The Marathon world record of Ruth Chepngetich: Was it legitimate, and how can data and analytics help us decide?

Richard L. Smith

Department of Statistics and Operations Research

University of North Carolina, Chapel Hill

Public Lecture sponsored by Sports Analytics Intelligence Laboratory

Joint work with Abigail Mabe

February 5, 2025

Introduction

- On October 13, 2024, the Kenyan runner Ruth Chepngetich ran a women's world record for the marathon of 2 hours, 9 minutes, 56 seconds

- This broke the previous world record by nearly 2 minutes and Chepngetich's personal best by nearly 5 minutes

- Her performance immediately raised questions about illegal drug use, but other commentators supported her record as "entirely plausible"

- In this talk, we show how statistical methods may be used to assess these claims.

Opinion: Why It's Hard To Trust Ruth Chepngetich's Marathon World Record

Last Updated: Oct 20, 2024 10:57 am

Amby Burfoot
EDITOR AT LARGE

Here's Why Ruth Chepngetich's World Record Is Completely Plausible: A Data-Driven Analysis

While the Kenyan has come under scrutiny and speculation, a look at the research suggests otherwise.

Last Updated: Oct 21, 2024 8:58 pm

Jessy Carveth
SENIOR NEWS EDITOR

Last weekend at the 2024 Chicago Marathon, **Ruth Chepngetich** rewrote the history books after breaking the world record in the marathon by nearly two minutes.

The entire race was a spectacle to behold, as she came through the halfway point in 1:04:16, which not only was the fastest half-marathon run by a woman on American soil but also the fifth-fastest half-marathon run by a woman in history.

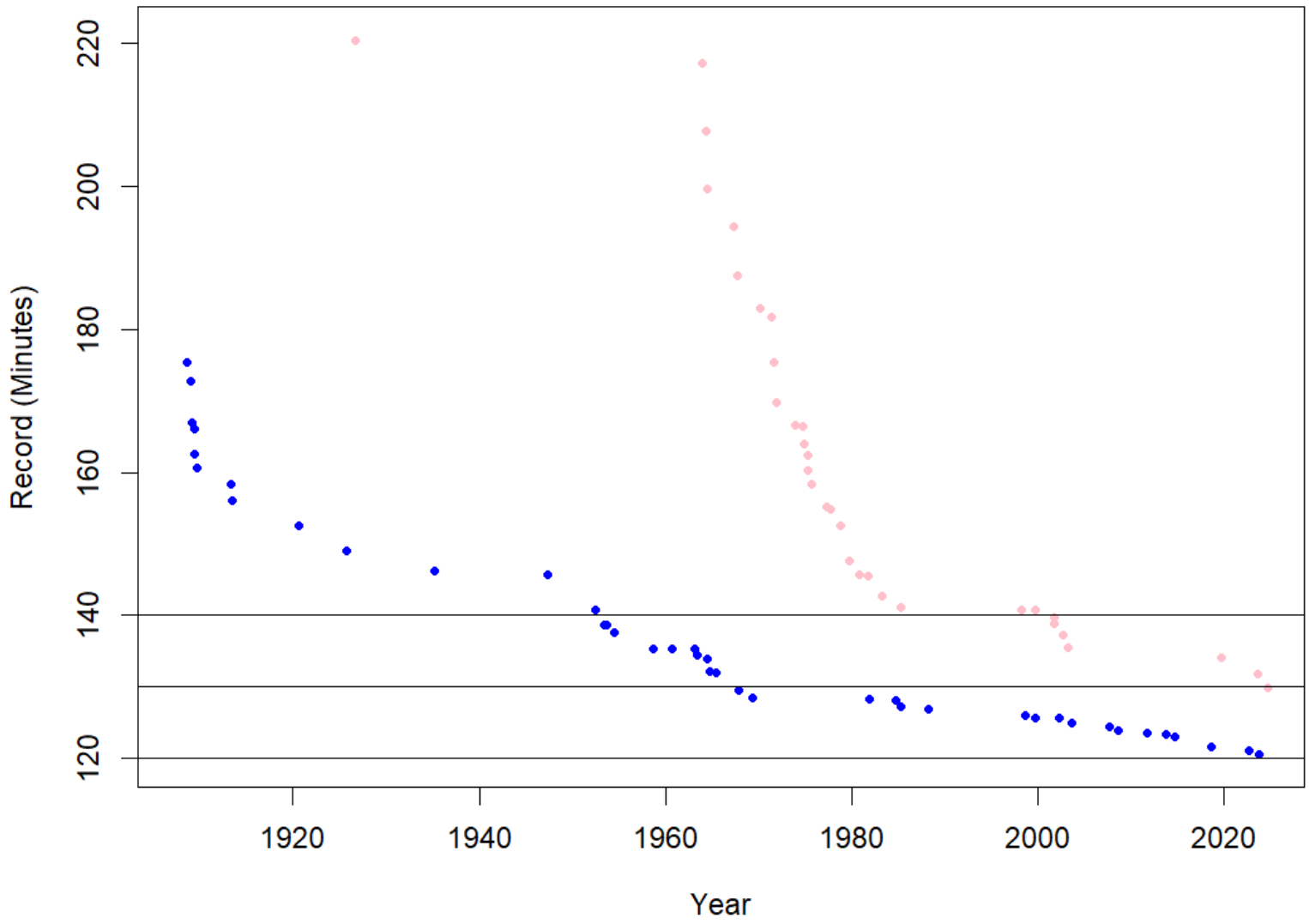
Although she slowed a bit in the second half, Chepngetich eventually blew the world record out of the water, crossing the line in 2:09:56, becoming the first woman in history to break 2:10:00 (or 2:11, for that matter).

RUTH CHEPNGETICH' S MARATHON WORLD RECORD OFFICIALLY RATIFIED BY WORLD ATHLETICS

Despite the controversy Chepngetich's 2:09:56 is officially official.

MARATHON HANDBOOK

Marathon world records (source: World Athletics)



Digression: My work with the Boston Marathon

- History: first run in 1897, long established as the top long-running race in the world
- Qualifying times were introduced in 1979 as a way to contain the field size. The initial qualifying time was 3 hours for men under 40
- Over the years, developed separate qualifying times for men and women and also separate times for each 5-year age group
- I ran the race 9 times between 1990 and 2004, qualified for 2010 but left it too late to enter, then did get in for 2011 race but race sold out in one day
- Much internet and social media chatter about what to do

2026 BOSTON MARATHON QUALIFYING TIMES

Age Group	MEN	WOMEN	NON-BINARY
18-34	2hrs 55min 00sec	3hrs 25min 00sec	3hrs 25min 00sec
35-39	3hrs 00min 00sec	3hrs 30min 00sec	3hrs 30min 00sec
40-44	3hrs 05min 00sec	3hrs 35min 00sec	3hrs 35min 00sec
45-49	3hrs 15min 00sec	3hrs 45min 00sec	3hrs 45min 00sec
50-54	3hrs 20min 00sec	3hrs 50min 00sec	3hrs 50min 00sec
55-59	3hrs 30min 00sec	4hrs 00min 00sec	4hrs 00min 00sec
60-64	3hrs 50min 00sec	4hrs 20min 00sec	4hrs 20min 00sec
65-69	4hrs 05min 00sec	4hrs 35min 00sec	4hrs 35min 00sec
70-74	4hrs 20min 00sec	4hrs 50min 00sec	4hrs 50min 00sec
75-79	4hrs 35min 00sec	5hrs 05min 00sec	5hrs 05min 00sec
80 and over	4hrs 50min 00sec	5hrs 20min 00sec	5hrs 20min 00sec

BOSTON MARATHON
April 18, 1994
Marathon Foto


this shoe represents the millions of people across the world who support the boston marathon and the entire running community.
this year, no matter where you are, we all run boston.
energy boost 2.0 limited edition, 2014
#weallrunboston
Boston Athletic Association
COMMEMORATIVE PROGRAM
26 MILES TO BOSTON
The Boston Marathon Experience from Hopkinton to Copley Square
Racers' Record Book

Qualifying Times for the Boston Marathon

• Ar http://chance.amstat.org/category/articles/

Richard L. Smith, Scott Powers, and Jessi Cisewski

Patriots' Day of 1897 marked the first of what is now the oldest annual marathon in the world: the Boston Marathon. The Boston Athletic Association (BAA) operates this marathon in which runners from around the world strive to participate. The Boston Marathon is not only historically appealing for runners, but is also considered a grueling course—topped off with "Heartbreak Hill" after mile 20 of 26.2. An extra twist to the allure of the Boston Marathon is that runners have to satisfy the BAA qualifying time standards in a certified marathon within about a year and a half of the desired Boston Marathon. Qualifying standards for the Boston Marathon were first established in the 1970s, when they were viewed as a means of controlling the size of the field while continuing to allow established marathon runners to participate. They have since been revised a number of times.

Table 1 contains the qualifying times used for the 2011 Boston Marathon (QT1), which had been in place since 2003. An additional 59 seconds is accepted for each standard. For instance, a man aged 18–34 could have run up to 3:10:59 and still been accepted into the race. Also, while race organizers may publish both "gun" and "chip" times, the chip time (representing net time between crossing the start and finish lines) is accepted as the basis of Boston

OPEN ACCESS Freely available online

PLOS ONE

Completing the Results of the 2013 Boston Marathon

Dorit Hammerling[1], Matthew Cefalu[2], Jessi Cisewski[3], Francesca Dominici[2], Giovanni Parmigiani[2,4], Charles Paulson[5], Richard L. Smith[1,6]*

1 Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, North Carolina, United States of America, **2** Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America, **3** Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America, **4** Dana Farber Cancer Institute, Boston, Massachusetts, United States of America, **5** Puffinware LLC, State College, Pennsylvania, United States of America, **6** Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

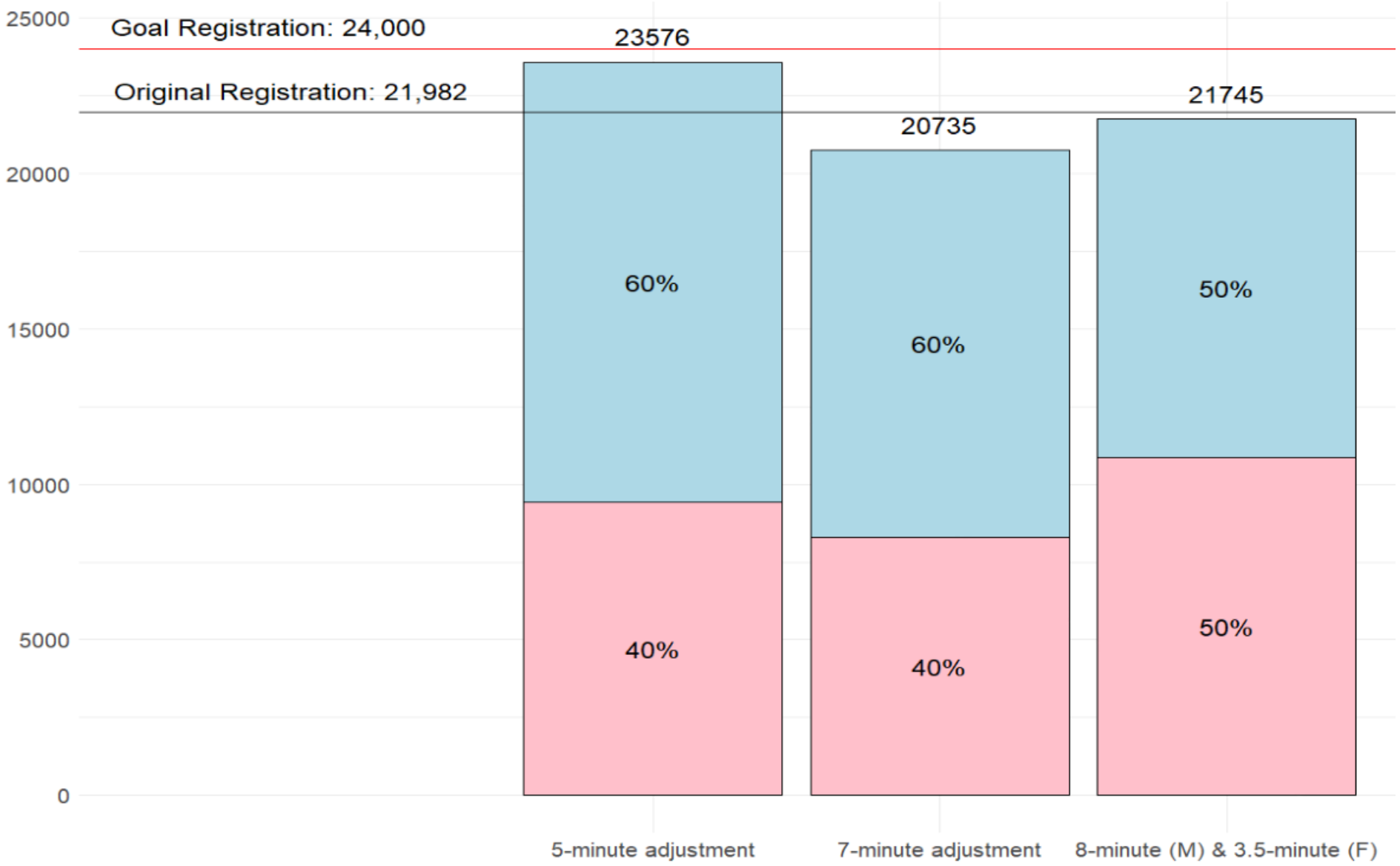
Abstract

The 2013 Boston marathon was disrupted by two bombs placed near the finish line. The bombs resulted in three deaths and several hundred injuries. Of lesser concern, in the immediate aftermath, was the fact that nearly 6,000 runners failed to finish the race. We were approached by the marathon's organizers, the Boston Athletic Association (BAA), and asked to recommend a procedure for projecting finish times for the runners who could not complete the race. With assistance from the BAA, we created a dataset consisting of all the runners in the 2013 race who reached the halfway point but failed to finish, as well as all runners from the 2010 and 2011 Boston marathons. The data consist of split times from each of the 5 km sections of the course, as well as the final 2.2 km (from 40 km to the finish). The statistical objective is to predict the missing split times for the runners who failed to finish in 2013. We set this problem in the context of the matrix completion problem, examples of which include imputing missing data in DNA microarray experiments, and the Netflix prize problem. We propose five prediction methods and create a validation dataset to measure their performance by mean squared error and other measures. The best method used local regression based on a K-nearest-neighbors algorithm (KNN method), though several other methods produced results of similar quality. We show how the results were used to create projected times for the 2013 runners and discuss potential for future application of the same methodology. We present the whole project as an

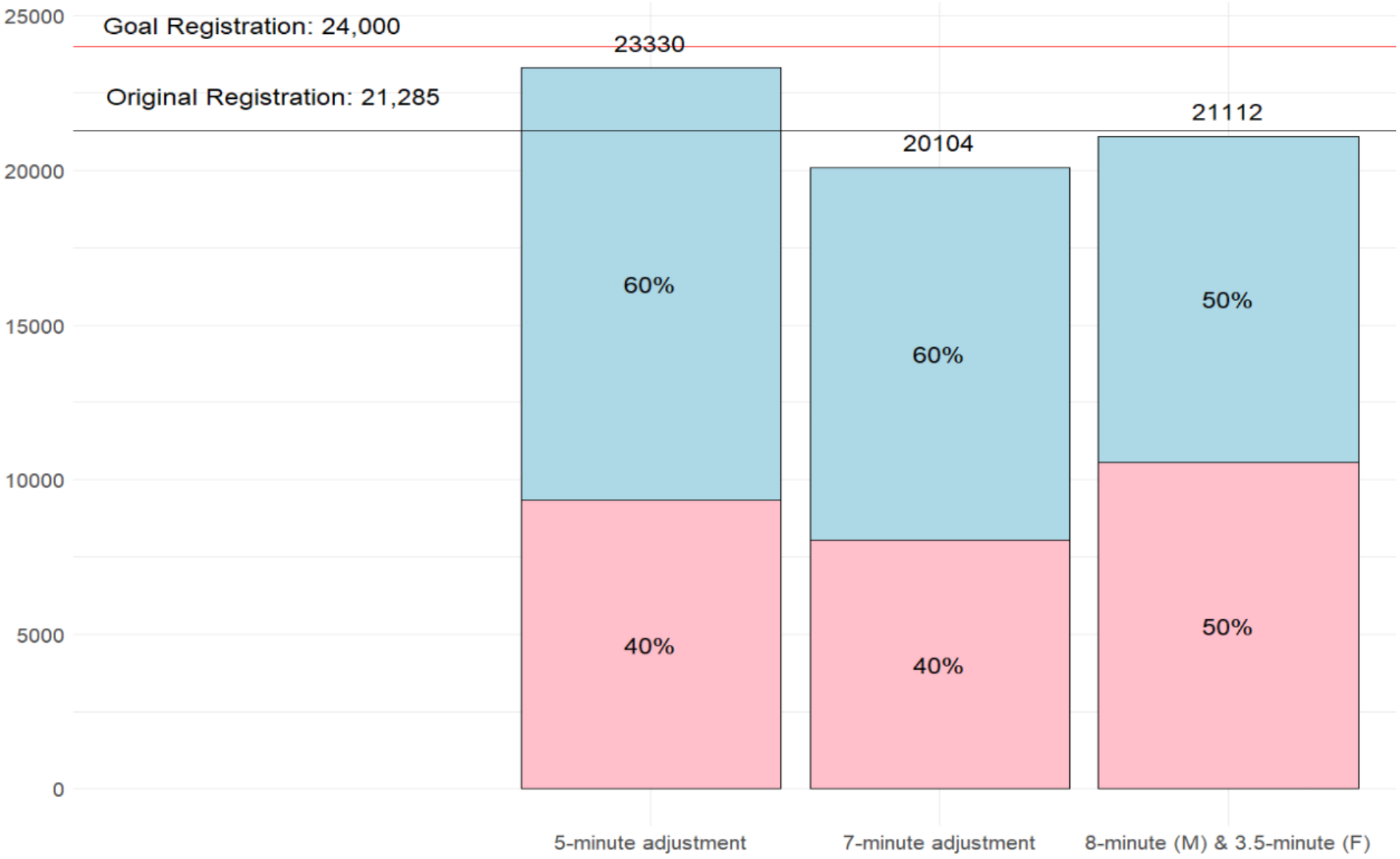
- We became involved again after the 2024 race was over-subscribed

- After several conference calls, the organizers decided to reduce all the qualifying times by 5 minutes for the 2026 race, except for runners over the age of 60

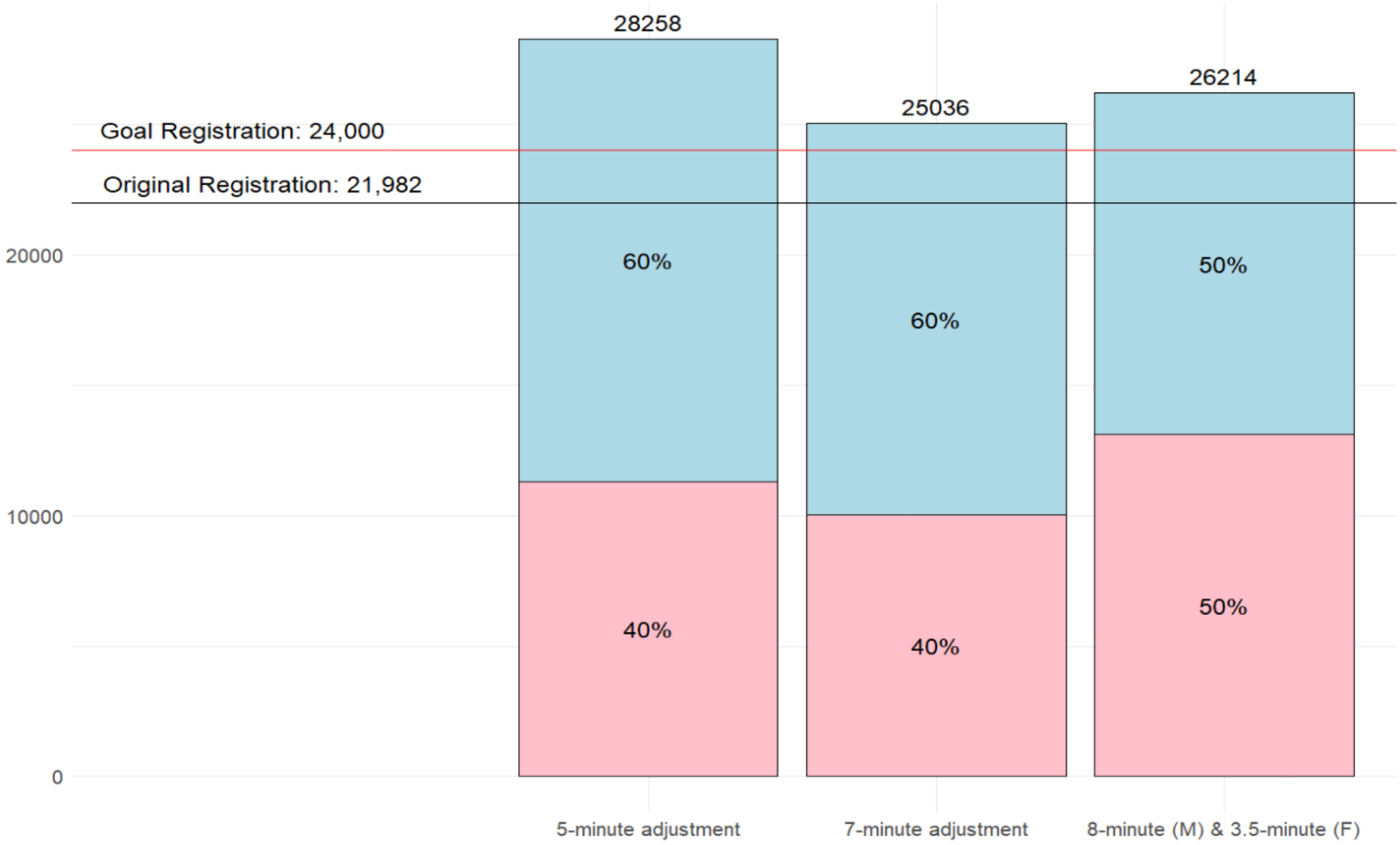
Adjustments pertaining only to 2024 Registration (Includes Withdrawn, Deferred, Approved)



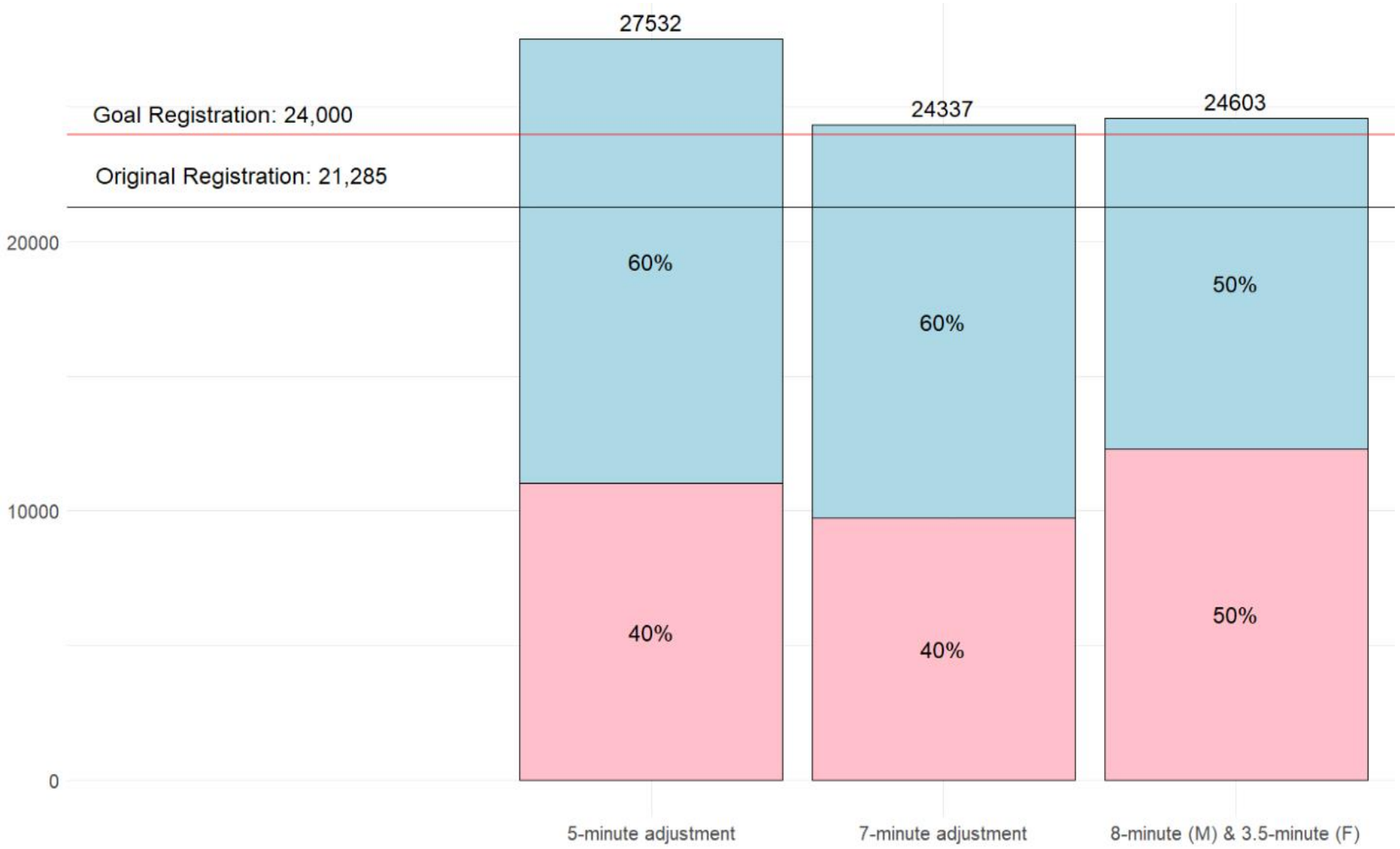
Adjustments pertaining only to 2024 Registration (Includes Deferred, Approved)



Adjustments with estimated 8% increase and 0-3 minutes faster across the board (Includes Withdrawn, Deferred, Approved)



Adjustments with estimated 8% increase and 0-3 minutes faster across the board (Includes Deferred, Approved)



Our Team

- Abigail Mabe (undergraduate statistics major, UNC)
- Jill Myler (undergraduate statistics major, UNC)
- Kellis Ward (graduate student in applied mathematics, Colorado School of Mines)
- Dorit Hammerling (associate professor of applied mathematics and statistics, Colorado School of Mines)
- Richard Smith (professor of statistics, UNC)

Second Digression: Wang Junxia's disputed records from 1993

- In 1993, a whole series of remarkable women's records were set at the Chinese national championships.
- Most notable were the performances of Wang Junxia: 29:31.78 for 10,000 m (first woman under 30 minutes) and 8:06.11 for 3,000 m (previous record was 8:22.62)
- Many suspicions were raised about possible illegal drug use
- How I posed the question (Smith 1997): given that a new world record occurred, what was the probability, based on data prior to 1993, that the record would be as good or better than the one actually achieved?
- A very small probability could be taken as circumstantial evidence or drug use

Appl. Statist. (1997)
46, *No.* 1, *pp.* 123–128

Letter to the Editors

Statistics for Exceptional Athletics Records

Robinson and Tawn (1995) analysed data from 1972 to 1992 in the women's 1500 m and 3000 m running events, with a view to establishing whether the remarkable performances achieved by Chinese athletes in 1993 were statistically inconsistent with previous performances, a conclusion that might be taken as evidence of illegal drug use. Particular attention was paid to the performance of Wang Junxia, who improved the 3000 m record from 502.62 s to 486.11 s. For this, they fitted a model to the five best performances by different athletes in each year. They then constructed 90% confidence intervals, under several variants of the model, for x_{ult}, a parameter representing the long-term limit of performance. Although the analysis provides some grounds for regarding Wang's time as extremely unusual, in all cases the reported confidence interval for x_{ult} included her record time, and to this extent the evidence is inconclusive.

In this letter, I argue that a simpler model, based on fitting part of the data without any trend component, produces very similar results to those of Robinson and Tawn with rather less effort. My main point, however, is that if we look at Wang's time from the point of view of prediction intervals for the observed value, rather than confidence intervals for the hypothetical x_{ult} parameter, then we indeed obtain strong evidence that the performance was a severe outlier.

Smith (1997)

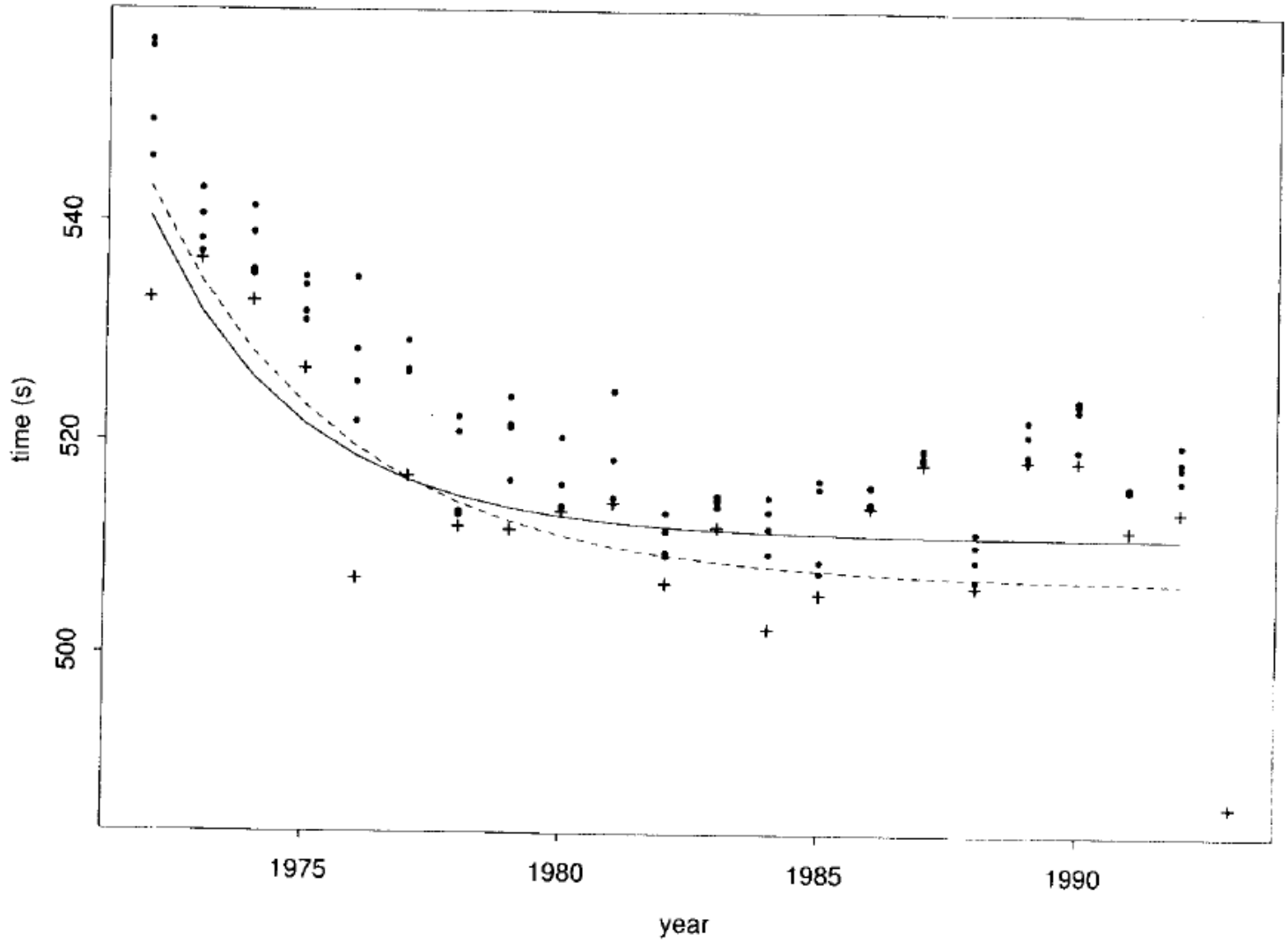


Fig. 1. Five best annual times for the 3000 m: +, annual minima; ———, regression function (1.5), estimated by using model A of Table 1; --------, regression function (1.5), estimated by using model C of Table 1

Robinson and Tawn (1995)

For the 3000 m data based on 1980–92, the posterior probability that $\theta < 486.11$ is 0.076—small, but hardly negligible. In contrast, when averaged according to the posterior distribution of the unknown parameters, the conditional probability of an observed record of less than 486.11, given that a record occurs at all, is 0.00047, which is much smaller.

Plots of the posterior and predictive density, shown in Fig. 2, are indeed of a very different

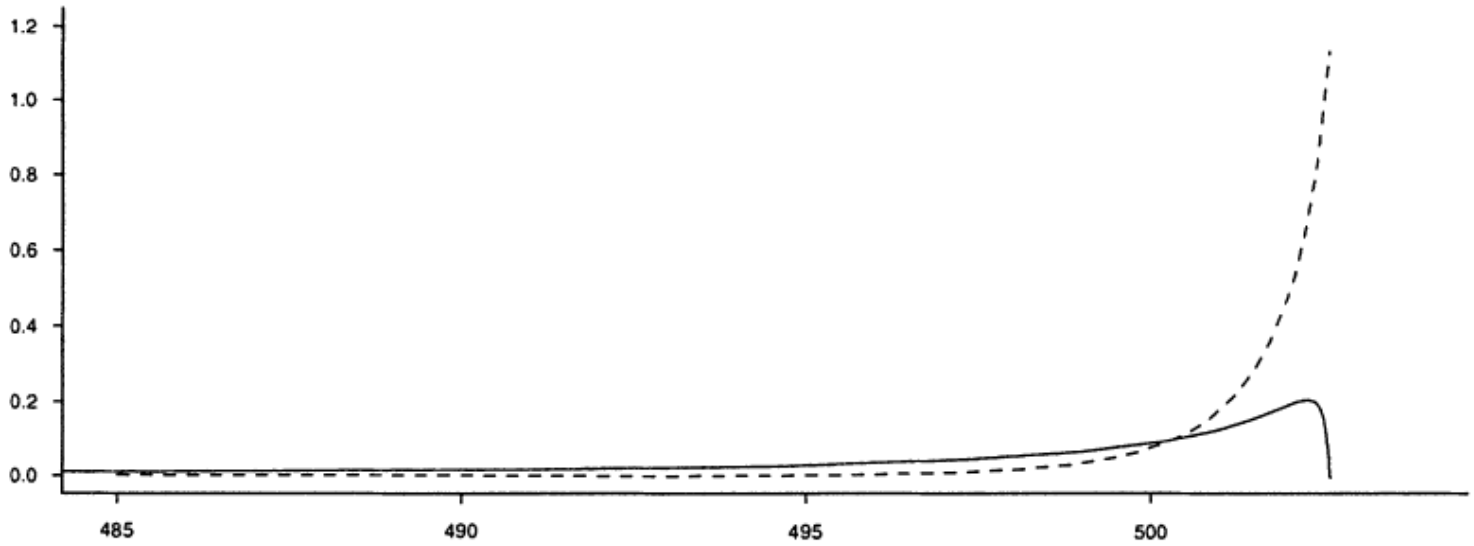


Fig. 2. Bayesian densities: ——— , posterior density; - - - - , predictive density

Athletics world records blow as Wang Junxia 'admits' being part of Chinese state-sponsored doping regime

Ben Bloom

READ MORE ABOUT:
IAAF (International Association of Athletics Federations), China

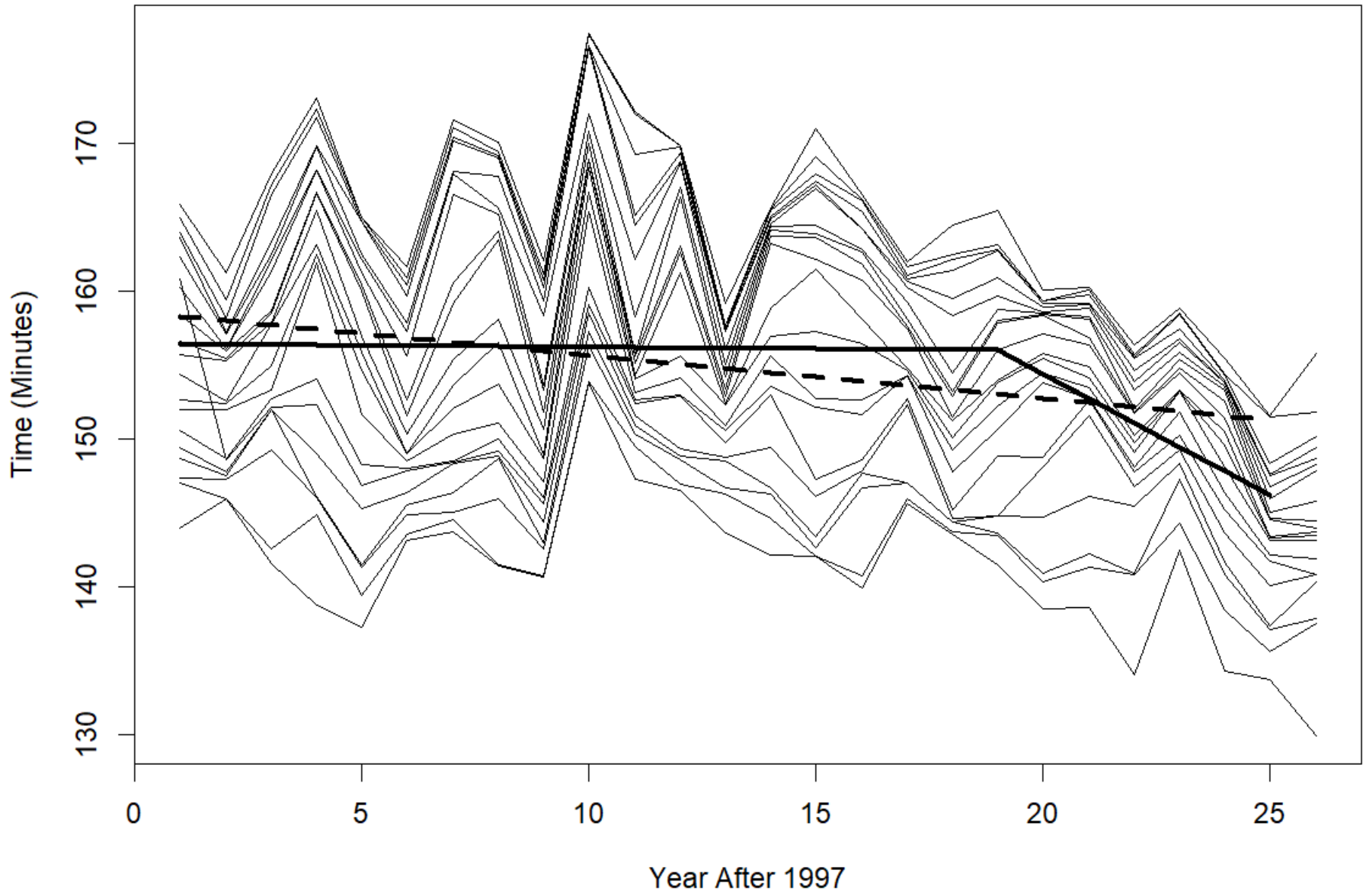
04 February 2016 10:47pm GMT

Gift this article free

Wang Junxia (right) currently holds the 10,000m and 3,000m world records Credit: REUTERS

Back to Chepngetich's Record

- We downloaded data from the Chicago Marathon website, best 20 women's times for each year from 1998–2024
- No race 2020 — we just left out that year
- 2007 is a possible outlier but we discuss that later
- Clear downward trend over the 27 years but strong suspicion of a changepoint in the mid-2010s
- Particular interest in a changepoint in 2016 as Nike Vaporfly shoes were only introduced in 2017


Time (Minutes)
130
140
150
160
170
0
5
10
15
20
25
Year After 1997

Statistical Model

- Generalized Extreme Value (GEV) density for sample minima

$$g(y;\mu,\sigma,\xi) \;=\; \frac{1}{\sigma}\left\{1+\xi\cdot\frac{\mu-y}{\sigma}\right\}^{-1/\xi-1}\exp\left[-\left\{1+\xi\cdot\frac{\mu-y}{\sigma}\right\}^{-1/\xi}\right]$$

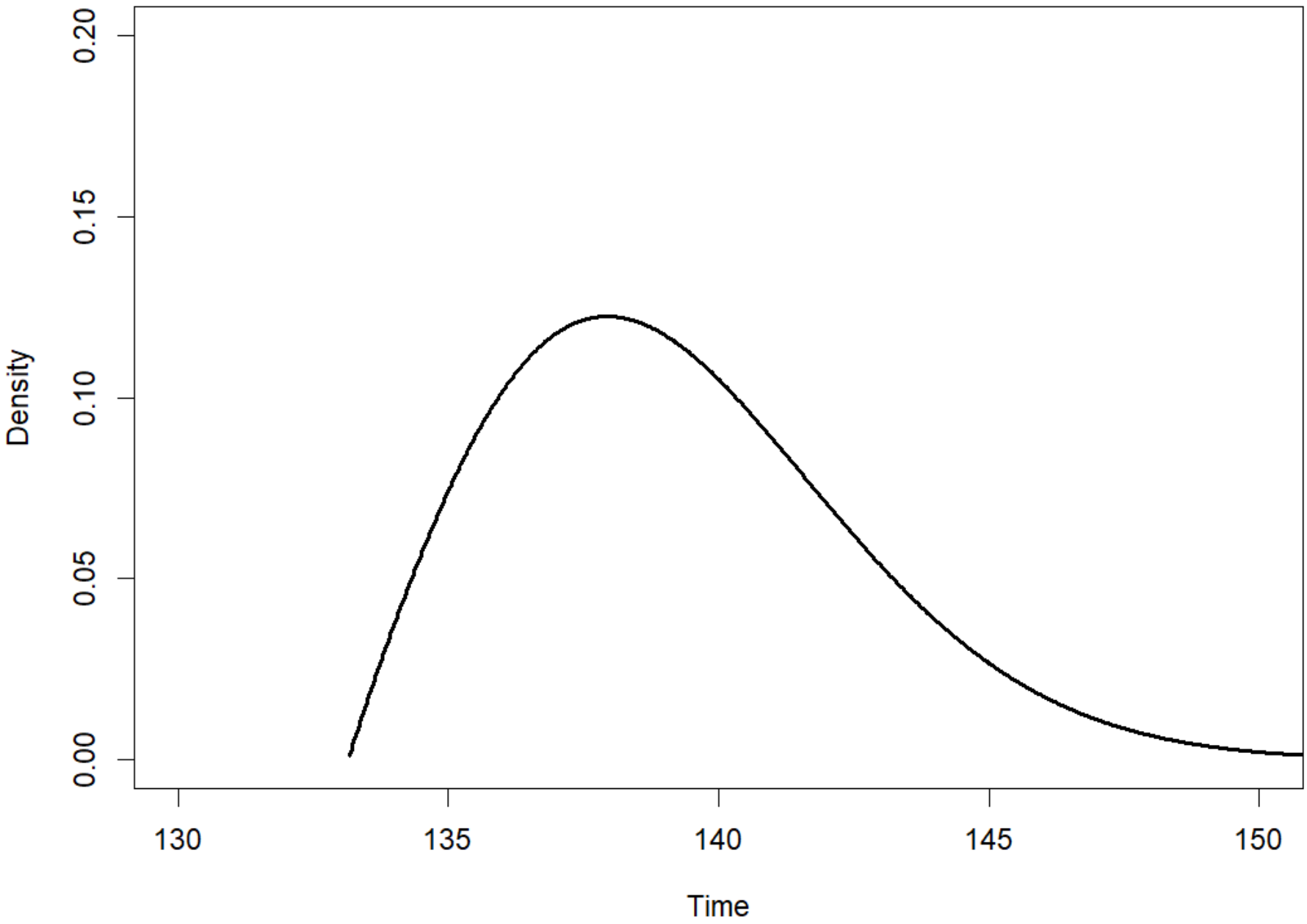
  valid whenever $1+\xi\cdot\frac{\mu-y}{\sigma}>0$.

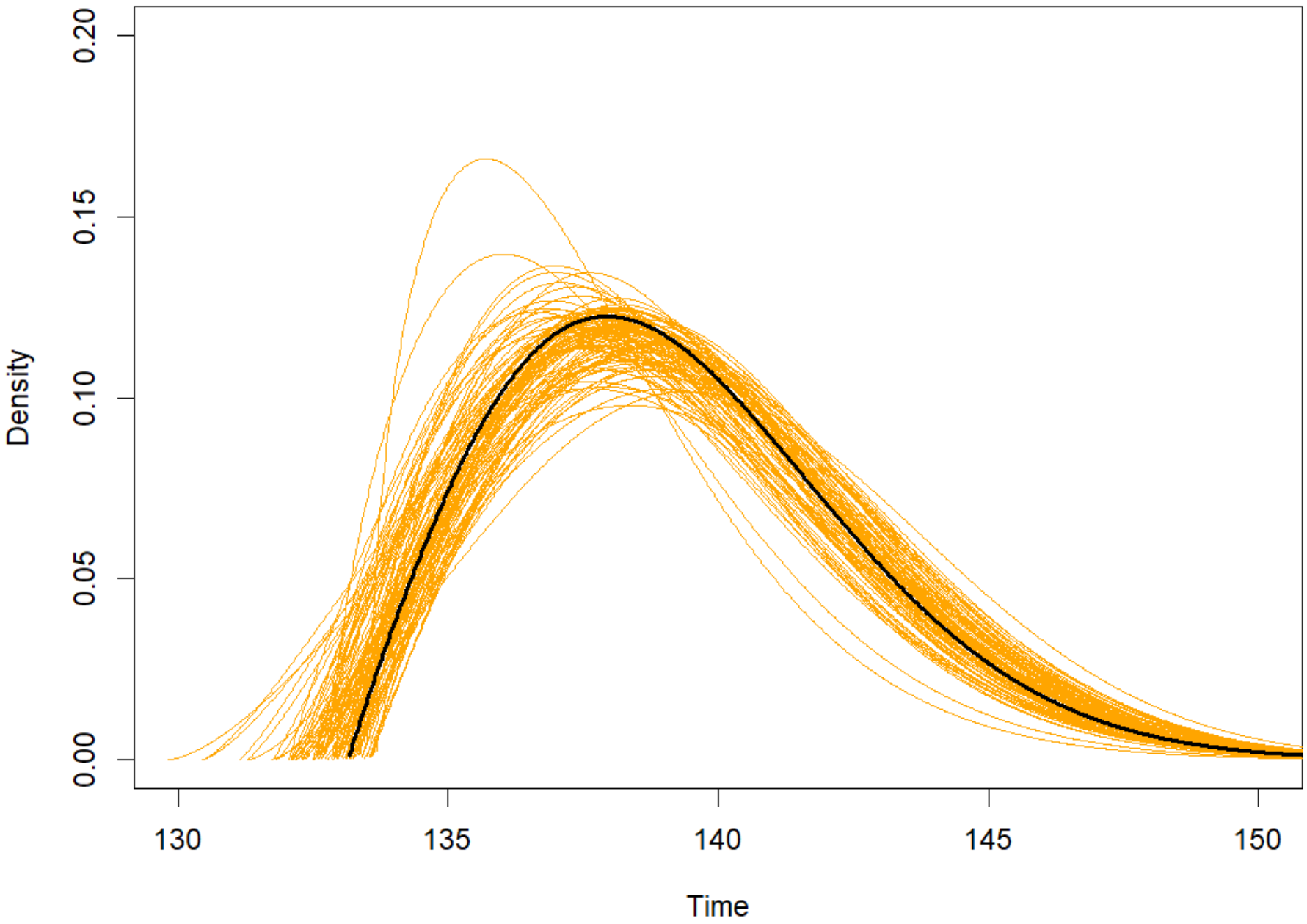
- y represents the winning time in a given year

- Location parameter μ

- Scale parameter σ

- Shape parameter ξ


Density
0.00
0.05
0.10
0.15
0.20
130
135
140
145
150
Time


Density
0.00
0.05
0.10
0.15
0.20
130
135
140
145
150
Time

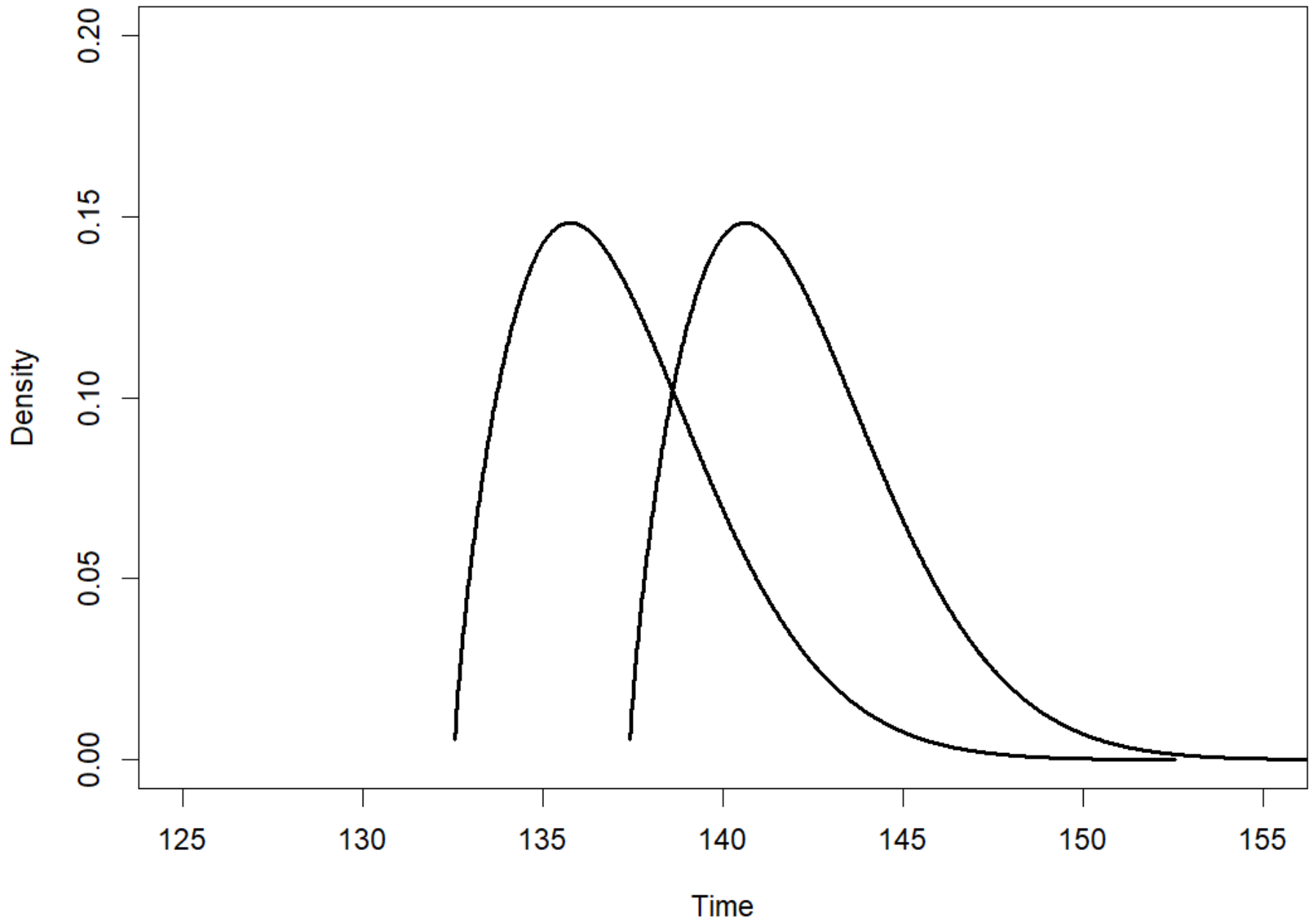
Extensions

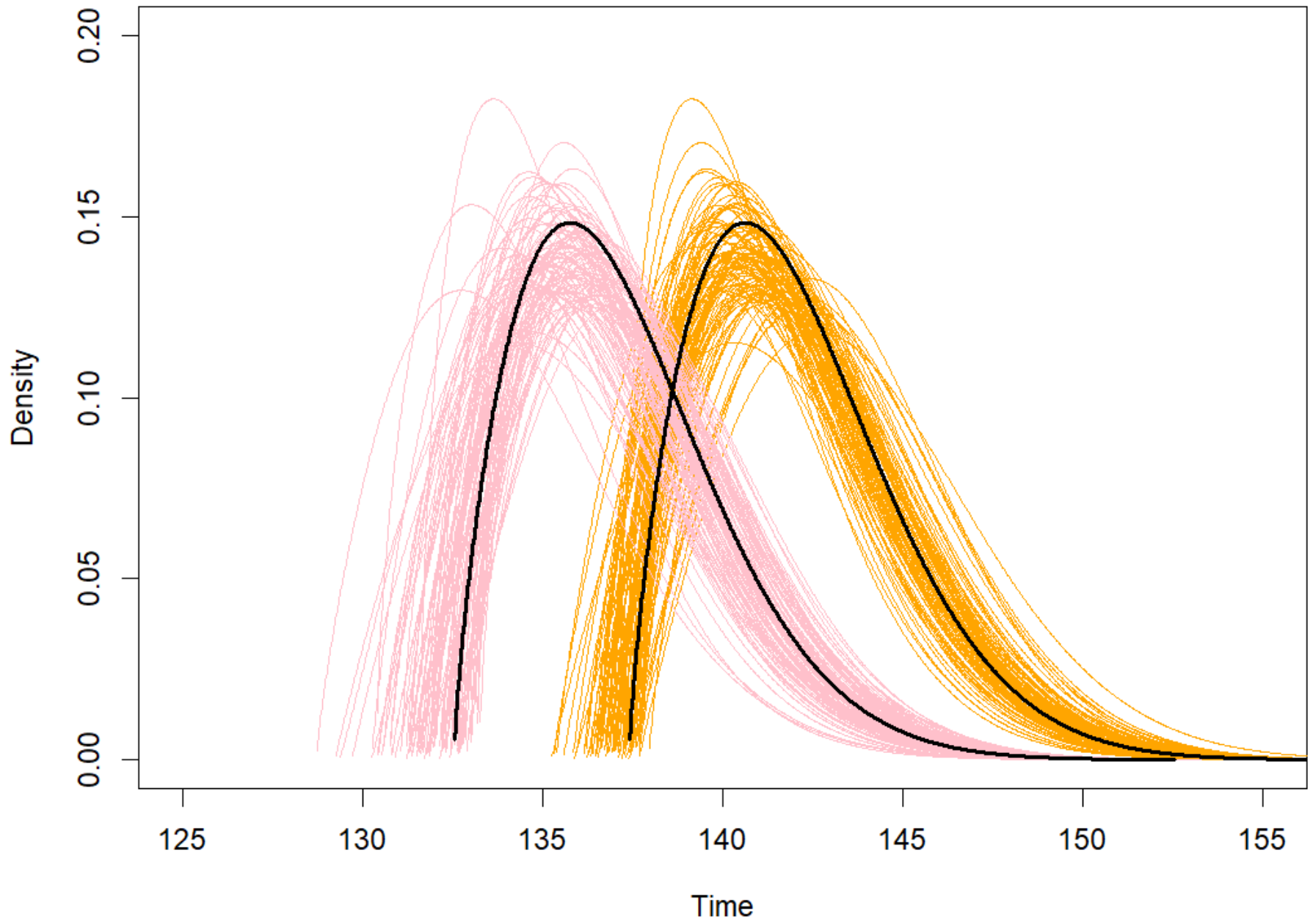
- In practice, we use an extension of this formula that allows us to incorporate the r fastest times in each year, where possible values for r included 5, 10, 15, 20 ("r-largest order statistics model")

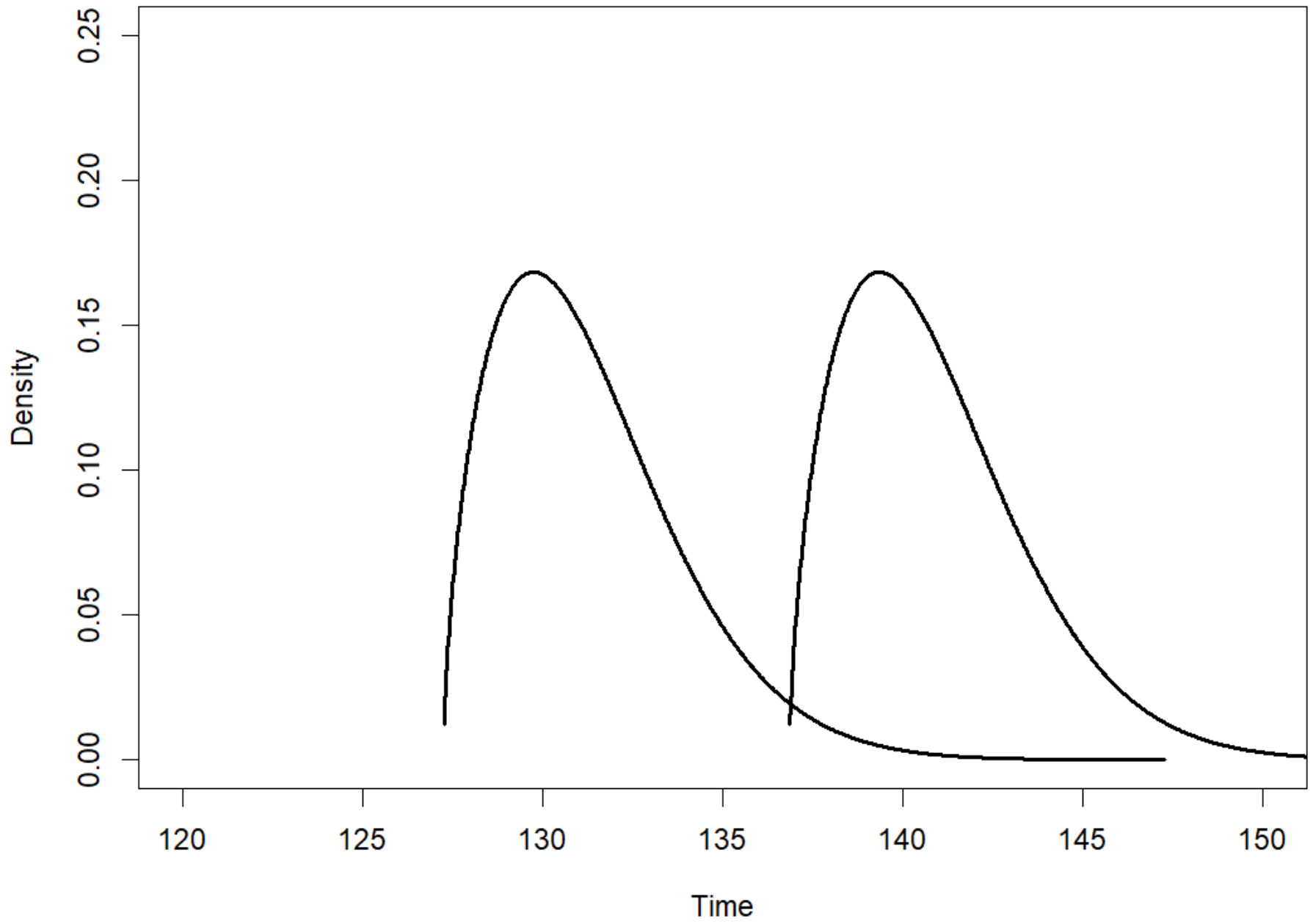
- We apply the formula to every year of the data, not just one year

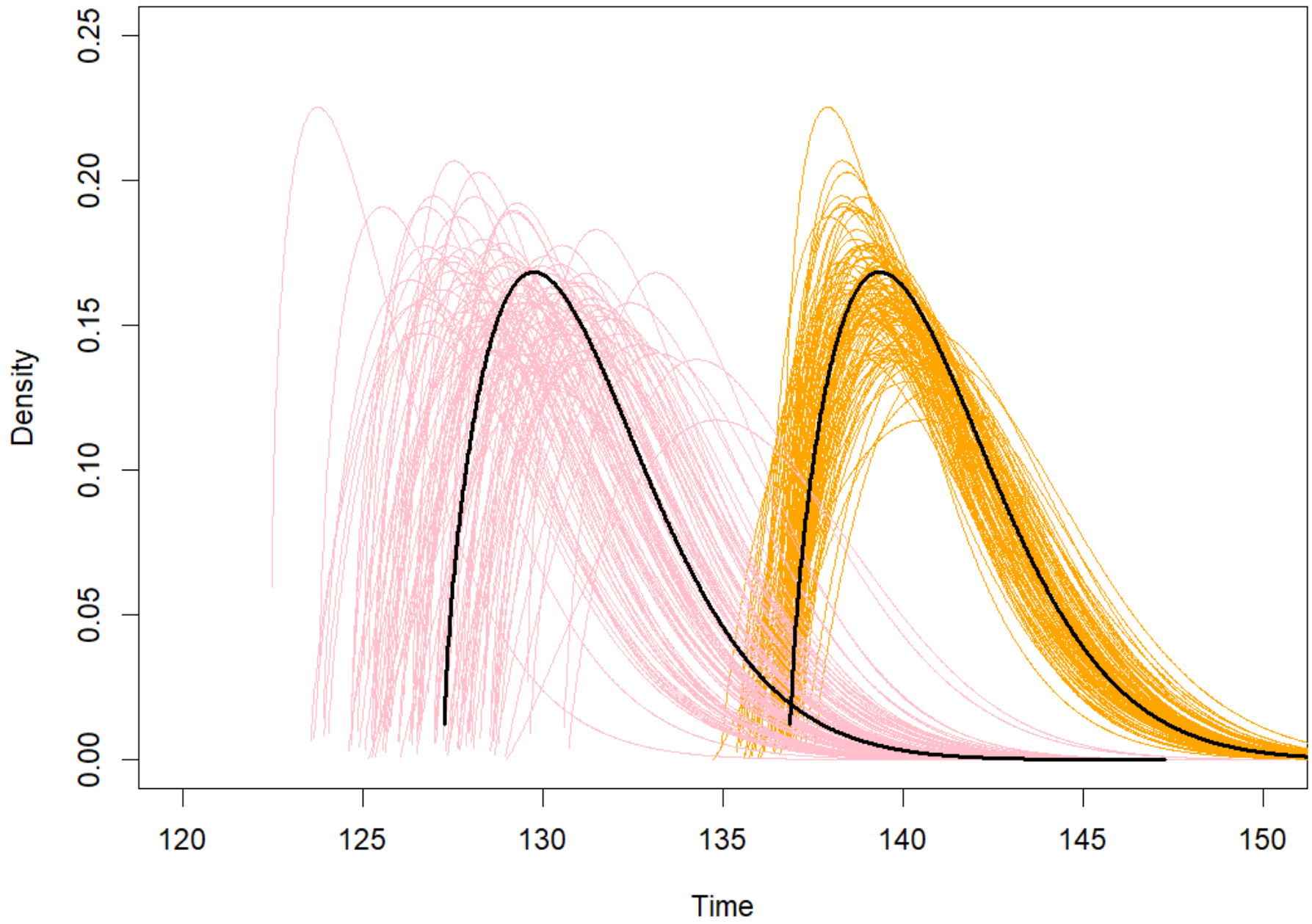
- We also allow the parameter μ to vary with time — write it μ_t in year t

- Possible models for μ_t:
  - Model 1: $\mu_t = \beta_0 + \beta_1 t$ (linear trend)
  - Model 2: $\mu_t = \beta_0 + \beta_1(t - t_0) + \beta_2(t - t_0)_+$ (changepoint in year t_0)


Density
0.00
0.05
0.10
0.15
0.20
125
130
135
140
145
150
155
Time


Density
0.00
0.05
0.10
0.15
0.20
125
130
135
140
145
150
155
Time


Density
0.00
0.05
0.10
0.15
0.20
0.25
120
125
130
135
140
145
150
Time


Density
0.00
0.05
0.10
0.15
0.20
0.25
120
125
130
135
140
145
150
Time

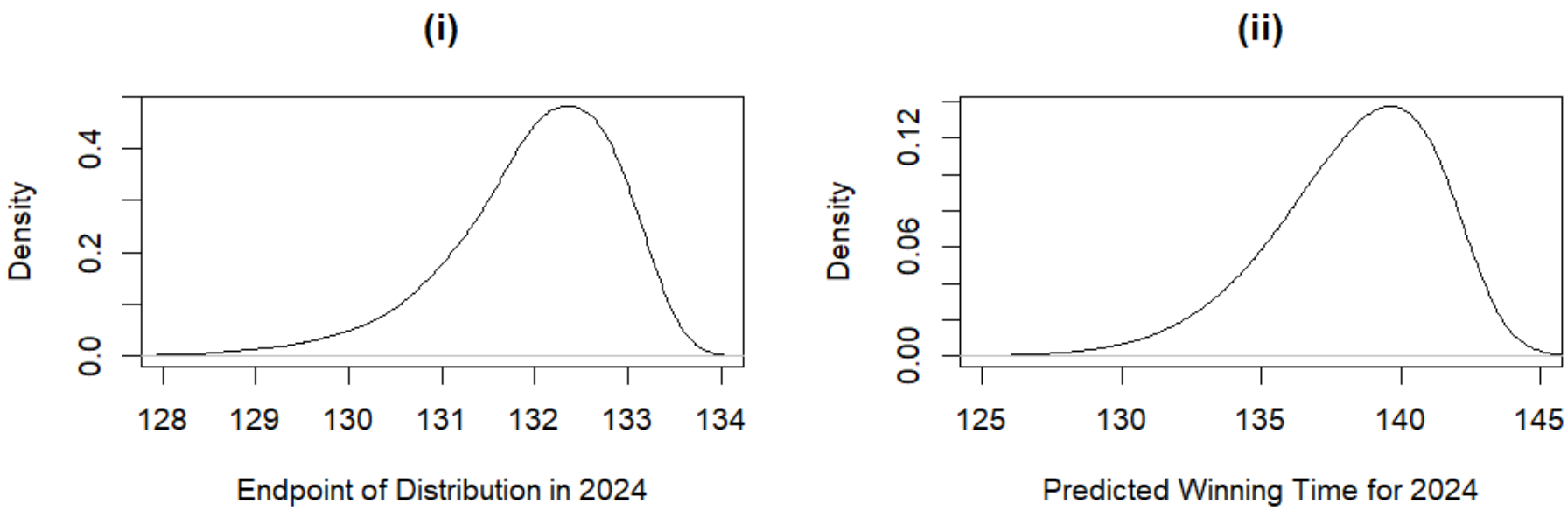
Outline of our method

- Focus on two quantities: the endpoint of the distribution (best possible time) and the actual winning time in 2024, conditional on all data prior to 2024.

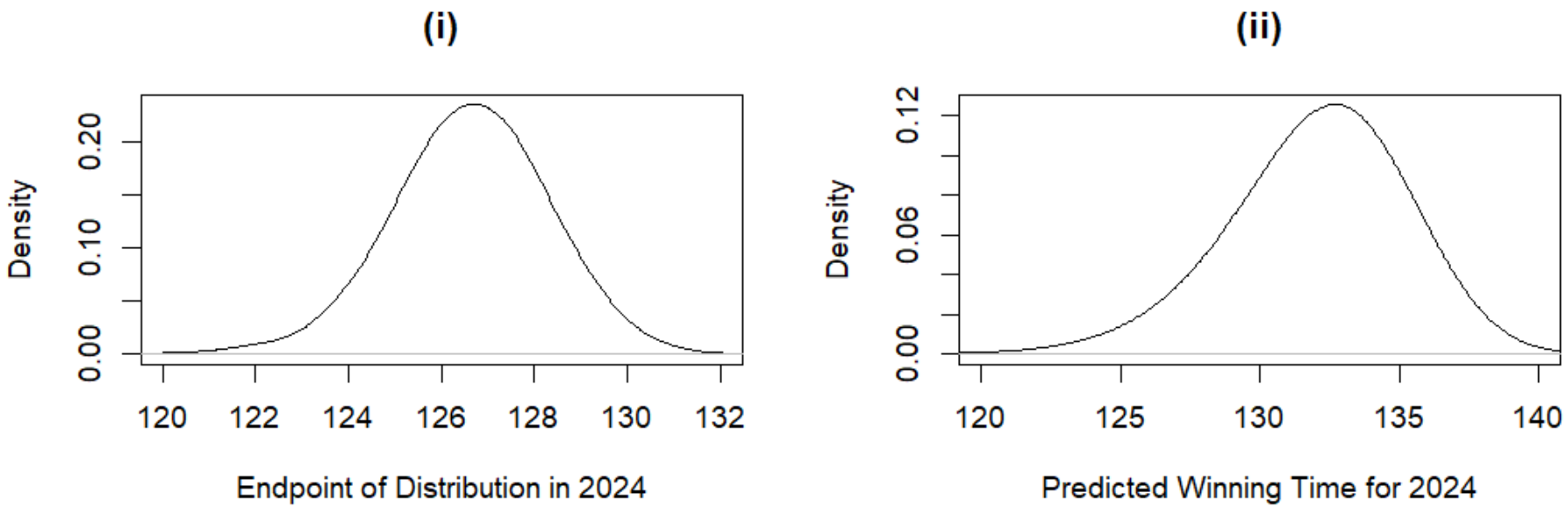
- Compute posterior density curves for both quantities

- Two trend models: linear or changepoint

- Other assumptions: $r = 10$, start year 1998, changepoint in 2016 (for the changepoint model)



Posterior density plots for the endpoint and the predicted 2024 winning time under linear trend. The (estimated) probability that the winning time is < 130 minutes is about 0.011 — "unlikely but not impossible"



Posterior density plots for the endpoint and the predicted 2024 winning time under changepoint model. The (estimated) probability that the winning time is < 130 minutes is about 0.25 — maybe even too large to be believable?

Sensitivity analysis I

Start Year	Endpoint				Winning Time			
	r=5	r=10	r=15	r=20	r=5	r=10	r=15	r=20
1998	0.17	0.034	0.031	0.091	0.004	0.011	0.013	0.014
2002	0.273	0.045	0.067	0.172	0.006	0.011	0.015	0.016
2005	0.211	0.07	0.121	0.213	0.015	0.022	0.024	0.024
2008	0.433	0.481	0.552	0.627	0.055	0.063	0.062	0.053
2005x	0.198	0.061	0.068	0.141	0.01	0.018	0.021	0.019
1998x	0.188	0.03	0.027	0.084	0.004	0.011	0.013	0.014
1998y	0.181	0.028	0.031	0.078	0.004	0.011	0.014	0.014

Posterior probability that the endpoint or the winning time is under 130 minutes, based on linear trend, for four values of r and different starting years.
2005x: analysis starting in 2005 but omitting 2007.
1998x: analysis with changepoint in 2013 instead of 2016.
1998y: analysis with changepoint in 2010 instead of 2016.

Sensitivity analysis II

Start Year	Endpoint				Winning Time			
	r=5	r=10	r=15	r=20	r=5	r=10	r=15	r=20
1998	0.931	0.983	0.992	0.996	0.097	0.25	0.306	0.315
2002	0.967	0.988	0.996	0.996	0.168	0.324	0.392	0.34
2005	0.918	0.964	0.978	0.981	0.182	0.283	0.283	0.256
2008	0.596	0.73	0.771	0.847	0.079	0.116	0.132	0.125
2005x	0.92	0.973	0.976	0.98	0.151	0.262	0.297	0.258
1998x	0.686	0.722	0.796	0.874	0.037	0.069	0.092	0.106
1998y	0.477	0.469	0.584	0.656	0.029	0.057	0.065	0.066

Same as previous table, but based on changepoint model.

Summary and Conclusions

- Trends in the data are clearly significant, and there is also strong evidence that a changepoint model fits better than a linear trend model, though not necessarily based on 2016 as the changepoint

- For estimated probabilities that the winning time for 2024 is below 130 minutes, conditional on previous times up to 2023, is never exceptionally small (smallest is 0.004, but most of our estimated probabilities are quite a bit larger than that), so we cannot say that Chepngetich's performance is "too good to be true"

- The results contrast strongly with similar analyses for Chinese women runners in 1993

- Final punchline:

Summary and Conclusions

- Trends in the data are clearly significant, and there is also strong evidence that a changepoint model fits better than a linear trend model, though not necessarily based on 2016 as the changepoint

- For estimated probabilities that the winning time for 2024 is below 130 minutes, conditional on previous times up to 2023, is never exceptionally small (smallest is 0.004, but most of our estimated probabilities are quite a bit larger than that), so we cannot say that Chepngetich's performance is "too good to be true"

- The results contrast strongly with similar analyses for Chinese women runners in 1993

- Final punchline: **no reason to dispute the record!!**