Is Computer Use Associated with BMI?

A Statistical Analysis

Roshini Balasubramanian

1 Introduction

The CDC found that by 2018, over 40% of Americans were obese, a number that has increased by over 10% in the past two decades (Center for Disease Control and Prevention 2020). Obesity is related to a variety of health conditions, increased medical costs, and decreased wellness. It is a condition that is recognized to disproportionately affect people based on age, race, and socioeconomic status. The emerging public health crisis caused by increasing pediatric obesity has often been tied to youth consumption of media, but others believe factors like diet are the main cause of rising rates. Accordingly, this study aims to investigate the question, **does spending time on the computer playing games or watching videos increase BMI?**

2 Theory

Research regarding the effect of media on obesity rates has been generally inconsistent. Many studies analyzing the relationship between the time children spend watching television and pursuing other activities found that increased screen media consumption had to little to no correlation with the time children spend on high-energy, outdoor activities (Vandewater 2006). It mostly displaced indoor activities and negatively correlated with time spent with family and friends. However, there are factors about increased media that could be contributing to the growing obesity rate. For instance, one theory suggests that increased snacking while watching videos or television and the large amount of food and beverage marketing advertisements could be affecting the diet of those who engage with media, especially for content unrelated to work or school, and thus contribute to obesity (Calvert 2002). Other factors, like duration of sleep, are related to obesity as well, and there are studies that show screen time negatively affects sleep habits while others show that there is limited impact (Robinson 2017). Finally, some studies believe that screen media could have a positive impact on the lives of some individuals, giving them access to interactive ways to improve diet and physical activity (Robinson 2017). Overall, while there is general consensus that time spent consuming screen media and obesity rates have both been on the rise, particularly among American youth, it is unclear whether there is a relationship between the two and if it is positive or negative.

3 Data

This analysis used the Add Health Data, a longitudinal study that looks at adolescents' social, economic, and physical characteristics from the years 1994 to 2008. In order to trace the appropriate variables, I used DS1 from Wave I (mainly for demographics) and DS22 from Wave IV (for more specific variables related to health and wellbeing).

The explanatory variable was coded as a binary variable, user, from this question: "In the past seven days, how many hours did you spend playing video or computer games, or using a computer? Do not count internet use for work or school," H4DA23, which ranged in values from 0 to 105. If the data value was 0, the respondent was assigned a user value of 0. Otherwise, the respondent was assigned a value of 1. While methods of measuring obesity are frequently contested, BMI is a common metric, as it is relatively inexpensive and non-invasive technique. The outcome variable is Body Mass Index, H4BMI. Sources classify the numeric cutoffs for obesity risks differently, as BMI is influenced by unrelated factors like height and body fat (American Heart Association, n.d.), so I considered BMI as a continuous variable with answers ranging from 14.40 to 889.00.

The controls were age, sex, race/ethnicity, household income, and diet. For age, I looked at the birth year of the respondent. For sex, I created a binary variable, "female," which was set to either 0 (male) or 1 (female). Dummies were included for hispanic, black, and asian race/ethnicity, with 0 being that the respondent did not identify with the group and 1 being that they did (the reference was set as white). It's important to note that the way the data was set up, respondents were able to identify with multiple races. Income was left as given, with with each unit representing $1000 and a response maximum of 999. To track diet, I looked at this question: "How many times in the past seven days did you eat food from a fast food restaurant, such as McDonald's, Burger King, Wendy's, Arby's, Pizza Hut, Taco Bell, or Kentucky Fried Chicken or a local fast food restaurant," H4GH8, which had a maximum response value of 99.

The Add Health dataset had some missing values. To avoid listwise deletion, I did imputation through Amelia for all of the variables. Thirteen of the observations reported BMIs of 888 or 889, which are unrealisticly high and do not fit on the known scale, so I removed and imputed these values as well. The implications of this are further investigated in the Discussion section. The visualizations in Figures 1 and 2 in the Appendix show the difference in missingness before and after imputation; after imputation, there were no more NA values. Additionally, because the data isn't collected through randomized control trials, I matched this observational data, so that covariates are balanced between the treatment and control groupl and can be compared "apples-to-apples" by finding observations that are similar on observed covariates. I matched on year of birth, female, hispanic, white, black, asian, income, and frequency of fast food. Figure 3 below can be used to check for covariate balance and shows it nearly matched for most covariates.

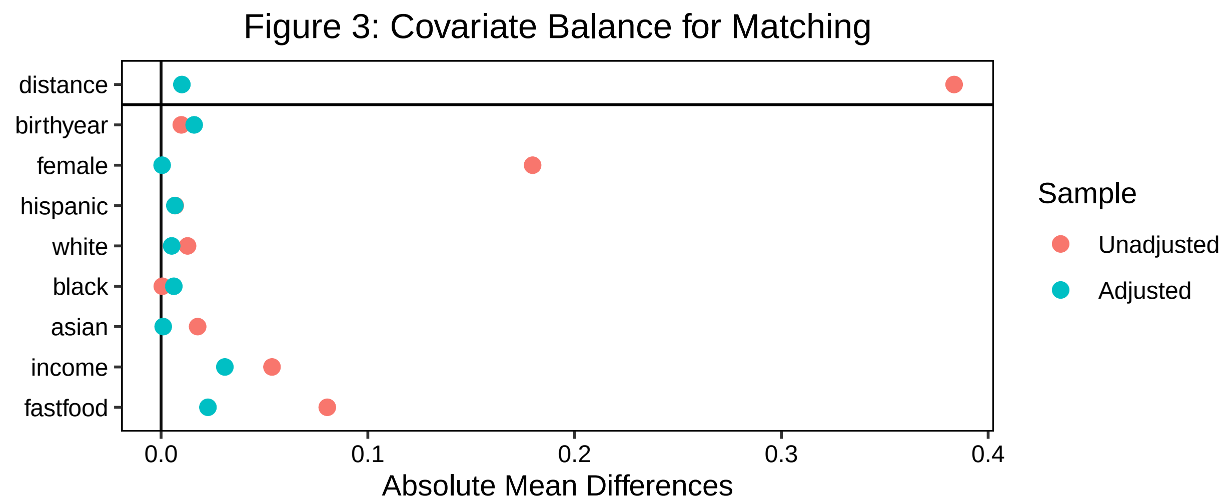


Table 1 below shows summary statistics, filtered by whether or not the repsondent used a computer to watch videos or play comuter games unrelated to work or school, of the data that was matched to correct for imbalance. One important part is the outcome variable, BMI, and its statistics–count, mean (SD), and range–that are included at the top.

Table 1: Summary Statistics (After Matching)

	Not a User (N=1949)	User (N=1949)	Total (N=3898)
BMI			
N	1949	1949	3898
Mean (SD)	28.86 (7.24)	29.69 (7.62)	29.28 (7.44)
Range	14.40 - 70.30	16.30 - 63.50	14.40 - 70.30
Year of Birth			

	Not a User (N=1949)	User (N=1949)	Total (N=3898)
N	1949	1949	3898
Mean (SD)	1978.99 (1.75)	1979.01 (1.78)	1979.00 (1.77)
Range	1975.00 - 1983.00	1974.00 - 1983.00	1974.00 - 1983.00
Sex			
Female	866 (44.4%)	867 (44.5%)	1733 (44.5%)
Male	1083 (55.6%)	1082 (55.5%)	2165 (55.5%)
Hispanic			
Hispanic	206 (10.6%)	193 (9.9%)	399 (10.2%)
Not Hispanic	1743 (89.4%)	1756 (90.1%)	3499 (89.8%)
White			
Not White	622 (31.9%)	632 (32.4%)	1254 (32.2%)
White	1327 (68.1%)	1317 (67.6%)	2644 (67.8%)
Black			
Black	469 (24.1%)	481 (24.7%)	950 (24.4%)
Not Black	1480 (75.9%)	1468 (75.3%)	2948 (75.6%)
Asian			
Asian	66 (3.4%)	68 (3.5%)	134 (3.4%)
Not Asian	1883 (96.6%)	1881 (96.5%)	3764 (96.6%)
Household Income			
N	1949	1949	3898
Mean (SD)	50.32 (40.95)	48.78 (41.97)	49.55 (41.46)
Range	0.00 - 600.00	0.00 - 600.00	0.00 - 600.00
Fast Food Frequency			
N	1949	1949	3898
Mean (SD)	2.32 (2.84)	2.41 (2.63)	2.37 (2.73)
Range	0.00 - 35.00	0.00 - 30.00	0.00 - 35.00

4 Methods

In order to analyze the data, I ran OLS regressions of BMI onto whether the respondent used a computer to play games/watch videos or not. OLS is the appropriate model, as opposed to logistic regression, because BMI is a continuous variable. The null hypothesis is that there is no relationship between playing computer games/watching videos and BMI. The data in this survey is observational, but analogizes to a real experiment that has random assignment of gaming/watching videos. The independent variable is manipulable so there is a treatment and control group.

Table 3 in the Appendix displays seven different models (the first three are included in the Table 3 below). The first three (Models 1, 2, and 3) are after imputing and matching, the next two (Models 4 and 5) are after imputing (no matching), and the final two (Models 6 and 7) are on the original data (no imputing, no matching), as I wanted to see if the modifications may have affected the significance of any regressors. Models 1, 4, and 6 are bivariate regressing BMI on computer game/video watching status. Model 2 included the same explanatory variable and also controlled for some demographics, namely gender and race. Models 3, 5, and 7 regresses BMI on the main explanatory variable along with a full set of controls.

There were shifts in the coefficients between the models on the imputed and unimputed data, suggesting that the missingness is systematic. In theory, imputing the model helps get rid off these biases present in the original data. A similar shift in coefficients was present between the models on matched and unmatched data. Matching the data helps better isolate the effect of the treatment.

The ANOVA tests included in the Appendix are a comparison of Models 1 and 2, Models 2 and 3, and Models 1 and 3, respectively, and these tables reveal which variables have the most explanatory power. The first and second ANOVA tests in Tables 4 and 5 demonstrate that the addition of gender and race/ethnicity explains more than the addition of the other controls (year of birth, income, fast food frequency), but the

Table 2: Regression Models

	Dependent variable:		
	BMI		
	Bivar(MI)	Demogr(MI)	Full(MI)
	(1)	(2)	(3)
User	0.828^{***} (0.238)	0.826^{***} (0.236)	0.805^{***} (0.235)
Year of Birth			-0.218^{***} (0.067)
Female		0.688^{***} (0.238)	0.747^{***} (0.239)
Hispanic		1.241^{***} (0.394)	0.963^{**} (0.396)
Black		1.912^{***} (0.280)	1.717^{***} (0.283)
Asian		-2.157^{***} (0.651)	-2.050^{***} (0.648)
Income			-0.016^{***} (0.003)
Fast Food Frequency			0.033 (0.044)
Constant	28.865^{***} (0.168)	28.041^{***} (0.215)	460.682^{***} (131.814)
Observations	3,898	3,898	3,898
R^2	0.003	0.022	0.033
Note:			*p<0.1; **p<0.05; ***p<0.01

third ANOVA in Table 6 reveals that the addition of all seven control is stastically significant (p<0.001). The table of linear regressions shows that the significance of the predictors does vary slightly depending on which version of the data you use, so the following results will be based on the imputed and matched data. Therefore, I favored Model 3: a linear model where BMI is a function of "user" status and a full set of controls on data that has been imputed and matched, as represented by the equation below.

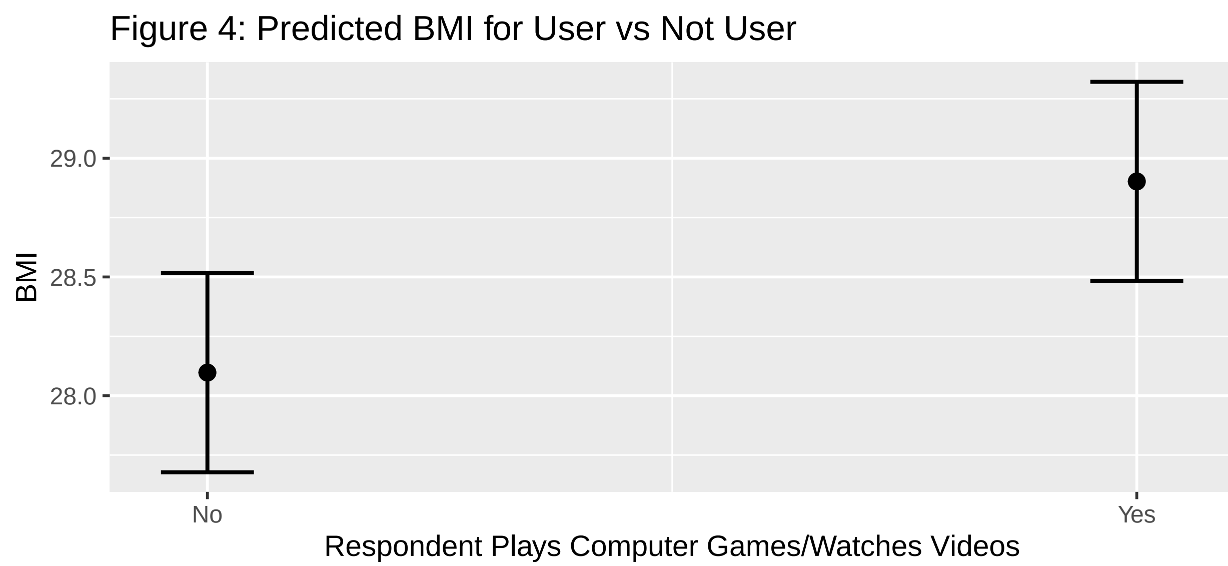
Equation:

$$BMI = \hat{\beta}_0 + \hat{\beta}_1 \times user + \hat{\beta}_2 \times birthyear + \hat{\beta}_3 \times female + \hat{\beta}_4 \times hispanic + \hat{\beta}_5 \times black + \hat{\beta}_6 \times asian + \hat{\beta}_6 \times income + \hat{\beta}_7 \times fastfood + \epsilon$$

There are many assumptions when running OLS. The dependent variable must be linear in parameters, as demonstrated by the linear relationships in the equation above. There is perfect collinearity between the regressors, and the observations must have random sampling. It is also assumed that the error term has an expected value and mean of zero and that they are uncorrrelated with all of the regressors, so that the error term is responsible for picking up any variation that is not otherwise accounted for.

5 Results

Because of matching, the coefficients across Models 1, 2, and 3 are similar for the attributes that are present in each model. All of the OLS regressions found whether the respondents plays computer games/watches videos to be statistically significant (p<0.01). Model 3, which has a full set of controls (Year of Birth, Gender, Race/Ethnicity, Income, and Fast Food Frequency) found that, holding all else constant, moving from "Not User" (control) to "User" (treatment) was associated with a 0.805 unit increase in BMI, with a standard deviation of 0.235, as represented in Figure 4 below. There is very little overlap between the 95% Confidence Intervals, suggesting that the intercept shift is very significant. The evidence is strong enough to reject the null hypothesis (p<0.01).



Many of the controls were statistically significant in Model 3. Asian race had the largest impact on the outcome variable in terms of magnitude on the outcome variable. Asians on average displayed a BMI 2.05 units lower than Whites, which is significant ($p<0.01$). Figure 5 in the Appendix displays a plot similar to before that again predicts BMI but considers Asian race. With the addition of just one control, the significance of the explanatory variable becomes more clear visually, as the 95% Confidence Intervals are even farther apart. The other binary variables for race represente significant intercept shifts as well: Blacks were 1.717 BMI unis higher ($p<0.01$) and Hispanics were 0.963 BMI units higher ($p<0.05$). Aside from these demographics, another important constant to note household income, with a thousand dollar increase in household income being associated with a 0.0016 decrease in BMI, which is significant ($p<0.001$). For all the constants, similar correlations are seen in the other models based off of the imputed (not matched) and original (not imputed, not matched) datasets, although the female indicator variable increases in significance after we modify the datasets and the hispanic indicator varies a little across the models.

The correlation between Fast Food Frequency and BMI was insignificant and varied between a positive and negative relationship depending on the dataset that was used, suggesting that it is not an important attribute. While this control was not significant on its own, it was still included in the model to soak up variation that may impact the relationship between the main explanatory variable and the dependent variable.

It is important to remember that these results only apply to a specific year (2008, when the data was collected) and population, (Americans who were in grades 7-12 during the 1994-95 school year, like in the study). These results have a limited scope of inference because distributions like BMI, diet, behavior, etc. of the population can change over the years, so it is possile that varying factors like location or time could yield different results than this study.

6 Discussion

First, using BMI as a measure of obesity is contested in the medical world, as it doesn't account for factors like height, body type, and body fat well. As seen in this report and in research studies, BMI is largely correlated with race and gender, which presents an opportunity for future improvements in data analysis, as I did not consider mixed race status or what percentage of the respondent was of each race due for simplicity's sake and due to a lack of available information. Additionally, weight is a health issue affected by a variety of social determinants of health, and while this report considers some controls, there are many other confounding variables including living situation, physical activity, sugar intake, etc. that could be biasing the results. A more thorough analysis would account for many other controls. While I tried to model random assignment through matching, the process was not perfect, as not all cofounders could be controlled, presenting a similar problem for the analysis. Finally, the main explanatory variable looked at time spent on the computer watching videos or playing computer games, which does not include many types of screen media nor distinguish between the two included. Further analysis should include different and more categories in order to more accurately pinpoint potential effects on obesity.

Regarding the models, in order to design an experiment with "control" and "treatment" groups and run a simple matching process, I turned the main explanatory variable into a binary one. This could potentially have been a source of error, so I tested similar regressions on a model with a continuous main explanatory variable, indicating the number hours spent on the computer playing games or watching videos, and the statistical significance for the variables was relatively similar. The full model based on imputed data (not matched) is included in the Appendix in Table 7. Additionally, there were a few entries with unreasonably high BMIs that I removed and imputed, and this drastically changed the results. Table 8 in the Appendix presents a model with full controls on imputed data that includes the unreasonably high numbers for comparison. This is important to note, because while I am sure the reported BMIs in the 800s are not possible, I do not know what actual values are, and the integrity of the data is lost with deletion.

The data allows me to make conclusions about the correlation between the variables for the average American adult between the ages included in the survey (those in grades 7-12 during the 1994-95 school year). Screen media and population social habits are constantly evolving, so the years included in this study are important to note. Location is important as well, because social, physical, and dietary habits vary drastically not only within the country but internationally as well. Due to these limnitations and the nature of observational data, the results should be taken as an extrapolation of the relationship between the main explanatory and outcome variables, not a causal claim.

7 Conclusion

From the ANOVA tests and summaries, I found that among the variety of linear models tested on imputed and matched data, the one containing full controls had the most explanatory power. Based on this model, as displayed in Table 2, moving from "Not User" (control) to "User" (treatment) was associated with a 0.805 unit increase in BMI if all else is held constant, where a "User" is a respondent who spends time on the computer playing games or watching videos unrelated to school or work; this result is significant ($p<0.01$). The largest regressor controls in terms of magnitude were asian race ($p<0.01$) and black race ($p<0.1$). Gender, birth year, and income were extremely significant as well.

In conclusion, this analysis suggests that increased screen time, particularly on a computer for computer games or watching videos unrelated to school or work, is associated with increased BMI, which was used as a measure of obesity, when controlling for other variables. This rejects the claim that increased access to screen media is uncorrelated with obesity, as Vandewater suggests, but her statistical analysis that reveals that screen time does not impact time spent engaging in intense, outdoor activity, was not investigated in this study. The results of this study do weaken Robinson's claim that increased access to screen media causes a decrease in obesity rates due to improved access to information regarding fitness and extra motivation online, as this did not prove to be the case at a population level. Calvert's reasoning is ultimately the one that is supported as he claims that increased screen time is associated with increased rates of obesity, but the exact reason behind this relationship remains unconfirmed. These results are important in health policy, as experts need to figure out what factors are most important for controlling obesity rates. Based on these results, limiting computer games and videos may be a directly effective solution. Targeting and controlling for other socioeconomic factors, like household income and race, may prove to be important as well for widely improving public health.

Appendix

Figure 1: Missing Values Before Imputation

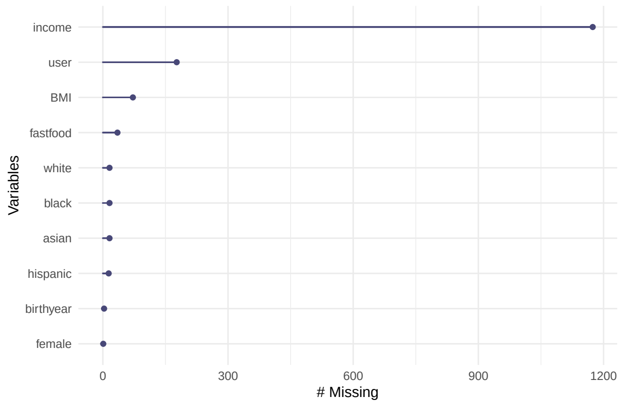


Figure 2: Missing Values After Imputation

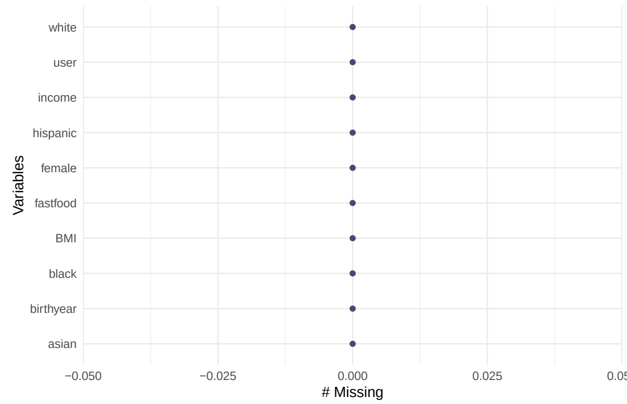


Figure 5: Predicted BMI Based on User vs Not User Including Asian Race

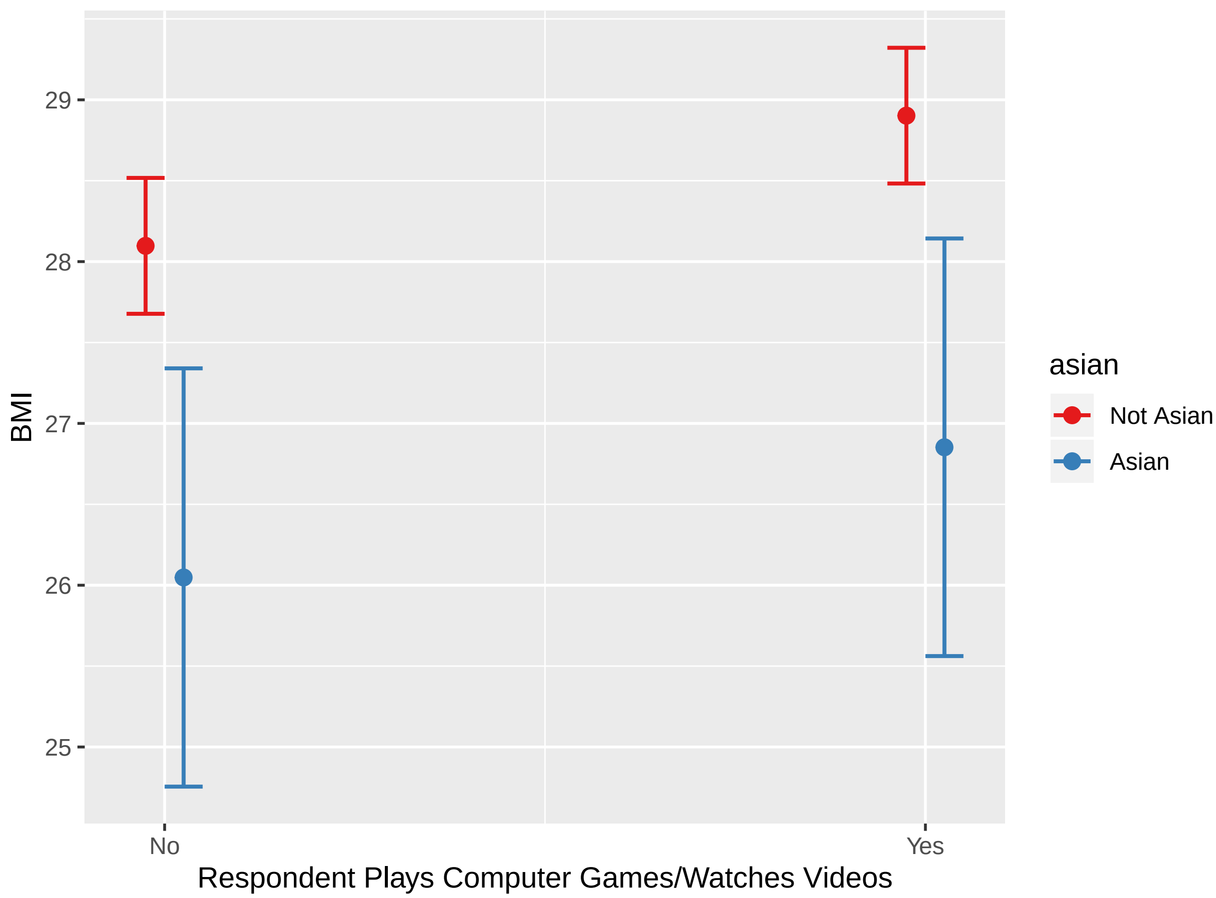


Table 3: Regression Models

	Dependent variable:						
	BMI						
	Bivar(MI)	Demogr(MI)	Full(MI)	Bivar(I)	Full(I)	Bivar	Full
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
User	0.828***	0.826***	0.805***	0.812***	0.915***	0.874***	0.995***
	(0.238)	(0.236)	(0.235)	(0.214)	(0.214)	(0.219)	(0.249)
Year of Birth			−0.218***		−0.239***		−0.209***
			(0.067)		(0.059)		(0.069)
Female		0.688***	0.747***		0.541**		0.258
		(0.238)	(0.239)		(0.212)		(0.247)
Hispanic		1.241***	0.963**		0.905***		0.825**
		(0.394)	(0.396)		(0.343)		(0.417)
Black		1.912***	1.717***		1.959***		2.167***
		(0.280)	(0.283)		(0.247)		(0.299)
Asian		−2.157***	−2.050***		−2.064***		−2.125***
		(0.651)	(0.648)		(0.555)		(0.662)
Income			−0.016***		−0.012***		−0.012***
			(0.003)		(0.002)		(0.002)
Fast Food Frequency			0.033		−0.005		0.060
			(0.044)		(0.031)		(0.044)
Constant	28.865***	28.041***	460.682***	28.832***	500.821***	28.796***	442.684***
	(0.168)	(0.215)	(131.814)	(0.135)	(116.016)	(0.138)	(137.286)
Observations	3,898	3,898	3,898	5,114	5,114	4,874	3,750
R^2	0.003	0.022	0.033	0.003	0.033	0.003	0.037
Note:					*p<0.1; **p<0.05; ***p<0.01		

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	3896.0000000000	215233.3824139466				
2	3892.0000000000	211086.2594023110	4.0000000000	4147.1230116356	19.1161220145	0.0000000000

Table 4: ANOVA 1 on Model 1 (Naive) vs. Model 2 (Deomgraphic Predictors)

	Res.Df	RSS	Df	Sum of Sq	Pr(>Chi)
1	3892.0000000000	211086.2594023110			
2	3889.0000000000	208836.3434554162	3.0000000000	2249.9159468948	0.0000000042

Table 5: ANOVA 2 on Model 2 (Demographic Predictors) vs. Model 3 (Full)

	Res.Df	RSS	Df	Sum of Sq	Pr(>Chi)
1	3896.0000000000	215233.3824139466			
2	3889.0000000000	208836.3434554162	7.0000000000	6397.0389585304	0.0000000000

Table 6: ANOVA 3 on Model 1 (Naive) vs. Model 3 (Full)

Table 7: Regression Model Based on Continuous Explanatory Variable (Imputed)

	Dependent variable:
	BMI Full(I)
User	0.052***
	(0.013)
Year of Birth	−0.245***
	(0.059)
Female	0.525**
	(0.212)
Hispanic	0.879**
	(0.344)
Black	1.889***
	(0.248)
Asian	−1.893***
	(0.557)
Income	−0.013***
	(0.002)
Fast Food Frequency	−0.011
	(0.031)
Constant	514.167***
	(116.197)
Observations	5,114
R^2	0.033
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 8: Regression Model Including Incorrect BMI Values

	Dependent variable:
	BMI
	Full(I)
User	1.244
	(1.281)
Year of Birth	−0.217
	(0.350)
Female	−1.047
	(1.263)
Hispanic	1.096
	(2.052)
Black	6.200***
	(1.477)
Asian	−3.665
	(3.337)
Income	−0.019
	(0.012)
Fast Food Frequency	−0.170
	(0.183)
Constant	461.475
	(693.291)
Observations	5,114
R^2	0.005
Note:	*p<0.1; **p<0.05; ***p<0.01

8 Bibliography

American Heart Association. n.d. "Body Mass Index (BMI) in Adults."

Calvert, Sandra L. 2002. "Electronic Gaming and the Obesity Crisis."

Center for Disease Control and Prevention. 2020. "Adult Obesity Facts."

Hlavac, Marek. 2018. "Stargazer: Well-Formatted Regression and Summary Statistics Tables."

Honaker, James, Gary King, and Matthew Blackwell. 2011. "Amelia: A Program for Missing Data."

Robinson, Thomas N. 2017. "Screen Media Exposure and Obesity in Children and Adolescents."

Vandewater, Elizabeth A. 2006. "Time Well Spent? Relating Television Use to Childrens Free-Time Activities."