A Comparison of Aging Effects Across Athletic Events

Ray C. Fair*

March 2004

Abstract

Records by age are used to estimate deterioration rates in track and field, road racing, and swimming events. Data for both men and women are used for swimming. Eighty one cases are analyzed. The results reveal that deterioration rates are fairly similar both across events and across men and women. In particular, running and swimming deterioration rates are similar after adjusting for distance. The three main differences across the 81 cases are: (1) the deterioration rates for the 100, 200, and 400 meter track events are smaller than for the other running events, (2) from about age 65 on the deterioration rates for swimming increase slightly as the distance increases, both for men and women, and (3) from about age 60 on the deterioration rates for women swimmers are larger than they are for men swimmers. In the end running was pooled into two categories (100-400 meter track events and all other running events), and swimming was pooled into six categories (50 meter or yard events, 100 meter or yard events, and all other events, both for men and women).

---

*Cowles Foundation and International Center for Finance, Yale University, New Haven, CT 06520-8281. Voice: 203-432-3715; Fax: 203-432-6167; e-mail: ray.fair@yale.edu; website: http://fairmodel.econ.yale.edu. I am indebted to William Sudderth for helpful comments, to Linda Honikman for help with the road racing data, to Suzanne Rague for help with the swimming data, and to Oxana Marmer for research assistance.

1 Introduction

In a previous study (Fair (1994)) I used records by age to estimate deterioration rates in a number of track and field events and road racing events. Since this study, many age records have been broken in these events, especially at the older ages. In addition, considerable data are now available for swimming. This paper uses the same methodology as in the previous study to update the earlier results and to add swimming. Data for both men and women are used for swimming.[1] Eighty one cases are analyzed. The results allow comparisons to be made across events and across men and women.

It will be seen that the results are fairly similar both across events and across men and women. In particular, running and swimming deterioration rates are similar after adjusting for distance. The three main differences across the 81 cases are: (1) the deterioration rates for the 100, 200, and 400 meter track events are smaller than for the other running events, (2) from about age 65 on the deterioration rates for swimming increase slightly as the distance increases, both for men and women, and (3) from about age 60 on the deterioration rates for women swimmers are larger than they are for men swimmers. In the end running was pooled into two categories (100-400 meter track events and all other running events), and swimming was pooled into six categories (50 meter or yard events, 100 meter or yard events, and all other events, both for men and women).

[1]Data for women for track and field and road racing were not used in the earlier study and are not used in this one. As discussed below, the number of women in their 90s who have competed in these events is probably not large enough to allow reliable results to be obtained.

2 The Model and Data

All deterioration rates are taken to be in percentage terms in this paper. The estimates in the paper are based on the following three assumptions: (1) deterioration has begun by at least age 35, (2) the deterioration rate is the same per year between age 35 and some age k^*, and (3) the deterioration rate increases by the same amount per year after age k^*. The solid line in Figure 1 reflects these assumptions.[2] (Ignore for now the points in Figure 1.) The line is linear before k^* and quadratic after that. b_k is the log of the biological minimum time for a person of age k for the particular event. The formula for b_k is

$$b_k = \begin{cases} \beta + \alpha k, & 35 \le k \le k^* \\ \gamma + \theta k + \delta k^2, & k > k^* \end{cases} \tag{1}$$

with the restrictions

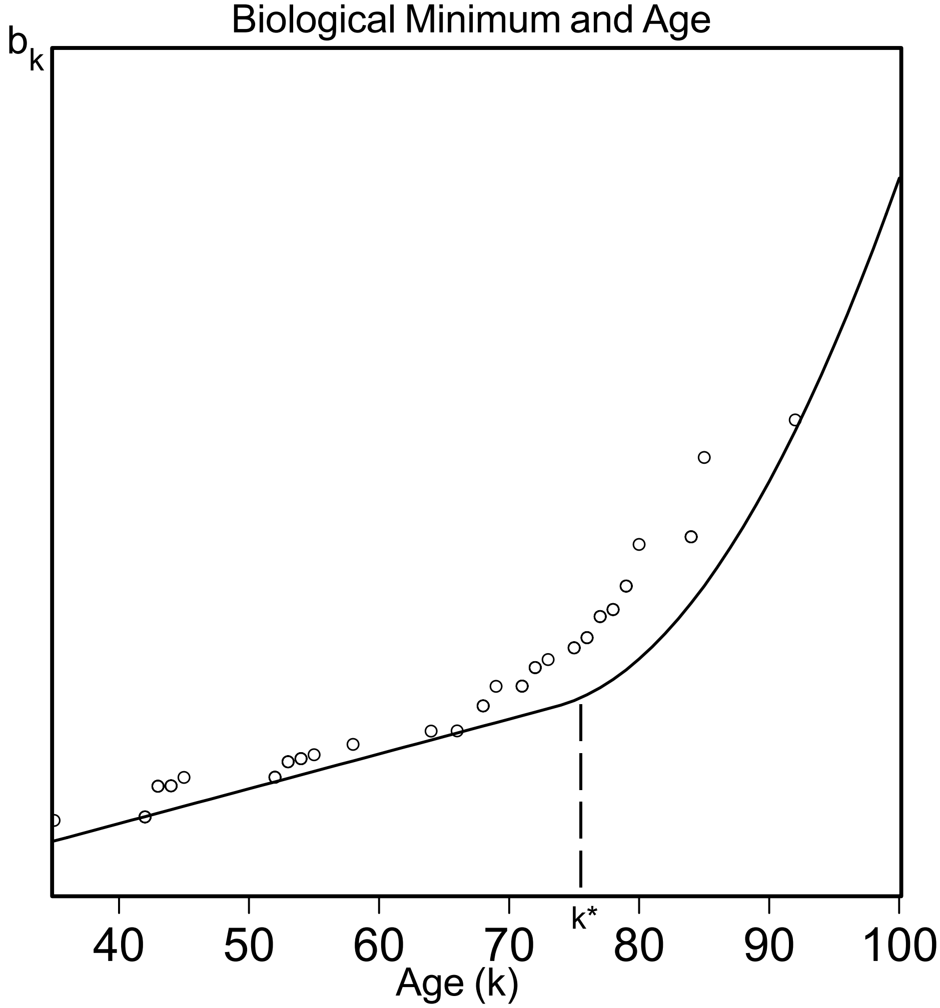
$$\begin{aligned} \gamma &= \beta + \delta k^{*2} \\ \theta &= \alpha - 2\delta k^* \end{aligned} \tag{2}$$

The two restrictions force the linear and quadratic segments to touch and to have the same first derivative at k^*. The unrestricted parameters to estimate are the intercept, β, the slope of the linear segment, α, the age at which the line changes from linear to quadratic, k^*, and the quadratic parameter, δ.

As discussed in Fair (1994), in the initial phase of this work a number of functional forms more complicated that the simple linear-quadratic form were tried, and none seemed to be an improvement. The linear-quadratic form has been

[2] For the high jump the measure of performance is distance, not time, where the larger the better instead of the smaller the better. For simplicity, this paper is written assuming time is the measure, but the switch to distance is straightforward.

Figure 1
Postulated Relationship between
Biological Minimum and Age



used exclusively in this study. The aim is to estimate the function in Figure 1 using records by age for each event and for various pooled events.

The track and field and road racing data are discussed in Fair (1994). For the present study the track and field data are from *Masters Age Records 2003 Edition*, and the road racing data are from TACSTATS/USA. These are data available as of 2003, as opposed to about 1992 for the earlier work. The track and field data give the world record by age for each event. The road racing data give the U.S. record by age for each event. Ideally, world records should be used instead of just U.S. records, but such data are not available for road racing. In one case, however, I did use a non-U.S. record. On September 23, 2003, Fauja Singh, a U.K. citizen, ran the Toronto marathon in 5:40:04. He was 92 at the time. This was too good an observation to pass up.[3]

One of the main differences between this study and the earlier one is the use of older ages here. In Fair (1994) the maximum age used was 83 except for 100 meters, where it was 89. For the present results the maximum age for the running events ranges from 94 to 98 except for the marathon, where it is 92 (Mr. Singh). This is an important difference because it allows better estimates of the quadratic curvature. Some of the road racing events used in Fair (1994) are not used here because of lack of good data at the very old ages. These events are 15K, 10 mile, 20K, half marathon, 30K, and 20 miles. For the same reason, of the eight field events used in Fair (1994) only the high jump is used here. The excluded events are: pole vault, long jump, triple jump, shot put, discuss throw, hammer throw, and javelin throw. The data are particularly poor at the very old ages for the shot put and the three throwing events because most of the competition in these events at the old ages uses lighter weights. Also, data for women are not used because there are still few records by women in their 90s for track and field and road racing.

The swimming data were obtained from the United States Masters Swimming (USMS) website. The site allows one to obtain (by much clicking) the top ten times by five-year age groups for each year between 1993 and 2003 for each swimming

[3]In Fair (2002), Chapter 8, equation (4) below was estimated for the marathon, where the maximum age used was 84. The estimated equation was then used to predict minimum times beyond age 84. The prediction for age 92 was 5:57:06, which at the time compared to the record of 9:23:25 by Paul Spangler in the New York City marathon in 1991. The predicted minimum time was thus 3:26:19 below the current best time. This seemed perhaps an overly optimistic prediction at the time, but Mr. Singh bettered the record by even more, 3:43:21!

event. Each of the ten times gives the age of the swimmer in that year. The times are for U.S. citizens only. (Again, it would be better to have world records, but these do not exist by age.) From these data one can find the best U.S. time by age for each event since 1993. Data are also available by five-year age groups for the best time ever by a U.S. citizen, although the age of the swimmer is not given (within the five-year group). In a few cases the best time ever was set before 1993, and in some of these cases the age of the swimmer was tracked down and this time was used instead of the time obtained from the top ten search.

Records for long course meters (LCM) and short course yards (SCY) were obtained for both men and women. For LCM there are 17 events, and for SCY there are 18 events. Records were thus obtained for 70 swimming cases. These are listed in Table 2 below.

The maximum age for men for swimming is as high as 100 in some events. The data for women are not as good, but in two cases the maximum age is 94 and in a number of others it is 92. Swimming is clearly better than running for capturing performances of very old women.

Regarding the use of the data, for a particular case let r_k denote the log of the record time for age k, and let ϵ_k denote the difference between r_k and the unobserved b_k:

$$r_k = b_k + \epsilon_k \quad . \tag{3}$$

ϵ_k is the measurement error. ϵ_k will be close to zero if the record time is close to the biological minimum. If a large number of people of age k have competed in the event, r_k is likely to be fairly close to b_k and thus ϵ_k close to zero. If, on the other hand, the number who have competed is fairly small, as it is at the very old ages, r_k may be above b_k and thus ϵ_k positive. This problem of a positive ϵ_k will be called the "small sample problem."

One way in which the small sample problem may manifest itself is for r_k to be larger than $r_{k'}$ where k' is larger than k. If this is true, r_k will be called a "dominated time." In the estimation work, dominated times have not been used. Under the assumption that people never get better after age 35, a dominated time cannot have a zero measurement error. Excluding these observations avoids using values that for sure have positive measurement errors. In this spirit, observations were excluded at the maximum ages when their times seemed unrealistically large. Also, in a few cases, mostly women's swimming, the first observation or the first few observations were excluded if they were close to being dominated by an observation at an older age. The age 35 observation for the marathon was excluded; the age 35 observation for the men's LCM 200FL was excluded; and early age observations were excluded for 12 of the 35 women's swimming events. To use the marathon as an example, the age 35 record time is 2:11:40 and the next non-dominated time is 2:12:47 at age 42. The closeness of the age 42 time to the age 35 time suggests that the age 35 time is soft (i.e., has a fairly large positive measurement error associated with it), and so it was not used. More will be said about this below.

The equation that is estimated for a given event is

$$r_k = \beta + \alpha k + \delta d_k (k^{*2} - 2k^* k + k^2) + \epsilon_k, \tag{4}$$

where $d_k = 0$ if $k \leq k^*$ and $d_k = 1$ if $k > k^*$. k ranges over the non-dominated observations. When events $i = 1, \ldots, n$ are pooled, the equation to be estimated is

$$r_{ik} = \beta_1 D_{1ik} + \cdots + \beta_n D_{nik} + \alpha k + \delta d_{ik}(k^{*2} - 2k^* k + k^2) + \epsilon_{ik}, \qquad (5)$$

where r_{ik} is the log of the record for event i and age k, D_{jik} is a dummy variable that is one when event i is equal to j and zero otherwise ($j = 1, \ldots, n$), $d_{ik} = 0$ if $k \leq k^*$ and $d_{ik} = 1$ if $k > k^*$, and ϵ_{ik} is the measurement error for event i and age k. Again, k ranges over the non-dominated observations. The n β_i parameters are the n different intercepts.

The estimation method for equation (4) is choose values of the parameters that minimize the sum of squared residuals subject to the restrictions that all the estimated errors are non-negative and that the estimated error for the first observation is zero. The equation is nonlinear in the parameters β, α, k^*, and δ. These parameters were estimated using a nonlinear optimization algorithm by minimizing the weighted sum $\sum_k \lambda_k \hat{\epsilon}_k^2$, where λ_k is equal to 1 if $\hat{\epsilon}_k \geq 0$ and is equal to a number greater than 1 if $\hat{\epsilon}_k < 0$, where $\hat{\epsilon}_k$ is the estimated error for observation k. This penalizes negative errors more than non-negative ones. In the estimation work a value of 100 was used for λ_k when $\hat{\epsilon}_k$ was less than zero. This was large enough to make nearly all the estimated errors non-negative at the optimum.[4] To insure that the estimated error for the first observations is close to zero, a value of 500 was used for λ_k when k is the first observation. The same procedure was used for the estimation of the pooled equation (5), where there are just more parameters to estimate. In this case the estimated error for the first observation for each of the n

[4]This procedure does not guarantee that all the estimated errors are non-negative because if $\hat{\epsilon}_k$ is negative but very close to zero, its contribution to the objective function is small even if λ_k is large. In practice the negative errors were very close to zero and were for all intents and purposes zero.

events was forced to be close to zero.

The estimates for a number of the cases are sensitive to whether or not the first observation is forced to be on the line (i.e., whether or not the estimated error for the first observation is forced to be close to zero). If this restriction is not imposed, some of the lines imply times that are unrealistically low for ages near 35 (e.g., times that are considerably below the current overall world record). If the measurement error in equation (3) is small for the first observation used, then the current procedure is justified. This is the reason that some of the early observations were excluded if they were nearly dominated by later ones. This avoids forcing the line to go through a soft observation.

The restrictions in (2) that are imposed in the estimation are examples of polynomial spline restrictions—see Poirier (1976). The restriction that all the estimated errors be non-negative is common in the estimation of frontier production functions—see Aigner and Chu (1968) and Schmidt (1976). The added complication here is that equation (4) is nonlinear in parameters. For linear equations the estimation problem can be set up as a quadratic programming problem and solved by standard methods, but for nonlinear equations a procedure like the one described above must be used. Other possible ways of estimating deterioration rates, including using an order-statistic approach, are discussed in Fair (1994), but these are not pursued here.

There is no obvious way to test the hypothesis that the coefficients for one event equal those for another. The assumption of iid errors is clearly not appropriate in this context. In practice, the estimated errors are much larger on average at the old ages, even after excluding the dominated times, which reflects the small sample problem. Comparisons have to be made by looking for patterns across the various cases rather than by formal hypothesis testing.

Estimates are presented below of the first derivative of equation (4) at various ages. The derivative, denoted ρ_k, is

$$\rho_k = \partial r_k / \partial k = \alpha + 2\delta d_k (k - k^*) \ . \tag{6}$$

It is also useful to have an estimate of cumulative percentage loss from some beginning age. Age 35 is usually the beginning age used in such measures, and it has been used here. Let $\hat{r}_k$ denote the predicted value of r_k from equation (4) using the estimated values of β, α, k^*, and δ and zero values for the error term for $k = 35, \ldots, 100$. The cumulative percentage loss, denoted R_k, is then taken to be:

$$R_k = e^{\hat{r}_k} / e^{\hat{r}_{35}}, \quad k = 35, \ldots, 100 \ . \tag{7}$$

R_k will be called the "age factor."

3 The Results

When examining the estimation results it is important to realize that the estimate of the transition age k^* and the estimate of the quadratic parameter δ are collinear. If one is low, the other tends to be low, and vice versa. In other words, sometimes the estimation gives an early transition age and low quadratic curvature, and sometimes it gives a late transition age and high curvature. The best way to see if two estimated equations are similar at the older ages is not to look at the estimates of k^* and δ, but at the age factors and estimated first derivatives.

It will be useful to begin with the pooled estimates. These are presented in Table 1. Presented in the table are the estimates of α, k^*, and δ along with the implied age factors at ages 55, 65, 75, and 85 and the implied first derivatives at ages 75 and 85. Also presented are the number of observations and the maximum age used. $\hat{\alpha}$ is the estimated yearly deterioration rate up to the transition age.

Consider first Sprint versus Run. For Sprint $\hat{\alpha}$ and $\hat{k}^*$ are 0.0063 and 77.3, and for Run they are 0.0082 and 73.7. The estimates of δ are similar. The deterioration rates for Sprint are thus less than they are for Run. By age 85 the age factor is 1.50 for Sprint and 1.83 for Run.

For men's swimming, the values for R_{55} and R_{65} are similar across the three distance categories, but for R_{75} and R_{85} the age factors increase with distance. The results for M200+ show the collinearity between $\hat{k}^*$ and $\hat{\delta}$ discussed above. The estimate of k^* is low relative to those for M50 and M100, as is the estimate of δ.

It is interesting to note that the age factors for M50 and Sprint are fairly similar, as are the age factors for M200+ and Run. These results suggest that the deterioration rates for swimming and running are similar after correcting for distance. It is the case, however, that R_{55} and R_{65} are noticeably larger for Run than for M200+ and slightly larger for Sprint than for M50. Swimmers are thus estimated to slow down less rapidly than runners until about the mid 60s.

The results for women swimmers are more problematic because of the small sample problem at the old ages. For W100 and W200+ the estimated value for k^* was 35, which means that there is no linear segment before the quadratic. The results are similar for women than for men in that the age factors increase with distance after a certain age. For women this age is somewhere between 55 and

Table 1
Pooled Estimates

Pooled Events	$\hat{\alpha}$	$\hat{k}^*$	$\hat{\delta}$	R_{55}	R_{65}	R_{75}	R_{85}	$\hat{\rho}_{75}$	$\hat{\rho}_{85}$	No. Obs.	Max Age
Sprint	0.0063	77.3	0.00154	1.13	1.21	1.29	1.50	0.63	3.00	119	98
Run	0.0082	73.7	0.00149	1.18	1.28	1.39	1.83	1.23	4.21	257	96
M50	0.0049	73.4	0.00124	1.10	1.16	1.22	1.51	0.90	3.39	256	100
M100	0.0059	70.2	0.00109	1.12	1.19	1.30	1.71	1.64	3.83	319	100
M200+	0.0045	54.8	0.00041	1.09	1.19	1.41	1.81	2.09	2.91	574	100
W50	0.0051	53.6	0.00036	1.11	1.22	1.44	1.83	2.04	2.75	231	92
W100	-0.0008	35.0	0.00028	1.10	1.26	1.52	1.93	2.16	2.72	263	94
W200+	0.0023	35.0	0.00022	1.14	1.31	1.56	1.94	1.98	2.42	542	94
HJ	-0.0096	73.3	-0.00085	1.21	1.33	1.47	1.81	1.24	2.94	34	96

Notes:
- Sprint = 100, 200, and 400 meter track.
- Run = all running except 100, 200, and 400 meter track.
- M50 = 50 meter and yard swimming events, men.
- M100 = 100 meter and yard swimming events, men.
- M200+ = all other swimming events, men.
- W50 = 50 meter and yard swimming events, women.
- W100 = 100 meter and yard swimming events, women.
- W200+ = all other swimming events, women.
- HJ = high jump.

65, whereas for men it is somewhere between 65 and 75. It is also the case that the age factors for women are greater than those for men beginning somewhere between age 55 and 65. Again, however, the results for women at the older ages must be taken with considerable caution, and it may be that some of the estimated age factors for women at the older ages are too large.

The results for the high jump are similar to those for Run.[5]

Table 2 presents the individual estimates for the 81 cases. The format is the same as that for Table 1. The first thing to look for are estimates that are out of line

[5]For ease of comparison for the high jump, the R_k values in Tables 1, 2, and 3 are the reciprocal of the values computed in equation (7) and the derivatives have had their sign reversed.

Table 2
Individual Estimates

Events	$\hat{\alpha}$	$\hat{k}^*$	$\hat{\delta}$	R_{55}	R_{65}	R_{75}	R_{85}	$\hat{\rho}_{75}$	$\hat{\rho}_{85}$	No. Obs.	Max Age
Sprint											
100M	0.0066	79.9	0.00187	1.14	1.22	1.30	1.46	0.66	2.56	36	98
200M	0.0073	76.8	0.00168	1.16	1.24	1.34	1.61	0.73	3.47	40	98
400M	0.0058	73.1	0.00162	1.12	1.19	1.27	1.68	1.18	4.43	43	98
Run											
800M	0.0085	73.3	0.00145	1.18	1.29	1.41	1.86	1.33	4.24	42	95
1500M	0.0090	77.1	0.00240	1.20	1.31	1.43	1.82	0.90	4.69	46	96
5000M	0.0079	71.8	0.00118	1.17	1.27	1.39	1.82	1.53	3.88	40	95
10000M	0.0090	79.1	0.00256	1.20	1.31	1.43	1.72	0.90	3.94	38	94
5K	0.0078	71.7	0.00150	1.17	1.26	1.39	1.93	1.78	4.78	34	95
10K	0.0073	74.1	0.00204	1.16	1.24	1.34	1.83	1.09	5.17	42	94
Marathon	0.0070	56.5	0.00047	1.15	1.28	1.55	2.08	2.44	3.38	25	92
Field Event											
High Jump	-0.0096	73.3	-0.00085	1.21	1.33	1.47	1.81	1.24	2.94	34	96

Notes:

- M = meters, K = kilometers.
- M events are track, K events are road racing.

with the others, and there are actually very few in Table 2. For the marathon $\hat{k}^*$ is noticeably lower than for the other running events. This is discussed below. For swimming the largest differences are for the butterfly (FL) for both men and women, where the age factors are generally larger than for the others. The maximum ages for FL are generally lower than for the others, which may reflect a more serious small sample problem for FL than for the others.

Regarding pooling, it seems clear from Table 2 that the 100 meter, 200 meter, and 400 meter results are close enough to warrant pooling. For the remaining running events the main question is what to do about the marathon. Aside from the results for the marathon, the results for the other running events are fairly close. The marathon is a case of a low estimated value of k^* going along with a

Table 2 (continued)

Events	$\hat{\alpha}$	$\hat{k}^*$	$\hat{\delta}$	R_{55}	R_{65}	R_{75}	R_{85}	$\hat{\rho}_{75}$	$\hat{\rho}_{85}$	No. Obs.	Max Age
Swimming, LCM, Men											
50FR	0.0032	78.1	0.00231	1.07	1.10	1.14	1.31	0.32	3.52	26	100
100FR	0.0056	74.9	0.00172	1.12	1.18	1.25	1.58	0.60	4.05	34	100
200FR	0.0054	68.8	0.00096	1.11	1.18	1.29	1.68	1.72	3.65	32	96
400FR	0.0035	58.8	0.00052	1.07	1.13	1.32	1.71	2.05	3.10	27	96
800FR	0.0033	55.8	0.00046	1.07	1.15	1.35	1.75	2.10	3.03	26	96
1500FR	0.0022	52.0	0.00044	1.05	1.15	1.38	1.81	2.26	3.14	25	96
50BA	0.0068	72.4	0.00118	1.14	1.22	1.32	1.69	1.29	3.65	33	100
100BA	0.0079	73.5	0.00122	1.17	1.27	1.38	1.74	1.15	3.60	35	100
200BA	0.0088	68.8	0.00082	1.19	1.30	1.46	1.92	1.89	3.54	31	100
50BR	0.0065	71.7	0.00129	1.14	1.22	1.32	1.74	1.51	4.09	33	96
100BR	0.0080	74.4	0.00178	1.17	1.27	1.38	1.82	1.01	4.56	35	96
200BR	0.0083	74.4	0.00160	1.18	1.28	1.39	1.81	1.01	4.20	36	95
50FL	0.0044	64.0	0.00094	1.09	1.14	1.34	1.89	2.52	4.41	24	91
100FL	0.0108	57.5	0.00047	1.24	1.42	1.78	2.45	2.74	3.69	25	91
200FL	0.0058	46.7	0.00037	1.15	1.35	1.70	2.31	2.70	3.45	24	91
200IM	0.0101	72.7	0.00125	1.22	1.35	1.51	2.00	1.59	4.10	29	91
400IM	0.0070	55.0	0.00041	1.15	1.29	1.56	2.05	2.33	3.15	28	91
Swimming, SCY, Men											
50FR	0.0036	74.0	0.00160	1.07	1.11	1.16	1.45	0.66	3.86	35	100
100FR	0.0033	67.2	0.00104	1.07	1.10	1.22	1.64	1.95	4.02	40	100
200FR	0.0051	67.5	0.00094	1.11	1.16	1.29	1.72	1.91	3.79	37	100
500FR	0.0033	51.8	0.00037	1.07	1.18	1.39	1.77	2.04	2.78	35	95
1000FR	0.0037	58.2	0.00060	1.08	1.15	1.37	1.85	2.39	3.60	33	96
1650FR	0.0048	56.9	0.00051	1.10	1.19	1.43	1.90	2.32	3.34	29	93
50BA	0.0049	59.0	0.00049	1.10	1.18	1.38	1.78	2.06	3.03	35	95
100BA	0.0082	70.7	0.00107	1.18	1.28	1.42	1.88	1.74	3.89	42	98
200BA	0.0081	72.7	0.00155	1.18	1.27	1.39	1.89	1.52	4.61	38	94
50BR	0.0076	74.1	0.00153	1.16	1.26	1.36	1.76	1.04	4.10	37	96
100BR	0.0072	70.5	0.00130	1.15	1.24	1.37	1.88	1.89	4.48	38	96
200BR	0.0089	72.6	0.00157	1.20	1.31	1.44	1.99	1.66	4.80	43	94
50FL	0.0048	68.9	0.00171	1.10	1.16	1.29	1.98	2.58	5.99	33	91
100FL	0.0040	56.3	0.00077	1.08	1.19	1.53	2.30	3.27	4.80	33	90
200FL	0.0091	62.2	0.00095	1.20	1.33	1.68	2.59	3.34	5.24	30	90
100IM	0.0071	65.8	0.00087	1.15	1.24	1.43	1.97	2.31	4.04	37	94
200IM	0.0097	74.9	0.00242	1.21	1.34	1.47	2.08	1.02	5.86	34	91
400IM	0.0088	67.5	0.00120	1.19	1.30	1.52	2.24	2.67	5.08	37	90

Notes:

- LCM = long course meters, SCY = short course yards.
- FR = free, BA = back, BR = breast, FL = fly, IM = individual medley.

Table 2 (continued)

Events	$\hat{\alpha}$	$\hat{k}^*$	$\hat{\delta}$	R_{55}	R_{65}	R_{75}	R_{85}	$\hat{\rho}_{75}$	$\hat{\rho}_{85}$	No. Obs.	Max Age
Swimming, LCM, Women											
50FR	0.0032	48.8	0.00033	1.08	1.20	1.42	1.79	2.02	2.67	35	92
100FR	0.0047	47.3	0.00029	1.12	1.26	1.51	1.91	2.08	2.67	29	94
200FR	0.0038	41.9	0.00025	1.13	1.28	1.52	1.91	2.00	2.49	33	94
400FR	0.0086	72.0	0.00144	1.19	1.30	1.43	1.96	1.73	4.61	32	92
800FR	0.0016	41.9	0.00029	1.09	1.23	1.47	1.86	2.09	2.67	31	92
1500FR	0.0122	77.8	0.00220	1.28	1.44	1.63	2.06	1.22	4.38	21	91
50BA	0.0036	50.2	0.00026	1.08	1.18	1.36	1.64	1.65	2.17	23	91
100BA	0.0083	65.4	0.00055	1.18	1.28	1.47	1.87	1.88	2.97	27	90
200BA	0.0105	70.3	0.00067	1.23	1.37	1.54	1.95	1.67	3.01	34	91
50BR	0.0086	66.8	0.00093	1.19	1.30	1.50	2.09	2.38	4.24	20	90
100BR	0.0097	61.7	0.00050	1.21	1.34	1.61	2.13	2.30	3.30	29	91
200BR	0.0104	64.7	0.00058	1.23	1.37	1.61	2.14	2.24	3.40	34	90
50FL	0.0059	56.4	0.00078	1.13	1.26	1.66	2.55	3.50	5.06	26	88
100FL	0.0194	69.5	0.00075	1.47	1.79	2.23	3.17	2.77	4.28	25	89
200FL	0.0049	45.7	0.00043	1.15	1.36	1.77	2.50	3.03	3.89	29	86
200IM	0.0053	49.4	0.00042	1.13	1.30	1.62	2.21	2.66	3.49	33	91
400IM	0.0092	64.1	0.00076	1.20	1.32	1.58	2.21	2.58	4.09	29	89
Swimming, SCY, Women											
50FR	0.0072	61.5	0.00047	1.15	1.25	1.45	1.86	1.99	2.94	33	91
100FR	0.0065	58.9	0.00057	1.14	1.24	1.50	2.03	2.47	3.60	33	91
200FR	0.0084	48.9	0.00022	1.19	1.36	1.63	2.03	2.00	2.45	32	92
500FR	0.0109	70.8	0.00095	1.24	1.38	1.57	2.08	1.88	3.79	32	92
1000FR	0.0095	65.3	0.00060	1.21	1.33	1.55	2.03	2.11	3.30	26	92
1650FR	0.0097	59.2	0.00031	1.21	1.35	1.59	1.99	1.94	2.56	25	90
50BA	0.0051	46.8	0.00026	1.13	1.27	1.50	1.88	1.96	2.47	29	91
100BA	0.0117	68.0	0.00060	1.26	1.42	1.64	2.13	2.00	3.19	33	91
200BA	0.0109	60.8	0.00032	1.24	1.39	1.65	2.08	1.99	2.63	28	90
50BR	0.0094	66.4	0.00087	1.21	1.33	1.55	2.16	2.44	4.17	28	90
100BR	0.0103	59.6	0.00044	1.23	1.38	1.68	2.23	2.40	3.29	30	88
200BR	0.0104	62.8	0.00059	1.23	1.37	1.65	2.25	2.47	3.65	33	90
50FL	0.0061	57.9	0.00080	1.13	1.25	1.62	2.45	3.36	4.97	37	91
100FL	0.0075	46.2	0.00038	1.20	1.43	1.85	2.58	2.94	3.70	27	90
200FL	0.0190	70.1	0.00074	1.46	1.77	2.18	3.05	2.62	4.09	24	90
100IM	0.0118	69.1	0.00099	1.27	1.43	1.66	2.32	2.35	4.32	30	92
200IM	0.0132	67.9	0.00080	1.30	1.48	1.76	2.44	2.45	4.05	32	92
400IM	0.0106	65.7	0.00079	1.24	1.37	1.64	2.28	2.53	4.11	34	90

Notes:

- LCM = long course meters, SCY = short course yards.
- FR = free, BA = back, BR = breast, FL = fly, IM = individual medley.

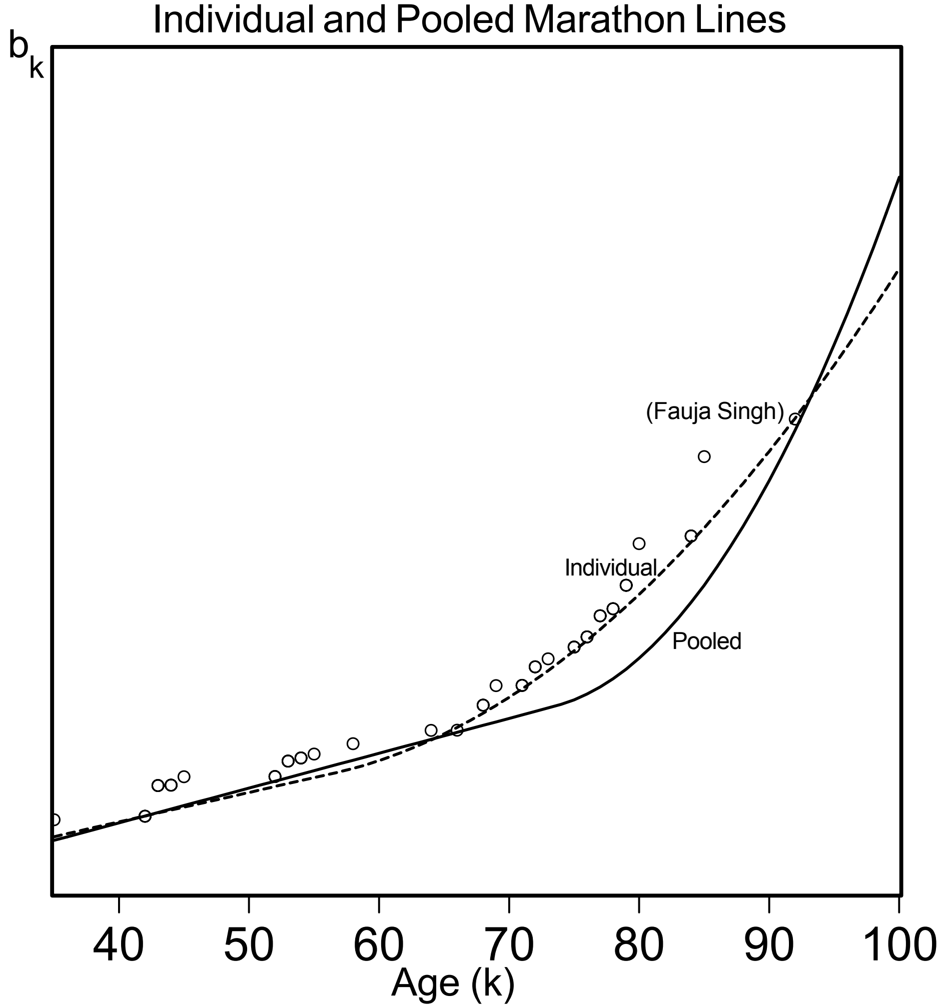
low estimated value of δ. Less confidence can be placed on the estimates for the marathon than for the other running events because the maximum age is only 92 for the marathon. It is the case, however, that the values of R_k for the marathon are fairly similar to those for the other running events, and primarily because of this, the marathon was pooled with the other running events. The individual marathon estimates probably imply deterioration rates that are too small in the middle and late 90s.

The marathon issue can also be seen graphically. In Figure 1 the solid line is the line for the pooled running events (excluding 100, 200, and 400 meters) and the points are the non-dominated marathon observations. Figure 2 is the same as Figure 1 except that it has the individual marathon line added. It is easy to see what is happening. Both lines essentially go through Mr. Singh's age 92 point, but the individual line gets closer to the points in the 70s and 80s. This is done by having the quadratic start earlier (age 56.5 instead of 73.7). The individual line is then more optimistic after age 92 than is the pooled line. This is probably not sensible—that people in their late 90s slow down less in the marathon than in shorter events. The marathon sample is probably not large enough to provide reliable estimates after 92. An advantage of pooling is that there are a number of observations for other events in the mid 90s, which can pin down the curvature better.[6]

Regarding swimming, it is generally the case in Table 2 that the age factors increase with age, especially at the older ages. The age factors are also generally

[6] Note in Figure 2 that both lines go through the age 42 point. Remember from the discussion in Section 2 that the age 35 observation was excluded for the marathon, which means that the next observation (in this case age 42) is forced to have a zero estimated error.

Figure 2
Individual and Pooled Marathon Lines



larger for women than for men, again especially at the older ages. The pooling in Table 1 is designed to pick up these differences.

4 Age-Graded Tables

Table 3 presents the age factors R_k for ages 35 through 100 using the pooled estimates in Table 1. These age factors have already been presented in Table 1 for ages 55, 65, 75, and 65. Although the estimates have been presented through age 100, not much confidence should be placed on the estimates in the 90s because of the small sample problem. The true curvature is not likely to be pinned down very

Table 3
Age Factors (R_k) using Estimates in Table 1

Age	Sprint	Run	M50	M100	M200+	W50	W100	W200+	HJ
35	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
36	1.006	1.008	1.005	1.006	1.004	1.005	1.000	1.003	1.010
37	1.013	1.017	1.010	1.012	1.009	1.010	1.000	1.006	1.019
38	1.019	1.025	1.015	1.018	1.014	1.015	1.000	1.009	1.029
39	1.025	1.034	1.020	1.024	1.018	1.020	1.001	1.013	1.039
40	1.032	1.042	1.025	1.030	1.023	1.026	1.003	1.017	1.049
41	1.038	1.051	1.030	1.036	1.027	1.031	1.005	1.022	1.059
42	1.045	1.059	1.035	1.042	1.032	1.036	1.008	1.027	1.069
43	1.052	1.068	1.040	1.048	1.037	1.041	1.012	1.033	1.080
44	1.058	1.077	1.045	1.054	1.041	1.047	1.016	1.039	1.090
45	1.065	1.086	1.051	1.061	1.046	1.052	1.020	1.046	1.101
46	1.072	1.095	1.056	1.067	1.051	1.057	1.026	1.053	1.111
47	1.078	1.104	1.061	1.073	1.055	1.063	1.031	1.061	1.122
48	1.085	1.113	1.066	1.080	1.060	1.068	1.038	1.070	1.133
49	1.092	1.122	1.072	1.086	1.065	1.074	1.045	1.078	1.144
50	1.099	1.132	1.077	1.092	1.070	1.079	1.053	1.088	1.155
51	1.106	1.141	1.082	1.099	1.074	1.084	1.061	1.098	1.166
52	1.113	1.150	1.088	1.105	1.079	1.090	1.070	1.108	1.177
53	1.120	1.160	1.093	1.112	1.084	1.096	1.080	1.119	1.188
54	1.127	1.170	1.098	1.118	1.089	1.101	1.090	1.131	1.200
55	1.134	1.179	1.104	1.125	1.094	1.107	1.101	1.143	1.212
56	1.141	1.189	1.109	1.132	1.099	1.115	1.113	1.156	1.223
57	1.148	1.199	1.115	1.138	1.106	1.123	1.125	1.170	1.235
58	1.155	1.209	1.120	1.145	1.113	1.131	1.139	1.184	1.247
59	1.163	1.219	1.126	1.152	1.122	1.141	1.153	1.199	1.259
60	1.170	1.229	1.131	1.158	1.131	1.152	1.168	1.215	1.271
61	1.177	1.239	1.137	1.165	1.141	1.163	1.184	1.232	1.283
62	1.185	1.249	1.143	1.172	1.153	1.176	1.201	1.249	1.296
63	1.192	1.260	1.148	1.179	1.165	1.190	1.218	1.267	1.308
64	1.200	1.270	1.154	1.186	1.179	1.204	1.237	1.286	1.321
65	1.207	1.281	1.160	1.193	1.193	1.220	1.256	1.306	1.333
66	1.215	1.291	1.165	1.200	1.209	1.236	1.277	1.326	1.346
67	1.223	1.302	1.171	1.207	1.226	1.254	1.299	1.348	1.359
68	1.230	1.313	1.177	1.214	1.244	1.273	1.321	1.370	1.372
69	1.238	1.324	1.183	1.221	1.264	1.293	1.345	1.394	1.386
70	1.246	1.335	1.189	1.229	1.285	1.315	1.370	1.418	1.399
71	1.254	1.346	1.194	1.237	1.307	1.337	1.397	1.444	1.412
72	1.262	1.357	1.200	1.248	1.331	1.362	1.424	1.470	1.426
73	1.270	1.368	1.206	1.262	1.357	1.387	1.453	1.498	1.440
74	1.278	1.380	1.213	1.278	1.384	1.414	1.484	1.527	1.454
75	1.286	1.395	1.222	1.298	1.412	1.443	1.516	1.557	1.471

Table 3 (continued)

Age	Sprint	Run	M50	M100	M200+	W50	W100	W200+	HJ
76	1.294	1.414	1.235	1.321	1.443	1.473	1.549	1.589	1.491
77	1.302	1.438	1.251	1.347	1.475	1.505	1.584	1.622	1.513
78	1.311	1.466	1.270	1.377	1.510	1.539	1.621	1.656	1.539
79	1.324	1.500	1.293	1.411	1.546	1.574	1.660	1.692	1.567
80	1.341	1.539	1.319	1.448	1.584	1.612	1.700	1.729	1.599
81	1.363	1.584	1.349	1.490	1.625	1.651	1.743	1.768	1.634
82	1.389	1.634	1.383	1.536	1.668	1.693	1.787	1.809	1.673
83	1.421	1.692	1.422	1.587	1.714	1.737	1.834	1.851	1.716
84	1.457	1.757	1.466	1.644	1.763	1.784	1.883	1.895	1.762
85	1.499	1.830	1.514	1.706	1.814	1.833	1.934	1.941	1.813
86	1.547	1.911	1.568	1.775	1.868	1.885	1.988	1.989	1.869
87	1.602	2.002	1.628	1.850	1.926	1.939	2.044	2.039	1.930
88	1.663	2.104	1.695	1.933	1.987	1.997	2.103	2.091	1.996
89	1.732	2.217	1.769	2.024	2.051	2.058	2.166	2.146	2.067
90	1.810	2.343	1.850	2.124	2.119	2.122	2.231	2.203	2.145
91	1.897	2.484	1.940	2.233	2.192	2.190	2.299	2.262	2.230
92	1.995	2.642	2.040	2.353	2.269	2.262	2.371	2.324	2.322
93	2.104	2.817	2.150	2.486	2.350	2.337	2.447	2.389	2.421
94	2.225	3.014	2.271	2.631	2.436	2.417	2.526	2.456	2.530
95	2.361	3.233	2.406	2.791	2.527	2.501	2.609	2.527	2.647
96	2.513	3.479	2.554	2.967	2.624	2.590	2.697	2.601	2.775
97	2.683	3.755	2.719	3.162	2.727	2.685	2.789	2.678	2.914
98	2.874	4.065	2.901	3.376	2.837	2.784	2.886	2.759	3.065
99	3.088	4.414	3.103	3.613	2.953	2.890	2.988	2.843	3.229
100	3.327	4.807	3.328	3.875	3.076	3.001	3.095	2.931	3.408

• See notes to Table 1.

well at the very old ages.

To take one example in Table 3, the age 80 row shows that Sprint and M50 are similar, that Run and M200+ are similar, that women's age factors are larger than men's holding distance constant, and that the age factors increase with distance. These results have already been discussed in the context of Table 1.

Although the deterioration rates in Table 3 are based on data for elite athletes, they can be used by a typical person under the assumption that deterioration rates are the same across people. Although a typical person is much slower than an elite athlete, it may be that the percent deterioration rates are similar between typical people and elite athletes. The age factors may thus be a useful guide on an individual basis, provided that one is not sick, injured, or in poor shape. Given, say, one's time at age 45, the age factors in Table 3 can be used to predict one's minimum time at other ages.[7]

One extra assumption is needed to use the age factors in Table 3 if the base age is less than 35, which is that deterioration does not start until age 35. The parameter estimates in this paper do not require that deterioration begins no earlier than age 35. On the other hand, if one's best time was achieved at, say, age 29, then using the age factors in Table 3 implicitly assumes that this same time could have been achieved at age 35. If deterioration in fact began before age 35, then the age factors in the table are too low for comparisons to the age 29 time.

5 Conclusion and Policy Implications

As more data for very old competitors become available, better estimates of the deterioration rates at the very old ages will be possible. The overall results in Table 2 do, however, seem close enough across similar events for fairly good estimates to be made now at all but the very old ages. It is interesting that while there are differences across distance and sex, they do not seem that large. In Table 1 the age-factor range is 1.16 to 1.33 for age 65, 1.22 to 1.56 for age 75, and 1.50 to 1.94 for age 85. Table 3 shows that the age factors double between ages 87 and

[7]The website http://fairmodel.econ.yale.edu allows a user to type in his or her age, time, and event and have returned the predicted minimum times at other ages. These computations use the age factors in Table 3.

93. If Sprint and M50 are excluded, the age factors double between ages 87 and 89. One conclusion is thus that while deterioration rates do differ across events and sex, the differences in general are fairly small.

A second conclusion is that the overall deterioration rates do not seem that large. For these vigorous activities, the loss from age 35 to age 65 is between 16 and 33 percent, and the loss from age 35 to age 75 is between 22 and 56 percent. Doubling of times occurs in the late 80s. The policy implications from this second conclusion are obvious. If deterioration rates are no larger than those reported here, old people have the capacity to make important contributions to society and to do most things for themselves, albeit a little more slowly. This requires, of course, that a person is not sick or injured and stays in shape.

An interesting question for future work is whether other events can be used. In the future enough data will be available to allow track and field and road racing events to be analyzed for women. For the present study I tried using Crash B rowing results (for both men and women), but there are not yet enough observations at the very old ages. The sport is only about 20 years old. Other possibilities in the future are bicycling and cross country skiing. Finally, a number of people have asked me about golf. I tried to get for some major golf course best scores by age for the course, but these data do not appear to be recorded. This is surprising given that so much handicapping data are available for golf, and it may be in the future that best scores by age will be recorded.

References

[1] Aigner, D. J., and S. F. Chu, 1968, "On Estimating the Industry Production Function," *The American Economic Review*, 58, 826-839.

[2] Fair, Ray C., 1994, "How Fast Do Old Men Slow Down?" *The Review of Economics and Statistics*, 103-118.

[3] _________, 2002, *Predicting Presidential Elections and Other Things*, Stanford: Stanford University Press.

[4] Mundle, Peter, Shirley Dietderich, Rex Harvey, and Beverly LaVeck, 2003, *Masters Age Records 2003 Edition*,

[5] Poirier, D. J., 1976, *The Economics of Structural Change with Special Emphasis on Spline Functions*, Amsterdam: North-Holland.

[6] Schmidt, Peter, 1976, "On the Statistical Estimation of Parametric Frontier Production Functions," *The Review of Economics and Statistics*, 53, 238-239.