The Role of Sport Participation in Adolescent Female Mental Health: A Propensity Score Matching Approach.

Zak Massey

Applied Microeconometrics (December 2020)

Abstract

The relationship between sports participation and mental health has been a highly examined topic which has had many differing opinions and outcomes. The purpose of this paper is to examine that relationship while focusing on adolescent females using a Nearest Neighbor Propensity Score Matching (PSM) technique along with the 2011/2012 National Survey of Children's Health (NSCH). The hypothesis is that sports participation can be used as a preventative measure against developing a depression, or that sports participation is associated with lower depression rates. First, I will discuss current mental health trends among Americans and their implications. This will include evidence that the state of mental health in the United States has been on the decline. Then I will discuss relevant literature that suggests sports participation may be a way to combat this issue. I will then cover the data used, the methodology, and finally the results. The results indicate that sports participation helps reduce the chances of suffering from depression among adolescent females, and that sports participation decreases the probability of experiencing depression by 3.7%.

Dr. Andres,

I know the paper seems kind of long. But if you don't count the two pages of references the word count is 2178. Also, I made a separate appendix which goes over the reasoning for all the control variables and where they were retrieved from. I can email you the appendix if you would like.

Best,

-Zak

Introduction

According to the National Alliance on Mental Illness (NAMI), 16.5% of US youths aged 6-17 experienced a mental health disorder in 2016. If we focus more on the state of Arkansas, in a single year, 40% of adolescent Arkansans experienced depression (Department of Health and Human Services). These numbers may seem relatively "normal", but the evidence shows that they are actually much higher than what they used to be. For example, in 1965, suicide was the 12th leading cause of death for all age groups (CDC). Fast-forward to 2019, and suicide has reached the 2nd leading cause of death for people aged between 10-34, and within the top 5 for all other age groups besides 55+. There are an abundance of fact lists and info-graphs concerning "suicide awareness" and "depression" that are often spammed across places like Twitter and Instagram during October, which has been named "Mental Health Awareness Month". But all of this information is given to us without a direct answer regarding preventative measures that can be taken to decrease the chances of developing a mental illness. Preventing the development of a mental illness is important because 90% of all suicide victims have a preexisting mental illness (NAMI). I like to think of this as a way to address an underlying issue that is often a contributing factor to one of the most tragic ways a life can end.

One of the most common practices behind preventative mental healthcare is physical activity, involvement in structured exercise helps alleviate symptoms of clinical depression (Craft and Perna). The scientific theory behind this practice is that physical activity aides the release of endorphins in the brain, which promote feelings of well-being (Mayo Clinic). Disregarding the theory aspect, participation in sports has also been shown to be a protective measure against depression and suicidal ideation (Babiss and Gangwisch). According to the previously mentioned study, Babiss and Gangwisch found as sports participation increases, the odds of suffering from depression decreases by 25%. For many, physical activity and involvement in sports typically starts at a young age, once the child has entered school. Research published by the *Journal of Adolescent Health* examined the relationship that adolescent sports participation had on early adulthood mental health. This longitudinal study found that sports involvement

during school was a statistically significant predictor of lower depression symptoms and higher self-rated mental health in early adulthood (Jewett, Sabiston, et al.). They suggest that polices to increase school sport participation may be warranted as a part of public health strategies to promote mental health. Our approach and hypothesis are very similar to that of the previously mentioned study (except this study does not use longitudinal data). This paper will examine the relationship between adolescent sports participation and the child's *current* state of mental health. The hypothesis is that sports participation can be used as a preventative measure against depression. I will now move on to the *Data* section of the paper to discuss the data used and our main variables of focus.

Data

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Depression	65165	.22	.414	0	1
Sports Participation	65148	.633	.482	0	1
Age	65164	11.731	3.485	6	17
Sex	65164	1.482	.5	1	2
Race	63521	1.402	.735	1	3
Poverty Level	59205	5.752	2.549	1	8
General Health	65162	1.541	.806	1	7
Sleep	65144	6.585	7.904	0	97
State Unemployment Rate	65164	7.438	1.728	3	11.3
Gross State Product	65164	319347.71	385902	28887	2144497
Active Physicians Per 10,000 Residents	65164	27.549	8.865	18.3	76.9
State Minimum Wage	65164	7.495	.431	7.25	9.04
State Population Density	65164	385.87	1356.595	1.2	9856.5
Racism	64328	.091	.554	0	7
Parental Domestic Abuse	64361	.137	.674	0	7

The data used for this analysis is based off of the 2011/2012 National Survey of Children's Health. The NSCH is conducted every few years (administered to different individuals) by the US Census Bureau and funded by the Health Resources and Services Administration's Maternal and Child Health Bureau. It was designed to produce national and state-level data on the participants physical and emotional health (SLAITS). This version of the NSCH was administered over the phone to the parent of youths between the ages of 0-17 for the 2011/2012 year. This specific year was chosen because it was the

richest year in terms of relevant questions asked, sample size, and accessibility. Overall, there were 95,677 observations included with the NSCH 2011/2012 data.

Variables:

The dependent variable, or the outcome variable, was based off a question regarding the selected child's mental health status and was only asked to the parents of children between the ages 6-17. The question asked the parent or guardian to describe how often the selected child was depressed, unhappy, or sad. Parents were able to give a ranked answer on a 1-7 scale (1 = Never…5 = Always). A score of 6 or 7 represented "Don't know" and "Refused" and were subsequently dropped from the dataset. Once the dataset contained only responses that ranged from 1-5, a dummy variable was created to represent whether or not the selected child was currently suffering from depression and was then used as the outcome variable (View Appendix).

The main independent variable was based off of a question that asked whether or not the selected child was on a sports team or had regularly participated in sports lessons during the past 12 months. This was used as a dummy variable in our analysis. The control variables aimed at making the best "apples to apples" comparison among all of our observations. They included many individual and state characteristics that may impact one's mental health. The demographic controls were the selected child's age, sex, race, and poverty level. Since this study is specifically focusing on females, the sex variable was omitted. The state controls were the unemployment rate, gross state product, minimum wage, number of active physicians per 10,000 residents, and population density. Other controls were chosen based off of their prevalence concerning mental health. These included a variable regarding amount of sleep, general health, whether or not the child had been a victim of racism, and whether or not the child had witnessed any form of domestic abuse. More information and the reasoning behind the inclusion of all of these variables can be found in the Appendix.

Methodology and Results

$$\textbf{Model: } Y_i = \alpha + \beta T_i + \gamma P(x_i) + \varepsilon_i$$

$$\textbf{Model: } Depression\ Outcome = \alpha + \beta_{Sports\ Participation} + \gamma P(x_i) + \beta_{Controls} + \varepsilon_i$$

Every microeconomic study has to overcome the fundamental evaluation problem and address the possible occurrence of selection bias (Caliendo and Kopeining). Since it is impossible to observe the difference in an observation's outcome with and without treatment, selection bias is an issue. PSM is a possible solution to this selection bias (Caliendo and Kopeining). The idea is that the bias is reduced when the comparison of outcomes is performed using treated and control subjects who are as similar as possible (Becker and Ichino). Adjusting for the propensity score allows us to remove the bias that is created by all the treatment confounders (Abadie and Imbens). The propensity score mentioned above, is the conditional probability of assignment to a particular treatment given a vector of observed covariates (Rosenbaum and Rubin). This basically means that all of the pre-treatment characteristics are summarized into a single index/likelihood variable through the use of a logit or probit model and then, that variable is controlled for.

In order to properly utilize PSM, I first needed to estimate the propensity score using a standard probability model. In this case I used a probit regression. Next, I needed to decide on a matching algorithm to use, I chose the Nearest Neighbor (NN) technique with replacement. This technique involves choosing an individual from the control group that is given a "matching partner" from the treated group based on their relative propensity scores. The "with replacement" aspect allows for an untreated individual to be used more than once as a match. This increases the average quality of match and also helps decrease bias (Becker and Ichino). After developing a propensity score, I needed to test whether or not the average propensity score of the treated and controls do not differ, this is known as the Balancing Condition (Becker and Ichino).

The results are shown below.

PSM: NEAREST NEIGHBOR:

----------PSM FOR FEMALES ONLY----------

Sex omitted because of collinearity

Probit regression — Number of obs = 28,330; LR chi2(12) = 3381.53; Prob > chi2 = 0.0000

Log likelihood = -17241.967 — Pseudo R2 = 0.0893

sports	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
Age	-0.006	0.002	-2.670	0.008	-0.011	-0.002
Sex				(omitted)		
Race	-0.058	0.011	-5.320	0.000	-0.079	-0.037
Poverty Level	0.141	0.003	43.620	0.000	0.135	0.147
General Health	-0.209	0.010	-20.210	0.000	-0.229	-0.189
Sleep	-0.002	0.001	-2.050	0.040	-0.004	-0.000
State Unemployment Rate	0.003	0.005	0.550	0.583	-0.007	0.013
Gross State Product	-0.000	0.000	-1.690	0.090	-0.000	0.000
Active Physicians Per 10,000 Residents	0.003	0.002	1.880	0.060	-0.000	0.007
State Minimum Wage	0.034	0.020	1.670	0.095	-0.006	0.074
State Population Density	-0.000	0.000	-2.050	0.040	-0.000	-0.000
Racism	0.010	0.016	0.660	0.508	-0.020	0.041
Parental Domestic Abuse	-0.038	0.012	-3.200	0.001	-0.062	-0.015
cons	-0.376	0.146	-2.580	0.010	-0.661	-0.090

Variable Sample	Treated	Controls	Difference	S.E.	T-stat
Depression	0.198	0.282	-0.084	0.005	-16.350
ATT	0.198	0.232	-0.034	0.008	-4.000

Note: S.E. does not take into account that the propensity score is estimated.

psmatch2: Treatment assignment	psmatch2: Common support On	suppor	Total
Untreated	11,020		11,020
Treated	17,310		17,310
Total	28,330		28,330

PSTEST

Variable	Mean Treated	Mean Control	%bias	t-test t	t-test p>|t|	V(T)/V(C)
Age	11.624	11.687	-1.8	-1.68	0.093	0.88*
Sex	2	2	.	.	.	.*
Race	1.3487	1.3421	0.9	0.89	0.376	1.04*
Poverty Level	6.4264	6.4212	0.2	0.22	0.824	0.99
General Health	1.3912	1.3907	0.1	0.07	0.943	0.96*
Sleep	6.2982	6.3702	-1.0	-1.06	0.289	0.97
Unemployment Rate	7.4392	7.4573	-1.0	-0.98	0.330	1.03
Gross State Product	3.1e+05	3.1e+05	-1.1	-1.03	0.301	1.00
Physicians Per 10,000	27.785	27.743	0.5	0.44	0.663	1.10*
State Minimum Wage	7.5058	7.5066	-0.2	-0.17	0.861	0.99
State Population Density	403.5	377.54	1.9	1.79	0.073	1.14*
Racism	.07839	.07614	0.4	0.42	0.672	0.94*
Parental Domestic Abuse	.10508	.09417	1.6	1.81	0.070	1.21*

* if variance ratio outside [0.97; 1.03]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.000	17.97	0.117	0.9	0.9	4.6	1.06	62

* if B>25%, R outside [0.5; 2]

Observations with the same propensity score must have the same distribution of observable (and unobservable) characteristics independently of treatment status. In other words, for a given propensity score, exposure to treatment is random and therefore treated and control units should be on average observationally identical (Becker and Ichino).

This condition is met because all of the p-values for the PTEST are above 0.05. This indicates that there is no statistical difference between the treated and control groups regarding their propensity score. The average treatment effect is -0.034, this indicates that there is a negative relationship between depression score and the treated group, who participated in sports. We can inspect the results further by executing PSM with the right standard errors using the “teffects” command. The results are shown below.

TEFFECTS COMMAND AND RESULTS

Treatment-effects estimation

Depression	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
r1vs0.sports	-.037	.007	-4.91	0	-.051	-.022	***
Mean dependent var		0.231	SD dependent var			0.421	

*** p<.01, ** p<.05, * p<.1

The estimated treatment-effects coefficient given is very close to the one previously given. In this case it is -0.037, whereas previously it was -0.034. To conclude, we can determine that sports participation decreases the probability of suffering from current depression by 3.7% for adolescent females. Both results suggest that sports participation is associated with lower depression scores[1].

Limitations

As with any approach, there are limitations that come with the PSM technique that was used to estimate the relationship between sport participation and depression. One of these limitations is that PSM only solves endogeneity issues dealing with selection bias, leaving the possibility of the presence of the omitted variable bias. This may occur if there were some unobservable confounders that might've affected our results. This potential issue would violate the exogeneity condition of the asymptotic properties. Unobservable confounders might include things like bullying or a genetic condition that would've had an impact on whether or not the child was experiencing depression. Another limitation would be the data itself. As mentioned before, the data was collected through an over the phone survey

[1] Depression was represented by a dummy variable where no-depression took the value of zero and having-depression took the value of 1. Sports was represented by a dummy variable which took the value of 0 if the child did not participate in sports and a 1 if the child did participate in sports. Therefore, a negative coefficient signals that sports participation is associated with lower depression rates.

These results are statistically significant at the 99% level.

- T score is outside of the -2, 2 range
- The confidence intervals do not include 0
- The p-value is <0.01

given to the parent of the selected child. This could be a limitation because the child might've been suffering from depression without the parent being aware.

Conclusion

Mental health has been an issue that has gained a lot of attention in recent years. This is because the state of mental health in the United States has been on the decline since the year 2000. The number of suicides and deaths of despair[2] have had a sharp increase in the US in the past 20 years (American Foundation for Suicide Prevention, 2019). There have been various approaches and hypotheses regarding how to most effectively treat depression and how to reduce the chances of developing this condition. One of the most researched areas of this topic was the relationship between sports participation and depression/mental health. The research suggests that sports participation is a way to reduce levels of depression and improve mental health. This paper examines that question using a Propensity Score Matching technique and the NSCH 2011/2012 data. I find that sports participation is related to lower depression rates in adolescent females, and sports participation decreases their probability of having depression by 3.7%.

2 Deaths caused by drug overdose, suicide, or alcoholic liver disease

References

Abadie, Alberto, and Guido Imbens. “MATCHING ON THE ESTIMATED PROPENSITY SCORE.” *Http://Www.nber.org*, NATIONAL BUREAU OF ECONOMIC RESEARCH, Aug. 2009, www.nber.org/papers/w15301.

Babiss, Lindsay A, and James E Gangwisch. “Sports Participation as a Protective Factor Against Depression and Suicidal Ideation in Adolescents as Mediated by Self-Esteem and Social Support.” *Https://Journals.lww.com/*, Journal of Developmental & Behavioral Pediatrics, Oct. 2009, journals.lww.com/jrnldbp/Abstract/2009/10000/Sports_Participation_as_a_Protective_Factor.2.aspx.

Becker, Sascha O, and Andrea Ichino. “Estimation of Average Treatment Effects Based on Propensity Scores.” *Https://Learn.uark.edu*, The Stata Journal, 2002, learn.uark.edu/bbcswebdav/pid-9068496-dt-content-rid-102078628_1/courses/1209-THEUA-ECON-5783-SEC001-10609/Becker%20and%20Ichino%20%282002%29%20Estimation%20of%20average%20treatment%20effects%20based%20on%20propensity%20scores.pdf.

Craft, Lynette L., and Frank M. Perna. “The Benefits of Exercise for the Clinically Depressed.” *Primary Care Companion to the Journal of Clinical Psychiatry*, Physicians Postgraduate Press, Inc., 2004, www.ncbi.nlm.nih.gov/pmc/articles/PMC474733/.

Caliendo, Marco, and Sabine Kopeinig. “SOME PRACTICAL GUIDANCE FOR THE IMPLEMENTATION OF PROPENSITY SCORE MATCHING.” *Wiley Online Library*, Journal of Economic Surveys, 31 Jan. 2008, onlinelibrary.wiley.com/doi/epdf/10.1111/j.1467-6419.2007.00527.x?saml_referrer=.

“Centers for Disease Control and Prevention.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 2020, www.cdc.gov/.

Craft, Lynette L., and Frank M. Perna. “The Benefits of Exercise for the Clinically Depressed.” *Primary Care Companion to the Journal of Clinical Psychiatry*, Physicians Postgraduate Press, Inc., 2004, www.ncbi.nlm.nih.gov/pmc/articles/PMC474733/.

“Exercise and Stress: Get Moving to Manage Stress.” *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 18 Aug. 2020, www.mayoclinic.org/healthy-lifestyle/stress-management/in-depth/exercise-and-stress/art-20044469.

Jewett, Rachel, et al. “School Sport Participation during Adolescence and Mental Health in Early Adulthood.” *The Journal of Adolescent Health : Official Publication of the Society for Adolescent Medicine*, U.S. National Library of Medicine, Nov. 2013, pubmed.ncbi.nlm.nih.gov/24928804/.

Mental Health By The Numbers.” NAMI, National Alliance on Mental Illness, 2019, www.nami.org/learn-more/mental-health-by-the-numbers.

“NIMH " Home.” National Institute of Mental Health, U.S. Department of Health and Human Services, 2019, www.nimh.nih.gov/index.shtml.

Rosenbaum, Paul R., and Donald B. Rubin. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” Biometrika, vol. 70, no. 1, 1983, pp. 41–55. JSTOR, www.jstor.org/stable/2335942. Accessed 30 Nov. 2020

“SLAITS - National Survey of Childrens Health.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 6 Sept. 2017, www.cdc.gov/nchs/slaits/nsch.htm.

“The State of Mental Health in America: Mental Health America.” The State of Mental Health in America | Mental Health America, Mental Health America, 2019, www.mhanational.org/issues/statemental-health-america.