Why is Dropping Out of High School Bad? Dropouts and Sexually Transmitted Infections

D. Mark Anderson
Department of Economics
University of Washington
dma7@u.washington.edu
www.dmarkanderson.com

Claus C Pörtner
Department of Economics
University of Washington
cportner@uw.edu
faculty.washington.edu/cportner

February 2011

Abstract

People who drop out of high school fare worse in many aspects of life. We analyze whether there is a causal effect of dropping out of high school on the probability of contracting a sexually transmitted infection (STI). Previous studies on the relationship between dropout status and sexual outcomes have not empirically addressed self-selection effects. Using individual fixed effects estimations we find strong evidence that dropping out increases the risk of contracting an STI for females. Furthermore, we present evidence illustrating differences between the romantic partners of dropouts versus enrolled students. These differences suggest female dropouts may be more susceptible to contracting STIs because they become sexually active with significantly different types of people. These results point to a previously undocumented benefit of encouraging those at risk to stay in school longer.

I. Introduction

People who drop out of high school are believed to do substantially worse in many aspects of life compared to those who graduate from high school. Yet, despite intense public and academic interest, we have only a rudimentary understanding of how dropouts differ from those who stay, why dropouts leave school, and what the causal impacts of dropping out of school are. The aim of this paper is to test whether a relationship exists between dropping out and the risk of contracting a sexually transmitted infection (STI) after controlling for self-selection effects.

Previous research has established that increasing the minimum drop out age leads to higher income, better health, higher self-reported happiness, less crime, and lower probability of having a teenage birth (Lochner and Moretti 2004; Oreopoulos 2007; Black, Devereux and Salvanes 2008; Anderson 2010). What is less well understood is why dropping out affects so many areas of life. A number of theories have been advanced to explain why dropouts do significantly worse than those who stay in school, but there is still little consensus on their relative importance. The most prominent theory is that dropping out of high school results in lower human capital. Lower human capital decreases expected future income and, as a result, lowers the cost of engaging in behaviors such as teenage child bearing (Black et al. 2008). The better outcomes could also arise from an incapacitation effect of schooling. The idea is that being in school leaves less time and opportunity for activities that may be detrimental for long-term outcomes (Jacob and Lefgren 2003; Luallen 2006; Black et al. 2008; Anderson 2010). Yet another theory draws on recent studies in neurology and psychology to argue that teenagers are less able to identify the long-run consequences of dropping out. Myopia results in young people undervaluing the cost of dropping out for life-time income (Oreopoulos 2007). It follows that those who drop out have a higher degree of myopia than those who stay in school. In other words, dropouts are inherently different across a set of (possibly unobservable) characteristics.

This paper takes a different approach from the previous literature in understanding the impact of dropping out. We are interested in whether there is something in the act of dropping out that changes conditions for those who leave school. In other words, what is the causal impact of dropping out? Dropping out may, for example, change the make-up of one's social circle and thereby affect behavior. Focus here is on sexual behaviors. Specifically, our outcome of interest is sexually transmitted infection (STI) diagnosis.

A major advantage of using information on STIs instead of other outcomes, such as pregnancies, is estimates are less likely to be biased due to reverse causality. A woman who becomes pregnant during high school has a substantially higher probability of dropping out. Contracting an STI is less likely to make somebody drop out because most STIs are easily treatable, but still provide an indicator that the individual has engaged in risky sex.[1] Beyond the negative effects of contracting an STI, simply engaging in intercourse and having more sexual partners have been shown to have adverse impacts on psychological well-being and academic performance (Sabia 2007; Sabia and Rees 2009). A second advantage of using STIs as an outcome is that diagnosis

---

[1] The exception to this is HIV/AIDS. We therefore present results both with and without HIV/AIDS cases included.

is easily measured and individual-level longitudinal data are available. This allows us to control for unobservable individual characteristics that simultaneously determine dropout status and sexual activity. Lastly, employing STI contraction as an outcome of risky behavior is preferred because males and females both participate in sexual activity at relatively high rates. Variables such as delinquency and crime have the drawback that females are much less likely to engage in these types of behavior (Sweeten et al. 2009).

Using both individual fixed effects estimations and a propensity score matching technique we find strong evidence that dropping out increases the risk of contracting an STI for females. There is some evidence that males who drop out are also more likely to contract an STI, but this result is not robust. A descriptive analysis shows that female dropouts have substantially older romantic partners and are more likely to have been verbally or physically abused by a partner than those who remain in school. These differences in partner attributes suggest a channel through which dropping out may increase the likelihood of contracting an STI.

This paper makes three important contributions to the literature. First, previous studies on the relationship between dropping out and sexual outcomes have failed to empirically address self-selection effects (see, e.g., Brewster et al. 1998; Manlove 1998; Darroch et al. 1999; Manlove et al. 2000). Our fixed effects approach controls for unobserved personal factors that may be simultaneously correlated with dropping out and sexual behavior. Second, to the authors' knowledge, this is the first paper to examine the effect of dropping out on STI contraction. Third, our results are policy relevant and point to a potential benefit of keeping kids in school.

II. Data

The data used in this paper come from the National Longitudinal Study of Adolescent Health (Add Health). The Add Health is a nationally representative sample of adolescents in the United States who were in grades 7 through 12 during the 1994-95 school year. Data collection started with the identification of over 26,000 schools that included an 11th grade and enrolled more than 30 students. From this sample frame, 80 high schools were selected to ensure representation of schools with respect to region of country, urbanicity, size, type, and ethnicity. Participating high schools were asked to identify feeder schools that included a 7th grade and sent at minimum five graduates to that high school. Feeder schools chosen to participate in the study were selected with probability proportional to the number of students it contributed to the high school. After including feeders, the total number of participating schools was 132.

In Wave I, data were collected from adolescents, their parents, siblings, friends, relationship partners, fellow students, and school administrators. The Add Health cohort has been followed with three subsequent in-home surveys in 1996, 2000-2001, and 2007-2008. The data includes information on respondents' social, economic, psychological, and health status. In addition to individual-level information, Add Health also contains contextual data on the family, neighborhood, school, and adolescents' peer networks.[2]

We use data from in-home surveys from Wave I and II because they contain information on self-reported sexual behaviors. The in-home component of Wave I closely followed the initial in-school interview. As a result, few students had dropped out

[2] For a full description of the Add Health data, see Udry (2003)

of high school during that short time period. By Wave II, a substantial number of students reported having dropped out of school. We therefore use Wave II for the cross-sectional analysis and incorporate Wave I when exploiting the panel nature of the data. In addition, respondents were still in their adolescent years during Wave II. Waves III and IV were fielded when respondents were adults.

We limit the sample to youths aged 15 to 19 at the time of Wave I. The lower limit is set at age 15 because some of the variables used in this analysis were constructed from questions that were only asked to respondents who were at least 15 years old. The upper age limit is 19 because individuals older than this are more likely to have been in high school at Wave I for atypical reasons. We also exclude individuals who report having been married at any time during Waves I or II and females who were not going to school because they were pregnant. Lastly, we omit respondents who had missing or inconsistent information on the key outcome variables used in the analysis. The sample sizes for the primary analyses are N = 4554 for males and N = 4470 for females.

Measures of sexual behavior

The measure of sexual behavior analyzed is based on respondent self-reports of having been diagnosed with an STI. The questions used to construct the binary variable for STI status was straightforward. Respondents were asked to identify if they had been told by a doctor or nurse they had any of the following STIs: chlamydia, syphilis, gonorrhea, HIV/AIDS, genital herpes, genital warts, trichomoniasis, and hepatitis B. Females were also asked if they had ever been diagnosed with bacterial vaginosis or nongonococcal vaginitis. A binary indicator was created to equal one for any respondent who had been diagnosed with at least one STI and equal to zero otherwise.

Dropout indicator

The explanatory variable of interest is a binary indicator for whether or not the respondent is a high school dropout. First, the adolescent was asked if he/she was currently attending school.[3] If the respondent was not attending school, the interviewer asked a follow-up multiple-choice question as to why the respondent was not in school. An available choice was the individual had "dropped out."

Table 1 illustrates the means of the dependent variables by dropout status. It is immediately apparent that dropouts are quite different than enrolled students by STI status. The problem with simple means is that dropouts are possibly different from non-dropouts along other dimensions than sexual behavior. These other differences, both observed and unobserved to the researcher, could be the driving factors behind the striking differences shown in Table 1. The goal of this paper is to determine the extent to which these disparities can be attributed to a causal effect of dropping out of high school.

[Table 1 about here.]

Covariates

The Add Health data allows us to control for a rich set of covariates that may be associated with dropout status and sexual behavior. These variables are described in Table 2. To retain sample size, dummy variables were created to indicate which respondents were missing information for specific variables. The explanatory variables

[3] If the respondent was interviewed in the summer, the question was worded in a retrospective manner. In this case, the respondent was able to identify if he/she had dropped out for the entire previous school year or only part of the year.

are grouped into three categories. The first group pertains to standard individual characteristics that describe age, ethnicity, race, and whether the respondent was born in the United States. The second group includes family attributes that are likely to be important for both the dropout decision and the decision to engage in sexual activity. These measures include whether the family has moved between survey waves, whether the respondent is an only child, rates of church attendance, parental education and income indicators, whether the adolescent's biological father was living in the household at the time of the Wave I interview, and a dummy variable indicating whether the mother strongly disapproves of her child engaging in sexual intercourse.

[Table 2 about here.]

The third group includes additional individual-level characteristics that are potentially important for dropout status and sexual behavior. Here, we include the respondent's score on the Add Health Picture and Vocabulary Test to serve as a proxy for cognitive ability. Variables for college aspirations and life expectancy were included because they may reflect levels of future orientation. Lastly, we also consider a measure for whether the respondent goes with his/her "gut feeling" and does not think about consequences when making decisions. This variable proxies for impulsive behavior.

III. Estimation strategy

Ordinary least squares

To model the effect of dropping out, this paper begins by estimating the following equation:

$$STI_i = \alpha + \boldsymbol{X_i\beta_1} + \beta_2 Dropout_i + c_i + \varepsilon_i \qquad (1)$$

where *i* indexes the individual respondent.

In Equation (1), *STI* refers to whether the respondent has been diagnosed with an STI since the date of last interview.[4] ***X*** is a vector of the personal and family characteristics described in Table 2. The variable *Dropout* is a binary indicator that is equal to one if the individual has dropped out of high school and equal to zero otherwise. The coefficient of interest, β_2, measures the effect of dropping out on sexual behavior. The variables c and ε represent unobserved individual effects and an error term, respectively.

Equation (1) is estimated with ordinary least squares (OLS) for ease of interpretation.[5] Standard errors are clustered at the school-level for all estimations. Clustering at the school-level is conservative because it takes into account any dependence of errors within schools. All regressions are weighted by the sample weights provided with the Add Health data.

If dropping out of high school was exogenous, then simple OLS regressions would yield consistent estimates of the influence of dropping out on the probability of STI contraction. Clearly, the assumption of exogeneity is unrealistic. Two possible sources of endogeneity exist when estimating Equation (1). First, there may be unobserved factors in the error term that are correlated with dropout status and sexual behavior. For example, an adolescent that discounts the future heavily may choose not to invest in education while concurrently caring little about the potential consequences of

---

[4] The date of last interview was approximately one year prior to the Wave II interview.

[5] Estimating Equation (1) with a probit model to explicitly account for the dichotomous nature of the dependent variable yielded qualitatively similar results to the OLS estimates. Results are available on request.

sexual behavior. Alternatively, one can imagine how home or school environments might simultaneously increase the likelihood a youth drops out and engages in sexual activity. Estimates of the impact of dropping out on sexual activity that omit these types of factors will be biased upward. Second, structural endogeneity in the form of reverse causality is a concern. That is, students engaging in risky sexual behaviors and incurring the costs associated with their actions may be more likely to leave school than their classmates who abstain from such activities. Again, if this type of endogeneity exists, results will be biased upward.

We take several approaches to control for endogeneity. First, we include in our models a rich set of covariates designed to capture important personal characteristics and to account for home and family environment. We compare models with and without these variables to gauge the extent to which the relationship between dropping out and STI diagnosis is influenced by contextual and background factors. Second, we employ a series of fixed effects models to eliminate time-invariant unobserved heterogeneity. Lastly, we discuss why reverse causality is not likely to be a concern and include robustness checks designed to address this issue.

School fixed effects

To account for school-level characteristics that may be correlated with both dropout status and sexual behavior, we expand Equation (1) to include school fixed effects as follows:

$$STI_{is} = \alpha + \boldsymbol{X_i \beta_1} + \beta_2 Dropout_{is} + c_i + \boldsymbol{\delta_s} + \varepsilon_{is} \qquad (2)$$

where s indexes the school respondent i attended during Wave I. School fixed effects are represented by $\boldsymbol{\delta}$, a vector of school indicators. In addition to accounting for unobserved time-invariant heterogeneity at the school-level, school fixed effects also serve as a proxy for important neighborhood and community characteristics that may bias estimates.

Individual fixed effects

Though school fixed effects absorb important differences, results may still be biased because of family or individual heterogeneity. Adolescents from poor family environments with parents who place relatively little value on their child's education may be more prone to engage in risky behaviors. Moreover, individuals with high rates of time preference, low expectations of positive future outcomes, or lack of motivation for academic pursuits may also be more likely to engage in sexual activities. To account for these sources of bias, we use Add Health data from Waves I and II and employ an individual fixed effects method that amounts to estimating the following first-differenced equation:

$$STI_{i(t+1)} - STI_{it} = \alpha + (\boldsymbol{X_{i(t+1)}} - \boldsymbol{X_{it}})\boldsymbol{\beta_1} + \beta_2(Dropout_{i(t+1)} - Dropout_{it}) + (\varepsilon_{i(t+1)} - \varepsilon_{it}). \quad (3)$$

Equation (3) is the preferred specification because it focuses explicitly on the changes in STI status that occur between Waves I and II, the time period where we observe individuals dropping out. In other words, Equation (3) estimates the causal impact of dropping out by removing individual specific characteristics that are correlated with contracting an STI.

This method does not purge bias due to unobserved time-variant heterogeneity. Because Waves I and II were only one year apart, it is likely that most unobserved personal characteristics that are correlated with STI and dropout status are constant over this period. Despite this, we examine the robustness of our results to the inclusion of time-varying observables. This addresses how sensitive the estimates are to factors that vary between survey periods. Lastly, an identifying assumption of Equation (3) is that the sexual outcomes we study do not cause individuals to drop out of high school. We address this issue of reverse causality in robustness checks below.[6]

Fixed effects propensity score matching

As a final estimation method, the robustness of the individual fixed effects results to a propensity score matching technique is examined.[7] After adolescents are matched, fixed effects estimates are obtained. This exercise essentially amounts to re-estimating Equation (3) on the matched sample, a subset of the general sample.[8] Because the propensity score represents the probability of becoming a dropout between Waves I and II, we exclude individuals from observation who reported being dropouts at Wave I.[9] As a result, one might think of our coefficient of interest as akin to the general difference-in-differences estimator. That is, we compare the change in STI status of an individual before and after he/she drops out to the change in STI status of a continuously enrolled individual, who had a similar probability of dropping out, over the same time span.

[6] An instrumental variables strategy could, in principle, be used to control for reverse causality and unobserved time-varying heterogeneity. This approach was attempted; however, given the Add Health data, identifying a valid set of instruments was difficult. Following other research (Campolieti et al. 2010), we used county and local unemployment rates as instruments, but these variables performed poorly in predicting dropout status in the first stage. We also considered county- and local-level industry composition rates, government expenditures on education, labor force participation rates by age, proportions of the population with high school diplomas, and lagged school-level dropout rates. None of these variables predicted adolescent dropout status sufficiently enough to avoid the problems common to weak instruments (Bound et al. 1995; Staiger and Stock 1997).

[7] See Sabia (2006) and Sabia (2007) for applications of this method to topics on sexual behavior.

[8] Observations "off the common support" are excluded from the analysis.

[9] For our sample, this amounted to dropping 44 individuals from observation.

Propensity score matching involves comparing dropouts and non-dropouts by a single index that captures information from observable pre-dropout characteristics (i.e. Wave I characteristics). The goal is to find a group among our comparison population (i.e. continuously enrolled students) that look as similar as possible to our treatment group (i.e. dropouts). Because there are relatively few treated individuals in the Wave II sample, it is all the more important that care is taken when selecting a control group. Accordingly, we identify the probability of dropping out, i.e. the propensity score, using the following probit model:

$$P(Dropout_{i(t+1)} = 1) = \Phi(\boldsymbol{X_{it}B}), \qquad (4)$$

where $\boldsymbol{X}$ represents a vector of pre-dropout personal characteristics.

Letting $P_{i(t+1)}$ denote the predicted probability of dropping out at time $t + 1$ for adolescent i, we match, by nearest propensity score, those whose dropout status did change between Waves I and II to those who remained enrolled for both periods. More formally, for each dropout i, a non-dropout j is selected such that

$$\lambda > | P_{i(t+1)} - P_{j(t+1)} | = \min_k\{| P_{i(t+1)} - P_{k(t+1)} |\} \qquad (5)$$

where λ is a pre-specified scalar.[10] This matching technique is referred to as the "within-caliper" method. Matching is done with and without replacement.[11]

IV. Results

OLS estimates

---

[10] For our analysis, adolescents are matched where the difference between each treated and untreated individual's propensity score is no greater than .10.

[11] Matching with replacement is arguably the more flexible approach when working with smaller samples as it allows each treated individual to be paired with his/her closest match, without limiting the set of possible matches (Sabia 2006).

Table 3 presents the baseline OLS estimates for Wave II. Each cell represents a separate regression; we only report the coefficient estimate on *Dropout*, our variable of interest. The full results are shown in Table A1 of the Appendix. Models are estimated separately for males and females. To examine the extent to which background characteristics explain the association between dropout and STI status, we examine increasingly rich specifications.

[Table 3 about here.]

Columns (1) and (2) consider models with only the baseline covariates listed in Table 2. Here, there exists a positive and statistically significant relationship between dropout and STI status. Male and female dropouts are 5.2 and 12.2 percentage points more likely to report having been diagnosed with an STI, respectively. However, these estimates may reflect differences between dropouts and non-dropouts rather than a causal relationship due to dropping out.

Columns (3) and (4) add parental and family covariates to the baseline specification. The coefficient estimates remain positive in sign, however, the estimate for males is no longer statistically significant. For females, including measures of family background explain part of the relationship between dropping out and STI contraction; yet, the magnitude of the coefficient estimate remains quite large.

Columns (5) and (6) include additional individual-level covariates likely to be correlated with dropout status. In these specifications we include the Add Health Picture and Vocabulary Test score to proxy cognitive ability, aspirations to attend college, life expectancy, and the degree to which adolescents make decisions based on their "gut feeling." The addition of these covariates decreases the magnitude of the coefficient estimate by slightly over one percentage point for females. The result remains significant at the 5% level and suggests that dropping out is associated with a 9.5 percentage point decrease in the likelihood a female will contract an STI. Again, the result for males is positive in sign, but not statistically significant at a conventional level. In sum, the results in Table 3 suggest robust evidence of a positive relationship between dropping out of high school and the risk of STI diagnosis for females.

School and individual fixed effects

Table 4 presents the fixed effects estimates. Columns (1) and (2) include school fixed effects to control for unobserved heterogeneity at the school- and neighborhood-levels. Estimates are similar to those presented in Columns (5) and (6) of Table 3.

[Table 4 about here.]

Columns (3) and (4) illustrate results based on estimation of Equation (3) and utilize data from Waves I and II of the Add Health in-home survey.[12] Initially, we exclude time-varying controls from the estimating equations. These fixed effects results purge the OLS estimates of any bias due to unobserved time-invariant heterogeneity at the individual-level. Intuitively, we expect the type of unobserved differences that would be important to this relationship to be the sort that would bias our OLS estimates upward.

For females, the coefficient estimate is slightly smaller in magnitude than the OLS results and is statistically significant at the 10% level. For males, the result is no longer large in magnitude and remains insignificant. Because contracting an STI requires

[12] The question in the Wave I interview asked if the respondent had ***ever*** been diagnosed with an STI, while the Wave II question asked if a positive diagnosis had occurred since the date of last interview. Fortunately, any bias due to this slight discrepancy would cause our estimates to understate the truth.

having had sex, we also considered models where virgins were excluded from the analysis. The results changed little under this alternative specification.

The identifying assumption underlying the individual fixed effects procedure is that $E(\varepsilon_{(t+1)} - \varepsilon_t \,|\, Dropout_{(t+1)} - Dropout_t) = 0$. This assumption is violated if time-varying unobservables exist that are correlated with dropout status and the likelihood of STI contraction. As mentioned above, because Waves I and II were conducted only one year apart, it is likely that most unobserved personal characteristics that are correlated with sexual behavior and dropout status are relatively constant over this period. In Columns (5) and (6), a set of time-varying controls are included to assess the sensitivity of the individual fixed effects models. While the estimate for females is slightly larger in magnitude than the estimate shown in Column (4), the results remain very much the same with or without time-varying controls. This gives us confidence that our estimates are not sensitive to unobserved characteristics that change over time.[13]

Robustness of individual fixed effects

In Table 5 we analyze the robustness of the individual fixed effects results. The baseline estimates from Columns (5) and (6) of Table 4 are shown in the first row of Table 5 for reference.

[Table 5 about here.]

[13] We also included time-varying characteristics for peer environment and substance use in separate regressions. The results remained robust to the inclusion of these additional (arguably endogenous) variables. These estimates are available from the authors upon request.

In Row 2 we modify the baseline specification to exclude cases of HIV/AIDS from the definition of having an STI.[14] This is done primarily to address concerns regarding reverse causality. Though evidence suggests that increases in sexual activity can have adverse impacts on academic performance (Sabia 2007; Sabia and Rees 2009), we believe there is little reason to be concerned that being diagnosed with a "less serious" STI would actually lead a youth to drop out of school entirely.[15] Dropping out of high school may be a likely outcome for an individual who has contracted a more serious STI such as HIV. The results in Row 5 take this concern into consideration by redefining "STI" to exclude cases of HIV/AIDS. The result for females is slightly smaller in magnitude than the baseline estimate but remains statistically significant at the 10% level.

In Row 3 we alter the baseline sample by excluding those individuals who had dropped out for only part of the school year as opposed to the entire year. This restriction also addresses some concerns about reverse causality in that it minimizes the likelihood an individual was diagnosed with an STI before dropping out. For females, the result is slightly larger in magnitude than the baseline estimate and is significant at the 5% level.

Next, we combine the restriction made in Row 3 with the constraint that individuals who were out of school for reasons other than having "dropped out" are excluded from the sample.[16] The purpose of this exercise is to compare dropouts with only individuals that have been in school the entire year. In this case, the coefficient

[14] We also considered models where we excluded anyone having been diagnosed with HIV from the sample entirely. The results changed little under this specification.

[15] Of the individuals in Wave II of Add Health that tested positive for an STI, around 65% were diagnosed with an STI that is curable with simple antibiotics. Roughly 20% tested positive for herpes or warts. While not curable, herpes and warts are readily treatable and can be managed quite effectively.

[16] Students could have been out of school for the following other reasons: graduated, suspended/expelled, sick/injured, or on leave.

estimate in the female equation remains large at 7.6 percentage points, but loses statistical significance.

In Row 5 we take a different approach and redefine the term "dropout" to include those individuals that were expelled from school. This helps us to better understand if it is the simple act of being out of school that matters for STI contraction. This specification suggests that being out of school leads to a 9.5 percentage point increase in the likelihood a female is diagnosed with an STI.

In Row 6 of Table 5 we exclude individuals who were dropouts at the time of the in-home Wave I interview in addition to excluding respondents who had dropped out for only part of the school year between survey waves. This restriction ensures that individuals are the same with respect to dropout status at baseline. An identifying assumption of the individual fixed effects model is that those who drop out and those who stay continuously enrolled share common unobserved time trends (Sabia 2007). This assumption may be violated if the effects of dropping out are cumulative. The findings in Row 6 suggest that dropping out leads to a 7 percentage point increase in the probability of STI diagnosis for females. While this point estimate remains relatively large in magnitude, it is not significant at conventional levels.

In the last row of Table 5, we estimate Equation (3) conditional on the respondents having had sex. Because STI contraction requires sexual activity, limiting the sample to those who report having had sex may be more appropriate than considering the sample in its entirety.[17] In making this restriction, the estimate for females remains very similar to the baseline result.

Table 6 investigates the sensitivity of the individual fixed effects results to the propensity score matching procedure described above.[18] An important feature of this technique is that it allows us to consider selection into sexual activity in a flexible manner. When estimating the propensity scores, we included virginity status as an observed characteristic on the right-hand-side of Equation (4). This method is arguably preferred to a basic individual fixed effects model that excludes virgins from observation (i.e. the model estimated in Row 7 of Table 5) because the latter approach does not explicitly model selection into sexual activity. However, it is worth mentioning that results were similar when excluding the indicator for virginity status from Equation (4).

Columns (1) and (2) show results where matching is done without replacement. Here, the coefficient estimate for females is quite large in magnitude though not statistically significant. Columns (3) and (4) consider results where matching is done with replacement. As mentioned above, matching with replacement is arguably the preferred approach when working with samples where the number of treated persons is relatively small as it allows each treated individual to be paired with his/her closest match, without limiting the set of possible matches. In this case, our results indicate that dropping out of high school increases the risk of STI diagnosis for females by 9.5 percentage points.

---

[17] It is important to note, however, that many of the STIs we study can be transmitted through means other than vaginal sex (e.g. oral sex).

[18] The results from the probit specification used to construct the propensity scores are available from the authors upon request.

[Table 6 about here.]

Overall, the robustness results in Tables 5 and 6 confirm that dropping out of high school increases the likelihood a female student will contract an STI. Moreover, the baseline estimate implies that over 70% of the difference in the mean rate of STI status between female dropouts and enrolled students can be explained by dropping out.

In an attempt to place the above estimates in context, Table 7 illustrates results where the following outcome variables are regressed on dropout status: number of reported sexual partners since Wave I, average age of romantic partner, and whether the respondent reports being physically or verbally abused by their romantic partner. The results are generated by simple OLS regressions using Wave II data and control for basic individual demographics. Unfortunately, due to Wave I data limitations, individual fixed effects analyses were not possible for these outcomes.

In Row 1 of Table 7 the estimates indicate that female dropouts are associated with having slightly more than 1 sexual partner than enrolled students. However, Row 2 illustrates this result loses statistical significance when conditioning on virginity status. For males, it appears there is little difference between the number of sexual partners for dropouts versus enrolled students. Row 3 highlights that female dropouts' partners are nearly 1.5 years older than those of female high school students. Even more striking is the difference in whether a female reports having been verbally or physically abused by a romantic partner. The estimate in Row 4 illustrates that female dropouts are over 10 percentage points more likely to report abuse.[19]

Although the results in Table 7 are descriptive in nature, they indicate that female dropouts engage in romantic relationships with significantly different types of partners. Their partners are not only older on average but are also more likely to be abusive. This statistical artifact serves as a possible explanation for the disparity in STI status among female dropouts and high school students. [20] Previous research has documented that females in abusive relationships are less likely than others to use condoms, more fearful of asking their partners to use condoms, and more worried about contracting STIs (Wingood and DiClemente 1997).

[Table 7 about here.]

V. Conclusion

The results presented here indicate there is an impact of dropping out on sexual behavior. The findings are robust to a number of different specifications and the use of different samples. Our preferred specification uses individual fixed effects to eliminate unobservable individual characteristics that are simultaneously correlated with the risk of dropping out and sexual behavior. For females, dropping out of high school leads to a significantly higher risk of contracting a sexually transmitted infection.

This paper is a first step toward understanding how the drop out decision is made and what the impact is at an individual-level. Individual characteristics clearly play a

---

[19] The sample sizes for the equations estimated in Rows 3 and 4 of Table 7 are greatly reduced because the "Relationship Information" module in the Add Health data set was fraught with missing data. In addition, numerous individuals reported not having been in a romantic relationship.

[20] In a related study, Harawa et al. (2003) find that differences in the characteristics of sexual partners does not explain the racial gap in STIs between young white and black women.

substantial role in determining behavior, including dropping out, but our findings highlight that there is also a significant causal effect of dropping out. To the extent that sexual behavior either directly or through other pathways affect future outcomes, this leaves an important role for public policy in inducing those at risk from dropping out to remain in school longer.

Acknowledgements

We would like to thank Shelly Lundberg and David Ribar for helpful comments and suggestions. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgement is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Appendix

[Table A1 about here.]

References

Anderson, D. Mark. 2010. “In School and Out of Trouble? The Minimum Dropout Age and Juvenile Crime.” Working Paper, University of Washington.

Bound, John; Jaeger, David and Baker, Regina. 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association* 90: 443-450.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2008). Staying in the Classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births. *The Economic Journal*, *118*(530), 1025-1054.

Brewster, Karin; Cooksey, Elizabeth; Guilkey, David and Rindfuss, Ronald. 1998. "The Changing Impact of Religion on the Sexual and Contraceptive Behavior of Adolescent Women in the United States." *Journal of Marriage and Family* 60: 493-504.

Campolieti, Michele; Fang, Tony and Gunderson, Morley. 2010. "Labour Market Outcomes and Skill Acquisition of High-School Dropouts." *Journal of Labor Research* 31: 39-52.

Darroch, Jacqueline; Landry, David and Oslak, Selene. 1999. "Age Differences Between Sexual Partners in the United States." *Family Planning Perspectives* 31: 160-167.

DeSimone, Jeffrey. 2010. "Binge Drinking and Sex in High School." NBER Working Paper No. 16132.

Harawa, Nina; Greenland, Sander; Cochran, Susan; Cunningham, William and Visscher, Barbara. "Do Differences in Relationship and Partner Attributes Explain Disparities in Sexually Transmitted Disease Among Young White and Black Women?" *Journal of Adolescent Health* 32: 187-191.

Jacob, Brian A. and Lefgren, Lars. 2003. "Are Idle Hands the Devil's Workshop? Incapacitation, Concentration, and Juvenile Crime." *American Economic Review* 93: 1560-1577.

Lochner, Lance and Moretti, Enrico. 2004. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review* 94: 155-189.

Luallen, Jeremy. 2006. "School's Out...Forever: A Study of Juvenile Crime, At-risk Youths and Teacher Strikes." *Journal of Urban Economics* 59: 75-103.

Manlove, Jennifer. 1998. "The Influence of High School Dropout and School Disengagement on the Risk of School-Age Pregnancy." *Journal of Research on Adolescence* 8: 187-220.

Manlove, Jennifer; Terry, Elizabeth; Gitelson, Laura; Papillo, Angela and Russel, Stephen. 2000. "Explaining Demographic Trends in Teenage Fertility, 1980-1995." *Family Planning Perspectives* 32: 166-175.

Oreopoulos, Philip. 2007. "Do Dropouts Drop Out Too Soon? Wealth, Health and Happiness from Compulsory Schooling. *Journal of Public Economics* 91: 2213-2229.

Renna, Francesco. 2008. "Teens' Alcohol Consumption and Schooling." *Economics of Education Review* 27: 69-78.

Sabia, Joseph. 2006. "Does Sex Education Affect Adolescent Sexual Behaviors and Health?" *Journal of Policy Analysis and Management* 25: 783-802.

Sabia, Joseph. 2007. "Reading, Writing, and Sex: The Effect of Losing Virginity on Academic Performance." *Economic Inquiry* 45: 647-670.

Sabia, Joseph. 2010. "Wastin' Away in Margaritaville? New Evidence on the Academic Effects of Teenage Bing Drinking." *Contemporary Economic Policy* 28: 1-22.

Sabia, Joseph and Rees, Daniel. 2009. "Does the Number of Sex Partners Affect School Attainment?" Forthcoming *Journal of Population Economics*.

Staiger, Douglas and Stock, James. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65: 557-586.

Sweeten, Gary; Bushway, Shawn and Paternoster, Raymond. 2009. "Does Dropping Out of School Mean Dropping into Delinquency?" *Criminology* 47: 47-91.

Udry, J. Richard. 2003. "The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994-1996; Wave III, 2001-2002 [machine-readable data file and documentation]. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.

Wingood, Gina and DiClemente, Ralph. 1997. "The Effects of an Abusive Primary Partner on the Condom Use and Sexual Negotiation Practices of African-American Women." *American Journal of Public Health* 87: 1016-1018.

Table 1: Means of dependent variable by dropout status

	High school students		High school dropouts	
	Males	Females	Males	Females
Diagnosed with STI	0.021	0.040	0.077	0.166
	(0.003)	(0.005)	(0.026)	(0.045)
	[4373]	[4347]	[181]	[123]

Notes: (1) Means are based on data from Wave II of the National Longitudinal Study of Adolescent Health. (2) Standard errors of the means are reported in parentheses. (3) Sample sizes are in brackets. (4) The STI variable refers to having been diagnosed with an STI since the Wave I interview.

Table 2: Descriptive statistics

Variable	Description	Males Mean	Males Std. Dev.	Females Mean	Females Std. Dev.
Dependent variable					
STI	Equal to 1 if respondent has been diagnosed with an STI, 0 otherwise.	0.023	0.151	0.045	0.208
Baseline respondent characteristics					
Dropout	Equal to 1 if respondent has dropped out of high school, 0 otherwise.	0.047	0.212	0.042	0.200
Age	Respondent's age.	17.55	1.078	17.43	1.005
Hispanic	Equal to 1 if respondent is Hispanic, 0 otherwise.	0.128	0.334	0.115	0.319
Black	Equal to 1 if respondent is black, 0 otherwise.	0.166	0.372	0.173	0.378
White	Equal to 1 if respondent is white, 0 otherwise.	0.730	0.444		
Other non-white	Equal to 1 if respondent is other non-white race, 0 otherwise.	0.146	0.354	0.138	0.345
U.S. born	Equal to 1 if respondent was born in the United States, 0 otherwise.	0.938	0.241	0.922	0.268
Family characteristics					
Moved	Equal to 1 if family moved between Waves I and II, 0 otherwise.	0.061	0.239	0.062	0.242
Only child	Equal to 1 if respondent is the only child in the family, 0 otherwise.	0.193	0.395	0.196	0.397
Only child missing	Equal to 1 if only child information is missing, 0 otherwise.	0.006	0.075	0.004	0.064
Church 1	Equal to 1 if did not attend church in last year, 0 otherwise.	0.279	0.448		
Church 2	Equal to 1 if went to church less than once per month, 0 otherwise.	0.188	0.391	0.191	0.393
Church 3	Equal to 1 if went to church at least once per month, 0 otherwise.	0.511	0.500	0.567	0.495
Church missing	Equal to 1 if church information is missing, 0 otherwise.	0.022	0.148	0.015	0.122
Mother's education 1	Equal to 1 if mother has less than high school degree, 0 otherwise.	0.144	0.351		
Mother's education 2	Equal to 1 if mother has high school degree or GED, 0 otherwise.	0.329	0.470	0.329	0.470
Mother's education 3	Equal to 1 if mother has more schooling than a high school degree, 0 otherwise.	0.408	0.491	0.411	0.492
Mother's education missing	Equal to 1 if mother's education is missing, 0 otherwise.	0.119	0.324	0.085	0.279
Father's education 1	Equal to 1 if father has less than high school degree, 0 otherwise.	0.109	0.311		
Father's education 2	Equal to 1 if father has high school degree or GED, 0 otherwise.	0.226	0.418	0.224	0.417
Father's education 3	Equal to 1 if father has more schooling than a high school degree, 0 otherwise.	0.352	0.478	0.330	0.470
Father's education missing	Equal to 1 if father's education is missing, 0 otherwise.	0.313	0.464	0.332	0.471
Bio father present	Equal to 1 if biological father was present during Wave I, 0 otherwise.	0.522	0.500	0.512	0.500
Bio father present missing	Equal to 1 if information on biological father's presence is missing, 0 otherwise.	0.120	0.325	0.127	0.333

Table 2 (continued): Descriptive statistics

Variable	Description	Males Mean	Males Std. Dev.	Females Mean	Females Std. Dev.
Parental income 1	Equal to 1 if total household income is less than 40k, 0 otherwise.	0.388	0.487		
Parental income 2	Equal to 1 if total household income is between 40k and 80k, 0 otherwise.	0.298	0.457	0.290	0.454
Parental income 3	Equal to 1 if total household income is greater than 80k, 0 otherwise.	0.093	0.290	0.100	0.300
Parental income missing	Equal to 1 if total household income information is missing, 0 otherwise.	0.221	0.415	0.225	0.417
Mother's disapproval	Equal to 1 if child thinks mother would strongly disapprove of him/her having sex, 0 otherwise	0.370	0.483	0.540	0.498
Mother's disapproval missing	Equal to 1 if information is missing about mother's disapproval of child having sex, 0 otherwise.	0.072	0.258	0.056	0.229
Additional respondent characteristics					
PVT score	Add Health Picture and Vocabulary Test score.	97.16	25.28	96.33	24.21
PVT score missing	Equal to 1 if Picture and Vocabulary Test score is missing, 0 otherwise.	0.045	0.207	0.039	0.194
College aspirations	Scale of 1 to 5 measure of how much respondent wants to go to college. 1=low aspirations. 5=high aspirations.	4.214	1.441	4.130	1.446
College aspirations missing	Equal to 1 if college aspirations information is missing, 0 otherwise.	0.036	0.186	0.053	0.224
Life expectancy	Scale of 1 to 5 measure reflecting odds respondent thinks he/she will live to be 35. 1=almost no chance. 5=almost certain.	4.214	0.950	4.323	0.891
Life expectancy missing	Equal to 1 if life expectancy information is missing, 0 otherwise.	0.003	0.053	0.004	0.060
Gut feeling	Scale of 1 to 5 measure reflecting if respondent, when making decisions, goes with his/her "gut feeling" without thinking about consequences. 1=Strongly agrees he/she goes with "gut feeling." 5=Strongly disagrees he/she goes with "gut feeling."	2.951	1.155	3.176	1.153
Gut feeling missing	Equal to 1 if "gut feeling" information is missing, 0 otherwise.	0.003	0.055	0.003	0.056

Note: All variables are from Wave II of the National Longitudinal Study of Adolescent Health with the exception of PVT score and the dummies for parental education, parental income, presence of the adolescent's biological father, and whether the respondent is an only child. These variables are from the Wave I survey.

Table 3: Dropping out and sexually transmitted infections (OLS estimates)

	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Dropout	0.052**	0.122***	0.038	0.107**	0.036	0.095**
	(0.026)	(0.044)	(0.025)	(0.044)	(0.025)	(0.045)
	[4554]	[4470]	[4554]	[4470]	[4554]	[4470]
1.) Baseline covariates	YES	YES	YES	YES	YES	YES
2.) Family covariates	NO	NO	YES	YES	YES	YES
3.) Additional individual-level covariates	NO	NO	NO	NO	YES	YES

Notes: (1) Each cell represents a separate regression. (2) Standard errors are in parentheses and are clustered at the school-level. (3) Sample sizes are in brackets. (4) Covariates included in the models are described in Table 2. (5) *, significant at 10% level; **, significant at 5% level; ***, significant at 1% level.

Table 4: Dropping out and sexually transmitted infections (Fixed effects results)

	School FE		Individual FE w/o time varying controls		Individual FE w/ time varying controls	
	(1)	(2)	(3)	(4)	(5)	(6)
	Males	Females	Males	Females	Males	Females
Dropout	0.033	0.102**	0.002	0.087*	0.001	0.091*
	(0.024)	(0.046)	(0.036)	(0.048)	(0.036)	(0.048)
	[4554]	[4470]	[9108]	[8940]	[9108]	[8940]

Notes: (1) Each cell represents a separate regression. (2) Standard errors are in parentheses and are clustered at the school-level. (3) Sample sizes are in brackets. (4) The school FE models include the same covariates as the OLS models in the last two columns of Table 3. The individual FE models in Columns 5 and 6 include a set of time-varying controls. The time-varying controls include church attendance, whether the respondent moved, sentiment of mother towards sexual behavior, college aspirations, life expectancy, and an indicator for making decisions based on "gut feeling." (5) *, significant at 10% level; **, significant at 5% level; ***, significant at 1% level.

Table 5: Dropping out and sexually transmitted infections (Robustness of individual FE)

	(1) Males	(2) Females
1.) Baseline estimates.	0.001 (0.036) [9108]	0.091* (0.048) [8940]
2.) Define "STI" to exclude cases of HIV/AIDS.	0.002 (0.036) [9108]	0.074* (0.043) [8940]
3.) Exclude those that dropped out for only part of the year.	0.004 (0.038) [9036]	0.105** (0.053) [8882]
4.) Restriction 3.) and exclude those who were out of school for other reasons.	0.017 (0.044) [8634]	0.076 (0.056) [8596]
5.) Restriction 3.) and define "dropout" as having dropped out or been expelled.	0.034 (0.029) [9036]	0.095** (0.046) [8882]
6.) Restriction 3.) and exclude those who were dropouts at wave I.	0.021 (0.040) [8970]	0.070 (0.057) [8820]
7.) Conditional on having had sex	-0.010 (0.047) [3692]	0.087* (0.053) [3492]

Notes: (1) Each cell represents a separate regression. (2) Standard errors are in parentheses and are clustered at the school-level. (3) Sample sizes are in brackets. (4) All models include a full set of time-varying controls. The time-varying controls include church attendance, whether the respondent moved, sentiment of mother towards sexual behavior, college aspirations, life expectancy, and an indicator for making decisions based on "gut feeling." (5) *, significant at 10% level; **, significant at 5% level; ***, significant at 1% level.

Table 6: Dropping out and sexually transmitted infections (Matching analysis)

	Matching without replacement		Matching with replacement	
	(1)	(2)	(3)	(4)
	Males	Females	Males	Females
Dropout	-0.017	0.120	0.003	0.095*
	(0.048)	(0.085)	(0.037)	(0.049)
	[714]	[476]	[6824]	[7588]

Notes: (1) Each column represents a separate regression. (2) Standard errors are in parentheses and are clustered at the school-level. (3) Sample sizes are in brackets. (4) All models include a full set of time-varying controls. The time-varying controls include church attendance, whether the respondent moved, sentiment of mother towards sexual behavior, college aspirations, life expectancy, and an indicator for making decisions based on "gut feeling." (5) *, significant at 10% level; **, significant at 5% level; ***, significant at 1% level.

Table 7: Differences between dropouts and non-dropouts (OLS results)

	(1) Males	(2) Females
Number of sexual partners since Wave I interview.	0.223 (0.515) [4557]	1.081*** (0.348) [4574]
Number of sexual partners since Wave I interview (Conditional on having had sex).	0.316 (0.508) [1814]	0.681 (0.457) [1786]
Average age of romantic partner.	-0.085 (0.153) [3035]	1.472*** (0.316) [3357]
Physically or verbally abused by romantic partner.	0.066 (0.053) [3062]	0.102** (0.044) [3415]

Notes: (1) Results are based on data from Wave II of the National Longitudinal Study of Adolescent Health. (2) Each cell represents a separate regression. (3) Standard errors are in parentheses and are clustered at the school-level. (4) Sample sizes are in brackets. (5) All models control for age, race, and whether the respondent was born in the United States. (6) *, significant at 10% level; **, significant at 5% level; ***, significant at 1% level.

Table A1: Dropping out and sexually transmitted infections (Full OLS estimates)

	(1) Males	(2) Females	(3) Males	(4) Females	(5) Males	(6) Females
Dropout	0.052**	0.122***	0.038	0.107**	0.036	0.095**
	(0.026)	(0.044)	(0.025)	(0.044)	(0.025)	(0.045)
Age 16	-0.001	-0.013	-0.005	-0.012	-0.004	-0.014
	(0.011)	(0.019)	(0.012)	(0.019)	(0.012)	(0.019)
Age 17	0.006	-0.002	0.003	-0.003	0.005	-0.005
	(0.012)	(0.020)	(0.013)	(0.021)	(0.013)	(0.021)
Age 18	-0.006	0.016	-0.007	0.014	-0.004	0.015
	(0.011)	(0.019)	(0.012)	(0.020)	(0.013)	(0.020)
Age 19	0.028	0.024	0.023	0.017	0.028	0.023
	(0.020)	(0.025)	(0.019)	(0.026)	(0.019)	(0.027)
Hispanic	0.013	-0.006	0.003	-0.010	0.001	-0.015
	(0011)	(0.016)	(0.012)	(0.015)	(0.013)	(0.016)
Black	0.036***	0.049***	0.032**	0.046***	0.024*	0.041***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)	(0.013)
Other non-white	0.004	0.019	0.005	0.021	0.004	0.023
	(0.009)	(0.018)	(0.008)	(0.018)	(0.008)	(0.017)
U.S. born	0.004	0.026**	0.002	0.026*	0.008	0.026*
	(0.014)	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)
Moved	---	---	-0.010	-0.003	-0.009	-0.008
			(0.011)	(0.021)	(0.011)	(0.021)
Only child	---	---	0.009	0.026**	0.012	0.027**
			(0.008)	(0.011)	(0.008)	(0.011)
Church 2	---	---	-0.003	0.018	-0.002	0.020*
			(0.012)	(0.012)	(0.012)	(0.012)
Church 3	---	---	-0.013	-0.002	-0.012	0.000
			(0.010)	(0.009)	(0.010)	(0.009)
Mother’s education 2	---	---	-0.030*	-0.011	-0.027*	-0.009
			(0.017)	(0.015)	(0.016)	(0.015)

Table A1 (continued): Dropping out and sexually transmitted infections (Full OLS estimates)

Mother's education 3	---	---	-0.046***	-0.013	-0.040**	-0.007
			(0.017)	(0.014)	(0.016)	(0.014)
Father's education 2	---	---	0.008	0.002	0.010	0.003
			(0.012)	(0.017)	(0.012)	(0.017)
Father's education 3	---	---	0.009	-0.008	0.013	-0.004
			(0.014)	(0.015)	(0.014)	(0.016)
Bio father present	---	---	-0.003	0.012	-0.002	0.013
			(0.008)	(0.014)	(0.008)	(0.014)
Parental income 2	---	---	-0.003	-0.001	-0.002	0.000
			(0.006)	(0.010)	(0.006)	(0.010)
Parental income 3	---	---	0.008	-0.008	0.011	-0.004
			(0.009)	(0.014)	(0.008)	(0.014)
Mother's disapproval	---	---	0.006	-0.023**	0.007	-0.020**
			(0.008)	(0.010)	(0.008)	(0.010)
PVT score/10	---	---	---	---	-0.005**	0.000
					(0.002)	(0.004)
College aspirations	---	---	---	---	0.001	-0.011
					(0.003)	(0.007)
Life expectancy	---	---	---	---	-0.008**	-0.010
					(0.004)	(0.006)
Gut feeling	---	---	---	---	-0.008***	-0.004
					(0.003)	(0.004)
N	4554	4470	4554	4470	4554	4470

Notes: (1) Each column represents a separate regression. (2) Standard errors are in parentheses and are clustered at the school-level. (3) Dummy variables for missing observations are included, but not reported. (4) *, significant at 10% level; **, significant at 5% level; ***, significant at 1% level.