4-21-2008

An Economic Analysis of the Effects of Steroids on Season Best Performances in Track and Field

Allison Fisher '08

"An Economic Analysis of the Effects of Steroids on Season Best Performances in Track and Field"

Allison Fisher
Faculty Advisor- Michael Seeborg
April 21, 2008

This paper investigates the relationship between steroids and an athlete's ability to run a season best time in track and field. Certain event groups have seen a faster drop in season bests than others. Medical research indicates which of these event groups (sprinters, distance, throwers, jumpers) would most benefit from steroids. I hypothesize that steroids allow sprinters to improve on their season bests more than other event groups. This hypothesis is based on a production theory from the economics literature where inputs such as coaching, facilities and steroids produce season best performances. Using ordinary least-squares regression, I use a dummy variable to determine if steroids were significant in track and field events. The regressions show that shot put had the highest marginal effects during the steroids period.

I. Introduction

Track and field has been a sport under much scrutiny as of late. Most recently, the poster girl of United States track and field, Marion Jones, has pleaded guilty to lying to federal investigators regarding allegations of steroid use. The Chicago Tribune reported that Jones admitted to taking steroids from September 2000 to July 2001 which encapsulates the 2000 Olympics in Sydney. At these games, Jones won three gold and two bronze medals. Jones also won several other medals at the world championships in 1999 and 2001, along with holding United States records in several relay events. Under statute of limitation rules, the International Olympic Committee (IOC) and other governing bodies have the ability to go back eight years to remove medals and nullify results. Jones returned her Olympic medals to the Anti-Doping Agency and has stated she will forfeit all results and medals since September 1, 2000 (Hersh). Even though she is trying to correct her actions, this brings into question the validity of the relay races in particular. In the most recent news, the IOC has removed the 2000 Sydney Olympic medals from her relay teammates despite their strong attempts to keep them (Wilson). The recent crackdown on steroid use in track and field and numerous positive tests by world class athletes have raised questions regarding the validity of race outcomes and their corresponding times/distances

Marion Jones may be the most recent athlete and biggest name to be affected by the crackdown on doping, but she is not the only one. Her ex-husband, C.J. Hunter, an Olympic gold medalist, was also busted for doping. Tim Montgomery, father of Jones' young son and former 100m world record holder, was stripped of his record after testing positive for steroids. Two of her teammates on the 4x100m relay from the Sydney Olympics, Chryste Gaines and Torri Edwards, have served bans for steroid use in the past few years (Hersh). Other athletes

found guilty and stripped of their world records include Ben Johnson and Justin Gaitlin, both 100m gold medalists and temporary world record holders in 1988 and 2006 respectively (IAAF). It seems that a pattern has appeared in that, the 100m record has consistently been broken and the athlete consequently tested positive for steroids. Is it just a coincidence, or are sprinters benefiting more from steroid use than other groups? If this is the case, are steroids and their benefits the reason that sprinters seem to have faster drops in season best times than other event groups?

This study investigates the relationship between season best times and several independent variables, including steroids. Before performing data analysis, a theoretical framework that is based on the theory of production is developed to explain the production of season best performances. It draws on medical research showing that sprinters benefit more from the use of steroids than other event groups. This framework includes research involving slow-twitch and fast-twitch muscles and an athlete's ability to increase the strength of these muscles through steroid use. From this framework, I use a data analysis to test the hypothesis that there will be a faster drop in season best times in the sprint events than in other track and field events, since I believe sprinters benefit the most from steroid use.

II. Literature Review

It is clear that anabolic steroids have been used to help athletes improve their performances, yet no studies have addressed the effect of steroids on season best times. Studies have addressed these subjects separately through the progression of season best times and using game theory to predict the probability that an athlete will use steroids (Haugen, 2004). One study completed, similar to my research was conducted, by Yuanlong Liu and Robert Schutz (1998). Liu and Schutz analyzed the trends of season bests of seven different events for track

and field using linear and exponential models. These events included 100m, 400m, 1500m, 5000m, 10000m, marathon and high jump. The purpose was to find the best model to fit the data, and also to predict new world records in the year 2010. The authors found that some of the events data took a linear shape while others conformed to a nonlinear model in that the times were decreasing at a decreasing rate or the distances were increasing at a decreasing rate (Liu and Schutz, 1998). The authors used the exponential model to predict the year 2010 world records, yet some of their projections of world records have already been surpassed by athletes indicating that a nonlinear model may not be the best option for forecasting. This finding leads me to believe that the linear model might provide a better fit for track and field data.

The *Performance-Enhancing Drug Game* by Kjetil Haugen (2004) addresses the topic of steroid use among athletes. Using game theory, Haugen shows how economic forces drive athletes to use drugs. One of the first points he identifies is that "athletes who focus on single abilities such as force or speed are more likely to have drug problems than athletes in whom a range of combined talents are needed (i.e. soccer)" (2004). This is consistent with my hypothesis since I believe sprinters, who rely heavily on force and speed, benefit most from steroids. Haugen calculated the probability of "getting caught" using steroids to be around 1% based on the number of drug tests performed and the number of positive tests. He found that, given the value of winning and the probability of being caught, most athletes would choose to dope, given the choice (2004). Thus, it is clear that there is a steroid problem among athletes, and my research will contribute to the literature regarding the consequent effects of steroid use on season best performances.

A study completed by Nikola Medic, Janet Starkes and Bradley Young (2007) highlights the effects of age on performance achievement in Masters athletes. The authors feel that the age

difference "may result in significant difference in performance" (Medic, Starkes and Young, 2007). This applies to the analysis of steroid use, since steroids result in significant differences in performance. The authors compared the time run by an athlete of, for example, 55 years old in an event in the 55-59 age bracket, to what time he would run if he was instead 59 years old and at the other end of the age bracket (Medic, Starkes and Young, 2007). Therefore, the authors were determining the effects of age on performances times. This is similar to my research since I will be determining the effects of steroids on season best performances among different track and field event groups.

A chapter from Sports Economics by Rodney Fort (2002) titled, "The Value of Sports Talent" also lends itself to my research. Primarily, Fort's discussion of marginal revenue product theory is pertinent to the issue of steroids in track and field. He identifies three factors that explain variation in pay among baseball players, but these factors may also apply to variation in track and field times run by athletes. One of these is innate ability, which is simply the fact that some people have a greater ability in the sports world (Fort, 207). The other two factors are training and experience which, as Fort puts it, "require investment of scarce time, energy, and monetary resources" (208). These factors directly relate to track and field since innate ability, training, and experience all contribute to investments in human capital which will be a critical part of my theoretical framework explained in the next section.

A final study that contributes to my paper is *Production Efficiency: Case of Professional Basketball* (Huang, Siegfried and Zak, 1979). The authors use a production function to determine a team's potential output given a number of inputs (1979). The authors used several different input factors and completed a special concentration on the "home-court advantage" (Huang, Siegfried and Zak, 1979). They focused mainly on ratios such as the ratio of field goal

percentages, which takes into account both the winning and losing team's field goal percentage (Huang, Siegfried and Zak, 1979). However, it would not be realistic to use a ratio for track since an athlete's performance generally does not depend on the performance of a competitor. An athlete has the ability to run a season best performance regardless of competition. The principle of measuring the marginal products of the production function will be useful in my study since I will identify a production function in the theoretical framework. A production function will not be estimated in the regression analysis, but will serve as a stepping stone from the theory to the empirical model.

III. Theory

Unlike sports such as baseball or football in which the opponent's or fellow teammate's ability may effect an athlete's performance, track and field relies solely on the individual athlete. The goal of a track and field athlete is to push his/her body to the ultimate limit, which will lead to season best or world record times. However, an athlete may reach a point in which he/she can no longer compete at the international level without the help of steroids. Steroid use in track and field is more prevalent than people may think. In reference to the 1988 Seoul Olympics, a Soviet coach reported to the New York Times, "I feel sorry for Ben Johnson. All sportsmen—not all, but maybe 90%, including our own—use drugs" (Yesalis, 51). The prevalence of steroids in the sport is apparent, the only question that remains is, who benefits the most?

A production function provides one useful framework for considering the effect of steroid use on performance. By definition, a production function indicates the highest output that a firm can produce for every specified combination of inputs (Pindyck and Rubinfield, 189). Thus, from an individual athlete's perspective, the highest output would be the fastest time possible, given a combination of inputs possessed by the athlete. Specifically, a production function may

be the most relatable in this case since it is used frequently in relating inputs to an output (Pindyck and Rubinfield, 190). The function is typically defined as:

$$Y = A\ F(L,K)$$

where Y= output, L= labor input, K= capital input and *A* is a constant determined by technology (Pindyck and Rubinfield, 190).

Therefore, when applying the production function to track and field, Y is defined as the season's best time/distance given labor inputs, capital inputs and technological advances. Labor inputs would include work done by people (Pindyck and Rubinfield, 190). Thus, the amount of time spent training, workout intensity, coaching, and nutrition would be labor inputs. Capital inputs refer to "already produced goods available for use as factors of production" (Pindyck and Rubinfield, 190). In the track and field context, capital inputs would include shoes, spikes, training equipment, track surfaces and, of course, steroids. The technological advances (*A*) represent the athlete's natural ability.

Labor inputs are directly related to the complementary theory of human capital. Human capital is the knowledge, skills and experience that make an individual productive and thereby able to earn a higher income over a lifetime (Pindyck and Rubinfeld, 564). When interpreted within the context of track and field, human capital is the athletic ability that can be created through investments made by the athlete with the goal of performing well. Thus, the more experience or technical knowledge an athlete gains, such as good mechanics out of the starting blocks, the more productive he will be in running a faster time. At a certain point, however, athletes may reach their maximum performance output and may turn to steroids in order to continue to improve performances. These human capital investments relate to the labor inputs in a production function since they make an athlete more experienced and skilled.

It is argued here that improvements in human capital and technological innovations occur rather continuously in track and field and that they should produce steady linear improvements in season best performances. Steroid use, on the other hand, could cause nonlinear improvements at the times when they are introduced. However, there are factors other than steroids that could lead to gradual linear improvements in performance. First, investments in human capital (e.g., training methods and coaching) improve systematically over time. Second, physical improvements in capital (e.g., track surfaces and shoe quality) improve systematically over time. Third, world class athletes tend to have children with other world class athletes, thus providing a higher level of natural talent to their children. Therefore, their children have the genes to develop a higher level of athleticism. This also contributes to the expectation that times/distances will improve linearly over time. However, nonlinearities in track and field performance over time are possible if steroids are introduced into the production function. The reason is that steroids were developed and introduced rapidly and they should have an abrupt effect on performance. Under the Empirical Models section, several nonlinear models are introduced to try to estimate the effects of the introduction of steroids on track and field performance.

To understand the likely effect of steroids on track and field performance requires some basic knowledge of the anatomy of muscle fiber. Athletes have specific percentages of slow-twitch and fast-twitch muscle fibers and can do nothing to change this. Fast-twitch muscles give athletes the ability to perform explosive work while slow-twitch muscles provide the ability for endurance work (Dare, 1979). Therefore, it would make sense that sprinters, jumpers and throwers all need a higher percentage of fast-twitch muscles since all of their actions require explosive force while distance runners require a high percentage of slow-twitch fibers. Fast-

twitch fibers react to training with significantly more hypertrophy (Dare). As far as training these fibers, both slow-twitch and fast-twitch fibers grow in thickness, although fast-twitch fibers respond with more growth (Dare). These findings contribute to the fact that throwers, jumpers and sprinters have more muscle mass than distance runners. Their higher percentages of fast-twitch fibers indicate a bulkier frame and the ability to build upon these fibers.

In detailing the percentages of slow-twitch fibers among different event groups, it is important to define the events contained in each event group. Sprinters are defined as those competing in the 100m-400m events. Middle distance events include the 800m-3000m, steeplechase and distance events are those over 5000m. Other event groups that will be discussed are throwers (shot put, javelin, weight throw, discus) and jumpers (long jump, triple jump, high jump). David Costill and his associates (Dare, 1979) discovered general relationships between event groups and muscle fiber percentages by analyzing several world class runners and athletes. The researchers found that sprinters typically have anywhere from 8-24% of slow-twitch fibers with the rest being fast-twitch. Middle distance runners have approximately 40-65% slow-twitch while distance runners have 80+% of slow-twitch fibers. Throwers and jumpers both had 40-50% slow-twitch fibers (Dare). The findings of this study verify the idea that sprinters, jumpers and throwers all require more fast-twitch fibers than distance runners. These percentages will prove useful in developing hypotheses regarding the determinants of best performances.

Given that athletes have the ability to increase the strength of fast-twitch fibers more than slow-twitch, it is interesting to see the effects of steroid use on these fibers. In a study of athletes using anabolic steroids for twenty-four weeks, researchers observed a significant increase in the muscle fiber area (Yesalis, 2000). Thus, steroids would double the effects of fast-twitch fiber

growth. In a study testing the effects of steroids on distance runners, rats running to exhaustion showed no gained benefit from the injection of anabolic steroids (Yesalis). Yesalis concluded that there was no evidence of any benefits to endurance athletes (2000). Dr. William Taylor (2002) found that steroids reduce an athlete's reaction time. Clearly, this is another benefit to sprinters since they rely heavily on their ability to get out of the blocks faster than the competition. Dropping even five hundredths of a second from the reaction time can be huge for a sprinter, whereas for other event groups, it would not be as significant. From these studies and evidence, it seems that anabolic steroids are more beneficial to those with more fast-twitch muscle fibers, mainly sprinters, given the faster reaction time.

Before continuing, it is critical to distinguish between testosterone and the anabolic steroids I will be addressing in this study. Testosterone is produced by both males and females, and is considered to be an anabolic steroid. It has the ability to bind to receptors contained in many tissues throughout the body (Taylor, 2002). This binding causes both androgenic and anabolic functions within the body. Androgenic functions include most of the occurrences during puberty. Anabolic functions are those that steroids are used for, such as increased muscle mass (Taylor). Thus, anabolic steroids are actually synthetic derivatives of testosterone (Taylor). Their purpose is to retain and magnify testosterone's anabolic effect while reducing the androgenic effects that most find to be the problem with using testosterone (Taylor). Also, anabolic steroids are much safer than injecting pure testosterone. Therefore, anabolic steroids ability to increase the anabolic functions makes them much more potent and helpful for athletes than testosterone.

In order for anabolic steroids to be the most productive, athletes generally take them in cycles to correspond with their training and meets they are planning to compete in. By cycling

the steroids, athletes have the ability to avoid a positive test at a competition since the clearance times for different steroids are known. Athletes typically take between three and five steroid cycles per year (Taylor, 2002). The first cycle (6-8 weeks) includes taking low doses of an oral steroid (Taylor). After the corresponding rest period and large increase in muscle mass, the second cycle (6-8 weeks again) is taken at a much higher dosage. In order to achieve the same amount of muscle gain as the first cycle, a much higher dose is required (Taylor). By the third cycle, athletes must begin stacking and pyramiding the steroids in order to achieve the required result. Stacking involves taking several different types of steroids, both oral and injectable, at the same time in order to accentuate the results (Taylor). Pyramiding is when athletes vary the dosage from week to week during a cycle in order to prevent crashing from dependence to the drugs (Taylor). Cycles four and five include the same actions as three only with higher doses.

The history of steroids dates back to ancient Greece when competitors were using testosterone to improve performances. However, my study will only include the use of anabolic steroids and their effect on certain time periods of track and field. Anabolic steroids were hardly used before or during the 1950's and even at the 1960 Olympic Games, steroids were limited to the Soviet strength athletes (Yesalis, 54). The first year that it was apparent that athletes were using steroids was 1964. It began mainly with throwers and power lifters, but eventually sifted to other athletes of the track and field community by the late 1960's (Yesalis). By 1972, as stated by Charlie Francis, "it was the insiders' consensus that 80% of the top male athletes were using steroids" (90). Finally in 1975, the International Association of Athletics Federation (IAAF) banned steroids from competition. From that point until the mid-80's, the IAAF continued to add more banned substances to the list (Francis). Athletes today have the ability to

tailor their steroid cycling to avoid testing and knowledge of steroids has become very advanced. It has been said that those that get caught have either very stupid coaches or very stupid doctors.

An article in The Economist (1998) discussed the use of steroids over the past few decades, and how it has been on the decline since the early nineties. The 1996 Atlanta Olympics were the cleanest ever, since only two of the two thousand athletes checked for steroids tested positive (Superhuman Heroes). Also, Judy Oakes, a British shot putter who opposed using drugs was ranked 27th in the world in 1988, which is believed to be one of the peaks of steroid use. In 1996, with the same mark, she was ranked 12th in the world (Superhuman Heroes). Due to these factors, it is believed that steroid use began to decline in the mid-1990's. Based on the history of steroids above and the theoretical framework provided, I hypothesize that use of anabolic steroids benefits sprinters most and thus, leads to faster changes in season best times for sprinters than any other event group in track and field. Specifically, athletes in events where fast twitch muscle fiber is most important will experience relatively better performance improvements compared to athletes in events where slow twitch muscle fiber is most important.

IV. Data

I will be using data provided by the International Association of Athletics Federations (IAAF), which is the governing body of world track and field. They provide top lists for each year in all events which are published in *Track and Field News* annually. I will focus on the time period from 1949 to 2007. The study will focus only on men's season bests since womens' data is not complete in all events for this time period. I will be analyzing six events that I feel encapsulate track and field as a whole. These events are: 100m, 400m, mile, 5000m, long jump and shot put. The 100m and 400m will be used to analyze both short sprinters and long sprinters. The mile will be used as the basis for middle distance runners, while the 5000m will be used for

distance runners. I selected long jump and shot put to represent the jumpers and throwers, respectively, since new technology has not changed the event greatly over the years. I ran separate regressions using the top season best performance for each event and then the median of the top 10 performances for each year per event. Using the median of the top 10 season best performances will allow additional analysis and serve as a control for "outlier" observations.

One data problem to be addressed is the use of Fully Automated Timing or FAT. It was not introduced to the track and field world until 1975 (Francis). Therefore, times before 1975 (especially in the short sprints) may be compromised. Before 1975, manual timing involved a person starting and stopping the watch based on their perception of the gun being fired and the athlete crossing the finish line. FAT removed the human reaction time factor altogether since it involves a system that is linked to the starting gun. The clock automatically starts when the gun fires, and a camera captures the precise moment when the athlete crosses the finish line. Experts have determined that the time would be much slower under FAT than manual timing. To adjust manual times, .24 seconds are added on to the times in events up to 1600m (Francis). Thus, times in the 100m and 400m before 1975 were adjusted for this fact. The mile and 5000m were not adjusted, since .24 seconds is not a large enough margin in these events to require correction. Officials feel that this adjustment takes into account the documented delayed reaction in starting the watch and stopping early (Francis, 162).

Based on historical evidence, I selected three time periods that I felt would define the stages of steroid use: 1) a control period, 2) a heavy steroid-use period and 3) a post-steroid period. The control period includes the years 1949-1968 for all events except shot put, which is defined by the years 1949-1966. The reason for this difference is that steroid use began in the throwing events in the mid-1960's and did not shift to other track events until the late 1960's.

The heavy-steroid use period is defined from 1969-1994 for all events except shot put, which includes the years 1967-1994. Despite testing being introduced in 1975, history has indicated that athletes still used steroids heavily until the early 1990's, since they learned to cycle the steroids to pass tests. The last period is from 1995-2007 and represents the time period in which steroid use has been believed to have decreased because of more effective testing. After adjusting the 100m and 400m for manual times previous to 1975, I plotted the season best times/distances for each event by year. A trend line was fit to the data and residuals were calculated for each stage of steroid use.

V. Empirical Models

This research develops in four models. The first model fits a simple regression to the annual data for each event and examines the deviations from the trend over three time periods. The model estimated using OLS regression is:

Model 1: $PERF = \alpha 1 + \alpha 2\ (YEAR) + \mu$

Where PERF= Measure of track and field performance (Season best performance or Median of top 10 season best performances)

YEAR= Actual year from 1949-2007

Model 2 of the research introduces dummy variables for the time periods (control period, heavy use period and post steroid period.) The model to be estimated using OLS regression is:

Model 2: $PERF = \alpha 1 + \alpha 2\ (YEAR) + \alpha 3\ (HEAVY\_USE) + \alpha 4\ (POST) + \mu$

Where HEAVY_USE= 1 if YEAR=1969-1994 (1967-1994 for Shot Put)
0 if otherwise

POST= 1 if YEAR=1995-2007
0 if otherwise

Model 2 allows the performance trend line to shift at the beginning of the heavy steroid use period (HEAVY_USE) and again at the beginning of the post-steroid use period (POST). Although the trend line can shift in this model, the slope within all three segments will remain constant.

Models 3 and 4 of the research explore whether there are non-linearities in the model beyond the shifts estimated in Model 2. The exploratory equations estimated using OLS regression techniques are:

Model 3: PERF= $\alpha 1+ \alpha 2$ (YEAR) + $\alpha 3$ (HEAVY_USE) + $\alpha 4$ (POST) + $\alpha 5$ ($YEAR^2$) + μ

Model 4: PERF= $\alpha 1+ \alpha 2$ (YEAR) + $\alpha 3$ (HEAVY_USE) + $\alpha 4$ (POST) +
$\alpha 5$ (HEAVY_USE*$YEAR^2$) + $\alpha 6$ (POST*$YEAR^2$) + μ

Where $YEAR^2$= Squared YEAR term
HEAVY_USE*$YEAR^2$= Interaction variable
POST*$YEAR^2$= Interaction variable

Model 3 tests whether there is a quadratic nonlinearity over the entire period from 1949 to 2007 while Model 4 tests whether there are quadratic non-linearities within the segments.

Using Models 1 and 2 described above, I will run two sets of regressions. The first set has season best time as the dependent variable. The second set of regressions has the median of the top 10 season best performances as the dependent variable. The coefficients to the two dummy variables, HEAVY_USE and POST, are interpreted in reference to the control period in the sample. Running events (100m, 400, Mile, 5000m) would be expected to have negative coefficients for YEAR, HEAVY_USE and POST since performance times should be decreasing over the years. The field events (LJ and SP) would be expected to have positive coefficients for YEAR, HEAVY_USE and POST since performance distances should be increasing over time. In events where fast-twitch muscle fiber is most important (100m, 400m, LJ and SP), I would expect the HEAVY_USE variable to be significant and have a larger coefficient than the POST variable. This is due to the fact that performances should show greater improvement during the HEAVY_USE period than during the POST period for these events.

Because of data limitations, it is important to note some possible biases. If, for example, the coefficient to HEAVY_USE is significant, we cannot conclude with absolute certainty that

this significance is due to steroid use. Other factors such as track surface improvements, new shoe technology and training/coaching techniques may also contribute to the sudden improvement in performances during the heavy use period. While it's nearly impossible to determine if certain training or coaching techniques were developed that substantially improved performances across the board, it is possible to discuss breakthroughs in track surfaces and new shoe technology. The most significant track surface improvement occurred in 1986 with the development of the "Mondo" track (Schulz). Mondo tracks are made of all-natural rubbers and are believed to be "fast tracks" (Schulz). Also, Bill Bowerman's development of the waffle-trainer shoe in 1975 was a huge improvement in running shoes (Shanks). Both of these breakthroughs occur during the HEAVY_USE period so it's possible that some of the performance improvements could be attributable to these events rather than steroids. However, if steroids were not an important factor in improving performance in the HEAVY_USE period, performance should not deteriorate in the POST period.

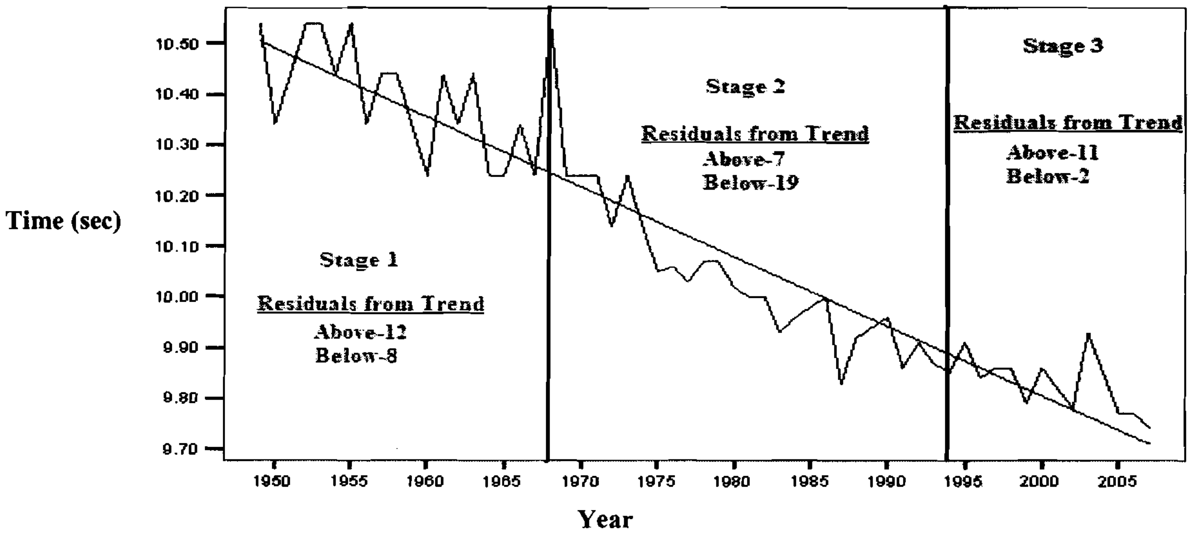
VI. Results

a. Residual Analysis from Regression Model 1

After plotting the data by event and fitting a trend line for each event, I divided the graph into three sections defining the stages of steroid use. Among the different stages, I counted the number of residuals that were above the trend line and the number that were below the trend line. These residuals represent difference between the actual season best performance and the predicted value. For running events, residuals above the trend line indicate that times were slower than predicted, while those below the trend line were faster than predicted. Field events are the reverse in that residuals above the trend line indicate farther distances than predicted while those below indicate distances shorter than anticipated.

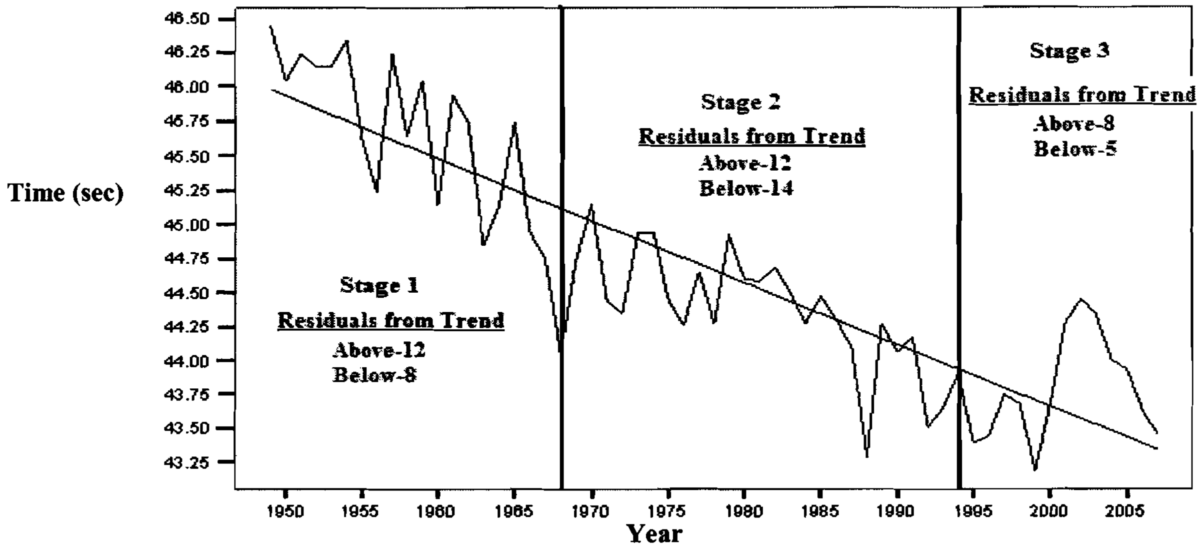
100m Figure 1 shows that Stage 1 and stage 3 both had more residuals above the trend line than below, indicating that times were slower than predicted. In stage 2, 19 of the 26 years had season best times below the predicted line. This is consistent with my hypothesis, since I expected the sprint events to have faster than predicted times during the heavy steroid use period, due to the fact that sprinters have the highest percentage of fast-twitch muscle fibers among event groups.

Figure 1- 100m Season Best Times



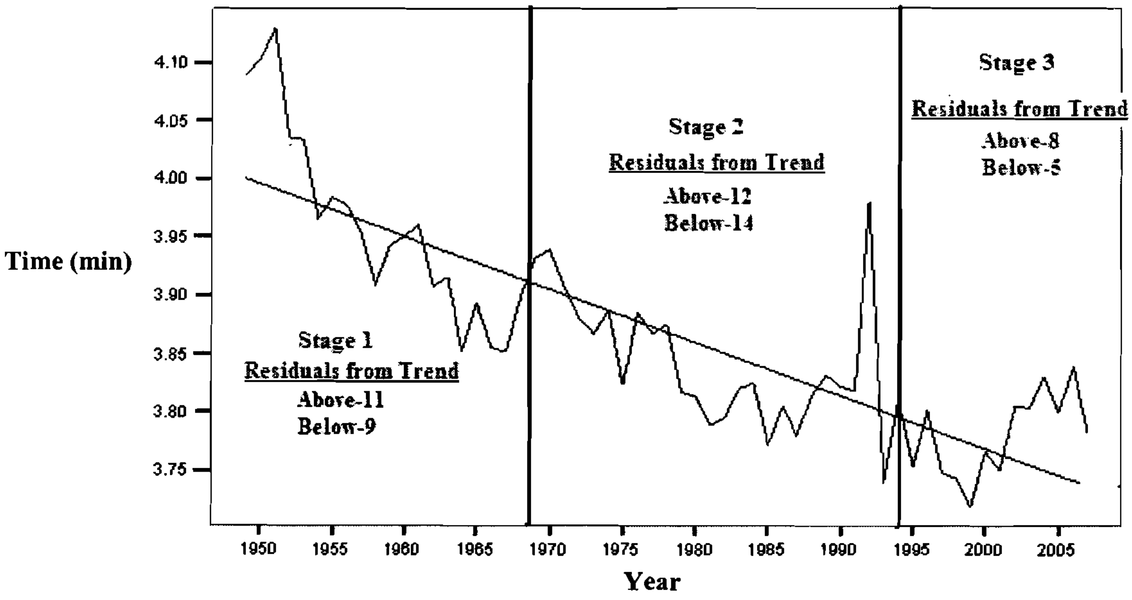
400m Figure 2 shows that as with the 100m, stage 1 had 12 residuals above the trend line and 8 below. Stage 3 had 8 above the trend line and 5 below. This indicates that times were generally slower than predicted in both of these periods. During stage 2, 12 residuals were positive while 14 were negative. Since this is a relatively even count, it's difficult to determine from the residual analysis the effect during the heavy use period. Looking at the overall pattern of residuals, it is not possible to conclude that there were substantial improvements in performance relative to trend during the heavy use period.

Figure 2- 400m Season Best Times



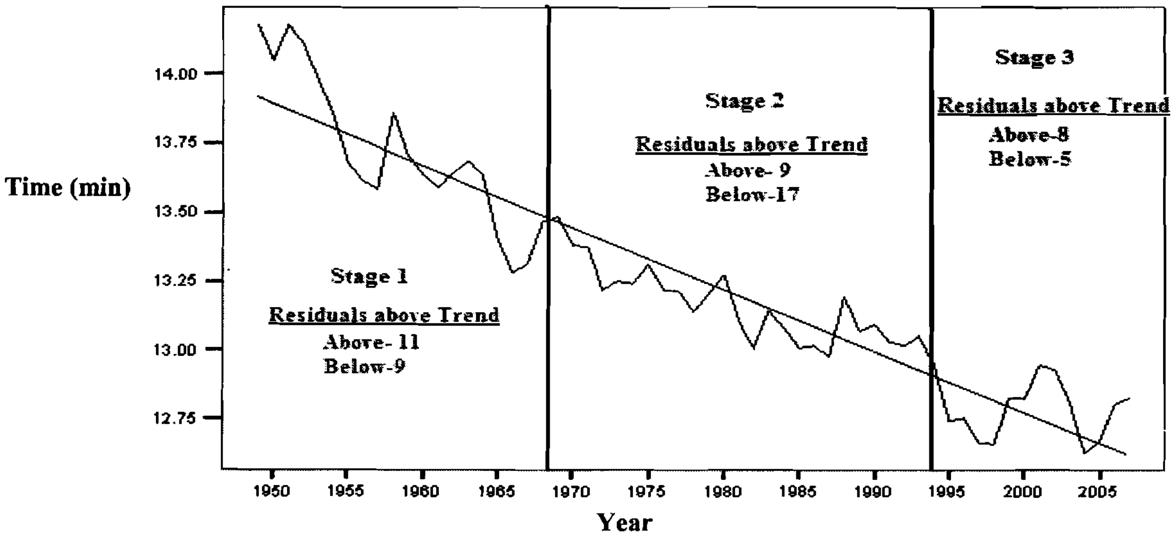
Mile Figure 3 illustrates the mile residuals. Stages 1 and 3 had 11 and 8 residuals above the trend line respectively. Stage 2 had 12 residuals above the trend line and 14 below which again, is fairly even. The mile is considered a middle distance event, and the participants in this event have less fast-twitch muscle fibers than sprinters. Therefore, under my hypothesis, steroid use should not have a significant impact on season best times in the mile.

Figure 3- Mile Season Best Times



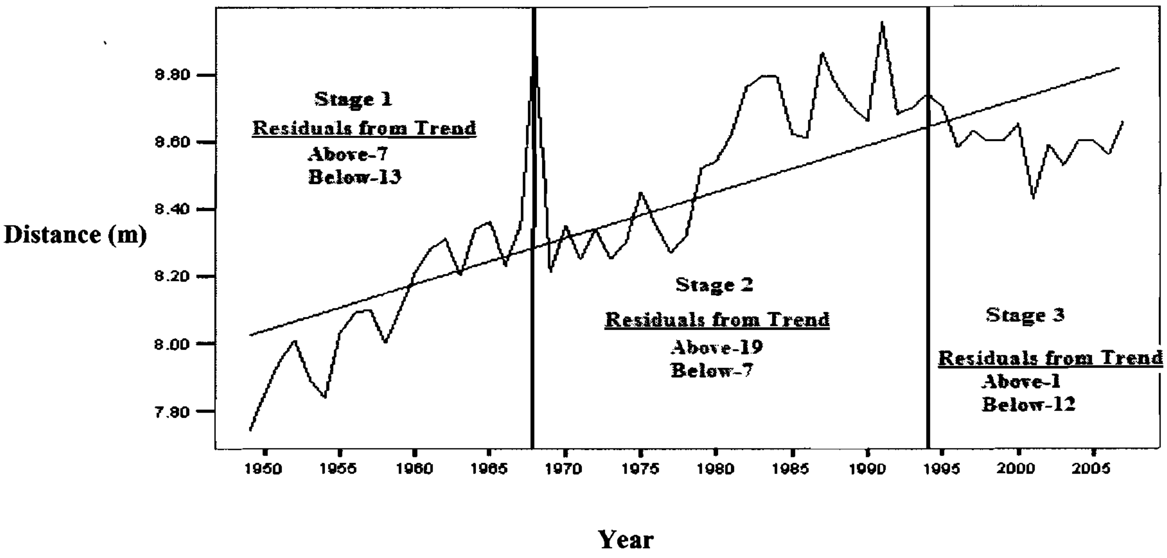
5000m Figure 4 shows that the residuals for the 5000m were not consistent with expectations. Stage 1 had 11 residuals above the trend line and 9 below while stage 3 had 8 residuals above the trend line and 5 below. Surprisingly, stage 2 had 9 residuals above the trend line and 17 residuals below. This follows the typical steroid pattern illustrated in the other events. Since the 5000m is a distance event and therefore employs a low percentage of fast-twitch muscle fibers, my hypothesis would not specify this event as one benefiting from steroids.

Figure 4- 5000m Season Best Times



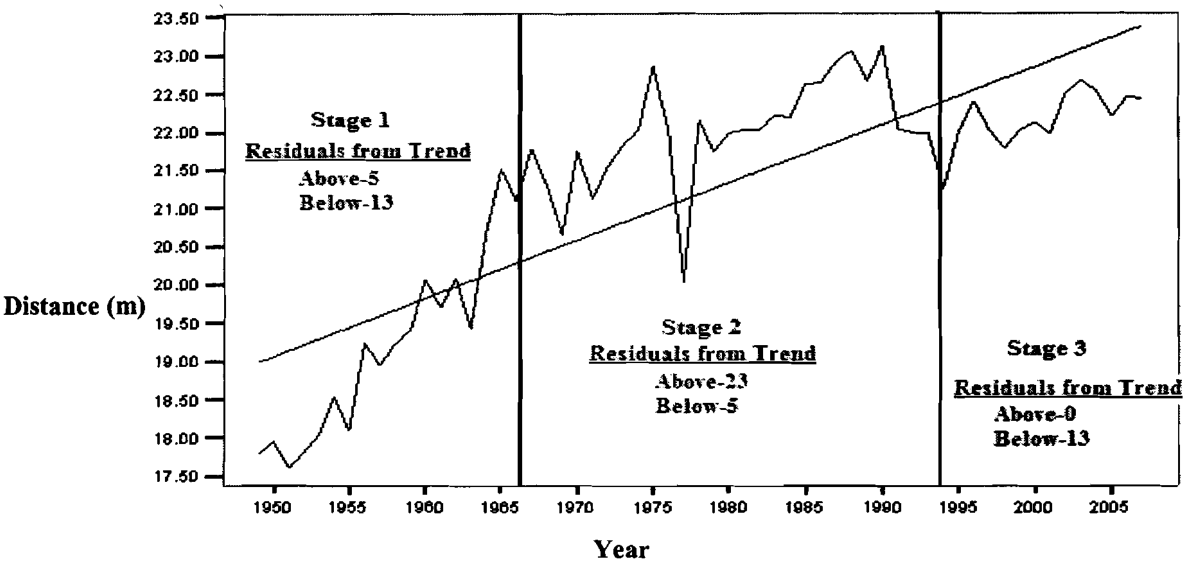
Long Jump Stage 1 had 7 residuals above the trend line and 13 below indicating that more season bests were shorter than the predicted distance (see Figure 5). In stage 3, especially, there was only one further than predicted jump, while the other twelve fell below predictions. Stage 2 had 19 jumps further than predicted, with only 7 jumps shorter than predicted. This is consistent with my hypothesis since long jumpers rely on fast-twitch muscles and explosiveness to achieve a season best distance.

Figure 5- Long Jump Season Best Distances



Shot Put Figure 6 shows the residuals for the shot put. The residuals for stages 1 and 3 were similar to long jump since they had more residuals below the trend line, indicating shorter than expected distances. Stage 1 had 5 residuals above the trend line and 13 below. Stage 3 did not have any performances above the trend line, which may be linked to a drop off in steroid use. In stage 2, 23 of the 28 residuals were above the trend line, indicating significant above expected performances. This again may be linked to the heavy use of steroids during that period. Shot put is an event similar to long jump in that its' participants rely on fast-twitch muscle fibers and explosiveness. My hypothesis denotes shot put as an event that would benefit from steroids.

Figure 6- Shot Put Season Best Distances



b. Model 1 Regression Results

Sections b, c, and d present the regression results obtained from running the four models described in the Empirical Models section. The dependent variable is the season best performance (time/distance) or median of the top 10 season best performances while the independent variables are YEAR, HEAVY_USE and POST when included.

The regression results for Model 1 include the coefficient for YEAR along with the R^2 and adjusted Durbin-Watson statistics. Autocorrelation is implied if the Durbin-Watson statistic is below 1.38 which represents 60 observations and one variable at the .01 significance level. This was a problem in many of the events so the Cochrane-Orcutt correction was used to correct for autocorrelation in these cases.

Table 1: Regression Results of Model 1: Dependent Variable is Season Best Performance (t-statistic in parentheses)

	Constant	YEAR	D-W	R^2
100m	38.74 (14.01)**	-.014 (-23.03)**	1.73	.90
400m	136.81 (14.15)**	-.046 (-10.03)**	2.03	.64
Mile	11.32 (7.00)**	-.005 (-7.29)**	2.26	.49
5000m	59.43 (11.92)**	-.023 (-11.96)**	1.67	.72
LJ	-14.99 (-3.13)**	.014 (6.01)**	2.20	.39
SP	-132.14 (-5.12)**	.077 (5.13)**	2.38	.32

*** Significant at the .05 level**
**** Significant at the .01 level**

The coefficient for YEAR was significant for all events at the .01 level and had signs as predicted. This shows that times do in fact decrease in the running events while distances increase in field events. R^2 was fairly high for all events, which indicates that the model explains some of the variation in season best marks.

The regression results for Model 1a are presented in Table 2 and include the coefficient for YEAR along with the R^2 and adjusted Durbin-Watson statistics. Model 1a utilizes the median of the top 10 performances for each year as the dependent variable.

Table 2: Regression Results of Model 1a: Dependent Variable is Median of Top 10 Season Best Performances (t-statistic in parentheses)

	Constant	YEAR	D-W	R^2
100m	36.69 (22.39)**	-.013 (-16.15)**	2.21	0.83
400m	132.49 (12.09)**	-.044 (-7.97)**	2.35	0.54
Mile	10.41 (6.35)**	-.003 (-3.97)**	2.09	0.22
5000m	54.07 (12.10)**	-.021 (-9.12)**	1.98	0.60
LJ	-14.15 (-2.88)**	.011 (4.55)**	2.26	0.27
SP	-63.83 (-2.16)**	.043 (1.54)**	2.49	0.04

*** Significant at the .05 level**
**** Significant at the .01 level**

As with Model 1, the coefficient for YEAR was significant for all events at the .01 level and had signs as predicted. The coefficients were consistent with the results from Model 1 indicating that the median value is representative of the season best value. R^2 was fairly high for

all events except SP. It seems that the median model doesn't explain the variation in season best performances as well as Model 1 which can be illustrated through the lower R^2.

c. Model 2 Regression Results

Model 2 adds on to Model 1 adding the dummy variables HEAVY_USE and POST dummy variables to the YEAR variable included in Model 1. These dummy variables take into account the different stages of steroid use and will determine if performances were significantly different from the original trend. It would be expected that events utilizing fast-twitch muscle fibers (100m, 400m) would have the most significant coefficients during the HEAVY_USE period. Data in Table 3 contains the coefficients for YEAR, HEAVY_USE and POST variables. The adjusted R^2 and adjusted Durbin-Watson statistics are also included for comparison. Corrections for autocorrelation were made to the mile, 5000m and shot put.

Table 3: Regression Results of Model 2: Dependent Variable is Season Best Performance (t-statistic in parentheses)

	Constant	YEAR	HEAVY_USE	POST	D-W	R^2
100m	39.23 (14.87)**	-.013 (-8.72)**	-.083 (-2.09)**	-.025 (-.37)	2.01	.92
400m	133.31 (14.11)**	-.04 (-5.88)**	-.444 (-2.03)*	-.013 (-.44)	1.49	.80
Mile	12.03 (8.45)**	-.005 (-3.80)**	-.009 (-.24)	.039 (.61)	2.25	.52
5000m	61.32 (12.53)**	-.021 (-5.82)**	-.051 (-.53)	-.09 (-.59)	1.67	.70
LJ	-16.68 (-4.25)**	.026 (9.74)**	-.19 (-2.60)**	-.66 (-5.31)**	1.53	.81
SP	-141.98 (-5.99)**	.062 (3.36)**	1.187 (2.33)**	.57 (.683)	2.11	.60

*** Significant at the .05 level**
**** Significant at the .01 level**

As with Model 1, the coefficient for YEAR is significant for all events at the .01 level and has the correct sign. These results are consistent with the Model 1 results reported in Table 1. The R^2 for all events either increased or remained fairly constant. The HEAVY_USE variable is significant at the .01 level for the 100m, LJ and SP. However, the coefficient to LJ has the incorrect sign and therefore is not consistent with expectations. The 100m and SP significance is consistent with my hypothesis since those events require a larger percentage of

fast-twitch muscle fiber. The 400m HEAVY_USE coefficient is significant at the .05 level. This is also consistent with my hypothesis since the 400m is considered a sprint event. The mile and 5000m coefficients to HEAVY_USE are insignificant. Since the mile and 5000m are more slow-twitch muscle fiber events, it is expected that they would not benefit from steroids and therefore be insignificant. All events have insignificant coefficients for the POST period except for LJ but once again, it has the wrong sign. In summary, most of these results are consistent with expectations since we would expect the sprinters and throwers to benefit the most from steroid availability due to the higher percentages of fast-twitch muscle fibers.

A possible explanation for the inconsistency in LJ could be Bob Beamon's jump in 1968. This was the last year in the control period when Beamon jumped 8.90 meters at the Mexico City Olympics. This broke the previous world record by a historic 55cm. The largest margin of breaking the LJ world record previous to this was 15cm and Beamon's previous personal best was 8.33m (Francis, 124). It is believed that the altitude (7400 ft.) may have been a contributing factor to this historic day and may be one of the reasons why no one had come close to breaking this record for another 20 years. I ran the LJ data in Model 2 without Bob Beamon's jump from 1968. The results are shown in comparison to the original data in Table 4.

Table 4: Regression Results of Model 2 for The Long Jump without Bob Beamon (t-statistic in parentheses)

	Constant	YEAR	HEAVY_USE	POST	D-W	R^2
Original	-16.68 (-4.25)**	.026 (9.74)**	-.19 (-2.60)**	-.66 (-5.34)*	1.53	.81
w/o BB	-16.93 (-4.52)**	.023 (7.74)**	.14 (1.13)**	.092 (1.01)	1.97	.76

*** Significant at the .05 level**
**** Significant at the .01 level**

Clearly, excluding Beamon's jump from the regression resulted in more logical results. The HEAVY_USE and POST coefficients are positive which is consistent with expectations.

Also, the HEAVY_USE variable is significant at the .01 level. Long jump is another event that generally utilizes more fast-twitch muscle fibers than slow-twitch. Therefore, it would be expected that jumpers would benefit from steroids. Excluding Beamon's mark lowers the R^2 slightly but does improve the Durbin-Watson statistic.

Model 2a is the same as Model 2 only with median of the top 10 season best performances as the dependent variable instead of season best performance. Data in Table 5 contains the coefficients for YEAR, HEAVY_USE and POST variables along with the adjusted R^2 and adjusted Durbin-Watson statistics.

Table 5: Regression Results of Model 2a: Dependent Variable is Median of Top 10 Season Best Performances (t-statistic in parentheses)

	Constant	YEAR	HEAVY_USE	POST	D-W	R^2
100m	36.48 (15.54)**	-.013 (-11.06)**	-.082 (-2.55)*	.006 (.10)	1.68	.95
400m	152.83 (14.91)**	-.054 (-10.39)**	-.17 (-1.21)	.46 (1.94)	1.45	.92
Mile	10.52 (4.63)**	-.003 (-2.90)**	-.002 (-.07)	.001 (.03)	2.09	.24
5000m	54.54 (8.54)**	-.021 (-6.43)**	.033 (.51)	.03 (.31)	1.97	.55
LJ	-17.54 (-2.99)**	.013 (4.37)**	.17 (1.06)*	.06 (.58)	2.21	.37
SP	-52.48 (-1.91)**	.037 (1.27)**	.29 (.72)*	.43 (.73)	2.48	.07

*** Significant at the .05 level**
**** Significant at the .01 level**

The results of the Model 2a regression show that the YEAR coefficient is significant at the .01 level for all events. All events have the correct sign, as well. In the HEAVY_USE period, the 100m, LJ and SP coefficients were all significant at the .05 level and have the correct signs. This is consistent with my hypothesis since these events utilize fast twitch muscle fibers. The 400m and mile coefficients were insignificant while the 5000m coefficient had the incorrect sign. These results confirm the results found in Table 3 even though the coefficients tend to be lower and less significant than the results in Model 2. None of the coefficients during the POST period were significant and all of the running events had positive coefficients which is not consistent with expectations. Also, the R^2 for some of the events was lower in the Model 2a

regressions involving medians than in the Model 2 regressions involving season best performances.

d. Model 3 and Model 4 Regression Results

Models 3 and 4 attempted to determine whether there were any quadratic nonlinearities across the trend or within segments. The results of these models were not consistent with expectations. Many of the signs were incorrect and the quadratic terms were generally insignificant. Also, the results within models were not consistent enough across events to deem them as fitting the data properly. We were not able to find support for nonlinear relationships between time and performance. For this reason, we concluded that only Models 1 and 2 were useful for analysis. However, results for the Model 3 and Model 4 regressions are presented for completeness in Appendix 3.

VII. Conclusion

The most important conclusion to be drawn from this analysis is that unregulated steroid use during the heavy use period (HEAVY_USE) from 1969 to 1994 (1967 to 1994 for shot put) resulted in significant improvements in performance relative to trend improvements over the entire period (1949-2007). Second, these improvements were statistically most significant in those events with high proportions of fast twitch muscle fiber (e.g. 100m, 400m and shot put). This result was consistent with expectations drawn from biological research on muscular anatomy. During the post heavy use period (POST) there seemed to be a return to trend since most coefficients on POST were statistically insignificant. This suggests that the more effective testing implemented during the POST period may be having some success in controlling excessive use of steroids in recent years.

The regression results from Models 1 and 2 supported my original predictions. Model 1 illustrated that there is a very strong relationship between the year and season best performance. Taking into account the different time periods of steroid use, it was apparent that the heavy use period showed significant jumps in some events rather than others. Due to inconsistencies in the measuring units, we cannot compare coefficients across events to determine the most significant impact of steroids. Based on the t-statistic, however, shot put had the most significant change during the heavy use period after controlling for trend. The 100m and LJ (without Bob Beamon) were also significant at the .01 level during the heavy use period, although not as significant as SP. The 400m was significant at the .05 level indicating that there was a substantial decrease in performance times after controlling for trend. I expected the 100m and 400m to have the most significant results given the amount of fast-twitch muscle fibers sprinters possess. However, SP was the most significant despite the fact that most athletes in the event possess less fast-twitch muscle fibers than sprinters. This indicates that there may be another factor not accounted for when determining what leads steroids to be the most beneficial. Even though the results show a significant jump during the heavy use period for the 100m, 400m, LJ and SP, we cannot conclude with absolute certainty that these improvements are attributable to steroid use as improvements in other factors. The results of Models 1a and 2a using the median data were not as significant as the top performance data but did confirm my hypothesis.

One of the largest questions this study raises is the effectiveness of Anti-Doping Agencies. Although drug testing has improved greatly over the years and the results of my study seem to show a decrease in use, many athletes are still testing positive. However, adding more drug testing administrations would be costly and probably not effective since athletes find ways to beat the system. Only those who are willing to risk getting caught will take steroids, but

consequently have a better chance to turn in a world performance time. As of late, many athletes who have taken the risk have been caught. Marion Jones was the poster girl of United States track and field, held several records, won many championship titles and possessed five Olympic medals. After admitting to steroid use, she lost these awards, her dignity and the respect of her fans.

Works Cited

Anabolic Steroids in Sport and Exercise. Ed. Charles Yesalis. Champaign, IL: Human Kinetics, 2000.

Dare, Bernie. Running and Your Body: Applying Physiology to Track Training. Los Altos, CA: Tafnews Press, 1979.

Fort, Rodney. "The Value of Sports Talent." Sports Economics. New York, NY: Prentice Hall, 2002. 177-219.

Francis, Charlie, and Jeff Coplon. Speed Trap: Inside the Biggest Scandal in Olympic History. New York, NY: St. Martin's Press, 1990.

Haugen, Kjetil. The Performance-Enhancing Drug Game. *Journal of Sports Economics*, Vol. 5 (2004), pp. 67-86.

Hersh, Philip. "On Fast Track to Infamy." Chicago Tribune 6 Oct. 2007, sec. B: 4+.

International Association of Athletics Federations. IAAF Athletics. 17 Nov. 2007. 11 Oct. 2007 <http://iaaf.org>.

Medic, Nikola, Janet Starkes and Bradley Young. Examining Relative Age Effects on Performance Achievement and Participation Rates. *Journal of Sports Sciences,* Vol. 25, No. 12 (Oct. 2007), pp. 1377-1384.

Pindyck, Robert S., and Daniel L. Rubinfield. Microeconomics. 6th ed. New Dehli: Prentice Hall, 2005. 189+.

Schulz, Nick. "Sprinters Love Athens' Mondo Track." Slate Magazine. 29 Aug 2004, pg 34.

Schutz, Robert, and Liu Yuanlong. "Measurement in Physical Education and Exercise

Science: Prediction Models for Track and Field Performance." (1998).

Shanks, Michael. "Timeline of Nike Shoes." Stanford: Nike Sneaker. Jan 2006.

"Superhuman Heroes." The Economist 347 (1998): 10-13.

Taylor, William. Anabolic Steroids and the Athlete. Jefferson, NC: McFarland and Company, Inc., 2002.

"World Best Lists." Track and Field News Jan. 1950-Jan. 2007.

Wilson, Stephen. "IOC Strips Medals from Jones' Teammates." Chicago Tribune. 10 Apr. 2008, sec. B: 5.

Zak, Thomas and Cliff Huang. Production Efficiency: The Case of Professional Basketball. *The Journal of Business,* Vol. 52, No. 3 (Jul., 1979), pp. 379-392.

Appendix 1

Slow-twitch Muscle Fiber Percentages for Selected Events

Event	Event Group Representation	Slow-twitch Muscle Fiber %
100m	Short Sprints	8-24%
400m	Long Sprints	8-24%
Mile	Middle Distance	40-65%
5000m	Distance	80+%
Long Jump	Jumps	40-50%
Shot Put	Throws	40-50%

Appendix 2

Selected Current World Records- provided by IAAF

Event	**Time/Distance***	**Athlete**	**Country**	**Date of Record**
100m	9.74	Asafa Powell	Jamaica	9-9-2007
400m	43.18	Michael Johnson	USA	8-26-1999
Mile	3:43.13	Hicham El Guerroui	Morocco	7-7-1999
5000m	12:37.35	Kenenisa Bekele	Ethopia	5-31-2004
Long Jump	8.95m	Mike Powell	USA	8-30-1991
Shot Put	23.13m	Randy Barnes	USA	5-20-1990

*Times are in mm:ss.0

Appendix 3

Regression Results of Model 3: Dependent Variable is Season Best Performance (t-statistic in parentheses)

	YEAR	HEAVY_USE	POST	$YEAR^2$	D-W	R^2
100m	-.01 (-3.96)**	-.08 (-1.51)	-.03 (-.36)	-.003 (-.55)	2.01	.92
400m	-.11 (-5.60)**	.15 (.66)	.008 (.02)	.001 (.26)	1.64	.83
Mile	-.02 (-9.61)**	.07 (3.22)	.004 (.11)	.000 (7.23)*	2.10	.85
5000m	-.05 (-7.76)**	.07 (1.23)	-.02 (-.32)	.001 (1.21)	1.85	.64
LJ	.04 (8.55)**	-.28 (-4.31)*	-.05 (-.41)	.000 (-5.06)*	1.71	.86
SP	.22 (7.80)**	.19 (.55)	.24 (.46)	-.003 (-5.68)**	1.42	.89

*** Significant at the .05 level**

****Significant at the .01 level**

Regression Results of Model 4: Dependent Variable is Season Best Performance (t-statistic in parentheses)

	YEAR	HEAVY_USE	POST	$YEAR^2$*HEAVY_USE	$YEAR^2$*POST
100m	-.01 (-3.63)**	-.06 (-.61)	-.18 (-1.22)	-.007 (-1.51)	.001 (.19)
400m	-.08 (.5.94)**	-.03 (-.13)	-.34 (-.76)	.001 (2.31)**	.000 (.76)
Mile	-.01 (-8.72)**	.02 (.85)	-.13 (-.54)	.000 (.19)	.000 (.77)
5000m	-.03 (-7.54)**	-.07 (-1.21)	-.38 (-.43)	.000 (.99)	.000 (.57)
LJ	.04 (9.17)**	-.22 (-3.41)**	-.07 (-.32)	.000 (-2.81)**	.000 (-4.57)**
SP	.22 (9.89)**	.21 (.51)	-1.65 (-.61)	-.001 (-.70)	-.002 (-4.12)**

	D-W	R^2
100m	2.10	.93
400m	1.74	.85
Mile	1.96	.84
5000m	1.41	.81
LJ	1.78	.87
SP	1.79	.91

***Significant at the .05 level**

****Significant at the .01 level**