Creating Road Running Age-Grade Tables

Alan Jones, 2020-06-20

Introduction

The history of the age-grading can be viewed here: Age-Grading. (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 run by Ken Young: 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:

36	40	40y268d	32:23	32	23	1943	1943	Christelle	Daunay	FRA	12/5/1974		A
37	41	41y169d	32:23	32	23	1943	1943	Christelle	Daunay	FRA	12/5/1974		M
38	42	42y036d	32:28	32	28	1948	1948	Veerle	Dejaeghere	BEL	8/1/1973		Ti
39	43		32:41	32	41	1961	1961	Tatyana	Pozdniakova	UKR	3/4/1955	Crescent City Classic	N
40	44	44y334d	33:08	33	08	1988	1988	Priscilla	Welch	USA	11/22/1944		B
41	45		33:29	33	29	2009	2009	Stephanie	Herbst-Lucke	USA	12/27/1965	Peachtree Road Race	A
42	46		34:01	34	1	2041	2041	Firaya	Sultanova-Zhadanova			Peachtree Road Race	A
43	47	47y019d	33:28	33	28	2008	2008	Tatyana	Pozdniakova	UKR	3/4/1955	Azalea Trail Run	M
44	48	48y115d	34:01	34	01	2041	2041	Evy	Palm	SWE	1/31/1942		G
45	49	49y142d	33:39	33	39	2019	2019	Linda	Somers	USA	5/7/1961	Heritage Oaks Bank 10K	P
46	50	50y141d	34:14	34	14	2054	2054	Linda	Somers	USA	5/7/1961	Heritage Oaks Bank 10K	P
47	51	51y021d	34:44	34	44	2084	2084	Tatyana	Pozdniakova	UKR	3/4/1955	Azalea Trail Run	M

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, 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.

40	96.4	95.4	32:23	Christelle	Daunay	FRA
41	97.0	96.0	32:23	Christelle	Daunay	FRA
42	97.5	96.4	32:28	Veerle	Dejaeghere	BEL
43	97.6	96.4	32:41	Tatyana	Pozdniakov	UKR
44	97.1	95.9	33:08	Priscilla	Welch	USA
45	96.9	95.7	33:29	Stephanie	Herbst-Luc	USA
46	96.3	95.0	34:01	Firaya	Sultanova-Zhadanova	
47	98.9	97.5	33:28	Tatyana	Pozdniakov	UKR
48	98.4	96.9	34:01	Evy	Palm	SWE
49	100.7	99.0	33:39	Linda	Somers	USA
50	100.1	98.4	34:14	Linda	Somers	USA
51	99.9	98.1	34:44	Tatyana	Pozdniakov	UKR
52	98.3	96.5	35:43	Nicole	Leveque	FRA

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.

```
32:23
32:23
32:28
32:41
33:08
33:29
34:01
33:28
34:01
33:39
34:14
34:44
35:43
```

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

```
0:32:23
0:32:23
0:32:28
0:32:41
0:33:08
0:33:29
0:34:01
0:33:28
0:34:01
0:33:39
0:34:14
0:34:44
0:35:43
```

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.

40	00:32:23	32.383
41	00:32:23	32.383
42	00:32:28	32.467
43	00:32:41	32.683
44	00:33:08	33.133
45	00:33:29	33.483
46	00:34:01	34.017
47	00:33:28	33.467
48	00:34:01	34.017
49	00:33:39	33.650
50	00:34:14	34.233
51	00:34:44	34.733

Creating the age standards

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

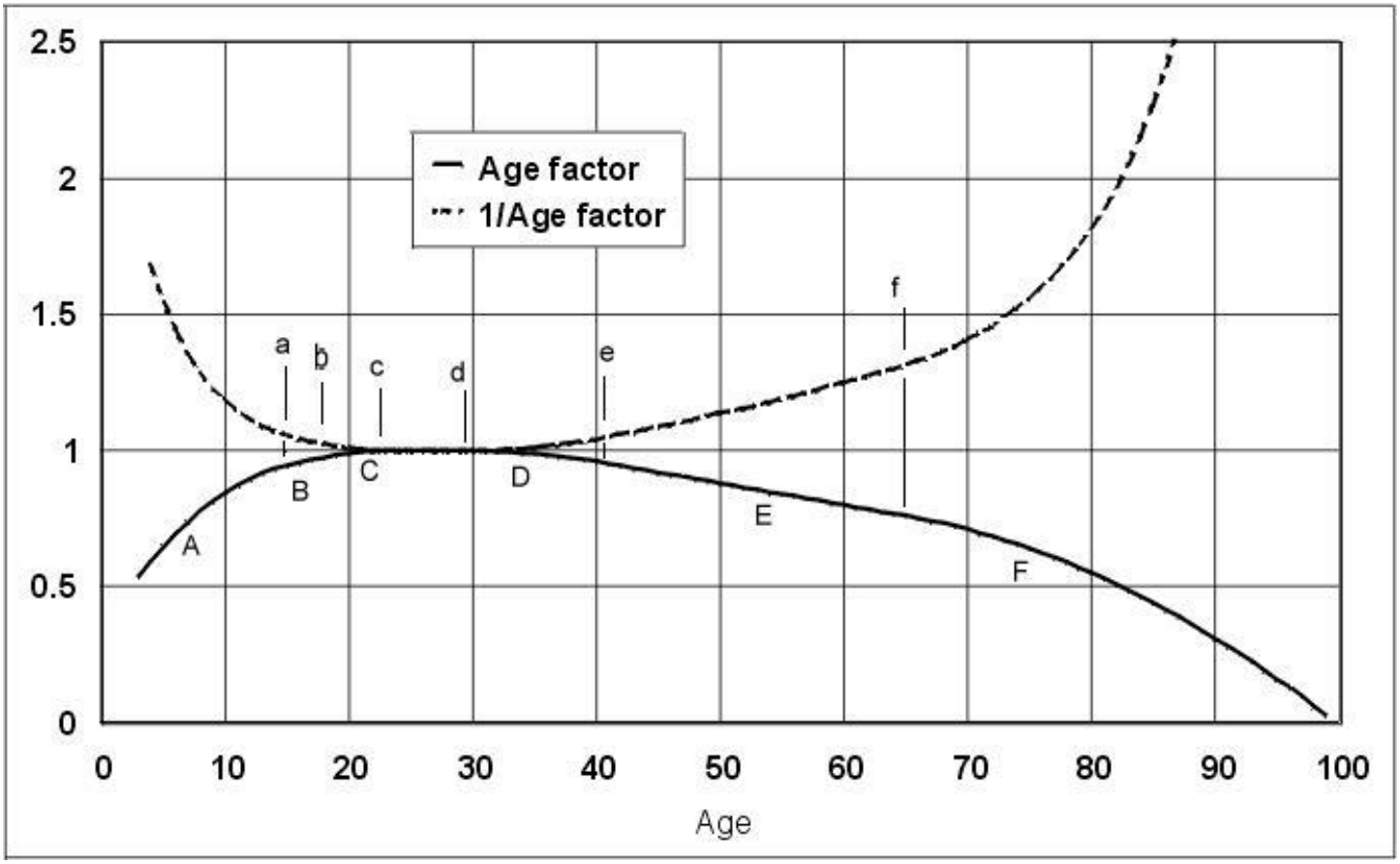
```
af = 1 - C(c - b)^2 - B(b - x) - A(a - x)^2 for x < a
af = 1 - C(c - b)^2 - B(b - x)              for x >= a and x < b
af = 1 - C(c - x)^2                         for x >= b and x < c
af = 1                                      for x >= c and x < d
af = 1 - D(x - d)^2                         for 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 be 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.



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

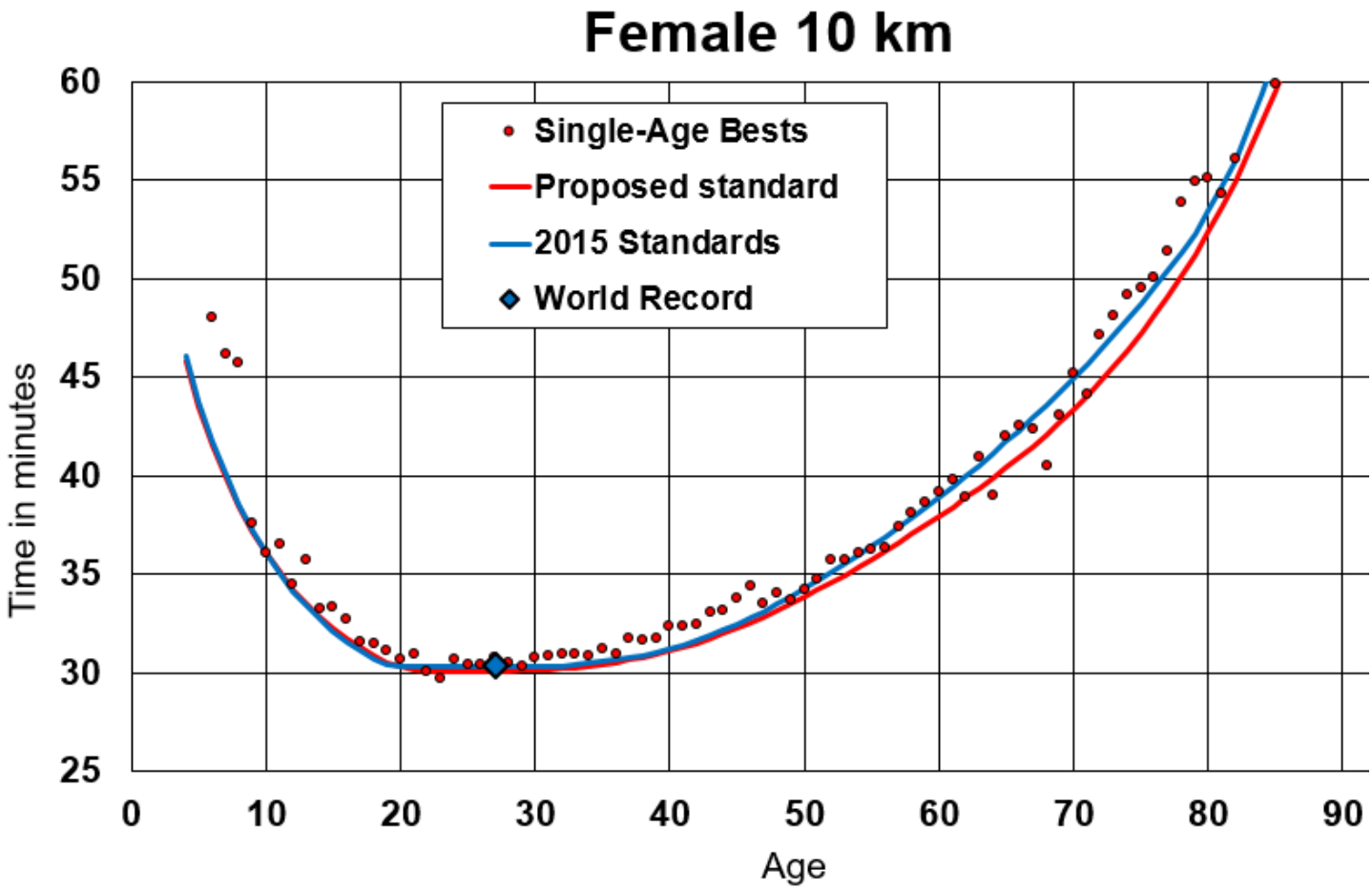
Record	Decimal record	Youth Coefficient	Masters Coefficient	Maximum Youth age	Minimum Masters age
		0.03200	0.110831		
		0.00200	0.000223	21	27.0
00:29:43	29.7167	0.01600	0.00994	17	49.3
	1783	0.00091	0.00038	15	75.0

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.

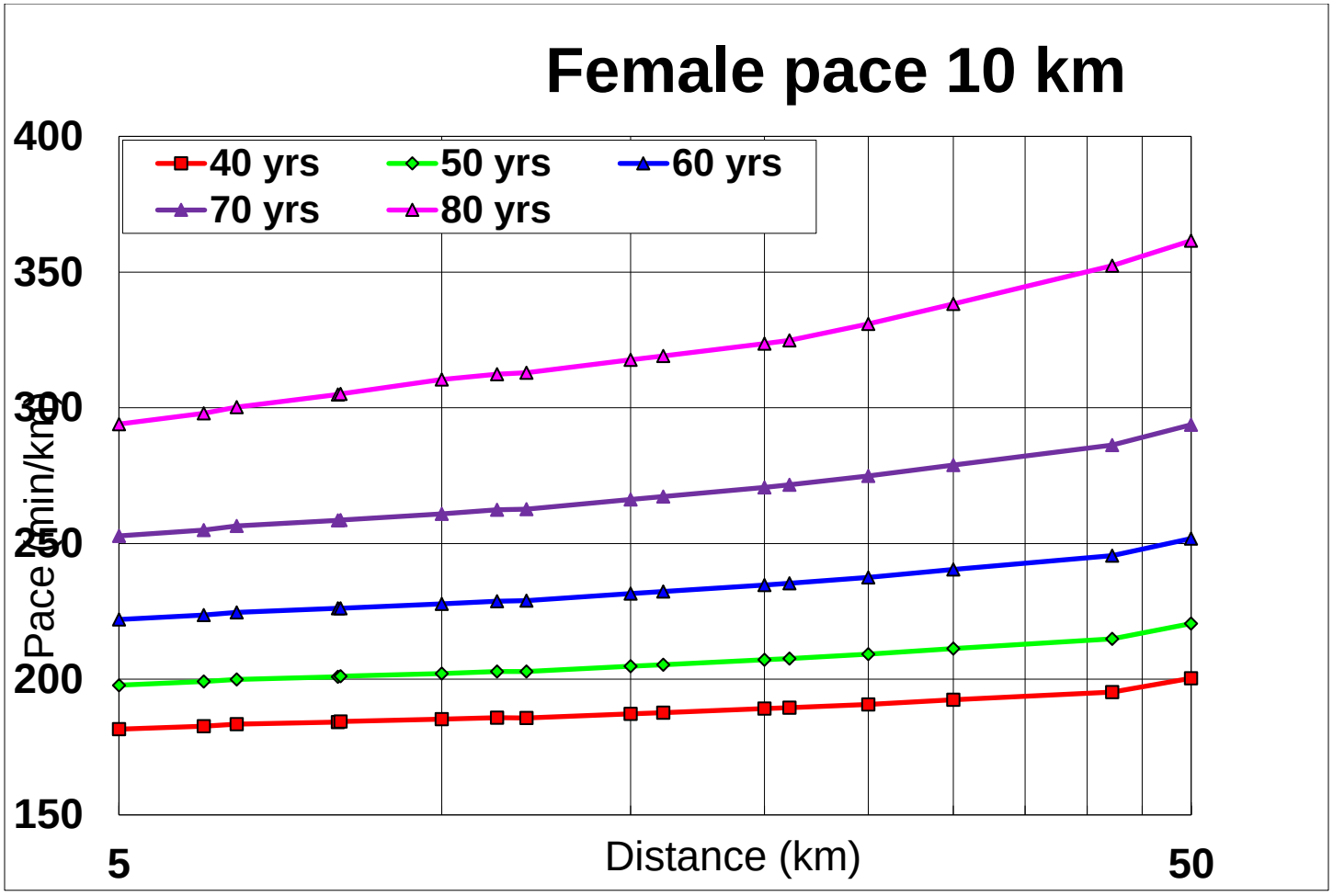


Notice the values for 64 and 68 years old. Here are the data for those points:

```
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:



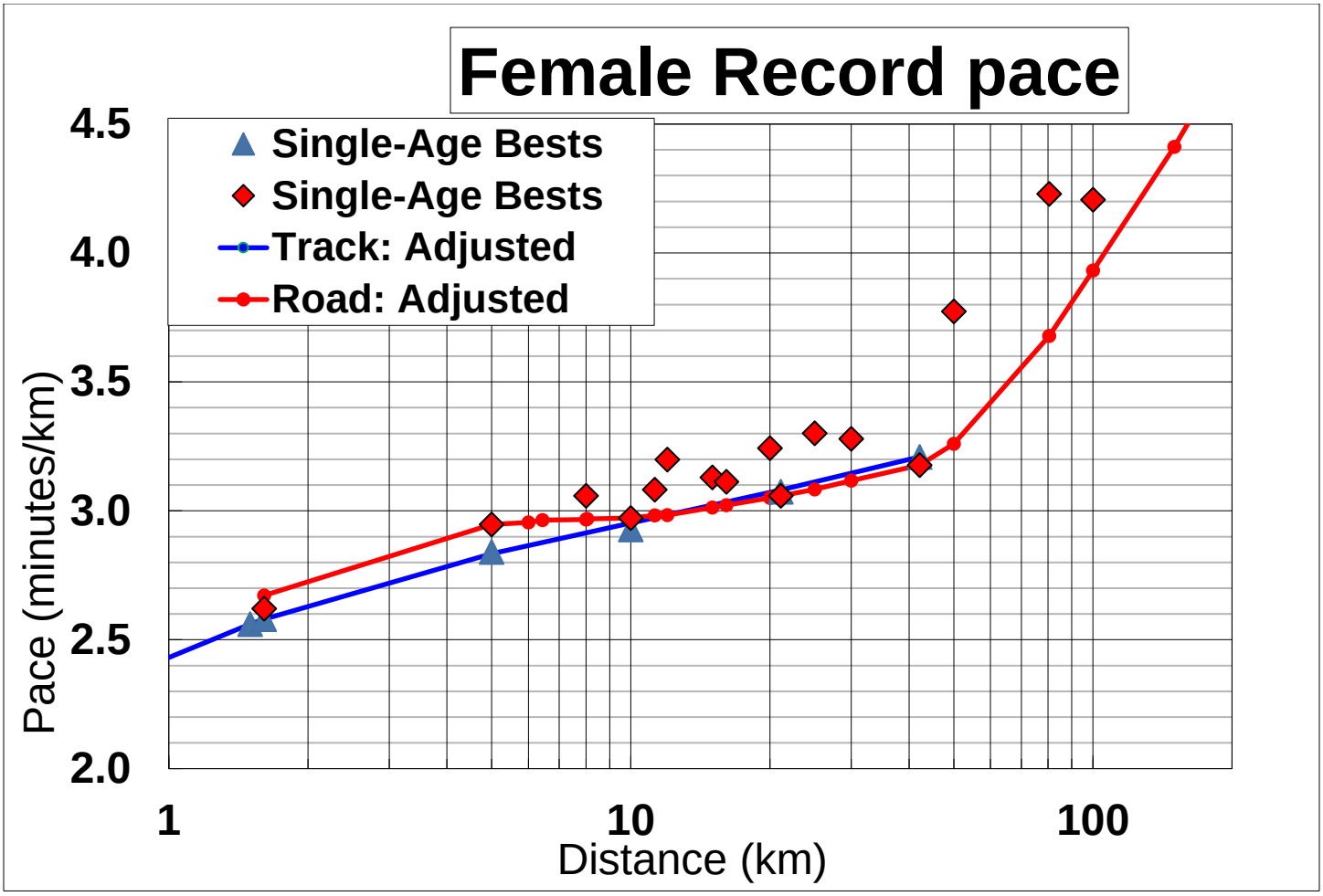
Creating Open Class times for intermediate distances

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.

	A	B	C	D	E	F	G
1	**Distance**	**km**	**Pace per km**	**Record (min)**	**Record**	**Record (s)**	**Adjusted record Road**
12	Mile	1.6093	62012	4.2167	00:04:13	253.00	00:04:18
13	5K	5.0000	2.94667	14.7333	00:14:44	884.00	00:14:44
14	6K	6.0000					00:17:44
15	4MI	6.4374					00:19:05
16	8K	8.0000	3.05833	24.4667	00:24:28	1468.00	00:23:44
17	5 Miles	8.0467					00:23:53
18	10K	10.0000	2.97167	29.7167	00:29:43	1783.00	00:29:43
19	7 Miles	11.2654	3.08171	34.7167	00:34:43	2083.00	00:33:36
20	12K	12.0000	3.19861	38.3833	00:38:23	2303.00	00:35:48
21	15K	15.0000	3.13000	46.9500	00:46:57	2817.00	00:45:12
22	10MI	16.0934	3.11203	50.0833	00:50:05	3005.00	00:48:38
23	20K	20.0000	3.24250	64.8500	01:04:51	3891.00	01:01:00
24	H. Mar	21.0975	3.05802	64.5167	01:04:31	3871.00	01:04:31
25	25K	25.0000	3.30067	82.5167	01:22:31	4951.00	01:17:05
26	30K	30.0000	3.27944	98.3833	01:38:23	5903.00	01:33:30
27	Marathon	42.1950	3.17731	134.0667	02:14:04	8044.00	02:14:04
28	50K	50.0000	3.77300	188.6500	03:08:39	11319.00	02:43:00
29	50MI	80.4672	4.22905	340.3000	05:40:18	20418.00	04:56:00
30	100K	100.0000	4.20617	420.6167	07:00:37	25237.00	06:33:11
31	150K	150.0000	5.50600	825.9000	13:45:54	49554.00	11:01:40
32	100MI	160.9344	5.14299	827.6833	13:47:41	49661.00	12:05:00
33	200K	200.0000	5.94800	1189.6000	19:49:36	71376.00	16:00:00

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.

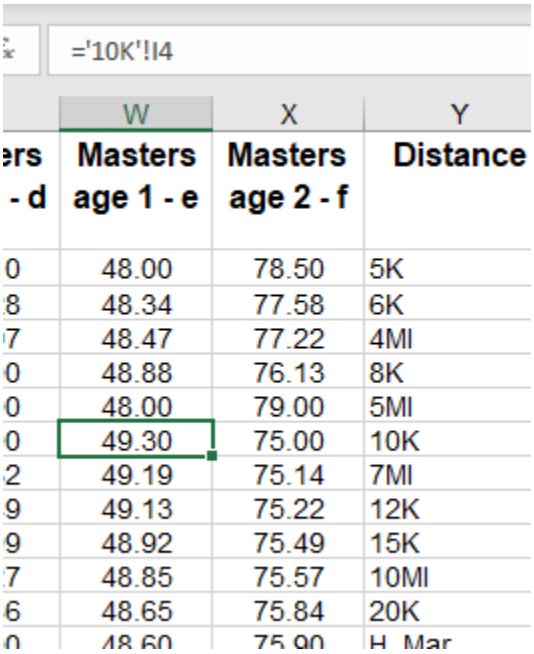
It seems strange that the pace for the female 5 km is very close to the pace for the 10 km.

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.

O	P	Q	R	S	T	U	V	W	X	Y
Youth Coef-1 - B	**Youth Coef-2 - A**	**Masters Coef-1 - E**	**Masters Coef-2 - F**	**Youth age 1 - c**	**Youth age 1 - b**	**Youth age 2 - a**	**Masters age 1 - d**	**Masters age 1 - e**	**Masters age 2 - f**	**Distance**
0.01600	0.00070	0.009710	0.000470	19.00	17.00	15.00	30.10	48.00	78.50	5K
0.01600	0.00085	0.009770	0.000446	19.00	17.00	15.00	29.28	48.34	77.58	6K
0.01600	0.00091	0.009794	0.000437	19.00	17.00	15.00	28.97	48.47	77.22	4MI
0.01600	0.00099	0.009866	0.000409	20.00	17.00	15.00	28.00	48.88	76.13	8K
0.01600	0.00099	0.009868	0.000600	20.00	17.00	15.00	29.00	48.00	79.00	5MI
0.01600	0.00091	0.009940	0.000380	21.00	17.00	15.00	27.00	49.30	75.00	10K
0.01600	0.00091	0.010013	0.000383	21.00	17.00	15.00	27.32	49.19	75.14	7MI
0.01650	0.00091	0.010052	0.000385	20.60	17.00	15.00	27.49	49.13	75.22	12K
0.01700	0.00091	0.010190	0.000390	20.80	17.00	15.00	28.09	48.92	75.49	15K
0.01720	0.00091	0.010233	0.000392	21.10	17.00	15.00	28.27	48.85	75.57	10MI
0.01850	0.00091	0.010367	0.000398	21.90	17.00	15.00	28.86	48.65	75.84	20K
0.01900	0.00091	0.010400	0.000399	22.00	17.00	15.00	29.00	48.60	75.90	H. Mar
0.01900	0.00091	0.010562	0.000399	22.00	17.00	15.00	29.64	48.45	75.43	25K
0.01800	0.00085	0.010735	0.000400	22.00	17.00	15.00	30.32	48.30	74.94	30K
0.01600	0.00091	0.011060	0.000400	22.00	17.00	15.00	31.60	48.00	74.00	Marathon

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.

=$V$13*(1-$Z14)++$V$18*$Z14

Y	Z	AA	AB	AC	AD	AE	AF	AG
Distance	Normalized log distance	Masters age 1 - d	Masters age 1 - e	Masters age 2 - f	Masters Coef-1 - E	Masters Coef-2 - F	Record (s)	Record (min)
5K	0.0000	30.10	48.00	78.50	0.009710	0.000470	884.0	14.733
6K	0.2630	29.28	48.34	77.58	0.009770	0.000446	1120.5	18.674
4MI	0.3645	28.97	48.47	77.22	0.009794	0.000437	1211.7	20.195
8K	0.6781	28.00	48.88	76.13	0.009866	0.000409	1493.6	24.893
5MI	0.6865	27.97	48.89	76.10	0.009868	0.000408	1501.1	25.019
10K	0.0000	27.00	49.30	75.00	0.009940	0.000380	1783.0	29.717
7MI	0.1596	27.32	49.19	75.14	0.010013	0.000383	2116.2	35.271
12K	0.2442	27.49	49.13	75.22	0.010052	0.000385	2292.9	38.215
15K	0.5431	28.09	48.92	75.49	0.010190	0.000390	2917.0	48.617
10MI	0.6374	28.27	48.85	75.57	0.010233	0.000392	3113.8	51.896
20K	0.9284	28.86	48.65	75.84	0.010367	0.000398	3721.6	62.026
H. Mar	0.0000	29.00	48.60	75.90	0.010400	0.000399	3871.0	64.517
25K	0.2449	29.64	48.45	75.43	0.010562	0.000399	4892.8	81.546
30K	0.5079	30.32	48.30	74.94	0.010735	0.000400	5990.4	99.840
Marathon	1.0000	31.60	48.00	74.00	0.011060	0.000400	8044.0	134.067

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.