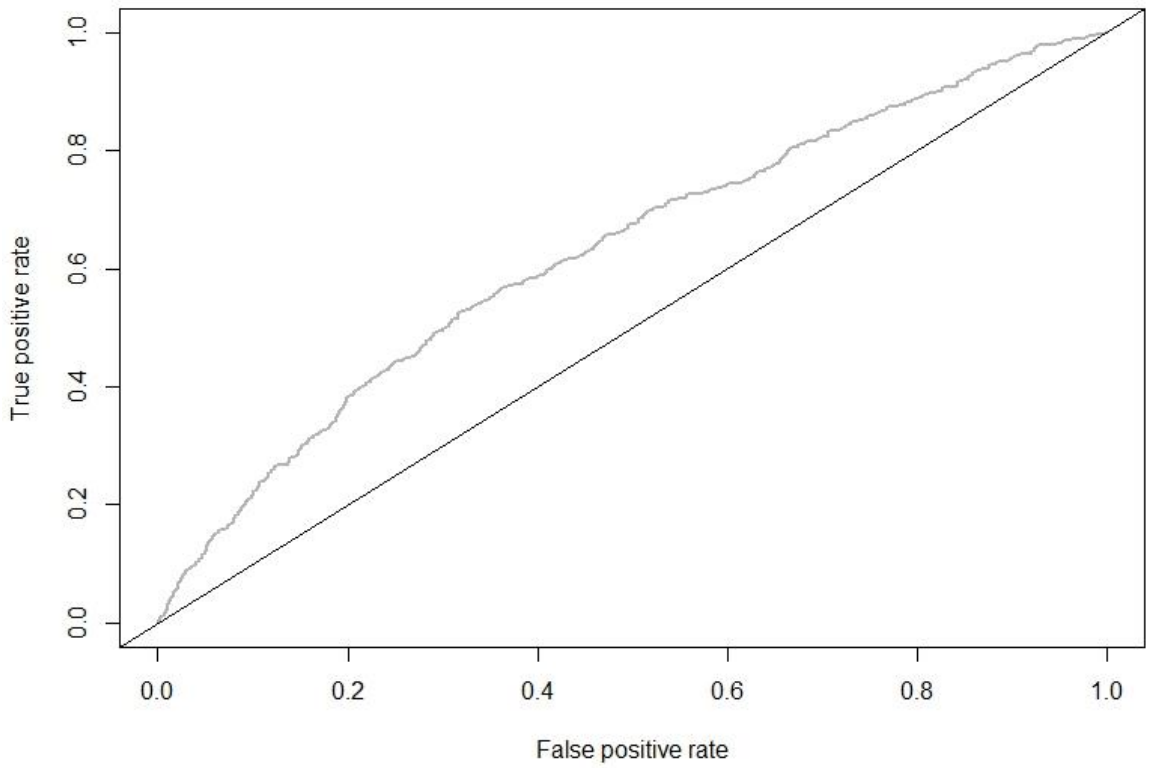

Figure 5-4 ROC curve of the artificial neural network model trained on starts from 2009 to 2014 for fatal injury prediction of 2015 starts

The 5% of starts that had the highest score in our models for 2015 were found to have 2.3 times (95% CI: 1.8 – 3.0) higher fatality prevalence than the fatality prevalence of 2015. Contrary to this, the 5% of starts that had the lowest score were found to have approximately 0.3 the risk (95% CI: 0.1 – 0.6) of the mean fatality prevalence of 2015.

The AUC for the stacked denoising autoencoder model was 62.7% (95% CI: 60.0%-65.3%).

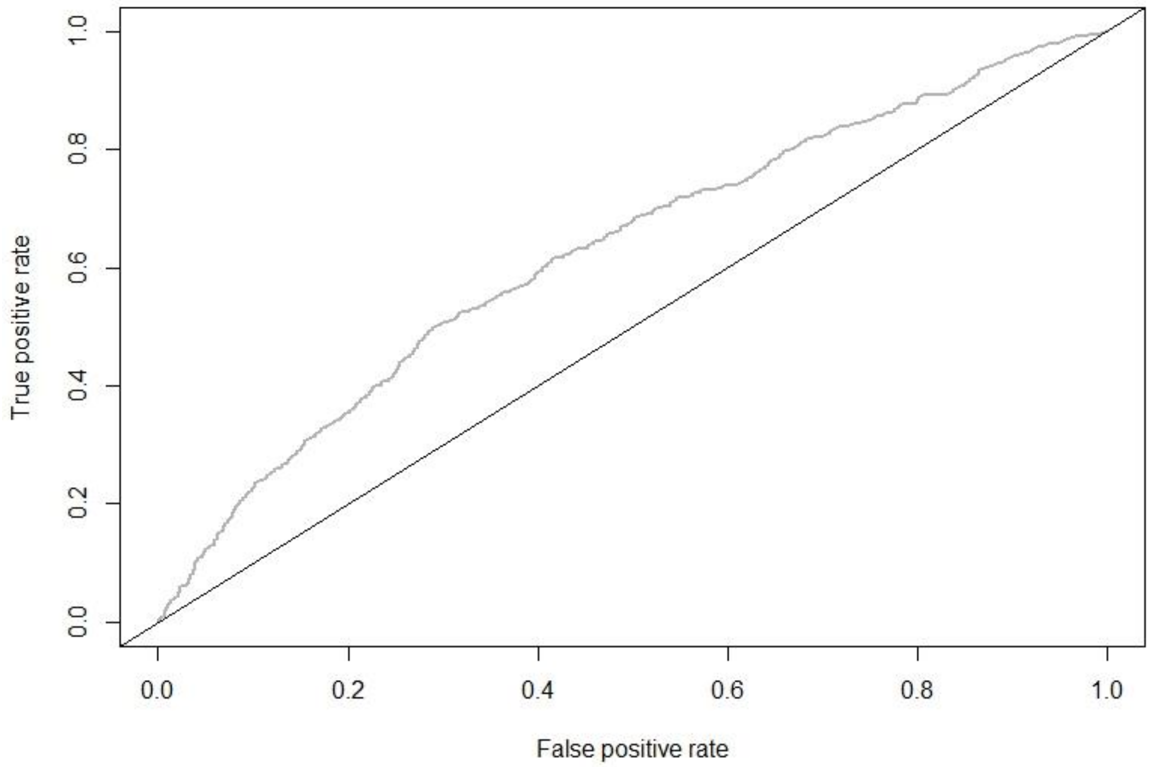

Figure 5-5 ROC curve of the stacked denoising autoencoder model trained on starts from 2009 to 2014 for fatal injury prediction of 2015 starts

The 5% of starts that had the highest score in the model for 2015 were found to have 2.4 times (95% CI: 1.8-3.0) higher fatality prevalence than the fatality prevalence of 2015. Contrary to this, the 5% of starts that had the lowest score were found to have approximately 0.4 the risk (95% CI: 0.1–0.6) of the mean fatality prevalence of 2015.

The AUC for the deep belief network was 63.1% (95% CI: 60.4%–65.6%).

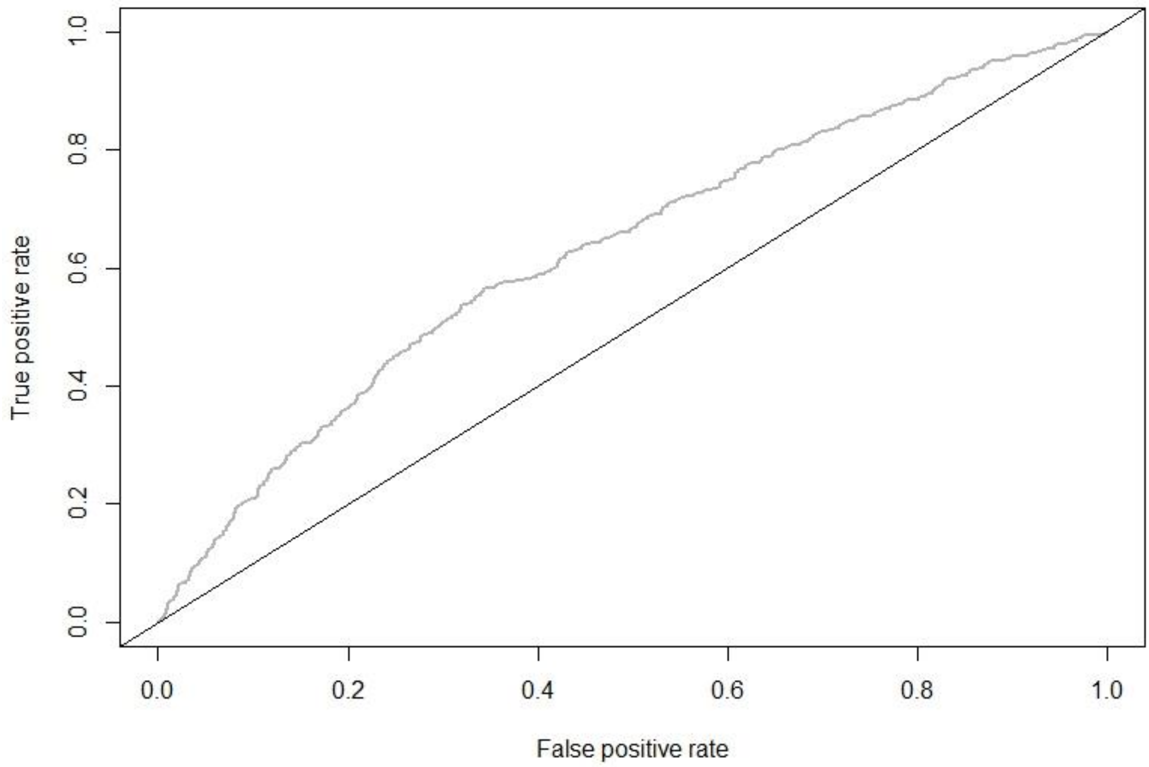

Figure 5-6 ROC curve of the deep belief network model trained on starts from 2009 to 2014 for fatal injury prediction of 2015 starts

The 5% of starts that had the highest score in the model for 2015 were found to have 2.3 times (95% CI: 1.7–2.8) higher fatality prevalence than the fatality prevalence of 2015. Contrary to this, the 5% of starts that had the lowest score were found to have approximately 0.4 the risk (95% CI: 0.2–0.7) of the mean fatality prevalence of 2015.

5.3.2.2. Fracture injuries

The AUC for the artificial neural network model was 62.8% (95% CI: 60.3%–65.1%).

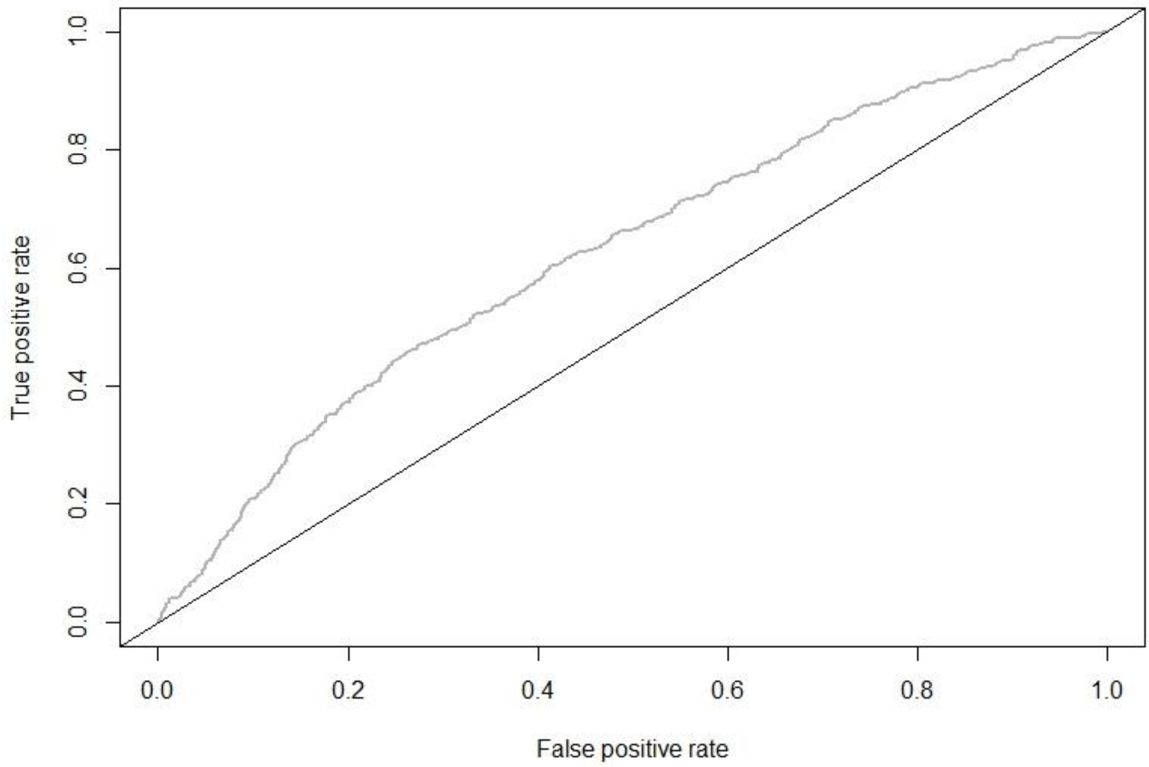

Figure 5-7 ROC curve of the artificial neural network model trained on starts from 2009 to 2014 for fracture injury prediction of 2015 starts

The 5% of starts that had the highest score in the model for 2015 were found to have 1.9 times (95% CI: 1.4–2.4) higher fracture prevalence than the fracture prevalence of 2015. Contrary to this, the 5% of starts that had the lowest score were found to have approximately 0.1 the risk (95% CI: 0.04–0.4) of the mean fracture prevalence of 2015.

The AUC for the SDA model was 62.3% (95% CI: 59.8%–64.7%).

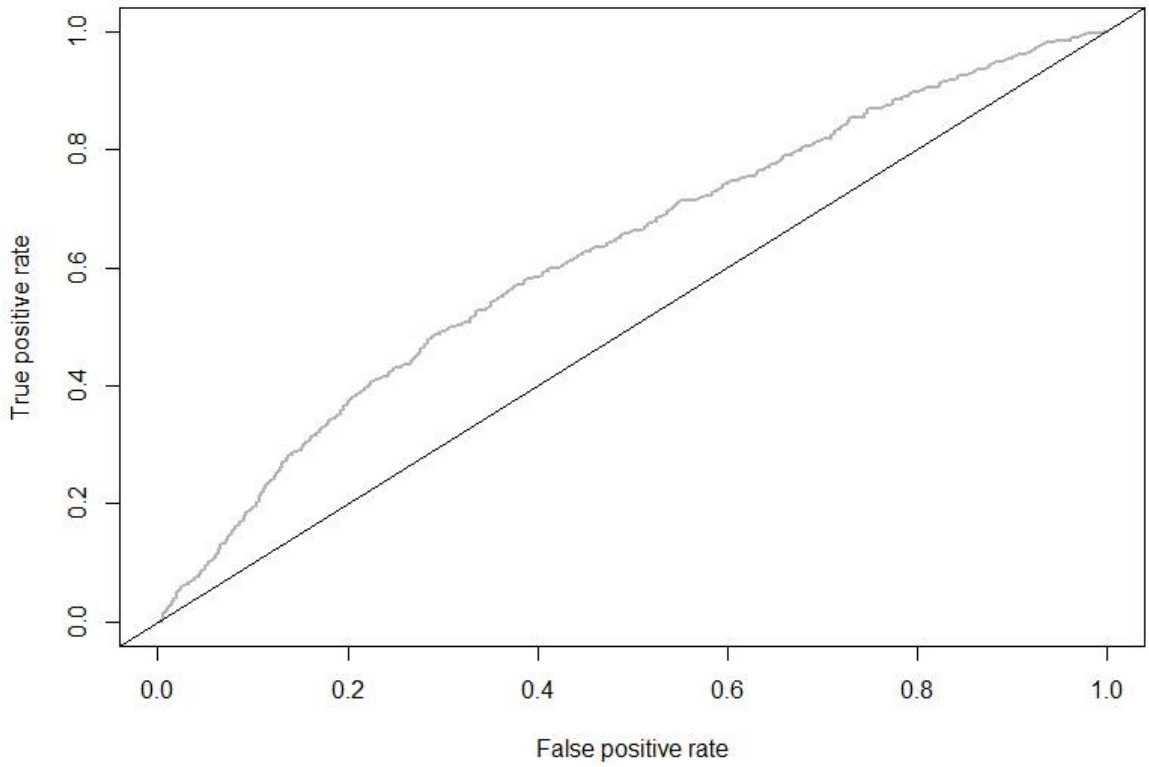

Figure 5-8 ROC curve of the stacked denoising autoencoder model trained on starts from 2009 to 2014 for fracture injury prediction of 2015 starts

The 5% of starts that had the highest score in the model for 2015 were found to have 1.8 times (95% CI: 1.4–2.3) higher fracture prevalence than the fracture prevalence of 2015. Contrary to this, the 5% of starts that had the lowest score were found to have approximately 0.3 the risk (95% CI: 0.1–0.5) of the mean fracture prevalence of 2015.

The AUC for the DBN model was 62.4% (95% CI: 60.1%–64.8%).

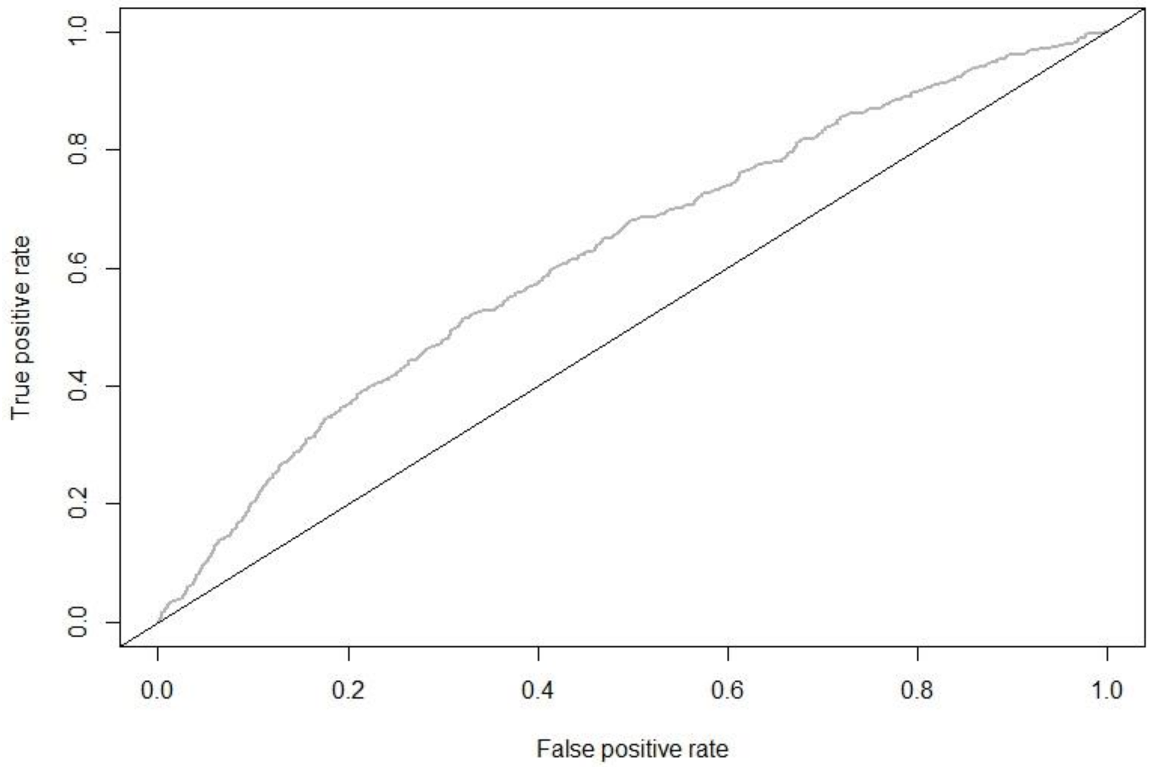

Figure 5-9 ROC curve of the deep belief network model trained on starts from 2009 to 2014 for fracture injury prediction of 2015 starts

The 5% of starts that had the highest score in the model for 2015 were found to have 2.0 times (95% CI: 1.5–2.5) higher fracture prevalence than the fracture prevalence of 2015. Contrary to this, the 5% of starts that had the lowest score were found to have approximately 0.5 the risk (95% CI: 0.2–0.7) of the mean fracture prevalence of 2015.

5.3.3. Improved Balanced Random Forest

5.3.3.1. Fatal injury

The AUC for the improved balanced random forest model was 62.8% (95% CI: 60.1%-65.3%).

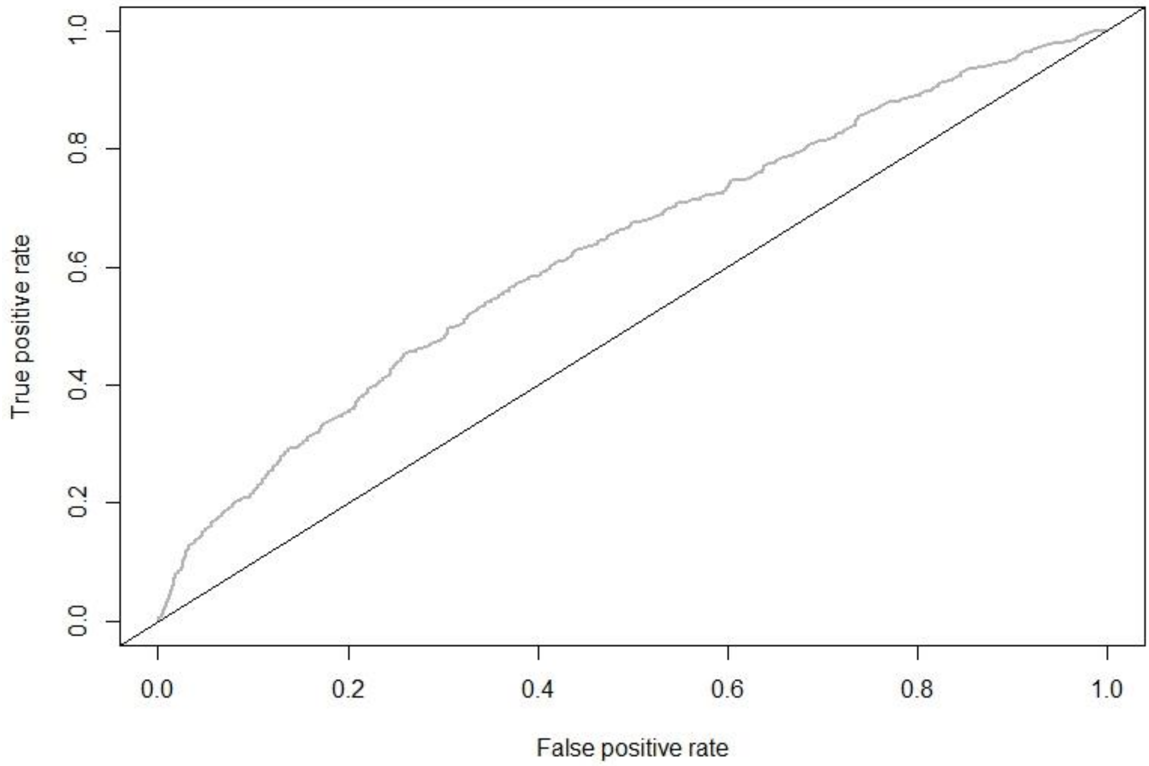

Figure 5-10 ROC curve of the improve balanced random forest model trained on starts from 2009 to 2014 for fatal injury prediction of 2015 starts

The 5% of starts that had the highest score in the model for 2015 were found to have 3.2 times (95% CI: 2.5–3.8) higher fatality prevalence than the fatality prevalence of 2015. Contrary to this, the 5% of starts that had the lowest score were found to have approximately 0.4 the risk (95% CI: 0.2–0.7) of the mean fatality prevalence of 2015.

5.3.3.2. Fracture injury

The AUC for the IBRF model was 62.9% (95% CI: 60.5%–65.2%).

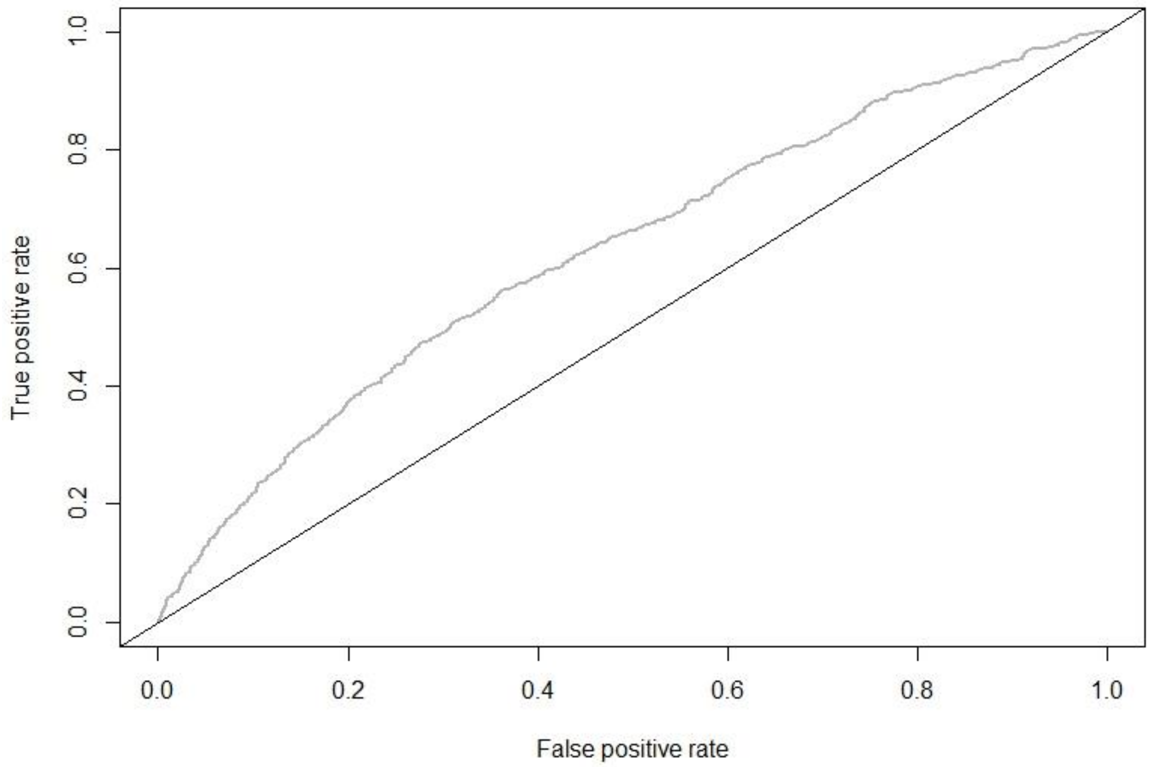

Figure 5-11 ROC curve of the improve balanced random forest model trained on starts from 2009 to 2014 for fracture injury prediction of 2015 starts

The 5% of starts that had the highest score in the model for 2015 were found to have 2.5 times (95% CI: 2.0–3.1) higher fracture prevalence than the fracture prevalence of 2015. Contrary to this, the 5% of starts that had the lowest score were found to have approximately 0.4 the risk (95% CI: 0.2–0.6) of the mean fracture prevalence of 2015.

5.3.4. Models comparison

5.3.4.1. Fatal injuries

The models produced similar predictive results with an AUC of approximately 63%.

All models were able to identify starts from low-risk horses that had approximately 0.3 of the risk of the average horse of sustaining a fatal injury and starts from high-risk horses that had approximately 2.3 times the risk of the average horse.

Table 5-1 Predictive Models - Area Under the Receiver Operating Characteristic Curve for predictions on 2015 fatal injuries

Model	AUC %	95% CI
Logistic Regression	63.5	60.9 - 66.0
Neural Network	62.9	60.3 – 65.5
IBRF	62.8	60.1 - 65.3
SDA	62.7	60.0 - 65.3
DBN	63.1	60.4 - 65.6

Table 5-2 Predictive Models - Ratio of the 5% of starts identified to have the least risk of fatal injury to average risk of 2015

Model	Ratio	95% CI
Logistic Regression	0.298	0.1 - 0.5
Neural Network	0.344	0.1 - 0.6
IBRF	0.426	0.2 - 0.7
SDA	0.388	0.1 - 0.6
DBN	0.431	0.2 - 0.7

Table 5-3 Predictive Models - Ratio of the 5% of starts identified to have the highest risk of fatal injury to average risk of 2015

Model	Ratio	95% CI
Logistic Regression	2.729	2.1 - 3.4
Neural Network	2.325	1.8 – 3.0
IBRF	3.155	2.5 - 3.8
SDA	2.368	1.8 - 3.0
DBN	2.239	1.7 - 2.8

5.3.4.2. Fracture injury

The models produced similar predictive results with an AUC of approximately 62%-63%.

All models were able to identify starts from low-risk horses that had approximately 0.3 of the risk of the average horse of sustaining a fracture injury and starts from high-risk horses that had approximately 2.3 times the risk of the average horse.

Table 5-4 Predictive Models - Area Under the Receiver Operating Characteristic Curve for predictions on 2015 fracture injuries

Model	AUC %	95% CI
Logistic Regression	63.2	60.9 - 65.5
Neural Network	62.8	60.3 - 65.1
IBRF	62.9	60.5 - 65.2
SDA	62.3	59.8 - 64.7
DBN	62.4	60.1 - 64.8

Table 5-5 Predictive Models - Ratio of the 5% of starts identified to have the least risk of fracture injury to average risk of 2015

Model	Ratio	95% CI
Logistic Regression	0.331	0.1 - 0.6
Neural Network	0.188	0.04 - 0.4
IBRF	0.367	0.2 - 0.6
SDA	0.300	0.1 - 0.5
DBN	0.450	0.2 - 0.7

Table 5-6 Predictive Models - Ratio of the 5% of starts identified to have the highest risk of fracture injury to average risk of 2015

Model	Ratio	95% CI
Logistic Regression	1.985	1.5 - 2.5
Neural Network	1.914	1.4 - 2.4
IBRF	2.537	2.0 - 3.1
SDA	1.839	1.4 - 2.3
DBN	1.989	1.5 - 2.5

5.4. Discussion

5.4.1. Performance of the Models

The models utilized in the study managed to achieve satisfactory predictive results for 2015. All models were able to achieve a statistically significant AUC score of more than 62% and identify a population of starts with approximately twice or more the average risk of 2015 and a population of starts with a risk of less than half the average risk of 2015.

Our research, also, shows that deep learning models can be used to achieve good results outside their usual realm of image classification and object recognition. It also suggests that our problem might be comprised by abstract low and high level features that the deep learning models were able to grasp. The shallow ANNs achieved results, similar to the deep learning models suggesting they can be used in difficult, imbalanced and complex, though data-rich, environments.

One important finding of this study is that we obtained similar predictive results from a wide array of different models. Machine learning techniques based on neural networks and deep learning, produced similar results with the improved balanced random forest, a machine learning technique based on classification tree, and the multivariable logistic regression model, a classical statistical approach. This shows that predictive models, with moderate predictive value indicated by their AUC scorers, can be learned and that maybe, this the highest possible predictive result that can be learned based on the available data, particularly considering the rare outcomes studied in this work.

5.4.2. Recommendations

Based on the results achieved in this part of the study the models could help identify horses at high risk on entering a race. An online system could be put in place identifying those horses with higher risk than the average and the information could be provided well before the race to veterinarians at the track and interested parties.

Furthermore, the implementation of such a system would only require the use of the easily programmable multivariable logistic regression model since predictive results are similar across models.

Interventions based on the predictive results of the statistical models should only follow with extreme care. On average, approximately two out of 1000 starts result in a fatal or fracture injury. We could consider a horse that has two or three times the risk of the average horse, of sustaining an injury during a racing start, as high-risk. However, the risk for even a high-risk horse would be below 1%. In our opinion, this cannot be the basis for regulation preventing a horse from not racing if identified as high-risk as this would disproportionately affect horses that could race without any problem.

If the information is provided to the track veterinarian it should be regarded as a helpful tool but not the basis on the decision to scratch a horse from a race. This could be problematic if a horse identified as ‘high-risk’ is not scratched and ends up sustaining an injury or if a ‘high-risk’ horse is scratched but on subsequent races, or even its whole career, participates without any problems. It is, of course, impractical to scratch approximately 15,000 starts per year to prevent 60 to 90 injuries which would be expected for high-risk starts according to the models.

The same arguments follow if the information is provided to the owner and trainer of the horse. This information should not be the basis for withdrawing a horse from a race. However, it might prove useful to owners and trainers to have this information in order to help them reduce the risk of the start of a racing horse.

Moreover, the results of the logistic regression model, which achieves similar predictive results with the machine learning models, are not based on black box methods but on multivariable logistic regression used for risk factor analysis. This makes it easy to additionally provide an explanation for why a horse is identified as 'high-risk' compared to the average horse based on interpretation of the model variables. We believe this could be helpful information for the owners and trainers to help them reduce the risk of the start of a racing horse.

Finally, we are not arguing that there is necessarily a causal link between the variables used to train the models and equine injuries. Variables that were used in the models which were associated with equine injuries might not be in the future. As the sport evolves and constantly improves so will the models change and adapt to provide useful predictions.

6. General Discussion

6.1. Introduction

The purpose of this chapter is to discuss the strengths and limitations of our study, as well as, highlight the important findings of the study. We also aim to make recommendations for reducing the risk of Thoroughbreds sustaining an injury during flat horse racing in the US and Canada and highlight areas of future research.

Our analysis was based on data provided by The US Jockey Club, for the years 2009 to 2015 and the work was funded through an Industry Partnership PhD provided by The US Jockey Club and the University of Glasgow. This is the first study to make use of the extensive information contained in the EID to identify risk factors associated with equine fatal and fracture injuries in the US and Canada for this period. To our knowledge, this is the largest retrospective observational study investigating the risk of equine fatal and fracture injuries during flat racing in the literature. This is also the first study to train logistic regression and machine learning models to predict equine injuries using such an extensive amount of data and a full year of horse racing events for prediction and evaluation.

6.2. Strengths and Limitations of the Study

One of the strengths of this study is the large number of observations available for analysis. This large size contributes to large power for identifying risk factors even when their impact is small. Furthermore, since for both fatal and fracture injuries are rare outcomes and there is an extreme imbalance in the dataset between cases and controls a large amount of observations provides a large enough amount of cases for statistical analysis.

Furthermore, we believe that the analysis is as representative as possible, since we have included in the statistical analysis 90% of racing starts from all official racing in the US and Canada for that period. A small source of bias could be the roughly 10% of starts which are not included in the study.

One of the limitations of the study is that the EID does not contain information on the medical history of each horse.

Another limitation of the study is that it includes only equine injuries recorded at the race track. Injuries that might have been found afterwards would not have been recorded and accounted for in our analysis. Furthermore, the EID does not contain any information on equine injuries sustained during training.

It is important to note that we did not make any attempt to differentiate the causes of fatal injury or the different types of fracture injury in the present study. Risk factors vary among types of fractures and it is likely that some of those risk factors were not identified in the present study. The types of injuries sustained and the reason for euthanasia have been accurately reported to the EID only recently. Thus, future analyses will be able to use more specific outcome variables to identify risk factors associated with the most common reasons for euthanasia of Thoroughbred racehorses following race-induced injuries.

Statistical significance does not necessarily translate to clinical significance. Although we identified several risk factors that were significantly associated with fracture injuries in Thoroughbred horses competing in flat racing, it is important to point out that the vast majority of race starts evaluated in the present study did not result in a fracture injury. Finally, because of the extremely large number of race starts evaluated and the resulting high statistical power of this study, the magnitude of effect for some of the risk factors was very small.

6.3. Important Findings of the Study

In Chapter 2 we estimated the annual average risk of fatal and fracture equine injuries for the period 2009 - 2015. We found that out of all fatalities 83.1% were fractures, 16.4% were soft tissue injuries, 15.2% were joint injuries and 7.5% were non-musculoskeletal injuries. Looking at fracture injuries we showed that in this 7-year period, out of all fractures sustained during racing 90.8% were fractures of the distal limb, 4.8% were fractures of the proximal bone and 2.6% were fractures of the axial skeleton. Furthermore, 74.9% of fractures resulted in fatalities. Finally, for each variable we considered as a possible risk factor, we showed the distribution of starts across its range of values, along with the number of fatal and fracture injuries sustained per 1000 starts.

In Chapter 3 we investigated the association between possible risk factors and fatal injuries. We identified 17 risk factors significantly associated with fatal injuries using all available starts. These risk factors include:

- the age of the horse at the beginning of its career
- the country the race was held
- the horse having entered the veterinarian's list sometime in its career
- Thoroughbreds racing for the first time
- Thoroughbreds racing in races with a purse lower or equal to $7500
- The number of layups horses had in their career.
- The number of previous injuries a horse had sustained during a racing start
- The odds rank of the horse
- The post position
- The distance of the race
- the season
- the sex of the horse
- the surface of the race track
- the average time between racing starts

- the time a horse has spent with the same trainer
- the size of the race track
- The first trainer of the horse

Furthermore, we assessed risk factors that summarize historical racing information prior to each race, using a sub-sample of the population consisting of all the starts from horses six months after their first recorded racing or exercise start. We identified 21 risk factors significantly associated with fatal injuries using those starts. These risk factors that were identified with this sample of the population that were not identified when using all starts include:

- Accumulated distance ran on racing starts
- Number of racing or exercise starts 30 days prior the race
- Number of racing or exercise starts 30 to 60 days prior the race
- Number of racing or exercise starts 90 to 180 days prior the race
- Number of racing starts 90 to 180 days prior the race
- Time spent in layup
- Races won per racing start 30 to 60 days prior the race
- Races won per racing start 90 to 180 days prior the race

In Chapter 4 we investigated the association between possible risk factors and fracture injuries. We identified 16 risk factors significantly associated with fracture injuries using all available starts. These risk factors include:

- the country the race was held
- the horse having entered the veterinarian's list sometime in its career
- Thoroughbreds racing for the first time
- Thoroughbreds racing in races with a purse lower or equal to $7500
- The number of layups horses had in their career.
- The number of previous injuries a horse had sustained during a racing start

- The number of non-veterinary scratches a horse had in its career
- The odds rank of the horse
- The distance of the race
- the season
- the sex of the horse
- the surface of the race track
- the average time between racing stars
- the time a horse has spent with the same trainer
- Time spent in layup
- The first trainer of the horse

Furthermore, we assessed risk factors that summarize historical racing information prior to each race, using a sub-sample of the population consisting of all the starts from horses six months after their first recorded racing or exercise start. We identified 22 risk factors significantly associated with fracture injuries using those starts. These risk factors that were identified with this sample of the population that were not identified when using all starts include:

- Accumulated distance ran on racing and exercise starts
- the age of the horse at the beginning of its career
- The time a horse has participated in racing
- Number of racing or exercise starts 30 days prior the race
- Number of racing or exercise starts 60 to 90 days prior the race
- Number of racing or exercise starts 90 to 180 days prior the race
- Number of racing starts 30 to 60 days prior the race
- the average time between exercise stars
- Races won per racing start 30 to 60 days prior the race
- Races won per racing start 90 to 180 days prior the race

Finally, in Chapter 5 we trained models to predict the risk of Thoroughbred sustaining a fatal or fracture injury during flat racing. We were able to train models that achieved statistically significant AUC scores of more than 62% and could identify a population of starts with more than 2 times the average risk of 2015 and a population of starts with a risk of less than half the average risk of 2015 for both fatal and fracture injuries.

We also showed that different machine learning models, ranging from random forest techniques to various implementations of artificial neural networks, can successfully be used to predict equine injuries. Furthermore, we showed that deep learning models can be used to achieve good results in difficult, imbalanced and complex, though data-rich, environments outside their usual realm of image classification and object recognition.

Finally, a very interesting finding regarding our predictive models is that all models trained were of similar predictive ability. This means that the logistic regression models that are commonly used in the industry could serve as the predictive models in an implementation system aimed at identifying horses at high risk without having to resort to the more complicated, computer-intensive machine learning techniques. Furthermore, since all different techniques performed at a similar level this might be an indication that we have achieved the highest possible predictive result that can be learned based on the available data.

6.4. Recommendations

Based on the results of this study, in order to further minimise the risk of equine injuries during flat racing in the US and Canada, we recommend that:

- Synthetic surfaces should be retained and where possible dirt surfaces should be replaced with turf and synthetic surfaces.
- Older horses should be considered for retirement earlier.
- Stallions should be considered for retirement and breeding earlier.
- Stallions that are not planned to be used for breeding should be considered for gelding.
- Thoroughbreds should not abruptly resume racing after a long time of absence without any exercise starts in the month prior the race.
- Extra care and consideration is given to Thoroughbreds that have already sustained an injury during their career or ever entered the veterinarian's list.
- The balance between subclinical bone damage and adaptation is considered by trainers as the risk of equine injuries was found to be lower for the number of racing and exercise starts a horse had in the periods 30 days and 30 to 60 days prior the race but higher in the periods of 60 to 90 and 90 to 180 days prior a race.
- Extra care and consideration is given to highly competitive horses that participate in races with higher than $7500 purse and are expected to perform well and win the race.
- An online predictive system that uses a multivariable logistic regression model is established at each participating track. As the predictive ability of a model is relatively low and the risk for even a high-risk horse would be well below 1%; we suggest that the system is used to provide information to the owners, trainers and the veterinarians on the track. However, before doing so it would be useful for the Jockey Club to consult and provide guidance on what steps stakeholders should consider when presented with a horse at higher risk of fatal or non-fatal injury.

- If possible, the EID is expanded to gather information on the medical histories of each racing horse as well as the medication they might be on at the time of the race.
- Further research is conducted for identifying risk factors for fatal and fracture injuries incorporating data from coming years as they become available.
- Further research is conducted for identifying risk factors for other type of injuries beyond fatal and fracture injuries, such as, soft tissue injuries, joint injuries and non-musculoskeletal injuries such as pulmonary hemorrhage and epistaxis.

7. Appendix

7.1. Descriptive statistics for risk factors for fatal injury

Table 7-1 Descriptive statistics for numerical risk factors possible associated with fatal injuries in Thoroughbred racehorses competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Controls - Cases	Controls mean (95% CI)	Cases mean (95% CI)
Age (years)	2,489,358 – 4,599	4.548 (4.546 - 4.55)	4.609 (4.565 - 4.653)
Age at first start (years)	2,489,358 – 4,599	3.297 (3.296 - 3.298)	3.391 (3.358 - 3.424)
Field size	2,489,358 – 4,599	8.448 (8.446 - 8.451)	8.485 (8.430 - 8.540)
Months since last racing start	2,489,358 – 4,599	1.227 (1.224 - 1.229)	1.224 (1.165 - 1.284)
Months since last racing or exercise start	2,489,358 – 4,599	0.478 (0.477 - 0.478)	0.564 (0.549 - 0.578)
No. of layups	2,489,358 – 4,599	1.040 (1.039 - 1.042)	0.972 (0.940 - 1.005)
No. of previous injuries	2,489,358 – 4,599	0.029 (0.029 - 0.029)	0.047 (0.040 - 0.053)
No. of previous vet scratches	2,489,358 – 4,599	0.406 (0.405 - 0.407)	0.462 (0.438 - 0.486)
No. of previous non-vet scratches	2,489,358 – 4,599	1.005 (1.003 - 1.007)	1.066 (1.014 - 1.118)
Odds at start of race	2,489,358 – 4,599	17.323 (17.297 - 17.349)	15.049 (14.503 - 15.595)
Odds rank in race	2,489,358 – 4,599	4.725 (4.722 - 4.728)	4.331 (4.255 - 4.408)
Post position	2,489,358 – 4,599	4.724 (4.721 - 4.727)	4.787 (4.709 - 4.864)
Purse ($1000)	2,489,358 – 4,599	24.965 (24.888 - 25.043)	21.821 (20.5 - 23.141)
Race distance (furlongs)	2,489,358 – 4,599	6.717 (6.716 - 6.719)	6.579 (6.541 - 6.618)
Time between exercise starts – avg (months)	2,489,358 – 4,599	1.262 (1.261 - 1.264)	1.289 (1.25 - 1.329)
Time between exercise starts - active – avg (months)	2,489,358 – 4,599	0.911 (0.91 - 0.913)	0.945 (0.916 - 0.973)
Time between racing starts – avg (months)	2,489,358 – 4,599	2.053 (2.05 - 2.056)	2.24 (2.157 - 2.323)
Time between racing starts – active - avg (months)	2,489,358 – 4,599	1.403 (1.402 - 1.405)	1.525 (1.488 - 1.562)
Time in racing – active (months)	2,489,358 – 4,599	13.3 (13.287 - 13.313)	13.348 (13.056 - 13.64)
Time in racing (months)	2,489,358 – 4,599	18.57 (18.552 - 18.588)	18.501 (18.094 - 18.908)

Table 7-1 (Continued)

Risk factor	Controls - Cases	Controls mean (95% CI)	Cases mean (95% CI)
Time with same jockey (months)	2,489,358 – 4,599	0.882 (0.879 - 0.884)	0.819 (0.764 - 0.873)
Time with same trainer (months)	2,489,358 – 4,599	6.904 (6.893 - 6.915)	5.807 (5.58 - 6.033)
Track size (furlongs)	2,489,358 – 4,599	7.752 (7.751 - 7.753)	7.668 (7.636 - 7.699)

Table 7-2 Descriptive statistics for categorical risk factors possible associated with fatal injuries in Thoroughbred racehorses competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Controls (%)	Cases (%)
Country		
Canada	200,465 (8)	225 (5)
US	2,288,893 (92)	4,374 (95)
Entered the vet list		
No	2,012,776 (81)	3,238 (70)
Yes	476,582 (19)	1,361 (30)
First Start		
No	2,310,661 (93)	4,369 (95)
Yes	178,697 (7)	230 (5)
Low purse race (<= $7500)		
No	2,126,510 (85)	3,965 (86)
Yes	362,848 (15)	634 (14)
Season		
Autumn	646,489 (26)	1247 (27)
Spring	610,600 (25)	1,058 (23)
Summer	775,590 (31)	1,335 (29)
Winter	456,679 (18)	959 (21)
Sex		
Mare/Gelding	2,179,652 (88)	3,874 (84)
Stallion	309,706 (12)	725 (16)

Table 7-2 (Continued)

Risk factor	Controls (%)	Cases (%)
Start with new jockey		
No	1,195.620 (48)	2,090 (45)
Yes	1,293,738 (52)	2,509 (55)
Start with new trainer		
No	2,250,492 (90)	4,034 (88)
Yes	238,866 (10)	565 (12)
Surface		
Synthetic	297,211 (12)	363 (8)
Dirt	1,837,893 (74)	3,684 (80)
Turf	354,254 (14)	552 (12)
Training with first trainer		
Yes	1,333,130 (54)	2,229 (48)
No	1,156,228 (46)	2,370 (52)

Table 7-3 Descriptive statistics for numerical risk factors possible associated with fatal injuries in Thoroughbred racehorses, six months after their first recorded racing or exercise start, competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Controls - Cases	Controls mean (95% CI)	Cases mean (95% CI)
Accumulated distance ran in career (Km)	1,958,722 - 3,696	40.519 (40.482 - 40.557)	37.998 (37.185 - 38.812)
Accumulated exercise distance ran in career (Km)	1,958,722 - 3,696	19.635 (19.615 - 19.655)	18.986 (18.525 - 19.447)
Accumulated racing distance ran in career (Km)	1,958,722 - 3,696	20.885 (20.86 - 20.909)	19.013 (18.505 - 19.52)
Age (years)	1,958,722 - 3,696	4.764 (4.762 - 4.766)	4.787 (4.739 - 4.834)
Age at first start (years)	1,958,722 - 3,696	3.211 (3.209 - 3.212)	3.308 (3.274 - 3.342)
Average speed change on previous race (m/s)	1,958,722 - 3,696	-0.007 (-0.009 - -0.005)	0.021 (-0.043 - 0.084
Average speed in previous race (m/s)	1,958,722 - 3,696	15.931 (15.928 - 15.935)	16.099 (16.03 - 16.168)
Field size	1,958,722 - 3,696	8.373 (8.370 - 8.375)	8.435 (8.373 - 8.496)
Months since last racing start	1,958,722 - 3,696	1.389 (1.386 - 1.392)	1.353 (1.28 - 1.426)

Table 7-3 (Continued)

Risk factor	Controls - Cases	Controls mean (95% CI)	Cases mean (95% CI)
Months since last racing or exercise start	1,958,722 – 3,696	0.500 (0.499 - 0.501)	0.584 (0.567 - 0.601)
No. of layups	1,958,722 – 3,696	1.311 (1.309 - 1.313)	1.198 (1.162 - 1.235)
No. of previous injuries	1,958,722 – 3,696	0.036 (0.036 - 0.036)	0.057 (0.048 - 0.065)
No. of previous vet scratches	1,958,722 – 3,696	0.499 (0.498 - 0.500)	0.555 (0.526 - 0.584)
No. of previous non-vet scratches	1,958,722 – 3,696	1.228 (1.225 - 1.230)	1.267 (1.204 - 1.329)
No. of racing and exercise starts (Present – 30 days prior race)	1,958,722 – 3,696	1.867 (1.865 - 1.868)	1.538 (1.508 - 1.569)
No. of racing and exercise starts (30 -60 days prior race)	1,958,722 – 3,696	2.002 (2.001 - 2.004)	1.854 (1.82 - 1.889)
No. of racing and exercise starts (60 -90 days prior race)	1,958,722 – 3,696	1.856 (1.854 - 1.858)	1.857 (1.819 - 1.894)
No. of racing and exercise starts (90 -180 days prior race)	1,958,722 – 3,696	4.794 (4.789 - 4.798)	5.224 (5.134 - 5.314)
No. of starts (Present – 30 days prior race)	1,958,722 – 3,696	0.805 (0.804 - 0.806)	0.734 (0.713 - 0.755)
No. of starts (30 – 60 days prior race)	1,958,722 – 3,696	0.933 (0.932 - 0.934)	0.970 (0.946 - 0.994)
No. of starts (60 – 90 days prior race)	1,958,722 – 3,696	0.834 (0.833 - 0.835)	0.91 (0.885 - 0.936)
No. of starts (90 – 180 days prior race)	1,958,722 – 3,696	2.035 (2.033 - 2.038)	2.298 (2.243 - 2.352)
Odds at start of race	1,958,722 – 3,696	16.81 (16.781 - 16.839)	14.39 (13.805 - 14.975)
Odds rank in race	1,958,722 – 3,696	4.654 (4.650 - 4.657)	4.241 (4.156 - 4.325)
Post position	1,958,722 – 3,696	4.687 (4.683 - 4.690)	4.756 (4.67 - 4.842)
Purse ($1000)	1,958,722 – 3,696	24.363 (24.276 - 24.45)	20.99 (19.588 - 22.392)
Race distance (furlongs)	1,958,722 – 3,696	6.788 (6.786 - 6.790)	6.608 (6.565 - 6.651)
Time between exercise starts - avg (months)	1,958,722 – 3,696	1.413 (1.411 - 1.415)	1.415 (1.368 - 1.462)
Time between exercise starts - active – avg (months)	1,958,722 – 3,696	0.975 (0.974 - 0.977)	0.997 (0.964 - 1.03)
Time between racing starts – avg (months)	1,958,722 – 3,696	2.319 (2.315 - 2.322)	2.48 (2.381 - 2.579)

Table 7-3 (Continued)

Risk factor	Controls - Cases	Controls mean (95% CI)	Cases mean (95% CI)
Time between racing starts – active - avg (months)	1,958,722 – 3,696	1.501 (1.499 - 1.503)	1.598 (1.558 - 1.637)
Time in layup (months)	1,958,722 – 3,696	6.669 (6.659 - 6.678)	6.381 (6.155 - 6.608)
Time in racing – active (months)	1,958,722 – 3,696	16.063 (16.049 - 16.077)	15.809 (15.494 - 16.124)
Time in racing (months)	1,958,722 – 3,696	22.732 (22.713 - 22.751)	22.191 (21.762 - 22.620)
Time with same jockey (months)	1,958,722 – 3,696	1.009 (1.006 - 1.012)	0.913 (0.846 - 0.980)
Time with same trainer (months)	1,958,722 – 3,696	8.417 (8.403 - 8.430)	6.863 (6.593 - 7.133)
Track size (furlongs)	1,958,722 – 3,696	7.741 (7.74 - 7.743)	7.657 (7.622 - 7.691)
Wins/starts (Present – 30 days prior race)	1,958,722 – 3,696	8.349 (8.312 - 8.386)	8.735 (7.848 - 9.621)
Wins/starts (30 – 60 days prior race)	1,958,722 – 3,696	9.24 (9.203 - 9.278)	11.884 (10.919 - 12.85)
Wins/starts (60 – 90 days prior race)	1,958,722 – 3,696	8.241 (8.206 - 8.277)	10.072 (9.178 - 10.965)
Wins/starts (90 – 180 days prior race)	1,958,722 – 3,696	9.173 (9.145 - 9.202)	11.511 (10.792 - 12.23)

Table 7-4 Descriptive statistics for categorical risk factors possible associated with fatal injuries in Thoroughbred racehorses, six months after their first recorded racing or exercise start, competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Controls (%)	Cases (%)
Country		
Canada	153,176 (8)	171 (5)
US	1,805,546 (92)	3,525 (95)
Entered the vet list		
No	1,525,066 (78)	2,471 (67)
Yes	433,656 (22)	1,225 (33)
Low purse race (<= $7500)		
No	1,661,562 (85)	3,195 (86)
Yes	297,160 (15)	501 (14)
Season		
Autumn	526,421 (27)	1,039 (28)
Spring	468,676 (24)	816 (22)
Summer	604,406 (31)	1,067 (29)
Winter	359,219 (18)	774 (21)
Sex		
Mare/Gelding	1,746,643 (89)	3,168 (86)
Stallion	212,079 (11)	528 (14)
Start with new jockey		
No	871,865 (45)	1,571 (43)
Yes	1,086,857 (55)	2,125 (57)
Start with new trainer		
No	1,748,839 (89)	3,184 (86)
Yes	209,883 (11)	512 (14)
Surface		
Synthetic	221,406 (11)	268 (7)
Dirt	1,447,562 (74)	2,977 (81)
Turf	289,754 (15)	451 (12)
Training with first trainer		
Yes	848,807 (43)	1,420 (38)
No	1,109,915 (57)	2,276 (62)

7.2. Descriptive statistics for risk factors for fracture injury

Table 7-5 Descriptive statistics for numerical risk factors possible associated with fracture injuries in Thoroughbred racehorses competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Controls - Cases	Controls mean (95% CI)	Cases mean (95% CI)
Age (years)	2,488,984 – 4,973	4.548 (4.546 - 4.55)	4.477 (4.436 - 4.517)
Age at first start (years)	2,488,984 – 4,973	3.297 (3.296 - 3.299)	3.318 (3.288 - 3.348)
Field size	2,488,984 – 4,973	8.448 (8.446 - 8.451)	8.428 (8.375 - 8.481)
Months since last racing start	2,488,984 – 4,973	1.227 (1.224 - 1.229)	1.175 (1.124 - 1.226)
Months since last racing or exercise start	2,488,984 – 4,973	0.478 (0.477 - 0.478)	0.553 (0.54 - 0.566)
No. of layups	2,488,984 – 4,973	1.040 (1.039 - 1.042)	0.910 (0.88 - 0.94)
No. of previous injuries	2,488,984 – 4,973	0.029 (0.029 - 0.029)	0.043 (0.037 - 0.049)
No. of previous vet scratches	2,488,984 – 4,973	0.406 (0.405 - 0.407)	0.434 (0.412 - 0.457)
No. of previous non-vet scratches	2,488,984 – 4,973	1.005 (1.003 - 1.007)	1.084 (1.032 - 1.136)
Odds at start of race	2,488,984 – 4,973	17.324 (17.298 - 17.351)	14.666 (14.145 - 15.188)
Odds rank in race	2,488,984 – 4,973	4.725 (4.722 - 4.728)	4.254 (4.180 - 4.328)
Post position	2,488,984 – 4,973	4.724 (4.721 - 4.728)	4.704 (4.629 - 4.778)
Purse ($1000)	2,488,984 – 4,973	24.958 (24.881 - 25.035)	25.834 (24.068 - 27.600)
Race distance (furlongs)	2,488,984 – 4,973	6.717 (6.715 - 6.719)	6.640 (6.603 - 6.678)
Time between exercise starts – avg (months)	2,488,984 – 4,973	1.263 (1.261 - 1.264)	1.209 (1.174 - 1.245)
Time between exercise starts - active – avg (months)	2,488,984 – 4,973	0.911 (0.910 - 0.913)	0.899 (0.873 - 0.924)
Time between racing starts - avg (months)	2,488,984 – 4,973	2.053 (2.05 - 2.056)	2.212 (2.135 - 2.289)
Time between racing starts - active - avg (months)	2,488,984 – 4,973	1.403 (1.402 - 1.405)	1.542 (1.506 - 1.579)
Time in racing – active (months)	2,488,984 – 4,973	13.301 (13.288 - 13.314)	13.095 (12.82 - 13.369)
Time in racing (months)	2,488,984 – 4,973	18.572 (18.554 - 18.590)	17.846 (17.469 - 18.224)
Time with same jockey (months)	2,488,984 – 4,973	0.882 (0.879 - 0.884)	0.843 (0.790 - 0.897)
Time with same trainer (months)	2,488,984 – 4,973	6.905 (6.893 - 6.916)	5.655 (5.443 - 5.867)
Track size (furlongs)	2,488,984 – 4,973	7.752 (7.751 - 7.753)	7.772 (7.741 - 7.803)

Table 7-6 Descriptive statistics for categorical risk factors possible associated with fracture injuries in Thoroughbred racehorses competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Controls (%)	Cases (%)
Country		
Canada	200,440 (8)	250 (5)
US	2,288,544 (92)	4723 (95)
Entered the vet list		
No	2,012,424 (81)	3,590 (72)
Yes	476,560 (19)	1,383 (28)
First Start		
No	2,310,330 (93)	4,700 (95)
Yes	178,654 (7)	273 (5)
Low purse race (<= $7500)		
No	2,126,103 (85)	4,372 (88)
Yes	362,881 (15)	601 (12)
Season		
Autumn	646,405 (26)	1,331 (27)
Spring	610,481 (25)	1,177 (24)
Summer	775,483 (31)	1,442 (29)
Winter	456,615 (18)	1,023 (21)
Sex		
Mare/Gelding	2,179,424 (88)	4,102 (82)
Stallion	309,560 (12)	871 (18)
Start with new jockey		
No	1,195,411 (48)	2,299 (46)
Yes	1,293,573 (52)	2,674 (54)
Start with new trainer		
No	2,250,162 (90)	4,364 (88)
Yes	238,822 (10)	609 (12)
Surface		
Synthetic	297,145 (12)	429 (9)
Dirt	1,837,708 (74)	3,869 (78)
Turf	354,131 (14)	675 (14)
Training with first trainer		
Yes	1,332,867 (54)	2,492 (50)
No	1,156,117 (46)	2,481 (50)

Table 7-7 Descriptive statistics for numerical risk factors possible associated with fracture injuries in Thoroughbred racehorses, six months after their first recorded racing or exercise start, competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Controls - Cases	Controls mean (95% CI)	Cases mean (95% CI)
Accumulated distance ran in career (Km)	1,958,464 - 3,954	40.520 (40.482 - 40.557)	38.024 (37.243 - 38.804)
Accumulated exercise distance ran in career (Km)	1,958,464 - 3,954	19.634 (19.614 - 19.654)	19.394 (18.953 - 19.835)
Accumulated racing distance ran in career (Km)	1,958,464 - 3,954	20.886 (20.861 - 20.91)	18.63 (18.138 - 19.121)
Age (years)	1,958,464 - 3,954	4.764 (4.762 - 4.766)	4.660 (4.616 - 4.704)
Age at first start (years)	1,958,464 - 3,954	3.211 (3.209 - 3.212)	3.241 (3.211 - 3.272)
Average speed change on previous race (m/s)	1,958,464 - 3,954	-0.007 (-0.009 - -0.005)	0.010 (-0.049 - 0.07)
Average speed in previous race (m/s)	1,958,464 - 3,954	15.931 (15.928 - 15.935)	16.102 (16.033 - 16.170)
Field size	1,958,464 - 3,954	8.373 (8.370 - 8.376)	8.372 (8.313 - 8.432)
Months since last racing start	1,958,464 - 3,954	1.389 (1.386 - 1.392)	1.294 (1.231 - 1.357)
Months since last racing or exercise start	1,958,464 - 3,954	0.500 (0.499 - 0.501)	0.574 (0.559 - 0.590)
No. of layups	1,958,464 - 3,954	1.311 (1.309 - 1.313)	1.132 (1.098 - 1.166)
No. of previous injuries	1,958,464 - 3,954	0.036 (0.036 - 0.036)	0.053 (0.045 - 0.060)
No. of previous vet scratches	1,958,464 - 3,954	0.499 (0.498 - 0.500)	0.525 (0.498 - 0.552)
No. of previous non-vet scratches	1,958,464 - 3,954	1.228 (1.225 - 1.230)	1.301 (1.239 - 1.364)
No. of racing and exercise starts (Present – 30 days prior race)	1,958,464 - 3,954	1.867 (1.865 - 1.868)	1.571 (1.541 - 1.601)
No. of racing and exercise starts (30 -60 days prior race)	1,958,464 - 3,954	2.002 (2.001 - 2.004)	1.900 (1.866 - 1.933)
No. of racing and exercise starts (60 -90 days prior race)	1,958,464 - 3,954	1.856 (1.854 - 1.858)	1.931 (1.895 - 1.968)
No. of racing and exercise starts (90 -180 days prior race)	1,958,464 - 3,954	4.793 (4.789 - 4.798)	5.455 (5.367 - 5.543)
No. of starts (Present – 30 days prior race)	1,958,464 - 3,954	0.805 (0.804 - 0.806)	0.724 (0.704 - 0.744)

Table 7-7 (Continued)

Risk factor	Controls - Cases	Controls mean (95% CI)	Cases mean (95% CI)
No. of starts (30 – 60 days prior race)	1,958,464 – 3,954	0.933 (0.932 - 0.934)	0.972 (0.949 - 0.995)
No. of starts (60 – 90 days prior race)	1,958,464 – 3,954	0.834 (0.833 - 0.835)	0.925 (0.9 00- 0.949)
No. of starts (90 – 180 days prior race)	1,958,464 – 3,954	2.035 (2.033 - 2.038)	2.301 (2.249 - 2.353)
Odds at start of race	1,958,464 – 3,954	16.811 (16.782 - 16.84)	14.055 (13.49 - 14.621)
Odds rank in race	1,958,464 – 3,954	4.654 (4.650 - 4.658)	4.169 (4.087 - 4.250)
Post position	1,958,464 – 3,954	4.687 (4.683 - 4.690)	4.659 (4.576 - 4.742)
Purse ($1000)	1,958,464 – 3,954	24.355 (24.268 - 24.442)	25.142 (23.171 - 27.113)
Race distance (furlongs)	1,958,464 – 3,954	6.788 (6.786 - 6.790)	6.673 (6.631 - 6.716)
Time between exercise starts – avg (months)	1,958,464 – 3,954	1.413 (1.411 - 1.415)	1.324 (1.282 - 1.367)
Time between exercise starts - active – avg (months)	1,958,464 – 3,954	0.975 (0.974 - 0.977)	0.946 (0.917 - 0.976)
Time between racing starts – avg (months)	1,958,464 – 3,954	2.319 (2.315 - 2.322)	2.474 (2.382 - 2.566)
Time between racing starts – active - avg (months)	1,958,464 – 3,954	1.501 (1.499 - 1.502)	1.639 (1.599 - 1.679)
Time in layup (months)	1,958,464 – 3,954	6.670 (6.660 - 6.679)	5.943 (5.738 - 6.148)
Time in racing – active (months)	1,958,464 – 3,954	16.063 (16.049 - 16.078)	15.612 (15.315 - 15.910)
Time in racing (months)	1,958,464 – 3,954	22.733 (22.714 - 22.752)	21.555 (21.156 - 21.955)
Time with same jockey (months)	1,958,464 – 3,954	1.009 (1.005 - 1.012)	0.946 (0.880 - 1.012)
Time with same trainer (months)	1,958,464 – 3,954	8.417 (8.404 - 8.431)	6.742 (6.486 - 6.997)
Track size (furlongs)	1,958,464 – 3,954	7.741 (7.740 - 7.743)	7.765 (7.730 - 7.800)
Wins/starts (Present – 30 days prior race)	1,958,464 – 3,954	8.348 (8.310 - 8.385)	9.362 (8.478 - 10.245)
Wins/starts (30 – 60 days prior race)	1,958,464 – 3,954	9.24 (9.202 - 9.277)	12.135 (11.193 - 13.078)
Wins/starts (60 – 90 days prior race)	1,958,464 – 3,954	8.24 (8.204 - 8.275)	10.633 (9.741 - 11.525)
Wins/starts (90 – 180 days prior race)	1,958,464 – 3,954	9.172 (9.144 - 9.2)	11.881 (11.169 - 12.593)

Table 7-8 Descriptive statistics for categorical risk factors possible associated with fracture injuries in Thoroughbred racehorses, six months after their first recorded racing or exercise start, competing in flat racing in the United States and Canada during the 7-year period from 2009 to 2015

Risk factor	Controls (%)	Cases (%)
Country		
Canada	153,161 (8)	186 (5)
US	1,805,303 (92)	3,767 (95)
Entered the vet list		
No	1,524,821 (78)	2,716 (69)
Yes	433,643 (22)	1,238 (31)
Low purse race (<= $7500)		
No	1,661,272 (85)	3,485 (88)
Yes	297,192 (15)	469 (12)
Season		
Autumn	526,367 (27)	1,093 (28)
Spring	468,582 (24)	910 (23)
Summer	604,336 (31)	1,137 (29)
Winter	359,179 (18)	814 (21)
Sex		
Mare/Gelding	1,746,498 (89)	3,313 (84)
Stallion	211,966 (11)	641 (16)
Start with new jockey		
No	871,735 (45)	1,701 (43)
Yes	1,086,729 (55)	2,253 (57)
Start with new trainer		
No	1,748,612 (89)	3,411 (86)
Yes	209,852 (11)	543 (14)
Surface		
Synthetic	221,336 (11)	308 (8)
Dirt	1,447,429 (74)	3,110 (79)
Turf	289,669 (15)	536 (14)
Training with first trainer		
Yes	848,649 (43)	1,578 (40)
No	1,109,815 (57)	2,376 (60)

8. References

Ackley, D. H. & Hinton, G. E., 1985. A Learning Algorithm for Boltzmann Machines. *Cognitive Science,* pp. 147-169.

Alonso, J.-M.et al., 2012. Determination of future prevention strategies in elite track and field: analysis of Daegu 2011 IAAF Championships injuries and illnesses surveillance. *British Journal of Sports Medicine,* pp. bjsports-2012.

Alpaydin, E., 2014. *Introduction to machine learning.* s.l.:MIT press.

Al-Rfou, R. et al., 2016. Theano: A Python framework for fast computation of mathematical expressions. *arXiv preprint arXiv:1605.02688.*

Amit, Y. & German, D., 1997. Shape quantization and recognition with randomized trees. *Neural Computation,* 9(7), pp. 1545-1588.

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. *American Journal of Veterinary Research,* 68(7), pp. 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.*

Bagley, S. C., White, H. & Golomb, B. A., 2001. Logistic regression in the medical literature:: Standards for use and reporting, with particular attention to one medical domain. *Journal of Clinical Epidemiology,* 54(10), pp. 979-985.

Bailey, C. J. et al., 1998. Flat, hurdle and steeple racing: risk factors for musculoskeletal injury. *Equine Veterinary Journal,* 30(6), pp. 498-503.

Bailey, C. et al., 1997. Risk factors associated with musculoskeletal injuries in Australian Thoroughbred racehorses. *Preventive Veterinary Medicine,* Volume 32, pp. 47-55.

Bengio, Y., 2009. *Learning Deep Architectures for AI (Foundations and Trends in Machine Learning).* s.l.:Now Publishers Inc.

Bengio, Y., 2012. Practical recommendations for gradient-based training of deep architectures. In: *Neural Networks: Tricks of the Trade*. Berlin, Heidelberg: Springer, pp. 437-478.

Bengio, Y. & Delalleau, O., 2009. Justifying and Generalizing Contrastive Divergence. *Neural Computation,* 21(6), pp. 1601-1621.

Bengio, Y. & Delalleau, O., 2011. On the Expressive Power of Deep Architectures. In: *Algorithmic Learning Theory*. Berlin Heidelberg: Springer, pp. 18-36.

Bhadesia, H. K. D. H., 1999. Neural Networks in Materials Science. *ISIJ International,* pp. 966-979.

Boden, L. A. et al., 2007a. Risk factors for Thoroughbred racehorse fatality in flat starts in Victoria, Australia (1989-2004). *Equine Veterinary Journal,* Volume 39, pp. 430-437.

Boden, L. et al., 2006. Risk of fatality and causes of death of Thoroughbred horses associated with racing in Victoria, Australia: 1989–2004. *Equine Veterinary Journal,* 38(4), pp. 312-318.

Boden, L. et al., 2007b. Risk factors for Thoroughbred racehorse fatality in jump starts in Victoria, Australia (1989–2004). *Equine Veterinary Journal,* 39(5), pp. 422-428.

Bourke, J. M., 1994. *Fatalities on racecourses in Victoria. A seven year study*. Stockholm, s.n., pp. 265-268.

Bozdogan, H., 1987. Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical expressions. *Psychometrica,* 52(3), pp. 345-370.

Bradley, A. P., 1997. The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognition,* 30(7), pp. 1145-1159.

Breiman, L., 1996. Bagging predictors. *Machine Learning,* 24(2), pp. 123-140.

Breiman, L., 2001. Random forests. *Machine Learning,* Volume 45, pp. 5-32.

Breiman, L., 2004. *Consistency For a Simple Model of Random Forests,* California: University of California at Berkeley.

Breiman, L., Friedman, J., Stone, C. J. & Olshen, R., 1984. *Classification and Regression Trees.* Boca Raton; London; New York; Washington DC: Chapman & Hall/CRC.

Carrier, T. K. et al., 1998. Association between long periods without high-speed workouts and risk of complete humeral or pelvic fracture in thoroughbred racehorses: 54 cases (1991-1994). *Journal of the American Veterinary Medical Association,* 212(10), pp. 1582-1587.

Chen, C., Liaw, A. & Breiman, L., 2004. *Using Random Forest to Learn Imbalanced Data,* s.l.: University of California Berkeley .

Ciresan, D. C., Meier, U., Gambardella, L. M. & Schmidhuber, J., 2010. *Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition,* s.l.: arXiv.

Ciresan, D., Meier, U. & Schmidhuber, J., 2012. *Multi-column Deep Neural Networks for Image Classification.* s.l., s.n.

Cohen, N. D. et al., 2000. Association of high-speed exercise with racing injury in Thoroughbreds. *Journal of the American Veterinary Medical Association,* 216(8), pp. 1273-1278.

Cohen, N. et al., 1997. Racing-related factors and results of prerace physical inspection and their association with musculoskeletal injuries incurred in thoroughbreds during races.. *Journal of the American Veterinary Medical Association,* 211(4), pp. 454-463.

Cook, R. D. & Weisberg, S., 1982. *Residuals and influence in regression.* New York: Chapman and Hall.

Development Core Team (2008). R: A language and environment for statisticalcomputing. R Foundation for Statistical Computing, Vienna, Austria. ISBN3-900051-07-0, URL http://www.R-project.org.

Dohoo, I., Martin, W. & Stryhn, H., 2003. *Veterinary Epidemiologic Research.* Charlottetown: AVC Inc..

Dunne, R. A. & Campbell, N. A., 1997. *On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function*. Melbourne, s.n.

Efron, B. and Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, pp.54-75.

Ely, E. R. et al., 2009. Descriptive epidemiology of fracture, tendon and suspensory ligament injuries in National Hunt racehorses in training. *Equine Veterinary Journal,* 41(4), pp. 372-378.

Enthought Canopy (Version 1.4.1) [Software]., 2014. Retrieved from http://www.enthought.com

Erhan, D., Courville, A., Bengio, Y. & Vincent, P., 2010. Why Does Unsupervised Pre-training Help Deep Learning?. *The Journal of Machine Learning Research,* Volume 11, pp. 625-660.

Erhan, D. et al., 2009. *The Difficulty of Training Deep Architectures and the Effect of Unsupervised Pre-Training*. Clearwater, s.n., pp. 153-160.

Estberg, L. et al., 1995. Cumulative racing-speed exercise distance cluster as a risk factor for fatal musculoskeletal injury in Thoroughbred racehorses in California. *Preventive Veterinary Medicine,* 24(4), pp. 253-263.

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. *Preventive Veterinary Medicine,* Volume 33, pp. 159-170.

Estberg, L. et al., 1996a. High-speed exercise history and catastrophic racing fracture in thoroughbreds. *Journal of the American Veterinary Medical Association,* 57(11), pp. 1549-1555.

Estberg, L. et al., 1996b. Fatal musculoskeletal injuries incurred during racing and training in thoroughbreds. *Journal of the American Veterinary Medical Association,* 208(1), pp. 92-96.

Estberg, L. et al., 1998b. Relationship between race start characteristics and risk of catastrophic injury in thoroughbreds: 78 cases (1992). *Journal of the American Veterinary Medical Association,* 212(4), pp. 544-549.

Frost, H. M., 1983. The minimum effective strain: A determinant of bone architecture. *Clinical Orthopaedics and Related Research,* Volume 175, pp. 286-292.

Frost, H. M., 1987. Bone “mass” and the “mechanostat”: A proposal. *The anatomical record,* 219(1), pp. 1-9.

Glorot, X. & Bengio, Y., 2010. *Understanding the difficulty of training deep feedforward neural networks.* Sardinia, s.n., pp. 249-256.

Henley, W., Rogers, K., Harkins, L. & Wood, J. L., 2006. A comparison of survival models for assessing risk of racehorse fatality.. *Preventive Veterinary Medicine,* Volume 74, pp. 3-20.

Hernandez, J. A. et al., 2005. Evaluation of horseshoe characteristics and high-speed exercise history as possible risk factors for catastrophic musculoskeletal injury in Thoroughbred racehorses. *American Journal of Veterinary Research,* Volume 66, pp. 1314-1320.

Hernandez, J., Hawkins, D. L. & Scollay, M. C., 2001. Race-start characteristics and risk of catastrophic musculoskeletal injury in Thoroughbred racehorses. *Journal of the American Veterinary Medical Association,* Volume 218, pp. 83-86.

Hill, A. E. et al., 2001. Risk factors for and outcomes of noncatastrophic suspensory apparatus injury in Thoroughbred racehorses. *Journal of the American Veterinary Medical Association,* 218(7), pp. 1136-1144.

Hill, T., Carmichael, D. & Maylin, G. K. L., 1986. Track condition and racing injuries in thoroughbred horses. *The Cornell Veterinarian,* 76(4), pp. 361-379.

Hinton, G., 2013. Where Do Features Come From?. *Cognitive Science,* pp. 1-24.

Hinton, G. E., Osindero, S. & Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. *Neural computation,* 18(7), pp. 1527-1554.

Hinton, M. A. C.-P. G. E., 2005. *On Contrastive Divergence Learning.* Barbados, s.n., pp. 33-40.

Hosmer, D. W. & Lemeshow, S., 2000. *Applied Logistic Regression.* 2nd ed. New York: John Wiley and Sons.

Hothorn, T., Hornik, K. and Zeileis, A., 2006. Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical statistics, 15(3), pp.651-674.

Johnson, B. et al., 1994. Causes of death in racehorses over a 2 year period. *Equine Veterinary Journal,* 26(4), pp. 327-330.

Kane, A. J. et al., 1998. Hoof size, shape, and balance as possible risk factors for catastrophic musculoskeletal injury of Thoroughbred racehorses. *American Journal of Veterinary Research,* Volume 59, pp. 1545-1552.

Kane, A. J. et al., 1996. Horseshoe characteristics as possible risk factors for fatal musculoskeletal injury of thoroughbred racehorses. *American Journal of Veterinary Research,* 57(8), pp. 1147-1152.

Kawcak, C. E. et al., 2000. Clinical effects of exercise on subchondral bone of carpal and metacarpophalangeal joints in horses. *American Journal of Veterinary Research,* 61(10), pp. 1252-1258.

Krook, L. & Maylin, G. A., 1988. Fractures in Thoroughbred race horses. *The Cornell veterinarian,* 78(2), pp. suppl-1.

Lanyon, L. E., 1982. Mechanical function and bone remodeling. In: G. Sumner-Smith, ed. *Bone in Clinical Orthopaedics.* Philadelphia: W. B. Saunders, pp. 273-304.

Lanyon, L. E. & Baggott, D. G., 1976. Mechanical function as an influence on the structure and form of bone. *Bone & Joint Journal,* 58(4), pp. 436-443.

Larochelle, H., Bengio, Y., Louradour, J. & Lamblin, P., 2009. Exploring Strategies for Training Deep Neural Networks. *Journal of Machine Learning Research,* pp. 1-40.

Larochelle, H. et al., 2007. *An Empirical Evaluation of Deep Architectures on Problems with Many Factors of Variation.* Corvallis, s.n.

LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P., 1998. *Gradient-Based Learning Applied to Document Recognition.* s.l., s.n.

LeCun, Y., Bottou, L., Orr, G. B. & Muller, K.-R., 2012. Efficient BackProp. In: *Neural networks: Tricks of the trade*. Berlin, Heidelberg: Springer, pp. 9-48.

Lee, H., Grosse, R., Ranganath, R. & Ng, A. Y., 2009. *Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations.* Montreal, s.n.

LeRoux, N. & Bengio, Y., 2008. Representational Power of Restricted Boltzmann Machines and Deep Belief Networks. *Neural Computation,* 20(6), pp. 1631-1649.

Liaw, A. & Wiener, M., 2002. Classification and regression by randomForest. *R news,* 2(3), pp. 18-22.

Liu, X.-Y., Wu, J. & Zhou, Z.-H., 2008. Exploratory Undersampling for Class-Imbalance Learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics.*

Loitz, B. & Zernicke, R., 1992. Strenuous exercise-induced remodelling of mature bone: relationships between in vivo strains and bone mechanics. *Journal of experimental biology,* 170(1), pp. 1-18.

Lyle, C. H. et al., 2012. Risk factors for race-associated sudden death in Thoroughbred racehorses in the UK (2000–2007). *Equine Veterinary Journal,* 44(4), pp. 459-465.

Lyle, C. et al., 2011. Sudden death in racing Thoroughbred horses: an international multicentre study of post mortem findings. *Equine Veterinary Journal,* 43(3), pp. 324-331.

McKee, S. L., 1995. An update on racing fatalities in the UK. *Equine Veterinary Education,* 7(4), pp. 202-204.

Mohammed, H. O., Hill, T. & Lowe, J., 1991. Risk factors associated with injuries in Thoroughbred horses. *Equine Veterinary Journal,* 23(6), pp. 445-448.

Nunamaker, D. M., Butterweck, D. M. & Provst, M. T., 1990. Fatigue Fractures in Thoroughbred Racehorses: Relationship with Age, Peak Bone Strain, and Training. *Journal of Orthopaedic Research,* Volume 8, pp. 604-611.

Parkin, T. D. et al., 2005. Risk factors for fatal lateral condylarfracture of the third metacarpus/metatarsus in UK racing. *Equine Veterinary Journal,* Volume 37, pp. 192-199.

Parkin, T. D. et al., 2006. Analysis of horse race videos to identify intra-race risk factors for fatal distal limb fracture. *Preventive Veterinary Medicine ,* Volume 74, pp. 44-55.

Parkin, T. D. H. et al., 2004a. Horse-level risk factors for fatal distal limb fracture in racing Thoroughbreds in the UK. *Equine Veterinary Journal,* 36(6), pp. 513-519.

Parkin, T. D. H. et al., 2004b. Race- and course -level risk factors for fatal distal limb fracture in racing Thoroughbreds. *Equine Veterinary Journal,* 36(6), pp. 521-526.

Parkin, T.D., Clegg, P.D., French, N.P., Proudman, C.J., Riggs, C.M., Singer, E.R., Webbon, P.M. and Morgan, K.L., 2004c. Risk of fatal distal limb fractures among Thoroughbreds involved in the five types of racing in the United Kingdom. *The Veterinary Record*, 154(16), pp.493-497.

Peloso, J. G., Mundy, G. D. & Cohen, N., 1994. Prevalence of, and factors associated with, musculoskeletal racing injuries of thoroughbreds. *Journal of the American Veterinary Medical Association,* Volume 204, pp. 620-626.

Pinchbeck, G. L. et al., 2004. Horse injuries and racing practices in National Hunt racehorses in the UK: the results of a prospective cohort study. *The Veterinary Journal,* Volume 167, pp. 45-52.

Poole, R. R. & Meagher, D. M., 1990. Pathologic findings and pathogenesis of racetrack injuries. *Veterinary Clinics of North America: Equine Practice,* Volume 6, pp. 1-29.

Priddy, K. L. & Keller, P. E., 2005. *Artificial neural networks: an introduction.* s.l.:SPIE press.

Reardon, R. J., 2013. *An investigation of risk factors associated with injuries to horses undertaking jump racing in Great Britain.* s.l.:Doctoral dissertation, University of Glasgow.

Riggs, C.M., 2002. Fractures–a preventable hazard of racing thoroughbreds?. The Veterinary Journal, 163(1), pp.19-29.

Riggs, C. M. et al., 1993. Mechanical implications of collagen fibre orientation in cortical bone of the equine radius. *Anatomy and Embryology,* 187(3), pp. 239-248.

Riggs, C. M., Whitehouse, G. H. & Boyde, A., 1999a. Pathology of the distal condyles of the third metacarpal and third metatarsal bones of the horse. *Equine Veterinary Journal,* 31(2), pp. 140-148.

Riggs, C. M., Whitehouse, G. H. & Boyde, A., 1999b. Structural variation of the distal condyles of the third metacarpal and third metatarsal bones in the horse. *Equine Veterinary Journal,* 31(2), pp. 130-139.

Rosanowski, S. M., Chang, Y. M., Stirk, A. J. & Verheyen, K. L. P., 2016. Descriptive epidemiology of veterinary events in flat racing Thoroughbreds in Great Britain (2000 to 2013). *Equine veterinary journal,* 49(3), pp. 275-281.

RStudio Team, 2015. RStudio: Integrated Development for R. RStudio Inc Boston,MA http://www.rstudio.com.

Rubin, C. T. & Lanyon, L. E., 1985. Regulation of bone mass by mechanical strain magnitude. *Calcified tissue international,* 37(4), pp. 411-417.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J., 1986. Learning internal representations by error propagation. In: *Parallel distributed processing: explorations in the microstructure of cognition.* Cambridge, MA, USA: MIT Press, pp. 318-362.

Simard, P. Y., Steinkraus, D. & Platt, J. C., 2013. *Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis.* s.l., IEEE Computer Society.

Stephen, J. O. et al., 2003. Risk factors and prevalence of injuries in horses during various types of steeplechase races. *Journal of the American Veterinary Medical Association,* 223(12), pp. 1788-1790.

Stover, S. M. et al., 1992. An association between complete and incomplete stress fractures of the humerus in racehorses. *Equine Veterinary Journal,* 24(4), pp. 260-263.

Takahashi, T. et al., 2001. Frequency of and risk factors of epistaxis associated with exercise-induced pulmonary hemorrhage in horses: 251,609 race starts (1992-1997). *Journal of the American Veterinary Medical Association,* 218(9), pp. 1462-1464.

Takahashi, T., Kasashima, Y. & Ueno, Y., 2004. Association between race history and risk of superficial digital flexor tendon injury in Thoroughbred racehorses. *Journal of the American Veterinary Medical Association,* 225(1), pp. 90-93.

The Jockey Club. 2017. *The Jockey Club.* [ONLINE] Available at: http://www.jockeyclub.com. [Accessed 06 April 2017].

Van Mechelen, W., 1992. Running injuries. *Sports Medicine,* 14(5), pp. 320-335.

Van Rossum, G., 2007, June. Python Programming Language. In USENIX Annual Technical Conference (Vol. 41, p. 36).

Verheyen, K. L. P., Newton, J. R., Price, J. S. & Wood, J. L. N., 2006. A case-control study of factors associated with pelvic and tibial stress fractures in Thoroughbred racehorses in training in the UK. *Preventive Veterinary Medicine,* 74(1), pp. 21-35.

Verheyen, K. L. P. & Wood, J. L. N., 2004. Descriptive epidemiology of fractures occurring in British Thoroughbred racehorses in training. *Equine veterinary journal,* 36(2), pp. 167-173.

Verheyen, K., Price, J., Lanyon, L. & Wood, J., 2006. Exercise distance and speed affect the risk of fracture in racehorses. *Bone,* 39(6), pp. 1322-1330.

Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.-A., 2008. *Extracting and composing robust features with denoising autoencoders.* s.l., ACM, pp. 1096-1103.

Vincent, P. et al., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *The Journal of Machine Learning Research,* Volume 11, pp. 3371-3408.

Williams, D. A., 1987. Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics,* pp. 181-191.

Williams, K. et al., 2008. Regional pulmonary veno-occlusion: a newly identified lesion of equine exercise-induced pulmonary hemorrhage. *Veterinary Pathology,* 45(3), pp. 316-326.

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 Veterinary Journal,* 33(5), pp. 478-486.

Wood, J., Harkins, L. & Rogers, K., 2000. *A retrospective study of factors associated with racehorse fatality on British racecourses from 1990-1999.* s.l., R&W Publications Newmarket, pp. 274-277.

Xie, Y., Li, X., E. N. & Ying, W., 2009. Customer churn prediction using improved balanced random forests. *Expert Systems with Applications,* Volume 36, pp. 5445-5449.

Wolter, K., 2007. Introduction to variance estimation. Springer Science & Business Media.

Zhang, C. & Ma, Y., 2012. *Ensemble Machine Learning.* Vol. 1 ed. New York: Springer.