Taking Risks On and Off the Court: How does the participation in team sports by high school students influence gambling?

Research Question: How does the participation of team sports by high school students influence the likelihood they participate in gambling related activities?

This research question seeks to determine the effect of an individual's regular participation in team sports on the probability of participation in gambling related activities. This is an important question to answer, since it would reveal how regular participation in extracurricular activities influences their attitudes and inclinations towards gambling, which can be thought of as risky behavior. If we find that students who regularly participate in team sports are less likely to participate in gambling related activities, then this can be promoted as one of the benefits of participation in team sports.

Moreover, high school students who suffer from gambling addictions may be recommended to participate in team sports regularly. Physical education classes in high schools can include more team sports (e.g. basketball) instead of individual sports like tennis. However, if the opposite was found to be true, and we find that regular participation in team sports increases the likelihood of participating in gambling-related activities, then special programs targeted at students who regularly participate in team sports can be provided to educate students regarding the potential risk of gambling.

Motivation:

We want to see whether or not the students regularly participating in a sports team involving a competitive and intense environment are more likely to gamble. By knowing the causal relationship, we could help with the creation of programs to help target those who are involved with underage gambling and prevent those kinds of behaviors.

Data Source:

We are using information from the High School Risk Survey from 2006. The data contains information from around 4,500 students in Connecticut public schools that were surveyed in 2006 about their backgrounds and history with risky behaviors, including drinking, drug use, and gambling.

Sample Size:

The sample size is 3,186 students. We dropped all observations whose income is "don't know", to help with the interpretation of the income dummy variable.

Variable Name	**Origin**	**Description**
d_gambling	Q13a, Q13b, …, Q13p	Indicator variable: 1 or 0 depending on whether the student has gambled at all in the past 12 months. 1 - gambled at least once in the past 12 months 0 - never gambled in the past 12 months
d_sports_team	Q7b	Indicator Variable: 1 - Participant does participate in a sports team regularly (at least 1-2 times a month) 0 - Participant does not participate in a sports team regularly
d_income	Q10	Indicator Variable: 1 - Middle income, High income 0 - On public assistance, Low income We decided to group middle income, and high income together since people who classify themselves as middle income often underreported themselves.
d_gender	Q2	Indicator Variable: 1 - participant is Male 0 - participant is Female
d_smoke_regularly	Q53	Indicator Variable: 1- participant smokes regularly now 0 - Otherwise

d_both_parents	Q9	Indicator Variable: 1 - Participant lives with both parents 0 - Participant lives with 1 or less parents
d_drink	Q75	Indicator Variable: 1 - Participant has drank in the past 30 days 0 - participant has never drank in the past 30 days
d_drug	Q80	Indicator Variable: 1 - Participant has used illegal drugs 0 - participant has never used illegal drugs
d_church	Q7d	Indicator Variable: 1 - Participant does attend church regularly (at least 1-2 times a month) 0 - Participant does not attend church regularly

Regression models:

Model 1

Linear Probability Model of Gambling based on Participation in Team Sport:

d_gambling = β0 + β1 d_sports_team + ε

```
. regress d_gambling d_sports_team
```

Source	SS	df	MS	Number of obs	=	3,186
				F(1, 3184)	=	17.81
Model	1.41635323	1	1.41635323	Prob > F	=	0.0000
Residual	253.151443	3,184	.079507363	R-squared	=	0.0056
				Adj R-squared	=	0.0053
Total	254.567797	3,185	.079927095	Root MSE	=	.28197

d_gambling	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
d_sports_team	.042169	.009991	4.22	0.000	.0225794	.0617585
_cons	.8913317	.007067	126.13	0.000	.8774754	.9051879

ε

participating in a sports team does have a positive effect on whether or not the participant has

gambled in the past year. The coefficient is positive and has value .0422. This means on average, regular participation in sports teams increases the probability of gambling by .042. Furthermore, the p-value is 0, which means the effect of regular participation in team sports on the probability of gambling is significant at any significance level.

However, we are unable to say that the association we observe between regular participation in sports teams and gambling is causal since there may be many confounding variables that cause bias in our coefficient estimates. We will attempt to address this in the next model that we estimate.

Proposed Control Variables

Looking through the questionnaire we found several variables that we suspect have a direct impact on the probability of gambling and also have an effect on regular participation in team sports. Thus, we suspect these variables to be confounders and hence are biasing the results we observed in model 1. For each variable, we explain our reasoning below, for why we think it is a confounder.

- d_income: We believe that if a student came from a family with higher income, they would be more likely to gamble because they would have less repercussions for losing money. Thus, we suspect income has a direct effect on gambling. We also believe that students with higher incomes are more likely to participate in team sports, as they likely have more resources to buy the equipment and training needed to participate in sports.
- d_smoke_regularly: We believe that if a student has smoked before, they are more likely to gamble because both smoking and gambling in high school are illegal and considered “risky”, thus smokers have a higher risk tolerance and are also more likely to gamble. Thus, smoking has a direct effect on gambling. Also, we believe those who smoke are likely to be more social, and thus more likely to be on a sports team.
- d_gender: We believe that males are more likely to take part in risky behavior such as gambling. Thus, we believe gender has a direct effect on probability of gambling. We also believe males are more likely to participate in team sports, as on average males tend to play more sports than females.
- d_both_parents: We suspect that not living with both parents, may cause a student to experience more psychological trauma and hence they may turn to gambling as a way

to relieve the stress. Hence, whether or not a student lives with both parents has a direct affect on gambling probability. We also believe that students who only have one parent may have to do more chores at home and have less time to participate in team sports. Thus, whether or not a student lives with both parents affects their participation in team sports.

- d_drug_use: As with d_smoking, we believe that if a high schooler is doing drugs, they are more prone to doing risky behavior and thus more likely to have participated in gambling. We also believe that these students who enjoy risky situations are more likely to participate in team sports, as there is a certain level of risk involved in sports.
- d_drink: We believe that if a student has exhibited a habit of drinking, they are more likely to gamble as well. This is because they are more prone to engaging in risky behaviors in general and thus also more likely to engage in sports.
- d_church: We believe that those attending church are less likely to be attracted to risky behaviors and thus less likely to gamble. We also believe that those attending church will have less time to participate in sports teams and thus those attending church are less likely to be on a sports team.

Model 2

Linear Probability Model of Gambling based on Participation in Team Sport with control variables:

$$d\_gambling = \beta0 + \beta1\ d\_sports\_team + \beta2\ d\_income + \beta3\ d\_gender + \beta4\ d\_both\_parents + \beta5\ d\_drink + \beta6\ d\_drug\_use + \beta7\ d\_church + \beta8\ d\_sports\_team + \varepsilon$$

```
. // LPM: gambling on sports team  with controls
.
. regress d_gambling d_sports_team d_income d_smoke_regularly d_gender d_both_parents d
> _drink d_drug_use d_church
```

Source	SS	df	MS	Number of obs	=	3,149
				F(8, 3140)	=	6.04
Model	3.82900529	8	.478625661	Prob > F	=	0.0000
Residual	248.804847	3,140	.079237212	R-squared	=	0.0152
				Adj R-squared	=	0.0126
Total	252.633852	3,148	.080252177	Root MSE	=	.28149

d_gambling	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
d_sports_team	.0447448	.01024	4.37	0.000	.0246671	.0648225
d_income	.0066804	.0174072	0.38	0.701	-.0274504	.0408111
d_smoke_regularly	.0430636	.0178655	2.41	0.016	.0080344	.0780928
d_gender	.0351318	.0101275	3.47	0.001	.0152745	.054989
d_both_parents	-.018237	.0113622	-1.61	0.109	-.0405152	.0040411
d_drink	-.0350001	.0105177	-3.33	0.001	-.0556223	-.0143778
d_drug_use	-.001206	.0113511	-0.11	0.915	-.0234623	.0210504
d_church	-.0171289	.0132386	-1.29	0.196	-.043086	.0088282
_cons	.9005463	.0196024	45.94	0.000	.8621114	.9389811

In this second model, we trained a linear probability model and included what we consider to be relevant confounding variables that we had measurements for in the dataset in order to remove the bias in the coefficient estimates caused by the confounding variable.

From the results, we see that even after adding the control variables for the confounders, the coefficient of d_sports_team is still positive and significant. The value of the coefficient is 0.044 meaning that regular participation in team sports, would on average, cause the probability of gambling to increase by 0.044, holding all the other independent variables constant. The value of the coefficient on d_sports_team in this model is also very similar to the coefficient obtained in model 1. Furthermore, the p-value is 0, meaning that even after adding controls for the potential confounders, the effect of regular participation in team sports on gambling is still significant at any significance level.

Even though we have added controls for the potential confounders, we are still hesitant to interpret the association between participation in sports teams and gambling as causal since there are many problems associated with the linear probability model. The first problem is that in a linear probability model, the error term can only take two distinct values for each value of the independent variable, and these two values are different for each value

of the independent variable. Thus, this breaks the exogeneity assumption for the linear probability model, meaning that our coefficient estimates may be biased. Furthermore, the error term has different variances for different values of the independent variable, meaning that it is heteroskedastic which would lead the SE to be wrong and hence we cannot rely on the t-statistic for hypothesis test. We will attempt to address these issues by estimating a logistic regression as our next model.

Model 3

Logistic Regression Model of Gambling based on Participation in Team Sport with control variables:

$log\ (\frac{p}{1-p})$ = β0 + β1 d_sports_team + β2 d_income + β3 d_gender + β4 d_both_parents + β5 d_drink + β6 d_drug_use + β7 d_church + β8 d_sports_team + ε

$p\ =\ probability\ of\ gambling$

```
. logistic d_gambling d_sports_team d_smoke_regularly d_gender d_both_parents d_drink d_drug_
> use d_church  d_income

Logistic regression                                     Number of obs =  3,149
                                                        LR chi2(8)    =  48.59
                                                        Prob > chi2   = 0.0000
Log likelihood = -913.48413                             Pseudo R2     = 0.0259
```

d_gambling	Odds ratio	Std. err.	z	P>|z|	[95% conf.	interval]
d_sports_team	1.768488	.2334709	4.32	0.000	1.365303	2.290738
d_smoke_regularly	1.826948	.4783073	2.30	0.021	1.093644	3.051944
d_gender	1.573829	.2060832	3.46	0.001	1.217581	2.03431
d_both_parents	.7903407	.1170471	-1.59	0.112	.5912265	1.056513
d_drink	.6373756	.0881544	-3.26	0.001	.4860345	.8358411
d_drug_use	.9832063	.1406482	-0.12	0.906	.7428135	1.301396
d_church	.8206154	.1302848	-1.25	0.213	.6011721	1.120161
d_income	1.09054	.2359074	0.40	0.689	.7136879	1.666383
_cons	9.642626	2.385417	9.16	0.000	5.937764	15.65913

We estimated a logistic regression model, to address the limitations of the linear probability model which we estimated in Model 2. We added control variables for all of the confounding variables that we discussed earlier in order to prevent them from biasing our coefficient estimates.

From the results of this model, we observe that the coefficient on sports teams is 1.768. Since we used the logistic command in Stata, this is the exponentiated coefficient. Thus, this can be interpreted as on average, participating in team sports increases the odds of gambling by 76.8% holding all of the other independent variables constant. Thus, in models 1, 2, and 3 the sign of the coefficient on sports_team is positive. Furthermore, we see that the p-value for the coefficient on sports teams is 0. This means that similar to models 1 and 2, the effect of participation in team sports on the odds of gambling is significant at any significant level.

Discussion/Conclusion:

Results

The main takeaway from the result is that in both the linear probability model and the logistic regression model with controls, the effect of participation in team sports on gambling is both positive and significant. In the LPM model with controls, we observed that those who participate in team sports, on average, have a .0447 higher probability of gambling compared to those who don't, holding all the other independent variables constant. Similarly, in the logistic regression model, we observed that participating in team sports is associated with a 76.8% increase in the odds of gambling. Thus, from these results we can conclude that there is suggestive evidence for a strong and positive effect of participation in team sports on gambling.

Limitations

Although we have suggestive evidence for a strong and positive relationship between high school sports participation and gambling, we are unsure if it is causal as other confounders may exist that bias our results and that we do not have measurements of. For instance, risk appetite may be a confounding variable, since those with a higher risk tolerance are more likely to participate in a risky activity such as gambling are also more likely to play sports, which has an element of risk. However, we do not have a measure for risk appetite in the dataset.

While there is not sufficient evidence in this dataset to prove causality, it would be beneficial to introduce an instrument in future studies. One such instrument we could explore could be whether family members are sports fans. As sports fans would likely incentivize their children to join a sports team, there is no apparent relation with gambling. Thus, this

instrument would only affect the dependent variable through the explanatory and wouldn't be correlated with the error term. Perhaps this could be done by conducting the same survey, with the added question of whether their family members are sport fans.

Policy implications

To address the limitation of this dataset in determining causality, we suggest that the CDC redesign their survey to include questions that could be transformed into instruments to study causality. This could help researchers study the causal effect of various factors that contribute to risky behavior.

Regardless of causality, a strong positive association exists between high school team sports participation and gambling. Schools should use this finding to implement programs to educate student athletes on the effects of excessive gambling. For example, all sports team coaches could be required to lecture their players on the potentially harmful effects of gambling. Furthermore, a class dedicated to lecturing students on risky behaviors can be added to the physical education curriculum for all students to target those students who participate in team sports, but are not in a varsity team. The aim of these programs is to deter students who regularly participate in team sports from gambling by teaching them about the implications of risky behavior.

Supplemental Material:

Data Analysis Do-file