Is There A Racial Difference in Swimming Speed? A Non-Linear Swim Hockey-Stick Mixed-Effects Model

Jenny K. Chen*

Abstract
This paper is to investigate whether or not racial differences affect swimmers' times as they age. For this purpose, swimming times for the 100-meter freestyle event were extracted from the USA Swimming Association and analyzed by a hockey-stick nonlinear mixed-effects model incorporating a transition age whereby swimming speeds start to plateau. We found that the common transition age for both Asians and Non-Asians was around 12 years old. At this age, the estimated times for the 100-meter freestyle in Asian swimmers was 65 seconds while Non-Asian swimmers will have had reached the faster 60 seconds. Before this transition age, the times for an average Asian swimmer dropped at a faster rate of 7.31 seconds per year as compared to 6.77 seconds a year for the average Non-Asian swimmer. After the transition age, the average rate of dropped time per year for Non-Asian swimmers was over twice as great at approximately 0.70 seconds than Asian swimmers at only 0.31 seconds, a considerably lower amount. Thus bearing useful information that dictates that on average, before the transition age, the rate at which Asian swimmers decrease in time is faster, but after the transition age, it is the Non-Asian swimmers who overall swim faster.

Key Words: Hockey-stick model, 100 free-style swim, nonlinear mixed-effects model, p-value.

1. The Question

My swimming career began when I was only five years old. From the very beginning, my performance improved at a rapid pace. It seemed that there were no boundaries for what could be achieved with the sport, as I would soon win my first state title and place first in South Dakota for my age group, a glorious title for a seven year old that would last for another two years. Soon after, my family and I moved to Georgia where I was determined to climb to greater heights and surpass my past accomplishments. Shortly after, at the age of ten, I regained my former state title and became the Short Course State Champion of Georgia. This fluctuating pattern of hard work and reward continued to oscillate as my life progressed, but as I approached fourteen, although I was still placing in the top 15 at state championships, it became clear that I was plateauing. The few years prior showed no extreme observable improvement like that of when I was younger, and circumstantially caused doubt to obscure the path to my goals. Soon, the sport that had propelled me through life and hardship seemed irrevocably gone. Something I had always loved to do, whether as an escape from the stresses of life, or just for fun, was torn away from me. It seemed that no matter how hard I worked, personal improvement and progress was limited. So while my teammates consistently dropped time in their events, I was left to ponder various questions: "Is this the end of my ten year swimming career? Why me?" It seemed that my progress was approaching a limit, and mine seemed to be approaching a fixed value much sooner than any of my other swimming counterparts. Coaches will always say that the harder you work, the faster you'll swim. After all, "practice makes perfect". Unfortunately, this did not seem to be the case for me, and as a result, my enthusiasm for everything in life was noticeably dampened. Something I had always loved to do, competitively, or just as a way to pass time, had been torn away from me, and I needed an answer.

*East Chapel Hill High School, 500 Weaver Dairy Road, Chapel Hill, NC, 27514, jennykc99@gmail.com

It is known that the success of competitive swimming is based off of many different factors such as strength, stamina, technique, and physical composition. Swimming is largely based off of the ability to surpass others in the pool, and some of these factors may limit a swimmer?s ability to do so. For example, a swimmer with less strength may find it much more difficult to outdistance another with a more ample muscular composition.

I already had some thoughts as to why I was beginning to plateau sooner than others. When I was told that there might be a racial difference regarding factors on strength, stamina and etc. it raised a scientific question as to whether or not there really is a racial difference in swimming speeds. There had to be an answer to this question; therefore, we decided to do some quantitative research by using data that was readily available.

This paper is organized as follows. Section 2 will describe the data acquisition, preliminary data analysis to propose a nonlinear hockey-stick swimming model, and the mixed-effects model fitting results. Finally, I will conclude the paper with a discussion in Section 3.

2. A Hockey-Stick Nonlinear Swimming Time Model

2.1 Data Acquisition

The data are assembled from USA Swimming Association, which can be accessed at `http://www.usaswimming.org`.

My swimming career includes the Brookings Swim Club (Brookings, SD), which lasted from 2005-2009, Statesboro Swim Club (Statesboro, Georgia) from 2009-2010, and PACK Swim Team (Pittsford, NY) from 2011-2014.

The extensive data list of my swimming teammates includes 35 Non-Asian swimmers and 39 Asian swimmers. I happen to know all the Non-Asian swimmers that are in this data set, but this wasn't the case for Asian swimmers. There were not nearly enough Asian swimmers in these three clubs that could match the numbers of the Non-Asian; so the only way to acquire more data was to search the USA Swimming Association webpage above for Asian swimmers, and those who had at least three years of experience with the 100-meter Freestyle event had their times added into the data.

In the sport of swimming, besides in special cases, each swimmer competes in multiple events. For myself, the events that are most compatible and where I typically place higher are the 100-meter Fly and the 100-meter Breaststroke. There was no way of knowing the events that everyone liked; but common sense dictated to steer clear of the mile or the 200-meter butterfly, for those are the events that the majority of the swimming population dreads to swim. For the sake of consistency, we chose the one event that most, if not all, swimmers have swum many times in their lives, the short course 100-meter Freestyle.

Among the 74 swimmers, only 1 Non-Asian (i.e. Non-Asians) swimmer had times at ages 6, and ages 18 to 21. Thus, we will restrict the age range from age 7 to 17. Table 1 summarizes the data for the age distribution along with the mean swimming times and the standard deviations. It can be seen from this table that Asian swimmers generally have slower times than the Non-Asian swimmers. The reader should keep in mind that, in competitive swimming, the faster the swimmer, the smaller the time is numerically.

2.2 Preliminary Analysis and Nonlinear Hockey-Stick Swim Time Model

We use R software (http://www.r-.org) for data analysis. To investigate the time speed trend, we use "lattice" library (Sarkar 2008) which can be seen from Figures 1, 2 and 4. Figure 1 is the boxplot by age for all the swimmers in the race category and Figures 2 and 4 are the data with the best model fit which will be explained in following sections.

Table 1: Data Summary for Swimming Times

	Asian Swimmers			Non-Asian Swimmers		
age	Num	Mean	SD	Num	Mean	SD
7	3	110.00	12.17	4	85.50	12.04
8	8	92.88	10.53	12	88.67	16.83
9	17	88.18	11.29	17	79.71	15.92
10	24	80.08	11.17	22	72.94	12.89
11	30	73.27	9.84	24	67.29	10.11
12	30	69.13	8.57	26	63.50	8.11
13	29	64.66	5.54	29	59.34	4.12
14	23	63.61	5.40	26	58.04	3.93
15	11	61.27	5.06	20	57.00	3.49
16	9	61.44	4.56	16	37.31	3.30
17	8	62.50	5.45	15	57.87	3.09

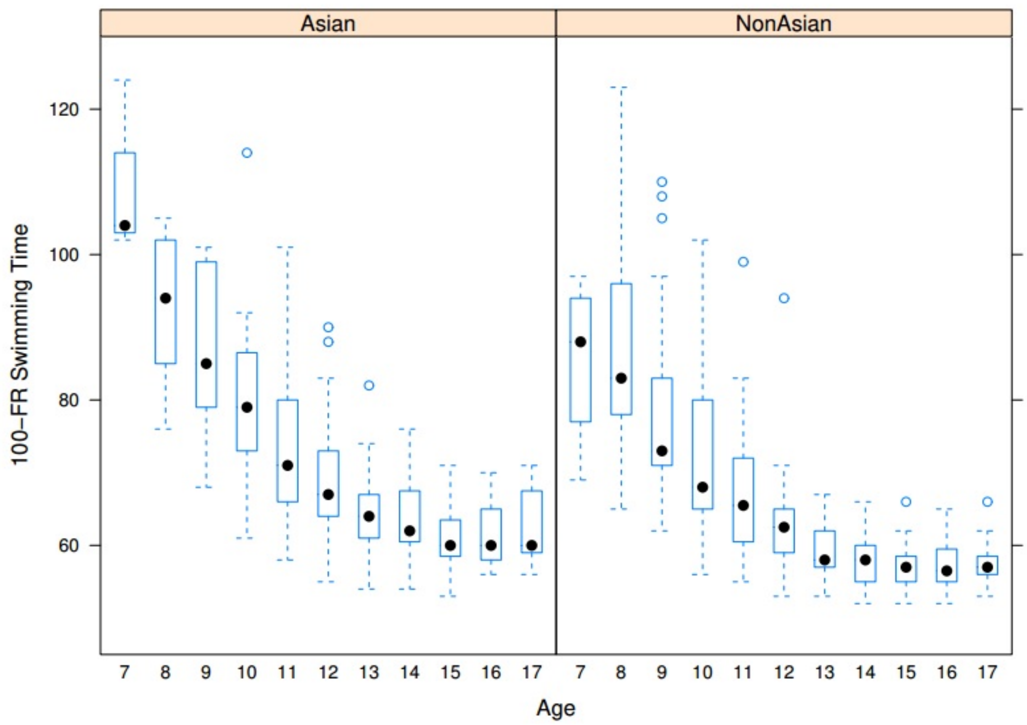

Figure 1: Boxplots for swimming speed for all ages and races.

As can be seen from all these figures, average swimming time will drop rather quickly from age 7 to approximately age 12 and then level-off after.

Many mathematical and statistical models can be used to model this type of data and their trends. To simplify the model fitting, I proposed a 4-parameter nonlinear hockey-stick swimming time model assuming that the swimming time (i.e. T) will drop from the beginning in a linear function with slope of K_1 to a transition age (denoted by "A_0" where the swimming time is denoted by "T_0") and then drop in a slower linear function with slope

K_2. The mathematical form of this nonlinear model can be written as follows:

$$\mathbf{T} = \begin{cases} T_0 + K_1 \times (age - A_0) & \text{if age} \leq A_0 \\ T_0 + K_2 \times (age - A_0) & \text{if age} > A_0 \end{cases} \quad (1)$$

with 4 parameters of A_0, T_0, K_1 and K_2 where A_0 and T_0 denote the transition age (A_0) and the associated swimming time (T_0). The other two parameters of K_1 and K_2 are for the slopes of the two linear functions indicting the swimming improvements. This model can be race-specific and therefore contains 8 parameters to be estimated for both races.

Since this 4-parameter model will be individually different, a statistical non-linear mixed-effects model will be used to fit the model to the data and we will use the R package "nlme" developed by Pinheiro and Bates (Pinheiro and Bates 2004). This model is well-used in sport statistics, for example, Kovalchik and Stefani (2013) used this mixed-effects model to investigate the gender differences in relative performance with the finding that there is no statistically significant average difference in performance improvement for events in running, jumping, throwing, and swimming. In addition, a fixed-effects regression of this mixed-effects model is used in Brander et al. (2014) to estimate the effects of age on NHL player performance trajectory.

2.3 Results

As the starting model, I first fit model (1) to all race combined to check whether or not the model (1) can fit the data and produce reasonable parameter estimates. With the "nlme" procedure, the estimated fixed-effects parameters are: $\hat{K}_1 = -8.5$, $\hat{K}_2 = -0.69$, $\hat{A}_0 = 11.96$ and $\hat{T}_0 = 62.83$, respectively. The estimated random-effects parameters are: $\hat{\sigma}^2_{K_1} = 3.22$, $\hat{\sigma}^2_{K_2} = 1.31 \times 10^{-4}$, $\hat{\sigma}^2_{A_0} = 0.98$, $\hat{\sigma}^2_{T_0} = 4.99$ and $\hat{\sigma}^2 = 2.31$. This model fitting indicates that generally the swimming time would drop about 8.50 seconds (i.e. $\hat{K}_1 = -8.50$) per year from age 7 to the transition age at 12 (i.e. $\hat{A}_0 = 11.96$). At this age, the average swimming time would be 62.83 seconds (i.e. $\hat{T}_0 = 62.83$) , whereafter the swimming time would drop at a much slower pace at a little less than a second per year (i.e. $\hat{K}_2 = -0.69$). From previous experience, these results are not only very reasonable for a group of this size and ability, but are also incredibly relevant to my own situation as well. My times are (104, 90, 78, 70, 64, 61, 61, 61) seconds from ages 7 to 14. I dropped 14, 12, 8, 6 and 3 seconds per year from age 8 to age 12 with average 8.6 seconds per year and after age 12 my time stayed at 61 seconds as can be seen in Figure 4.

With these encouraging results, I fitted a series of models as follows:

- **Model 1**: race-specific model with 8 fixed-effects parameters and 8 random-effects parameters where I found that the transition age for both races was quite close (i.e. $\hat{A}_{0,Asian} = \hat{A}_{0,NonAsian}$),
- **Model 2**: a reduced model with the same transition age for both races (i.e. $A_{0,Asian} = A_{0,NonAsian} = A_0$),
- **Model 3**: the same K_1 for both races,
- **Model 4**: the same K_2 for both races,
- **Model 5**: the same T_0 for both races as well as some of the combinations of these restrictions.

A series of maximum likelihood ratio tests for model selections were performed and we found the Model 2 is the best model fit among all the models.

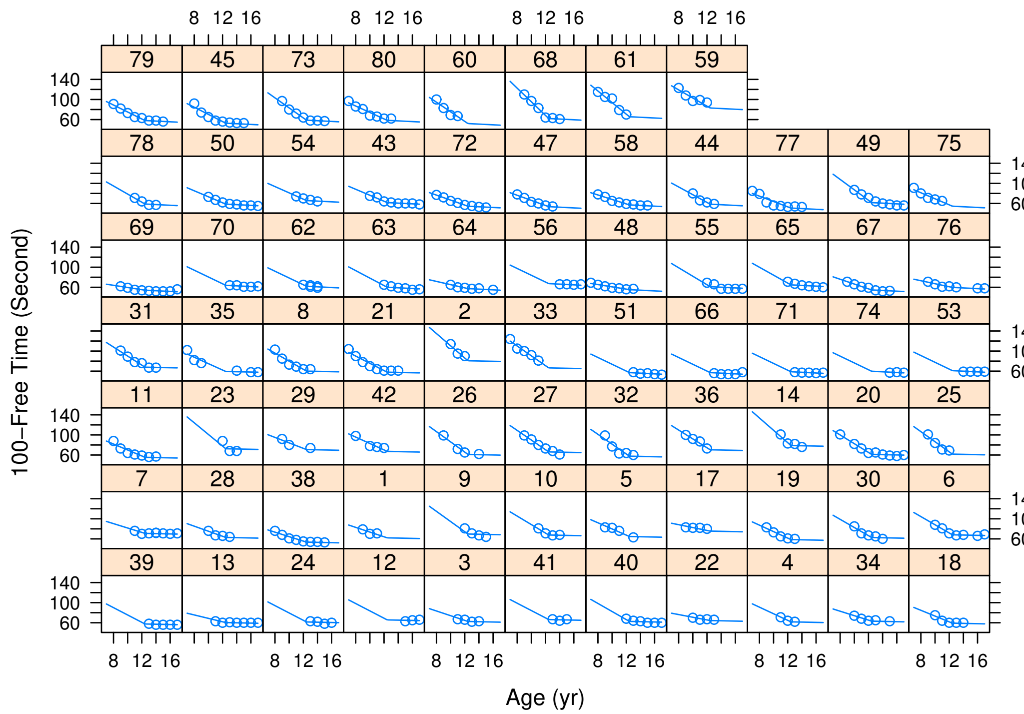

Figure 2: Data (in dots) and the best model fit (solid line) for the 74 swimmers.

This model fits the data very well, and can be seen graphically in Figures 2 and 3. Figure 2 illustrates the model fitting for each of the 74 swimmers with this nonlinear hockey-stick model. In this figure, the observed swimming times for each swimmer are plotted in dots and the model fits in solid lines. It can be seen that this nonlinear hockey-stick model fits the data very well. To further illustrate the model fits, the predicted swimming times were extracted from the model fitting which was then plotted with the observed swimming times as seen from Figure 3. As seen from Figure 3, the model predicted, and the observed swimming times are scattered around the perfect 1-to-1 line and again indicating reasonable model fitting.

With this model, the estimated common transition age is about 12 years old (i.e. $\hat{A}_0$= 12.45 with SE = 0.09), where $\hat{T}_{0,Asian}$= 64.56 (SE= 1.17) and $\hat{T}_{0,NonAsian}$= 59.47 (SE= 1.15). Other estimated parameters are $\hat{K}_{1,Asian}$=−7.31 (SE= 0.61), $\hat{K}_{2,Asian}$=−0.31 (SE= 0.29), $\hat{K}_{1,NonAsian}$=−6.77 (SE= 0.67), and $\hat{K}_{2,NonAsian}$=−0.70 (SE= 0.22), respectively.

All the parameters are highly statistically significantly different from zero except $K_{2,Asian}$ indicating that the time improvement for Asian group will be very minimal after the transition age. From the model fitting, the swimming time for Asian group initially dropped 0.54 seconds faster (i.e. $\hat{K}_{1,Asian}$=−7.31) than the Non-Asian group (i.e. $\hat{K}_{1,NonAsian}$=−6.77) to the transition age 12. Interestingly after the transition age (i.e. age 12), the swimming time for the Non-Asian group drops over twice as fast (i.e. $\hat{K}_{2,NonAsian}$=−0.70) as the Asian group (i.e. $\hat{K}_{2,Asian}$=−0.31), which makes sense given that typically after this age, Non-Asian swimmers typically grow to be stronger and taller than Asian swimmers.

To graphically illustrate the data and model, Figure 4 is the well-known spaghetti plot for the data from the 74 swimmers superimposed with the best model fits (the solid lines in both panel). I also plotted my own swimming progression with a thick dashed line in the left panel. It can be seen that I have been a pretty good swimmer.

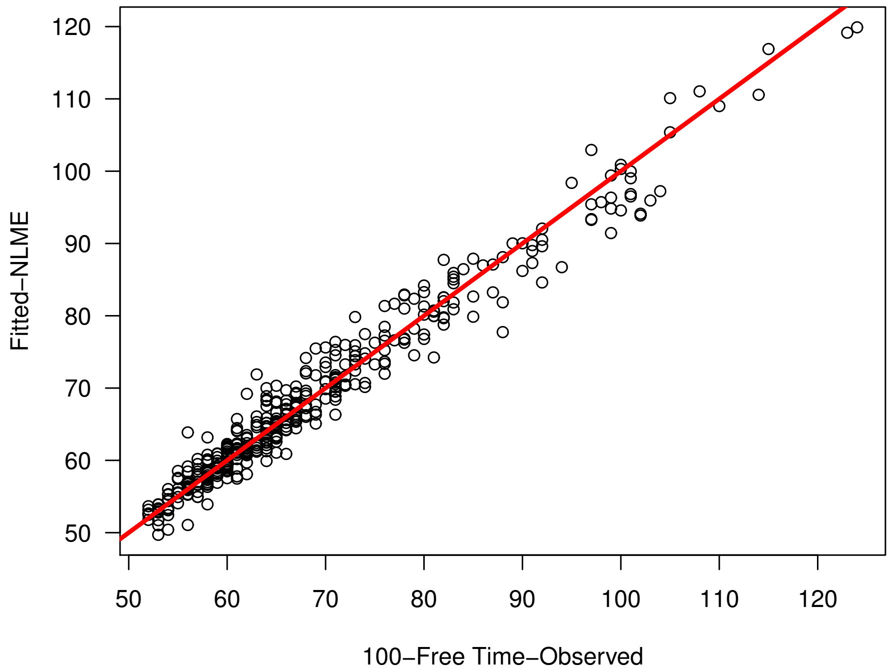

Figure 3: The observed swimming time from the 74 swimmers vs. the model predicted swimming time. The solid line is the one-to-one line indicating perfect model fitting and prediction.

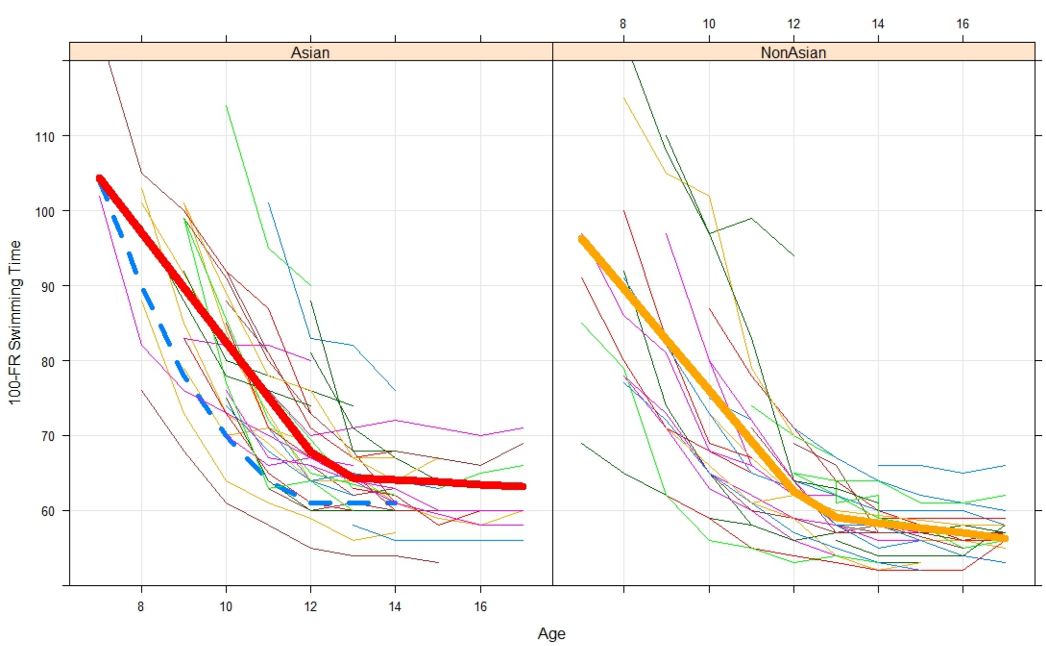

Figure 4: Spaghetti plots for all 74 swimmers (thin lines) by Race. The dashed line in the Asian panel is my swimming time. The thick lines in both panels are the best model fit.

3. Discussions

In this paper, I proposed a nonlinear hockey-stick swimming time model to analyze the data for this certain event. The model fit the data sufficiently, but not only that, the results point towards the knowledge of puberty and the changes it inflicts on the body as a reason for the results.

With this model, the estimated common transition age for both races was about 12 years old, which is consistent with our common knowledge on average ages when girls hit puberty as reported in Anderson et al. (2003), Al-Sahab et al. (2010) and Hamilton-Fairley (2013). Before the transition age, it can be seen that Asian swimmers demonstrate a faster rate of improvement as compared to Non-Asian swimmers. However, Asian swimmers on average share an earlier transition age whereafter they will plateau to a larger degree than Non-Asians.

The definition of the onset of puberty may vary on different perspectives (such as hormonal versus physical, etc.), but a common definition for the onset is the physical changes to a person's body. This may vary between individuals, but typically begins between 10 and 13 years of age, which is consistent to the finding of this analysis. In a systematic review and meta-analysis, Lieberoth et al (2014), confirmed age 12 as the age of menarche for girls which is linked with risk of asthma as well. From this analysis, at this age, the estimated times for Asians will on average reach about 65 seconds while the Non-Asians reach about 60 seconds. Initially, Asians dropped 0.54 seconds more per year than Non-Asians. Before the transition age, the average swimming time for Asians drops at 7.31 seconds per year while times for the Non-Asian group drops at 6.77 seconds a year. After the transition age, the average time dropped per year was 0.31 seconds for Asians, which is considerably lower than the Non-Asians at 0.70 seconds. The effects of menarche are different for all races. Since Non- Asian swimmers usually become stronger and taller than the Asian swimmers earlier, it is evident that those advantages will work to their benefit.

The effects of puberty also include fat distribution of the body, changes in body proportions, and growth spurts. But what some may not be aware of is that along with growth spurts, come periods of great increase in fat and muscle growth. Basically, puberty is a time of increased strength and endurance. In addition to the increase of muscle mass, there is a maturation of the circulatory and respiratory organs such as the lung and the heart. The maturation increases the capacity of both organs and most surely benefits the performance of an athlete of any sport.

Before this, I had various ideas as to why the majority of my Non-Asian counterparts were improving at a faster rate, but none of them were backed up by proof. However, as the reason has unfolded itself, it is better known now what truly backs up a phenomenon. In reality, the fact is that no one can change their own ethnicity, no one can change the ethnicity of their friends, and it simply isn't possible to completely alter what happens to each person naturally. But knowledge is one thing that can always be expanded upon; that is one thing that we can change. Even though my competitors frequently possess shoulder widths twice as wide as mine, considerably more muscle mass, and always seem to tower over my measly 5' 5", the sport has been a substantial part of my life and will continue to be. Recently, my family has moved to Chapel Hill, North Carolina where the all too normal, extremely stressful, junior year was waiting. A move before junior year, a much more competitive high school, and a new agenda to make new friends with my extremely sarcastic personality has been challenging task. However throughout the year, there have been many upsides, and being by far the most important of them all: making it to NCHSAA States as the captain of the East Varsity swim team.

Acknowledgements

I would like to thank John D. Chen, a vice president of ”Credit Suisse” (New York), for his technical support. I’d also like to thank my math teacher at East Chapel Hill High School, Jay Wilson, for his support and guidance this whole year.

REFERENCES

Anderson, S.E., Dallal, G.E. and Must, A. 2003. Relative weight and race influence average age at menarche: results from two nationally representative surveys of US girls studied 25 years apart. Pediatrics 111 (4): 844-850.

Brander, J.A., Egan, E. J. and Yeung, L. 2014. Estimating the effects of age on NHL player performance. Journal of Quantitative Analysis in Sports. 10(2), 241-259.

Al-Sahab, B., Ardern, C.I., Hamadeh, M.J. and Tamim, H. 2010. Age at menarche in Canada: results from the National Longitudinal Survey of Children & Youth. BMC Public Health 10 (736) 1-8.

Hamilton-Fairley, Diana. 2013. Obstetrics and Gynaecology (Second edition). Blackwell Publishing. Retrieved 2013-11-09.

Kovalchik, S. A. and Stefani, R. 2013. Longitudinal analyses of Olympic athletics and swimming events find no gender gap in performance improvement. Journal of Quantitative Analysis in Sports. 9(1): 1559-0410.

Lieberoth, S., Gade, E.J., Brok, J., Backer, V. and Thomsen, S.F. 2014. Age at Menarche and risk of asthma: systematic review and meta-analysis. Journal of Asthma; 51(6) 559-565.

Pinheiro, J. C. and Bates, D. M. 2004. Mixed-Effect Model in S and S-Plus. Springer, New York, NY.

Sarkar, D. 2008. Lattice: Multivariate Data Visualization with R. Springer, New York, NY.