Sangyoon Park

Prof. Ming-Wen An

Mathematics 242

9 December 2015

Gender, Race, and Illegal Drug Experience Are Persistent Predictors of Sexually Transmitted Diseases:

Logistic Regression Modeling Based on a Representative Sample of the U.S. Population

Abstract

This study examines the relationship between STD experience and previously identified risk factors of STDs based on a representative sample of the U.S. population from NHANES 2007-2008 data. The study also examines other relevant variables as potential predictors of STDs. The result indicates that gender, race, and illegal drug experience are persistent predictors of STDs and that education level is another potentially significant predictor of STDs. The result also shows non-significance for alcohol use and the number of sexual partners, suggesting that these variables may not be as reliable as other previously identified predictors of STDs. Implications and limitations of the study are discussed.

Sex is a wonderful thing. It gives us pleasure, it gives us a new life, and it gives us a family. But it also gives us a disease. Sexually transmitted diseases (STDs) still form a significant health problem in the U.S. but are largely unrecognized by many due to the social stigma associated with them. Since STDs are more easily preventable than other diseases, researchers have tried to identify risk factors of STDs. Past studies indicate that STDs are more associated with certain behavioral factors such as substance abuse, drug injection, having unprotected sex, and having multiple sexual partners (Mayo Clinic; Cook & Clark, 2005; Finer, Darroch, & Singh, 1999; Ericksen & Trocki, 1992). Studies have also identified association between STDs and certain demographic characteristics such as poverty, being an African American,

being a female, and being an adolescent or a young adult (Healthy People 2020; Laumann & Youm, 1999). However, many of these studies are based on a sample of a certain population subgroup (e.g., African Americans in a Southern state), which limits the generalizability of their findings. Hence, this study aims to examine whether these previously identified risk factors of STDs can still be good indicators of STDs in a more representative sample of the U.S. population. Furthermore, the study probes whether other relevant variables including marital status, education level, smoking experience, and sleep problems[1] are also associated with STDs.

<div align="center">**Methods**</div>

National Health and Nutrition Examination Survey (NHANES) 2007-2008 data files were used to derive 15 variables of interest including: STD experience (STD), gender (GENDER), age (AGE), race/ethnicity (RACE), educational level (EDUC), marital status (MARRIED), annual family income (INCOME), number of sexual partners (PARTNERS), condom use (CONDOM), illegal drug use (DRUG), illegal drug injection experience (INJECT), alcohol use (ALCOHOL), smoking experience (SMOKING), serious sleep trouble (SLEEP1), and sleep disorder (SLEEP2).

Some variables such as AGE and GENDER were straightforward and could be directly used from the original NHANES dataset. But other variables such as STD and PARTNERS did not have precisely corresponding survey items, so had to be derived from relevant items in the NHANES dataset. For such variables as EDUC and INCOME, the NHANES dataset contained precisely corresponding items but with too many response levels. These variables were also adapted to fit the purpose of this study.

For practical reasons, the final dataset was derived to include only fully complete observations for the 15 variables of interest. It included a total of 2052 males and females across the U.S. aged 20-59. Of the 15 variables, 2 were continuous, 9 were binary, and 4 were multilevel categorical.

Pie chart was used to explore baseline characteristics of binary and other categorical variables. Histogram and box plot were used to examine baseline characteristics of continuous variables. Then, bar

---

[1] Sleep research shows that sleep problems are associated with psychological disorders, which in turn are associated with higher risks of STDs (Magidson et al., 2014).

graphs were constructed to explore the relationship between STD and other categorical (binary and multilevel) variables. The relationship between STD and continuous variables were explored by plotting logit p vs. deciles of each continuous variable.

An unadjusted logistic regression model was fitted for each variable as the single predictor for STD, and the resulting model was put to the Wald Test or Likelihood Ratio Test (LRT) to examine the statistical significance of the corresponding predictor variable. Fuller logistic regression models were then constructed through the forward selection procedure: statistically significant predictor variables were added one by one, in their order of significance level, and each resulting model was put to the Wald Test or LRT to ensure each newly added variable is still statistically significant. The process was repeated until the model has as many significant predictor variables as possible, which was then adopted as the final model.

Finally, the C-index was calculated to examine the final model's predictive ability.

Results

Of the 2052 observations in the dataset, about 9% reported having been ever diagnosed with a STD. About 67% reported having ever used any illegal drug but only about 4% reported having injected any illegal drug. There was an almost equal divide between two genders (49% male, 51% female), but the majority of the observations were white (50%). The average age was 38.4 ($SD = 10.9$). Other baseline characteristics are summarized in Table 1.

In the exploratory analysis with bar graphs, DRUG, EDUC, GENDER, INJECT, MARRIED, and RACE each showed considerable group differences in STD, suggesting these variables may be significant predictors of STD. Similarly, mild linearity was observed in plotting logit p vs. deciles of PARTNERS (Figure 3), suggesting PARTNERS may be a significant predictor of STD.

The subsequent statistical testing through the unadjusted logistic regression largely agreed with these exploratory observations: DRUG, GENDER, and RACE were found to be highly significant ($p < 0.001$); and EDUC, INJECT, and MARRIED to be moderately significant ($p < 0.05$). However, the

unadjusted logistic regression reported PARTNERS as statically non-significant. Other variables were also found to be statically non-significant for predicting STD.

The forward selection procedure resulted in a logistic regression model with 5 predictor variables: GENDER, DRUG, RACE, EDUC, and INJECT. The model has 11 coefficient terms as summarized in Table 2.

The model predicts a minimal odds (hence a minimal probability) of STD for an American white male adult educated below 9th grade who has not ever used or injected illegal drug. The odds of STD for an American female adult is predicted to be twice as high as that for her male counterpart with the same race/ethnicity, education level, and illegal drug experience. The model also predicts that, with other conditions (gender, education level, and illegal drug experience) being equal, the odds of STD for an African American adult is more than 1.5 times higher than that for an American adult of a different race/ethnicity. An American adult with college education or above is also predicted to have a higher—more than 1.5 time—odds of STD than his or her less educated counterpart (with the same gender, race/ethnicity, and illegal drug experience). Finally, the odds of STD for an American adult with any illegal drug experience is predicted to be twice as high as that for his or her counterpart without such experience. This difference is predicted to be even bigger if the drug use has engaged needle injection.

The model's C-index is about 0.68, indicating a moderate level of predictive capacity (0.50 = chance level).

Discussion

The final model includes several predictor variables previously identified as risk factors of STDs. Consistent with past research, the model indicates that being a female or an African American is associated with higher risks of STDs. The model also agrees with the previous findings that illegal drug use and drug injection are associated with higher risks of STDs. Notably, the model predicts association between STDs and drug injection to be greater than association between STDs and drug use, accrediting the common opinion that needle sharing is more risky than insufflation. Overall, these results suggest that gender, race, and illegal drug experience are persistent and reliable predictors of STDs.

However, the model rules out several other predictor variables previously identified as risk

factors of STDs. Specifically, the current logistic regression modeling found age, annual family income,

number of sexual partners, condom use, and alcohol use to be non-significant predictors of STDs even

though past research identifies them as indicators of STDs (Mayo Clinic; Healthy People 2020; Cook &

Clark, 2005; Finer, Darroch, & Singh, 1999). Some explanations are in order. First, the discrepancy

between what this study intends to measure and what was actually measured in the original NHANES

dataset seems to the have contributed to non-significance of some of these variables. For instance, the

actual survey item for condom use asks of recent practice (i.e. unprotected sex in the past 12 months),

which may not be representative of the respondent's overall condom use history (it may well be that the

past contraction of a STD has since made the person particularly careful about condom use). A similar

problem is relatable to non-significance of age but in a different way: in this case, the survey item for

STD experience seems to have been inadequate as it simply asks of the total experience of any STD ("Has

a doctor or other health care professional ever told you that you had [disease name]?"). The respondent's

reported age hence does not necessarily reflect the time of contraction, and this makes it impossible to

distinguish STD contraction at a young age from that at an older age.

However, annual family income, number of sexual partners, and alcohol use are free of this

problem, so their non-significance requires different explanations. A possible explanation for the number

of sexual partners is that having sex with just one partner who has had multiple sexual partners can also

increase the risk of STD contraction (Healthy People 2020; Laumann & Youm, 1999; Finer, Darroch, &

Singh, 1999) but this qualitative aspect is not captured in the sheer number of sexual partners. Yet, studies

still indicate that this sheer number is associated with higher risks of STDs (Mayo Clinic). Furthermore,

the exploratory plot indicates linearity between the number of sexual partners and STD experience

(Figure 3). It need be also noted that the distribution of the variable displays a strong positive skew

(Figure 2; some observations excluded for illustrative purpose only), with several extreme outliers (Figure

1). These considerations then point to a different possibility that the non-significance of the variable is

due to the peculiarity of the current dataset and that the NHANES data from a different year may produce

a different result. We may also resort to this latter explanation to account for the non-significance of annual family income and alcohol use as no better alternatives seem available for these variables.

It is noteworthy that education level was found to be a significant predictor of STDs since past STD research does not explicitly identify it as a risk factor. Even more noteworthy is the direction of the prediction: as mentioned above, current data suggest that college or more advanced education is associated with higher risks of STDs, which contradicts the common opinion that lower education level is associated with more health problems. Could it be that the college educated learn more about many things including sexuality but not necessarily more about the dark side of sex? Future studies are called upon to decide whether this counterintuitive finding is a result of sampling error or reflective of some unfamiliar truth.

Finally, this study indicates some significance for marital status even though it was ruled out from the final model. Hence, future studies are necessary to decide whether this exclusion was a valid step or another instance of sampling error.

Conclusion

This study began with two questions. First, would previously identified risk factors of STDs still maintain their status in a more representative sample of the U.S population? Second, can other relevant variables be also significant predictors of STDs? The current analysis answers these questions in both the positive and the negative. The logistic regression modeling suggests that gender, race, and illegal drug experience are persistent predictors of STDs and that education level is another potentially significant predictor of STDs. On the other hand, the result rules out alcohol use and the number of sexual partners from the list of significant predictors of STDs with no particularly relatable confounds, implying these variables may not be as persistent and reliable as gender, race, or illegal drug experience. Hence, this study opens up a new possibility that different predictors of STDs may considerably differ in their levels of significance, especially in a more representative population sample. Future studies are called upon to address and critically examine this possibility, which will certainly contribute to enhancing our knowledge and prevention of STDs.

References

Cook, R. L., & Clark, D. B. (2005). Is there an association between alcohol consumption and sexually transmitted diseases? A systematic review. *Sexually Transmitted Diseases, 32*(3), 156-164.

Ericksen, K. P., & Trocki, K. F. (1992). Behavioral risk factors for sexually transmitted disease in American households. *Social Science & Medicine, 34*(8), 843-853.

Finer, L. B., Darroch, J. E., & Singh, S. (1999). Sexual partnership patterns as a behavioral risk factor for sexually transmitted diseases. *Family Planning Perspectives,* 228-236.

Healthy People 2020 (2015, December 7). *Sexually transmitted diseases.* Retrieved from http://www.healthypeople.gov/2020/topics-objectives/topic/sexually-transmitted-diseases

Laumann, E. O., & Youm, Y. (1999). Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: a network explanation. *Sexually Transmitted Diseases, 26*(5), 250-261.

Magidson, J. F., Blashill, A. J., Wall, M. M., Balan, I. C., Wang, S., Lejuez, C. W., & Blanco, C. (2014). Relationship between psychiatric disorders and sexually transmitted diseases in a nationally representative sample. *Journal of Psychosomatic Research, 76*(4), 322-328.

Mayo Clinic (2014, August 19). *Risk factors.* Retrieved from http://www.mayoclinic.org/diseases-conditions/sexually-transmitted-diseases-stds/basics/risk-factors/con-20034128
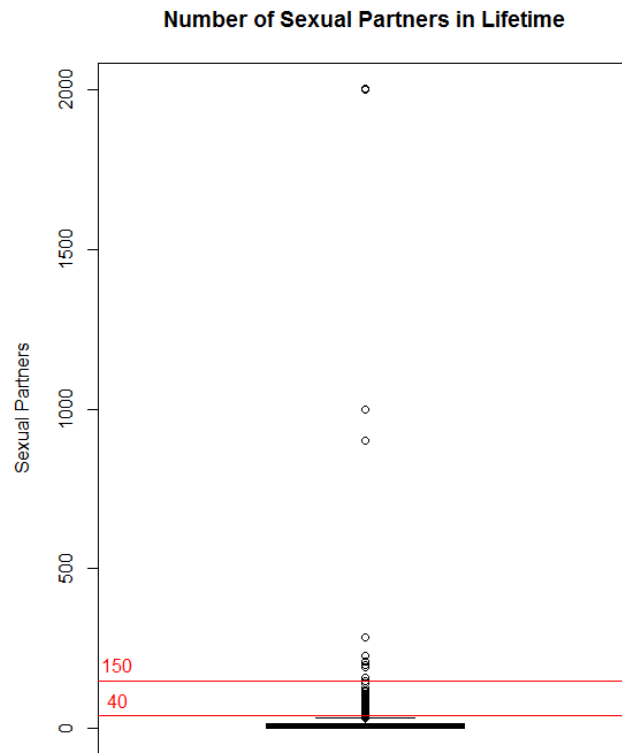
Number of Sexual Partners in Lifetime

Figure 1. Distribution of Number of Sexual Partners: Box Plot

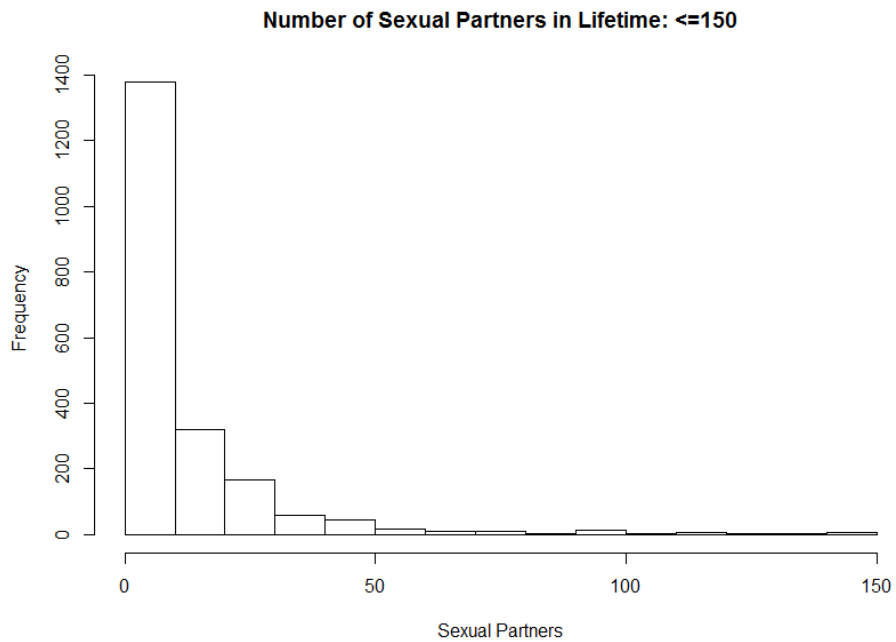**Number of Sexual Partners in Lifetime: <=150**

Figure 2. Distribution of Number of Sexual Partners: Histogram

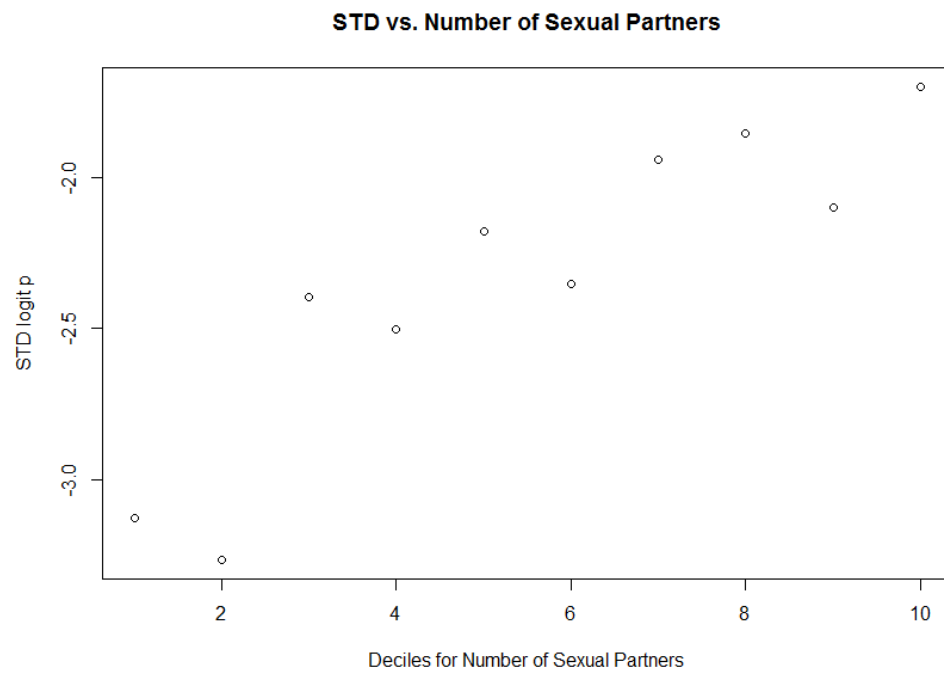**Figure 3. STD vs. Number of Sexual Partners**

<p align="center">**Table 1. Baseline Characteristics**</p>

Variable	Observations ($n = 2052$)
Age, *yr*	$M = 38.4, SD = 10.9$
Gender	
Male	49%
Female	51%
Race/Ethnicity	
White	50%
Black	18%
Hispanic	29%
Other	3%
Educational Level	
Below 9th Grade	5%
9-11th Grade	17%
High School Grad / GED	23%
Some College / AA	31%
College Grad or Above	23%
Annual Family Income	
Poverty Level	19%
Working Class	23%
Lower Middle Class	28%
Upper Middle Class	30%
Married	
Yes	56%
Number of Sexual Partners in Lifetime	$M = 18.6, SD = 95.1$
Ever diagnosed with STD	
Yes	9%
Sex with Condom during Past Year	
Never	27%
Sometimes	33%
Always	40%
Ever used any illegal drug	
Yes	67%
Ever injected any illegal drug	
Yes	4%
Ever had 5 or more drinks everyday	
Yes	16%
Smoked at least 100 cigarettes in life	
Yes	50%
Ever told doctor about trouble with sleeping	
Yes	23%
Ever diagnosed with sleep disorder?	
Yes	7%

Table 2. Summary of the Multiple Linear Regression Model

Coefficient	Estimate	OR	Standard Error	z-value	Probability (> | z |)
Intercept[†]	-3.3421	0.0354[‡]	0.4922	-6.790	1.12e-11***
Female	0.8101	2.2482	0.1652	4.905	9.32e-07***
Illegal Drug Use	0.6851	1.9839	0.1938	3.536	0.000406***
Black	0.5750	1.7771	0.1854	3.101	0.001926**
Hispanic	-0.5052	0.6034	0.2302	-2.195	0.028179*
Other Race	0.0305	1.0310	0.4502	0.068	0.945958
9-11th Grade	-0.0904	0.9136	0.4852	-0.186	0.852276
High School Grad / GED	-0.1956	0.8223	0.4769	-0.410	0.681652
Some College / AA	-0.1428	0.8669	0.4678	-0.305	0.760178
College Grad or Above	0.4831	1.6211	0.4696	1.029	0.303617
Illegal Drug Injection	0.8360	2.3071	0.3446	2.426	0.015278*

† Reference group: Male, No Illegal Drug Use, White, Less than 9th Grade, and No Illegal Drug Injection
‡ Odds for the baseline (i.e. reference group) in this case
* p < .05; ** p < .01; *** p < .001