Bayesian Networks for the Prediction of Ranking in Male Tennis Players

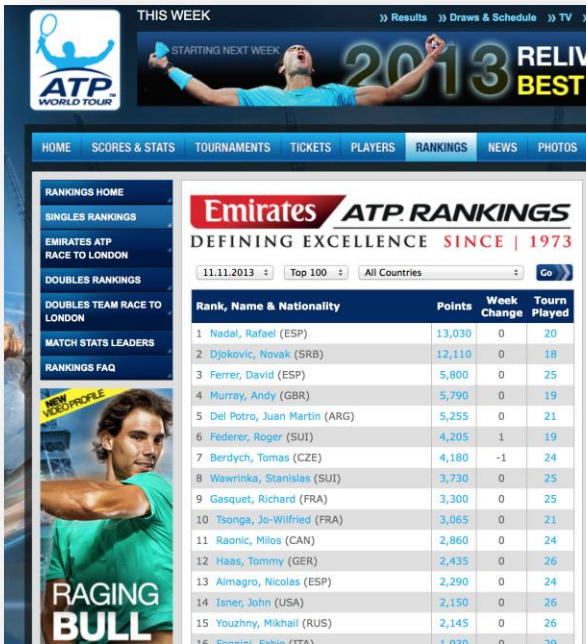
Dr. Michael Bane
Dr. Stuart Morgan
Prof. Ann Nicholson
Dr. Machar Reid

tennis Australia's Favourite
AUSTRALIAN INSTITUTE OF SPORT
VICTORIA UNIVERSITY MELBOURNE AUSTRALIA
ISEAL INSTITUTE OF SPORT, EXERCISE AND ACTIVE LIVING
MONASH University

Overview

- *Athletes only break even if ranked within the top 100.*
- *Investment in athletes is big business, LTA and TA high-performance expenditure of £12M and $24M respectively*
- *Can we improve our return on investment?*

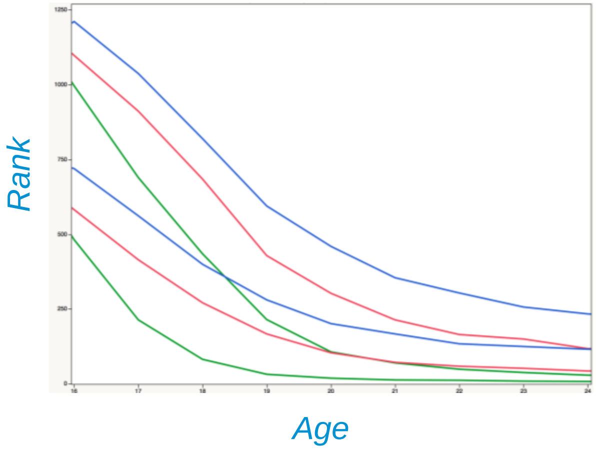
Data



Rank, Name & Nationality	Points	Week Change	Tourn Played
1 Nadal, Rafael (ESP)	13,030	0	20
2 Djokovic, Novak (SRB)	12,110	0	18
3 Ferrer, David (ESP)	5,800	0	25
4 Murray, Andy (GBR)	5,790	0	19
5 Del Potro, Juan Martin (ARG)	5,255	0	21
6 Federer, Roger (SUI)	4,205	1	19
7 Berdych, Tomas (CZE)	4,180	-1	24
8 Wawrinka, Stanislas (SUI)	3,730	0	25
9 Gasquet, Richard (FRA)	3,300	0	25
10 Tsonga, Jo-Wilfried (FRA)	3,065	0	21
11 Raonic, Milos (CAN)	2,860	0	24
12 Haas, Tommy (GER)	2,435	0	26
13 Almagro, Nicolas (ESP)	2,290	0	24
14 Isner, John (USA)	2,150	0	26
15 Youzhny, Mikhail (RUS)	2,145	0	26
16 Fognini, Fabio (ITA)	1,930	0	29

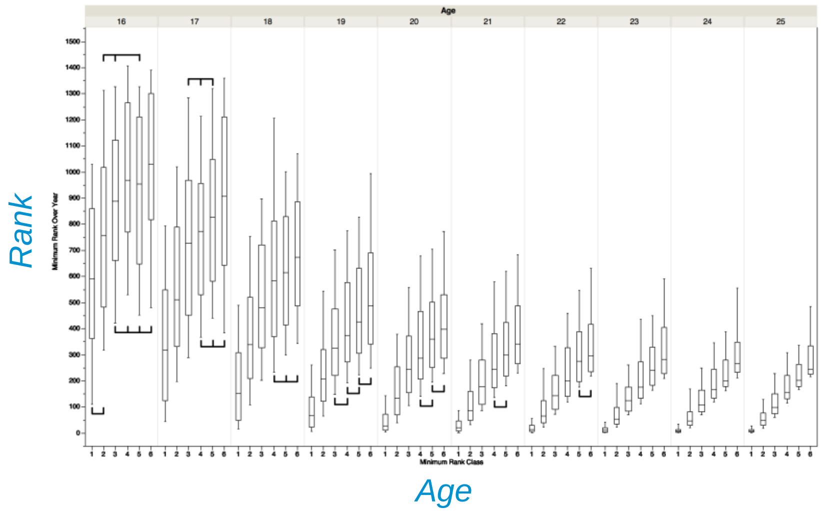
- *Data is sourced from weekly ranking lists provided by the ATP.*
- *Data relates to the accumulated ranking points and net ranking.*
- *Date of birth of athletes also sourced.*
- *Rankings obtained as far back as 1973 (data is complete from 1984)*
- *2338 athletes whom achieved at least Top 500 (80/20 train-test split)*

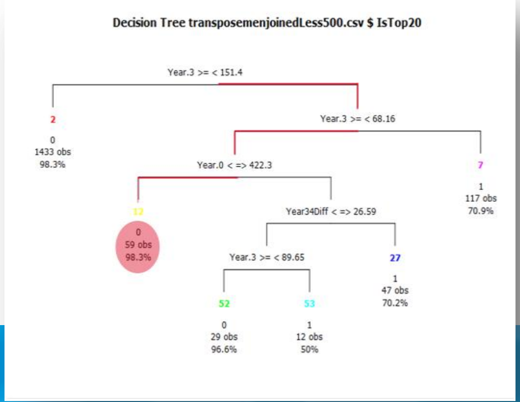
Putting a toe in the water



🡐 *The 25th/75th quartiles of ranking, conditioned on best career ranking (Top 10/50/100). Prosecutors fallacy at play!*

🡒 *Trends have been modelled by ANOVA*





🡐 *Decision trees were fitted with limited success*

Why BNs

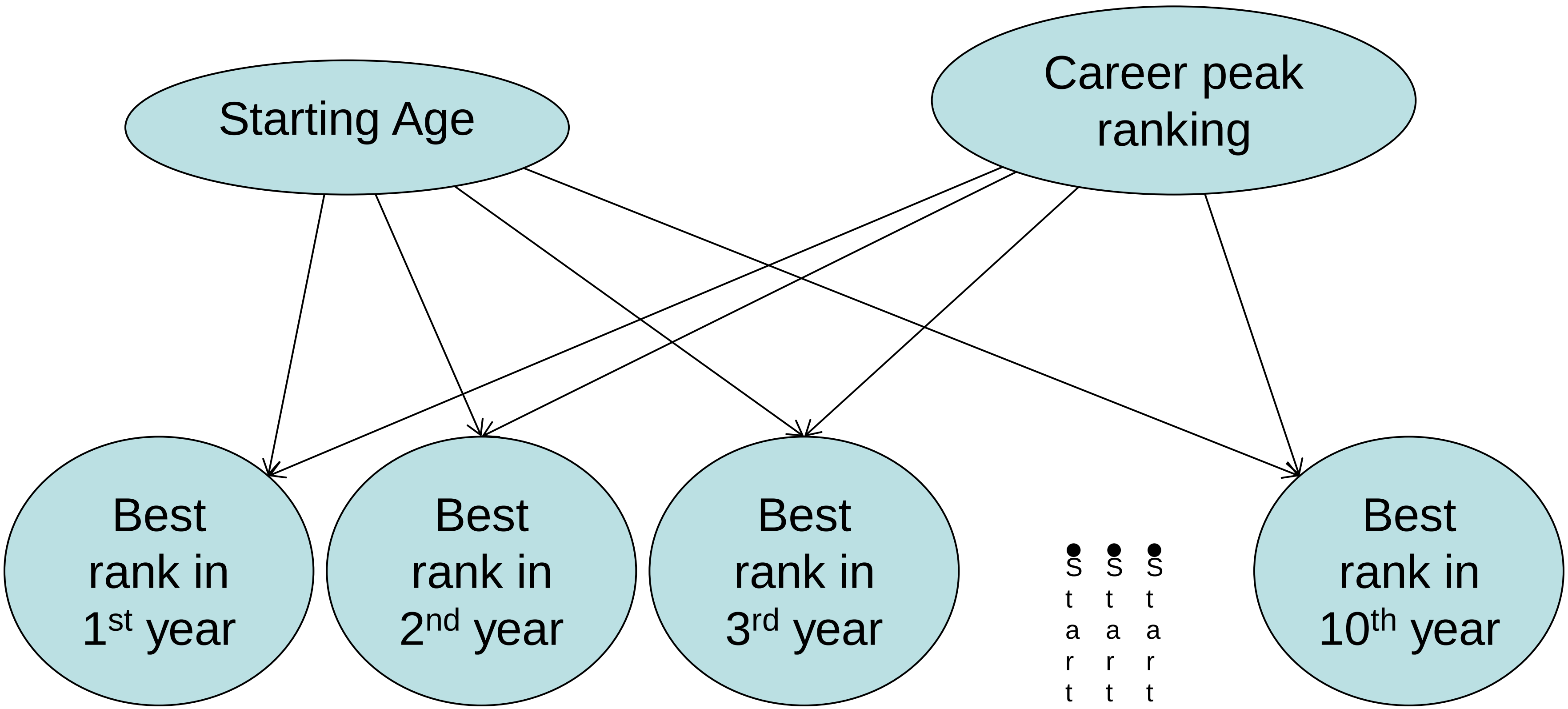
- *Intuitive, relative to other machine learning techniques.*

- *Versatile applications.*

- *Works well with variable amounts of data.*

- *Simple to implement to a wide audience.*

Naive Network



- Starting Age
- Starting Age
- Starting Age

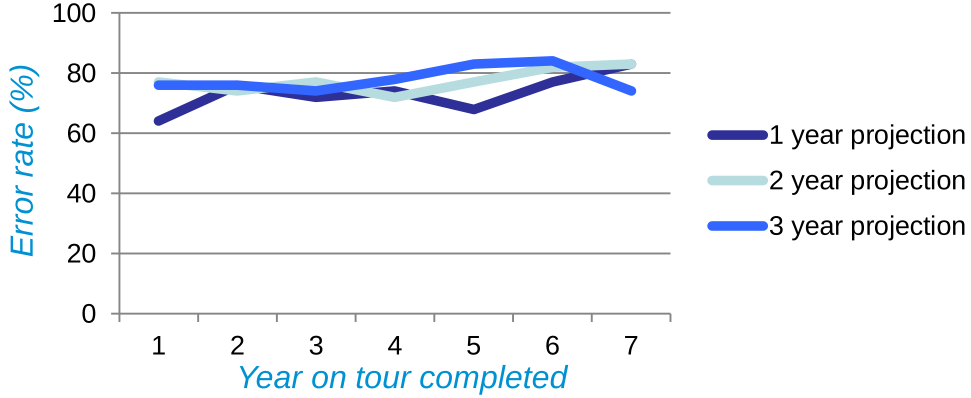
Can't use current athletes

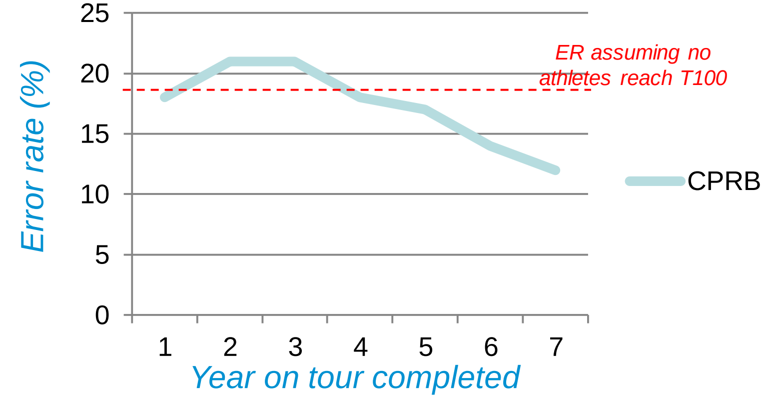
Naive Network

Demonstration

Naive Network (results)

The Bad





The Good

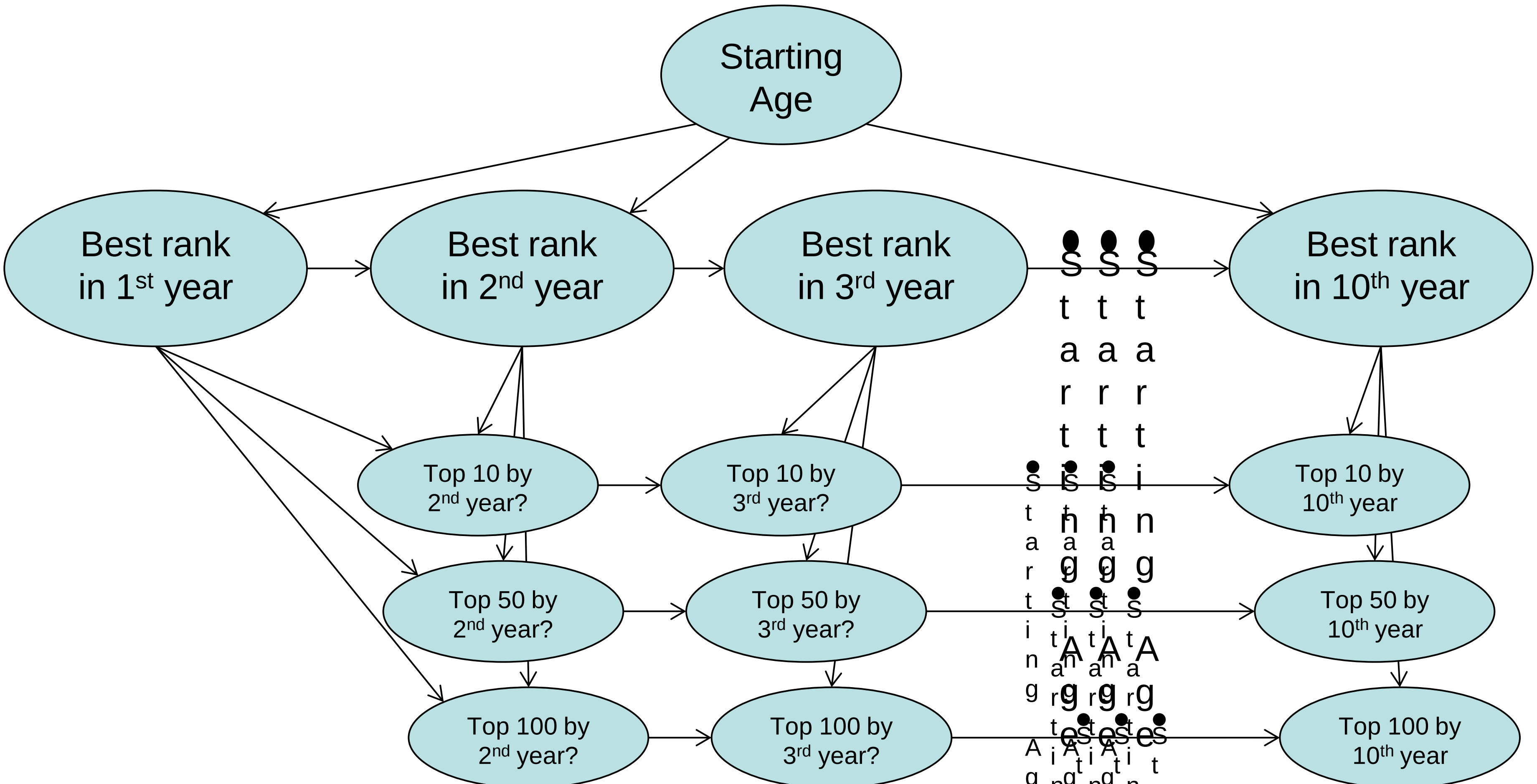
Calibration

Y1-2,+1 0-0.5: 0 | 0.5-1: 0 | 1-2: 0.279| 2-5: 2.46 | 5-10: 6.76 | 10-15: 14.4 | 15-20: 13.2 | 20-30: 26.5 | 30-40: 34.7 |
Y1-3,+1 0-0.5: 0 | 0.5-1: 0.844| 1-2: 1.18 | 2-5: 3.33 | 5-10: 6.8 | 10-15: 12.9 | 15-20: 20.6 | 20-30: 24.4 |
Y1-2,+2 0-0.5: 0 | 0.5-1: 0.379| 1-2: 1.3 | 2-5: 2.73 | 5-10: 6.89 | 10-15: 11.9 | 15-20: 18.4 | 20-30: 26.7 | 30-100: 50 |
Y1-7,+2 0-0.5: 0 | 0.5-1: 0.413| 1-2: 0.498| 2-5: 2.29 | 5-10: 5.69 | 10-15: 14.2 | 15-20: 19.4 | 20-30: 23.7 | 30-100: 66.7 |
Y1-3,+3 0.2-0.5: 6.52 | 0.5-1: 0 | 1-2: 4.65 | 2-5: 3 | 5-10: 8.61 | 10-15: 11.4 | 15-20: 15.9 | 20-30: 22.6 | 30-40: 25 | 40-100: 50 |

Times surprised

Ex.	<1%	<10%	>90%	>99%
Y1-2, +1	0 (0/437)	2.94 (58/1974)	0 (0/0)	0 (0/0)
Y1-3, +1	0.53 (2/378)	3.66 (70/1911)	0 (0/0)	0 (0/0)
Y1-2, +2	0.19 (1/526)	3.21 (62/1930)	0 (0/0)	0 (0/0)
Y1-7, +3	0.21 (1/473)	2.56 (43/1681)	0 (0/0)	0 (0/0)
Y1-3, +3	1.57 (8/511)	3.30 (28/848)	0 (0/0)	0 (0/0)

Temporal Network



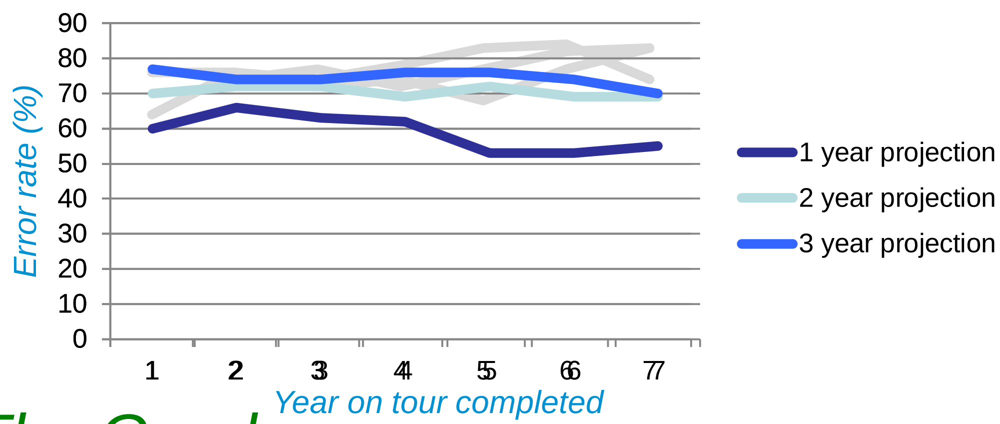
Can us current athletes

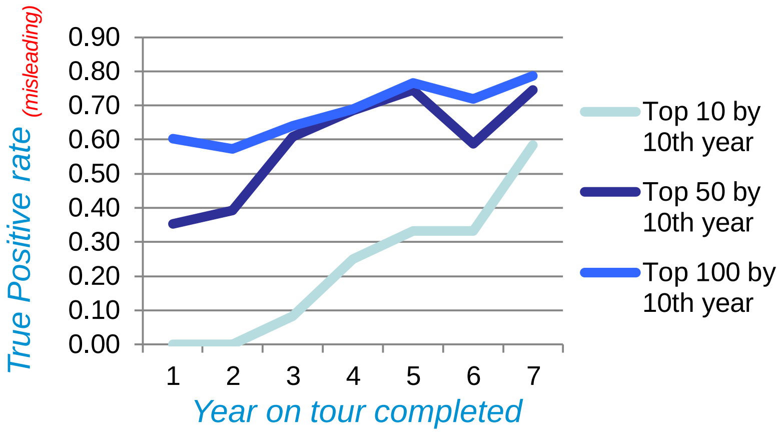
Temporal Network

Demonstration

Temporal Network (results)

The Bad





The Good

Calibration

Y1-2,+1 0-0.5: 0 | 0.5-1: 0.14 | 1-2: 0.607| 2-5: 2.69 | 5-10: 9.64 | 10-15: 9.52 | 15-20: 18 | 20-30: 28.2 | 30-40: 34.9 | 40-50: 34 |
Y1-3,+1 0-0.5: 0 | 0.5-1: 0 | 1-2: 0.836| 2-5: 2.05 | 5-10: 9.92 | 10-15: 6.11 | 15-20: 24.8 | 20-30: 28.2 | 30-40: 35.6 | 40-50: 44.3 | 50-100: 48 |
Y1-2,+2 0-1: 0 | 1-2: 0.563| 2-5: 2.26 | 5-10: 8.63 | 10-15: 13.4 | 15-20: 20.4 | 20-30: 23.7 | 30-40: 32.9 |
Y1-7,+2 0-2: 0 | 2-5: 0.901| 5-10: 5.38 | 10-15: 13.3 | 15-20: 21 | 20-30: 29.2 | 30-50: 43.2 |
Y1-3,+3 0-2: 0 | 2-5: 1.19 | 5-10: 5.54 | 10-15: 13.4 | 15-20: 20.5 | 20-30: 26.3 |

Times surprised

Ex.	<1%	<10%	>90%	>99%
Y1-2, +1	0.12 (1/845)	2.29 (73/3181)	0 (0/0)	0 (0/0)
Y1-3, +1	0 (0/691)	2.46 (78/3175)	0 (0/0)	0 (0/0)
Y1-2, +2	0 (0/120)	3.73 (107/2870)	0 (0/0)	0 (0/0)
Y1-7, +3	0 (0/0)	3.14 (55/1750)	0 (0/0)	0 (0/0)
Y1-3, +3	0 (0/0)	3.13 (73/2329)	0 (0/0)	0 (0/0)

Problems

- *Athletes follow junior/senior/mixed pathways throughout their early career*
- *Lack of variety in data. Relying on historical accumulation.*
- *Nature of benchmarking implies only rank can be used.*
- *Very long term predictions.*
- *Variability in ranking pathways is huge.*

Successes

- *We can learn from history, and objectively calculate accurate probabilities of events*
- *An improvement from previous prediction attempts (there are none!)*
- *Identify how poor ranking is as a predictor success at peak*

Acknowledgements

Dr. Stuart Morgan AUSTRALIAN INSTITUTE OF SPORT

Prof. Ann Nicholson MONASH University

Dr. Machar Reid tennis Australia's Favourite

Thanks for listening