University of Glasgow

Equine Injury Database – models, risk factors and prediction

Tim.Parkin@Glasgow.ac.uk
@ThoroughbredHN

University of Glasgow

Introduction

- EID since 2008
- Raw descriptive statistics
- Modelling to identify risk factors
- Testing the predictive ability of the models
- The next 12 months

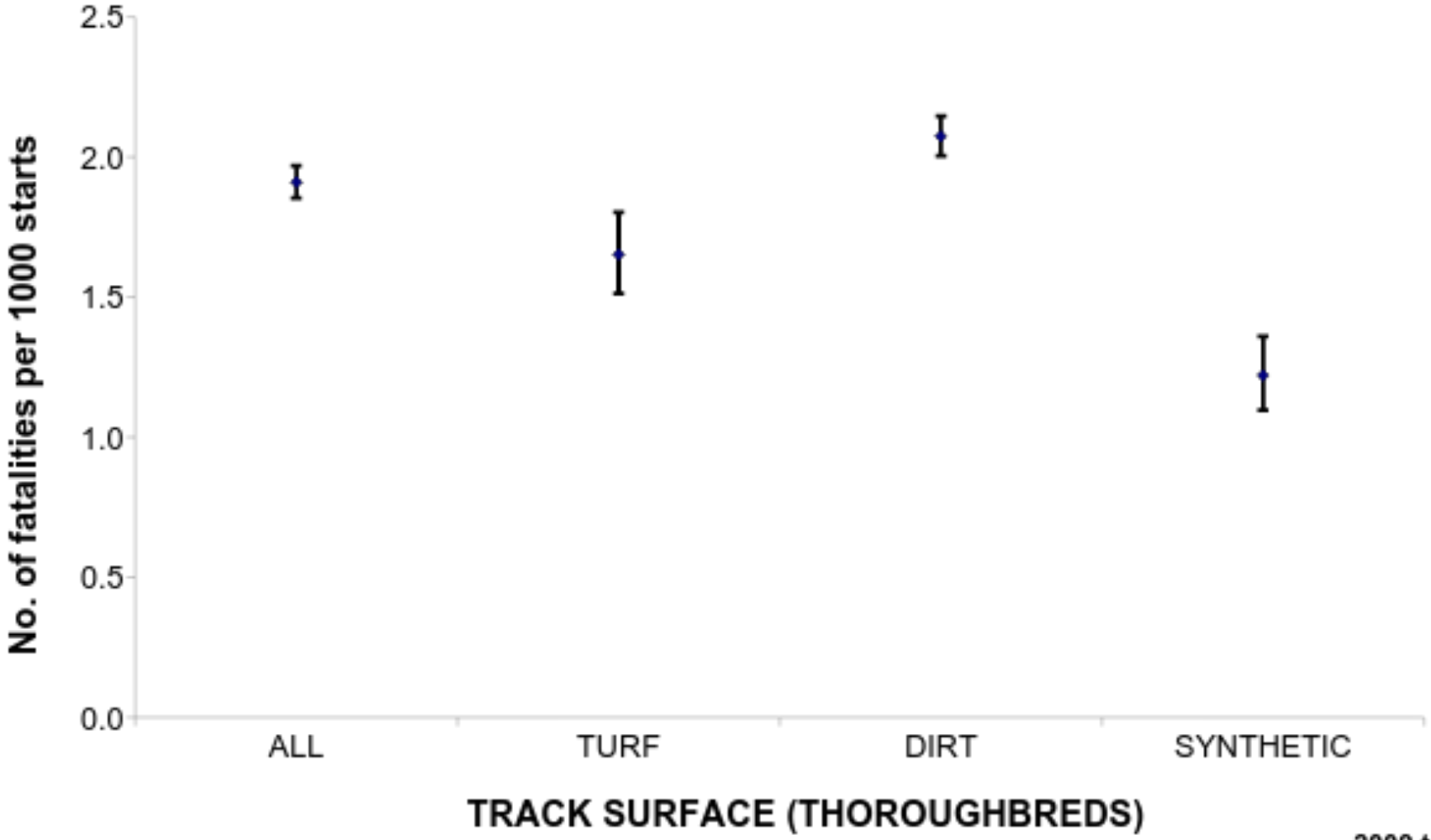
University of Glasgow

Definitions of race day fatalities

- Within 72 hours of race
- Estimates now by calendar year
- Point estimates and 95% confidence intervals
- Now producing multivariable models that account for inter-relationships between variables

University of Glasgow

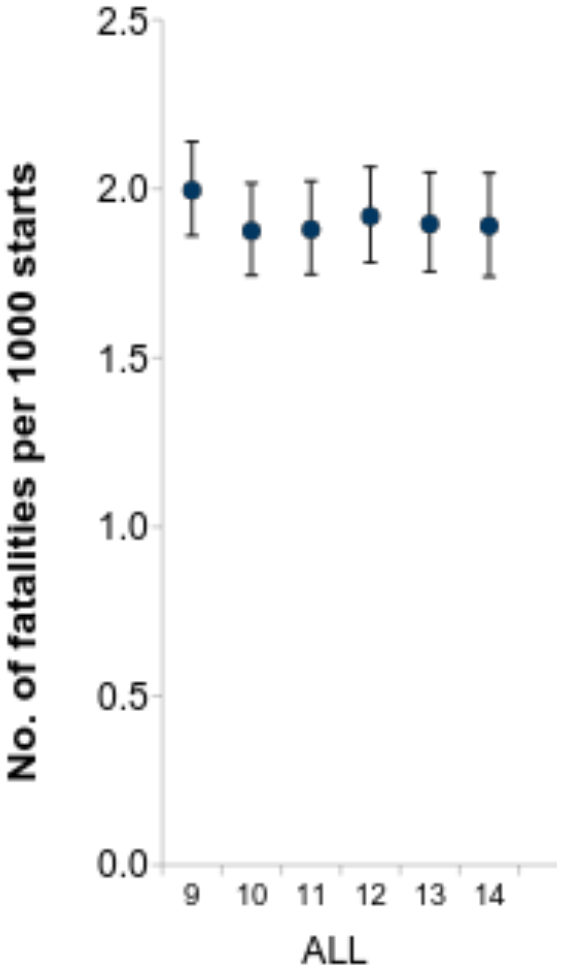
By surface type



2009 to

University of Glasgow

By surface type 2009 - 2014

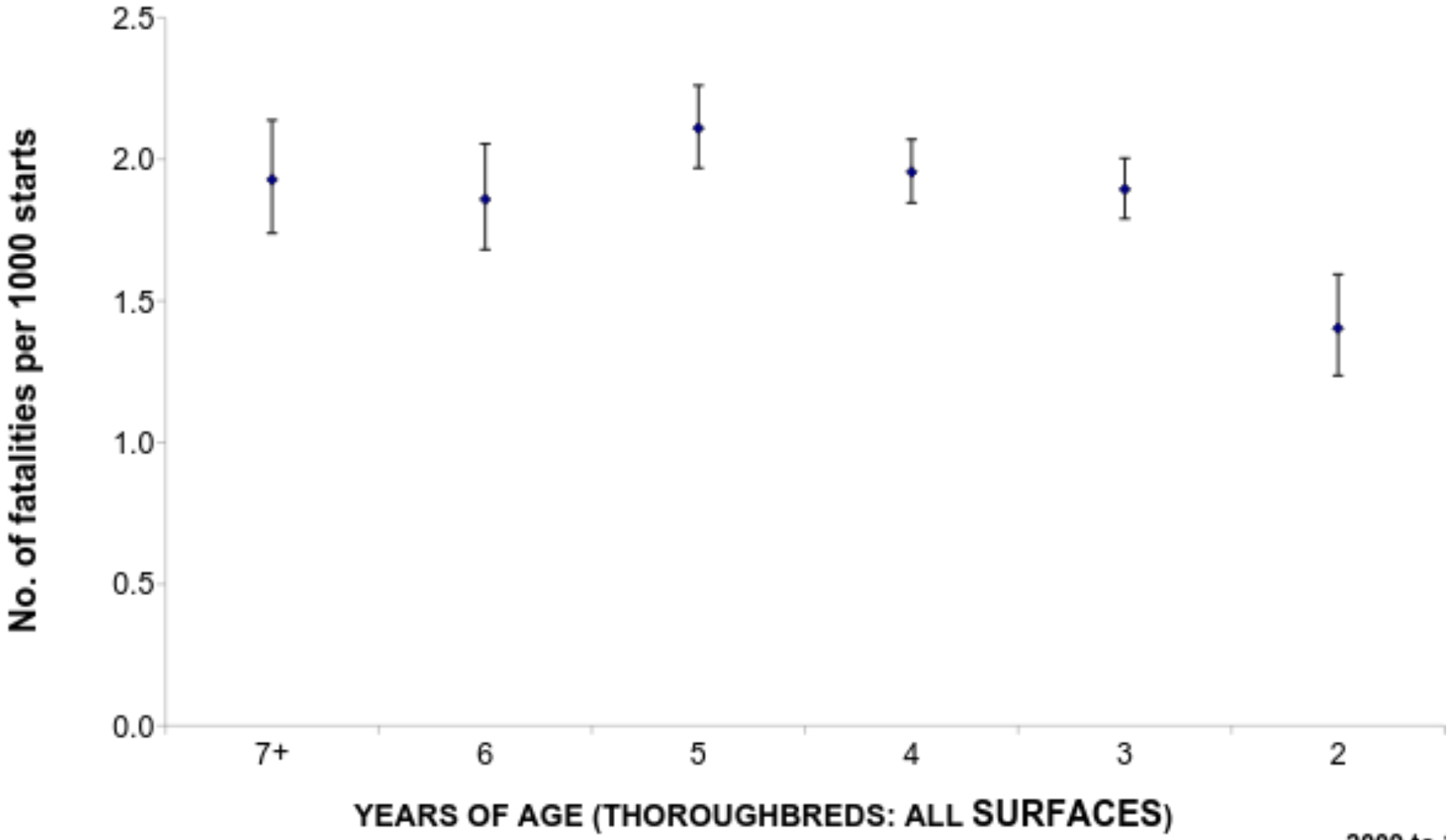


TRACK SURFACE (THOROUGHBREDS)

2009 to 2014

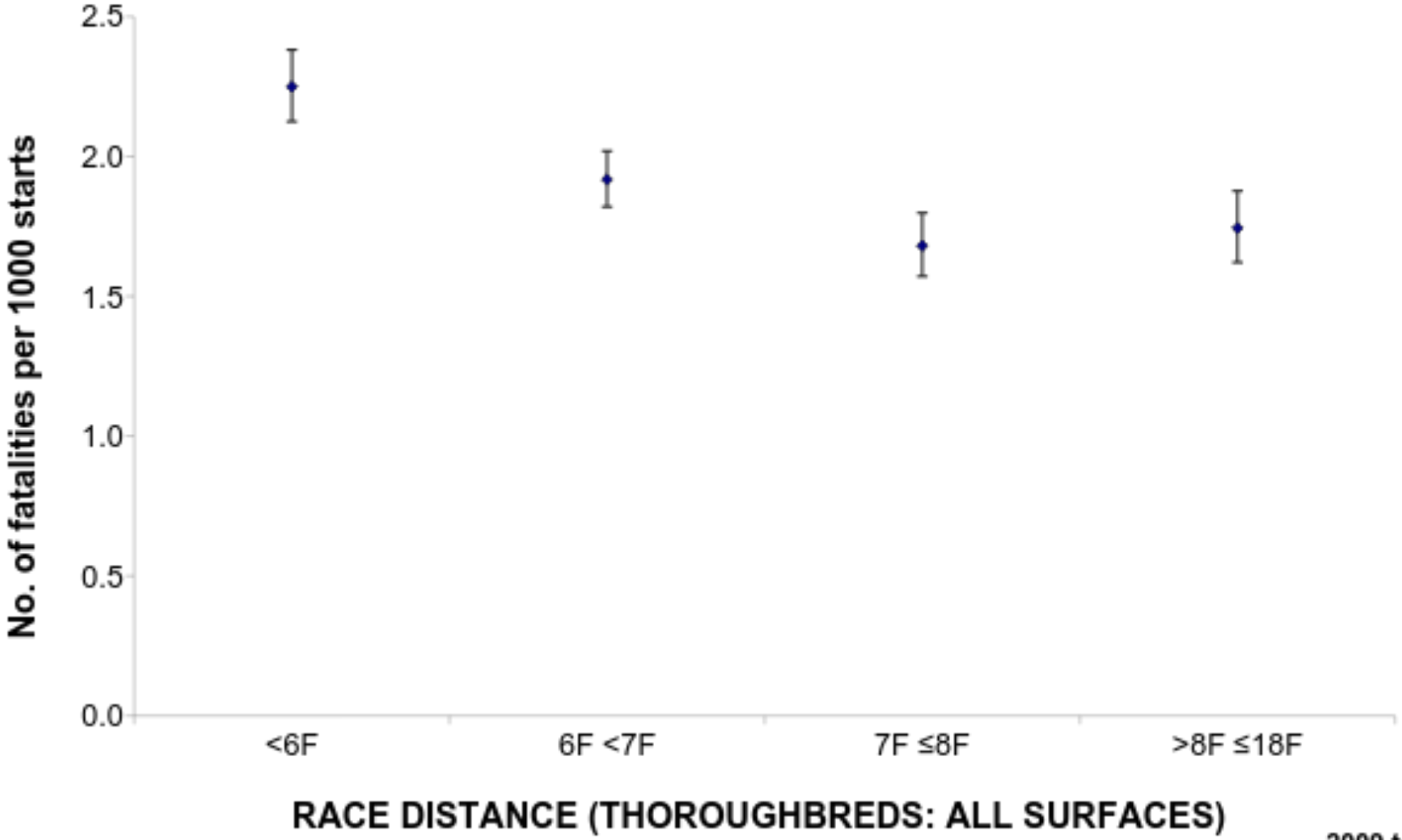
University of Glasgow

By age



University of Glasgow

By race distance



University of Glasgow

Models

- Account for effect of risk factor upon each other and the risk of fatal injury
- National and Track Specific models
- National models built using 6-years of data
  - All races and claiming races only
- Track-specific models for 8 tracks
  - Dependent on sufficient number of starts at these tracks to provide adequate statistical power

University of Glasgow

National and track-specific models

- 2.2 million starts
- 150,000 horses
- 94% of all starts in North America (2009 to 2014)

- A selection of important risk factors:
  - Previous EID injuries
  - Appearance on a vet list
  - Time with same trainer
  - Race distance
  - Surface
  - Previous race history
  - Drop in claim price since previous race
  - Age at first race

University of Glasgow

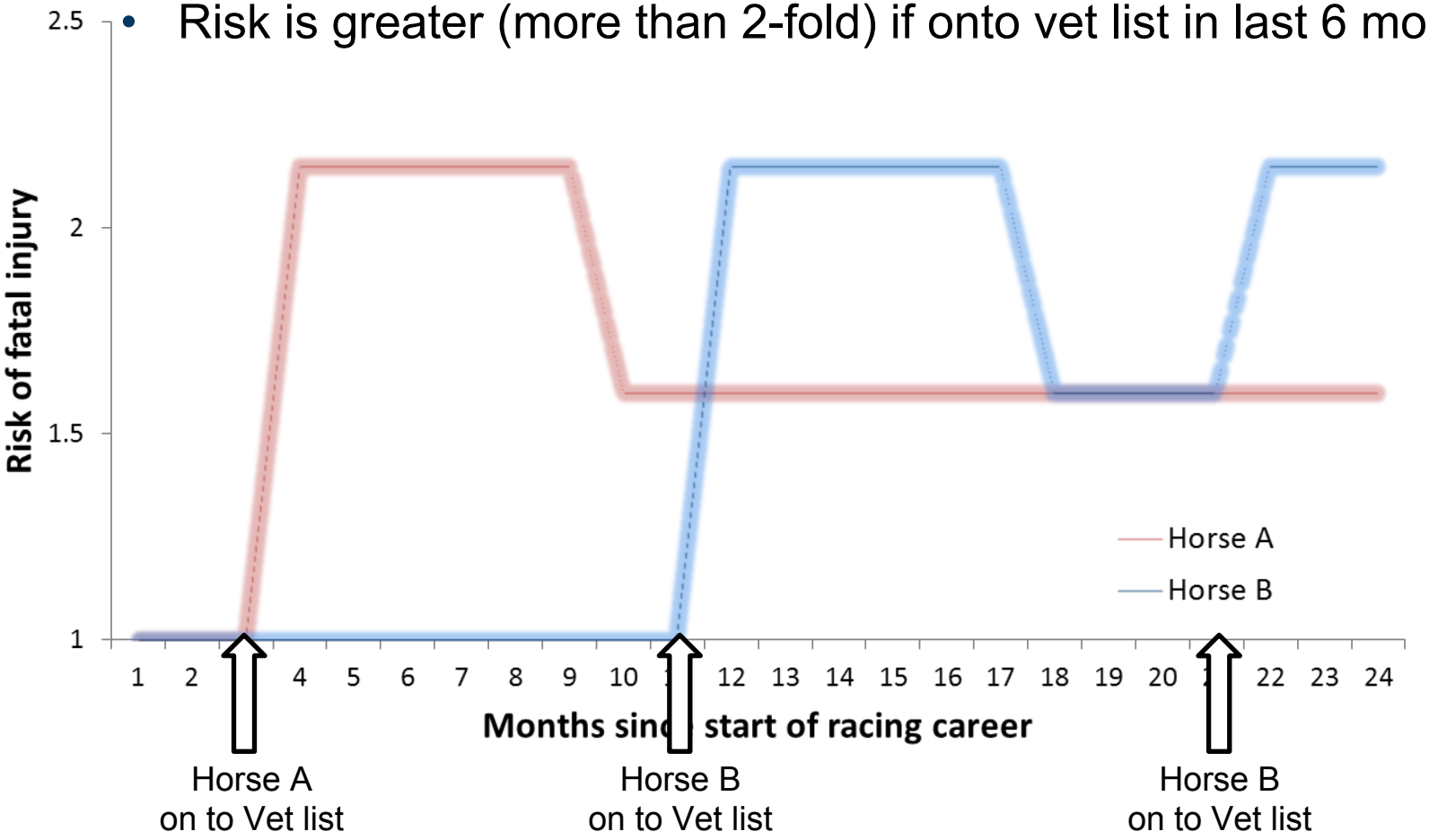
Previous injuries

- Note: Only EID reported injuries
  - Actual relationship could be much bigger
- For every extra previous injury the risk of fatal injury during racing increases by 30%
  - Compared with a horse with no previous EID injury:
    - 1 previous injury – 30% greater risk (about 2% of starts)
    - 2 previous injuries – 70% greater risk (0.1% of starts)
    - 3 previous injuries – 110% greater risk (0.01% of starts)
- Could be much more valuable IF we could include injuries that are not recorded on EID

University of Glasgow

Vet list

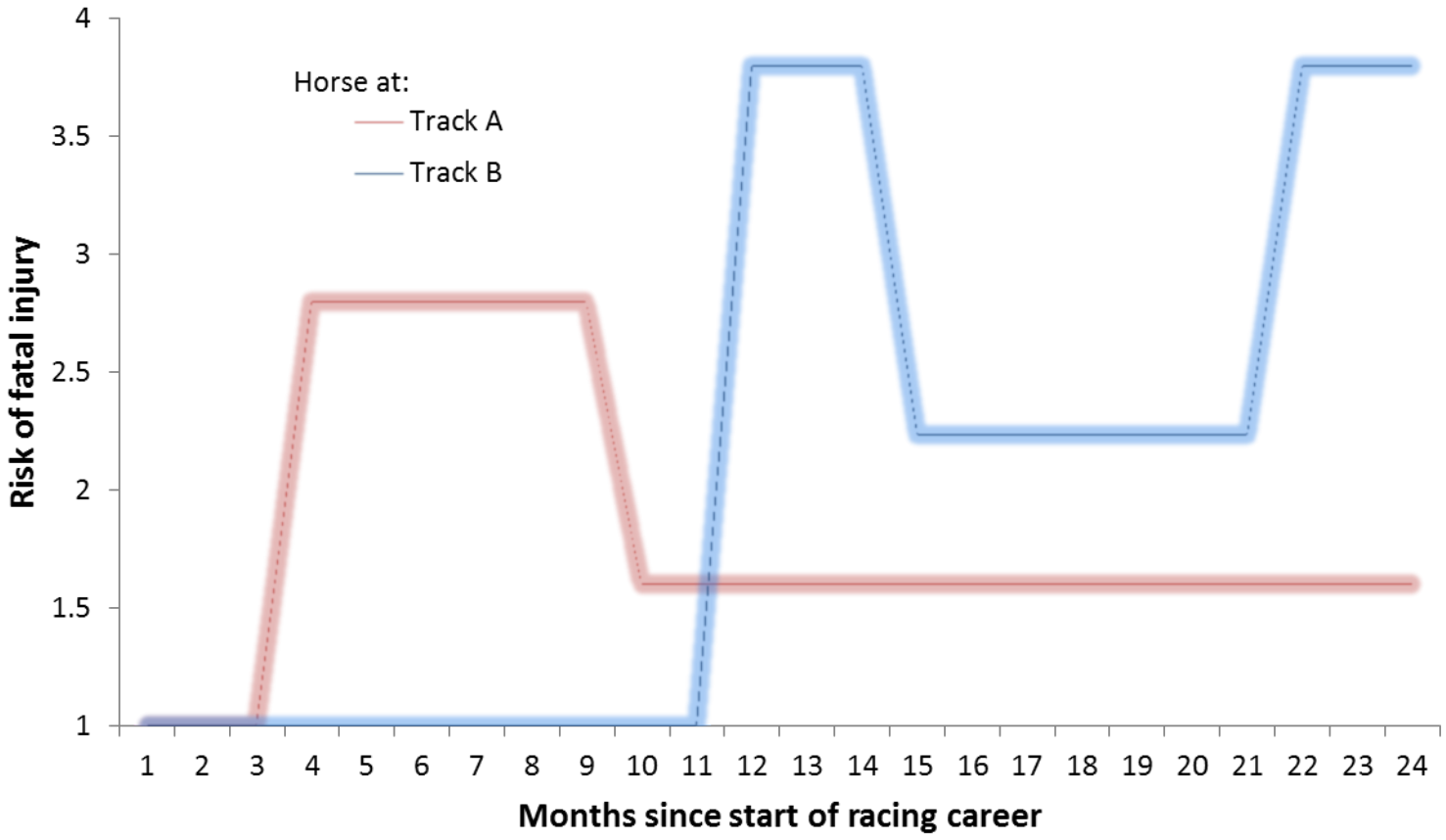
- No difference if include when come off the vet list
- Risk does not return to 'base line' once been on the vet list
- Risk is greater (more than 2-fold) if onto vet list in last 6 months



University of Glasgow

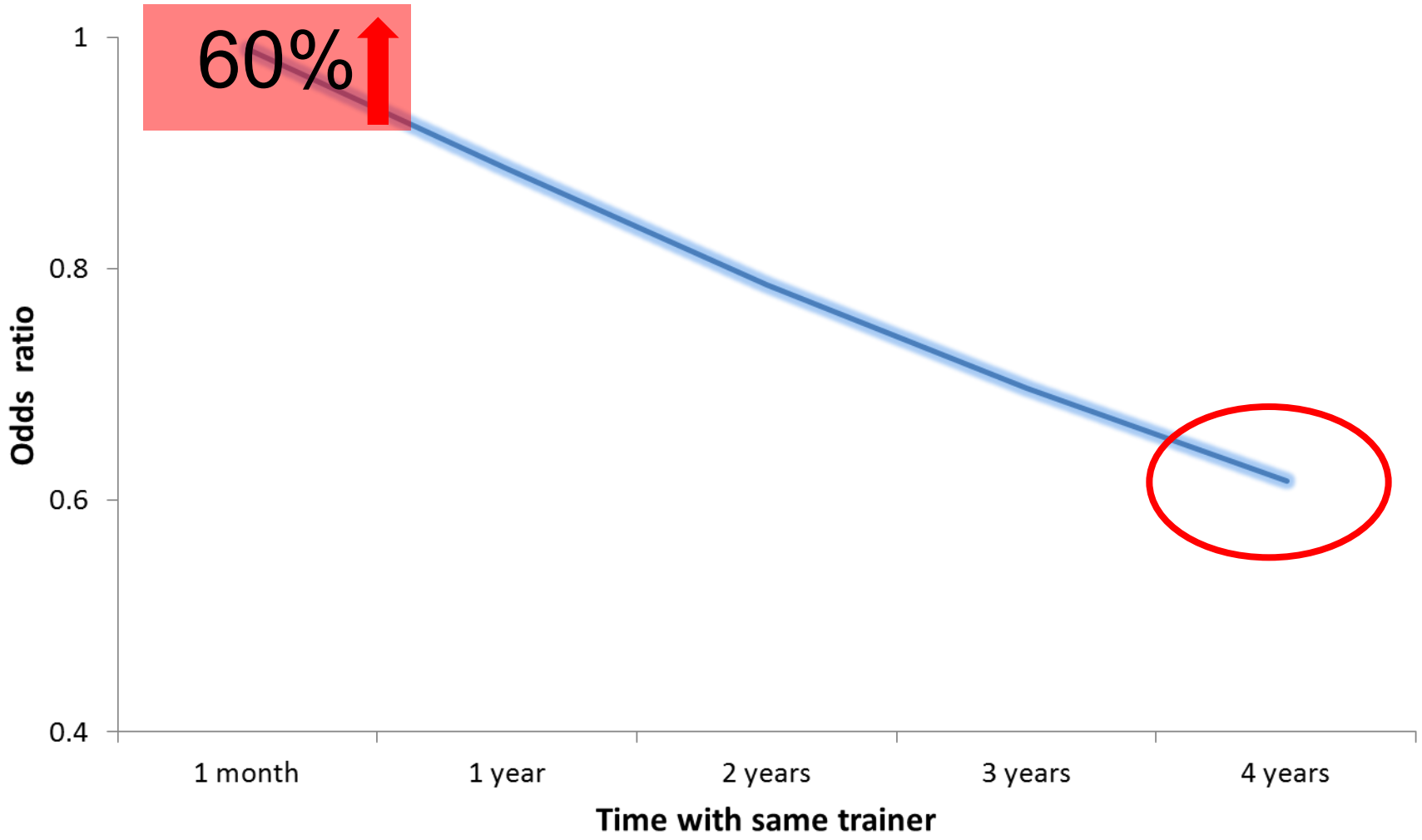
Vet list

- Each track is different
  - Amount of time after onto vet list that risk is increased
  - After onto vet list 'baseline' risk



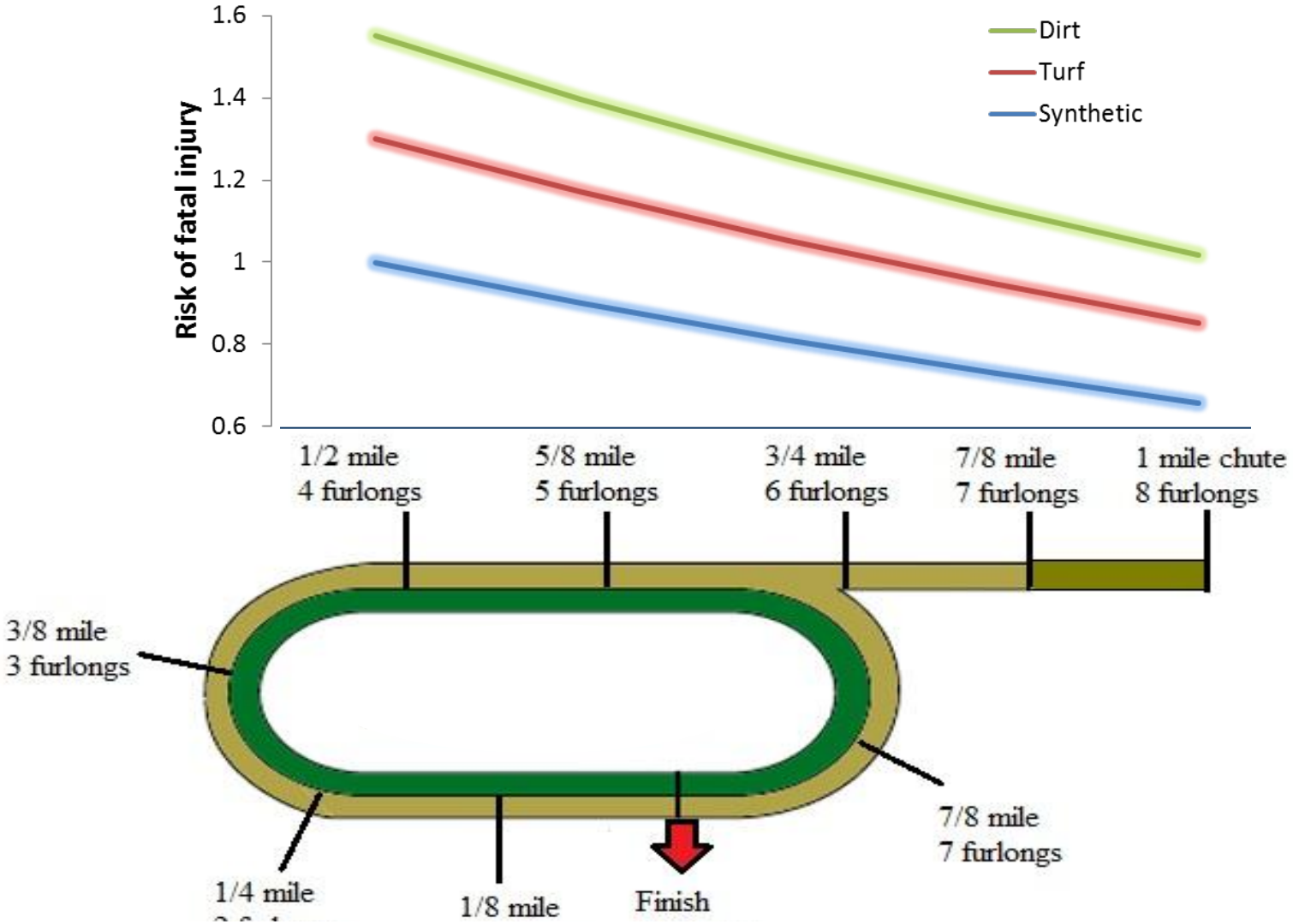
University of Glasgow

Time with same trainer



University of Glasgow

Surface and race distance



University of Glasgow

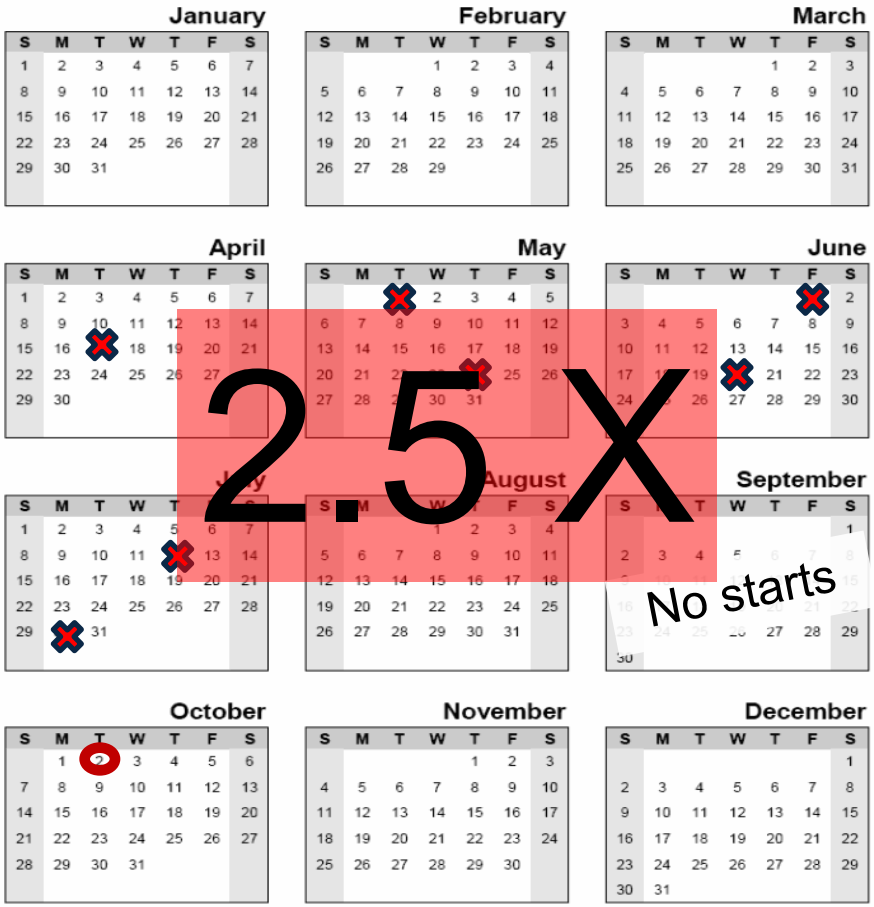
Previous race history

Horse A

2014

Horse B

2014



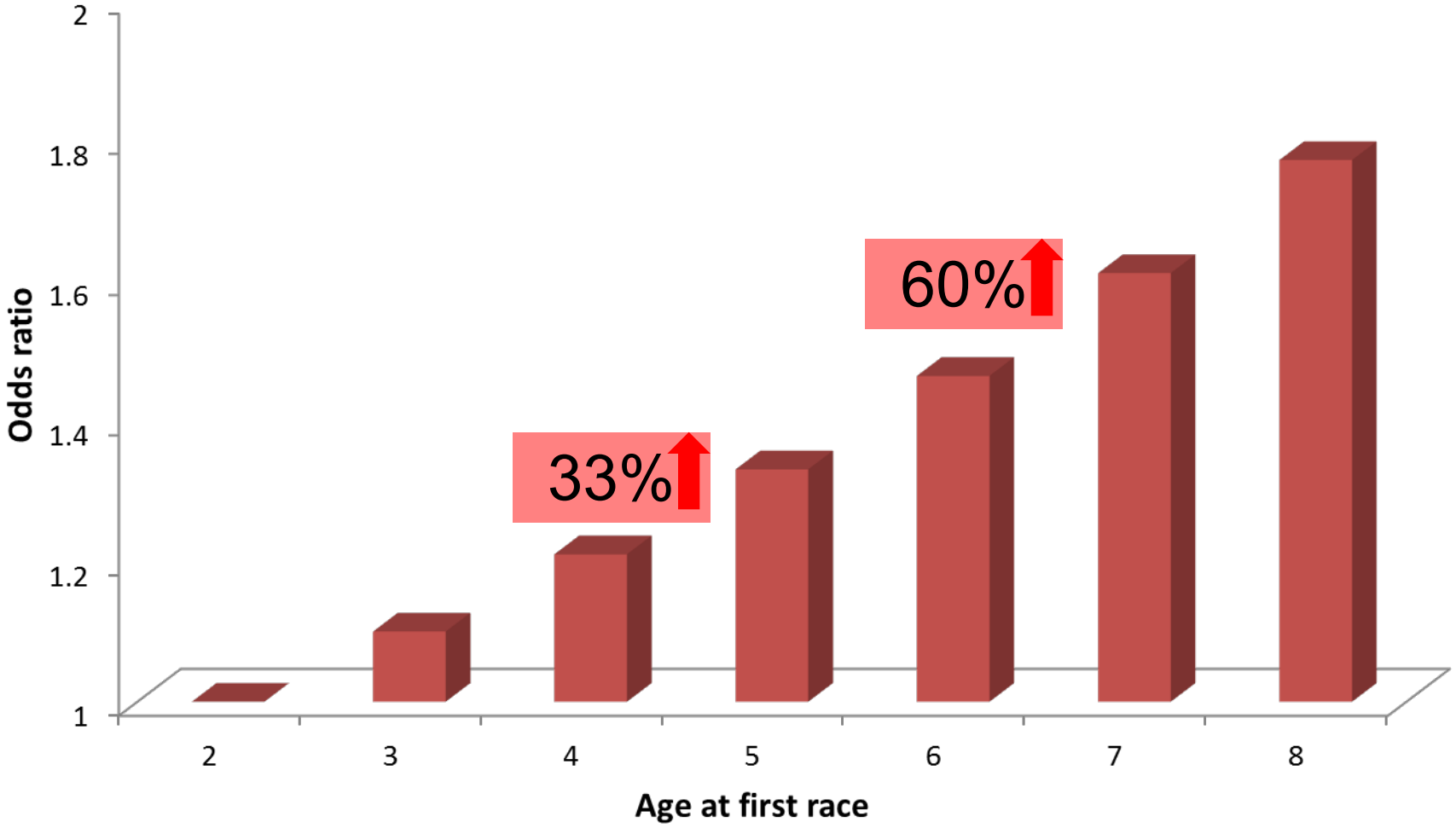
University of Glasgow

Drop in claiming price since last race

Little change since last race (+/- $500)	Drop of between $500 and $10,000	Drop of more than $10,000
Reference	14%	16%

University of Glasgow

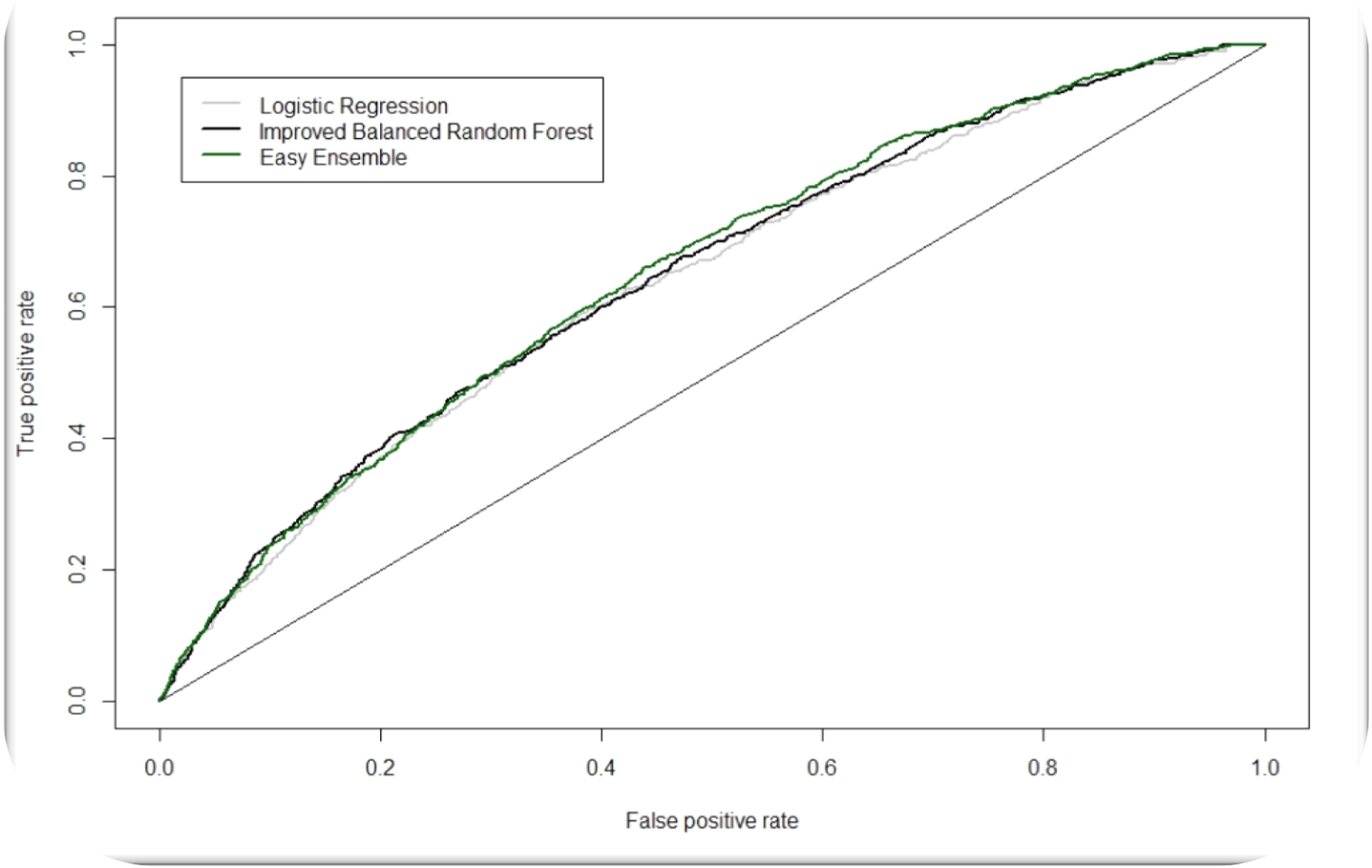
Age at first race



University of Glasgow

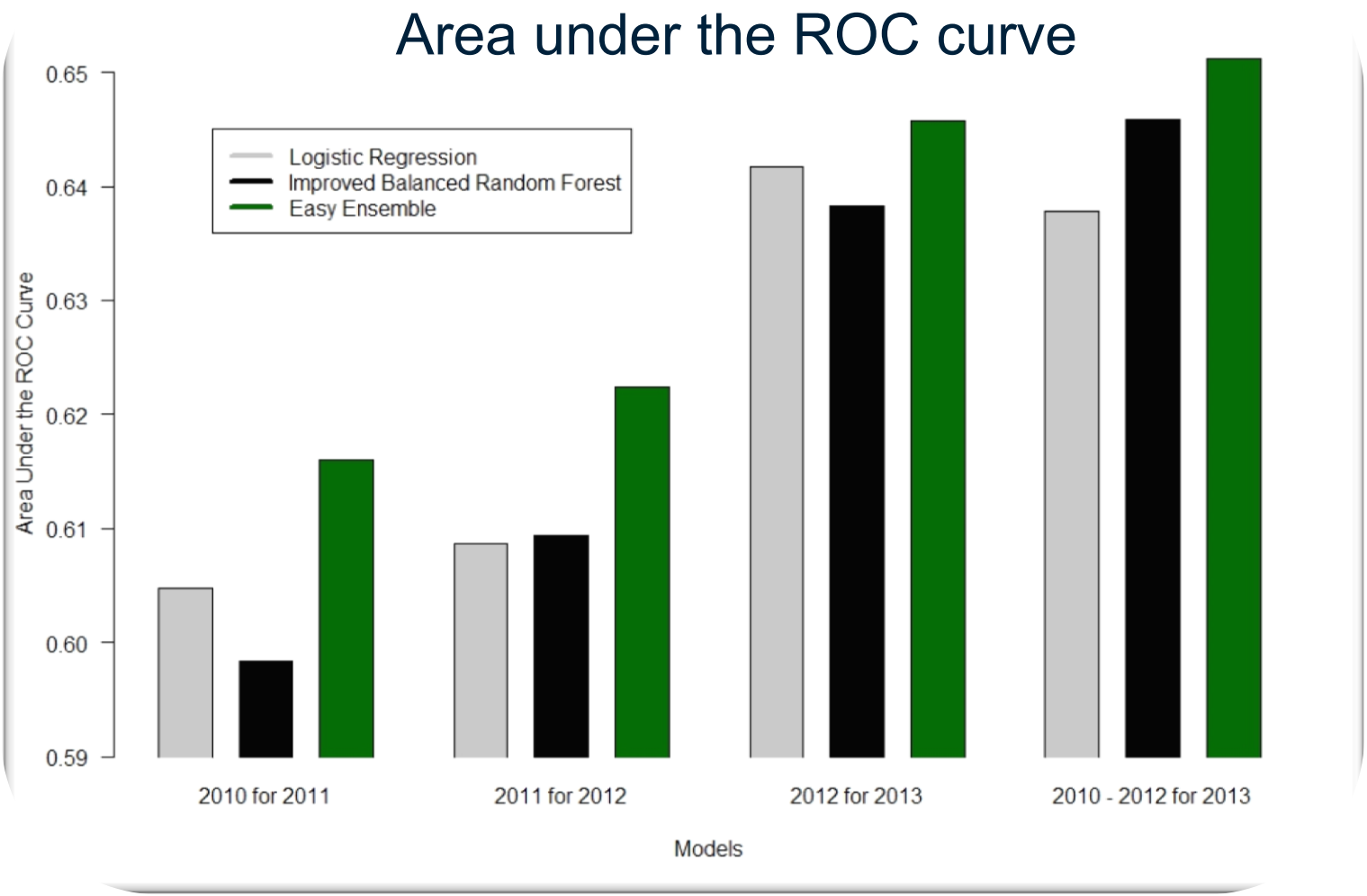
Predictive ability of models ~ 65%

Area under the ROC curve



University of Glasgow

Predictive ability



University of Glasgow

Variable predictive ability at different tracks

- AUC at different tracks:
- Range from 53% to 68%
  - Most individual track models are slightly less predictive
- A lot of ‘local’ factors that are simply missed in EID or not recorded at all
- Importance of ‘local’ knowledge and working with those on the ground at different tracks

University of Glasgow

Overall 3-fold greater risk for top 5%

Multivariable Logistic Regression								
Quantiles of Score	0-5%	5-20%	20-35%	35-50%	50-65%	65-80%	80-95%	95-100%
Relative Horse Risk	0.46	0.47	0.66	0.87	0.94	1.09	1.60	2.71
Improved Balanced Random Forest								
Quantiles of Score	0-5%	5-20%	20-35%	35-50%	50-65%	65-80%	80-95%	95-100%
Relative Horse Risk	0.49	0.45	0.62	0.72	1.05	1.14	1.46	3.26
Easy Ensemble								
Quantiles of Score	0-5%	5-20%	20-35%	35-50%	50-65%	65-80%	80-95%	95-100%
Relative Horse Risk	0.43	0.47	0.67	0.74	0.94	1.20	1.48	3.10

University of Glasgow

Predictive ability of the models

- How close are we at being able to more accurately find horses of interest BEFORE they race?
- Topping out at 65% on predictive models?
  - Maybe best possible
  - Unmeasured variables
  - Inherent variability i.e. unmeasurable variables
- Risk factors & predictive models for injuries/triage 2+
- Keep with analysis from all tracks
- Focus in on tracks with available training data
- Availability of medical/treatment records?
  - Importance of being on the vet list/previous injuries and from work we have done with BHA

University of Glasgow

Further analyses

Variables

- Number of times on vet list
- Work to get off vs. automatically off vet list
- Type of previous injury (fetlock)
- Vet scratches vs. trainer scratches
- Length of meet

Fast work data models

Use of “National” model

- Examine predictive ability of National model for each track

University of Glasgow

What to do with this information?

- Is a three-fold difference in risk important for you to be aware of?
  - 3-fold difference in risk between ‘average’ horse and horse in ‘top 5%’
- Which outcome would be best to try to embed within automatic risk profiling for each start?
  - Fatality – clearly important but rare
  - Injury/triage 2+ – important and more common, but case definition will include a lot of variation
  - Fracture of distal limb (fatal and non-fatal)

University of Glasgow

Acknowledgements

- US Jockey Club
  - Matt Iuliano
  - Kristin Werner Leshney
  - Jamie Haydon
- University of Glasgow
  - Stamatis Georgopoulos (who did all the work!)