**Effect of Personal Relationships on MLB Athletes’ Performance**

Connor Johnson

**Abstract**

Being a professional athlete is one of the most physically and mentally demanding job that a person can have. Because of this, many things in an athlete’s personal life can affect their performance. This study looks at the effect of getting married and having children on an athlete’s stats, specifically the batting average of baseball players. Using the averages of 117 athletes, two statistical models were used to analyze the short- and long-term effects of these decisions. The data was controlled for age as older players tend to be married and have more kids. The data showed a small, temporary effect on players’ batting average after having his first child. The magnitude of this change was found to be a .01 drop in batting average in the season after the first child was born. There was no significant change found for marriage or children after the first. This change is small, but it does show that an athlete’s personal life can have negative effects in their careers. Implications, limitations, and potential future studies on this topic are discussed in the conclusion.

**Introduction**

Getting married and/or having kids are some of the biggest decisions a person can make in their life. These decisions have long lasting effects on the person’s life. Since personal relationships like these require time and commitment, they can often pull people away from other things like their career, especially for women. Studies show that women often see their salaries stagnate after having kids. This is more likely due to the fact that they spend more time away from work than their performance in their careers worsening. This study is going to analyze the effect of personal relationships on male athletes. Since an athlete must commit much more time and energy to their career compared to the average person, it is likely that personal relationships could have a larger effect on their performance. The primary research questions in this study are as follows:

* Does getting married affect a male professional baseball player’s performance?
* Does having children affect a male professional baseball player’s performance?

The original idea for this study came from a comment by NHL commentator Mike Milbury. He said that NHL players were performing better while isolated during the COVID-19 pandemic because they have “no women to distract them.” This study will attempt to answer the question of whether person relationships negatively affect performance even though relationships tend to be a positive “distraction” from their careers. This question is important because people should understand the overall effects of these types of decisions, especially those in careers that require a lot of commitment. Most studies that have been done on this topic have focused on the average person’s career and income instead of performance. This study will provide more insight on the topic by concentrating on actual performance in a much more intensive career.

**Literature Review**

While looking for past research on this topic, I was only able to find one study that specifically analyzed the effect of personal relationships on athletes’ performance. So, I decided to look more generally in relationships effect on one’s career in the workforce since a sport at the major league level is a career for athletes. I found three papers that may be relevant towards the topic of this one.

**The Role of Romantic Relationships on Athletes’ Performance and Well Being**

The goal of this study was to find evidence of spillover in athletic performance from interference with maintaining a romantic relationship (Jowett and Cramer pg. 58). It also analyzed the effects on the athletes’ level of depression. The method of study involved 4 antecedents: interpersonal trust, interpersonal commitment, hostile interactions, and communication quantity. These were then used to predict the athletes’ sport satisfaction and depression (Jowett And Cramer pg. 62). The study was done on 87 athletes male and female athletes among many different sports (Jowett and Cramer pg. 62). All the variables were measured using surveys that involved questions like “I am satisfied with the improvement in my skill level thus far” and a response scale range to represent their satisfaction (Jowett and Cramer pg. 63). The study found that negative factors of a relationship like hostile interactions cause lower satisfaction in sports performance. They also found that higher levels of commitment also lowered performance satisfaction suggesting that relationships may pull athletes away from their sports (Jowett and Carter pgs. 65-67). Some problems with this study are that the sample size is small and the data is based on very subjective surveys (Jowett and Carter pgs. 68-69).

**The Motherhood Penalty at Midlife: Long-Term Effects of Children on Women’s Careers**

This was a study done to determine how having children affects the careers of women. Unlike past studies in this field which have been on younger women, this study also looked at woman up to 54 and the long-term effect of having children (Kahn, Garcia-Mangalo, and Bianchi). The method of study was interviewing 5159 women multiple times and measure labor force participation, hourly wages, and occupational status throughout their career (Kahn, Garcia-Mangalo, and Bianchi). The values of these stats were measured throughout the women’s careers so changes could be seen before and after the women had children. The changes in the wage, status, and participation would then show the effects of children on a woman’s career. Using this, the authors concluded that having children does have a temporary negative effect on their career, that is reduced as the woman gets older (Kahn, Garcia-Mangalo, and Bianchi). The two main limitations of the study are some of the data could no longer be socially relevant since it’s taken from an old source and that the data doesn’t consider a woman’s change in preferences for her career throughout her lifetime (Kahn, Garcia-Mangalo, and Bianchi).

**The Fatherhood Bonus and The Motherhood Penalty**

The goal of this report was to find if having children is the cause of the current wage gap in the country. That is, how are careers and income affected by children based on gender (Budig pgs. 6-7). Using the National Longitude Survey of Youth 1979, Budig analyzed the effects of children on wages from 1979-2006 (Budig 9). Changes in men and women’s wages were measured when they had children. From this data, it was found that fathers get a 6 percent wage increase from children on average and women lose 4 percent of their wages per child (Budig pgs. 9-17). It is concluded in the report that this difference is more likely due to social stigmas around each gender having kids instead of actual changes in performance.

**Compare and Contrast**

The second and the third articles are the most similar. They both analyze the effect of wages and careers for parents after they have kids. Article 2 concentrates on only women while Article 3 analyzes the effects for both men and women. Since Article 2 uses more dependent variables in its analysis of careers over a greater period of time, it can be argued that the conclusion is more likely to be true. Since Article 3 doesn’t analyze the wage effect as long Article 2, the conclusion that the effects of children diminish over time is not found. For both of these articles, not much can be concluded about the actual performance changes as a result of having children because income isn’t necessarily an accurate representation of work performance. The do however make the assumption that the change in wage is likely not purely performance based.

Article 1 is much different than the other two because it is based on athletes and doesn’t consider income. This study goes more into how athletes view their own performance based off their personal relationships. While quantitative analysis is used to make a conclusion, surveys that are qualitative in nature are used to get the data. Because of this, the conclusion made in this study is weaker than those in Articles 2 and 3. In general, data based on people’s qualitative view provides much weaker evidence than quantitative data like income. This study also did not consider time in its data. The surveys were given to the athletes only once. Because of this, there are many more confounding variables that could affect the data. Overall, this study is the most similar to what I plan on doing because it concentrates more on the actual performance changes that come from personal relationships, but its data is much weaker compare to the other two articles.

**Methodology**

This section includes information on the sample and how the data will be gathered as well as the statistical methods used to analyze it

**Sample**

The sample of this study will include players in the MLB. Four starters from each team with families will be randomly selected to give a sample size of 120 players. Because of injuries and lack of public information, some players had to be removed so the final sample size is 117. The following variables will be collected about each player including dates.

Variables

* Batting average: quantitative variable that will be split into categories

Categories of Variables

* Pre-marriage: categorical variable for the above variables before the player is married
* Post-marriage: categorical variable for the above variables after the player is married
* No Kids: categorical variable for the above variables while the play has no kids
* One Kids: categorical variable for the above variables while the play has one kid
* Two Kids: categorical variable for the above variables while the play has two kids

The batting average of the players for each year they played is publicly provided by the MLB. The data will specifically be coming from baseball-refernce.com which provides accurate stats on players which was obtained from the MLB. Information on the marriage date and birthdates of athletes’ children can also be easily found on online bios and social media accounts. Instagram and Twitter have been found to be the best source for finding this information so these two social media sites will be used the most. Since anyone can post this data, more research needs to be done to verify the accuracy.

**Method**

The study analyzes the change of variables among the categories provided. There will be two methods of analysis using this data. The first will be a simple analysis of the difference in batting average between the seasons before and after the events described above. The second analysis will be an ANACOVA model using the batting averages of players in 2019. Players will be split into groups based on the categorical variables. Analysis of the data will then show if there is a significant difference among these groups of players.

**Analysis of Difference in Batting Average**

The first analysis will be a simple t-test on the means of the change in batting average that resulted from marriage, 1st child, and the 2nd child. This data was found by finding the difference in the averages of players in the MLB between the season before and after each of these three events occurred. The null hypothesis of the test will be that the mean is equal to zero. Since it was predicted earlier that the batting averages will decrease, the alternate hypothesis is the mean is less than zero.

Age will also be checked as a confounding variable on the response. This will be tested using a linear regression shown below.

Y = β0 + β1\*Age

* β0 – constant for the regression
* β1 – change in the response variable as age increases by one

The null hypothesis for this test is β1 = 0 and the alternate hypothesis is β1 ≠ 0. If the null hypothesis is false, then age will need to be considered as a confounding variable above in showing if the difference is significant. If age isn’t a significant predictor, then it won’t need to be included in the above model.

**ANOVA Regression**

The second analysis to be done is an ANOVA regression using the batting averages of all players from the sample in 2019. This will give more info on the long-term effects of the independent variables being studied. The analysis above only used data from the season before and after the players developed a new relationship. Because of this, the results only show what immediate effects their decisions may have. In this analysis, the players in the sample will be split into the following groups: Not married, married, no kids, one kid, two or more kids. Since these groups have multiple overlaps, they will be analyzed in pairs. The pairs of groups that will be tested are Married/Not Married, One Kid/No Kids, Two or More Kids/No Kids, Two or More Kids. The regression equation used for this is below.

Y = β0 + β1\*X1 + β2\*Age

* X1 – dummy variable that represents which category the player is in
* β0 – constant for the regression
* β1 – difference in the mean batting average between the two groups
* β2 – change in the response variable as the player’s age is increased by one

Since the goal of this analysis is to look for a difference between the groups being looked at, β1 is the variable that will be tested. The null hypothesis is β1 = 0 and the alternate hypothesis is β1 ≠ 0. A p-value that is less than 0.05 will show a statistically significant difference. Age acts as a control variable to avoid it confounding with the data so β2 shouldn’t be a significant coefficient. The significance of β1 will show us how significantly different the groups are.

**­Results**

This section contains the results of both the models used to analyze the data collected.

**Analysis of Difference in Batting Average**

The sample for this model has 54 players in the marriage sample, 48 players in the first child sample, and 32 players in the second child sample. This sample is smaller because this model doesn’t include the players that aren’t married and don’t have children. This model also requires the exact date of the marriage and children’s birth which couldn’t be obtained for some of the players. Because of this, the sample size is much smaller.

The first model is the regression that tests if age is a significant predictor in the change in batting average. All three regression models are shown below.

**Figure 1**

*Linear Regressions of each response variable versus Age of the player.*

**Coefficients**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Term** | **Coef** | **SE Coef** | **T-Value** | **P-Value** | **VIF** |
| Constant | -0.0475 | 0.0612 | -0.78 | 0.441 |  |
| Age (Marriage) | 0.00167 | 0.00236 | 0.71 | 0.482 | 1.00 |
|  |  |  |  |  |  |

**Coefficients**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Term** | **Coef** | **SE Coef** | **T-Value** | **P-Value** | **VIF** |
| Constant | 0.0065 | 0.0400 | 0.16 | 0.871 |  |
| Age (1st Child) | -0.00061 | 0.00149 | -0.41 | 0.682 | 1.00 |
|  |  |  |  |  |  |

**Coefficients**

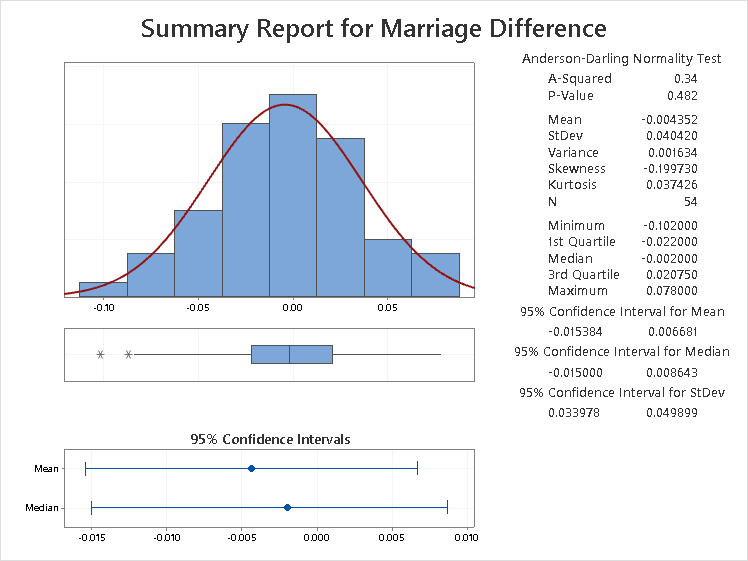
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Term** | **Coef** | **SE Coef** | **T-Value** | **P-Value** | **VIF** |
| Constant | -0.0481 | 0.0488 | -0.99 | 0.332 |  |
| Age (2nd Child) | 0.00159 | 0.00169 | 0.94 | 0.356 | 1.00 |
|  |  |  |  |  |  |

The p-value for all three variables is much greater than 0.2. Because of this, we can assume that that Age does not have a significant effect on the response variable and does not need to be considered in the model.

Now, the difference in batting average for all three tests can be analyzed for its significance. The figures below show the distribution and descriptive statistics of the three.

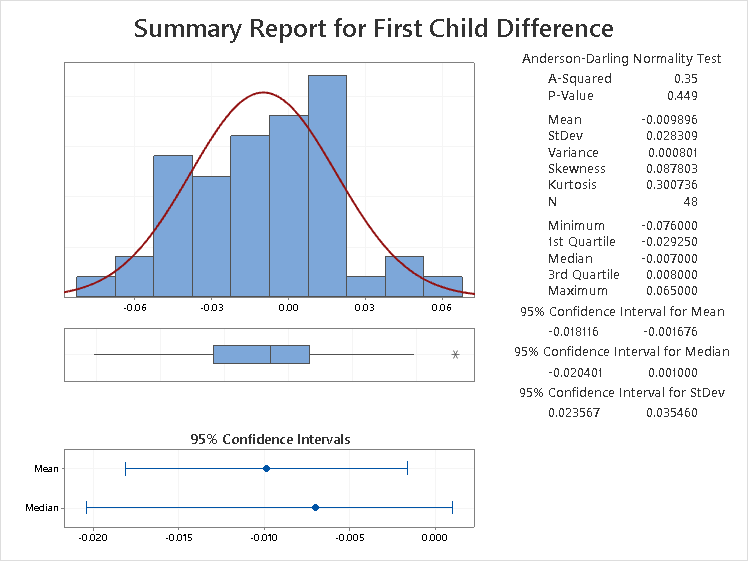
**Table 1**

*Histogram and descriptive statistics of the Marriage Difference response variable*



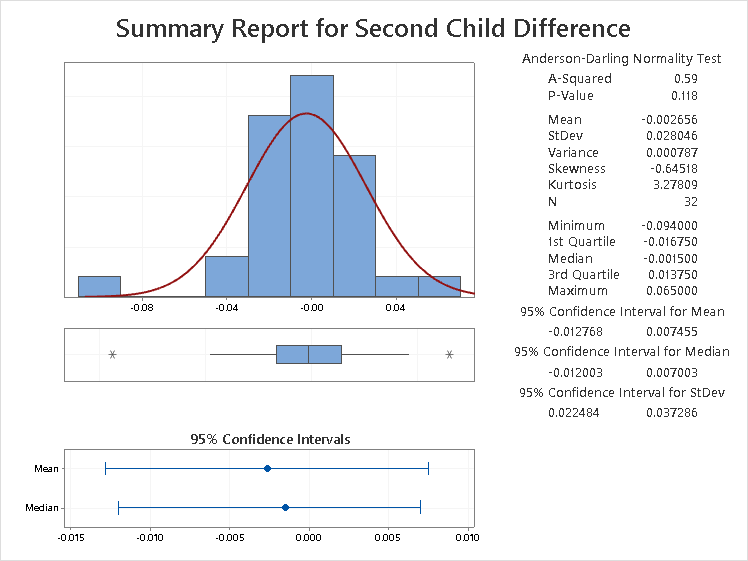
**Table 2**

*Histogram and descriptive statistics of the First Child Difference response variable*



**Table 3**

*Histogram and descriptive statistics of the First Child Difference response variable*



From figures 2, 3, and 4, it can be seen that the distributions of all three variables are relatively normal. So, the t-test statistic can be used to show if they are significantly less than zero. These tests are shown below.

**Figure 2**

*Test Statistics and p-values for all response variables.*

**Marriage Dif. 1st Child Dif. 2nd Child Dif.**

|  |  |  |  |
| --- | --- | --- | --- |
| Null hypothesis | | | H₀: μ = 0 |
| Alternative hypothesis | | | H₁: μ < 0 |
| **T-Value** | **P-Value** |
| -0.54 | 0.298 |

|  |  |  |  |
| --- | --- | --- | --- |
| Null hypothesis | | | H₀: μ = 0 |
| Alternative hypothesis | | | H₁: μ < 0 |
| **T-Value** | **P-Value** |
| -0.79 | 0.216 |

|  |  |  |  |
| --- | --- | --- | --- |
| Null hypothesis | | | H₀: μ = 0 |
| Alternative hypothesis | | | H₁: μ < 0 |
| **T-Value** | **P-Value** |
| -2.42 | 0.010 |

Since we’re looking for 95% confidence in the test, the p-value must be less than 0.05 for the mean of the response variable to be significantly less than zero. As seen in Figure 5, the First Child Difference is the only variable to have a p-value less than 0.05 with p = 0.010. Thus, the null hypothesis is rejected and the mean is presumed to be less than zero. This tells us that the first child of a baseball player has an immediate effect on their performance. This effect is small as shown by the data, but still statistically significant. For the marriage response variable, the T-value is -0.79 with p = 0.216. For the second child variable, the T-value is -0.54 and p = 0.298. Because the values are both much greater than 0.05, the null hypothesis can’t be rejected meaning there isn’t enough evidence to show the difference is less than zero.

**ANOVA Regression**

In this model, all 117 players are sorted into groups depending on the categorical variables described earlier. An ANOVA regression is then calculated to check if there is a significant difference between the means of this group with Age used as a control. The models are shown in the figures below.

**Figure 3**

*Regression and Analysis of the Married/Not Married groups of players*

**Regression Equation**

|  |  |  |
| --- | --- | --- |
| Average | = | 0.2762- 0.00077 Married - 0.00056 Age |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Age | 1 | 0.000345 | 0.000345 | 0.30 | 0.586 |
| Married | 1 | 0.000012 | 0.000012 | 0.01 | 0.919 |
| Error | 113 | 0.130467 | 0.001155 |  |  |
| Lack-of-Fit | 24 | 0.019788 | 0.000824 | 0.66 | 0.874 |
| Pure Error | 89 | 0.110680 | 0.001244 |  |  |
| Total | 115 | 0.131011 |  |  |  |

**Figure 4**

*Regression and Analysis of the One Kid/No Kids groups of players.*

**Regression Equation**

|  |  |  |
| --- | --- | --- |
| Average | = | 0.2679 - 0.00559 One Kid\_1 - 0.00020 Age |

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Age | 1 | 0.000039 | 0.000039 | 0.03 | 0.854 |
| One Kid | 1 | 0.000647 | 0.000647 | 0.56 | 0.455 |
| Error | 113 | 0.129833 | 0.001149 |  |  |
| Lack-of-Fit | 24 | 0.034348 | 0.001431 | 1.33 | 0.167 |
| Pure Error | 89 | 0.095484 | 0.001073 |  |  |
| Total | 115 | 0.131011 |  |  |  |
|  |  |  |  |  |  |

**Figure 5**

*Regression and Analysis of the Two or More Kids/No Kids groups of players.*

|  |  |  |
| --- | --- | --- |
| Average | = | 0.2660 - 0.00728 Two or More Kids - 0.00015 Age |

**Regression Equation**

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Age | 1 | 0.000024 | 0.000024 | 0.02 | 0.884 |
| Two or More Kids | 1 | 0.000825 | 0.000825 | 0.72 | 0.398 |
| Error | 113 | 0.129654 | 0.001147 |  |  |
| Lack-of-Fit | 23 | 0.026445 | 0.001150 | 1.00 | 0.471 |
| Pure Error | 90 | 0.103209 | 0.001147 |  |  |
| Total | 115 | 0.131011 |  |  |  |

**Figure 6**

*Regression and Analysis of the Two or More Kids/No Kids groups of players.*

|  |  |  |
| --- | --- | --- |
| Average\_1 | = | 0.3157 - 0.0006 Two or More Kids - 0.00197 Age |

**Regression Equation**

**Analysis of Variance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Age\_1 | 1 | 0.001748 | 0.001748 | 1.23 | 0.273 |
| Two or More Kids\_1 | 1 | 0.000004 | 0.000004 | 0.00 | 0.959 |
| Error | 48 | 0.068253 | 0.001422 |  |  |
| Lack-of-Fit | 20 | 0.035158 | 0.001758 | 1.49 | 0.164 |
| Pure Error | 28 | 0.033095 | 0.001182 |  |  |
| Total | 50 | 0.070381 |  |  |  |

*Note:* This regression doesn’t include players with no kids so N = 51 instead of N=117.

***Married/Not Married***

The regression shows a drop in average for players that are married. Age has a p-value p = 0.586 so it isn’t significant enough to be a major confounding variable. The p-value of the marriage coefficient is p = 0.919. A p-value this high shows the null hypothesis cannot be rejected. There isn’t enough evidence to show a difference in batting average between the groups.

***One Kid/No Kids***

The regression shows players with one kid having a lower batting average. The F-value for the age coefficient is F = 0.03 with a p-value of p = 0.854. So, age doesn’t need to be considered as a confounding variable. The F-value for the One Child coefficient is F = 0.56 with a p-value p = 0.455. The data for these groups is more significant than the marriage groups, but not significant enough to reject the null hypothesis. It can’t be concluded that the difference between the means of the two groups is significantly less than zero.

***Two or More Kids/No Kids***

This regression also shows a negative difference between the two groups. The age coefficient has an F-value F = 0.02 and p-value p = 0.884. So, age isn’t significant enough to be considered. The F-value of the Kids coefficient is F=0.72 with p-value p=0.398. This is more significant than the last two analyses but most of this likely comes from the One Kid factor. However, this p-value is too small to suggest that the difference is significant.

***Two or More Kids/One Kid***

Finally, this last regression also shows a negative difference between the groups. Age has F-value F = 1.23 and p-value p = 0.273. This is more significant than the other regressions but still too weak to be considered. The coefficient for the difference between the groups has F-value F = 0.00 and p-value .959. This is an extremely high p-value which shows almost no difference between the groups. The null hypothesis can’t be rejected and the means of the groups can’t be assumed to be different.

Out of all the methods used to analyze the changes in players’ batting averages, the only statistically significant method was the One Kid Difference response variable. The ANOVA methods all showed statistically insignificant results. With a p-value of p = 0.01, the difference in batting average in the season before and after having a first kid is the only significant response variable with a mean difference of -.0099.

**Conclusion**

The initial questions of this paper were whether personal relationships such as marriage or children had an effect on baseball players’ performance. This was checked using their batting average as a representation of their performance. The data shows that the first kid an athlete has is the only relationship studied that may have an impact on his performance.

**Interpretation of Results**

The first method of analyzing the data showed the change in batting average in the season before and after an athlete got married/had kids. The change in batting average was only significant for when an athlete had his first child. The change was a drop of 0.01 or 1 fewer hit for every 100 at bats. While this change is small, it is statistically significant. The ANOVA method did not return any significant results. Because this method used data that was longer after the athletes’ started these relationships, it more accurately shows the long-term effects. Since the data shows nothing significant, it can’t be concluded that marriage or children have long term effects on baseball players’ performance. The data and two methods of this study show the first kid a player has is likely to have a minimal negative and temporary effect on his performance.

**Implications**

Since the results of the study don’t show any large effects, the implications aren’t too large. It does show to players that the decision to have children can affect their career performance. In my opinion, the effect is too small and short-lived to be an important factor in an athletes’ decision. Since baseball isn’t a very physically demanding sport, it is possible that the impact is larger in other sports. It’s also possible that this same effect could be seen in other demanding careers outside of sports. Overall, this study does show that seemingly harmless decisions athletes make outside of their careers can have impacts on how they play, even if they are small.

**Limitations**

The main limitation of this study is that the data taken for batting averages is only taken from individual seasons than throughout the athletes’ careers. Limited by time, I wasn’t able t collect more data to use and gather more accurate results. Another big limitation is that batting average may not be the most accurate predictor of performance. There are many players in the MLB that are defensive players so their hitting doesn’t fully represent how they are playing. There are more batting statistics like on-base percentage and OPS (on-base plus slugging) that weren’t considered in this study. Including these stats as well could give much more insight to the overall performance of each player. Finally, baseball isn’t the most demanding sport that athletes play. It was picked because I thought the stats in baseball better represent the players performance compared to other sports where many factors in the game can affect stats. It is likely that the effects seen in this study could be greater in more physically demanding sports like football and basketball. With these limitations, there is definitely more research that can be done to get more accurate results on the effects on athletes’ performance.

**Future Research**

Based on the limitations just described, I would suggest a similar study to be done with more data collection. Using data on stats throughout a player’s career, it could be shown more n depth the exact effects of personal relationships. I’s also suggest using more stats so that he performance of the players can be quantitatively represented more accurately. Simply because of time restraint, I was not able to do this in the study, but I believe it could bring about much more interesting results. I also think the results found in this study would definitely be larger in a sport that is more demanding than baseball. A study using similar methods in other sports could bring insight to that.

This study investigated the effect that personal relationships can have on athletes’ performance. It was shown that baseball players’ have a small drop in batting average after their first child. Because of the limitations of this study and the small size of the effect found, I don’t think it’s enough evidence for athletes to start changing their life choices. It does show that there is more to be investigated on this topic. The results of this study show there could be a larger negative effect that needs to be looked at. More research on the topic could greatly benefit the lives of professional athletes by giving them more information about the personal decisions they make.

**References**

Budig, Michelle1 J. “The Fatherhood Bonus and The Motherhood Penalty: Parenthood and the Gender Gap in Pay – Third Way.” *Third Way*, 2014, www.thirdway.org/report/the-fatherhood-bonus-and-the-motherhood-penalty-parenthood-and-the-gender-gap-in-pay.

Jowett, Sophia & Cramer, Duncan. (2009). The Role of Romantic Relationships on Athletes' Performance and Well-Being. Journal of Clinical Sports Psychology. 3. 58-72. 10.1123/jcsp.3.1.58.

Kahn, Joan & García-Manglano, Javier & Bianchi, Suzanne. (2014). The Motherhood Penalty at Midlife: Long-Term Effects of Children on Women's Careers. Journal of Marriage and Family. 76. 56-72. 10.1111/jomf.12086.

“MLB Stats, Scores, History, & Records.” *Baseball*, 2020, www.baseball-reference.com/.