

# Virginity Pledges & Abstinence Revisited: A Propensity Score Approach

WORKING PAPER

July 2006

Norbert J. Michel  
Nicholls State University  
[Norbert.michel@nicholls.edu](mailto:Norbert.michel@nicholls.edu)

Abstract:

The effectiveness of abstinence-only sexual education is an important public policy issue for at least two reasons. First, federal tax dollars are being used to fund these programs at an increasing rate. Second, research suggests participants of these programs may be less likely to use condoms at initial intercourse, thus increasing their risk of unwanted pregnancy and sexually transmitted disease (STD) infection. Research using the latest wave of the National Longitudinal Study of Adolescent Health (Add Health) data suggests that STD infection rates are no different among young adults who, earlier in their lives, had taken a pledge to remain abstinent until marriage and those who did not take such a pledge. This research was attacked by proponents of abstinence-only education for being analytically flawed. The present paper aims to show that, when correcting for endogeneity problems, the conclusions of the earlier research stand. That is, we cannot say there is any difference in the STD infection rates between young adults who, as adolescents, pledged to remain abstinent until marriage and those who did not.

## I. Introduction

For more than a decade advocacy groups have been promoting sexual education programs that encourage young people to abstain from sexual activity. Those programs that encourage “abstinence only” have drawn criticism for failing to provide a more complete sexual education. For instance, the *WAIT Training* program promotes abstinence and does not promote condom use as an effective way to reduce the risk of pregnancy or STD infection. Because some evidence suggests that adolescents who take virginity pledges are less likely to use condoms at first intercourse, critics have suggested that a better approach to sexual education would be to present students with a full range of options, including both abstinence and condom use. This approach, it is argued, would more effectively prevent unwanted pregnancies and the spread of sexually transmitted diseases (STDs). Because abstinence programs are receiving more and more federal funding, and because unwanted pregnancies and STDs are a public health concern, resolving this matter is a nontrivial task.

At the heart of this controversy is a paper by Bearman and Bruckner (2001), a study which used data from the first wave of the National Longitudinal Study of Adolescent Health (Add Health). In the 2001 paper, the authors reported that taking abstinence pledges appeared to delay adolescents’ first sexual intercourse, but individuals taking the pledges (referred to as “pledgers” in the present paper) were unlikely to delay intercourse until marriage and unlikely to use a condom at first intercourse. For example, Bearman and Bruckner (2001) estimate the odds that pledgers will use contraceptives at first intercourse is about one-third lower than for non-pledgers. While this lower

probability for condom use may have upset some people, the controversy was rather mild until the release of a second paper.

In Bruckner and Bearman (2005), the authors used data from a third wave of the Add Health survey, one that included the results of (many) respondents' STD tests. In this third wave, Add Health collected urine samples from respondents (now between 19 and 25 years old) and tested for the presence of three STDs – Chlamydia (CH), Gonorrhea (GC) and Trichomoniasis (TR).<sup>1</sup> Bruckner and Bearman (2005) reports “Contrary to expectations, we found no significant differences in STD infection rates between pledgers and nonpledgers, despite the fact that that they transition to first sex later, have less cumulative exposure, fewer partners, and lower levels of nonmonogamous partners.” Although some may view the opportunity to measure STD rates based on actual urine tests as opposed to relying on self reported STD diagnoses or symptoms (which are included in waves 1 and 2 of Add Health), some critics of these new findings do not share this view.

A Heritage Foundation paper, for example, criticized Bruckner and Bearman for using only the STD urine-test results to measure STD infection rates for pledgers and non-pledgers. The Heritage Foundation authors, Rector and Johnson, also present their own analyses which appears to contradict the results in Bruckner and Bearman (2005). Rector and Johnson (2005, June) states that “Bearman and Bruckner’s [sic] conclusion that virginity pledgers have the same STD rates as non-pledgers is clearly the result of serious limitations in their analytic methods. Our current paper shows that taking a virginity pledge in adolescence is associated with a substantial decline in STD rates in

---

<sup>1</sup> Additionally, 7,000 females who reported they had sex as of their wave 3 interviews were randomly selected for Human Papilloma Virus (HPV) analyses.

young adult years.” While Bruckner and Bearman (2005) relied on the results from a urinalysis to code their STD infection variable, Rector and Johnson employ additional measures, such as self reported diagnoses and symptoms. The Rector and Johnson paper also differs from Bruckner and Bearman (2005) in that it employs logistic regression, whereas the Bruckner and Bearman paper employs descriptive statistics instead.

The present paper argues that the findings presented in Rector and Johnson (2005) are likely due to a problem known as self-selection bias. This problem stems from the fact that adolescents are not randomly assigned to groups of “pledgers” and “non-pledgers.” In other words, if adolescents who take virginity pledges are more likely to avoid sexual intercourse due to some unobservable characteristics, then any observed correlation between virginity pledges and STDs may be attributable to these unobservable characteristics *instead* of the respondents’ pledge status. Rector and Johnson (2005) does not address this issue.

In the present paper, the propensity score method (introduced by Rosenbaum and Rubin (1983)) is used to control for self-selection bias. The propensity score method is used in both a descriptive statistic and logistic regression framework. To implement the propensity score method, the probability that an individual will take a virginity pledge is estimated, reflecting that individual’s “propensity” for taking the pledge. This method allows researchers to compare STD rates for pledgers vs. non-pledgers who exhibit a similar propensity for taking the pledge, thus controlling for unobservable characteristics.<sup>2</sup>

---

<sup>2</sup> I will also use “twin-matching” method which examines differences in STD rates between twins with differing pledge status. This method controls for endogeneity because twins with opposite pledge status should still possess most of the same unobservable characteristics. As explained below, the current version of this paper uses a public-use version of Add Health, a limitation of which is that twins cannot be matched.

## II. Data and Methodology

### A. Overview of Data

There are two versions of the National Longitudinal Study of Adolescent Health (Add Health)<sup>3</sup> data, the “public use” and the “restricted use” versions. The restricted use version can be thought of as the “full” data set, for which use is restricted due to sensitive respondent information. The public use version of the Add Health data is a random sample from the restricted use file. The present paper uses data from the public use version of Add Health (all three waves) because researchers have to undergo a review process and enter into a contract to gain access to the full file. The present author is currently seeking the necessary funding to use the restricted use file, and expects that using the full data set will improve the quality of the paper.

The aforementioned papers by Bearman and Bruckner, as well as by Rector and Johnson, employed the restricted use version of the Add Health data. The key innovation of this paper is the use of the propensity score method (introduced by Rosenbaum and Rubin (1983)) to control for possible endogeneity problems, such as the self-selection bias of adolescents choosing to take virginity pledges. Because it is impossible to match siblings in the *public use* version of the file, the current draft of this paper does not include results employing the twin-matching method.

---

<sup>3</sup> The National Longitudinal Study of Adolescent Health (Add Health) is a program project designed by J. Richard Udry and Peter Bearman and funded by a grant HD31921 from the National Institute of Child Health and Human Development to the Carolina Population Center, University of North Carolina at Chapel Hill, with cooperative funding participation by the following agencies: The National Cancer Institute; The National Institute of Alcohol Abuse and Alcoholism; the National Institute on Deafness and other Communication Disorders; the National Institute on Drug Abuse; the National Institute of General Medical Sciences; the National Institute of Mental Health; the Office of AIDS Research, NIH; the Office of Director, NIH; The National Center for Health Statistics, Centers for Disease Control and Prevention, HHS; Office of Minority Health, Centers for Disease Control and prevention, HHS, Office of the Assistant Secretary for Planning and Evaluation, HHS; and the National Science Foundation. Additional funding for the analyses reported in this paper were provided by the Office of Population Affairs.

## B. Overview of Propensity Score

One common approach to the propensity score method is to estimate the probability that individuals will engage in a certain behavior by employing a probit model. This estimated “propensity” for each individual to engage in the given behavior can then be included in a regression. Alternatively, respondents’ propensity score can be used to split the observations into equal strata (typically, at least five equal groups). Then, differences can be taken for the treatment and control groups within each propensity score strata. For example, we would test for a difference in STD rates between pledgers and non-pledgers within the first quintile of the propensity score, the second quintile, and so on. In the present paper, a probit analysis is used to predict the likelihood that an adolescent will take a virginity pledge. Differences between pledgers and non-pledgers are then examined using both difference-of-proportion tests and logistic regressions.

Although very little research has been produced specifically to *predict* whether an adolescent will take a virginity pledge, a substantial body of research exists which examines adolescents’ determinants of first intercourse. One theme that runs through the literature is that the social context that adolescents live in is an important determinant of their behavior.<sup>4</sup> Adolescents’ families, schools, peer groups and romantic relationships interact with adolescents’ individual characteristics – such as pubertal development, gender and cognitive skills – to influence behavior. Teens living in single-parent families (or with stepparents), for example, tend to initiate sexual activity sooner than teens living in two-parent families.

---

<sup>4</sup> For a summary of the sexual initiation research, see Hofferth (1987). A few additional references (for various themes in the related literature) are as follows: Bearman and Bruckner (1999), Resnick et al. (1997), Thornton and Camburn (1989), Miller (1998) and Udry (1988).

In the present paper, the variables included in the propensity score probit model are as follows: race, gender, age (at wave 1 interview), whether the respondent is romantically involved, whether the respondent has had sexual intercourse, the respondent's attitude toward sex (an index of how strongly respondents believe they will gain more respect, be viewed as more attractive, or feel less lonely if they have sex), whether the respondent believes that birth control is morally wrong, a "mutuality" index (an Add Health constructed network variable), how badly the respondent would like to be romantically involved (in the next year), whether the respondent has had suicidal thoughts (in the past year), number of times the respondent admits to drinking and driving, how bad the respondent believes it would be if she (he) became (got someone) pregnant, a religiosity index, a friend network "density" index (an Add Health constructed network variable), and the following "context" variables: urbanicity, proportion of area with 1989 income below poverty line, and proportion of "type" of family (married, non-married, etc). These variables serve as covariates in a weighted probit model, with the outcome variable set to one for anyone who has taken a virginity pledge, and to zero for anyone who responds they have not.<sup>5</sup>

One anomaly in the Add Health data is that some respondents give inconsistent answers to the question of whether they took a virginity pledge. For example, a respondent may have reported in waves 1 and 2 that she took a pledge, but reported in wave 3 that she had not taken a pledge. A possible cause of this inconsistency is that the wave 3 survey question was altered slightly (see Bruckner and Bearman (2005)).

Therefore, in the present paper, I follow the convention in Bruckner and Bearman (2005)

---

<sup>5</sup> An additional shortcoming of the public use file is that all clustering and design effects cannot be controlled for as well as with the restricted use file.

and report statistics for both “inconsistent” and “consistent” pledgers.<sup>6</sup> Inconsistent pledgers are those respondents who claimed, in any of the three waves, that they took a virginity pledge, while consistent pledgers are those for whom positive responses are not contradicted in either of the later waves. As for defining “STD rates,” I rely primarily on the measure used in Bruckner and Bearman (2005), which relies on respondents’ urinalysis results instead of relying on self reported data.<sup>7</sup> Of course, all appropriate weights are used for all statistics presented below.

NOTE: UNABLE TO TWIN ANALYSIS WITH PUBLIC USE DATA: The Add Health public use data includes responses from twins for approximately 200 adolescents. Additionally, 26 twins responded that they had taken a virginity pledge, while 73 twins responded they had *not* taken a pledge. To control for possible unobservable characteristics, we will test for significant differences in STD rates between twins who have/have not taken a pledge.<sup>8</sup>

### III. Results

The following results were derived from a preliminary analysis of the public-use version of the Add Health data. Using the probit model discussed above, the mean probability (propensity score) for predicting whether an adolescent would take a virginity pledge was approximately 22 percent, with a maximum predicted probability of approximately 73 percent. Only about 1,300 respondents have complete data for fitting

---

<sup>6</sup> {For purposes of this draft, only “inconsistent” pledgers were uses due to small sample problems}

<sup>7</sup> I will, possibly, expand to one other measure used in Rector and Johnson, but the broader this measure, the less reliable it becomes.

<sup>8</sup> Because siblings cannot be matched in the public use file, this portion of the analysis will be performed later, when access to the restricted use version is obtained.

the probit model, another shortcoming associated with using the public use file. Overall, the public-use file contains only 1,045 “inconsistent” pledgers, far short of the 1,622 reported in Bruckner and Bearman (2005), a paper that uses the restricted use file.

As seen in Table 1, a basic cross-tabulation of STD rates broken out by pledgers and nonpledgers, 7.27 percent of adolescents who report that they did *not* take a virginity pledge tested positive for an STD, while 8.45 percent of adolescents who report that they did take a virginity pledge tested positive for an STD. Table 2 presents the same type of cross-tabulation, only this time the statistics make use of the propensity score.

Specifically, the cross-tabulations are run for each of the five quintiles of the predicted propensity score. As seen on Table 2, there is only one pledger in the first strata (for those adolescents with the lowest predicted probability of taking a pledge) who tested positive for an STD (although this number represents a higher percentage than for the non-pledger group).

In the second-lowest propensity strata, a slightly higher percentage (9.09 percent) of pledgers than non-pledgers (8.99 percent) have an STD. This difference reverses for the third strata, with 5.88 percent of pledgers testing positive, and 6.82 percent of non-pledgers testing positive. In the fourth strata, the a higher percentage of pledgers than non-pledgers tested positive for an STD (almost 11 percent vs. 6 percent). Finally, in the fifth strata, there is also a higher percentage of pledgers than non-pledgers that tested positive for an STD (9 percent vs. 7.55 percent). Given the very small sample sizes, however, none of these differences are statistically significant. In other words, using the public use file, we cannot say that there is a difference in the STD rates between pledgers

and non-pledgers, even once we account for respondents propensity to take a virginity pledge.

For the logistic regression analysis, the results presented below suggest that controlling for self-selection bias among virginity pledgers may be necessary. First, a base logistic regression model is run, with a dichotomous variable for “STD test result” as the dependent variable, and only one covariate, a dichotomous variable for “pledge status.” To test whether the propensity score may be controlling for self-selection bias, a second model is run, one that includes the propensity score (*phat*) as an additional covariate. Table 3 shows that the addition of the propensity score increases the p-value of “pledge status” from .778 to .968 – a substantial reduction in statistical significance (to more fully control for the loss of observations, however, it will be necessary to use the restricted data set).

Table 4 presents the same type of analysis, but with additional covariates included in the base model. These additional covariates are age (at wave 1 interview), race and gender.<sup>9</sup> With the additional covariates, the difference is not as dramatic when the propensity score is included in the model, and the directional impact is not the same. However, the pledge status variable remains statistically insignificant, with a p-value of .95 in the base model, and of .94 in the model with the propensity score included.

{Hopefully, using the full version of the data will strengthen these results and allow for further analysis.}

#### IV. Conclusions – to come

---

<sup>9</sup> This approach is very similar to that of Rector and Johnson (2005). Unfortunately, because Rector and Johnson use the restricted use version of the Add Health data, I am unable to replicate their analysis.

**Table 1**

-> incpldg = 0			
Stdtest	Freq.	Percent	Cum.
0	2,118	92.73	92.73
1	166	7.27	100.00
Total	2,284	100.00	
-> incpldg = 1			
Stdtest	Freq.	Percent	Cum.
0	672	91.55	91.55
1	62	8.45	100.00
Total	734	100.00	

**Table 2**

```
. by incpldg: tabulate stdtest if quint ==1
```

-> incpldg = 0

Stdtest	Freq.	Percent	Cum.
0	197	94.71	94.71
1	11	5.29	100.00
Total	208	100.00	

-> incpldg = 1

Stdtest	Freq.	Percent	Cum.
0	10	90.91	90.91
1	1	9.09	100.00
Total	11	100.00	

```
. by incpldg: tabulate stdtest if quint ==2
```

-> incpldg = 0

Stdtest	Freq.	Percent	Cum.
0	172	91.01	91.01
1	17	8.99	100.00
Total	189	100.00	

-> incpldg = 1

Stdtest	Freq.	Percent	Cum.
0	20	90.91	90.91
1	2	9.09	100.00
Total	22	100.00	

```
. by incpldg: tabulate stdtest if quint ==3
```

-> incpldg = 0

Stdtest	Freq.	Percent	Cum.
0	164	93.18	93.18
1	12	6.82	100.00
Total	176	100.00	

-> incpldg = 1

Stdtest	Freq.	Percent	Cum.
0	48	94.12	94.12
1	3	5.88	100.00
Total	51	100.00	

```
. by incpldg: tabulate stdtest if quint ==4
```

-> incpldg = 0

Stdtest	Freq.	Percent	Cum.
0	153	93.87	93.87
1	10	6.13	100.00
Total	163	100.00	

-> incpldg = 1

Stdtest	Freq.	Percent	Cum.
0	50	89.29	89.29
1	6	10.71	100.00
Total	56	100.00	

```
. by incpldg: tabulate stdtest if quint ==5
```

-> incpldg = 0

Stdtest	Freq.	Percent	Cum.
0	98	92.45	92.45

**Table 3****Base Model**

Number of obs = 3018  
 Wald chi2(1) = 0.08  
 Prob > chi2 = 0.7777  
 Log psedulikelihood = -733.42215  
 Pseudo R2 = 0.00001

Stdtest	Odds Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
incpldg	1.052722	.1915811	0.28	0.778	.7368968	1.503907

**Model with Propensity Score**

Number of obs = 1082  
 Wald chi2(2) = 0.25  
 Prob > chi2 = 0.8840  
 Log Pseudolikelihood = -245.9304  
 Pseudo R2 = 0.0006

Stdtest	Odds Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
incpldg	1.01322	.3277739	0.04	0.968	.5374556	1.91014
phat	1.504189	1.318567	0.47	0.641	.2698622	8.384222

**Table 4****Base Model (with covariates)**

Number of obs = 3016  
 Wald chi2(4) = 34.18  
 Prob > chi2 = 0.0000  
 Log Pseudolikelihood = -719.04288  
 Pseudo R2 = 0.0193

Stdtest	Odds Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
incpldg	.9890255	.1755845	-0.06	0.950	.6983791	1.400631
age_1	1.051445	.0460096	1.14	0.254	.9647274	1.145303
race	1.196631	.0435346	4.93	0.000	1.114276	1.285073
gender	.6933057	.1140393	-2.23	0.026	.5022428	.9570524

**Model with Propensity Score**

Number of obs = 1082  
 Wald chi2(4) = 10.23  
 Prob > chi2 = 0.0691  
 Log Pseudolikelihood = -242.50731  
 Pseudo R2 = 0.0145

Stdtest	Odds Ratio	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
incpldg	1.024014	.3244835	0.07	0.940	.5502777	1.90559
age_1	.940701	.1066953	-0.54	0.590	.7531951	1.174886
race	1.196187	.072594	2.95	0.003	1.062042	1.347276
gender	.9643697	.229951	-0.55	0.584	.5131587	1.455953
phat	.998145	.8703998	-0.00	0.998	.1806887	5.513866

## References

- 2005 Brückner, Hannah, and Peter S. Bearman: After the Promise: The STD Consequences of Adolescent Virginity Pledges. *Journal of Adolescent Health* 36:271-278.
- 2001 Bearman, Peter, and Hannah Brückner: Promising the Future: Abstinence Pledges and the Transition to First Intercourse. *American Journal of Sociology* 106: 859–912.
- Bearman, Peter S. and Hannah Brückner. 1999. Peer Effects on Adolescent Girls' Sexual Debut and Pregnancy. Washington, D.C.: National Campaign to Prevent Teen Pregnancy.
- Hofferth, Sandra L. 1987. "Factors Affecting Initiation of Sexual Intercourse." Pp. 7-35 in *Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing*, Vol. II, vol. II, eds. Sandra L. Hofferth and Cheryl D. Hayes. Washington, DC: National Academy Press.
- Miller, Brent C. 1998. *Families Matter: A Research Synthesis of Family Influences on Adolescent Pregnancy*. Washington, D.C.: National Campaign to Prevent Teen Pregnancy.
- 2005 (August) *Teenage Sexual Abstinence and Academic Achievement*, Heritage Foundation White Paper, Robert Rector and Kirk Johnson.
- 2005 (June) *Adolescent Virginity Pledges, Condom Use and Sexually Transmitted Diseases*, Heritage Foundation White Paper, Robert Rector and Kirk Johnson.
- Resnick, Michael D., Peter S. Bearman, Robert Wm. Blum, Karl E. Bauman, Kathleen M. Harris, Jo Jones, Joyce Tabor, Trish Beuhring, Sieving Renee E., Marcia Shew, Majorie Ireland, Linda H. Bearinger, and Richard J. Udry. 1997. "Protecting Adolescents From Harm: Findings From the National Longitudinal Study on Adolescent Health." *Journal of the American Medical Association* 278(10):823-32.
- Rosenbaum, Paul R. and Donald B. Rubin, 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70, April, 41–55.
- Thornton, A. and D. Camburn. 1989. "Religious Participation and Adolescent Sexual Behavior and Attitudes." *Journal of Marriage and the Family* 51:641-53.
- Udry, Richard J. 1988. "Biological Predispositions and Social Control in Adolescent Sexual Behavior." *American Sociological Review* 53:709-22.

**Appendix A - Propensity Score Model**

```

Probit Regression
Number of obs =      1307
Wald chi2(17) =     162.05
Prob > chi2 =       0.0000
Log pseduolikelihood = -566.43558
Pseudo R2 =        0.1544
    
```

Incpldg	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
Race	.0029666	.0321435	0.09	0.926	-.0600336	.0659667
Gender	-.0659614	.103211	-0.64	0.523	-.2682513	.1363284
Age_1	.048804	.0438891	1.11	0.266	-.0372171	.1348252
Rominvol	-.306239	.2233175	-1.37	0.170	-.7439333	.1314552
Sex1	-.8035764	.1113417	-7.22	0.000	-1.021802	-.5853507
Sexattit	.2021369	.0641631	3.15	0.002	.0763795	.3278944
Bcmorbad	-.0255853	.044257	-0.58	0.563	-.1123275	.0611569
Mutualin	-1.431991	1.046598	-1.37	0.171	-3.483286	.6193044
Longrom	-.1355137	.0655442	-2.07	0.039	-.2639779	-.0070494
Suicdlth	-.1462597	.1495205	-0.98	0.328	-.4393146	.1467952
Xsdrinkd	-.435837	.2363848	-1.84	0.065	-.8991426	.0274687
Pregbad	-.071622	.064155	-1.12	0.264	-.1973634	.0541195
Religndx	-.3899843	.0670367	-5.82	0.000	-.5213739	-.2585947
Totnetde	-.9915866	2.20174	-0.45	0.652	-5.306918	3.323745
Bst90p19	.1348354	.0629454	2.19	0.029	.0144647	.2612062
Bst90p01	-.1899434	.0818731	-2.32	0.020	-.3504118	-.029475
Bst90p13	-.0610691	.0413104	-1.48	0.139	-.142036	.0198979
_cons	-.0610924	.9448104	-0.06	0.948	-1.912887	1.790702

**Appendix A - Summary Statistics for Propensity Score**

```

. predict phat
(option p assumed; Pr(incpldg))
(2462 missing values generated)
    
```

```

. sum phat
Variable      Obs      Mean      Std. Dev.      Min      Max
phat          1366      .2198877      .1635425      .0006488      .7314441
    
```