Creating Road Running Age-Grade Tables

Alan Jones, 2024-10-23

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.

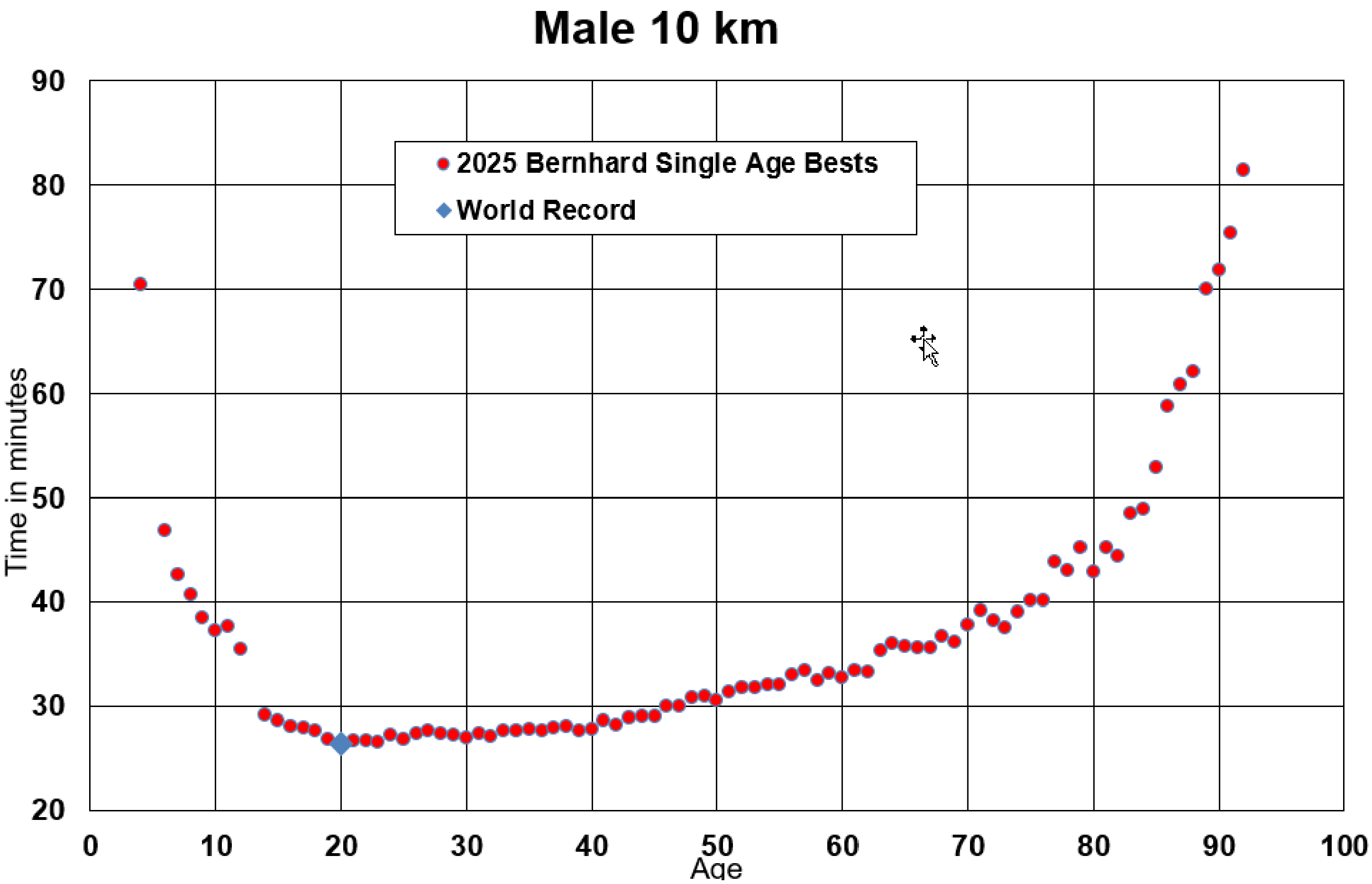
Single-Age Bests

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

10 km	F	21	30:55	1855	Peres	Chepchirchir	KEN	9/27/1993		Prague CZE	9/5/2015	
10 km	F	22	28:46	1726	Agnes	Ngetich	KEN	1/23/2001	Valencia Ibercaja 10K	Valencia, ESP	1/14/2024	WR
10 km	F	23	28:57	1737	Emmaculate	Achholi	KEN	4/2/2000	Valencia Ibercaja 10K	Valencia, ESP	1/14/2024	
10 km	F	24	30:39	1839	Liz	McColgan	SCO	5/24/1964		Orlando FL USA	3/11/1989	
10 km	F	25	30:24	1824	Viola	Jepchumba	KEN	10/23/1990		Prague CZE	9/10/2016	
10 km	F	25	30:01	1801	Agnes	Tirop	KEN	10/23/1995	Adizero Road to Records	Herzogenaurach, Ger	9/12/2021	
10 km	F	26	30:25	1825	Viola	Jepchumba	KEN	10/23/1990		Prague CZE	9/9/2017	
10 km	F	27	30:47	1847	Paula	Radcliffe	ENG	12/17/1973		New York NY USA	6/9/2001	
10 km	F	28	30:30	1830	Tirunesh	Dibaba	ETH	6/1/1985		Tilburg NED	9/1/2013	
10 km	F	29	30:21	1821	Paula	Radcliffe	ENG	12/17/1973		San Juan PUR	2/23/2003	
10 km	F	30	29:38	1778	Kalkidan	Gezahegne	BRN	5/8/1991	The Giants Geneva 10K	Geneva, SWI	10/3/2021	
10 km	F	31	30:50	1850	Lornah	Kiplagat	NED	5/1/1974		San Juan PUR	2/26/2006	
10 km	F	32	30:56	1856	Gladys	Cherono	KEN	5/12/1983		Ottawa ON CAN	5/23/2015	
10 km	F	33	30:59	1859	Ingrid	Kristiansen	NOR	3/21/1956		Boston MA USA	4/9/1989	
10 km	F	34	30:52	1852	Shalane	Flanagan	USA	7/8/1981		Boston MA USA	6/26/2016	
10 km	F	35	31:13	1873	Elana	Meyer	RSA	10/10/1966		Budapest HUN	10/14/2001	
10 km	F	36	30:57	1857	Irina	Mikitenko	GER	8/23/1972		Karlsruhe GER	9/13/2008	
10 km	F	37	31:45	1905	Helen	Clitheroe	ENG	1/2/1974		Manchester ENG	5/15/2011	
10 km	F	38	31:37	1897	Edith	Masai	KEN	4/4/1967		Tilburg NED	9/4/2005	
10 km	F	39	31:44	1904	Irina	Mikitenko	GER	8/23/1972		Paderborn GER	4/7/2012	
10 km	F	40	32:23	1943	Christelle	Daunay	FRA	12/5/1974		Arras FRA	8/30/2015	
10 km	F	41	32:23	1943	Christelle	Daunay	FRA	12/5/1974		Manchester ENG	5/22/2016	
10 km	F	42	32:28	1948	Veerle	Dejaeghere	BEL	8/1/1973		Tilburg NED	9/6/2015	
10 km	F	42	32:09	1929	Edna	Kiplagat	KEN	11/15/1979	BAA 10K	Boston, MA	6/26/2022	
10 km	F	43	32:41	1961	Tatyana	Pozdniakova	UKR	3/4/1955	Crescent City Classic	New Orleans, LA	4/11/1998	
10 km	F	44	33:08	1988	Priscilla	Welch	USA	11/22/1944		Baltimore MD USA	10/22/1989	
10 km	F	45	33:29	2009	Stephanie	Herbst-Lucke	USA	12/27/1965	Peachtree Road Race	Atlanta, GA	7/4/2011	
10 km	F	46	34:01	2041	Firaya	Sultanova-Zhadanova	USA	4/29/1961	Peachtree Road Race	Atlanta, GA	7/4/2007	
10 km	F	47	33:28	2008	Tatyana	Pozdniakova	UKR	3/4/1955	Azalea Trail Run	Mobile AL USA	3/23/2002	

Tom always marks the World Record for the distance with yellow.

Graph of Single-Age Bests



The task is to develop age-grade standards that will be less than or the same as any of the dots in the graph. In looking at these data, I thought a multi-section curve would be the job. The line would start out as a decreasing curved parabola followed by a decreasing linear section followed by a flat section which included the world record, followed by a upward trending linear section, and followed by an upward increasing parabola.

See below.

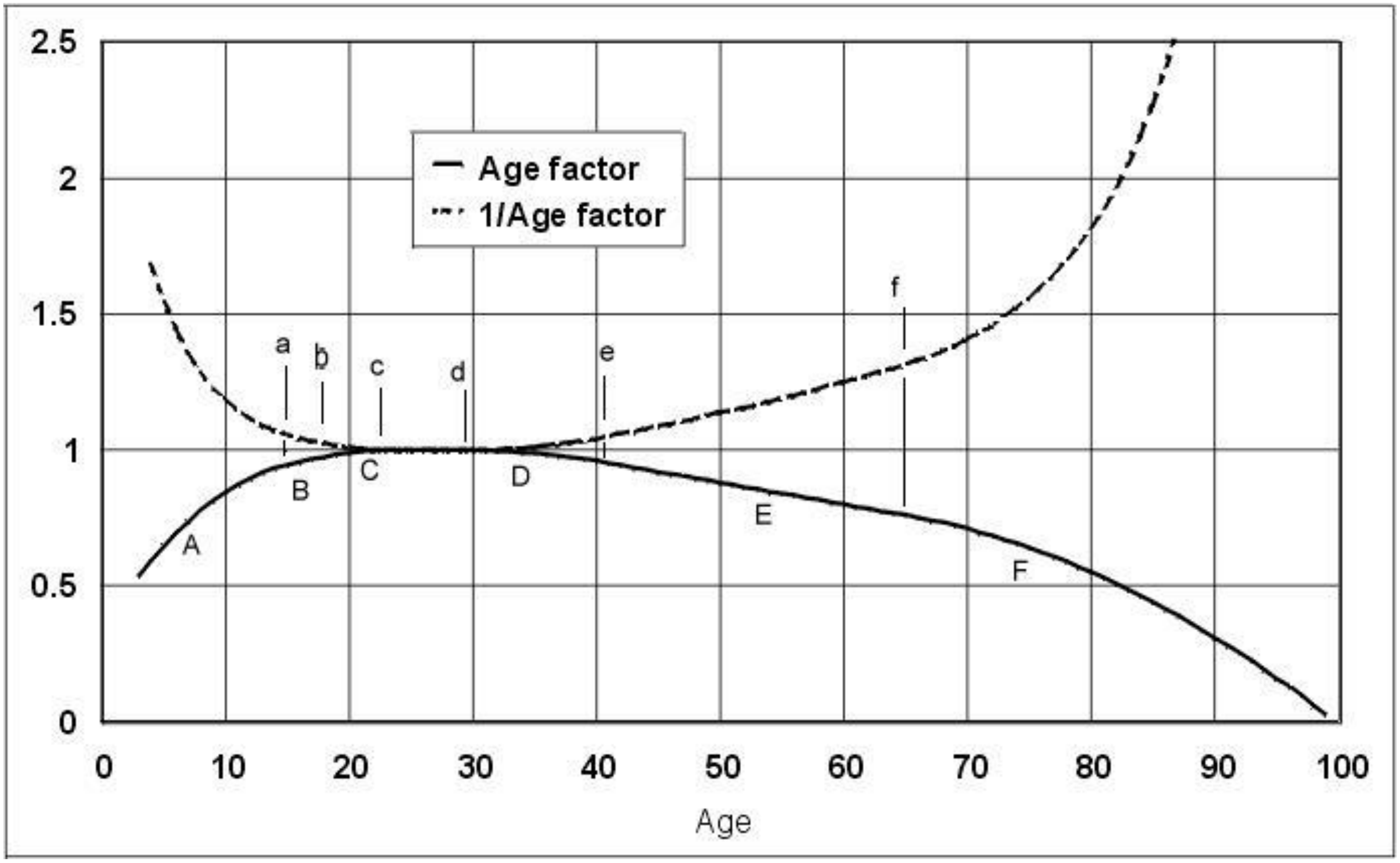


Figure 1: Proposed age-grade curve

The Excel Spreadsheets

The male sheet is called **`maleRoad2025.xlsx`** and the female called **`femaleRoad2025.xlsx.`**

Each file is composed of many sheets:

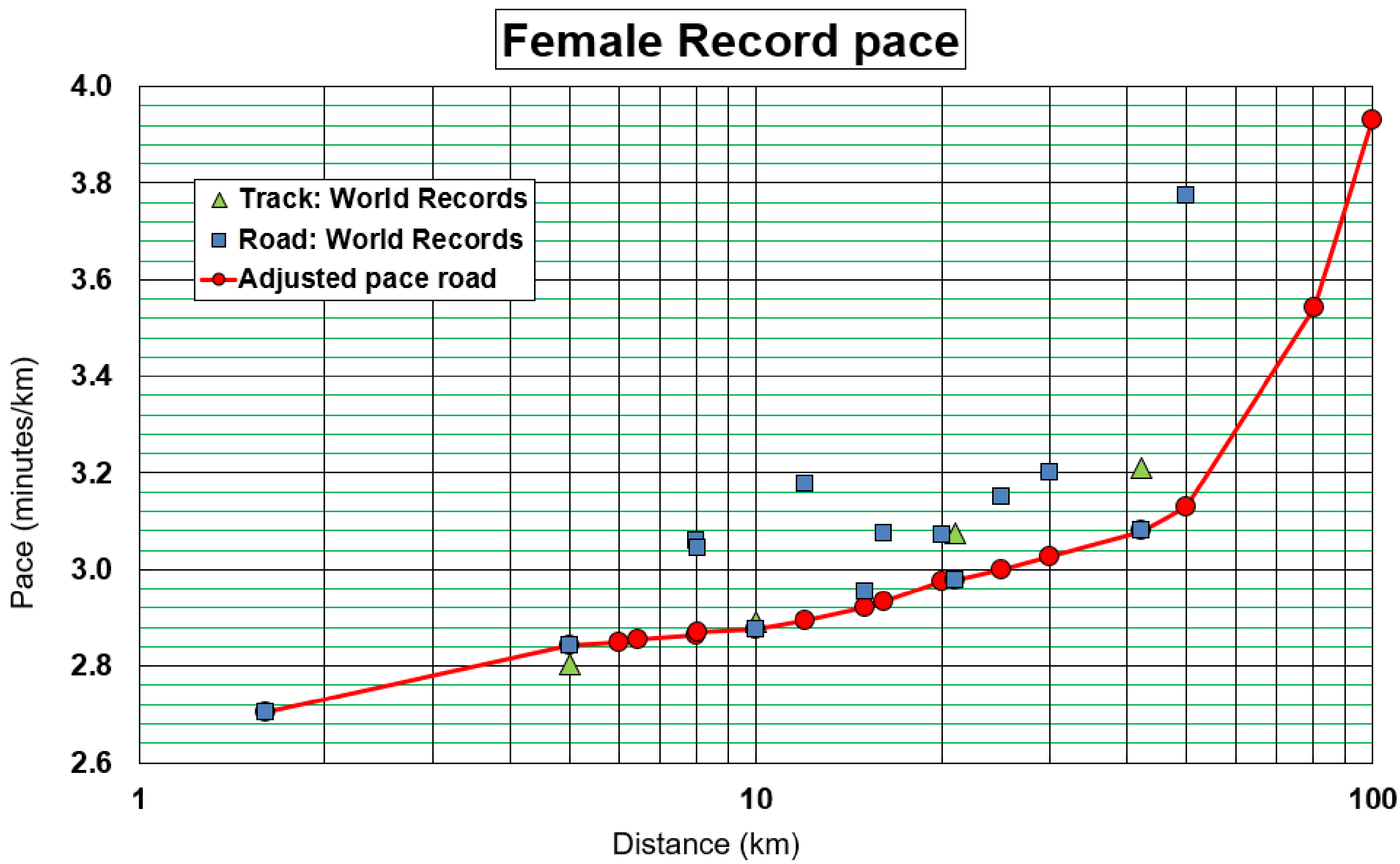
- Parameters: Open times for each distance plus a lot of other data. Includes interpolation controls.
- One each for each of the distances: mile, km, 6 km, 4 mi, 8 km, 5 mi, 10 km, 12 km, 15 km, 10 mi, 20 km, Half-Marathon, 25 km, 30 km, Marathon, 50 km, 100 km, 200 km.
- Age-factors. There is an age-factor 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.

Determine single age Open Class Times

The **Open Class Time** for each age is an estimate of the fastest time possible for that age. For the 5 km, 10km, Half-Marathon, and Marathon, it is the world record. We determine the other distances by plotting pace vs distance. See below. The first step is to enter the world record for each distance. That is done on the **Parameters** sheet. There is a table for track distances and another for road distances. I'll only show the road times here.

	A	B	C	D	E	F	G
1	Distance	km	Pace per km	Record (min)	Record	Record (s)	Adjusted record Road

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 blue squares are the road records. The blue triangles are track records for those same distances. 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.

Import single-age bests

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

Age	Performance 2025 data vs 2020 standards	Performance 2025 data vs 2025 standards	2025 Bernard Single Age Bests	FNAME	LNAME	Country
20	93.9	94.5	30:41	Gladys	Cheshire Kipt	KEN
21	93.0	93.2	30:55	Peres	Chepchirchir	KEN
22	95.6	100.0	28:46	Agnes	Ngetich	KEN
23	96.8	99.4	28:57	Emmaculate	Achholi	KEN
24	93.9	93.9	30:39	Liz	McColgan	SCO
25	94.6	95.8	30:01	Agnes	Tirop	KEN
26	94.6	94.6	30:25	Viola	Jepchumba	KEN
27	93.4	93.6	30:47	Paula	Radcliffe	ENG
28	94.3	94.6	30:30	Tirunesh	Dibaba	ETH
29	94.9	95.2	30:21	Paula	Radcliffe	ENG
30	93.7	97.6	29:38	Kalkidan	Gezahegne	BRN
31	93.6	94.0	30:50	Lornah	Kiplagat	NED
32	93.5	94.0	30:56	Gladys	Cherono	KEN
33	93.6	94.1	30:59	Ingrid	Kristiansen	NOR
34	94.2	94.7	30:52	Shalane	Flanagan	USA
35	93.5	94.0	31:13	Elana	Meyer	RSA
36	94.7	95.1	30:57	Irina	Mikitenko	GER
37	92.7	93.1	31:45	Helen	Clitheroe	ENG
38	93.5	93.9	31:37	Edith	Masai	KEN
39	93.7	94.0	31:44	Irina	Mikitenko	GER
40	92.3	92.6	32:23	Christelle	Daunay	FRA
41	92.9	93.1	32:23	Christelle	Daunay	FRA
42	93.3	94.4	32:09	Edna	Kiplagat	KEN

The times in the above, Single-Age Bests are copied into a text editor – not a word processor. The editor, TextPad, can do "block select". We had to add two zeros and a colin to the time data in order to have a consistent time format for all the distances. Part of the above becomes:

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

This is then copied into column B, "2025 Bernhard Single Age Bests", 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	00:32:09	32.150

Creating the age standards

Figure 1 shows how we fit the age-grade performance factor to a series of lines: a parabola (A), a linear line (B), a constant section (C), another linear line (E) and, finally, another parabola (F) as shown in this figure. The points on the x-axis where the curve changes are a, b, c, d, e, and f. The coefficient of each section is shown as capital letters: A, B, C, D, E, and F.

These mathematical expressions govern the different segments where y represents the age factor.

$y = 1 - C(c-b)^2 - B(b-x) - A(a-x)^2$	for x < a
$y = 1 - C(c-b)^2 - B(b-x)$	for x >= a and x < b
$y = 1 - C(c-x)^2$	for x >= b and x < c
$y = 1$	for x >= c and x < d
$y = 1 - D(x-d)^2$	for x >= d and x < e
$y = 1 - D(e-d)^2 - E(x-e)$	for x >= e and x < f
$y = 1 - D(e-d)^2 - E(x-e) - F(x-f)^2$	for x >= f

The age factor is always equal or less than 1. The upper curve is 1 divided by *y*. This is the same shape as that for the age standards for each distance since they are computed by dividing the world record for that distance by the age factor.

There are 12 parameters: *a, b, c, d, e, f, A, B, C, D, E, and F*. We want to require that the slope be continuous from the quadratic sections to the linear sections. To do this, we compute the derivative of the above expression and then set them equal at points *a, b, e* and *f*.

A: $\frac{dy}{dx} = B + 2A(a-x)$

B: $\frac{dy}{dx} = B$

C: $\frac{dy}{dx} = 2C(c-x)$

D: $\frac{dy}{dx} = -2D(x-d)$

E: $\frac{dy}{dx} = -E$

F: $\frac{dy}{dx} = -E - 2F(x-f)$

Set the slopes equal at:

x = a: $B = B$

x = b: $B = 2C(c-b)$

x = e: $-2D(e-d) = -E$

x = f: $-E = -E$

From the x = b condition, we obtain: $C = \frac{B}{2(c-b)}$

From the x = e condition, we obtain: $D = \frac{E}{2(e-d)}$

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.

The ten parameters can be adjusted to fit the single-age records. The factor (*af*) 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 to obtain a time that this person should be able to run as an open athlete.

Below is an image from a portion of the spreadsheet. There is an image for both male and female and for each distance. Those parameters are represented by cells in the spread sheet.

	A	B	C	D	E	F	G	H	I
1	**Female 10 km**			Record	Decimal record	Youth Coefficient	Masters Coefficient	Maximum Youth age	Minimum Masters age
2						0.04725	0.172200		
3						0.00189	0.000160	22	24.0
4				00:28:46	28.7767	0.01890	0.01050	17	56.8
5					1726	0.00091	0.00051	15	76.7

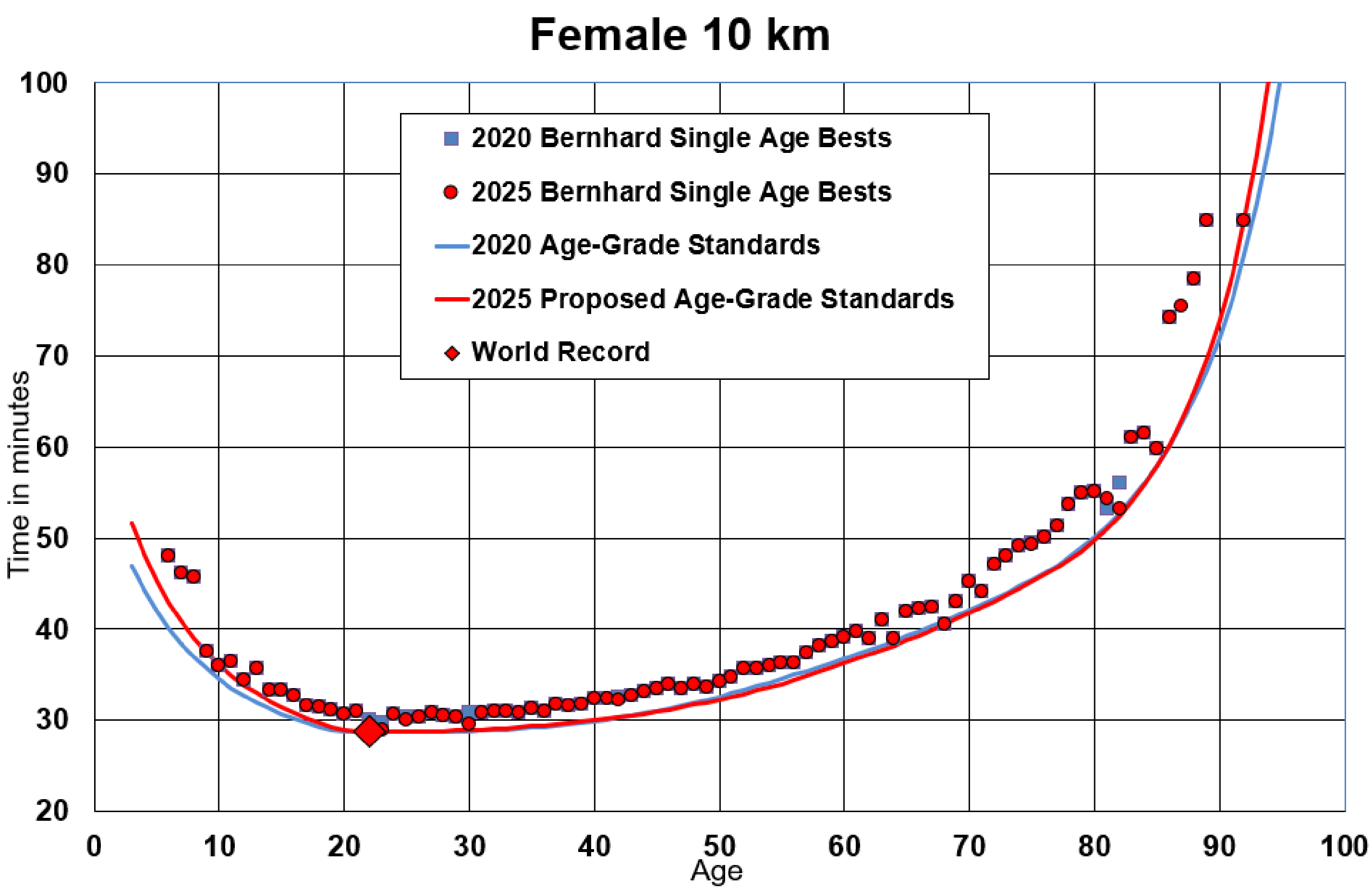
	F	G	H	I
2	$\frac{B(c-b)}{2}$	$\frac{E(f-e)}{2}$		
3	$\frac{B}{2(b-a)}$	$\frac{E}{2(e-d)}$	c	d
4	B	E	b	e
5	A	F	a	f

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

The parameters for youths, 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.

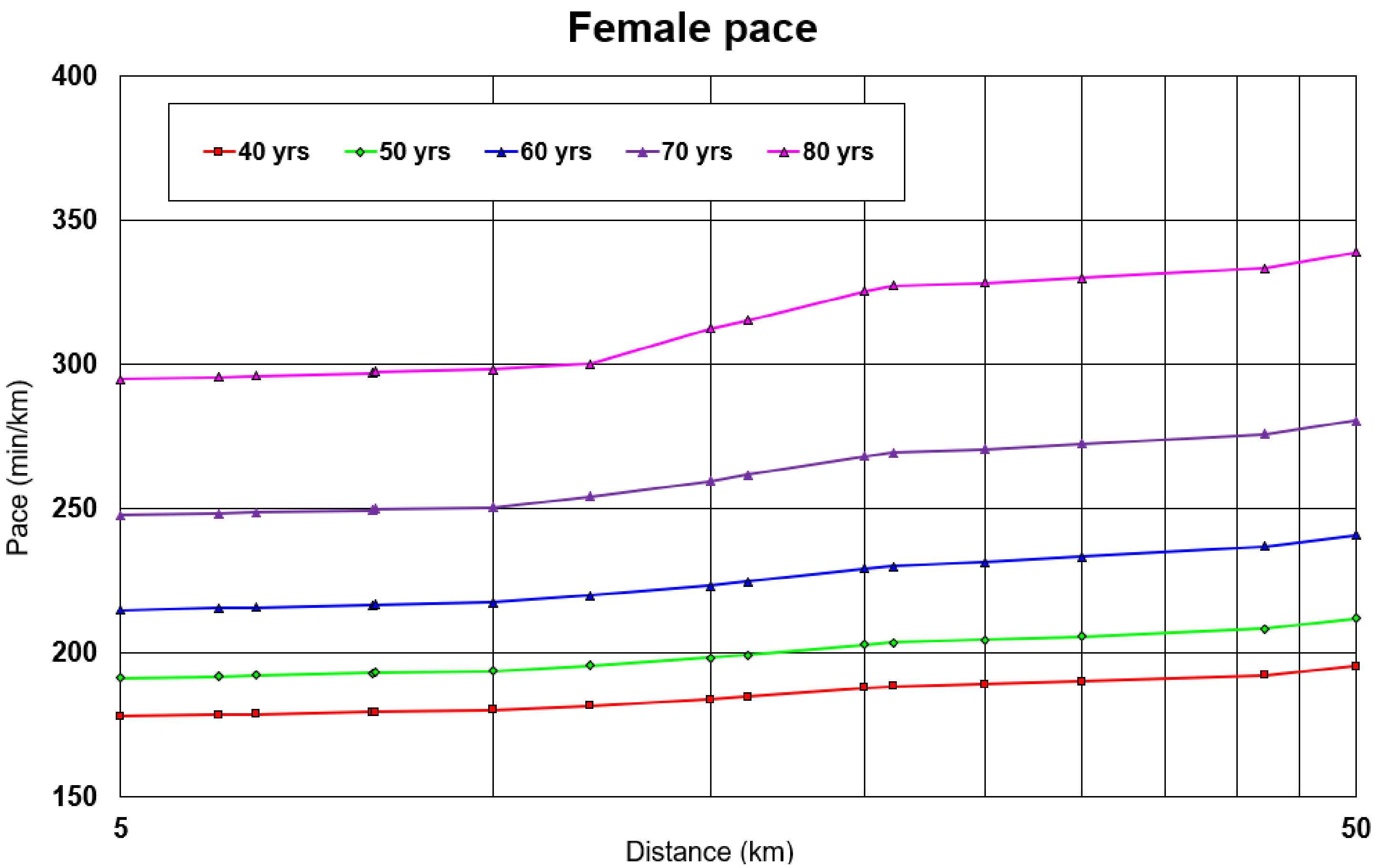
Note that four of the cells are represented by a calculation. These equations represent intermediate computations of constants that are used in the computation of the age factor. I have locked the cells so they cannot be modified and have colored them gray to indicate the lock.

Here is the plot for the female road 10 km.



We have included Tom Bernhard's single-age bests from 2020. As you can only a few times have changed and the curve is almost identical. An exception is the youth values where we improved on the fit.

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. 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:



The pace seems reasonable except for age-80. Also, the interpolation does not do as well for the youth ages. In the 2020 version, we used the same age-factor for all distances for the youth ages. It did well.

Interpolation

After we have good standards for the distances 5 km, 10 km, Half Marathon, and Marathon, we use interpolation for the other distances. In my first attempt, I interpolated the parameters, A, B, etc. which are then used to compute age-factors for the other races. This did not produce the desired result. It seems the different parameters affect the age-factors in different ways.

We then decided to interpolate from the actual age factors. That is, for example, we would compute the age factor for 5 km and the 10 km for each age. We then used those "end points: to determine that age-factor for that age for distances.

To do the interpolation, we start with two endpoints: (x_a, y_a) and (x_b, y_b). To compute the y-value for any x-value, we use the equation:

(2) $$y = y_a(1-u) + y_b u$$

The variable *u* varies between 0 and 1. I used u to be the $\log_{10}$ of the distance. Otherwise the distances between each distance would vary too much.

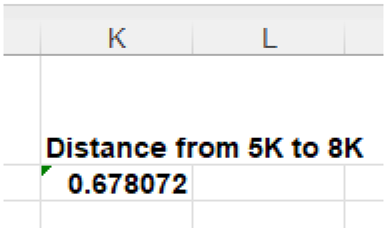
The determination of the parameter ***u*** for each distance, is computed on the Parameters sheet. It is the log to the base 10 of the distance *d* from last key distance $\mathbf{d_n}$ (5K, 10K, Half Marathon, or Marathon)

(3) $$u = ((\log_{10}(d) - \log_{10}(d_n))/(\log_{10}(d_{n+1}) - \log_{10}(d_n))$$

		u
5K	5.0000	0.000000
6K	6.0000	0.263034
4MI	6.4374	0.364545
8K	8.0000	0.678072
5 Miles	8.0467	0.686473
10K	10.0000	0.000000
12K	12.0000	0.244212
15K	15.0000	0.543104
10MI	16.0934	0.637351
20K	20.0000	0.928443
H. Mar	21.0975	0.000000
25K	25.0000	0.244856
30K	30.0000	0.507890
Marathon	42.1950	1.000000

The parameter *u* goes from 0 to 1 starting at 10K. This *u* does the interpolation for distances 6 km, 4 Miles, 8 km, and 5 Miles. At 10K, we start a new value of *u* increasing from 0 starting at 10 km for the distances 12 km, 15 km, 10 Miles, and 20 km. The last *u* starts at the Half-Marathon distance reaching the value of 1 at Marathon.

Here is an example of how the interpolation is done. Say we want to determine the interpolated values for the age factor for the 8 km distance. We take *u* for that distance, 0.678072, and insert it in cell K2 in the 8K sheet.



In the Age Factor column, for each age, we insert the Excel version of Equation (1):

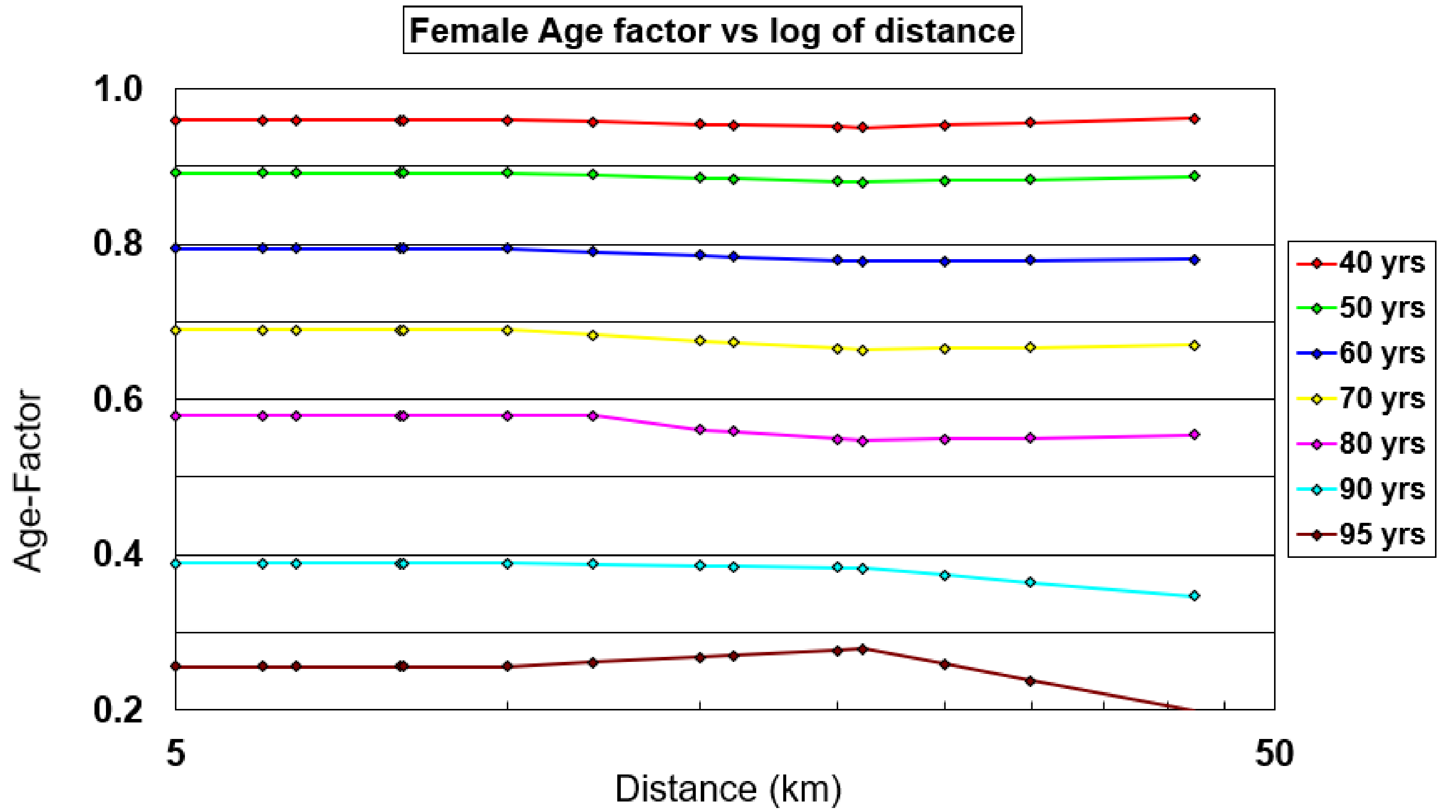
```
='5K'!$E41*(1-$K$2)+'10K'!$E41*$K$2
```

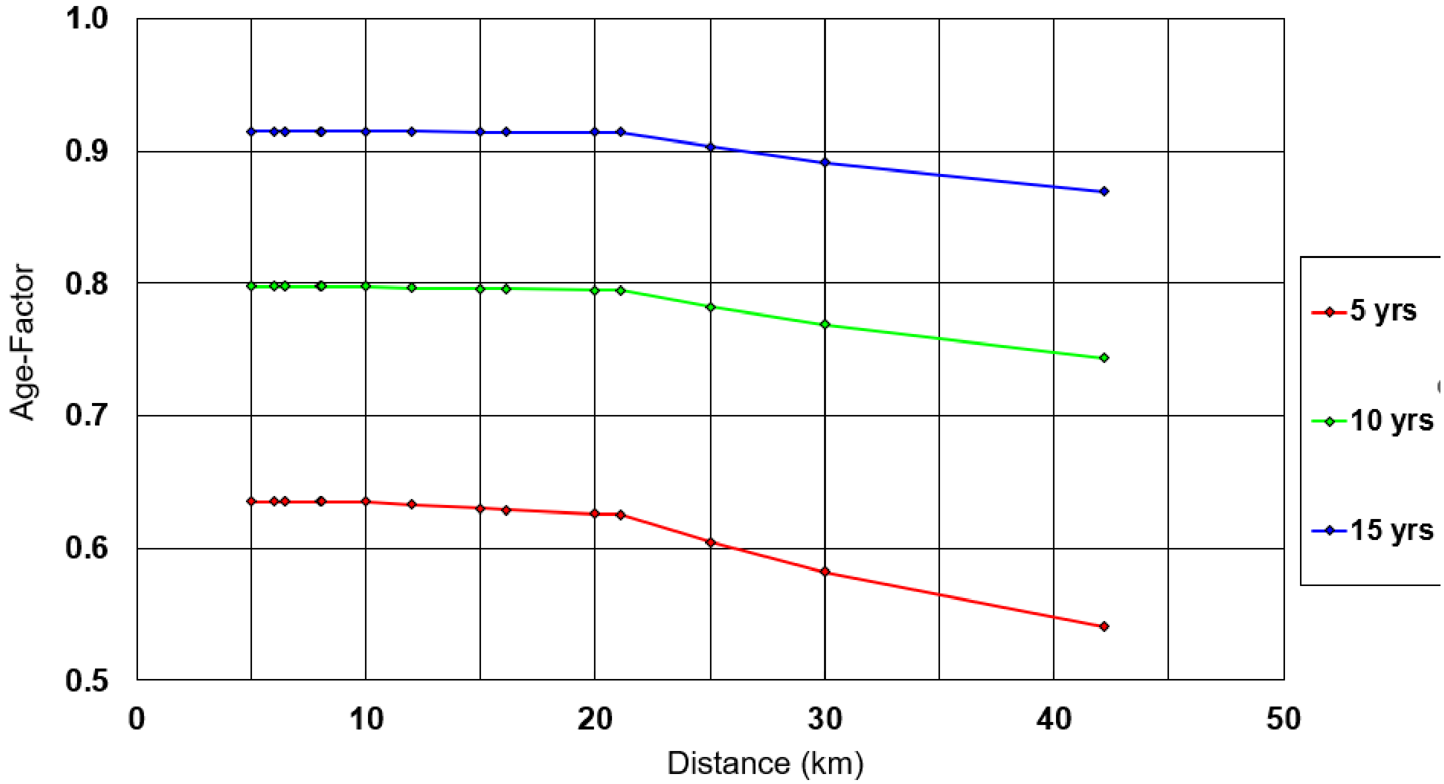
Where "'5K'!$E41" is the value of the age factor for 35 years at 5 km, K2 is u, and "'10K'!E41" is the age factor for 35 years for 10 km.

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 chart below shows the effectiveness of the interpolation. Before we did it, there were irregularities at a number of places, in particular between 20 km and Half-Marathon. The first chart is for master ages; the second one for youths.


Female Age factor vs log of distance
1.0
0.8
0.6
0.4
0.2
Age-Factor
5
50
Distance (km)
40 yrs
50 yrs
60 yrs
70 yrs
80 yrs
90 yrs
95 yrs


1.0
0.9
0.8
0.7
0.6
0.5
Age-Factor
0
10
20
30
40
50
Distance (km)
5 yrs
10 yrs
15 yrs