Group 13 BIOS650 Project

Aonan Chen, Malcolm Risk, Xiaochun Shao, Xuheng Qiang, Nicolas Gargurevich, Hua Jiang

December 7, 2021

The Effect of Physical Activity on Testosterone

Abstract

Background: Testosterone and physical activity each play important roles in human health. However, previous studies are mixed as to whether physical activity is associated with increased testosterone. Our goal was to build a single model and clarify the true relationship adjusting for the most important confounders and covariates in previous studies. **Methods:** Our study is a secondary analysis of the 2011 and 2012 National Health and Nutrition Examination Survey (NHANES) data with a sample size of 2788. We compared physically active individuals with inactive individuals by sex, age, obesity status (BMI), race, smoking status, alcohol intake, and testosterone levels (ng/dl). A general linear regression was used to quantify the relationship between testosterone and physical activity as well as interactions among covariates. Model diagnostics were conducted using residual and influence plots. **Results:** Except gender and testosterone level, all other baseline statistics differred between active and inactive groups with significant p-values ($p<0.001$). Sequential models of testosterone with physical activity using univariate linear regression, multiple linear regression without interaction, and multiple linear regression with interaction terms all show insignificant p-values ($p=0.104$, $p=0.130$, $p=0.311$) for the physical activity term. Regression diagnostics show that the final model (including interaction terms) satisfied linearity, independence, and normality assumptions, but violated the equal variance assumption. A robust standard error method was conducted to refit the model to address this problem and both the interaction terms and main effect for physical activity were significant ($p=0.012$ for main effect of physical activity, $p=0.030$ for interaction between age and physical activity, $p=0.002$ for interaction between gender and physical activity). Influence diagnostics show no influential points exist. **Conclusion:** The study's finding indicates relationship between testosterone and physical activity, which helps us understand the two health factors better and raises people's awareness of being active.

Background

Testosterone is one of the most significant sex hormones and is the primary sex hormone in men. Low testosterone levels in men have been associated with low energy, poor concentration, depression, low libido, and erectile dysfunction [1]. The health effects of low testosterone in women are not as well-documented, but there are associations between testosterone and favourable cardiovascular effects as well as improved musculoskeletal health [2]. Testosterone levels decrease with age and so low testosterone levels are a particular concern in older adults [3].

Some research has associated increased physical activity with higher testosterone levels in men [4]. However, a 2016 study on the 1999-2004 NHANES dataset found no overall association between physical activity and testosterone levels [5]. In studies conducted exclusively on women, no association between physical activity and testosterone has been found [6].

Our goal was to use the 2011-12 NHANES data to assess whether there was an association between higher physical activity and testosterone levels, while controlling for confounding factors such as age, gender, weight, alcohol intake, and smoking status. We also set out to determine whether the effect of physical activity on testosterone differed based on gender and age.

Methods

Study Population

The 2009-10 NHANES dataset did not include data on testosterone, so testosterone levels are only reported for 4,126 patients in the 2011-12 dataset. We removed 1338 patients missing data on either physical activity, gender, age, alcohol intake, or BMI, and 2788 patients were included in the final analysis. Characteristics of the study population are shown in Table 1.

Table 1: Study Population Characteristics

Characteristic[a]	**All Subjects (n=2788)**	**Physically Active (n=1571)**	**Not Active (n=1217)**	**P-Value** [b]
Male	1511 (54.2%)	852 (54.2%)	659 (54.1%)	1
Age (y)	46.4 (17.2)	43.5 (16.7)	50 (17)	<0.001
BMI	28.6 (6.3)	27.7 (5.7)	29.6 (6.9)	<0.001
Smoker[c]	561 (20.1%)	210 (13.4%)	351 (28.8%)	<0.001
Race	-	-	-	<0.001
White	1969 (70.6%)	1164 (74.1%)	805 (66.1%)	-
Black	275 (9.9%)	139 (8.8%)	136 (11.2%)	-
Other	544 (19.5%)	268 (17.1%)	276 (22.7%)	-
Testosterone (ng per dl)	233.5 (233.7)	239.8 (235.1)	225.3 (231.8)	0.1
Among males	410.4 (179.2)	421.4 (171.3)	396.1 (188)	-
Among females	24.1 (20.3)	24.6 (23.2)	23.6 (15.9)	-
Alcohol Intake[d]	74.9 (101.9)	86.2 (103.6)	60.3 (97.7)	<0.001

[a] mean (sd) for continuous variables, n (%) for categorical

[b] Chi-square test for correlation for categorical variables, t-test for continuous variables

[c] Refers to participants who currently smoke and have smoked at least 100 cigarettes in their lifetime.

[d] Number of days drinking alcohol over the past year.

To account for potential bias in the data, we compared adult subjects with complete data to subjects with incomplete data. There were significant differences between complete and incomplete cases in terms of gender, testosterone level, and race but complete and incomplete cases were similar in BMI, alcohol intake, smoking status, and age. Bias is not a major concern as our analysis controls for gender and race, and complete and incomplete cases had similar testosterone levels when comparing within each gender.

Models

Variables *Predictor(s) of Interest* PhysActive is a binary variable which answers the survey question whether a participant does moderate or vigorous-intensity sports, fitness or recreational activities.

Outcome(s) Testosterone is a continuous variable which measures the total testosterone in ng/dL.

Covariates For adjusted analysis, we included demographic variables (i.e. age, gender, and race), BMI, smoking status, and alcohol intake. Previous research has shown that age affects physical activity [7] and aging is associated with declining testosterone [3]. Therefore, we included age (continuous, years) as a confounder. As testosterone is found at greater concentrations in men [9] and the other previous studies have shown that men were more active than women, we also included gender as a confounder. A 2006 study showed that obesity may affect testosterone levels[10]; The relationship between physical activity and body mass index is unclear among non-obese individuals[14]. Therefore, we choose to introduce BMI to test obesity as a precision variable. As the result of a study about race, there is variation in testosterone levels across different racial groups[11] and we introduce race as an another categorical precision variable (i.e., Black, White, Other) to our model. A report about smoking and testosterone levels found that smoking negatively affects testosterone levels[12] and then we include current smoking status (categorical, yes/no) as a precision variable. We also chose AlcoholYear (The estimated number of days intaking alcohol over the past year) as a precision variable, as some studies reports alcohol intake in low amounts can increase testosterone levels[13].

Statistical Analyses The association between physical activity and testosterone was analyzed using multiple linear regression models. The fitted models were fitted sequentially as follows:

1) Unadjusted analysis
2) Adjusted analysis without interaction
3) Adjusted analysis with interactions
    - As a similar study included interaction between age group and physical activity, we also included this interaction in our analysis
    - We included interaction between gender and physical activity as we are interested in the adjusted effect of physical activity on testosterone for males and females.
    - We excluded alcohol intake from our model for two reasons: 1) the p-value is not significant 2) the days of alcohol consumption in the past year might not properly account for the actual values for alcohol intake and the results of previous studies were sensitive to the exact quantity of alcohol consumed
4) Refitting the Model with Robust standard Error after diagnostic analysis revealed a violation of the equal variance assumption.

For all models, we used a significance level of 0.05 to assess associations for each regression term.

Results

From crude analysis, we found that there was no statistically significant association between high physical activity and testosterone levels (p = 0.104). After adjusting for all confounders and precision variables, there was still no statistically significant association between high physical activity and testosterone levels (p = 0.13). However, this conclusion could be invalid due to violation of constant variance assumption. Once we refit the model with robust standard errors, our p-value slightly increased to 0.133 but there was no impact on our conclusion. From our third model, we found that the equal variance assumption is also violated. Therefore, we refit the model with robust standard errors. The main effect of physical activity is now significant associated with testosterone levels with a p-value of 0.0116.

Diagnostics

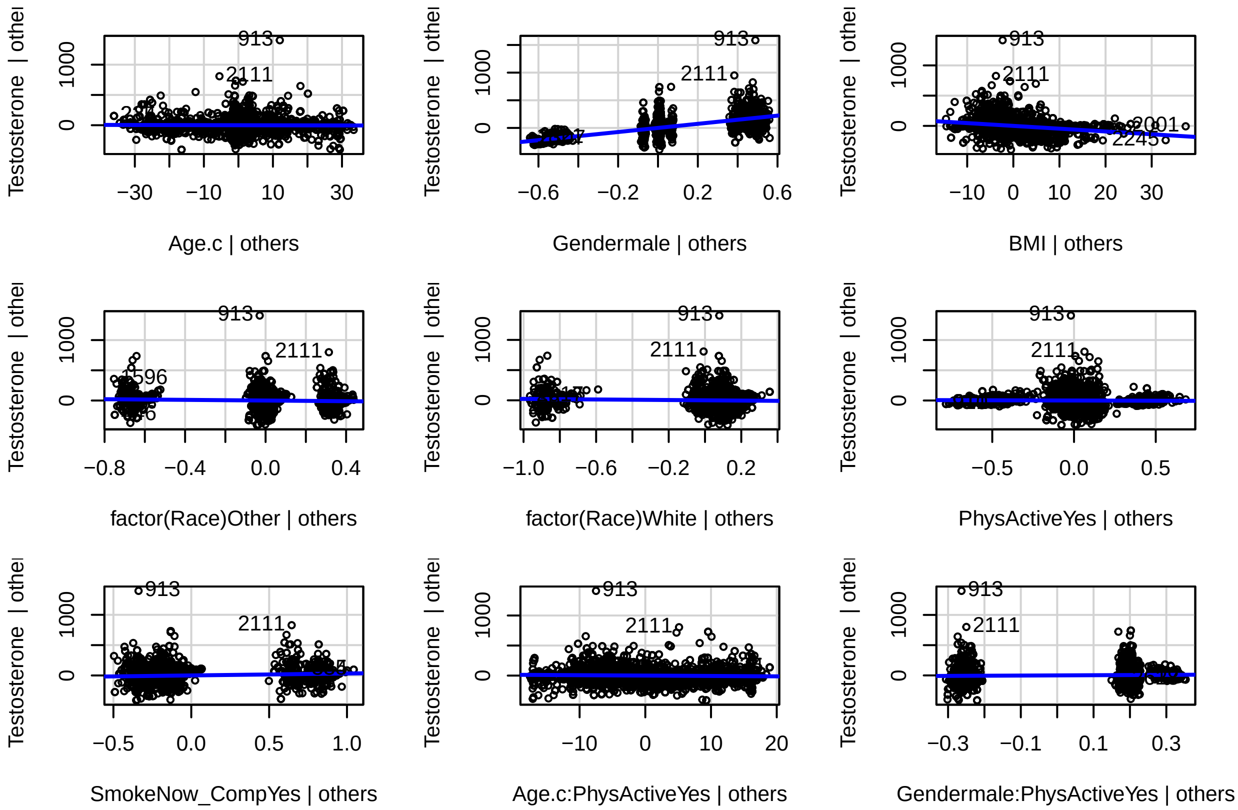

Figure 1: Partial regression plots for full model with interactions

Linearity Given that we observed significant interactions during modeling analysis, we decided to move forward with diagnostic analysis for the third model. From the partial regression plots we observe consistently linear trends between the residuals of testosterone and covariates of interest after adjusting for all other covariates. This indicates that the linearity assumption has not been violated, providing a foundation for sound estimation and inference. The most noticeably non-zero linear associations are observed with variables Gender and BMI, both in the full model as well as the model including interactions of interest.

Independence The independence assumptions pertain to the design of the NHANES study and the manner in which participants were selected. NHANES participants are recruited via random sampling methods, and thus we can assume that each observation is independent from each other.

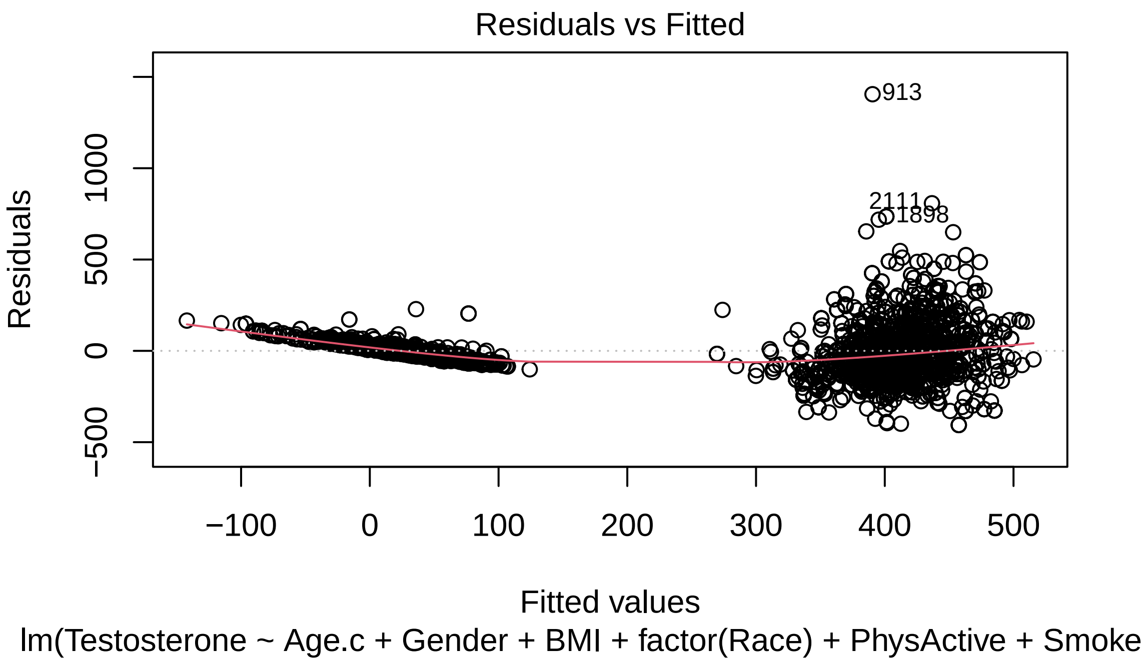

Figure 2: Fitted values vs. residuals for full model with interactions

Equal Variance In evaluating the equal variance assumption in a fitted values vs. residuals plot,we observe two distinct clusters with noticeably different variances in the residuals. This clustering is due to the large effect Gender has on Testosterone and differences in variation between these two groups.As a result, we would classify this as a violation of the equal variance assumption and propose to implement robust standard errors to account for the unequal variances. This is necessary to address, given that though the beta coefficient estimates remain unbiased, the variance being ill-defined results in biased hypothesis testing and invalid confidence intervals.

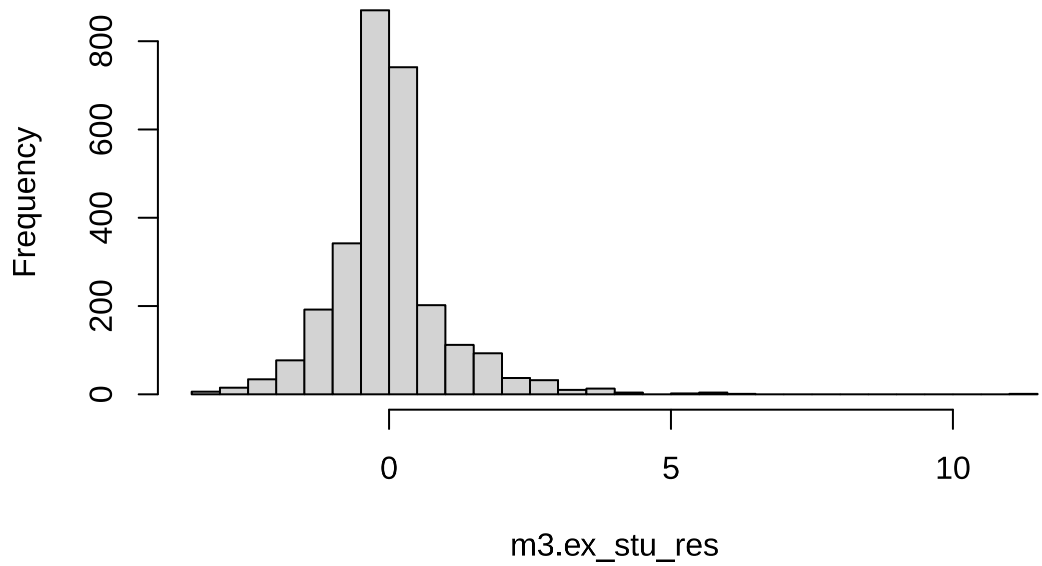

Figure 3: Histogram of externally studentized residuals for full model with interactions

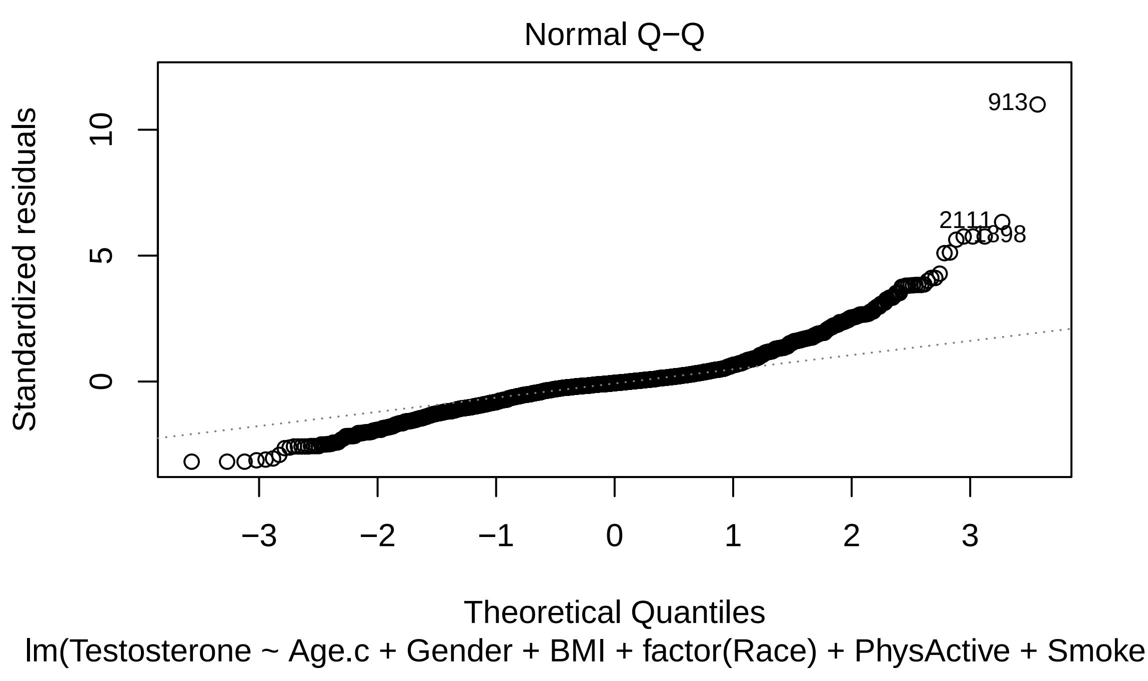

Figure 4: QQ-plot of externally studentized residuals for full model with interactions

Normality The histogram and QQ-plot of the externally studentized residuals of the adjusted analysis including interactions shows that the residuals follow a normal distribution fairly closely. However, we do observe a slight deviation near the right tail. We might consider the tails to be slightly too heavy. However, due to the robust nature of the OLS estimators and the effects of the CLT with such a large sample size, we are not concerned about the validity of the estimates.

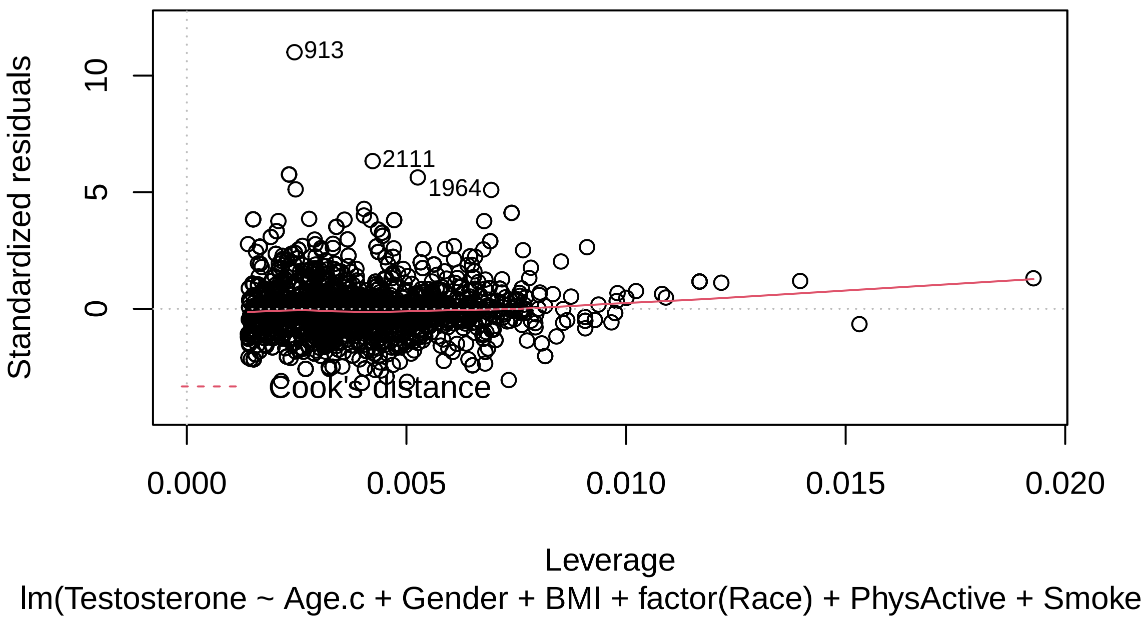

Figure 5: Residuals vs. leverage for full model with interactions

Influence Diagnostics There are no points that we considered to be both outliers and leverage points in our data set as can be seen by the Residuals vs Leverage graphs. These would be points with extreme residual values as well as points of high leverage.One outlier of interest is observation 913, a male participant with very high testosterone (1795.6). Without further information it is difficult to discern if this was a result of an error in recording, measurement, etc. There is no justification to remove any of these points as these do not greatly impact the modeling, nor is there indication that these observations are due to error.

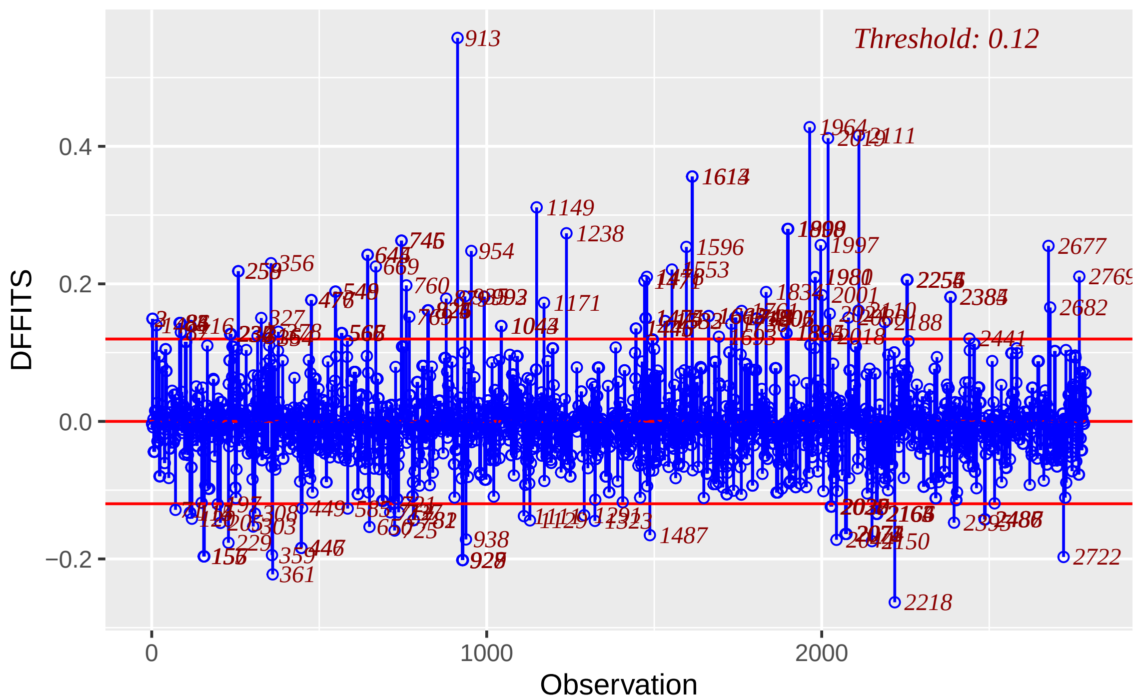

Figure 6: DFFITS for Testosterone, Full Model with Interactions

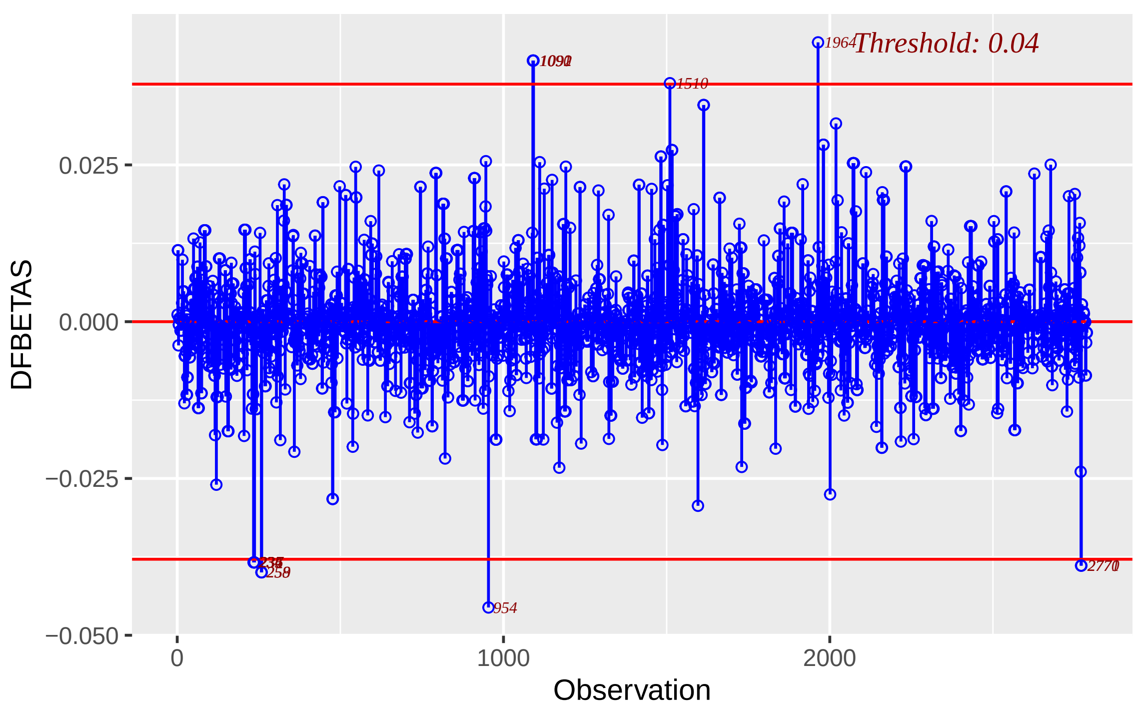

Figure 7: DFBETAS Plot for Physical Activity, Full Model with Interactions

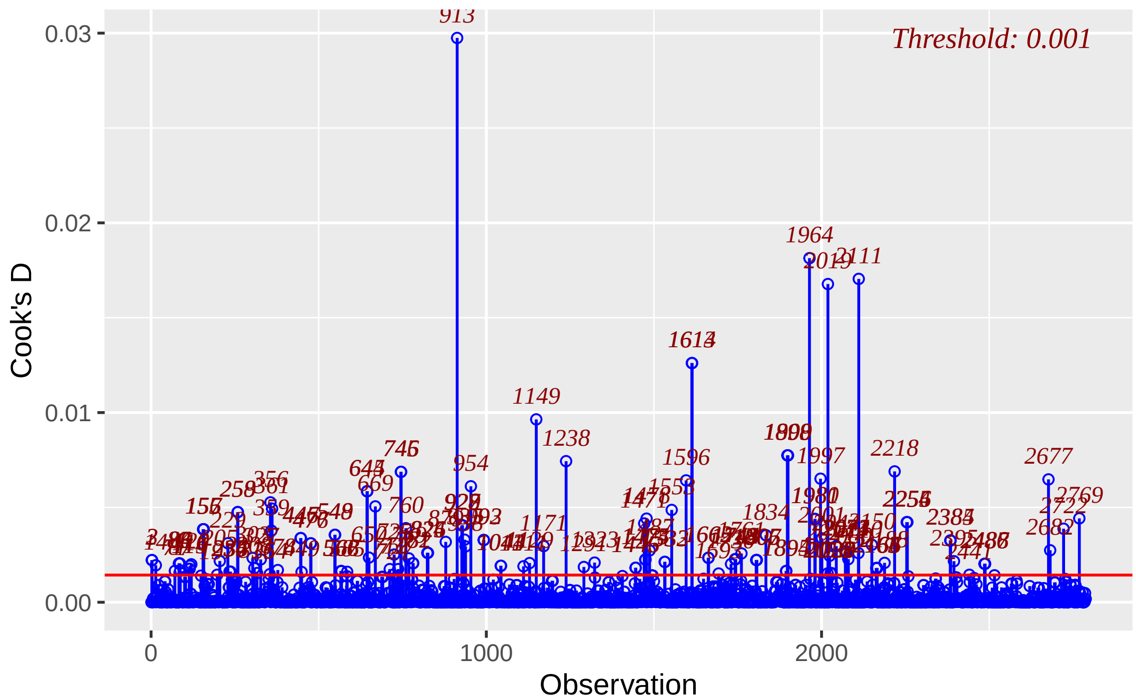

Figure 8: Cook's D Plot, Full Model with Interactions

Next, we evaluated the impact each observation had on the model more closely. The DFFITS figure demonstrates which observations have the highest impact on the fitted values by comparing the fitted value before and after removal of the observation of interest. The observation that had the highest impact was observation 913, the outlier identified above.

We also evaluated DFBETAS visualizations to discern the impact each observation had on each estimated beta coefficient. For our primary predictor of interest, we found no observations that were drastically impacting the estimated beta coefficient.

Lastly, we compared the cook's distance of each point, a holistic measure that evaluates the impact each observation on all the estimated beta coefficients. Again, we see that observation 913 is a point of interest as it has the largest cook's distance.

Refitting the Model

Discussion

Among the descriptive statistics, the gender difference between active and inactive groups is not statistically significant. Both groups have around 54.2% males which equals to the original percentage of males in the study. People being physically active is younger than not active groups which is easy to understand since the elderly are not as energetic as the youth. People being active have lower BMI can be attributed to the effect of being active like doing sports. There are 28.8% of smokers in the inactive groups which is significantly higher than the 13.4% in the active group. This raises another interesting topic—it seems that smoking is influencing people's willingness of being active. The difference of race constitution between groups is also significant. This cannot be explained straightforwardly and is probably due to the variation of sampling. The alcohol intake levels are different between groups and the physically active group is even higher than the

Table 2: Model 3 Summary with Robust standard errors

	Beta Estimate	SE before RSE	P-value before RSE	SE with RSE	P-value with RSE
Intercept	175.0338	15.4889	< 0.0001	14.047	< 0.0001
Age Centered	-0.1202	0.2188	0.583	0.2183	0.582
Gender (male)	369.5555	7.4015	< 0.0001	7.1927	< 0.0001
BMI	-4.625	0.393	< 0.0001	0.3673	< 0.0001
Race (other)	-27.81	9.5434	0.0036	10.1125	0.006
Race (White)	-23.0682	8.3371	0.0057	8.8935	0.0095
Physically Active	-7.5531	7.4469	0.3105	2.9899	0.0116
Smoking	31.9573	6.2456	< 0.0001	6.6967	< 0.0001
Age*Physical Activity	-0.6406	0.2896	0.027	0.2959	0.0305
Gender*Physical Activity	28.5713	9.8182	0.0036	9.1859	0.0019

inactive ones. This result is just the opposite to our previous assumption. The relationship between them can be either just a correlation or a potential causal interpretation.

The univariate linear regression shows that the testosterone levels between activity groups is not statistically significant, which implies there may not be any association between active or not and testosterone levels.

Then the result of the multiple linear regression without interaction terms shows that physical activity is still not significantly associated with testosterone levels adjusting for all other covariates (p=0.130). Number of days drinking alcohol is a covariate we included from previous studies, but the result is not significant.

Multiple linear regression with interaction terms includes the interaction between age and physical activity and the interaction between gender and physical activity. This changed the significance of original age term (p=0.583).

The final refitted model had a change on the significance of the main effect of physical activity—it is now significantly associated with testosterone levels. The coefficient of the interaction between age and physical activity means the estimated mean difference is -0.641 ng/dl of testosterone level in adjusted age effect comparing physically active individuals with physically inactive individuals (p=0.030). The coefficient of the interaction between gender and physical activity means the estimated mean difference is 28.571 ng/dl of testosterone level in adjusted effect of sex comparing physically active individuals with physically inactive individuals (p=0.002). Other covariates used in the model proved a validation of previous similar studies on testosterone and physical activity. For example, adjusting for all other covariates, the estimated mean of testosterone level will decrease by 4.625 for every one unit higher BMI. We can finally make the conclusion that the adjusted mean effect of physical activity for men is 21.018 ng/dl of testosterone level. The adjusted mean effect of physical activity for women is -7.553 ng/dl of testosterone level.

The study has both strengths and limitations. We introduced interaction terms to help find out the potential relationship between the two indices, which is not commonly seen in other similar studies. Our model meets the 3 assumption of linear regression and then a robust standard error modification is conducted to meet the last criterion and the data is generated from a nationwide sampling with large sample size. Therefore, the result and interpretation are trustworthy.

Since the model is now used to find the relationship between two specific indices, a further implementation of the model would be making predictions. However, since the data is collected by a cross-sectional study, we do not know the causality between them. In other words, we do not know whether it is inherent testosterone level influencing the physical activity behavior or it is physical activity behavior affecting testosterone level. Therefore, making predictions will not be convincing. Moreover, we only used the complete data in the dataset, the result may not be able to extrapolate to the entire population. Furthermore, possible bias due to the self-reported physically active variable rather than a precise measured one may have an effect on the final results. We did not run another two models stratified by gender since this will change our scientific question. And since we have found that the interaction term to be significant, this will also explain the difference between gender. Further studies on this topic can focus on building different gender models and conduct a cohort study to build an inference or prediction model to find causal relationships. Possible modifications to deal with missing data such as multiple imputations can be done in further studies.

Conclusions

Our study validated some of the confounders in previous studies of the relationship between physical activity and testosterone level. We also discovered that testosterone level is associated with physical activity and its interaction with gender and age. Being physically active is associated with increased testosterone level for male. While for female, being physically active is associated with decreased testosterone level. For older people, being physically active is associated with decreased levels of testosterone.

References

1. Halpern JA, Brannigan RE. Testosterone Deficiency. JAMA. 2019;322(11):1116. doi:10.1001/jama.2019.9290

2. Davis SR, Jacobsen SW. Testosterone in women - the clinical significance. The Lancet Diabetes and Endocrinology. 2019. doi.org/10.1016/S2213-8587(15)00284-3

3. Harman SM, Metter EJ, Tobin JD, Pearson J, Blackman MR; Baltimore Longitudinal Study of Aging. Longitudinal effects of aging on serum total and free testosterone levels in healthy men. Baltimore Longitudinal Study of Aging. J Clin Endocrinol Metab. 2001 Feb;86(2):724-31. doi: 10.1210/jcem.86.2.7219. PMID: 11158037.

4. Kumagai H, Zempo-Miyaki A, Yoshikawa T, Tsujimoto T, Tanaka K, Maeda S. Increased physical activity has a greater effect than reduced energy intake on lifestyle modification-induced increases in testosterone. J Clin Biochem Nutr. 2016 Jan;58(1):84-9. doi: 10.3164/jcbn.15-48. Epub 2015 Nov 27. PMID: 26798202; PMCID: PMC4706091.

5. Steeves JA, Fitzhugh EC, Bradwin G, McGlynn KA, Platz EA, Joshu CE. Cross-sectional association between physical activity and serum testosterone levels in US men: results from NHANES 1999-2004. Andrology. 2016 May;4(3):465-72. doi: 10.1111/andr.12169. Epub 2016 Mar 16. PMID: 26991734; PMCID: PMC5808848.

6. Ennour-Idrissi et al. Breast Cancer Research. 2015 17:139 DOI 10.1186/s13058-015-0647-3

7. Suryadinata, Rivan Virlando et al. "Effect of age and weight on physical activity." Journal of public health research vol. 9,2 1840. 3 Jul. 2020, doi:10.4081/jphr.2020.1840

8. Azevedo, Mario Renato et al. "Gender differences in leisure-time physical activity." International journal of public health vol. 52,1 (2007): 8-15. doi:10.1007/s00038-006-5062-1

9. Anders, Sari M. van, et al. "Effects of Gendered Behavior on Testosterone in Women and Men." PNAS, National Academy of Sciences, 10 Nov. 2015, https://www.pnas.org/content/112/45/13805.

10. J. A. Osuna C., R. Gómez-Pérez, G. Arata-Bellabarba & V. Villaroel (2006) RELATIONSHIP BETWEEN BMI, TOTAL TESTOSTERONE, SEX HORMONE-BINDING-GLOBULIN, LEPTIN, INSULIN AND INSULIN RESISTANCE IN OBESE MEN, Archives of Andrology, 52:5, 355-361, DOI: 10.1080/01485010600692017

11. Ellis, L. and Nyborg, H. (1992), Racial/ethnic variations in male testosterone levels: A probable contributor to group differences in health. Steroids, Volume 57, Issue 2, 72-75, ISSN 0039-128X. https://doi.org/10.1016/0039-128X(92)90032-5.

12. Svartberg, J. and Jorde, R. (2007), Endogenous testosterone levels and smoking in men. The fifth Tromsø study. International Journal of Andrology, 30: 137-143. https://doi.org/10.1111/j.1365-2605.2006.00720.x

13. Sarkola, T. and Eriksson, C.J.P. (2003), Testosterone Increases in Men After a Low Dose of Alcohol. Alcoholism: Clinical and Experimental Research, 27: 682-685. https://doi.org/10.1111/j.1530-0277.2003.tb04405.x

14. Hemmingsson E, Ekelund U. Is the association between physical activity and body mass index obesity dependent? Int J Obes (Lond). 2007 Apr;31(4):663-8. doi: 10.1038/sj.ijo.0803458. Epub 2006 Sep 5. PMID: 16953254.