Senior Swim Competition Times

David P. Doane[1], Kevin Murphy[1]
[1]Oakland University, 2200 N. Squirrel Road, Rochester, MI 48309-4401

Abstract
This paper utilizes 565 observations on 500-yard freestyle swim times in the biennial U.S. National Senior Games (ages 50 and up) in three successive competitions (2009, 2011, 2013). Five topics are discussed: (1) What is the relationship between age and swim times, and how should this be modeled? (2) Do men and women exhibit the same patterns of decline by age? (3) Do the years differ, perhaps as a result of stiffening competition? (4) What is the pattern of split times (i.e., in each of the 10 laps)? (5) Do starting platform reaction times vary by age or gender?

Three models are proposed for the time-age relationship: quadratic, semi-log, and age category binaries. We obtain parameter estimates using OLS and quantile regression (25%, 50%, 75%). For a “typical” swimmer, the most realistic predictions are from the semi-log model using quantile regression with bootstrap standard errors (there are high extremes). Women’s average times are slower than men’s, and may show a somewhat steeper decline with age. There is evidence that women’s times have improved in successive biennial competitions, but not men’s. Split times are faster in the first and last laps of the race. Reaction times do not differ by gender, and only slightly by age.

Key Words: Senior Athletes, Quantile Regression, Logarithmic Model

1. Introduction

The National Senior Games Association (NSGA) sponsors competitions during odd years (e.g., 2009, 2011, 2013). To participate in the NSGA Nationals, one must first qualify through NSGA State Game during the preceding even year (e.g., 2008, 2010, 2012). Participants must be at least 50 years old during the qualifying year in the state where they live, or any state that allows out-of-state competitors. Summer Games medal sports include archery, badminton, basketball, bowling, cycling, golf, horseshoes, pickleball, race walk, racquetball, road race, shuffleboard, softball, swimming, table tennis, tennis, track and field, triathlon, and volleyball. In most sports (including swimming) the top 4 finishers in each age group qualify for Nationals. Competition is by age bracket in 5-year intervals (50-54, 55-59, 60-64, ...). This investigation examines swim times in the 500-yard freestyle (25-yard, short course) event in the Summer National Senior Games (ages 50 and up) in three successive biennial competitions (2009, 2011, 2013).

Because this is more of an endurance event, rather than a sprint, individual performance is less affected by random variation and is more reflective of stamina and training. On average, performance is expected to decline with age, but how should this be modeled and estimated? Do men and women show the same patterns of decline? We explore alternative model specifications and estimation techniques, and create benchmarks for swimmers of either gender to compare their times against recent NSGA competitors. Regarding effects of age and gender, our findings are broadly consistent with existing studies of elite competitors in the Olympics and U.S. Master’s Swimming competitors.

However, because our goals, methods, and database are different, our results should not be generalized beyond the biennial NSGA competitions.

Aside from effects of age and gender, our rich data set permits us to study additional questions. (1) Does performance change over time, perhaps as a result of stiffening competition? (2) What is the pattern of split times (i.e., in each of the 10 laps)? (3) Do age and gender affect starting platform reaction times (time from the starting buzzer to when the swimmer's weight actually leaves the platform)? (4) How well do qualifying seed times predict actual meet times?

1. Data

1.1 Variables

Our data set has 565 observations (248 men, 317 women) on 20 variables. Our data set and documentation are accessible on from the *Journal of Statistics Education* archives "Data Sets and Stories" (see the direct web links at the end of this paper).

Obs	observation in sorted list (year, gender, age group, place)
Place	finish order in age group for that year
State	participant's qualifying state
Gender	0 = male, 1 = female
Age	age in years
Age2	age squared (for non-linearity tests)
AgeGrp	age group (5-year bins 50-54, 55-59, etc)
Year	competition year (2009, 2011, 2013)
Seed	qualifying time prior to national competition
Time	time in race (seconds)
Split	time (seconds) in each 50 yard split (ten data columns)

1.2 Data Accuracy

While data accuracy is generally high, a few issues exist. For example, some split times are missing. The 500-yard race consists of ten 50-yard laps (20 lengths of 25 yards). Each 50-yard lap is a *split*. The swimmer touches an electronic pad at the end of each lap. Occasionally, a swimmer touches with hands above the pad or pushes off so lightly that the touch is not recorded. This is more common with older swimmers, who also may avoid flip turns. In addition, electronic touchpads may vary in their sensitivity, may be imperfectly calibrated, or may have "dead spots." Human spotters will report if the swimmer actually fails to touch, which would disqualify the time and hence would not be part of our data. If only one touch is unrecorded, we estimate the missing split time as half the time between the adjacent touches. Otherwise, a split time is recorded as missing.

As a further accuracy check, the ten split times should sum to the total time. If not, it was sometimes possible to reconcile the difference by examining raw timer data. We had access to some detailed data provided by individual participants, which allowed us to retain almost all of the discrepant observations.

Another problem is high extremes. While there is a physiological lower limit on swim times, there is no upper limit. For example, a swimmer may suffer cramps or may simply need to slow down. Senior swimmers are philosophical about their limitations, and are more willing to "back off" than a younger athlete might be.

The age distribution of participants in our database is shown in Figure 1. Male and female swimmers have similar patterns, with two modes (ages 60-64 and 70-74). As would be expected, there are fewer participants in higher age categories.

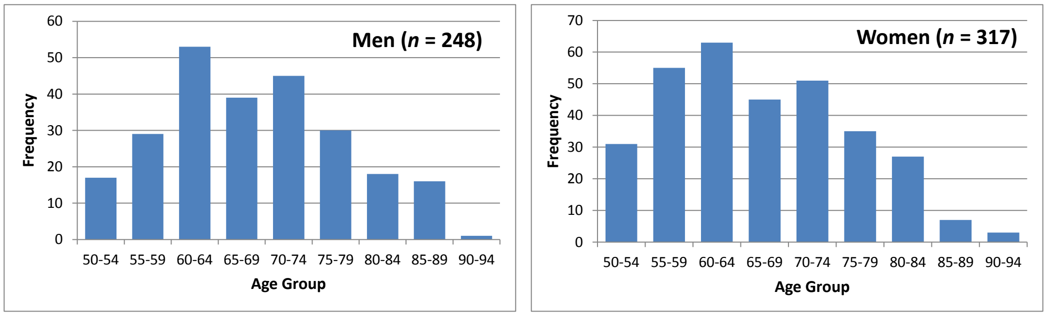

Figure 1: Distribution of swimmers by age group

One lurking concern is selection bias. Many eligible swimmers (the top four per state) decline to participate in the nationals. This is partly a financial matter. The cost of hotel, meals, and travel to the host city (2009 San Francisco, 2011 Houston, 2013 Cleveland) can be daunting. The NSGA summer games last for two weeks, and many athletes compete in multiple events (up to six). Health problems may force an athlete to drop out, or to focus on the events in which she/he has the best chance. Some may judge their medal chances are too low to justify the trip. Some states have well-organized senior swim teams that encourage and support participation, while others have none.

2. Age Group Binaries

Our first model defines age group dummy variables (a1 = 50-54, a2 = 55-59, ... , a9 = 90-94). This has the advantage of avoiding a specific model form. We omit a1 so that age group 50-54 is the base. We define dummy variables for the years (d9 = 2009, d11 = 2011, d13 = 2013) and omit d9, so that 2009 is the base year. We estimate regressions separately for each gender, first with OLS and then with quantile (median) regression (bootstrap standard errors) as shown in Table 1.

Table 1: Estimated Age Binary Regression Coefficients

	OLS Regression Coefficients				*Median Coefficients*			
Predictor	*Men*	*t*	*Women*	*t*	*Men*	*t*	*Women*	*t*
a2	-26.0	-0.80	24.6	0.93	-9.8	-0.49	33.0	1.82
a3	36.2	1.22	98.9	3.85	46.6	1.89	94.2	6.78
a4	62.2	2.01	161.2	5.92	82.1	3.63	164.5	4.58
a5	99.7	3.29	207.8	7.80	123.4	5.40	185.8	7.48
a6	151.1	4.69	262.2	9.07	177.9	7.17	262.5	9.44
a7	191.5	5.29	347.7	11.31	241.7	8.42	346.4	8.43
a8	389.5	10.52	595.0	12.10	408.0	5.13	577.8	11.43
a9	535.2	4.87	590.7	8.38	586.9	2.28	614.9	12.29
d11	22.4	1.33	-29.9	-1.90	2.4	0.13	-25.5	-1.55
d13	-6.3	-0.38	-40.7	-2.44	-7.9	-0.67	-22.2	-1.39
_cons	413.0	15.38	470.7	20.92	381.3	24.68	447.8	29.49

Except for a2, the age binary coefficients are positive, significant, and increasing with age as would be expected. The a2 coefficient for men is effectively zero, suggesting that times for age 50-54 (a1) are the same as for age 55-59 (a2). For women, coefficients for the year binaries suggest that swim times averaged 20-40 seconds faster in 2011 and 2013 than in 2009. For men, 2011 times may have been slower than in 2009 and 2013. Because 2013 is more typical for both genders, we will use 2013 to illustrate the predictions for each model.

Figure 2 shows predicted times for 2013. Expected times deteriorate more than linearly. For either gender, the OLS prediction (conditional mean) is usually higher than the 50% quantile prediction (conditional median) because median coefficient estimates are less affected by high extremes. While the differences for between OLS and 50% quantile may appear small on this scale, they may represent a full pool length in some age groups.

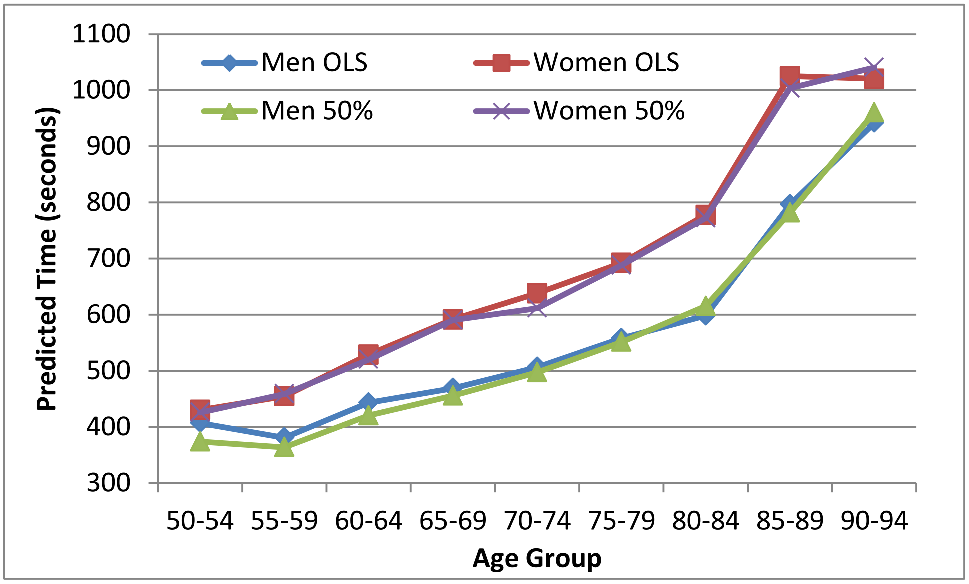

Figure 2: Predictions for 2013 Using Age Binaries

3. Quadratic Age

Next, we specify the model as $Time$ = f(Age, Age^2, d_{11}, d_{13}). We estimate the genders separately, again using OLS and quantile (median) regression, as summarized in Table 2.

Table 2: Estimated Quadratic Regression Coefficients

	OLS Regression Coefficients				*Median Coefficients*			
Predictor	*Men*	*t*	*Women*	*t*	*Men*	*t*	*Women*	*t*
age	-35.12	-3.71	-15.68	-1.80	-27.57	-2.54	-22.78	-2.83
age2	0.3260	4.800	0.2139	3.35	0.2732	3.23	0.2646	4.51
d11	24.28	1.46	-31.73	-2.06	11.83	0.93	-35.14	-2.11
d13	-9.46	-0.58	-36.76	-2.28	-8.76	-0.75	-27.90	-1.89
_cons	1355.3	4.16	705.7	2.40	1072.4	3.17	928.8	3.41

The year binaries d11 and d13 suggest 2013 as a reasonable basis for predictions (2009 was slower for women and 2011 was perhaps slower for men). Non-linearity is evident in the 2013 quadratic predictions (Figure 3), although for men between 50 and 60 the predictions are nearly flat. Women's times deteriorate at a steadier rate after age 50. Predicted times (OLS) exceed predicted median times (50% quantile), although for women they are strikingly similar. For the "typical" swimmer, the 50% predictions might be a better guide to the competition.

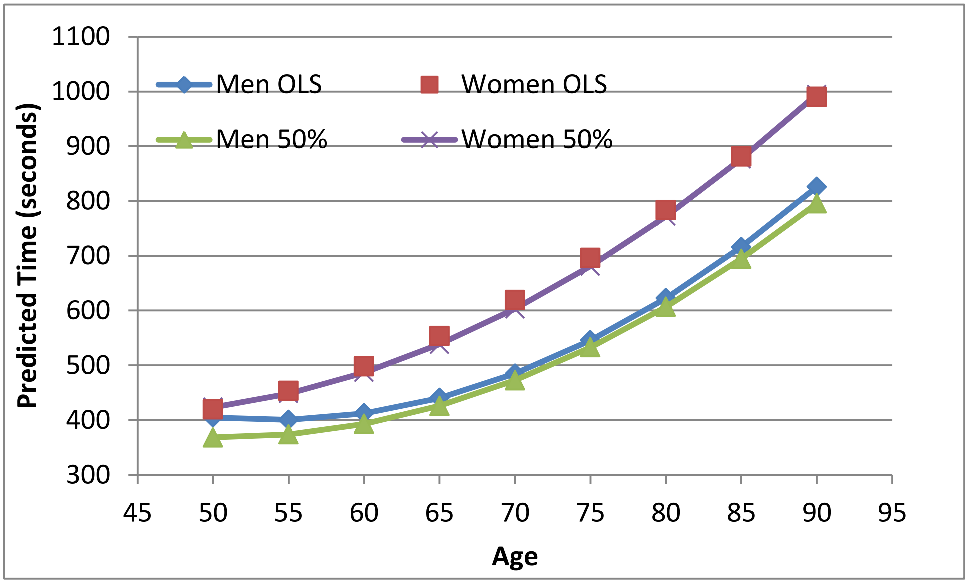

Figure 3: Predictions for 2013 Using Quadratic Model

4. Semi-Log Age

Next, we specify the model as ln(*Time*) = f(*Age*, d_{11}, d_{13}). We estimate the genders separately, again using OLS and quantile (median) regression, as summarized in Table 3. The year coefficients support the previous conclusions (2009 was slower for women, 2011 was slower for men).

Table 3: Estimated Semi-Log Regression Coefficients – Separate Genders

	OLS Regression Coefficients				*Median Regression Coefficients*			
Predictor	*Men*	*t*	*Women*	*t*	*Men*	*t*	*Women*	*t*
age	0.01860	15.25	0.02134	20.67	0.01954	15.14	0.02239	15.92
d11	0.04239	1.49	-0.05909	-2.51	0.01289	0.36	-0.05731	-1.71
d13	-0.01467	-0.53	-0.06480	-2.63	-0.01627	-0.54	-0.05499	-1.61
_cons	4.92325	57.78	4.99093	71.83	4.84114	48.83	4.90022	52.06

As in the previous models, the 2013 semi-log predictions (Figure 4) show that swim times deteriorate non-linearly with age, although not as steeply as the quadratic model. While not apparent in this plot, predicted mean times (OLS) are slower than predicted median times (50% quantile) except for the oldest women. The differences, however, are small (-12 to +12 seconds). A "typical" swimmer would regard the median times as a reasonable guide to the competition. An attraction of the semi-log model is its conformance to what one might expect based on human physiology (i.e., steady deterioration with age past 50). However, unlike the quadratic model, the log model forces a steady increase in times (it allows no point of inflection).We also estimated a semi-log model for combined genders, using a gender binary and an interaction term: ln(*Time*) = f(*Age*, *Gender*, *Age***Gender*, d_{11}, d_{13}). While the results were similar, we feel that estimating each gender separately makes it easier to explain.

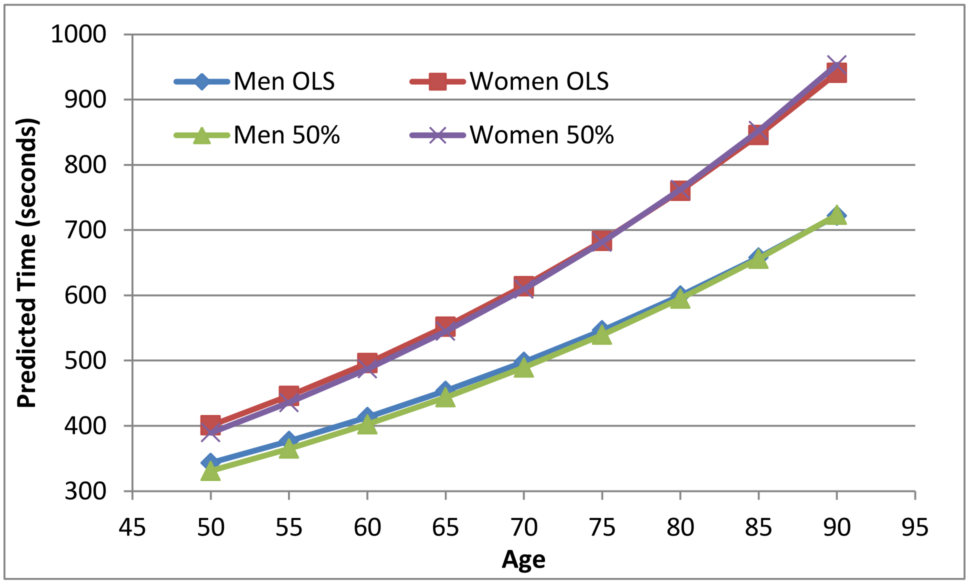

Figure 4: Predictions for 2013 Using Semi-Log Model

We then proceeded to estimate 25%, 50%, and 75% quantiles. In Figure 5, the resulting quantile predictions for 2013 are plotted on the entire 2009-2013 data set for each gender. For the men, we re-estimated the model after omitting one high outlier (five standard errors). 2013 predictions are slightly optimistic because 2009 times (women) and 2011 times (men) were a bit slower. Given the recent number of competitors in each age group, a swimmer in the fastest 25 percent would have a reasonable chance of placing in the top 8, thereby qualifying for recognition on the winner's dais (but only the top 3 receive medals).

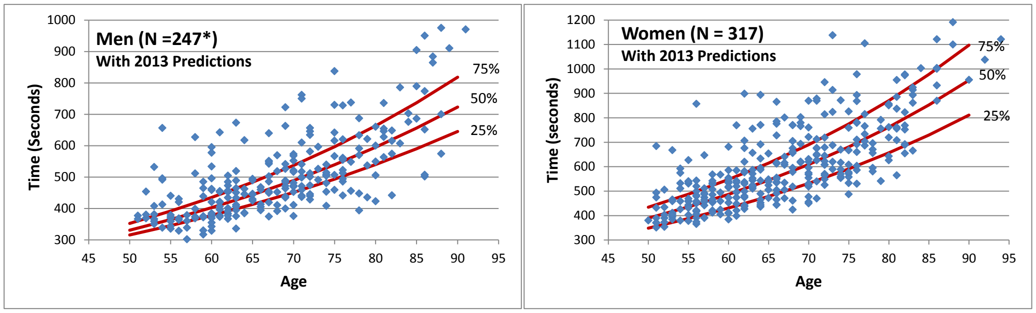

Figure 5: Quantile Regression Predictions for 2013 (*one outlier omitted)

A swimmer who plans to compete in NSGA Summer Games can assess his/her relative standing against competitors in recent biennial NSGA competitions using the estimated 2013 models shown here. We plan to assess these predictions when results become available from the 2015 NSGA games in Minneapolis/St. Paul (July 3-16). It will be interesting to see whether we can predict who would likely earn a medal.

Men:

25% quantile:	*Time* = exp(4.8610816 +0.0178765 *Age*)
50% quantile:	*Time* = exp(4.8248709 +0.0195431 *Age*)
75% quantile:	*Time* = exp(4.7931116 +0.0213336 *Age*)

Women:

25% quantile:	*Time* = exp(4.8135287 +0.0210377 *Age*)
50% quantile:	*Time* = exp(4.8452264 +0.0223859 *Age*)
75% quantile*:	*Time* = exp(4.9148751 +0.0231770 *Age*)

*75% quantile was re-estimated omitting one extreme outlier

5. Comparison with Existing Research

The effects of age and gender on athletic performance have been analyzed extensively, including swim times, both in cross-sectional and longitudinal studies. In all these studies, data become sparse toward the highest ages. Swimming studies have utilized results from U.S. Master's Swimming competitions (Rubin and Rahe 2010, 2013) and focus on best times by elite swimmer (e.g., Fairbrother 2007; Donato *et al* 2003; Konig *et al* 2014). Studies of the 1500 m event (e.g., Tanaka and Seals, 1997; Fairbrother, 2007) to exemplify an endurance race are most comparable to ours. A quadratic regression model is often used to capture nonlinearity, though we also see the linear model (e.g., Rahe and Arthur, 1975), semi-log model (e.g., Rubin *et al* 2013), and correlation analysis (e.g., Konig *et al* 2014). Attention in is often given to age 70 as a perceived discontinuity or "break point" past which times deteriorate non-linearly.

Our results cannot easily be compared with other studies, given differences in ages covered, race length, and data sources. But our scatter plots, fitted functions, and gender relationships do resemble those in previous studies, although we did not find age 70 to be a clear discontinuity. We feel that our use of quantile regression adds a useful new perspective for other researchers, and our NSGA data provide useful insights into a broader cross section of senior swimmers.

6. Split Times

The first 50-yard split (denoted S1 in Figure 6) is fast because of the dive from the starting platform. An exception would be swimmers who choose (because of age or health issues) to start in the water rather than risk a dive, or who have trouble getting up onto the platform. The second split (S2) usually is also fast because of the "adrenalin" factor at the race's start (crowd noise, etc). The middle laps (S3-S9) tend to be "just swimming." The final 50-yard split (denoted S10) typically is faster as the swimmer makes a strong finishing effort. To allow individual comparisons, we standardized each swimmer's split times as ratios to her/his average (i.e., *Time*/10 where *Time* is the swimmer's overall time in the 500-yard race). Figure 6 shows standardized split times for 100 randomly-chosen swimmers (both genders) that illustrate these general patterns. The reference point 1.00 would be an "average" split time for that swimmer.

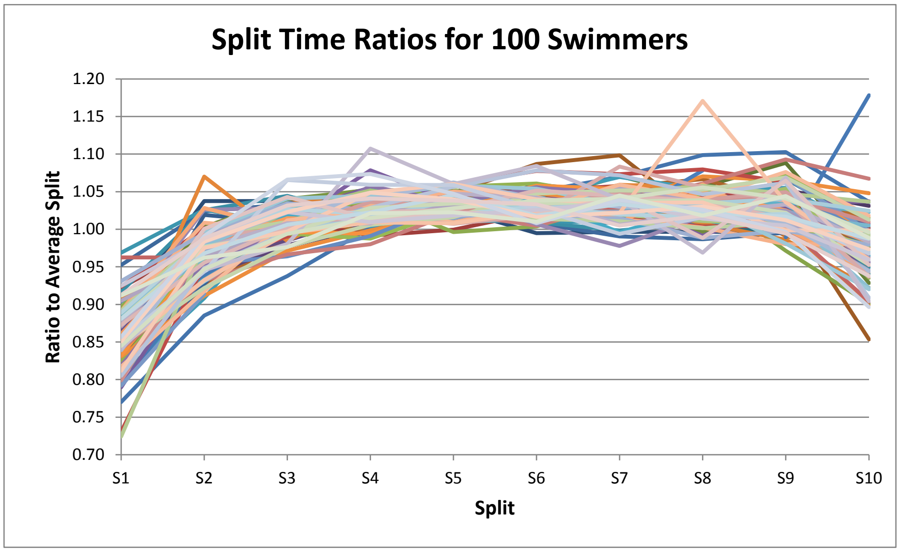

Figure 6: Split Time Ratios for 100 Swimmers

To see the patterns more clearly, we take averages over all swimmers (by gender). Women's and men's patterns are almost the same. Omitting unusual split times (there are a few examples in Figure 6) had little effect. Good strategy requires a swimmer to know her/his capabilities and to regulate the pace. One must balance "type I error" (over-doing it in early laps, then fading) against "type II error" (conserving too much, then unable to catch up). Experienced swimmers are swimming against themselves as much as against others (who may be hard to see anyway).

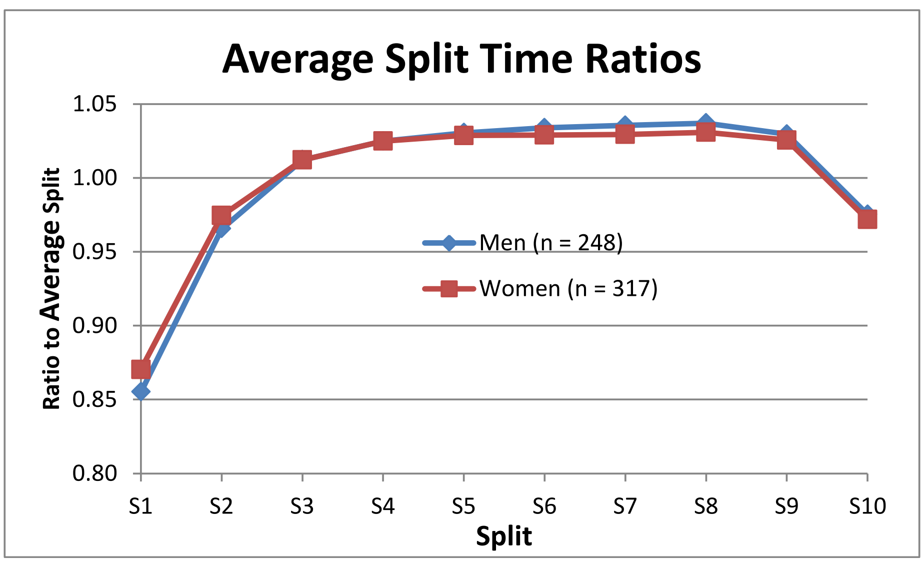

Figure 7: Split Time Averaged Over All Swimmers

7. Reaction Times

We have a small data set (only for 2011) with included the time (seconds) from the sound of the starting buzzer until the time the swimmer's weight leaves the starting platform. Figure 8 shows the distributions of these reaction times. The statistics for women and men are almost identical. For men, the Anderson-Darling test suggests a normal distribution (p = .67) but not for women (p = .049).

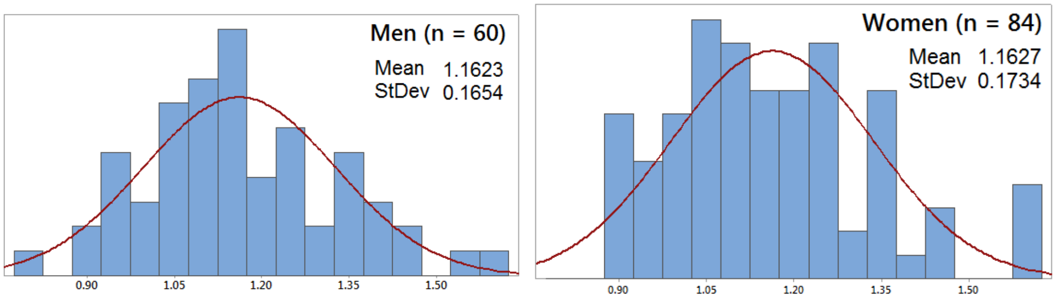

Figure 8: Reaction Times for Starting Platform

Does reaction time deteriorate with age? Not to a noticeable degree, as shown in Figure 9. While the regression has a positive slope, the scatter plot says that the effect is of no practical importance. It is interesting that the oldest competitor (a 94-year old woman) got off the platform quicker than some younger swimmers. In this illustration, both genders are combined because tests showed no significant differences by gender (t-tests of means were insignificant and fitted regressions were almost identical).

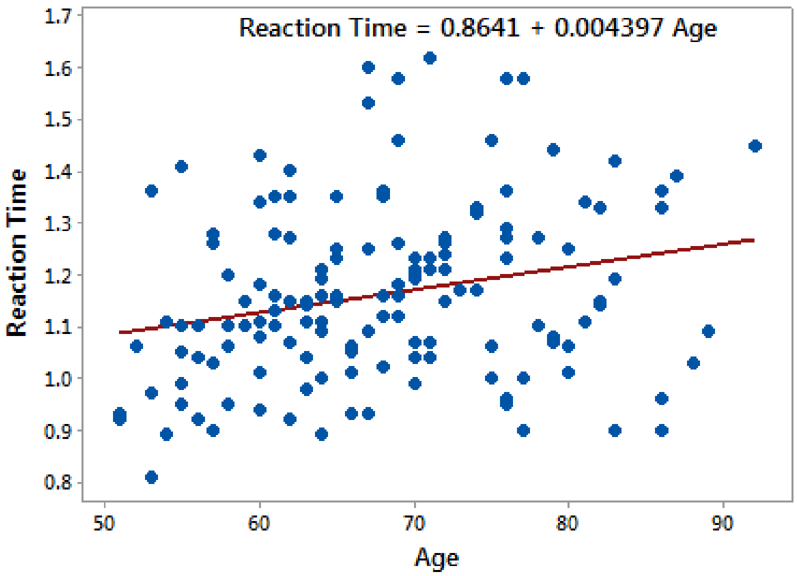

Figure 9: Reaction Times versus Age (both genders combined)

8. Seed Times

One more question that we studied was how well a swimmer's qualifying time ("seed") predicts the actual time. The seed time is usually the best time in state competitions prior to the nationals. Seed times are provided to NSGA competitors prior to the finals in "psych sheets" (this title emphasizes the psychological aspect of competitive swimming). Figure 10 shows that seed time predicts national time very well. However, for both genders, the national competition time is better, on average, than seed time (the green 45-degree line would represent perfect predictions). Presumably, the national competition brings forth a little extra effort in most swimmers.

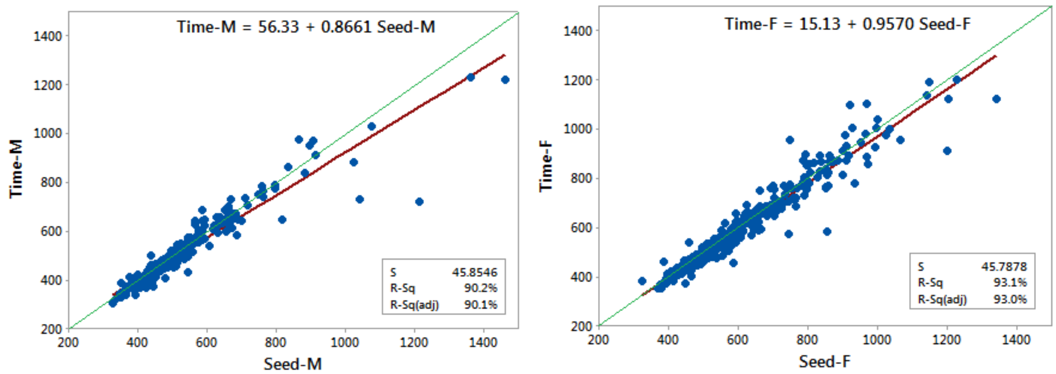

Figure 10: Seed Time as Predictor of Actual TIme

9. Conclusions

While we cannot generalize beyond NSGA competitions, our empirical observations about age, gender, split times, reaction times, and seed times are reasonable *a priori*. At a minimum, our analysis provides a reference point for future empirical research. It will be interesting to see how the 2015 NSGA times (July 3-16) compare with predictions from our quantile regressions based on the 2009-2013 NSGA competitions.

Over time, we speculate that Title IX may reduce the male-female gap in conditional means at all ages, as more women participate in competitive swimming. However, swimming is a sport where equality of opportunity already has a fairly long history. Over time, we would expect tougher competition in *senior* swim meets, as more seniors become aware of formal competitions. Currently, many senior competitors are self-taught. Decades ago, there were fewer high schools with first-rate aquatic facilities, training, and support for varsity travel teams. Competitive swim training nowadays starts as early as young as age 6 (e.g., USA Swimming, YMCA Live Y'ers). When these youths become seniors (over age 50) they are likely to be formidable swimmers.

Acknowledgements

The authors would like to thank the National Senior Games Association for permission to utilize results from the biennial NSGA competitions. Swimmer names have been removed. Quantile regressions were performed using Stata's *reg* and *qreg* procedures with 400 iterations for the bootstrap standard errors. Our data set is posted in the data archive of the *Journal of Statistics Education*. You can download our data set in text form, as well as full documentation.

References

Donato, A. J., Tench, K., Glueck, D. H., Seals, D. R., Eskurza, I., and Tanaka, H (2003), "Declines in physiological functional capacity with age: a longitudinal study in peak swimming performance," *Journal of Applied Physiology*, 94, 764–769.

Fairbrother, J. T. (2007), "Prediction of 1500-m freestyle swimming times for older Masters All-American swimmers," *Experimental Aging Research*, 33, 461–471.

König, S., Valeri, F., Wild, S., Rosemann, T., Rüst, C. A., and Knechtle, B. (2014), "Change of the age and performance of swimmers across World Championships and Olympic Games finals from 1992 to 2013 – a cross-sectional data analysis," SpringerPlus 2014, 3:652, 1-13.

Rahe, R. H. and Arthur, R .J. (1975), “Swim performance decrement over middle life,” *Medicine and Science in Sports*, 7(1), 53–58.

Rubin, R. T., Lin, Sonia, Curtis, Amy, Auerbach, Daniel, and Win, Charlene (2013), “Declines in swimming performance with age: a longitudinal study of Masters swimming champions,” *Open Access Journal of Sports Medicine*, 4, 63-70.

Rubin, R. T. and Rahe, R. H. (2010), “Effects of aging in Masters swimmers: 40-year review and suggestions for optimal health benefits,” *Open Access Journal of Sports Medicine*, 8, 39-44.

Tanaka H. and Seals D. R. (1997), “Age and gender interactions in physiological functional capacity: insight from swimming performance,” *Journal of Applied Physiology*, 82(3), 846–851.