The Effect of Age on Performance in Marathon Running

Richard L. Smith*

July 31, 2014

Abstract

This paper presents a systematic study of how running times in foot races vary with age. Most existing methods are based either on *ad hoc* comparisons of race results in different age groups, or so-called age-graded performances which are based on world records. I argue that a more appropriate methodology is based on the individual performances of runners who have run races over many years. Using archived online results from the Boston Marathon, I have assembled a database of 750 runners [exact number to be inserted later] who have run the Boston marathon at least six times in the case of men or five times for women. Statistical analysis of these results leads us to an age profile for all runners combined. The analysis is carried out separately for men and women, and can be further subdivided in various ways, for instance, constructing different age-performance curves for different ability groups. The results are compared both with current Boston qualifying times and with the USATF age-graded tables.

1 Introduction

Understanding the aging process is important for anyone who competes in long-distance running events. Races are traditionally divided into five-year age groups for the purpose of awarding prizes as well as for other purposes, including setting qualifying times in the Boston marathon and other races that require or invite qualification[1]. Therefore, it is important to have some basis for de-

*University of North Carolina and SAMSI. Mailing address: Statistical and Applied Mathematical Sciences Institute, 19 T.W. Alexander Drive, P.O. Box 14006, Research Triangle Park, NC 27709-4006. Email: rls@email.unc.edu.
SAMSI is supported by the National Science Foundation under Grant DMS-1127914.

[1]The Boston marathon is the only major road race that requires qualifying times of all runners, though there is also a charity division that does not require a qualifying time. Other races use qualifying times for certain categories

ciding whether a proposed set of qualifying times is fair across different age groups. Some races publish "age-graded" performances in addition to finish times, and there are several online calculators available by which any runner can translate his performances into age-graded equivalents. However, as we discuss here, the methods by which age-graded performances are calculated are questionable. An individual runner may be interested in knowing how her personal profile of race results (how rapidly she slows down with increasing age) compares with other runners. Finally, a better statistical understanding of how race times vary with age could be valuable input into research looking at physiological factors such as VO_2 max[2].

As a concrete example, the current set of Boston Marathon qualifying times is given in Table 1.

Age Group	Men	Women
18-34	3hrs 05min	3hrs 35min
35-39	3hrs 10min	3hrs 40min
40-44	3hrs 15min	3hrs 45min
45-49	3hrs 25min	3hrs 55min
50-54	3hrs 30min	4hrs 00min
55-59	3hrs 40min	4hrs 10min
60-64	3hrs 55min	4hrs 25min
65-69	4hrs 10min	4hrs 40min
70-74	4hrs 25min	4hrs 55min
75-79	4hrs 40min	5hrs 10min
80 and over	4hrs 55min	5hrs 25min

Table 1: Boston Marathon Qualifying Times

These standards were originally introduced in the 1970s, and have changed over time. For example, in 2002 the standards for runners aged 45 and over were slightly relaxed, presumably

of entries, e.g. the New York marathon uses a set of qualifying times that lead to guaranteed entry if certain other conditions are met.

[2]The maximum capacity of an individual's body to transport and use oxygen during incremental exercise; a common measure of physical fitness.

to counter an impression that the previous standards unfairly discriminated against older runners. The most recent updating of Boston qualifying standards, completed in 2011 and fully implemented for the 2013 race, tightened all the standards by five minutes and also introduced a staggered entry system, to allow runners who had beaten their qualifying standards by a larger margin to enter first, but did not change the increments in the qualifying times across different age groups.

The present paper proposes a systematic method for comparing race times across different age groups. To the best of my knowledge, it is the first attempt to do this by a longitudinal study, i.e. one that is based on the performances of the same runner across many years, rather than comparing race results or world record times that, by definition, are comparing different runners at different ages. A longitudinal study should be able to determine more precisely how a "typical" runner slows down with increasing age. For this purpose, we have constructed a dataset that consists of repeat times in the Boston marathon by runners who have completed the race at least six times (five for women — we used a slightly different criterion for women to increase the number of women in the database, though even with this difference, men outnumber women by about 3:1 in our database). We then propose a method of statistical analysis that, in effect, pieces together portions of the age-performance curve from individual runners, and combines them to create an overall age-performance curve. The analysis is performed separately for men and women, and may be further subdivided, e.g. analyzing fast and slow runners separately.

Section 2 describes the method of constructing the dataset. Section 3 describes the analysis of the data, Section 4 discusses the application to qualifying times, and Section 5 summarizes the paper and my conclusions.

2 Data

The possibility of constructing a dataset along the lines described in Section 1 arose after a previous project [1], that involved projecting the finish times of Boston marathon runners who were unable to finish the 2013 race as a result of the bombs that exploded at the finish line. The previous paper described several methods of constructing such projections based on split times taken at intermediate points along the course, and recommended a method based on K nearest neighbors (KNN method).

As a result of that work, I had complete results for the 2010, 2011 and 2013 Boston marathons in a common Excel format. A search for common names among the runners in all three years yielded a preliminary list of 1,272 runners who had run the race all three years. This initial list is surely incomplete for various reasons; for example, the same runner registering under two slightly different names (such as "Richard Smith" and "Richard L. Smith") would not count as a match under this algorithm, but conversely, it would not distinguish two runners who had exactly the same name. However, these issues were corrected in the next phase of the data compilation.

Using the "archive" feature of the Boston marathon website (http://registration.baa.org/cfm_Archive/iframe_ArchiveSearch.cfm), I then searched for each of the 1,272 names during the races for which full results are available online, 2000–2013. For the subsequent analysis, I retained runners who had run at least six times during that period, or five in the case of women. Overall, men outnumber by about 3:1 in this database, and using a slightly weaker inclusion criterion for women partially compensated for that. For each runner in each year, the database included age, sex and city of residence, and these were used, as far as possible, to determine when repeat performances under the same name were in fact the same runner. Inevitably there are still ambiguous cases, usually arising when two runners with the same name and the same age appear with different cities of residence. In such cases, I counted them as a match unless there was some other reason to think they might be different runners. Altogether, this algorithm identified 453 men who had run the race at least six times between 2000 and 2013, and 179 women who had run the race at least five times between 2000 and 2013, a total of 632 runners (28% women). [Note as of 11/10/2013: these are provisional numbers as the database is still being constructed.] For runners who ran in 2013 but did not finish, we used the reconstructed finish times from [1] under the KNN method. Analyses within that paper suggest that 90% of the results by this method are accurate within 4 minutes, which is easily sufficient accuracy for the analyses of the present paper.

As a result of this process, we have a total of 5,632 [still being updated] individual race results for a total of 632 runners over the period 2000–2013. Each record includes year, age, sex, finish time and an indicator of whether the result was official or (in the case of 2013) reconstructed.

3 Analysis of Data

In section, I analyze the data using the following model:

$$\log t_{ij} \quad = \quad \alpha_i + \beta_{y_{ij}} + S(a_{ij}; df) + \epsilon_{ij}, \tag{1}$$

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.

The idea behind model (1) is to take out the two major sources of variability that are not due to age, namely, the overall ability of the runner and the effect of running conditions in a particular year. During the period covered by the study, there were two years, 2004 and 2012, during which the weather was much hotter than usual, and finish times were much slower in those years. Thus, we would expect β_{2004} and β_{2012} to be much larger than in the other years, and this was confirmed by our analysis. Other sources of year to year variability do not appear to be important, e.g., 2011 had a strong following wind, as well as cool conditions, and this resulted in the fastest winning time in any marathon ever — Geoffrey Mutai of Kenya won the race in 2 hours, 3 minutes and 2 seconds, nearly a minute under the then-current world record, but it was not counted as a new world record because the Boston course is not accepted as a world record course (not only because it is a point to point couse, which makes possible a wind-assisted result, but also, because the course has more than a 42 meter drop from start to finish). However, by our results, the 2011 conditions did not especially benefit ordinary runners.

The function $S(a;K)$ is a nonlinear function of age a, of the form

$$S(a;K) \quad = \quad \sum_{k=1}^{K} \gamma_k s_k(a)$$

where $\gamma_1, ..., \gamma_K$ are unknown coefficients and $s_1(\cdot), ..., s_K(\cdot)$ are K fixed basis functions. These could be, for example, the basis functions for either a B-spline or natural spline representation — in this work, I have used a natural splines representation based on the `ns` function in R [3]. The number of degrees of freedom, K, is chosen subjectively, but the main results presented here are for $K = 6$ which was chosen subjectively but which seems appropriate for the amount of nonlinearity in the curves being fitted.

In practice, I have replaced the function $S(a;K)$ by $S(a;K) - S(a;30)$ so that what is actually being modeled is the difference (on a logarithmic scale of time) between performance at age a and performance at age 30. The age 30 was chosen because it is near the minimum of the performance-age curve and so is an appropriate reference point as the runner gets older.

The final step in the specification of the model is to make α_i and β_j random effects. Thus $\alpha_i \sim N[\mu_\alpha, \sigma^2_\alpha]$ with unknown mean and variance $\mu_\alpha,\ \sigma^2_\alpha$, $\beta_j \sim N[\mu_\beta, \sigma^2_\beta]$ with unknown mean and variance $\mu_\beta,\ \sigma^2_\beta$, and finally $\epsilon_{ij} \sim N[0, \sigma^2_\epsilon]$, all of these random terms being mutally independent.

The motivation for making α_i and β_j random effects is as follows. Recall that there are between 5 and 13 races per runner: therefore, as the total sample size (of race times) grows, the number of runners and hence the number of α_i to be estimated grows approximately linearly in the sample size. This is an instance of the classical Neyman-Scott problem [2], which can lead to inconsistent estimators. However, the problem is solved by making the α_i's random effects. The case for making β_j's random as well is less clear-cut, but I argue that we are not primarily interested in estimating the year effect for any given year: rather, our intention in including this effect is to make an overall adjustment for the year to year variability of race conditions. It seemed to me that this was best taken into account by treating year as a random effect: the resulting estimates of the age effect take into account the existence of a year effect but are not adapted to any specific year.

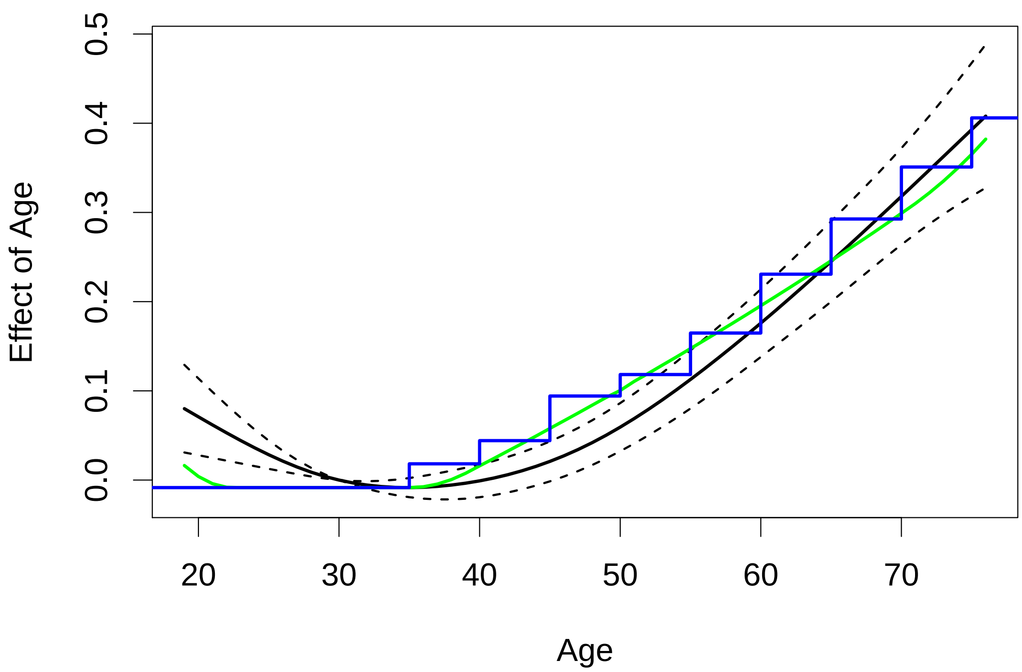

Figure 1: Age-performance curve for men. The solid black curve is the estimated spline curve of the model (1), relative to age 30, and the dotted black curves are pointwise 95% confidence bounds. The blue step function is the assumed relationship in the Boston marathon quaifying times, and the green curve is the implicit age-performance curve used by the USATF for age-graded times. Both the blue step function and the green curve are normalized so that their minimum is the same as that for the estimated spline curve.

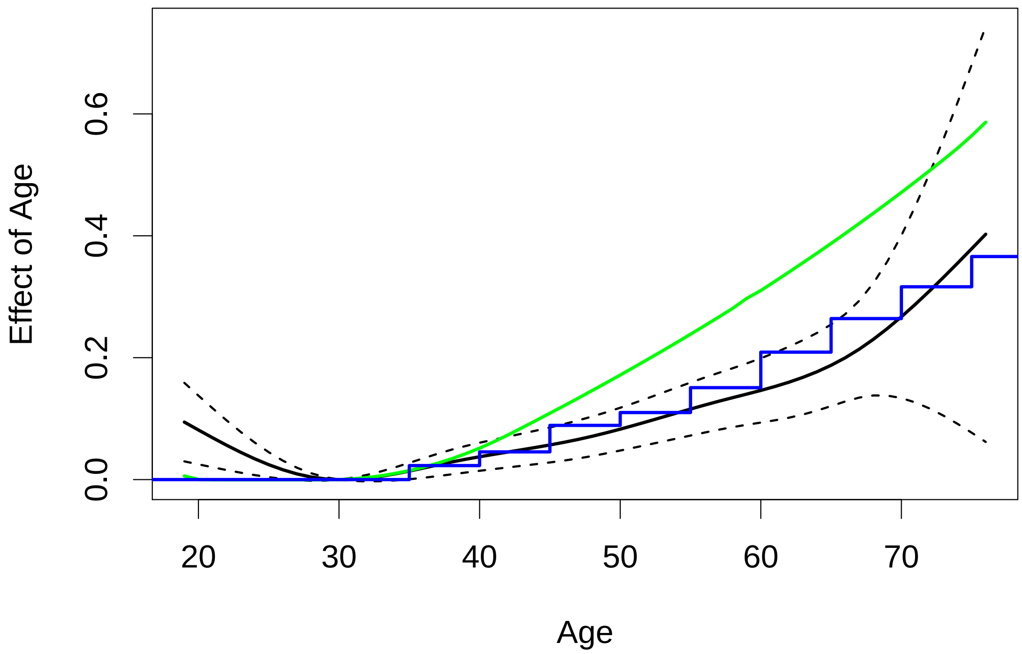

Figure 2: Same as Figure 1, but for women's performances.

4 Discussion

5 Conclusions

References

[1] D. Hammerling, M. Cefalu, J. Cisewski, F. Dominici, G. Parmigiani, C. Paulson, and R.L. Smith. Completing the results of the 2013 Boston Marathon. *PLOS ONE*, xx:xx–yy, 2013.

[2] J. Neyman and E.L. Scott. Consistent estimates based on partially consistent observations. *Econometrica*, 16:1–32, 1948.

[3] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0.