Creating Road Running Age-Grade Tables

# Alan Jones, 2020-04-22

Rough draft

# Introduction

The history of the age-grading can be viewed here: [Age-Grading](http://www.runscore.com/Alan/AgeGrade.html). (This information will be moved to the Age-Grade Tables GitHub repository.)

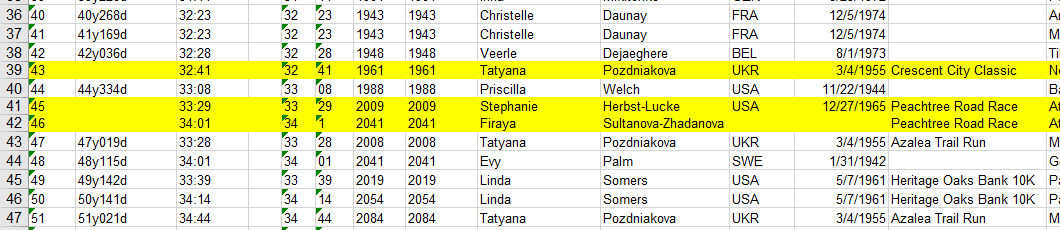
The tables are created in two Excel spread sheets – one for males and one for females.

Until late 2019, data for the tables are from the Association of Road Racing Statisticians: [ARRS.net](arrs.net) run by Ken Young: [kcy@frontiernet.net](mailto:kcy@frontiernet.net). Unfortunately, Ken died in early 2018. Since then Tom Bernhard has been providing tables of single age bests.

In this document, only the Female tables are described. The same procedure is used for the Male tables.

# Single-Age Bests

Tom’s Excel files have separate sheets for each distance. Here is an example for the female 10 km bests:



When Tom produces these files, if there is a new record for a given age and distance, he leaves the old record in and colors that line red. The new record line is colored yellow. In his files in this repository, the red lines have been removed so that only the current records are included and can be easily copied to the Age-Grade Excel files.

# The Excel Spread sheets

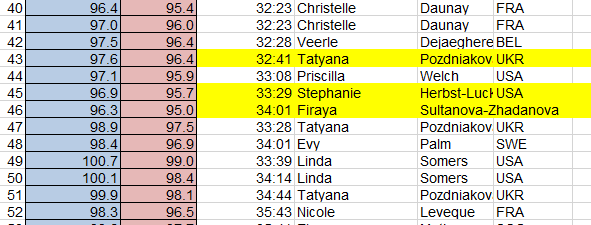
The male sheet is called **maleRoad2020.xlsx** and the female called **femaleRoad2020.xlsx**.

Each file is composed of many sheets:

* Parameters: World records for each distance plus a lot of other data.
* One each for each of the distances: 5 km, 6 km, 4 mi, 8 km, 5 mi, 10 km, 7 mi, 12 km, 15 km, 10 mi, 20 km, Half-Marathon, 25 km, 30 km, Marathon, 50 km, 100 km, 200 km.
* Age-factors. There is an age-factor for each age for each distance. When and age-factor is multiplied by a runner’s time, it will produce a time that this person should have been able to run the given distance when 27 years old.
* Age Standards in Seconds
* Age Standards in H:MM:SS format
* Pace: The pace is computed for several ages over the entire distance range. If there are any errors, they will show up as bumps on the pace plot.

# Import single-age bests

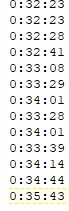
The first step is to copy Tom’s sheet for the event into femaleRoad2020.xlsx. See Sheet 10K.



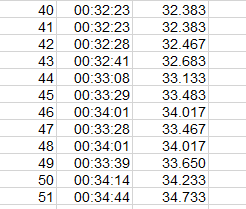
The data above, Single-Age Bests are copied into a text editor – not a word processor. The editor, TextPad, can do “block select”. In this way one can cut out the time column in the table and round the fractional seconds.



The must be hour digits. They are added and fractional seconds rounded up – if there are any.



This is then copied into column B, Records, of the appropriate sheet. The first column is the age, the second the time in h:mm:ss format, and the third is the time in minutes to three decimal places. The time in minutes is produced by multiplying the time by 1440. Why 1440? Because Excel stores times as fractions of a day. For example, a time of 1 hours is stored as 1/24 = 0.041666… Multiply that by 1440 and you get 60, the number of minutes in an hour.



# Creating the age standards

The age-grade standards are based on the mathematical expressions:

af = 1 - C(c - b)^2 - B(b - x) - A(a - x)^2 for x < a

af = 1 - C(c - b)^2 - B(b - x) for x >= a and x < b

af = 1 - C(c - x)^2 for x >= b and x < c

af = 1 for x >= c and x < d

af = 1 - D(x - d)^2 for x >= d and x < e

af = 1 - D(e - d)^2 - E(x - e) for x >= e and x < f

af = 1 - D(e - d)^2 - E(x - e) - F(x - f)^2 for x >= f

See figure below. Notice that there are 12 parameters: *a, b, c, d, e, f, A, B, C, D, E, and F*. However, we want to require that the slope is continuous from the quadratic sections to the linear sections. With this requirement, we can determine the value of two of these parameters in terms of the others reducing the number of independent variables to ten. These two additional equations are:

D = E/(2\*(e - d))

C = B/(2\*(c - b))

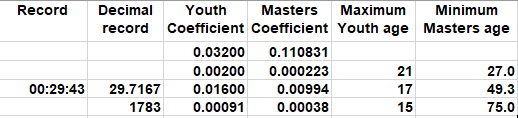
The ten parameters can be adjusted to fit the single age records. The factor (*f*) in the age-graded tables is always a number equal to or less than one. When doing age-grading, a person's time is multiplied by the factor *af* to obtain a time that this person should be able to run as an open athlete.

Below is a graph of the above expression along with its inverse. Note that the linear sections are linear only in the age-factor. When the inverse is plotted, the curve is always concave upwards as shown in the figure.

A close up of a map

Description generated with very high confidence

Those parameters are represented by cells. Below is the 5K sheet:

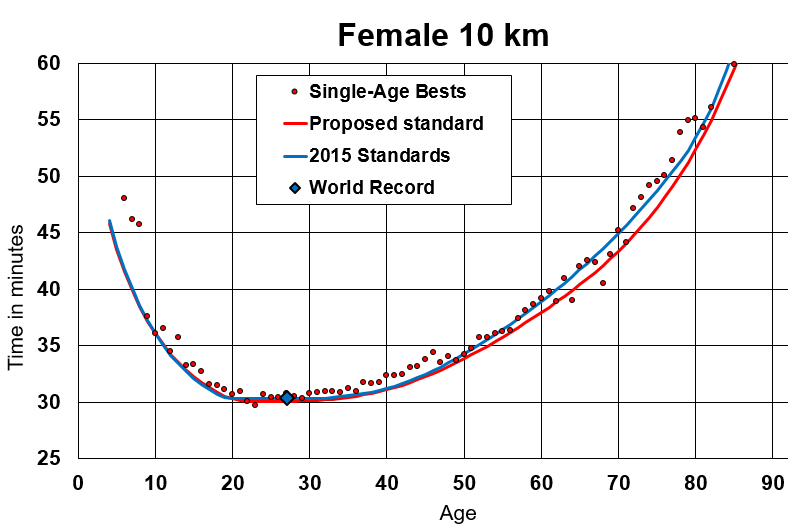


The parameters for masters, *e* and *f*, are represented by cells G4 and G5 respectively and *D*, *E*, and F by cells in the J column: I3, I4, and I5 respectively.

The parameters for youth, *b* and *a*, are represented by the cells F4 and F5 while the parameters *C*, *B*, and *A* are represented by the cells H3, H4, and H5 respectively.

We will set aside the youth parameters for now. The youth data are rather sparse; we find w can fit all distances with the same youth parameters. We should look more closely at this in the future as the data improve.

Here is the plot for the road 10 km. This distance seems to need the most adjustment from the 2015 values.



Notice the values for 64 and 68 years old. Here are the data for those points:

64 64y014d 38:57 38 57 2337 2337 Angela Copson ENG 4/20/1947

68 68y020d 40:30 40 30 2430 2430 Angela Copson ENG 4/20/1947

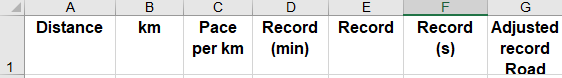
We are declaring these as outliers since they are so far out of the trend of all the other ages. We make no statement about the age of the runner, the length of the course, or drug use. Just that they are outliers.

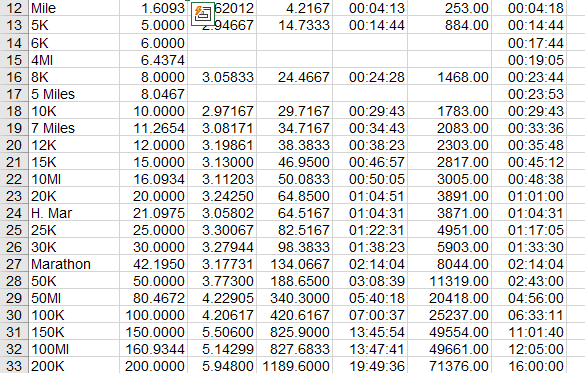
As mentioned above, we only perform the fit for the distances 5 km, 10 km, half-marathon, and marathon. These distances are contested much more than any of the others. Therefore, there exists a large body of data. The age-factors for the other distances are obtained by interpolation in semi-log space. To check how well the interpolation has done, go to the Pace sheet where you will see the following graph:

# Creating Open Class times for intermediate distances

As mentioned above, we only fit the age-factor curve to the data for distances: 5 km, 10 km, half-marathon, and marathon. The other distances are handled by interpolation.

The first step is to compute the Open Class time for each distance.





The record for each distance is in column “Record” (E). The adjusted records are in column “Adjusted record Road” (G). Notice that these values are the same for distances 5 km, 10 km, H. Mar, and Marathon. All of the other records are adjusted, by eye, so that the pace vs distance is a smooth curve. Here is that plot:

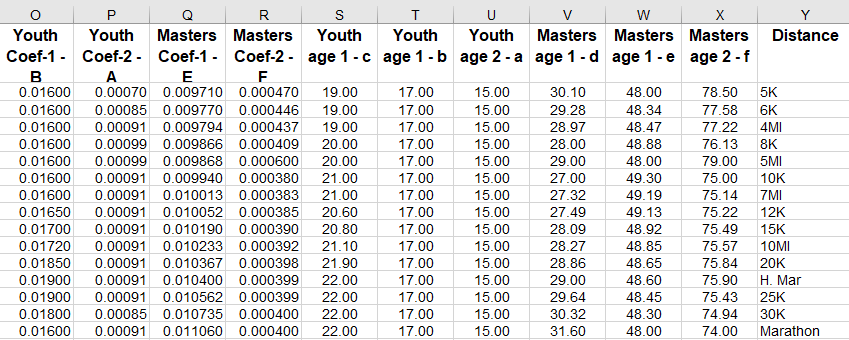
The red diamonds are the records for the 5 km, 10 km, half-marathon, and marathon. The blue triangles are track records for those same distanced. They are included for reference. We don’t actually use them.

As one adjusts the record, a very small adjustment, particularly for the shorter distances, will result in noticeable movements on the plot. The goal is to produce a smooth line connecting the records.

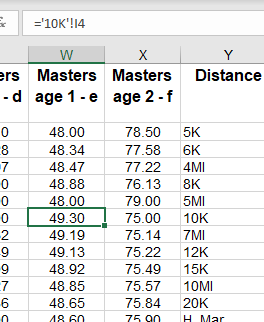
It seems strange that the pace for the female 5 km is very close to the pace for the 10 km.

# Interpolation

The interpolation is done on the first sheet – Parameters. Below you can see how the different parameters used in the curve-fitting a indicated in the top row. These values are filled in automatically from each distance sheet.

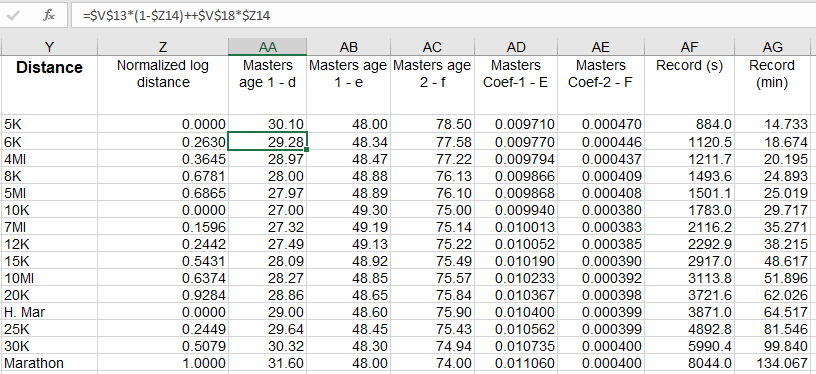


For example, the screen grab below, you can see that the cell highlighted in the Masters age-2 f column gets its value from the 10K sheet, cell “I4”.



To the right of the array above is another array containing the interpolated values.

Columns AA to AE contain the interpolated values. An interpolation is done from 5 km to 10km, 10 km to half-marathon, and half-marathon to marathon in log space.



The highlighted cell AA13 shows, in the command line, the interpolation formula. It used the log of the distance in each of the above intervals. Those log distances are in column Z.