Masters Handicapping

Over the years, Green Lake Crew has occasionally had questions from masters about our handicapping algorithm. We have always used the US Rowing equation/methodology, but could provide no other information about it. On the last inquiry, we thought that is this day and age where data is so plentiful and the ability to collect data was nothing more than a website, we decided to run some numbers.

Initially, the intent was to create a site where master rowers would come in droves to provide data about their erg times and we would run the numbers. After setting up the website, we found out that masters weren't coming in droves to we also realized we needed a lot of data. It was recommended by Tif Wood, that we contact Concept2 to see if they would allow access to their data. They definitely accommodated my requests.

What we wanted

The original plan was to collect our own data so that we could control the quality. Our request is:

- Personal best erg time for 1K and 5K pieces within the previous six (6) months.
- We requested the contributor to have at least two years racing experience or have rowed as a junior or collegiate level.
- Of course, over the age of 27.
- That they are honest about their times.

We are interested in 1K erg times because that is the masters sprint race distance. We are interested in 5K erg times because that is roughly the head race distance. We want both because how an individual rows, depends on how far he/she needs to row. We are interested in personal best time because in a race condition, we believe races will strive to that level of effort.

What we got from Concept2

Concept2 has been collection erg time since 2004. They have been collecting data for the purpose of ranking. What this means is that the quality of data would not be the same as what we had in mind. We don't know if their data is representative of the entire rowing population or a self-selected subset composed on only those who are interested enough to see how they rank. Clearly, they represent the more serious athletes. There are also issues of duplicate entries, . . . although for masters erg times, we are not certain that is an issue. Even in our own request for data, we had no conceptual problem with an individual entering their personal best (previous six months) every year because we reason that as a rower age, their ability changes.

Nevertheless, the data provided is data that we simply would not have access to and at the end of the day, it's better than nothing.

The Data

Concept2 provided two (2) sets of data:

- 1K for ages >=27 from 2004 through present, all sources. A total of 36817 records.
- 5K for ages>27 from 2004 through present, all sources. A total of 90459 records.

The sources of data that Concept2 tracks are: Race, IND, IND_V, C2Log, ErgData and RowPro.

What we did

With the separate data groupings we first looked at the distribution of records by age, gender, data source and weight classification. Basically, for each age, we wanted to see the number of men/women, the distribution from each data source and the number of lightweights. The summary table is as follows:

Report counts the number of data points in the data file by age, male/female counts, male/female counts from each data source, male/female counts for light and heavyweight rowers.

Age	Male Count	Female Count	Data Source (Male/Female) Counts							Weight Classification (Light/Heavy)	
			None	Race	IND	IND_V	C2Log	ErgData	RowPro	Male	Female
28	1266	213	0 / 0	1 / 1	1065 / 191	19 / 3	139 / 14	15 / 1	27 / 3	195 / 1071	81 / 132
29	1304	273	1 / 0	0 / 2	1083 / 242	20 / 4	163 / 18	12 / 1	25 / 6	176 / 1128	90 / 183
30	1495	291	1 / 0	1 / 1	1232 / 254	26 / 3	182 / 25	15 / 2	38 / 6	239 / 1256	85 / 206
31	1662	277	0 / 0	2 / 1	1336 / 250	42 / 6	213 / 15	33 / 2	36 / 3	243 / 1419	74 / 203
32	1788	290	0 / 0	2 / 0	1467 / 254	27 / 4	204 / 24	32 / 3	56 / 5	243 / 1545	87 / 203
33	1987	346	0 / 0	1 / 0	1610 / 306	18 / 7	241 / 25	32 / 1	85 / 7	288 / 1699	124 / 222
34	1865	382	1 / 0	3 / 2	1487 / 338	27 / 3	242 / 31	36 / 4	69 / 4	280 / 1585	117 / 265
35	2192	396	2 / 0	2 / 0	1723 / 330	49 / 12	293 / 37	36 / 9	87 / 8	287 / 1905	99 / 297
36	2373	423	1 / 0	1 / 1	1855 / 374	32 / 5	308 / 37	67 / 1	109 / 5	340 / 2033	122 / 301
37	2499	430	0 / 0	0 / 0	1988 / 371	35 / 11	303 / 30	64 / 7	109 / 11	336 / 2163	128 / 302
38	2477	429	2 / 0	0 / 0	1941 / 380	33 / 5	336 / 32	52 / 5	113 / 7	327 / 2150	105 / 324
39	2585	425	2 / 0	0 / 0	2017 / 370	38 / 8	371 / 37	48 / 2	109 / 8	356 / 2229	122 / 303
40	2721	493	2 / 0	0 / 1	2140 / 418	49 / 9	348 / 49	56 / 8	126 / 8	363 / 2358	146 / 347
41	2800	524	3 / 0	1 / 1	2231 / 464	43 / 8	342 / 39	62 / 5	118 / 7	437 / 2363	162 / 362
42	2908	547	2 / 0	2 / 0	2224 / 475	61 / 7	408 / 42	72 / 13	139 / 10	413 / 2495	159 / 388
43	2850	550	0 / 0	3 / 0	2252 / 466	54 / 11	369 / 59	52 / 6	120 / 8	399 / 2451	150 / 400
44	2833	547	1 / 0	3 / 1	2238 / 462	34 / 5	361 / 62	68 / 6	128 / 11	446 / 2387	149 / 398
45	2706	571	1 / 0	1 / 0	2098 / 500	37 / 10	377 / 45	51 / 8	141 / 8	393 / 2313	177 / 394
46	2776	521	0 / 0	1 / 0	2155 / 465	37 / 10	361 / 33	60 / 8	162 / 5	415 / 2361	178 / 343
47	2631	511	0 / 0	3 / 2	2072 / 451	33 / 8	339 / 35	62 / 5	122 / 10	372 / 2259	168 / 343
48	2489	510	1 / 0	2 / 0	1968 / 443	50 / 5	266 / 39	63 / 6	139 / 17	335 / 2154	146 / 364
49	2257	463	0 / 0	0 / 0	1763 / 406	28 / 8	290 / 29	56 / 14	120 / 6	333 / 1924	133 / 330
50	2467	513	0 / 0	2 / 2	1977 / 444	35 / 12	283 / 37	52 / 5	118 / 13	358 / 2109	159 / 354
51	2192	461	1 / 0	0 / 0	1771 / 398	42 / 7	250 / 40	37 / 4	91 / 12	334 / 1858	158 / 303
52	2147	444	0 / 0	3 / 2	1738 / 389	36 / 7	239 / 31	40 / 6	91 / 9	347 / 1800	144 / 300

53	2009	426	1 / 0	0 / 1	1592 / 360	29 / 10	257 / 40	24 / 3	106 / 12	302 / 1707	130 / 296
54	1863	366	0 / 0	2 / 1	1505 / 316	21 / 7	233 / 36	25 / 1	77 / 5	294 / 1569	112 / 254
55	1716	331	0 / 0	0 / 1	1434 / 290	14 / 5	173 / 21	38 / 8	57 / 6	248 / 1468	114 / 217
56	1480	301	0 / 0	2 / 0	1229 / 258	20 / 3	158 / 26	15 / 4	56 / 10	238 / 1242	96 / 205
57	1349	253	0 / 0	1 / 1	1139 / 220	18 / 2	126 / 20	23 / 1	42 / 9	253 / 1096	78 / 175
58	1252	216	0 / 0	1 / 0	1045 / 184	24 / 3	140 / 24	10 / 0	32 / 5	216 / 1036	68 / 148
59	1046	184	0 / 0	0 / 1	856 / 158	19 / 2	133 / 14	14 / 2	24 / 7	201 / 845	66 / 118
60	1277	194	0 / 0	1 / 0	1076 / 154	31 / 11	122 / 24	15 / 2	32 / 3	258 / 1019	75 / 119

The above table is for 1K data set. It is provided to illustrate what the data looks like.

This told us that most of the data came from the IND data source. After looking at the actual distributions, it was decided to just lump IND, IND_V and C2Log erg times into the computation for each age.

It was also clear that there was a lot more erg time data for men over 135 lbs than women in the same category. In some case 10 times more.

Looking at outliers

Because of the questions we had about how representative the Concept2 data would be of the entire rowing population, we decided to do an outlier analysis the data at each age. We also hoped to identify elite athletes. The results were surprising.

Data filtered on: Data filtered on Source = IND, IND_V, C2Log, Sex = Male, Weight = Heavy Weights, all data points considered.

Age	Median Erg Time	Q1	Q3	Low Inner Fence (seconds)	Hi Inner Fence (seconds)	Low Outer Fence (seconds)	Hi Outer Fence (seconds)	Count	Inner Outlier Count (Low/High time)	Outer Outlier Count (Low/Hight time)
28	1185.25	1122.1	1261.6	912.85	1470.85	703.6	1680.1	1038	2 / 50	0 / 11
29	1182.1	1122.9	1275.05	894.675	1503.275	666.45	1731.5	1099	2 / 38	0 / 9
30	1184.9	1124.5	1277.2	895.45	1506.25	666.4	1735.3	1214	0 / 41	0 / 11
31	1182.9	1121.4	1270	898.5	1492.9	675.6	1715.8	1359	0 / 53	0 / 13
32	1185	1124.1	1279.8	890.55	1513.35	657	1746.9	1474	0 / 55	0 / 19
33	1182.95	1125.6	1265	916.5	1474.1	707.4	1683.2	1598	0 / 67	0 / 14
34	1188	1128.65	1282.5	897.875	1513.275	667.1	1744.05	1498	0 / 56	0 / 13
35	1182.8	1125.45	1270.25	908.25	1487.45	691.05	1704.65	1801	0 / 75	0 / 11
36	1188.75	1132.1	1284.45	903.575	1512.975	675.05	1741.5	1890	0 / 73	0 / 20
37	1187.9	1126.5	1273	906.75	1492.75	687	1712.5	2024	0 / 65	0 / 6
38	1190.55	1126.55	1279.6	896.975	1509.175	667.4	1738.75	2008	0 / 61	0 / 13
39	1192.7	1136.1	1286.2	910.95	1511.35	685.8	1736.5	2105	0 / 67	0 / 12
40	1191.15	1133.25	1275.65	919.65	1489.25	706.05	1702.85	2210	0 / 86	0 / 21
41	1196.7	1133.35	1293.15	893.65	1532.85	653.95	1772.55	2210	1 / 66	0 / 12
42	1193.4	1134.2	1282.45	911.825	1504.825	689.45	1727.2	2314	0 / 73	0 / 17
43	1197.3	1138.3	1292.45	907.075	1523.675	675.85	1754.9	2307	0 / 72	0 / 19
44	1198.95	1137.35	1298.9	895.025	1541.225	652.7	1783.55	2216	0 / 72	0 / 20
45	1200.2	1143.1	1301.4	905.65	1538.85	668.2	1776.3	2151	0 / 70	0 / 11

46	1200	1142.15	1300.25	905	1537.4	667.85	1774.55	2179	0 / 67	0 / 16
47	1208	1152.7	1301.9	928.9	1525.7	705.1	1749.5	2094	0 / 87	0 / 16
48	1214.9	1151.8	1302.9	925.15	1529.55	698.5	1756.2	1974	0 / 70	0 / 17
49	1221	1159.9	1314.8	927.55	1547.15	695.2	1779.5	1774	0 / 61	0 / 13
50	1203.6	1150.1	1291.9	937.4	1504.6	724.7	1717.3	1974	2 / 77	0 / 13
51	1223.9	1157.9	1316.2	920.45	1553.65	683	1791.1	1748	0 / 67	0 / 17
52	1222.05	1156.5	1317	915.75	1557.75	675	1798.5	1678	0 / 62	0 / 11
53	1228.65	1164.35	1312	942.875	1533.475	721.4	1754.95	1594	0 / 73	0 / 12
54	1245	1173.95	1325.3	946.925	1552.325	719.9	1779.35	1483	0 / 61	0 / 12
55	1237.2	1176.7	1315.1	969.1	1522.7	761.5	1730.3	1387	0 / 74	0 / 10
56	1253.9	1178.4	1338.1	938.85	1577.65	699.3	1817.2	1182	0 / 52	0 / 13
57	1255.8	1179.5	1350.55	922.925	1607.125	666.35	1863.7	1043	0 / 40	0 / 8
58	1271.65	1193.1	1354.7	950.7	1597.1	708.3	1839.5	998	0 / 40	0 / 5
59	1267	1196.6	1370	936.5	1630.1	676.4	1890.2	810	0 / 36	0 / 5
60	1254	1186.1	1343.9	949.4	1580.6	712.7	1817.3	983	0 / 37	0 / 8

The above table is for 1K erg times. We computed both the inner and outer fences and the number of data points within those ranges.

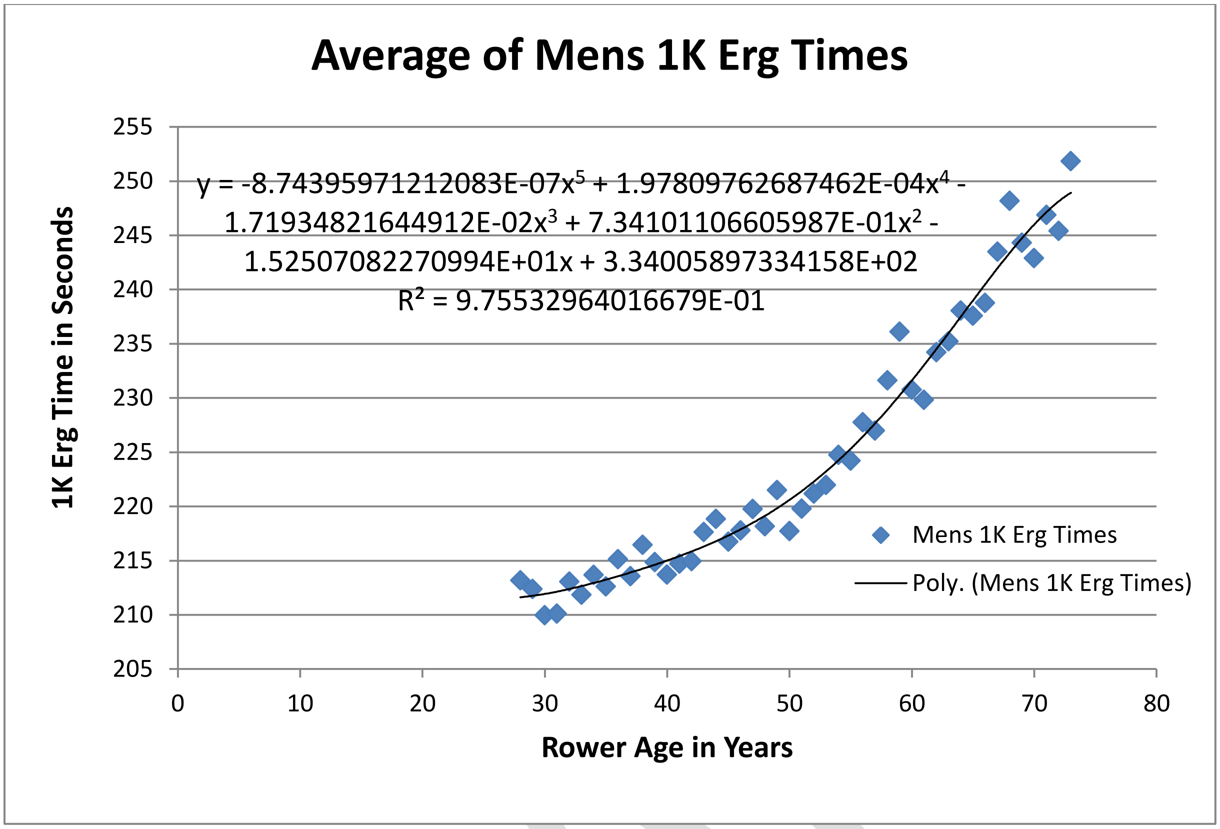
We found that there were very few low outliers (a low erg time means noticeably fast times) and lots of high erg time, unusually slow rower erg times). Since we had so much data, we decided to exclude all outliers from the data analysis where there was sufficient data. This means that the computed average erg time for each age will be faster than the average if all data points were included in the computation.

1000 meter Results

We computed average erg times for each age and put them into Excel where we used their curve fitting equations to derive the regression equation that most closely fit the data.

Men, over 135lbs, all outliers excluded

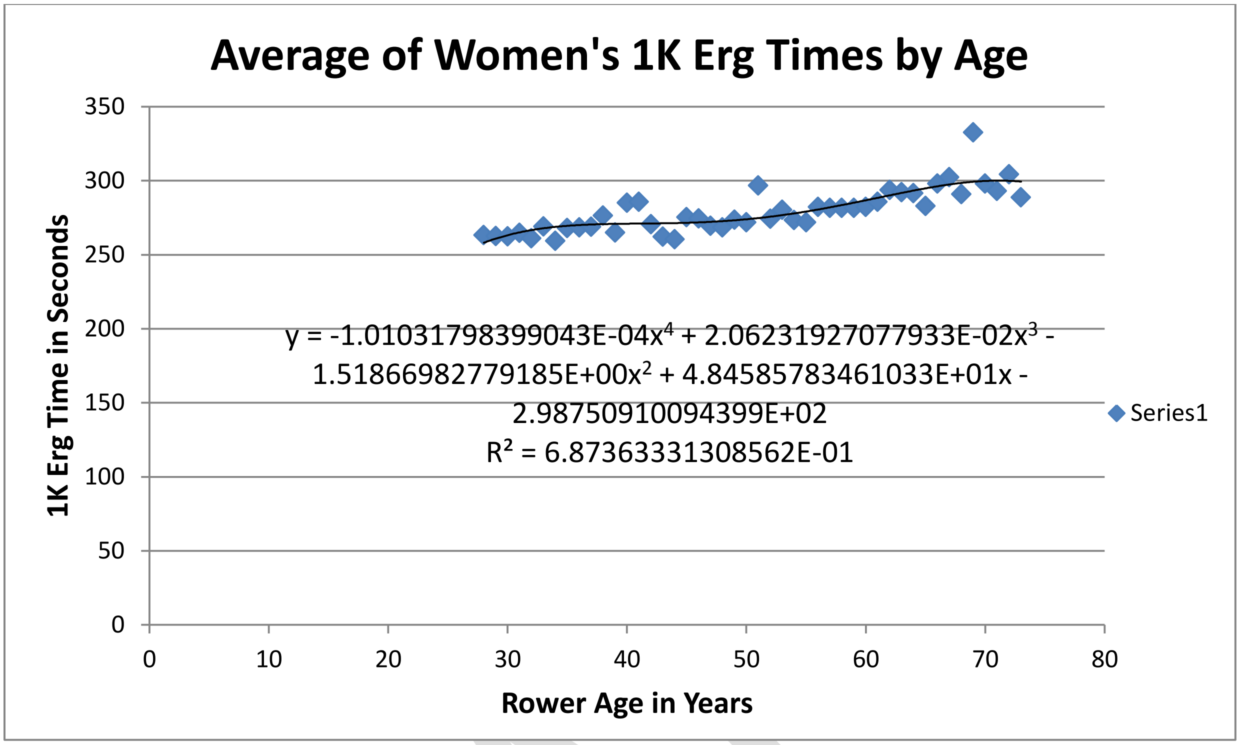
In total 21584 data records were used from the 1K data set provided by Concept2. The equation that was generated is based on average ages up to 73.



Because of the large number of records and the fact that we removed outliers, it is no surprise that the Regression coefficient is 0.9755 out of a possible 1.0. The above equation is a very good fit up to age 73. Since there is little data above 73, it is not recommended that the equation be used above that age.

Women, over 135lbs, all outliers INCLUDED

There simply was not enough data. A total of 4904 data points were available (including outliers) over 47 years, or about 100 data point a year. To be consistent with the men, we used averages up to 74 years of age to generate the regression equation. The result is that the regression coefficient is only 0.687, or a "goodness of fit" rating of about 70%.



Men's lightweight Results

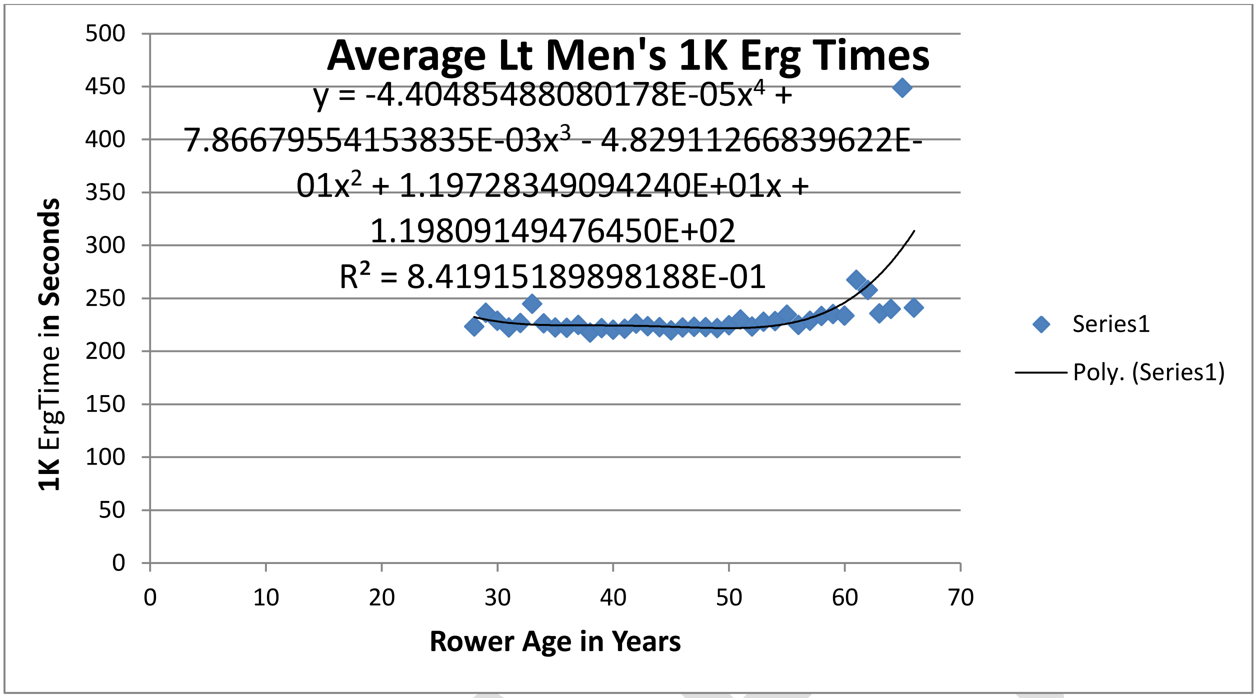
This is actually the most interesting result. Considering all data point, there was marginally enough data to see how erg times for lightweight men changed with age. The equation that we came up with is:

Considering that the regression coefficient is about 85%, it is actually a more reliable equation than the women's equation, if we actually look at the graph:

Age	Avg Erg Time	Median Erg Time	Variance	Std Dev	Count	4th Deg Poly Time	
28	223.1566	217.7	641.7748	25.3333	76	222.0633	
29	236.2821	222.05	7853.702	88.6211	78	221.6016	-0.46172
30	228.689	218.3	1653.24	40.66	91	221.0982	-0.50335
31	222.3125	216.1	744.6076	27.2875	96	220.5693	-0.52896
32	226.6495	216.6	1369.888	37.012	103	220.0296	-0.53962
33	244.7021	216	54808.65	234.1125	95	219.4933	-0.53637
34	226.2099	220.9	790.2807	28.1119	111	218.973	-0.52028
35	222.2146	217.8	633.1417	25.1623	96	218.4806	-0.49241
36	222.0652	219	753.2405	27.4452	112	218.0268	-0.4538
37	224.7964	219.3	672.5507	25.9336	111	217.6212	-0.40553
38	217.5703	213.8	381.1715	19.5236	101	217.2726	-0.34864
39	221.7504	214.6	876.7061	29.6092	119	216.9884	-0.2842
40	220.2298	216.9	683.5844	26.1454	141	216.7751	-0.21325

41	220.9542	212.8	864.6896	29.4056	131	216.6383	-0.13686
42	225.9288	214.2	3487.786	59.0575	156	216.5822	-0.05609
43	223.3765	217.1	1554.251	39.424	132	216.6102	0.028013
44	222.5169	214.8	712.74	26.6972	160	216.7246	0.114386
45	219.6227	213.85	1202.177	34.6724	150	216.9266	0.201974
46	222.2104	216.55	677.7909	26.0344	134	217.2163	0.289718
47	222.8726	215.75	676.7823	26.015	124	217.5928	0.376562
48	222.6424	218.6	470.5729	21.6927	118	218.0543	0.461449
49	221.7967	216.65	443.3665	21.0563	92	218.5976	0.543321
50	224.5504	216.5	1186.279	34.4424	137	219.2187	0.621122
51	229.9504	220.4	1849.203	43.0024	121	219.9125	0.693793
52	223.2113	219.7	561.6226	23.6986	141	220.6728	0.760279
53	227.8017	219.95	743.5532	27.2682	118	221.4923	0.819521
54	228.3454	221.1	744.6171	27.2877	108	222.3628	0.870463
55	234.3078	219.5	5279.505	72.6602	103	223.2748	0.912047
56	224.8441	218.4	586.0958	24.2094	93	224.2181	0.943217
57	228.6436	223	643.4593	25.3665	101	225.181	0.962914
58	233.064	227.8	947.6182	30.7834	89	226.1511	0.970083
59	235.0915	230.15	774.7526	27.8344	82	227.1147	0.963665
60	233.4481	227.25	767.5044	27.7039	106	228.0573	0.942604
61	267.4426	230.4	90287.85	300.4794	115	228.9632	0.905843
62	257.789	230.1	35633.93	188.7695	73	229.8155	0.852324
63	235.7259	228	667.0749	25.8278	81	230.5965	0.78099
64	239.7295	228.7	1670.147	40.8674	78	231.2873	0.690783
65	448.6531	232.9	2774112	1665.567	64	231.8679	0.580648
66	241.12	233.5	1371.022	37.0273	65	232.3174	0.449526

Looking at the last column we can see that up to about age 43, there should not be a handicap and after that, there is only a 1 second difference between ages 43 and 58. The equation starts to fall apart above 58 because of the lack of data



It is obvious that performance is pretty flat to about 60. In this case, we will actually show the table:

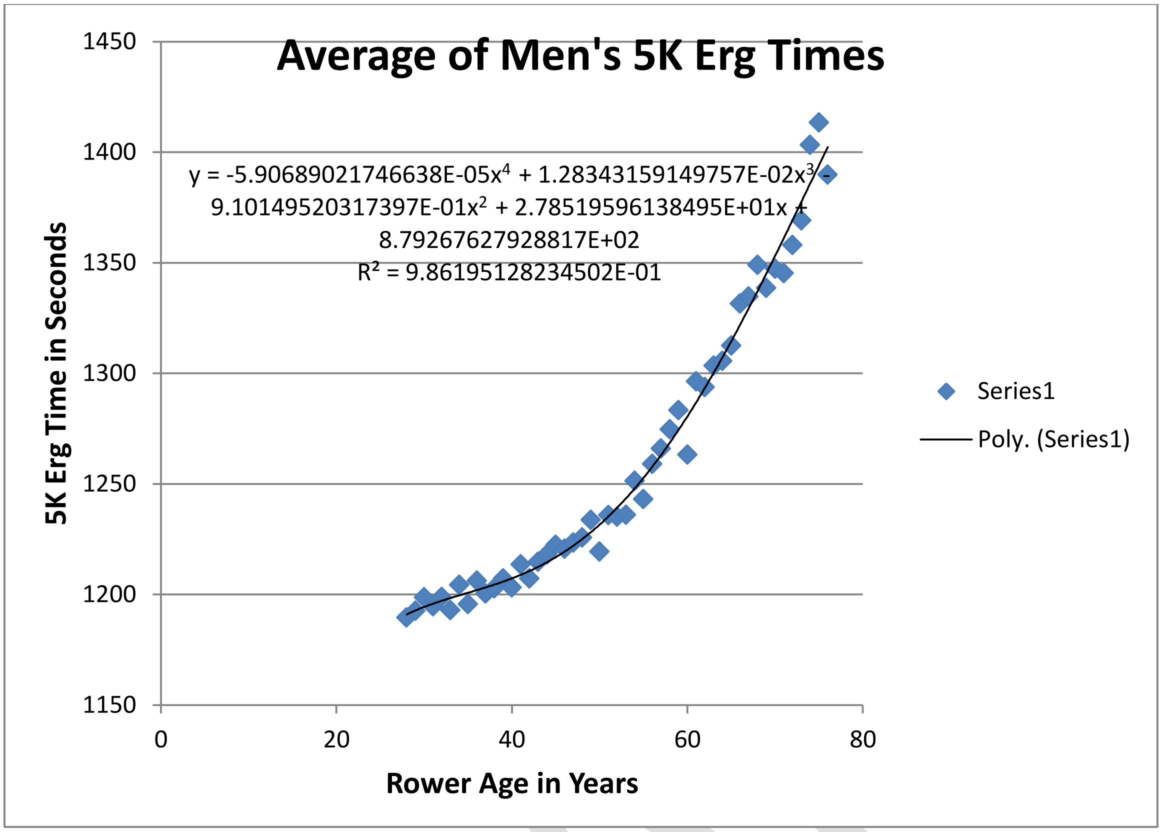
5000 meter Results

One of the primary reasons for doing the analysis was not really for the 1K sprint handicaps, but to get an idea of what a 5K handicap might look like. Any rower knows that they will pull a different average split for a 1K distance than for a 5K distance.

5K Men's Results

Based on the data provided by Concept2, there is no problem with having enough data. After removing outliers, we still had about 58,200 data point! The equation looks like:

With a goodness of fit in the range of 99%, the results are about as good as one can hope for. The Graph looks like:

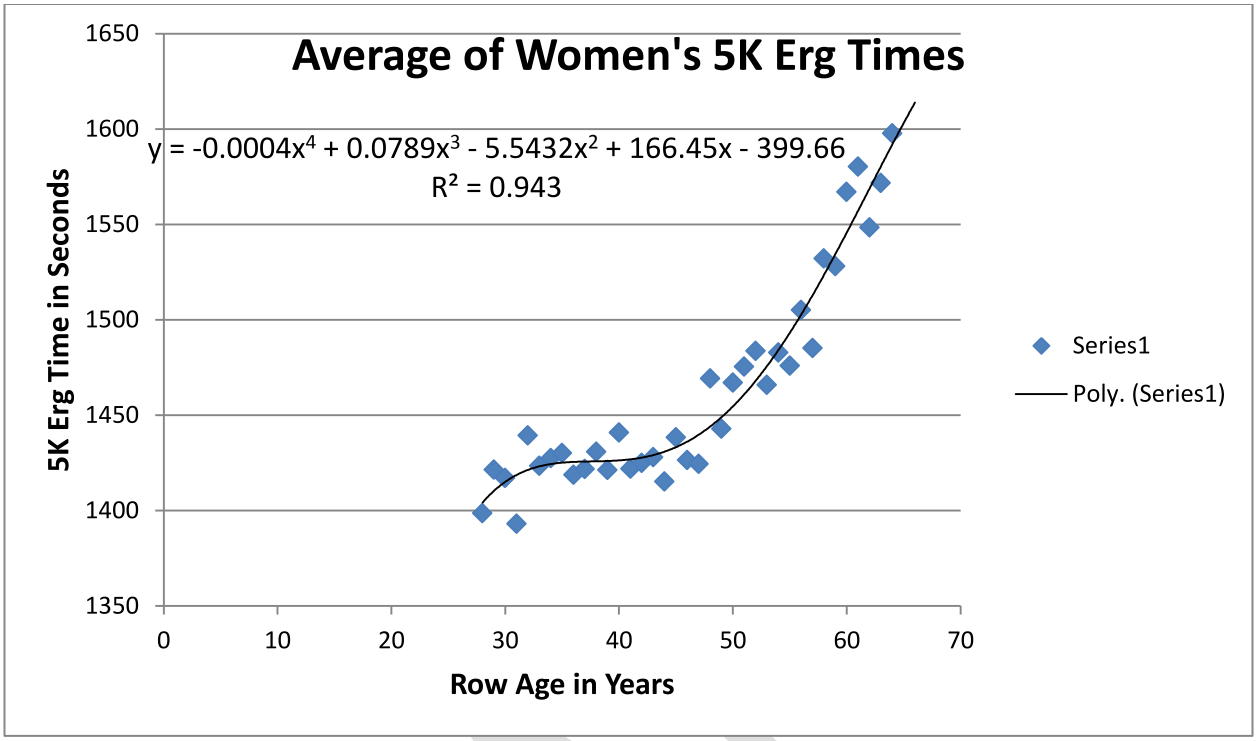


The data spans from 27 to 76. Don't rely on results for ages greater than 76, these polynomials do funny things outside of their limits.

5K meter Women's Results

The numbers for women's 5K results are better than their 1K results. This is due to the fact that excluding outliers, we still had 8900 data points spanning ages 27 to 66.

With an R squared of about 94%, it is a pretty good equation. Since both the women's and men's equation are polynomials, they fit pretty well within the range of ages. It might not be a good idea to use it to estimate handicaps over age 64. The graph looks like:



We don't know what to make of the steep rise from age 27 to about 33 then a leveling off from ages 33 to about 46. After 46, there seems to be a pretty quick decline in women's average 5K erg times.

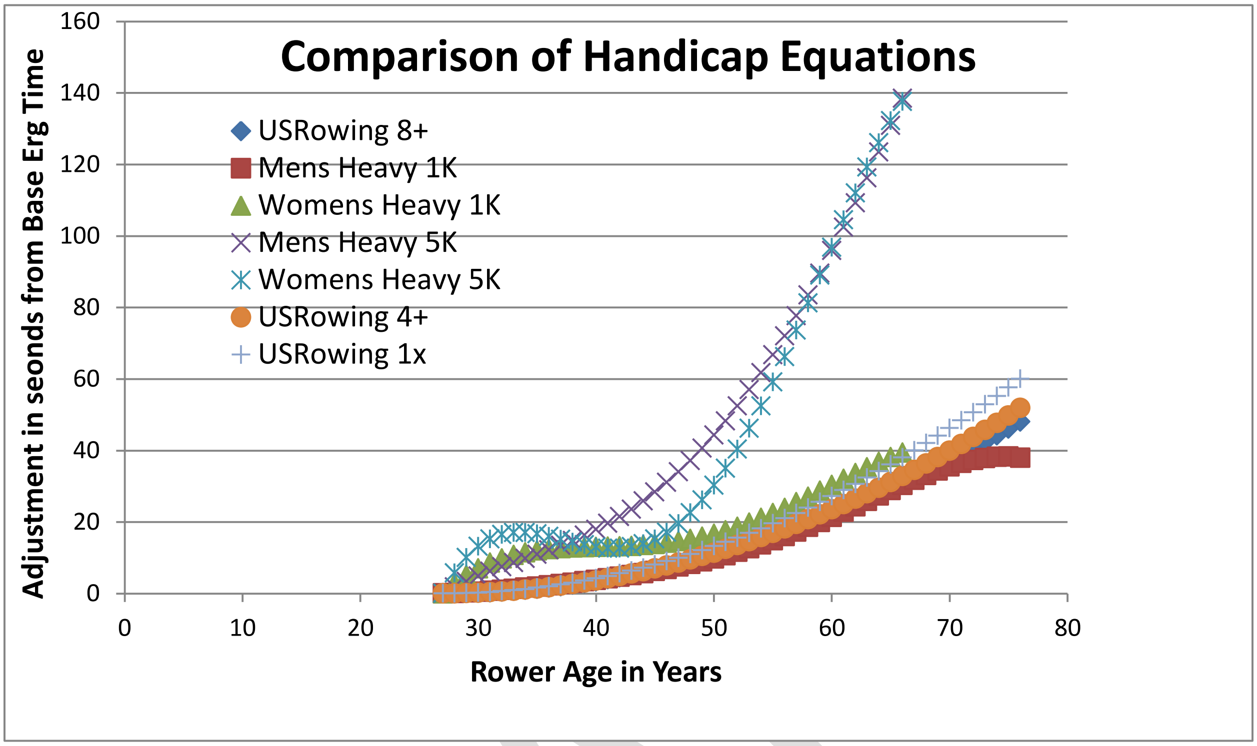
Great, how does it compare with US Rowing's Equation?

The US Rowing equation actually, compares very well with the actual data we compared against. We were surprised at the similarity between the results and US Rowing's equation when one looks at the comparison graph.

There are differences in how the US Rowing equation is applied, but looking at the rate of change (slope) of the US Rowing equation and comparing against the rate of change of the above polynomials (1K), we think they are within the confidence level of the data. That is certainly the case of a direct comparison between the US Rowing 8+ and the men's heavy 1K slopes. For one, the US Rowing equation makes a distinction based on boat, hence they use the same power formula and change the coefficient based on eights, fours, doubles and singles, whereas our approach looks solely at the individual effort. Their equation would be applied to races of any distance (we assume) whereas we have looked at individual performance at different distances.

The benefit of using polynomials for regression instead of using a power function (as US Rowing) is that polynomials are able to capture differences in sub-ranges, for example in the women's 5K graph see following graph), where there appears to be a leveling off between ages 34 and 44. A power function would effectively take an average rate of change through the same range.

If one were to plot the different equations against the US Rowing equations it would look like:



What we see is that the US Rowing 8+, 4+ and 1+ are essentially the same power curve with different coefficients, the all follow the same power shape. The men's 1K curve for all practical purposes is the same as the US Rowing 8+ curve. The US Rowing 8+ and 4+ curve misses the steep increase in women's 1K erg times from age 27 to about 34, then the flattening out to about age 45 before increasing at essentially the same rate as the US Rowing curve.

> Note: Looking at the graph can be a little misleading. It is more important to look at the slope (rate of change) at any point compared with the slope of the other lines than it is to look at the where the points lie relative to one another. The X axis is age and the Y axis is time in seconds. We are not interested in showing that women have larger erg times than men at any age, we knew that. We are looking at the rate of change between age erg times each year.

Where the US Rowing equation really differs is in how quickly the slope changes (increases) with age. The data indicates that for the longer distance (5K), the handicap adjustment increases much more dramatically with each year. We think this is very interest, but not at all surprising. US Rowing never said (at least we never saw any documentation to the effect) that their equation was for both sprint and head races.

Again, what is particularly interesting is the shape of the women's 5K curve. It initially rise fairly steeply then levels off (actually turns negative) between approximately ages 34 to 44, then rises steeply. There isn't even a hint of that in the men's data.

Thoughts about that

To be honest, even though the R Squared for the equations are fairly high, we still don't know, or have a way of determining if the data provided by Concept2 is really representative of the population of rowers, it could be that only rowers (men and women) with certain characteristics are inclined to provide erg data. On the other hand, it could be real and women between 34 and 44 are able to sustain a consistent level of fitness that men can't. We don't know.

If the data is really a self-selected subset of the population, then one could surmise that if the entire population of rowers were surveyed, there would be slower times for each age group. We could also surmise that the rate of change for the equation slopes would be steeper with age, meaning that in the general population, as people get older, there would be a greater percent that get more out of shape than at a younger age. Or maybe the opposite, as rowers get older (and more out of shape) they simply stop rowing, and those that continue will still row slower, but at the same rate as younger rowers who are still rowing. We don't know.