FAIR QUALIFYING TIMES ACROSS AGE AND GENDER CATEGORIES FOR THE BOSTON MARATHON

Richard L Smith and Spencer Siegel (University of North Carolina); Dorit Hammerling, Laura Albrecht and Ross Ring-Jarvi (Colorado School of Mines)

Joint Statistical Meetings: August 4, 2020

BACKGROUND

- The Boston Marathon is the only major marathon to require qualifying standards of the majority of participants
- Qualifying standards were first introduced in the 1970s and have been revised several times since
- The standards were most recently revised in 2019, in preparation for the 2020 race. However, this was still not sufficient to allow them to accept every qualified runner.
- Recent project: predicting the number of entries for the 2021 race (assuming this is held!)
- This talk is not about that, but a more philosophical question: what standard would really be fair, taking age and sex into account?

2020 Boston Marathon qualifying

Qualifying standard and actual qualification time by age group and gender

Age group	men's standard	men's qualification	women's standard	women's qualification
18-34	3:00:00	2:58:21	3:30:00	3:28:21
35-39	3:05:00	3:03:21	3:35:00	3:33:21
40-44	3:10:00	3:08:21	3:40:00	3:38:21
45-49	3:20:00	3:18:21	3:50:00	3:48:21
50-54	3:25:00	3:23:21	3:55:00	3:53:21
55-59	3:35:00	3:33:21	4:05:00	4:03:21
60-64	3:50:00	3:48:21	4:20:00	4:18:21
65-69	4:05:00	4:03:21	4:35:00	4:33:21
70-74	4:20:00	4:18:21	4:50:00	4:48:21
75-79	4:35:00	4:33:21	5:05:00	5:03:21
80 and older	4:50:00	4:48:21	5:20:00	5:18:21

Source: BAA (reprinted by the Boston Globe)

YEAR	FIELD SIZE	"CUT-OFF TIME"*	QUALIFIERS NOT ACCEPTED
2012	27,000	1:14	3,228
2014	36,000	1:38	2,976
2015	30,000	1:02	1,947
2016	30,000	2:28	4,562
2017	30,000	2:09	2,957
2018	30,000	3:23	5,062
2019	30,000	4:52	7,248
2020	31,500	1:39	3,161

Source: BAA

IDEA MOTIVATING THE PRESENT TALK

- All runners slow down as they get older, but there is a lot of individual variability
- Try to use statistical methods to characterize the age-graded performance of a "typical" runner
- The standard method used for age-graded performances is nominally based on world records in different age groups, but this may not reflect typical runners' performances
- This talk is focused mainly on the age question: separate work by Dorit and her students has examined the gender-equity issue, though the two questions are closely related
- I will motivate the method by describing a method of analysis I first worked out several years ago, and then describe our more recent work to extend the results

My original analysis

- About 500 runners have run the Boston marathon at least 10 years in succession (BAA)

My original analysis

- About 500 runners have run the Boston marathon at least 10 years in succession (BAA)
- No easy way to reconstruct the list

My original analysis

- About 500 runners have run the Boston marathon at least 10 years in succession (BAA)
- No easy way to reconstruct the list
- I used the datasets I had available to find 1,272 runners who had run each of 2010, 2011, 2013

My original analysis

- About 500 runners have run the Boston marathon at least 10 years in succession (BAA)
- No easy way to reconstruct the list
- I used the datasets I had available to find 1,272 runners who had run each of 2010, 2011, 2013
- I then used the BAA archive to find all of those runners who had run at least 6 times (men) or 5 times (women) during 2001-2013

My original analysis

- About 500 runners have run the Boston marathon at least 10 years in succession (BAA)
- No easy way to reconstruct the list
- I used the datasets I had available to find 1,272 runners who had run each of 2010, 2011, 2013
- I then used the BAA archive to find all of those runners who had run at least 6 times (men) or 5 times (women) during 2001-2013
- Runners who did not finish in 2013 were estimated using Hammerling et al. (2014)

My original analysis

- About 500 runners have run the Boston marathon at least 10 years in succession (BAA)
- No easy way to reconstruct the list
- I used the datasets I had available to find 1,272 runners who had run each of 2010, 2011, 2013
- I then used the BAA archive to find all of those runners who had run at least 6 times (men) or 5 times (women) during 2001-2013
- Runners who did not finish in 2013 were estimated using Hammerling et al. (2014)
- Result: 547 men and 249 women identified (806 runners; 7,219 individual race results)

Longitudinal Data Approach

- Each individual runner record is a part-trace of the performance v. age curve for that runner
- Allow for a random “runner effect”
- Also allow for a random “calendar year” effect (2004 and 2012 were very hot)
- Separate men’s and women’s performance
- A refinement (later): also distinguish runners of different ability levels

Statistical Model:

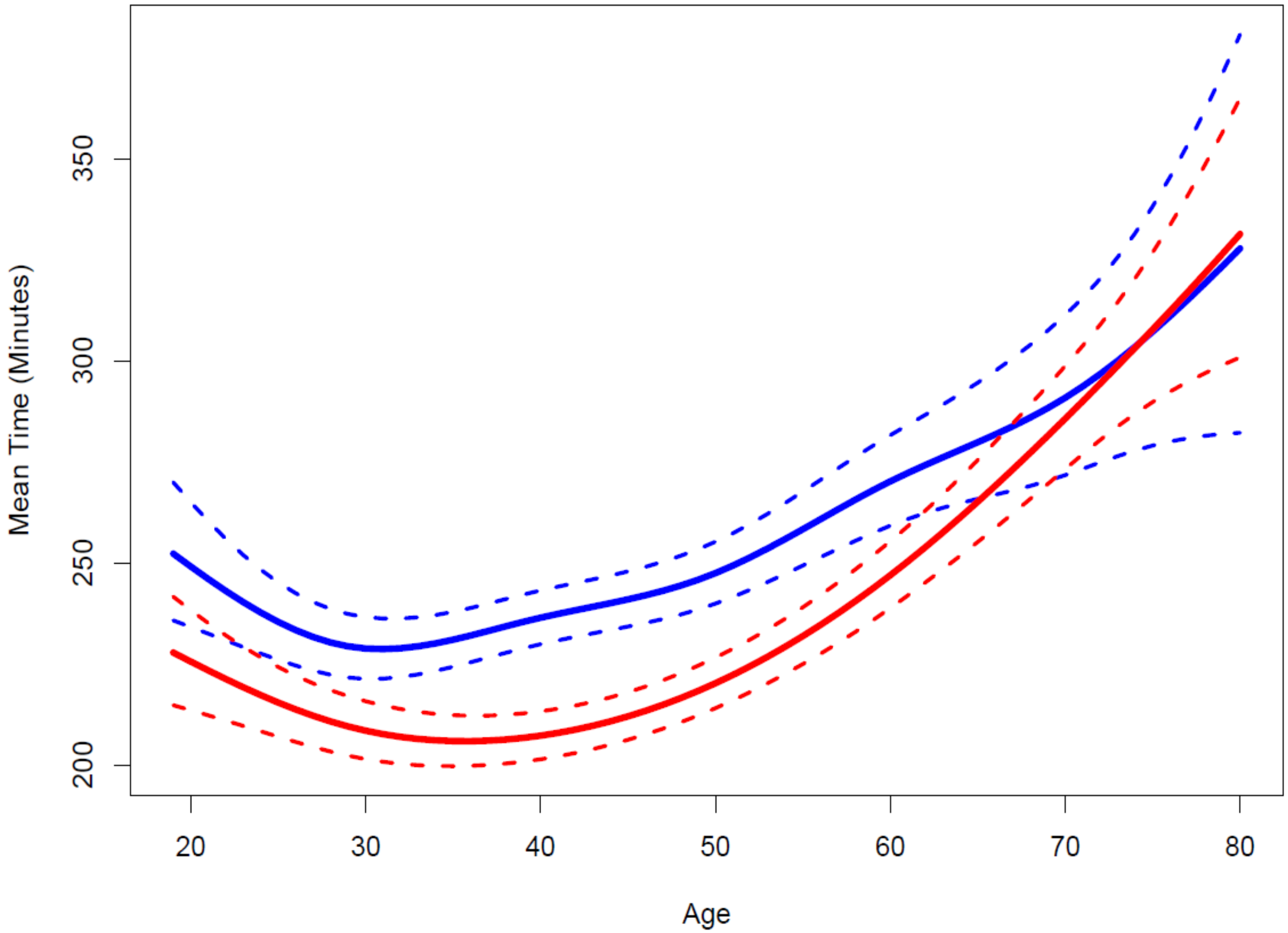
$$\log t_{ij} = \alpha_i + \beta_{y_{ij}} + S(a_{ij}; K) + \epsilon_{ij},$$

where

- t_{ij} is the jth finish time of runner i,
- y_{ij} is the year of the jth finish time of runner i,
- a_{ij} is the ith runner's age in her jth finish time,
- α_i represents the overall ability level of runner i (small α_i means a faster runner),
- $\beta_{y_{ij}}$ is a year effect,
- $S(a_{ij}; K)$ represents a nonlinear function of age with K degrees of freedom,
- ϵ_{ij} is a random error.
- Computation: use function `lmer` within R package `lme4`.

Results

- Men's curve (red, with confidence limit)
- Women's curve (blue, with confidence limits)
- Crossover above age 70 almost certainly an artifact
- Other anomalies need to be explained



Extension of These Results

- By web-scraping, we were able to download nearly complete results for Boston, Chicago, New York and several other major US marathons
- This allowed us to apply the foregoing analysis to much larger datasets

Race	Years	# of Unique Runners	# of Observations
Boston Marathon	2001-2017 except 2015	51,119	146,570
Chicago Marathon	2000-2017 except 2015	72,588	194,370
New York Marathon	2000-2019 except 2012	84,515	236,526
Los Angeles Marathon	2000-2019	39,446	127,721
Marine Corps Marathon	2000-2018	44,630	128,535
Twin Cities Marathon	2000-2019 except 2003-2005	21,262	65,617
Philadelphia Marathon	2000-2019	19,784	53,741
Houston Marathon	2000-2018 except 2011	17,288	60,889
Grandma's Marathon	2000-2019 except 2006	15,382	46,530
California International Marathon	2000-2019	14,667	44,933

*There is no explicit reason why certain years were omitted from the data

Table 2.1 - Overview of Datasets

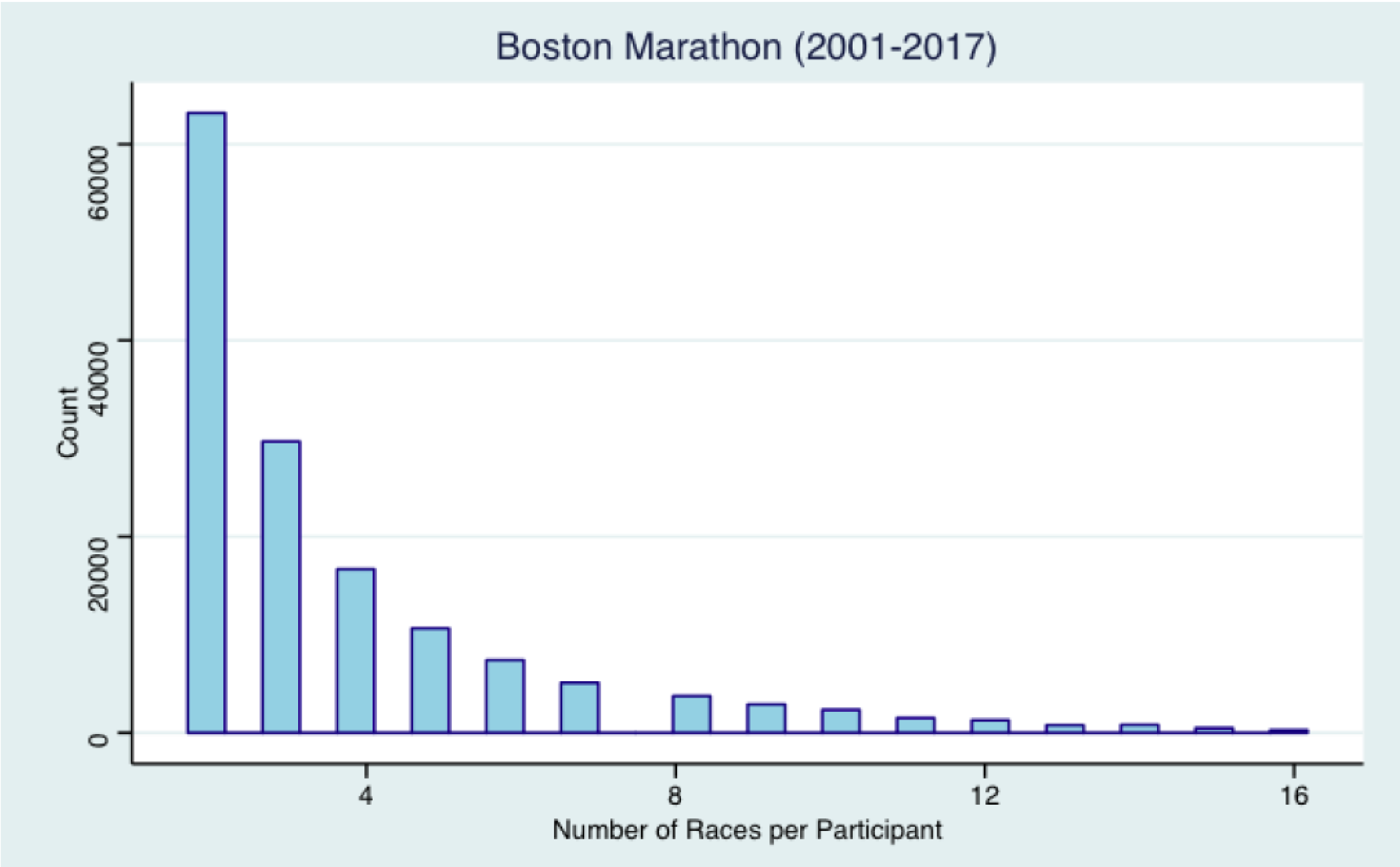


Figure 2.1 – Boston Marathon Data # of Races per Participant

Selecting the Number of Knots in the Spline

- Used 5-fold cross-validation
- Little evidence we need more than 2 or 3 knots (surprising)
- Similar results using orthogonal polynomials instead of splines, and for other large races
- Also considered splitting runners by ability level but similar results

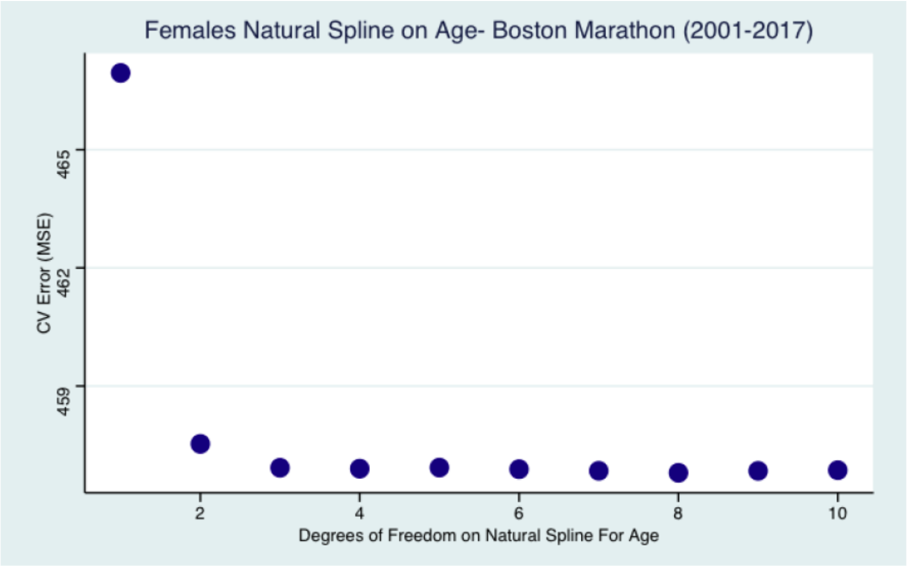


Figure 3.1 – Cross-Validation plot of Natural Splines for Boston Female data

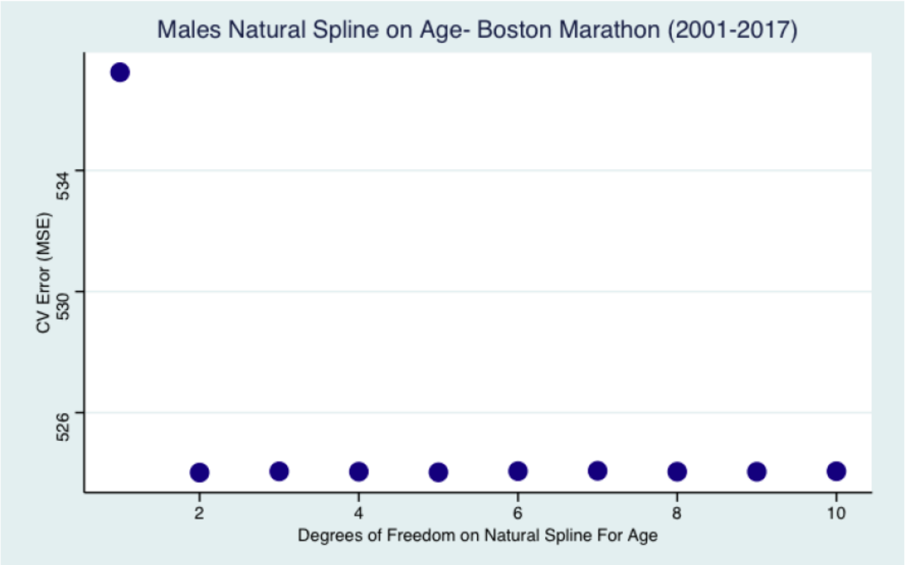


Figure 3.2 - Cross-Validation plot of Natural Splines for Boston Male data

Results: Age-Time Curves for Boston

- Combined all Boston Data 2001-2007
- Also split into quartiles based on standardized times

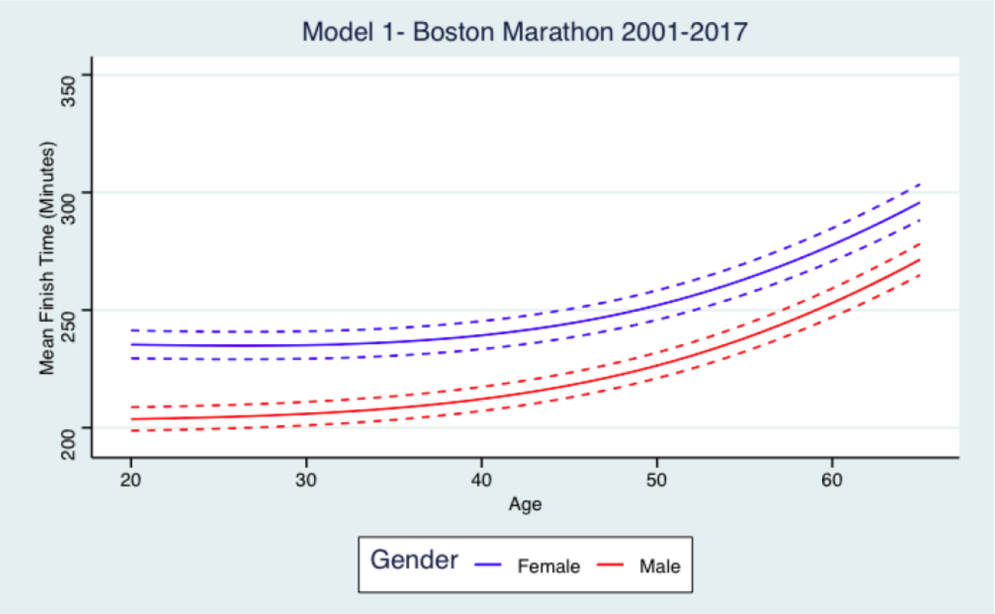


Figure 4.3 – Age-Performance curve for Model 1 on Boston data

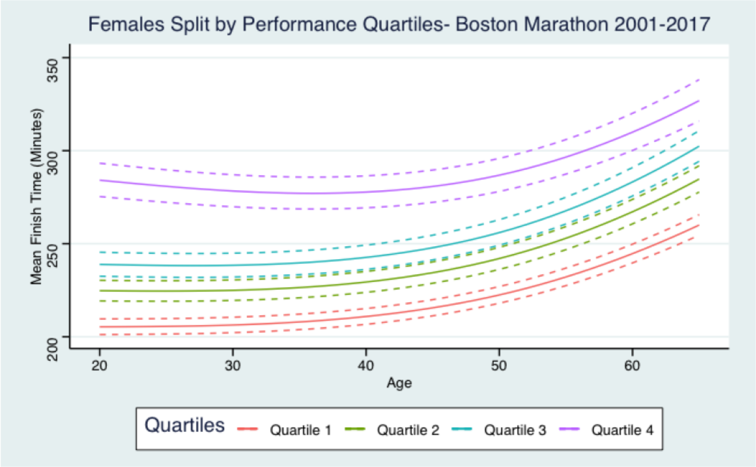


Figure 4.4 - Age-Performance curve for Boston Female data split by Quartiles

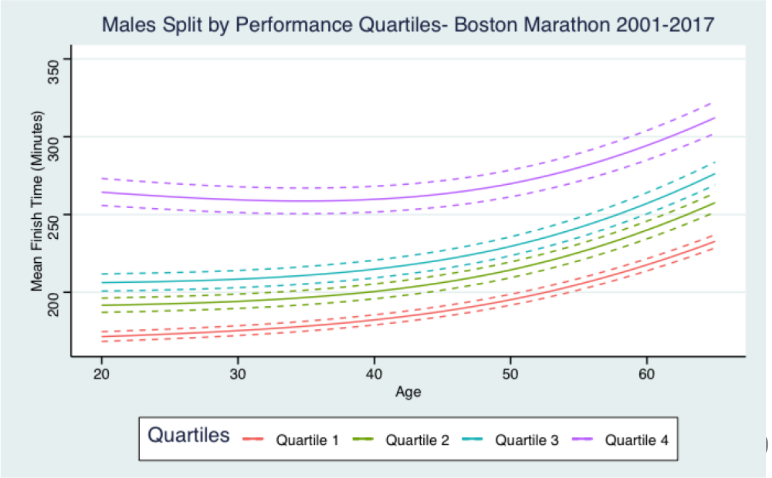


Figure 4.5 - Age-Performance curve for Boston Male data split by Quartiles

Results: Age-Time Curves for Chicago

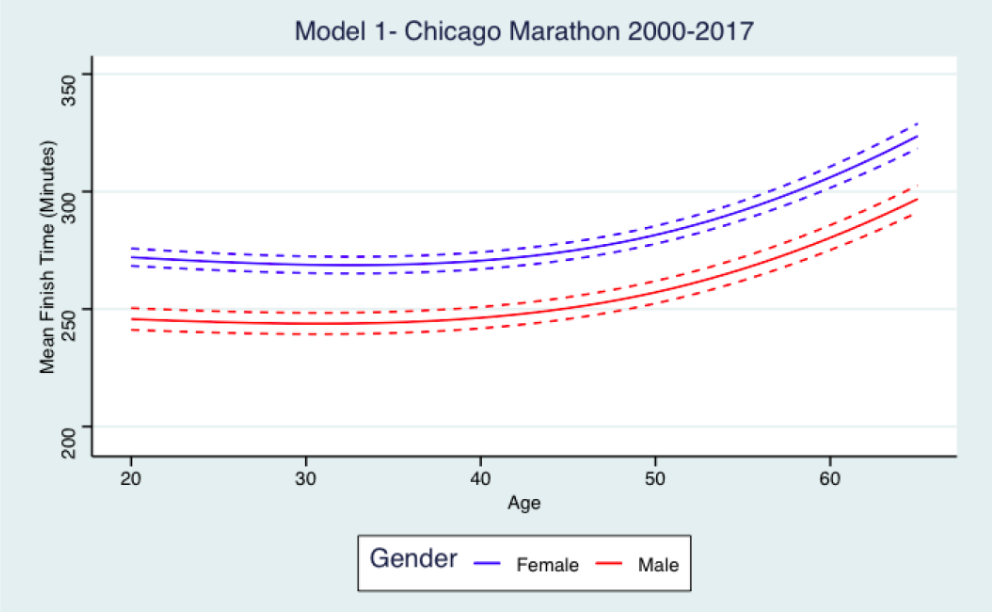


Figure 4.10 - Age-Performance Curve from Model 1 on Chicago data

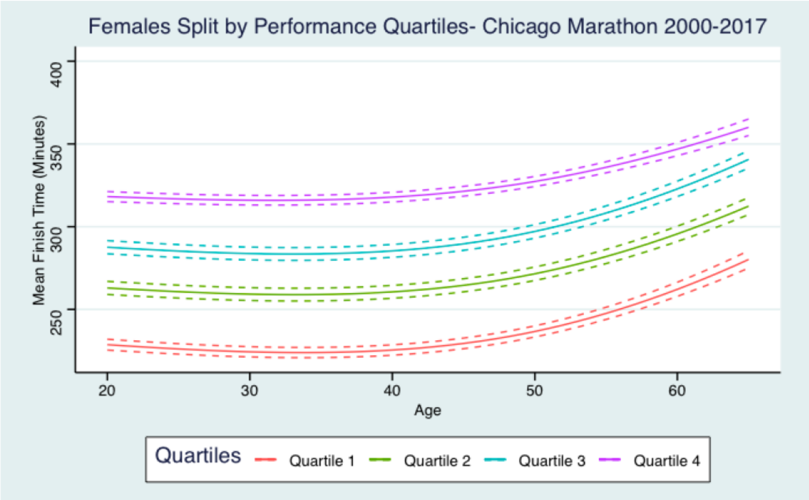


Figure 4.15 - Age-Performance curve for Chicago Female data split by Quartiles

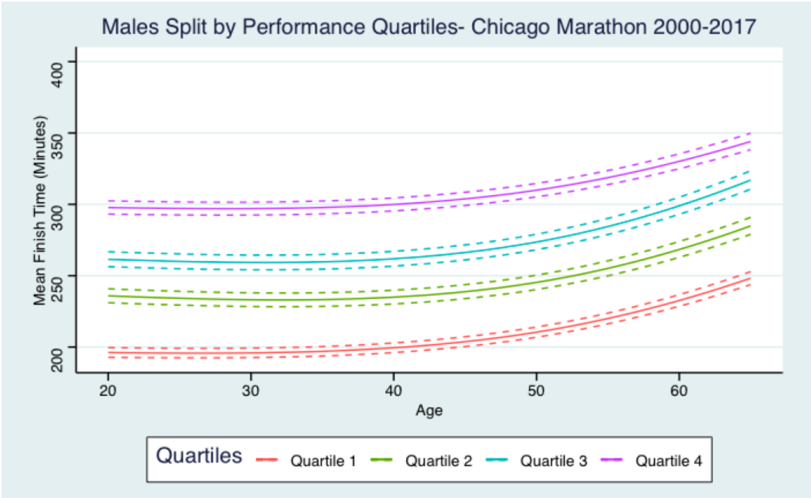


Figure 4.16 - Age-Performance curve for Chicago Male data split by Quartiles

Results: Age-Time Curves for New York

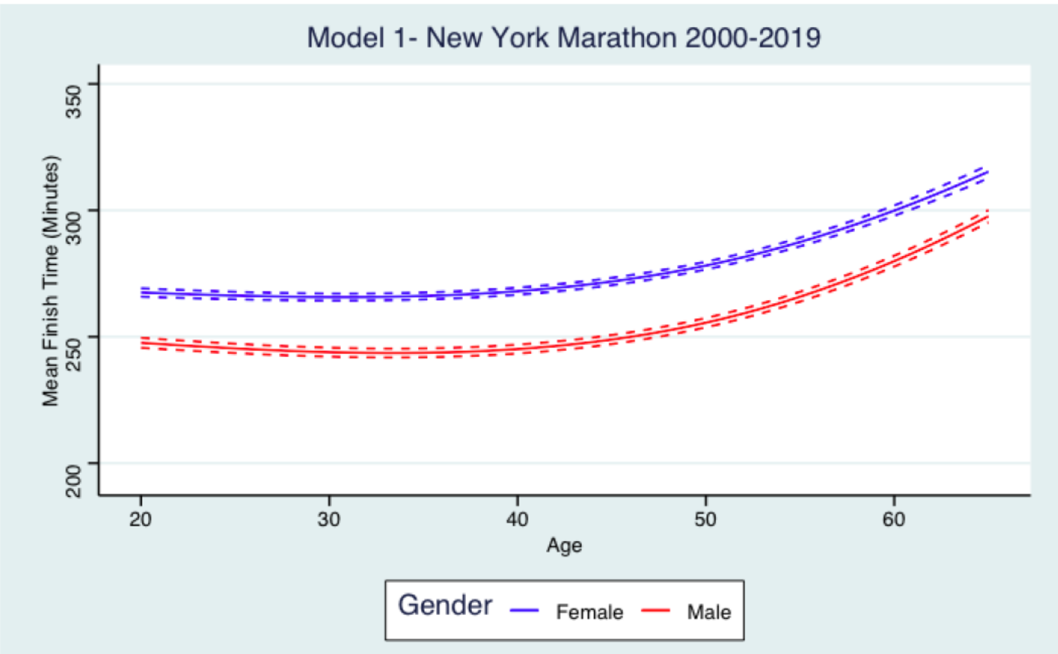


Figure 4.21 - Age-Performance Curve from Model 1 on New York data

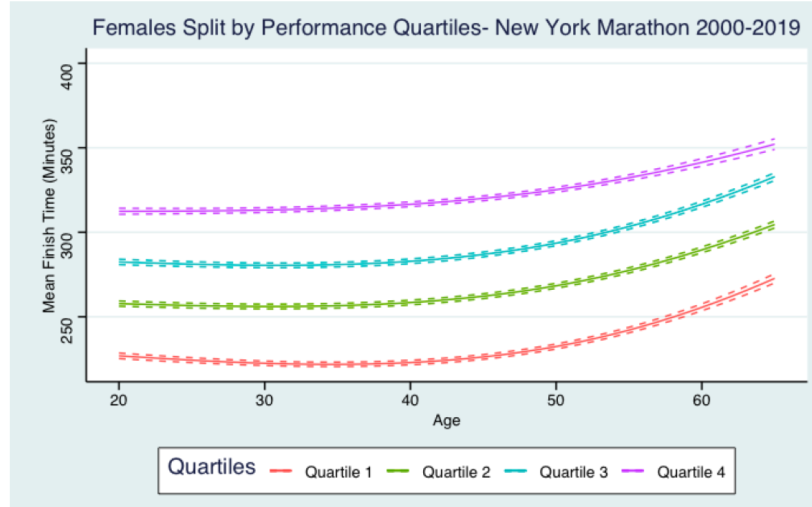


Figure 4.22 - Age-Performance curve for New York Female data split by Quartiles

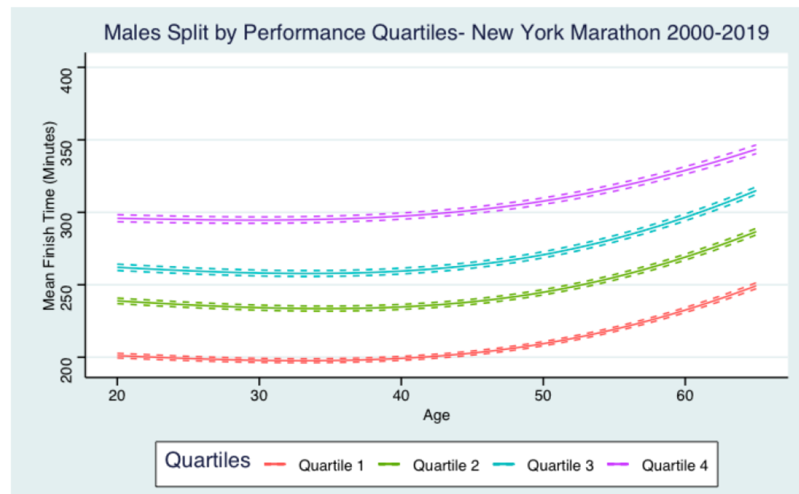


Figure 4.23 - Age-Performance curve for Chicago Male data split by Quartiles

What would these results mean for qualifying times?

- Assumed fixed 3:00 and 3:30 for male and female 18-34, as at present
- Compute "equivalent" times for other age groups

Age Groups	Male Times	Female Times
18-34	03:00:00	03:30:00
35-39	03:04:01	03:31:55
40-44	03:07:57	03:35:08
45-49	03:13:46	03:40:24
50-54	03:22:00	03:48:17
55-59	03:33:03	03:59:11
60-64	03:46:59	04:13:06
65-69	04:03:48	04:29:58
70-74	04:23:29	04:49:43
75-79	04:45:59	05:12:13

Table 4.4 - Qualifying Standards produced from Model 1 on Boston Marathon data

Current:

Age group	men's standard	women's standard
18-34	3:00:00	3:30:00
35-39	3:05:00	3:35:00
40-44	3:10:00	3:40:00
45-49	3:20:00	3:50:00
50-54	3:25:00	3:55:00
55-59	3:35:00	4:05:00
60-64	3:50:00	4:20:00
65-69	4:05:00	4:35:00
70-74	4:20:00	4:50:00
75-79	4:35:00	5:05:00
80 and older	4:50:00	5:20:00

Equivalent Results for Other Races

- Maybe Boston results are too closely tuned to current qualifying times
- Try same analysis for New York (top 50% of runners)

Age Groups	Male Times	Female Times
18-34	03:00:00	03:30:00
35-39	02:58:30	03:28:08
40-44	03:00:11	03:29:49
45-49	03:03:55	03:33:38
50-54	03:10:14	03:40:09
55-59	03:19:35	03:49:50
60-64	03:32:00	04:02:39
65-69	03:47:29	04:18:33
70-74	04:06:02	04:37:27
75-79	04:27:33	04:59:14

le 4.11 - Qualifying Standards produced from Model 1 on top 50% of New York data

Current:

Age group	men's standard	women's standard
18-34	3:00:00	3:30:00
35-39	3:05:00	3:35:00
40-44	3:10:00	3:40:00
45-49	3:20:00	3:50:00
50-54	3:25:00	3:55:00
55-59	3:35:00	4:05:00
60-64	3:50:00	4:20:00
65-69	4:05:00	4:35:00
70-74	4:20:00	4:50:00
75-79	4:35:00	5:05:00
80 and older	4:50:00	5:20:00

CONCLUSIONS

- Still work in progress!
- Longitudinal approach aims to reproduce how "typical runners" perform in races such as Boston, Chicago, New York
- Doesn't rely on age-group world records (disadvantage of age-graded performances)
- We got similar results for other large races (e.g. Los Angeles, Marine Corps) using the same methods
- Have not yet tried on any race outside US
- But, some caveats:
  - Still have to define the "population of interest" (e.g. use all runners in a race, top 50%, top 25%, etc.)
  - We have tried to build a model for dropout probabilities but without changing the results very much
  - Other forms of the random effects model have been tried and are still being explored
  - Current analysis doesn't directly address equity between men and women but that question is also being explored