



A Better Age-Grading Framework for Marathon Runners



[Alexis Spinetta](#)

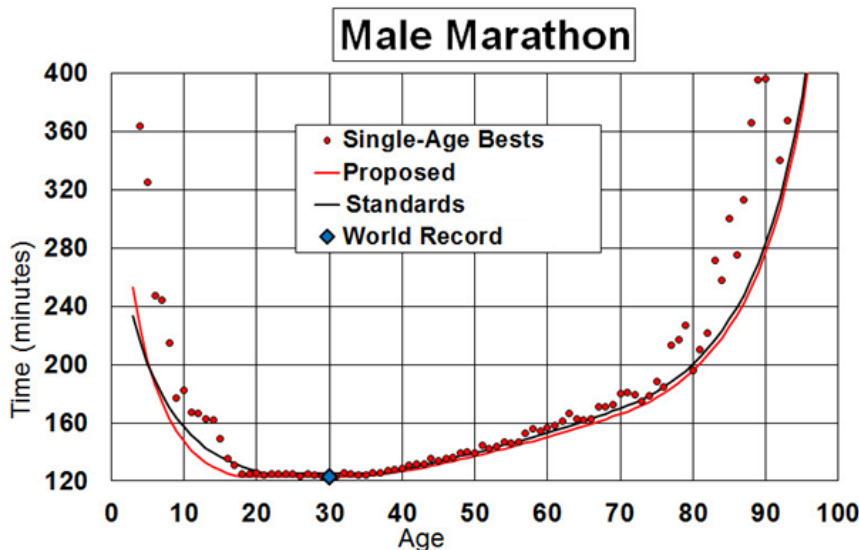
Runner. Scholar. Innovative Thinker.

Many sports, including golf and bowling, calculate "handicaps" to allow athletes of varying abilities to compete. Similarly, runners rely on a framework called "age-grading."

History and Shortcomings of the WMA Framework

The World Association of Veteran Athletes, an organization later renamed [World Masters Athletics \(WMA\)](#), developed what would become the standard for age-grading in 1989. To calculate age grades, runners start by referencing tables of "ideal" times, which approximate the "best possible times" athletes should be able to run after considering age and gender. Next they divide these "ideal" times by actual finishing times. For example, a 47-year-old male with an "ideal" marathon time of 2:13:43 who completes a race in 3 hours would earn an age-grade of $(2:13:32 / 3:00:00) = 74\%$.

While the WMA should be applauded for pioneering age-grading, its [approach](#) suffers significant shortcomings. First, because the WMA subjectively determines "ideal times," so too are its age grades. Every few years, the organization graphs world records and "single-age bests" on scatterplots and then manually draws curves to define top-performance frontiers. Below is one such example:



Second, because "ideal times" are based on world records and "single-age bests," [WMA tables](#) become "stale" and thus increasingly inaccurate with time. Worse, the WMA has a practice of selectively ignores times that it deems outliers, which is odd because world records and single-age "bests" are, by definition, the most extreme outliers on the fastest end of the performance spectrum. Indeed, the seemingly arbitrary justification the WMA offers for excluding times from its model led John Davis, the author of ["How to Use Age Grade Calculators"](#) and critic of WMA age-grading, to complain, "This is crazy!" Lastly and most importantly, if runners are interested in being graded against peers, it makes no conceptual sense to compare performances against those of the world's most elite runners.

Critics have long questioned the accuracy and validity of the WMA's "ideal times." In fact, that is why Alan Jones, an engineer, took over

responsibility for creating WMA tables in 2005. Shortly after the WMA released its 2002 update, Jones became convinced younger runners were not being treated fairly. Jones contacted Rex Harvey and Chuck Phillips, two researchers who were instrumental in creating the tables, and asked them to re-check their work. Jones recalls, "Rex took a close look at Chuck's tables and found some errors due to some fast performances that Chuck did not know about." Jones proposed that he re-draw the curves, tweaking the model to include previously excluded times. The WMA accepted Jones's offer, endorsing him as the organization's age-grading standard bearer.

Jones was eventually forced to revise his 2005 tables after 65 runners achieved age-graded scores over 100%. Stale "ideal" times led a handful of statisticians to create their own performance charts, each promising more accuracy based on proprietary methods of computing "ideal" times. For example, [Howard Grubb](#) offers one popular age-grading calculator. [Ray Fair](#) offers another. Because each derivative of the WMA framework relies on different "ideal" times, they produce slightly different age-grades for the same finishing time.

Jones recognizes the WMA framework has significant deficiencies. He says "stale" tables are, to a certain extent, unavoidable because frequent revisions after every new world record or "single-age best" would trash historical consistency. Furthermore, he says he needs to be subjective in how he draws top-performance curves so runners are guaranteed smooth progression as they age. For example, Figure 1 shows that the fastest 58-year old marathon time is slower than the time posted by the top 59-year old. That means the same time delivered by an older runner would earn a lower age grade, which does not make sense.

Davis observes, "Any method based on single-age all-time best performance will be significantly impacted by outlier performers." He continues, "Often, great Masters athletes will go on a multi-year tear, setting single-age records for every age they pass through does not make sense."

In fact, that was exactly what happened in 2015, which caused Jones to discard the times of one senior athlete. Referring to Olga Kotelko, Jones says, "The times of one [Canadian] female athlete are so good that they had to be ignored. By ignoring her performances, it can be shown that females slow, with age, at a rate slightly higher than males. However, if her times are included, the reverse would be the case." In addition to discarding Kotelko's times, Jones removed other times from his model. Specifically, he discarded times from young African runners. He also decries the performances of young Chinese runners, both male and female, however the narrative that accompanies the release of his latest tables does not make clear their fate (i.e., whether he also chose to discard their times.)

Frankly, it is not clear what criteria other than intuition Jones employs to ignore times he considers outliers. Mostly, it seems he discards times when they skew his smoothly drawn curves. Indeed, when considering whether to discard Olga Kotelko's times, Davis says, "[She] would have skewed the entire prediction curve for women's performances!" Referring to another runner's record, Jones says, "I ignored it because it was so out of line with the clustering of other performances." Furthermore, he says he ignored the time "because it is necessary to have a smooth transition as one goes from one distance to the next." Jones's subjectivity in determining what times to include and exclude from his model led Davis to make his aforementioned complaint, "This is crazy!"

Interestingly, when discussing the release of his latest tables, Jones admits the fallibility of his intuition when it comes to discarding times. "In 2004, I ignored some performances by Tatyana Pozdnyakova because they were apparently outliers but I have to admit with the addition of new records in nearby ages, her times do not seem as far out of line as they did previously."

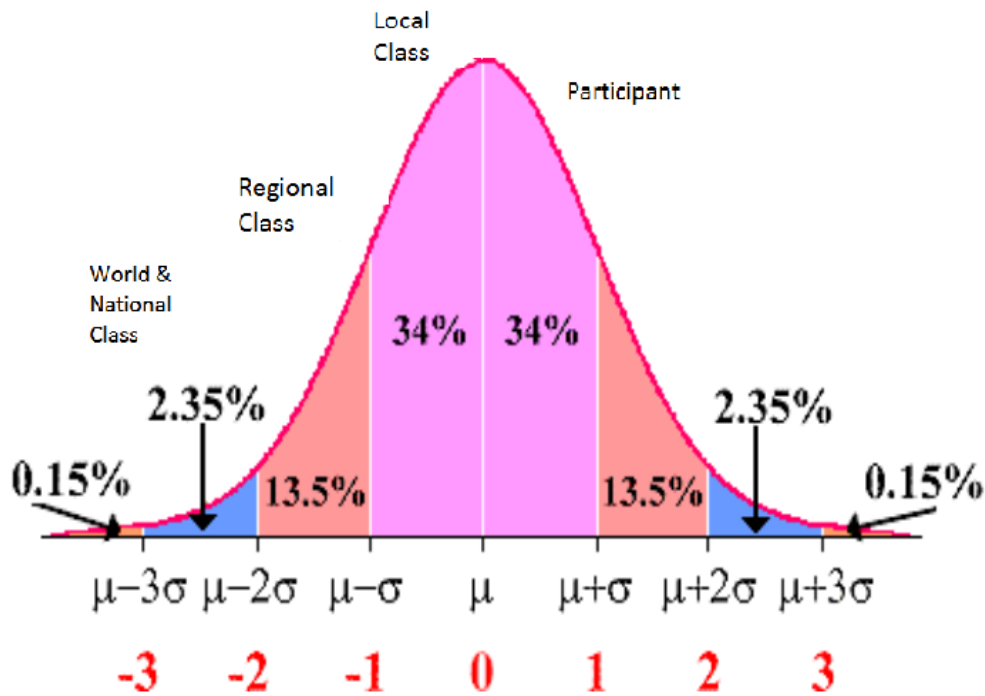
Dennis Kimetto's October 2014 world-record marathon performance pushed Jones to revise and release new charts in 2015. Kimetto's time caused a significant shift in Jones's curve. Nevertheless, runners continued to question the accuracy of WMA "ideal" times. "So even after plugging in the new world record, age-graded percentages for folks at the older and younger ends of the spectrum weren't reflecting recent outstanding performances in their respective age groups," complained Runner's World.

On September 2018, Eliud Kipchoge beat Kimetto's world record time by 18 seconds, which means another potential substantial shift in age grades. No word yet if WMA plans to yet again release new tables.

A Better Approach

"[WMA age-grades] are commonly misunderstood as being percentiles--e.g. an age-grade of 68% meaning you are faster than 68% of all runners your age," observes Davis. "What the percentage actually does is measure the fraction of your race time that is equivalent to the predicted all-time best for your age and sex." In other words, WMA age-grading only provides runners with a sense of performance relative to the most elite (i.e., all-time best) runners of their age and gender categories. This limitation led Davis to implore, "It would be nice if there were another method of rating performances . . . your position relative to all runners your age, not just the best . . . Maybe that can be the next big project for running statisticians!"

My new framework answers Davis plea. It computes [Z-scores](#), which measures the number of standard deviations a finishing time is from the mean. A z-score tells you where the score lies on a normal distribution curve (i.e., "[Bell Curve](#)") to assess runner performance.



To earn World and National Class recognition, athletes would have to earn a Z-score two standard deviations faster than the mean. Based on the characteristics of a normal distribution, that means they are faster than 97.55% of all other runners. Similarly, to qualify for Regional Class, runners would have to be one standard deviation faster than the mean, which means they are faster than 84% of all other runners. This scale replaces the following arbitrarily defined WMA categories that seemingly were modeled after a typical elementary school grading system:

100% = Approximate World-Record Level

90+% = World Class

80+% = National Class

70+% = Regional Class

60+% = Local Class

<60% = Participant

Instead of assessing finishing times against world record and single-age "bests", this new framework assesses an individual's performance against all runners of the same gender and/or age. Accordingly, it essentially eliminates the need for frequent revisions because the mean and standard deviation of marathon runners slowly (perhaps generationally) changes. Moreover, unlike the WMA's framework where new world records and single-age "bests" cause its "ideal" times to become stale, outliers marginally effect mean and standardization benchmarks because they are weighed in the context of performances delivered by masses of non-elite runners.

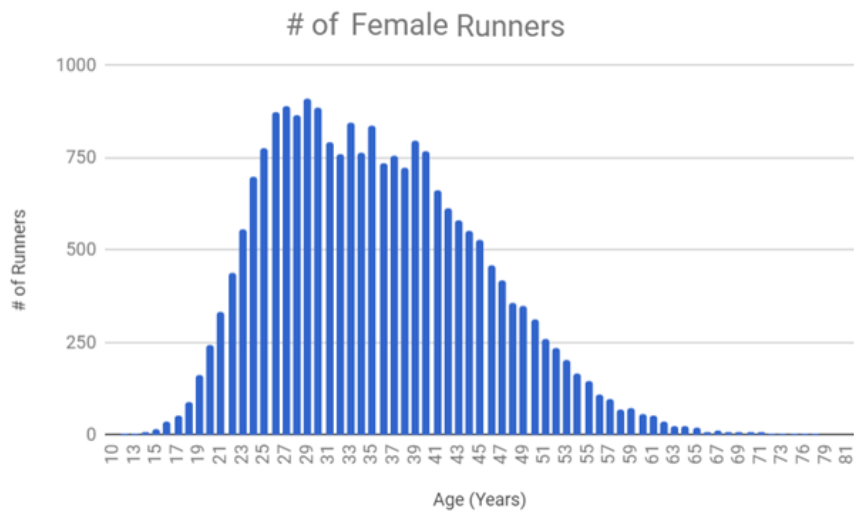
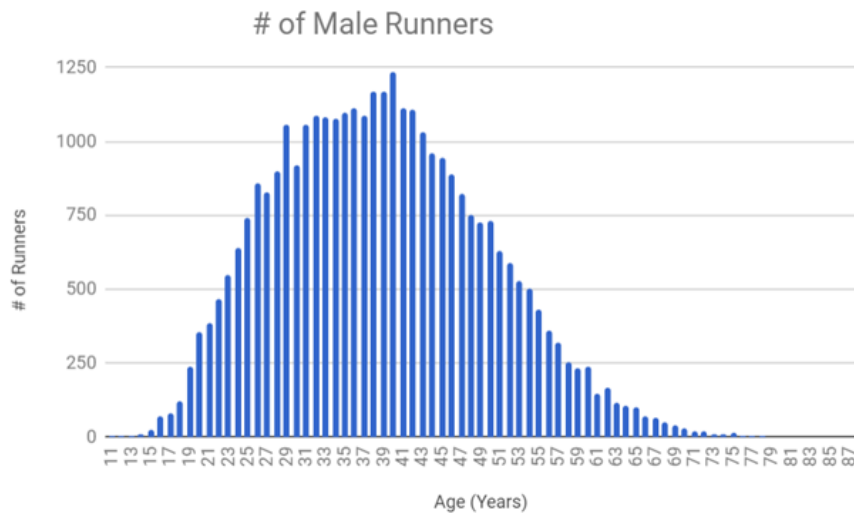
Calculating Gender Grades

To enable runners to compare themselves against others of the same gender regardless of age, my new age-grading framework calculates a Z-score using the following formula:

$$z = (x - \mu) / \sigma$$

(finishing time - mean for all males or female runners regardless of age / standard deviation)

To calculate the mean and standard deviation for men and women for the first iteration of this model, I investigated the finishing times of 55,683 (33,587 male, 22,096 female) runners, aged from 18 to 80 who completed the Austin Marathon over a 16-year span (2003-2018, excluding the year 2012, which was unavailable). See the graphs below for a distribution of runners by age:



The Austin Marathon, a race strategically located in the center of the United States, attracts a large pool of runners from across the nation. Because officials do not do not impose qualifying times for entry, the race is not dominated by elite runners who would skew results. Additionally, analyzing the Austin Marathon was attractive because the race offered a relatively long history of results, thus ensuring continuity of data. Indeed, the Austin Marathon benefits from having a high percentage of repeat runners, which adds to this study's validity because it tracks the performances of large groups of people as they age. Additionally, because there have been no significant changes to the route, runners have repeated the same course year after year.

Based on 15 years of Austin Marathon results, that means the following:

	World and National Class	Regional Class	Local Class	Participant	
	-2 Standard Deviation	-1 Standard Deviation	Mean	+1 Standard Deviation	Standard Deviation
Male	2:36:58	3:28:46	4:20:35	5:12:24	51:48.5
Female	3:03:06	3:56:15	4:49:25	5:42:35	53:09.8

Calculating Age Grades

Calculating age grades requires first calculating the average finishing times for every age. Another researcher, Niklas Lehto, a statistician from the Lulea University of Technology in Sweden, did exactly that. He completed a groundbreaking large-scale study of age-related performance changes. His [2016 journal article](#) analyzes the times of 312,342 male runners who competed in 36 Stockholm Marathons (1979-2014). Lehto concludes, "The effect of age on performance . . . is clearly measurable, quantifiable, and possible to describe." Moreover, he says it can be modeled using a second-order polynomial, $t = a + bx + cx^2$, where time t as a function of age x . His calculation suggests the average male marathoner slows 4.2 minutes 10 years after his peak age, which he pegs at 34.3 (+ or - 2.6) years.

While Lehto's study provides a great start, it only examines male runners and thus neglects gender differences.

Similar to Lehto, this study performed second-order polynomial least-squares regression analysis except it quantifies the effects of age on marathon performance for both genders. The regressions yield the following equations:

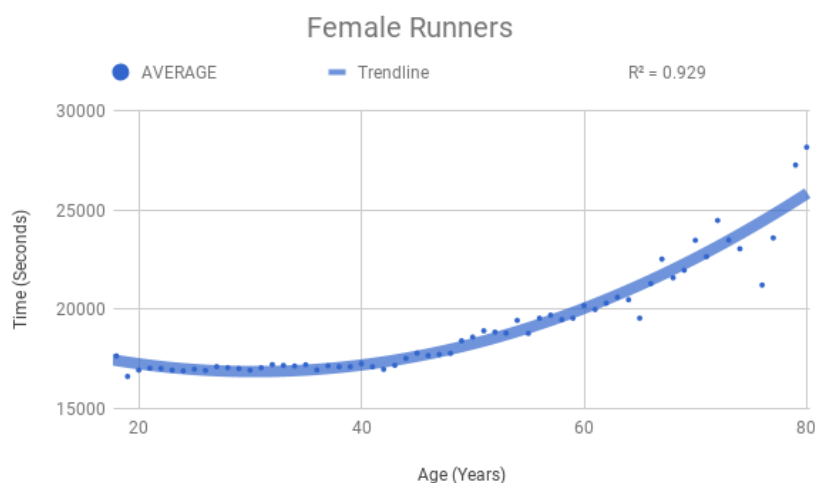
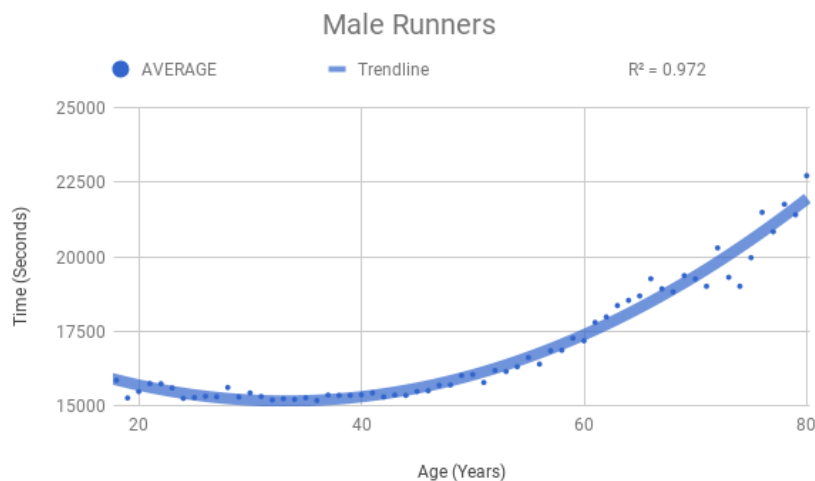
$$\text{Men: } t = 18564 - 205x + 3.1x^2$$

$$\text{Women: } t = 18564 - 2222x + 3.65x^2,$$

where time t in seconds is a function of age x in years.

The two equations respectively explain 97.2% and 92.9% of the variation in the average male and female finishing times as a function of age.

The following graphs respectively plots mean finishing times for men and women in one-year increments (blue dots) and their corresponding "best-fit" regression curves (blue line):



These "best-fit" curves nearly overlay the average times for runners aged from 18 through 60 but, as expected, show more deviation for senior runners. This is likely due to decreasing participation in the most mature age categories, thus contributing to higher volatility in finishing times at the extreme. As a result, the best-fit curves in the above figures are perhaps slightly flatter than they otherwise would be because, by definition, only the most elite elderly athletes run marathons. In other words, self-selection bias serves to slow the rate at which the slope of the curves increases at the geriatric extremes.

To calculate age grades, my new age-grading framework calculates Z-scores using the same formula for gender grades, however it substitutes the mean for runners of a specific age and gender (e.g., 29-year-old male) as modeled by the second-order least squares regression equations.

Insights Into Age and Marathon Performance

In addition to enabling a new and improved age-grading framework, the results of the second-order least squares regressions yield three interesting insights. First, solving for the minimums of each equation (i.e., taking the derivative of each, setting them equal to zero, and solving for x) suggests the optimal age for marathon runners is 33.1 for men and 30.4 years for women. The finding agrees with Lehto's conclusion that the average male Stockholm Marathon runner's performance improves up to 34.3 (+ or - 2.6) years and declines thereafter. This suggests, as Lehto highlighted, that "runners can focus their training to obtain maximal performance during their mid-30s [for males and early-30s for females] instead of late 20s" Second, calculating rates of change (i.e., slopes) along the curves shows year-over-year seconds gained or lost. For example, a 39-year-old male runner can expect, all else being equal, a 39.9 second slower time when he turns 40. Similarly, a female runner can expect a 66.4 second slower time.

Year-Over-Year Change In Performanc (Seconds)

AGE	MALE	FEMALE
19	-90.3	-87
20	-84.1	-79.7
21	-77.9	-72.4
22	-71.7	-65.1
23	-65.5	-57.8
24	-59.3	-50.5
25	-53.1	-43.2
26	-46.9	-35.9
27	-40.7	-28.6
28	-34.5	-21.3
29	-28.3	-14
30	-22.1	-6.7
31	-15.9	0.7
32	-9.7	8
33	-3.5	15.3
34	2.7	22.6
35	8.9	29.9
36	15.1	37.2
37	21.3	44.5
38	27.5	51.8
39	33.7	59.1
40	39.9	66.4
41	46.1	73.7
42	52.3	81
43	58.5	88.3
44	64.7	95.6
45	70.9	102.9
46	77.1	110.2
47	83.3	117.5
48	89.5	124.8
49	95.7	132.1
50	101.9	139.4
51	108.1	146.7
52	114.3	154
53	120.5	161.3
54	126.7	168.6
55	132.9	175.9

56	139.1	183.2
57	145.3	190.5
58	151.5	197.8
59	157.7	205.1
60	163.9	212.4
61	170.1	219.7
62	176.3	227
63	182.5	234.3
64	188.7	241.6
65	194.9	248.9
66	201.1	256.2
67	207.3	263.5
68	213.5	270.8
69	219.7	278.1
70	225.9	285.4
71	232.1	292.7
72	238.3	300
73	244.5	307.3
74	250.7	314.6
75	256.9	321.9
76	263.1	329.2
77	269.3	336.5
78	275.5	343.8
79	281.7	351.1
80	287.9	358.4

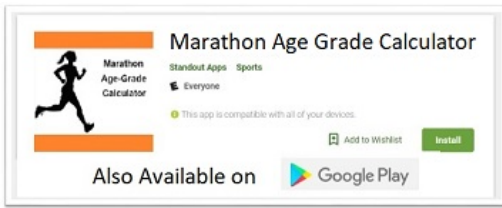
Third, as shown in the above table, women, on average, suffer higher performance degradation as they age than men. This finding is similar to [Jones's conclusion](#) after analyzing the performances of elite runners that women slow with age at a slightly higher rate than males.

Future Research

While the distribution and mean finishing times associated for male and female Austin Marathon runners finishing times is likely representative of many marathon races, future research should expand the volume of data examined. In other words, adding more races to the data included in the regressions would help verify and validate coefficients generated by the least squares regressions. It would also allow age grades to be tailored for specific races. For example, it would be possible for a runner to receive a worldwide age grade as well as ones calculated for specific races (e.g., the New York Marathon). Lastly, with more research and modeling, the framework can be extended to races of all lengths.

<http://www.agegradecalculator.com>

Updated 02/03/2019



The image shows a promotional card for the 'Marathon Age Grade Calculator' app. On the left, there is a silhouette of a runner. To the right of the runner, the text reads 'Marathon Age-Grade Calculator'. Further right, it says 'Standout Apps Sports' and 'Everyone'. Below this, a green checkmark icon is followed by the text 'This app is compatible with all of your devices.' At the bottom right of the card, there are two buttons: 'Add to Wishlist' and 'Install'. At the bottom left, it says 'Also Available on' followed by the Google Play logo and the text 'Google Play'.