Popul Res Policy Rev (2016) 35:757–789
DOI 10.1007/s11113-016-9410-4

Pregnancy and Dropout: Effects of Family, Neighborhood, and High School Characteristics on Girls' Fertility and Dropout Status

Nathan Berg[1,2] · **Teresa D. Nelson**[3]

Received: 3 December 2015 / Accepted: 5 August 2016 / Published online: 31 August 2016

Abstract Administrative data from multiple sources are combined to measure pregnancy (excluding those ending in abortion or miscarriage) and high school dropout in a cohort of girls who were 9th graders in the 1994–1995 academic year. Rates of pregnancy (as identified in the data) and dropout are substantially higher among Hispanic high school students than among African-Americans or non-Hispanic whites. Previous studies of teen pregnancy and dropout typically focus on pregnancy rates conditional on dropout status, or dropout rates conditional on fertility. This paper presents estimates of pregnancy and dropout as a joint-dependent variable. Estimates of their joint probability distribution conditional on individual, family, neighborhood, and high school characteristics are reported. The estimates use longitudinal administrative data collected as annual censuses of all public school students in Texas with individual-level ids. Neighborhood characteristics (from the US Census data geographically linked to Texas high schools) have large effects on pregnancy and dropout. Immigrant Hispanic girls' pregnancy rates are significantly lower than native-born Hispanic girls' pregnancy rates. Above-normal-age status in the 9th grade is among the strongest predictors of pregnancy and dropout in later years. Ethnic differences in age distributions within grade level appear to explain a large share of ethnic differences in pregnancy and dropout rates.

✉ Nathan Berg
Nathan.Berg@otago.ac.nz

Teresa D. Nelson
dr.tdnelson@gmail.com

[1] Economics Department, University of Otago, Dunedin, New Zealand

[2] University of Newcastle, Callaghan, Australia

[3] Carrollton, USA

Keywords Adolescent motherhood · Teenage childbearing · High school completion · Latina fertility · Neighborhood · School and peer effects

JEL Classification I19 · I21 · Z1

Introduction

As recently as the 1970s, US schools typically required pregnant girls to leave school and become dropouts (Upchurch and McCarthy 1990). Those policies led to statistical associations between pregnancy and dropout that no longer necessarily hold today. The policy environment in recent decades has shifted to encouraging pregnant teens to stay in school and graduate from high school. The mix of federal, state, district, and school policies aimed at reducing pregnancy and dropout rates have contributed to more heterogeneous statistical associations between pregnancy and dropout in recent decades, with large differences across individual, family, neighborhood, and school-specific characteristics.

With the understanding that abortions and miscarriages are not identified as pregnancies in our data, the first goal of this paper is to document the heterogeneous rates of identified pregnancy and dropout by age, ethnicity, and immigrant status, using a large administrative database that includes records for all public school students in the State of Texas during the years 1994–2000. In addition to descriptive bivariate contrasts in mean rates of pregnancy and dropout, the goal of providing precise descriptive evidence about the characteristics that influence the likelihood of joint pregnancy and dropout outcomes requires estimation of conditional probabilities. The resulting empirical conditional probabilities provide a descriptive mapping from observable information about variation in individual and family characteristics, high school characteristics, and neighborhood characteristics, to each cell in the joint distribution of pregnancy and dropout.

Whereas much of the relevant literature on pregnancy and dropout focuses on fertility conditional on high school completion or dropout conditional on fertility (see Ribar 1994, for a discussion of econometric issues raised by these two, contrasting methodological approaches), this paper follows di Tommaso and Weeks (2000) in treating pregnancy and dropout as a jointly determined bivariate random variable consisting of four mutually exclusive combinations of two binary events: Pregnant & Dropped, Pregnant & Not Dropped, Not Pregnant & Dropped, and Not Pregnant & Not Dropped. The statistical analysis that follows provides estimates of the marginal effects of differences in family, school, and neighborhood characteristics on these four joint outcomes that classify each girl's eventual outcome, estimated among a cohort of girls that began as 9th graders in the same school year. The data follow each girl's academic progress over five academic years (1994–1995 through 1999–2000 academic years) to track fertility before graduation and school completion (allowing 5 years to record graduation outcomes for girls who repeat up to 2 years or leave and return). A key aspect of the design of the cohort analysis—including all girls enrolled in the 9th grade in the State of Texas in a particular year—is the grade level rather than age-level cohort analysis. Variation in age in the

9th grade by ethnicity turns out to be a key driver of ethnic differences in pregnancy rates in subsequent years.[1]

Perhaps unsurprisingly given the vast body of previous research into ethnic gaps in pregnancy and dropout, the data reveal ethnicity to be an important characteristic that conditions likelihoods of pregnancy and dropout. Hispanic girls become pregnant while in high school significantly more frequently than girls in any other ethnic group.[2] Therefore, a second goal of this paper is to quantify how much of Hispanic girls' above-average rate of pregnancy can be accounted for by differences in age, family characteristics, ethnic composition of high schools, and observable characteristics of neighborhoods where students reside. The third goal of this paper is to use the conditional probability functions we estimate (mapping right-hand-side variables into the joint distribution of pregnancy and dropout probabilities) to assess the relative importance of various channels through which policy could effectively influence rates of pregnancy and dropout. One concern is, of course, that girls may leave school early in their pregnancy and report another reason for their dropping out. These girls enter the "not pregnant, dropped out" coding of the dependent variable. If this pattern of measurement error is correlated with any of the variables included in the models, then that correlation will contribute to misspecified associations. Therefore, our empirical modeling and interpretations thereof give special consideration to the methodological challenges of estimating marginal effects.

The paper proceeds as follows. "Related Literature" section attempts to contextualize this paper's main goals in relation to the voluminous literatures on teen pregnancy, high school dropout, and policy approaches to these issues. "Data" section describes data sources, and "Descriptive Statistics" section provides descriptive statistics, unconditional contrasts, and details on the construction of the cohort used in subsequent estimates of the conditional joint distribution of pregnancy and dropout. "Results" section presents the main results, which consist of estimated conditional probabilities for each cell in the joint distribution of pregnancy and dropout. Finally, "Discussion" section undertakes a prospective interpretation of the relative magnitudes of coefficients corresponding to different channels through which new policies intended to reduce pregnancy and dropout rates could make a meaningful difference.

[1] Selection into above-age status is difficult to control for given our focus on estimating conditional effects on the joint outcomes of pregnancy and dropout. Rather than explaining all pathways into being old for one's grade level (which is beyond the scope of this paper), the results we report should be interpreted as follows. Being old for one's grade level is a risk factor rather than causal with respect to pregnancy and dropout, whose measured effects are subject to the caveat of unmeasured selection.

[2] *Hispanic*, as opposed to *Latino*, is the ethnic label used by public schools in Texas (through 2009) when collecting ethnic information from students. Unlike recent waves of the US Census, ethnicity in Texas public schools was, until 2009, measured as a forced choice survey item that did not allow for multiple categories. Beginning in 2009, the Texas Education Agency changed the coding of ethnicity and race to conform with new requirements from the US Department of Education (issued in 2007) to adopt ethnic/racial taxonomies at the state level to match that of the 2000 US Census. This paper refers to individuals coded as African-American as *black*, and those coded in the pre-2009 data as Caucasian as *white*.

Related Literature

The literature provides detailed empirical characterizations of the probability of dropout conditional on pregnancy (Moore and Waite 1977; Upchurch and McCarthy 1990; Astone and Upchurch 1994) and of pregnancy conditional on being a high school dropout (Crane 1991; Manlove 1998; Upchurch and McCarthy 1990; Hill and Johnson 2004). Motivated by concern that these conditional probability models might suffer from endogeneity bias of unknown sign (Ribar 1994), di Tommaso and Weeks (2000) put forward a practical approach for modeling two jointly determined binary outcomes. In their study of labor market participation and fertility, di Tommaso and Weeks (2000) address an endogeneity problem very similar to ours by modeling those two binary outcomes as a single integer-valued dependent variable coding the four possible combinations. This paper adopts di Tommaso and Weeks' (2000) approach using an integer-valued dependent variable with a range of 1–4, whose associated probabilities follow the multinomial logit specification. This technique for empirically modeling the conditional relationship between a large set of right-hand-side variables and the four cells of the joint distribution of pregnancy and dropout is intended to complement previous work that focused on one conditional probability generated by these high-stakes life events.

The effects of different policy interventions such as abstinence messages and efforts to boost school attendance are studied extensively, pointing toward the general finding that no silver bullet among institutional experiments attempted so far has succeeded in reliably and reproducibly reducing high school pregnancy and dropout (Oettinger 1999; Sawhill 2001; Tanner et al. 2009). Some studies suggest that parenting styles and family structure (single-parent families, adoptive parents, or a having a stepparent) and individual characteristics may rival the importance of incentives, information, specialized services received by students at school and social capital linked to neighborhoods, as tools for influencing rates of pregnancy and dropout (O'Connor 1998; Moore et al. 1998; Upchurch et al. 1999). Interventions using tools ranging from contraception to family discussion lead to substantively different results depending on demographic factors (Santelli et al. 2007; Brewster et al. 1998) that also play a primary role in our empirical models.

Subtle and perhaps counterintuitive effects concerning socioeconomic and immigrant status are another result from the pregnancy and dropout literatures that are useful to keep in mind when trying to interpret the statistical models presented subsequently. For example, despite the clear finding that the poorest neighborhoods produce the highest chances of teen pregnancy (Crane 1991), important non-linearities with respect to income are present in many datasets. For example, adolescents residing in working-class neighborhoods appear to take fewer risks than those in higher income neighborhoods (Upchurch et al. 1999). In a study of Mexican teens in Los Angeles, Afable-Munsuz and Brindis (2006, p. 208) find that immigrant teenagers "were less likely to initiate sex than their US-born… counterparts, but those who initiated sex were more likely to get pregnant and to give birth."

Pregnant Hispanic teens are less likely than black or white girls to seek abortions (Guttmacher Institute 2002). There is an unresolved puzzle, however, in that

Hispanics have higher rates of pregnancy than blacks (in our data and in other sources) but lower rates of sexual intercourse (Santelli et al. 2007). Hispanic girls also tend to be significantly older than blacks (though at ages similar as whites) at the time of first sexual intercourse (Warren et al. 1998).[3] Rates of contraceptive use among teenage Hispanics are lower than for whites and blacks (Manlove and Franzetta 2004). Investigations of teen pregnancy reveal links to cognitive development (Crane 1991; Manlove and Franzetta 2004; Moore et al. 1998), culture (Ogbu 1992; Upchurch et al. 1999), academic ability, having been retained in grade (Moore et al. 1998), and religiosity (Regnerus 2007).

Lofstrom (2007) reports dropout rates of 33 % among Texas public school students scheduled to graduate in 1995 and 38 % among those scheduled to graduate in 1999. By ethnicity, these rates are roughly 40 % among Hispanics, 39 % among blacks, and 26 % among whites. Studies of high school dropout identify a variety of factors as potential causes: disengagement from school (Rumberger and Larson 1998; Rumberger 2001); grade retention (Jimerson et al. 2002; Jimerson 2001); poor academic performance (Barro and Andrew 1987); economic disadvantage (Wilson 1987); family structure or dynamics (Manlove and Franzetta 2004; Rumberger 2001); and school policies (Rumberger 2001). Some scholars consider retention a more important determinant of dropout than family interactions or family structure (Jimerson et al. 2002; Jimerson 2001) or socioeconomic level and gender (Rumberger 1995). Portes (1998) and Rumberger (2001) present evidence showing that lower levels of human and social capital among parents elevate the chance of dropout. Large statistical associations indicate that families and neighborhoods can powerfully influence pregnancy and dropout outcomes (Rumberger 1995; Alexander et al. 1997; Jimerson et al. 2002). These studies together motivate our focus on ethnicity, age, socioeconomic status, immigrant status, ethnic and immigrant composition of high schools, and neighborhood characteristics.

Data

Cohort Definition Based on Grade Level

Our study focuses on a cohort of females defined by having attended ninth grade in the 1994–1995 academic year. This cohort definition is based on grade level rather than age. The cohort therefore includes substantial heterogeneity with respect to girls' ages. Age heterogeneity turns out to be an important driver of ethnic differences in rates of pregnancy and dropout with this cohort.

[3] Abortion rates in Texas in 1996 are lower among Hispanics than among blacks (Henshaw and Feivelson 2000).

Data Sources

Administrative data are provided by the Texas Education Agency (TEA) and housed at the University of Texas-Dallas (UTD) Texas Schools Project (TSP).[4] With individual identifiers for all public school students in the state, and a rich set of demographic and academic information (including statewide standardized test scores in reading and mathematics), these data enable us to follow students over multiple years with various data files that include more than 15 million individual records. The cohort data record attendance, standardized test scores, graduation, and proxies for pregnancy for 6 years from the beginning of the 9th grade. Graduation can therefore be observed in the expected year of graduation or 1 or 2 years late. The unit of analysis is the individual. Associated with each individual's school records are school-wide demographics and neighborhood demographics based on the US Census data.

TSP data provide a wealth of information showing which courses each student takes in high school, dropout records (contained in what is sometimes referred to as *Leaver* files), graduation records, standardized test scores from the Texas Assessment of Academic Skills (TAAS) tests, and Career and Technology files indicating participation in specialized courses for pregnant teens, parenting teens, or daycare service provided by schools. This information is spread out across multiple files. The data reported below draw on no fewer than ten different files observed over 5 years. Enrollment files provide information on race/ethnicity, gender, date of birth, low-socioeconomic status (i.e., qualifying for reduced price or free lunch programs), and participation in special education, gifted, bilingual, and English as a Second Language (ESL) programs.

TEA provides multiple source files to TSP for measuring pregnancy, all of them requiring some auxiliary assumptions for coding pregnancy outcomes. Sources and definitions are described below. Ethnic taxonomies include the designations *American Indians/Alaskan Natives* and *Asian/Pacific Islanders.* Individuals coded in these ethnic categories were excluded from the analysis because their numbers in most schools were insufficient to generate sufficient variation in exposure to own-ethnic classmates, which is one of the key variables in the empirical models.

Dropout information comes from three TEA files with overlapping, but not always consistent, records. The so-called *Drop* files list encrypted student identification numbers and years in which students dropped out. Separate *Dropout Reason* files indicate reasons for dropping out. And *Leaver* files list all students who leave the campus, which includes not only dropouts, but also attrition to private or out-of-state schools.[5]

After deleting a small number of individuals with duplicate personal identifiers in the same year, the cohort consists of 145,175 females enrolled in ninth grade during the 1994–1995 academic year. The US Census data provide information about

[4] The TSP is one of the State of Texas' officially designated "State of Texas Education Research Centers." Further detail can be found at http://www.utdallas.edu/research/tsp-erc/.

[5] The TEA replaced *Drop* and *Dropout Reason* files with *Leaver* files beginning in the 1999–2000 school year, which is the temporal midpoint of our 6-year horizon for recording pregnancies and dropouts.

neighborhoods in which the schools are located (Jargowsky and Bane 1991; Jargowsky 1997). Tract-level year-2000 Census data include percent Hispanic, percent black, percent Spanish-speaking, percent immigrant, median household income, and the percentage of homes with grandparents living in them.[6]

Measuring Pregnancy

A number of courses specifically designed for teen parents and pregnant girls are offered in Texas public schools. Some of these courses take place on school campuses that are exclusively for pregnant girls. In cases where the names of campuses make explicit reference to pregnant teens or teen parents, attendance at these campuses is used to indicate a girl is pregnant according to our coding scheme. In addition to specialized campus names, the coding scheme uses TSP data files showing coursework taken, dropout reasons (listing pregnancy as one of the reasons), and Career and Technology course attendance records (in courses for pregnant teens, parenting teens, or teens using on-campus daycare services) to generate an "identified pregnancy" indicator for each female identified as ever having been pregnant or a parent. The event of identified pregnancy in our data excludes pregnancies that end in abortion or miscarriage. For simplicity, we refer to an identified pregnancy as the event of "pregnancy" with the understanding that abortions and miscarriages are not identified as pregnancies in our data.

Unfortunately, the TSP data do not allow us to identify pregnancies that occur after a student dropped out.[7] To summarize, four criteria are used to identify status as pregnant while in high school: (1) attending a school designated as providing education exclusively to pregnant females or parenting students; (2) participating in a Career and Technology program for students identified as single-parents, pregnant teens, or utilizing daycare services; (3) taking a course (outside the Career and Technology program) that identifies the student as pregnant or parenting; or (4) pregnancy as a reason for dropping out of school.[8]

[6] This is an imputed measure estimated by the US Census Bureau of the number of grandparents (in households with residents age 30 or older) living with grandchildren less than 18 years old. Using these counts, we calculated the percentage within tracts of grandparents, referred to as the variable *Pct grandparents in neighborhood.*

[7] Within the pregnant group, there are likely important differences between those who drop out after entering schools for pregnant girls or special classes for young parents compared to those who drop out without attending special classes. Similarly, problematic is the subset of students who were pregnant but may have never enrolled in a special school or special program for pregnant students or teenage moms, and hence were not recorded as pregnant. There are few options for dealing with such concerns given the information available in the data.

[8] Our pregnancy rates consequently differ from those reported by TEA in two ways. TEA relies solely on a variable in Reason for Leaving (or Leaver) files when reporting pregnancy statistics even though information about additional pregnancies is available. Therefore, the pregnancy rates reported by TEA describe only the fraction of girls who drop out and list pregnancy as the reason for their dropout. Our indicator variable *pregnant* is constructed as follows. First, our Stata code scans each year from 2000–2001 back to 1989–1990 and codes the student as pregnant ("ever identified as pregnant" would be more descriptive of this definition) if she is listed as having left school because of pregnancy in one or more years from 1989 through 2001. Another way that individuals in our sample get coded as pregnant = 1 is if their reason for leaving school in any particular year was marriage, under the

Measurement Error in the Pregnancy Variable

Girls may leave school early in their pregnancy and report another reason for dropping out. In our taxonomy of outcomes, these girls enter the category of "not pregnant, dropped out." To the extent that misclassification (i.e., truly pregnant girls who report a reason other than pregnancy for their dropout) covaries with any of the variables included in the empirical models, then, of course, the statistical associations we report will be misspecified with respect to the true (i.e., causal) data-generating process and its marginal effects. With this caveat held steadfastly in mind, we proceed with the exploratory data analysis, mindful of the stringent orthogonality assumptions required for causal interpretations to be valid.[9]

Dropout

The dropout indicator, labeled *dropped*, is constructed as follows. First, girls are coded as dropped if their id appears in TEA *Drop* or *Dropout Reason* files. Despite the TEA's long list of reasons coded in dropout reason files, underreporting is an acknowledged problem (TEA 2001). Estimates using the under-counted dropout variable based solely on the two dropout-related TEA files are available upon request. To deal with what is suspected to be severe under-reporting of dropout, an overly inclusive data coding convention is adopted that counts all those who attrit from the cohort (without appearing in the *Drop* or *Dropout Reason* files) as dropouts. This assumption no doubt generates false positives.

Footnote 8 continued
assumption that marriage is a proxy for pregnancy. A third way for a girl in our cohort to be coded as pregnant is if the girl was identified by the Career and Technology files as participating in programs for single parents, pregnant teens, or daycare services. Then enrollment files are checked to see if girls attended a school designated as providing education exclusively to pregnant females or parenting students. Finally, course attendance files are checked to see if girls ever took classes (outside the Career and Technology Program) identified as preparing pregnant teens for parenthood.

[9] While our "identified pregnancy" variable undoubtedly undercounts pregnancies that terminate in abortion or miscarriage, we undertook as a crude robustness check to reestimate all models reported subsequently using alternative assumptions that almost surely over-count true pregnancy. The idea was to consider a range of estimates in which correlations between the pattern of non-random misclassification or measurement error and other predictors would cover a range on the real line including zero. Pregnancy was based on: females raising children while in high school (confirmed pregnancies); and females who dropped out and were coded by the school as having done in the so-called "leaver file" (interpreted at face value and counted as pregnancies regardless of whether the girl went on to have an abortion, a miscarriage, or carried the birth to term—none of which can be observed in the available data). Girls that became pregnant, had an abortion, and did not drop are not counted as "identified pregnancies." Unidentified pregnancies are not counted in our dependent variable, which therefore contains a non-zero measurement error term. Girls who got pregnant after they dropped (or before returning to high school) and who did not utilize any funded program for pregnancy or parenthood are similarly miscoded and not counted by our "identified pregnancy" variables. As constructed based on the available data, our dependent variable's coding of "identified pregnancy" aggregates information from up to four data sources. Using successively more inclusive (and speculative) mappings from names of schools and classes taken into the construction of the coding of "identified pregnancy" outcomes, reestimation of the empirical models reveals little evidence of marginal effects that contradict those reported in this paper using the most conservative pregnancy measure that suffers from undercounting of true pregnancies.

Alternative estimates of the models reported below provide several robustness checks by beginning with the most restrictive definitions of Pregnant & Dropped and then successively changing the dependent variable using more inclusive definitions based on different proxies for pregnancy and dropout. The results reported subsequently are qualitatively robust to these variations except where explicitly mentioned otherwise.

The binary indicators described above for pregnant and dropped were then combined to form a four-valued discrete dependent variable coding the four joint outcomes: Pregnant & Dropped ($y = 1$); Pregnant & Not Dropped ($y = 2$); Not Pregnant & Dropped ($y = 3$); and Not Pregnant & Not Dropped ($y = 4$).

Descriptive Statistics

Table 1 shows the joint distribution of pregnancy and dropouts coded by the variable y among all girls in the cohort and broken out by ethnicity, immigrant status, and geographic proximity to the border between Texas and Mexico. Overall, 1.6 % of the cohort has an identified pregnancy while in high school and 39.3 % leave high school without graduating. Pregnant & Dropped is more than twice as common as Pregnant & Not Dropped (1.1 vs. 0.5 %) among the 145,175 girls in the cohort. We note that the event of pregnancy we report on is observed imperfectly. Pregnancies that end in abortion or miscarriage are unobservable and therefore not included in the identified pregnancies in our data.

The marginal pregnancy rate among all girls (1.6 %) breaks down by ethnicity as 0.6 % pregnant among whites, 1.7 % among blacks, and 2.8 % among Hispanics. If there were no ethnic differences in abortion rates, then the 2.2 percentage-point difference between Hispanics and whites would imply that pregnancy rates among Hispanic girls are four times larger than among whites and 67 % larger than among blacks. Different abortion rates by ethnicity imply that the differences we measure should be interpreted as differences in rates of becoming pregnant *and* not having an abortion.[10] Immigrants' pregnancy rate of 1.3 % is lower than for non-immigrants. And proximity to the Texas-Mexico border (in the row labeled "Border residents") correlates strongly with pregnancy, with a pregnancy rate of 3.8 %.[11]

Table 1 shows that the unconditional dropout rate among all girls in the cohort is 39.3 %, consistent with estimates from independent non-TEA researchers (Secada et al. 1998). Just under one-third of whites (31.0 %) fail to complete high school during the period of five academic years starting in 9th grade, whereas 45 % of

[10] From Table 3 (p. 277) in Henshaw and Feivelson (2000): "Per 1000 in Texas in 1996 for females aged 15–19 (abortions ending in abortion or live births): White Non-Hispanic: Pregnancy 78 per 1000, births 45 per 1000 and abortion 21 per 1000 Blacks: Pregnancy 141 per 1000, births 88 per 1000 and abortions 32 per 1000 Hispanic: Pregnancy 154 per 1000, births 109 per 1000 and abortions 20 per 1000."

[11] Proximity to the US–Mexico border in south Texas codes girls living in counties designated by the Texas Secretary of State as *colonias,* which depends on proximity to the border (e.g., within 150 miles of the US–Mexico border) and economic distress (U.S. Department of Housing 2000; Texas Secretary of State 2005). *Colonias* typically have low median income, high concentrations of Hispanic residents (although not necessarily large concentrations of immigrants), and sometimes suffer from inadequate sewage treatment, potable water, electricity, paved roads, and housing.

Table 1 Pregnancy and dropout by ethnicity, immigrant status, and proximity to Texas-Mexico border

	Number of observations	Joint distribution of pregnant and dropped outcomes				Marginal distributions	
		Pregnant & Dropped	Pregnant & Not Dropped	Not Pregnant & Dropped	Not Pregnant & Not Dropped	Pregnant	Dropped
Among							
All	1,45,175	1.1	0.5	38.2	60.2	1.6	39.3
Whites	68,221	0.4	0.2	30.6	68.8	0.6	31.0
Hispanics	54,017	1.8	0.9	45.4	51.9	2.8	47.2
Blacks	22,937	1.2	0.5	43.8	54.5	1.7	45.0
Immigrants	9204	0.8	0.5	51.5	47.1	1.3	52.4
Border residents	20,921	2.2	1.6	39.4	56.8	3.8	41.6

In the taxonomy of ethnic categories available in the administrative school data, "Whites" and "Hispanic" are mutually exclusive. In other words, the label "Whites" should be interpreted as non-Hispanic white

blacks and 47.2 % of Hispanics leave without graduating. Immigrants have the highest dropout rate at 52.4 %, and immigrant Hispanics' dropout rate (not show in Table 1) is virtually identical at 53.0 %.

Hispanic girls' unconditional estimated probability of an identified pregnancy and dropping out (Pregnant & Dropped) is 1.8 %, which is 4.5 times that of white girls and 1.5 times that of black girls. Immigrants' rate of identified pregnancy and dropping out (Pregnant & Dropped) is less than half of that of Hispanics and of Border Residents. A surprisingly uniform pattern is revealed by comparing ratios of rates of Pregnant & Dropped to Pregnant & Not Dropped, which are roughly 2:1 across all ethnic groups. Within each ethnic group, roughly 2/3 of pregnant girls wind up also dropping out and 1/3 succeed at remaining in school through graduation. The fraction of pregnant girls who stay in school through graduation is slightly larger for immigrants and much larger for Border residents. The column in Table 1 corresponding to girls who are not pregnant and have dropped out reveals a large gap between whites and non-whites in rates of graduation among those who are not pregnant. Hispanics and immigrants have especially low rates of high school completion while coded as non-pregnant.

Table 2 presents age distributions within ethnic group. More than 40 % of Hispanics and blacks in the cohort are above normal age, which means their dates of birth fall outside date-of-birth ranges published by TEA corresponding to their grade level.[12] The large age gap by ethnicity in Table 2 suggests the possibility that the biology of fertility as a function of age–together with an as-yet unspecified mechanism generating different age distributions for different ethnic groups–might explain a substantial portion of ethnic differences in Table 1.

Table 2 shows the percentages of Hispanic and black girls who are above the normal age definitions for 9th graders given by TEA (43.0 and 40.0 %, respectively) are twice that of white girls (20.1 %). Because there are relatively few below-normal-age girls and because the contrasts we are interested in primarily concern the above-normal-age girls, subsequent tables group normal-age and below-normal-age girls under the label "normal-age." While it is beyond the scope of this paper to diagnose the causes of ethnic differences in age distributions among 9th graders, we simply note that the data show there exists a highly non-uniform process with respect to ethnicity determining age among ninth graders in Texas public schools.

Table 3 shows that pregnancies are 5 times more frequent (3.5 vs. 0.7 %) among above-normal-age girls, and dropouts are 2–3 times more frequent (67.7 vs. 26.0 %). Among normal-age girls, there are smaller ethnic differences in rates of Pregnant & Not Dropped: 76.8, 68.7, and 71.5 %, respectively, among whites, Hispanics, and blacks. Marginal rates of pregnancy are 0.2, 1.5, and 0.5 %, respectively.

The lower portion of Table 3 shows that when above-normal-age girls become pregnant, a much larger fraction of them drops out than among normal-age girls.

[12] According to TEA age guidelines, girls not yet 15 years old by September 1, 1994, are classified as below normal age (of whom there were few); those who are 15 years old by September 1, 1994, but not older than 16 by August 31, 1994, are classified as normal age; and those at least 16 years old by August 31, 1995, are classified as above normal age. In the next section, we consider selection problems into Above-Age status within each grade.

Table 2 Age distributions within ethnic groups

	White	Hispanic	Black
Age status			
Below-normal-age	2.8	2.0	3.2
Normal-age	77.2	55.0	56.8
Above-normal-age	20.1	43.0	40.0
Total	100	100	100

Table 3 Pregnancy and dropout percentages by age and ethnicity, immigrant status, and border resident status

	Joint distribution				Marginal distributions	
	Pregnant & Dropped	Pregnant & Not Dropped	Not Pregnant & Dropped	Not Pregnant & Not Dropped	Pregnant	Dropped
Among						
Normal-age						
All	0.3	0.4	25.8	73.6	0.7	26.0
White	0.1	0.1	22.9	76.8	0.2	23.0
Hispanic	0.6	0.9	29.8	68.7	1.5	30.4
Black	0.2	0.3	28.0	71.5	0.5	28.2
Immigrant	0.3	0.5	37.0	62.2	0.8	37.3
Border resident	1.1	1.9	25.4	71.6	3.0	26.5
Above-normal-age						
All	2.7	0.8	64.9	31.6	3.5	67.7
White	1.5	0.5	61.3	36.7	2.0	62.8
Hispanic	3.5	0.9	66.0	29.6	4.4	69.5
Black	2.7	0.7	67.6	29.0	3.4	70.2
Immigrant	1.3	0.5	65.2	33.0	1.8	66.5
Border resident	4.0	1.1	62.4	32.5	5.1	66.3

Table 4 similarly shows that immigrants have a lower likelihood of pregnancy and a substantially higher risk of dropout. The table indicates that (at least without accounting for other factors in the conditional models presented in the next section) the higher rate of pregnancy among Hispanics (relative to blacks and whites) cannot be attributed to immigrant status. In Table 4, Hispanic–black differences in the rates of pregnancy similarly cannot be explained as the result of immigrant behavior. Residing at or near the border also is associated with large differences across all outcomes. Lower pregnancy rates among immigrants hold even among border area residents. Further cross tabulations of border resident and immigrant status (among immigrant vs. non-immigrant Hispanics, not shown in Table 4) reveal that increased rates of pregnancy among border residents hold within both subsets of the Hispanic girls in the cohort.

Table 4 Pregnancy and dropout percentages by immigrant status and ethnicity and border resident status

	Joint distribution				Marginal distributions	
	Pregnant & Dropped	Pregnant & Not Dropped	Not Pregnant & Dropped	Not Pregnant & Not Dropped	Pregnant	Dropped
Among						
Non-immigrant						
All	1.1	0.5	37.3	61.1	1.6	38.4
White	0.4	0.2	30.6	68.8	0.6	31.0
Hispanic	2.0	1.0	44.1	52.9	3.0	46.1
Black	1.2	0.5	43.8	54.6	1.7	45.0
Border resident	2.5	1.8	36.6	59.1	4.3	39.1
Immigrant						
All	0.8	0.5	51.5	47.1	1.3	52.4
White	0.0	0.6	37.7	61.7	0.6	37.7
Hispanic	0.9	0.5	52.2	46.5	1.4	53.0
Black	0.9	0.5	49.3	49.3	1.4	50.2
Border resident	1.2	0.8	49.4	48.7	1.9	50.5

Non-immigrants are a little less than 2/3 times as likely as immigrants to be dropped out, and about one quarter more likely to be pregnant

These statistics show that it is not so much being immigrant compared with non-immigrant for being pregnant, though more so for being dropped out

Most of this is being driven by Hispanics and border residents

Non-immigrant Border residents are significantly more likely to be pregnant than others, as well for immigrants, border and non-border alike

Given that the marginal distributions by each outcome and cumulatively for pregnancy and dropped out are quite different among Hispanics by immigrant status, it is quite likely that second or higher generation Hispanics have higher fertility rates than do immigrant Hispanics

Taken together, these numbers indicate that cultural effects at the border may be strongly influential, especially in generating pregnancy outcomes

Table 5 suggests that being poor greatly increases the likelihood of pregnancy and dropout. Ethnic differences in Table 5 are large. Being poor raises pregnancy rates by a factor of 2.5 among all Hispanics (1.2–3.0), by a factor of five among blacks (0.4–2.0), and by a factor of seven (0.2–1.4) among whites. Among the poor, white-black differences are generally small. Table 5 suggests that economic disadvantage and age may exert a stronger joint effect than ethnicity alone, raising questions addressed by the conditional probability models in the next section.

Results

The empirical models presented in this section are intended to provide descriptive statistical associations relevant to the question of explaining Hispanic girls' higher than average pregnancy and dropout probabilities in terms of individual, family, high school, and neighborhood characteristics. Marginal effects reported in this section are scaled in units of percentage points on a 100-point scale, representing the difference in percentage points associated with a 1-unit change in the right-hand-side variable for a girl with otherwise average characteristics. Among the four dependent variable outcomes, the omitted category is not pregnant and not dropped (NPND). Probabilities across the four outcomes must sum to 1: prob(NPNP) = 1 − prob(Pregnant & Dropped) − prob(Pregnant & Not Dropped) − prob(Not Pregnant & Dropped), which implies that the sum of marginal effects across these four categories must be zero. Results are presented only for the first three dependent variable outcomes. Except where otherwise noted, all statistical results are relative to whites as the omitted reference category.

Descriptive results showed that above-normal-age status is associated with much higher chances of pregnancy and dropout. Previous discussion of theoretical material made explicit our assumption that age potentially shifts the entire mapping of other characteristics into pregnancy and dropout outcomes. In other words, exposure to a particular family, neighborhood, or school characteristics was predicted to have distinct effects on the probability of Pregnant & Dropped among above-normal-age versus normal-age girls. We ran the models presented below fully interacted with the above-normal-age indicator variable. The null of equality of coefficients in above-normal-age and normal-age girls was easily rejected (p-val = 0.0000). Accepting that age influences virtually all other marginal effects, our analysis turned to the subpopulation of above-normal-age girls. Although the 46,099 above-normal-age girls represented just under 32 % of girls in the cohort, they produced nearly 70 % of all pregnancies and 55 % of all dropouts. The marginal effects linking individual, family, neighborhood, and school characteristics to these outcomes were, in general, larger in magnitude for above-normal age girls. This result implies that characteristics of the decision-making environment that these girls were exposed to had greater influence among older girls. In consideration of space, our reporting of marginal effects below focuses on models estimated using exclusively above-normal-age girls. Standard errors were computed using Stata's robust command, which has the effect of increasing standard errors and making it less likely (i.e., more evidentially demanding) to find statistical significance.

Table 5 Pregnancy and dropout by family income, ethnicity, and age

	Joint distribution				Marginal distributions	
	Pregnant & Dropped	Pregnant & Not Dropped	Not Pregnant & Dropped	Not Pregnant & Not Dropped	Pregnant	Dropped
Among						
Non-poor whites						
All ages	0.1	0.1	23.2	76.6	0.2	99.8
Normal-age	0.1	0.1	18.3	81.6	0.1	18.4
Above-normal-age	0.7	0.4	53.1	45.8	1.0	53.8
Poor whites						
All ages	1.0	0.4	48.6	50.0	1.4	49.6
Normal-age	0.3	0.3	37.6	61.8	0.6	37.9
Above-normal-age	2.3	0.7	69.3	27.8	2.9	71.6
Non-poor Hispanics						
All ages	0.8	0.5	32.8	66.0	1.2	33.6
Normal-age	0.3	0.4	19.2	80.1	0.7	19.5
Above-normal-age	2.0	0.7	67.6	29.7	2.6	69.6
Poor Hispanics						
All ages	2.0	1.0	47.6	49.3	3.0	49.6
Normal-age	0.7	1.1	32.3	66.0	1.7	33.0
Above-normal-age	3.6	1.0	65.8	29.6	4.6	69.5
Non-poor blacks						
All ages	0.2	0.2	29.3	70.3	0.4	29.5
Normal-age	0.1	0.2	18.8	81.0	0.2	18.8
Above-normal-age	0.6	0.4	63.6	35.3	1.0	64.2
Poor blacks						
All ages	1.5	0.6	47.8	50.2	2.0	49.3

Table 5 continued

	Joint distribution				Marginal distributions	
	Pregnant & Dropped	Pregnant & Not Dropped	Not Pregnant & Dropped	Not Pregnant & Not Dropped	Pregnant	Dropped
Normal-age	0.3	0.4	31.5	67.9	0.7	31.8
Above-normal-age	3.0	0.8	68.2	28.1	3.7	71.1

Notes As you move from younger to older, there is a significant increase in any outcome involving dropout or pregnancy

As you move from non-poor to poor, there is an even larger increase in outcomes involving dropout or pregnancy

Blacks are not greatly different from Whites for same econ-age categories

But Hispanics have much larger marg distr for outcomes involving dropout or preg than whites or blacks for same econ-age categories

And in one case, a non-poor normal age Hisp has a larger uncondit prob of being not dropped & preg than a poor normal-age white or black female

Pregnant & Dropped

Table 6 reports expected changes in the probability of Pregnant & Dropped for above-normal-age girls. The first column of Table 6 serves as a benchmark with only two ethnic indicator variables. These ethnic differences, in the absence of any further conditioning information about characteristics of girls' decision-making environments, revealed that Hispanic and black girls' rates of pregnancy were 1.987 and 1.182 percentage points higher, respectively, than white girls' rate (of 1.5 %, based on Table 2). Adding additional individual-level demographic and family characteristics, Model 2 shows that economic disadvantage (i.e., qualifying for the reduced price lunch program) and having a missing standardized test score (which is presumably a proxy for academic or other family problems that degrades academic performance) had large conditional effects. These individual-level characteristics reduced Hispanic girls' differential from 1.987 to 1.190 percentage points and almost fully explained the rate differential between black and white girls. The variable Math z-score is the z-score for each girl's math score on the statewide standardized academic achievement test. Black girls' marginal effect for Math z-score of -0.240 and economically disadvantaged girls' marginal effect for Math z-score of 1.704 imply that the mean black girl would need a math score seven standard deviations above average to compensate for being economically advantaged.

In Model 3, we added school characteristics. Variables labeled "Pct" are percentages on a 0–100 percentage point scale. The marginal effect on Pct Black at high school of -0.007 implies that the mean girl in the cohort's probability of Pregnant & Dropped was expected to fall by 0.7 percentage points if moved from a school with 0 % to a school with 100 % black classmates. In contrast, moving the mean girl from a school with no economically disadvantaged classmates to a school whose students are 100 % economically disadvantaged was expected to raise the rate of Pregnant & Dropped by as much as 3.8 percentage points. Campus size was measured as the natural logarithm of the number of students at each girl's school. This variable was negatively correlated with the rate of Pregnant & Dropped, reflecting a generally small-magnitude but still statistically significant risk factor for pregnancy at small schools that would be difficult to discover without the multivariate econometric analysis.[13] In Model 3, both Hispanic and black girls' rate differentials were around 0.85 percentage points, which was about half the size of Hispanic girls' benchmark rate differential in Model 1, and just slightly smaller than the benchmark for black girls.

Once neighborhood characteristics were included (Model 4) or both school and neighborhood characteristics were included (Model 5), the marginal effects of ethnic group on Pregnancy & Dropped shrank (relative to the benchmark Model 1)

[13] The coefficient on campus size of -0.214 implies that doubling the campus size (which in log-approximated units of the right-hand-side variable is equivalent to a one-unit increase in the log number of students) is associated with a reduction in the rate of pregnant and dropped of 0.214 percentage points. Extending the log approximation perhaps too far, Model 3 would predict that campuses whose student body is four times larger than average would enjoy a smaller rate of Pregnant & Dropped of approximately 1 percentage point lower.

Table 6 Marginal effects on the probability of being Pregnant & Dropped (unconditional = 2.71 %) among above-normal-age females ($N = 46{,}099$)

Variables	Model 1: ethnicity and age	Model 2: + individual-level demographics	Model 3: individual demographics + high school characteristics	Model 4: individual demographics + neighborhood characteristics	Model 5: individual, high school and neighborhood characteristics
Hispanic	1.987***	1.190***	0.864***	0.704***	0.796***
Black	1.182***	0.136	0.844***	0.480***	0.793***
Economically disadvantaged		1.704***	1.396***	1.559***	1.400***
Math *z*-score		−0.240***	−0.209***	−0.219***	−0.206***
No math score		3.005***	2.770***	2.866***	2.700***
ESL or bilingual classroom		−1.328***	−0.968***	−1.095***	−0.927***
Gifted		−0.781*	−0.270	−0.559	−0.265
Special ed		−1.210***	−1.164***	−1.174***	−1.133***
Immigrant		−1.342***	−0.875***	−1.180***	−0.866***
Border resident		0.949***	1.100***	1.159***	1.719***
Pct Hispanic in neighborhood				0.891***	0.420**
Pct Black in neighborhood				−0.006***	0.016
Family income in neighborhood				−0.366***	0.306
Pct immigrant in neighborhood				−0.005	0.021
Pct Spanish speakers in neighborhood				−0.002**	−0.006***
Pct mobility in neighborhood				0.046***	0.071***
Pct grandparents in neighborhood				−0.002	−0.038***
Missing census				−1.756***	15.152
Pct Hispanic at high school			0.002		−0.006**
Pct Black at high school			−0.007***		−0.007***
Pct economically disadvantaged at high school			0.038***		0.045***

Table 6 continued

Variables	Model 1: ethnicity and age	Model 2: + individual-level demographics	Model 3: individual demographics + high school characteristics	Model 4: individual demographics + neighborhood characteristics	Model 5: individual, high school and neighborhood characteristics
Pct immigrant at high school			0.000***		−0.001**
Campus size			−0.214***		−0.266***

Note "Campus size" is measured as the natural logarithm of the number of students, both male and female, attending the individual's high school

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

but did not disappear. At 0.796 and 0.763 in Model 5 of Table 6, Hispanic and black girls both appeared to face substantial additional risk of Pregnant & Dropped. In percentage terms, this risk, relative to the unconditional mean of 2.71 %, implies increased relative risk by 29 % for both Hispanic and black girls. All models indicate that immigrant girls' risk of Pregnant & Dropped and that of girls in ESL programs were lower than among non-immigrants and non-ESL students. Proximity to the border was associated with substantially elevated relative risk across all models. As more controls were added (moving left to right in Table 6), the marginal effect or being a Border resident increased, reaching 1.719 additional percentage points in Model 5. Among neighborhood effects, it was easy to notice that concentrations of Hispanic residents were associated with higher risk. And among school effects, concentrations of low-income students were another important risk factor according to results in Table 6.

There was a potentially serious multicollinearity problem in model specifications that included both high school and neighborhood composition variables. We expected that the ethnic compositions of schools and neighborhoods might very well have pairwise correlations over 0.90. That expectation was not borne out by the data, however, in pairwise correlations or using Variance Inflation Factor (VIF) statistics. Unreported correlations between characteristics of schools and the neighborhoods in which they are located were rarely larger than 0.50 across the entire State of Texas. These surprisingly modest correlations between ethnic composition of neighborhoods and schools, for example, implies that these variables provide distinct conditioning information and represent distinct components of girls' school-and-neighborhood environments, which are perhaps more complex, socially and demographically, than is commonly appreciated.

Pregnant & Not Dropped

Table 7 presents marginal effects on girls' rates of Pregnant & Not Dropped. Model 1 in Table 7 provides a benchmark for ethnic differences, indicating elevated risks of 0.405 and 0.207 percentage points for Hispanics and blacks, respectively. Relative to the unconditional mean of 0.76, these risk differentials were 53 and 27 % when expressed as relative risks.

For Hispanic girls, adding individual demographics into Model 2 raises risk only slightly and, even with the full set of right-hand-side variables in Model 5, this elevated risk level persisted at 0.289 or 38 % relative risk. Among black girls, including more control variables increased the magnitude of the marginal ethnic effect, which was 51 % relative risk (with respect to the unconditional mean rate). Being an immigrant was neither statistically nor economically significant in any of the models. Being a border resident was, however, always statistically significant and associated with increased rates of Pregnant & Not Dropped of well over 50 % relative risk across all models. Most other contextual variables did not appear to have important effect sizes. When all factors were introduced in Model 5, only the percent black at the high school was statistically significant with a modest risk-lowering marginal effect. Among neighborhood effects, the only two statistically

Table 7 Marginal effects on the probability of Pregnant & Not Dropped (unconditional = 0.76 %) among above-normal-age females ($N = 46{,}099$)

Variables	Model 1: ethnicity and age	Model 2: + individual-level demographics	Model 3: individual demographics + high school characteristics	Model 4: individual demographics + neighborhood characteristics	Model 5: individual, high school and neighborhood characteristics
Hispanic	0.405***	0.464***	0.303***	0.326***	0.289***
Black	0.207*	0.208**	0.402***	0.283**	0.384***
Economically disadvantaged		0.291***	0.276***	0.291***	0.280***
Math *z*-score		0.153	0.155	0.161	0.148
No math score		−0.670***	−0.588***	−0.625***	−0.575***
ESL or bilingual classroom		−0.516***	−0.399***	−0.455***	−0.394***
Gifted		0.043	0.252	0.061	0.190
Special ed		0.319**	0.241**	0.281**	0.242**
Immigrant		−0.166	0.033	−0.143	0.028
Border resident		0.484***	0.513***	0.470***	0.547***
Pct Hispanic in neighborhood				0.086	0.006
Pct Black in neighborhood				0.004	0.026
Family income in neighborhood				0.090	0.072
Pct immigrant in neighborhood				−0.002	0.010
Pct Spanish speakers in neighborhood				−0.002	−0.004***
Pct mobility in neighborhood				0.008	0.008
Pct grandparents in neighborhood				0.000	−0.015***
Missing census				3.836	2.862
Pct Hispanic at high school			0.020*		0.010
Pct Black at high school			−0.002*		−0.003***
Pct economically disadvantaged at high school			−0.003		0.000

Table 7 continued

Variables	Model 1: ethnicity and age	Model 2: + individual-level demographics	Model 3: individual demographics + high school characteristics	Model 4: individual demographics + neighborhood characteristics	Model 5: individual, high school and neighborhood characteristics
Pct immigrant at high school			0.000		0.000
Campus size			−0.029		−0.037

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

significant marginal effects were percent Spanish speakers and percent grandparents in the neighborhood.

Not Pregnant & Dropped

Finally, Table 8 shows marginal effects for girls who were not pregnant and have dropped out. The ethnic effects in the benchmark Model 1 are 4.700 for Hispanics and 6.259 for blacks. These effects were not extremely large relative to the unconditional mean rate of 64.94 among above-normal-age girls. The variable No math score, which codes girls who did not take the statewide test on the given test day, absorbed a very large portion of variation in dropouts among non-pregnant girls. In Models 2–5, the ethnic marginal effects switched signs relative to Model 1, indicating that Hispanic and black girls actually had a lower risk of dropping out while non-pregnant than would have been predicted by their other characteristics. As in Tables 6 and 7, Table 8 shows, once again, that all else equal, immigrants (who succeeded at having a valid math test score in the 9th grade) had lower dropout risk than native-born students with identical demographic, school, and neighborhood characteristics.[14]

Discussion

The results in Tables 6, 7, and 8 reflect the complex overlay of policies in individual schools and districts, including different school sizes, approaches to reducing pregnancy and dropout, and services offered to pregnant teens. The data presented in this paper (in Tables 2, 3, 4, 5) show that above-normal-age status in the 9th grade is one of the most powerful modulators of pregnancy and dropout risk. Therefore, policy efforts to strengthen academic performance of at-risk girls that have no explicit connection to pregnancy would be expected to provide extremely valuable mitigation of pregnancy risk. Academic support for girls at risk of grade retention early on in their public school careers would seem to provide important reduction of pregnancy risk in the years following, as many observers have previously noted.

Perhaps surprisingly, girls attending larger-than-average schools across virtually all models presented enjoyed reduced risks (both joint and marginal) of pregnancy and dropout.[15] One can imagine that the additional services that larger schools

[14] Multilevel modeling could also shed new empirical light on the joint outcomes of pregnancy and dropout. We follow Ribar (1994) and di Tommaso and Weeks (2000), however, who argue for the estimation of a single conditional expectation rather than the multilevel approach. None of the empirical results are overturned (in terms of statistical significance) by clustering at different levels of disaggregation by school campus, district, census block, or joint spatial partitions using the school attendance zones and US Census blocks.

[15] We are grateful to a referee for pointing out the following important caveats. Our multivariate results pertain primarily to the relatively small sample of students who were old for grade—both because this is part of the analytic sample we use and because the significant interactions between age for grade and school/neighborhood factors were determined to be meaningful. These empirical findings suggest that students who are old for their grade level have particularly heightened sensitivities or vulnerabilities to environmental factors that may be less consequential for students in the normative age group. If being old

provide, in terms of academic support and specialized programs encouraging pregnant girls to finish high school, have succeeded to some degree, although mean rates of pregnancy and dropout remain alarmingly large. The Dallas Independent School District has considered shifts in policies for above-normal-age students that would combine them into one large high school with services targeted at particular age groups. The empirics we report would lend at least tentative support to such policies.[16] However, the risk of missing out on beneficial heterogeneity in policy experimentation—nurturing discovery of successful new institutions and maintaining place- and school-specific cultural norms already in place at schools with high-risk characteristics but low rates of realized pregnancy and dropout risk—raise concerns about drawing any prescriptive inferences. Descriptive analysis identifying the school-specific characteristics in greater detail than is available in administrative data collected by the Texas Education Agency would provide useful information about best practices and sociocultural practices observable in neighborhoods that are far below mean rates of pregnancy and dropout. The finding that inclusion of neighborhood and school characteristics absorb up to half the effect size attributed to own-ethnic status in the benchmark models points strongly to distinct school and neighborhood social norms as a driver of pregnancy and dropout risk. The simultaneous inclusion of both school and neighborhood variation uncovers suggestive evidence, such as percent grandparents providing protective risk-mitigating effects net of school ethnic composition, perhaps as conduits for transmitting religiosity through family ties (although there is no direct information on religiosity in the model).

A potentially important caveat concerns the age of the data which are by now nearly 20 years old. Teen pregnancy rates have fallen substantially since these data were first collected. One explanation for declines in rates of teen pregnancy is increased use of contraception. Substantial gaps between Hispanic and black girls' pregnancy rates persist, however, which leads to our preferred interpretation that the data reported in this paper have continued relevance to pregnancy and dropout especially in Texas. Insofar as the true data-generating process has shifted—perhaps because of changes in education policy, increased use of contraception or other likely shifts in girls' perceived benefits and costs of pregnancy and dropout—then the marginal effects reported in this paper would, of course, change, too.

Once Hispanics become sexually active, previous work has indicated that they are less likely to use contraceptives (Solorio et al. 2004; Schuster 2003) and, once

Footnote 15 continued
within one's grade level is such a powerful risk factor as our data suggest, does it imply that we advocate social promotion to mitigate pregnancy and dropout risk associate with above-average-age status? We do not interpret our findings as giving tacit support for social promotion given the many other unknown social benefits and costs that are affected. Rather, our goal was to let the data speak descriptively, thereby providing a new empirical characterization of risk factors in the presence of peer-group, school, and neighborhood effects.

[16] An important caveat for empirical models and the extent to which they speak to real-world policy environments concerns changes in (i.e., the likely instability of) the data-generating process. The age of the data used in our study count as further grounds for interpreting the empirical results with caution.

Table 8 Marginal effects on the probability of Not Pregnant & Dropped (unconditional = 64.94 %) among above-normal-age females ($N = 46{,}099$)

Variables	Model 1: ethnicity and age	Model 2: + individual-level demographics	Model 3: individual demographics + high school characteristics	Model 4: individual demographics + neighborhood characteristics	Model 5: individual, high school and neighborhood characteristics
Hispanic	4.700***	−3.970***	−4.464***	−4.147***	−4.506***
Black	6.259***	−6.627***	−9.625***	−7.624***	−9.698***
Economically disadvantaged		10.259***	9.600***	10.180***	9.682***
Math $z-$score		−12.445***	−12.483***	−12.441***	−12.478***
No math score		36.759***	36.640***	36.695***	36.583***
ESL or bilingual classroom		−3.675***	−4.179***	−3.825***	−4.093***
Gifted		−18.869***	−20.914***	−19.239***	−20.553***
Special ed		−31.575***	−31.353***	−31.488***	−31.370***
Immigrant		−5.176***	−6.023***	−5.049***	−5.793***
Border resident		−4.926***	−5.715***	−3.634***	−3.273***
Pct Hispanic in neighborhood				−0.689***	−0.207
Pct Black in neighborhood				−0.004	−0.164***
Family income in neighborhood				−0.515	0.761
Pct immigrant in neighborhood				0.097***	0.072**
Pct Spanish speakers in neighborhood				−0.377***	−0.456***
Pct mobility in neighborhood				−0.047*	−0.034***
Pct grandparents in neighborhood				0.126***	0.367
Missing census				−2.692	−1.592
Pct Hispanic at high school			0.010		0.008
Pct Black at high school			0.072***		0.123***
Pct economically disadvantaged at high school			−0.003		−0.007

Table 8 continued

Variables	Model 1: ethnicity and age	Model 2: + individual-level demographics	Model 3: individual demographics + high school characteristics	Model 4: individual demographics + neighborhood characteristics	Model 5: individual, high school and neighborhood characteristics
Pct immigrant at high school			0.026		0.043
Campus size			−0.985***		−1.113***

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

pregnant, more likely to marry and (perhaps) drop out.[17] Some attribute these tendencies to culture-specific attitudes toward fertility and contraception (Fernández and Fogli 2005) and the availability of family planning or contraceptives in the neighborhood. According to this idea, cultural norms that run more deeply than simply identifying oneself as Hispanic (e.g., the degree to which having a family and children are emphasized as desirable markers of social success) could predispose Hispanic females to seek out motherhood (Oropesa et al. 1994; Kao et al. 2007). See, however, Ginther and Sundström (2010) for evidence from Sweden of asymmetric gender norms in leading to married women's specialization into household production (as theorized by Gary Becker). Ginther and Sundström's findings provide surprising evidence of persistent gender asymmetry even in social environments whose social norms are widely praised as offering men and women fairly equal opportunities. Their findings provide interesting context—from what would seem to be a very different environment *vis-à-vis* socioeconomic disparities by gender in Sweden—for the possibility of gendered expectations affecting Hispanic girls' education outcomes.

The empirical results presented above for immigrants are consistent with earlier papers suggesting that positive self-selection into immigration (e.g., higher levels of ability or motivation), religious institutions linked to Hispanic immigrants that spend a great deal of effort preventing girls from becoming sexually active, and stricter parenting provide a partial explanation as to why immigrant Hispanic girls tend to have lower rates of pregnancy than native-born Hispanics. This matches Stephen and Bean's (1992) findings for Hispanic females of Mexican origin aged 15–19 (based on 1970 US Census data). Hill and Johnson (2004) report that immigrant females (the so-called 1.0 generation) gave birth to more children than Hispanic females with one native- and one immigrant-parent (2.0 generation) but fewer children than were born to the US-born parents (3.0 generation). Because the women in Hill and Johnson's data were aged 15–44, their results would likely have captured a catch-up effect (i.e., information about the timing of fertility) that is missing in our data. If pregnancy rates for the third generation are indeed larger than for first- and second-generation Hispanics, this would be consistent with higher pregnancy rates among native-born Hispanic girls than among immigrant girls in our data.

For young females in poor neighborhoods who perceive few opportunities in the labor markets they will face, Edin and Kefalas (2005) suggest that becoming a parent is a rational choice because of the benefits it provides in terms of meaning and purpose in life in what otherwise is a bleak environment with few hopes of economic and social advancement. Observing that older peers have not achieved income security as workers (relative to their white peers) that they had expected, Hispanic girls still in school could interpret this as a signal that damps expected returns on effort invested into academic success and tips the balance in favor of pregnancy and dropout.

[17] The idea that dropping out to get married can be interpreted as a proxy for pregnancy is considered and empirically confirmed by Galimbertti (2004) and appears indirectly in the raw statistics on pregnancy intention in Rocca et al. (2010). See also Oropesa et al. (1994) and Landale et al. (2006).

Immigrants face a number of unique stresses (Harris et al. 2005; Rothe 2005; Afable-Munsuz and Brindis 2006). Given the higher rates of pregnancy and dropout among native-born Hispanics in our data, it is possible that these stresses become more severe for subsequent generations with parents or grandparents who immigrated. According to Rothe (2005), native-born Hispanic girls living under the same roof with foreign born parents or grandparents are likely to experience tensions that threaten to change family roles and the dynamics of expectations and norms. Rothe states that the process of migration can debilitate parents and diminish their ability to provide nurturing support when children and adolescents most need them. In contrast, Rothe's work implies that children who immigrated with their parents may have closer family ties and stronger desire for education and economic advancement relative to their home country. Our interpretation is that the stresses of being Hispanic and adaptation to US culture appear to have a delayed onset for many families, increasing significantly among those in the second generation.

Chiswick and Miller's (2002) analysis implies that we should expect that going from a neighborhood with no immigrants to one that is completely populated by immigrants would increase the likelihoods of both pregnancy and dropout. The school data from Texas appear to show the opposite, however, raising questions about conditions generating outcomes in Chiswick and Miller's data that are missing or importantly different in the cohort that we studied. The risk-mitigating effect of being surrounded by a community of peers and neighbors with similar immigrant backgrounds does enjoy some empirical corroboration in previous studies of protective effects of school peers (Upchurch et al. 1999).

Neighborhoods and High Schools

Once neighborhood effects and characteristics of high school peers are included in the models, Hispanic–black differences become nearly indistinguishable, all else equal. As previously discussed, this does not mean that racial/ethnic effects disappear entirely. The ethnic composition of high school student bodies has a significant influence on risks of pregnancy and dropout although the magnitude of these effects is modest in most cases. Higher percentages of black students reduce the probability of getting pregnant and dropping out. Unreported runs with interactions between percent black and own ethnicity do not reveal differences across own-ethnic status, implying that this protective effect applies to girls from each ethnic category. In contrast, concentrations of Hispanic students were associated with less precise and generally smaller marginal effects on pregnancy and dropout risk.

Proximity to the border with Mexico captures a geographic effect broader than neighborhood effects that some have interpreted as a proxy for distinct cultural norms (Leiske 1993). Insofar as the measure of family poverty available in our data controls for differences in family-specific poverty, the large border effects we find indicating higher risk of pregnancy and dropout cannot be attributed to low income. Our findings show that proximity to the US–Mexico border exerts a stronger effect increasing the likelihood of pregnancy than on dropout without concomitant pregnancy. Some observers familiar with Texas' border areas who have seen our

finding suggest that proximity to the border, net of all other controls in the model, could be capturing the so-called *familism* (Lugo Steidel and Contreras 2003; Piña-Watson et al. 2015) and *marianismo* (Castillo et al. 2010).

With so many Spanish-speaking immigrants in Texas, a large number of students require school services to help with English language deficiencies. Fortunately, students who receive these services are less likely to become pregnant or drop out. ESL status is also commonly interpreted as a proxy for culture (Afable-Munsuz and Brindis 2006). English language ability is influenced to some extent by students' effort applied to study (and therefore endogenous). Language is also a reflection of family background, however, and exposure to successful ESL programs in school can therefore be regarded as exogenous. Questions therefore remain regarding how to interpret the negative association of ESL participation and pregnancy and dropout risk. This association can be interpreted as simply picking up selection effects based on high-ability and high-motivation students; as the result of schools' ESL programs; or as a proxy for slow assimilation that mediates other environmental factors driving native-born Hispanics' large rates of pregnancy and dropout.

Conclusion

The empirical results reported in this paper provide new descriptive evidence regarding how characteristics of families, schools, and neighborhoods, influence high school girls' pregnancy and dropout risk. The data reveal very large differences in unconditional rates of pregnancy and dropout by ethnic status, although conditional models with the full set of controls, which include exposure to ethnic composition of neighborhoods and schools, show that differences between Hispanic and black girls' rates of pregnancy and dropout virtually disappear after accounting for other important features of girls' decision-making environments. The data also show a modest negative association between large schools and pregnancy and dropout risk. Elevated pregnancy risks for above-normal-age status (or, equivalently, the protective effect of helping girls stay at age level through their public school tenure) are perhaps underappreciated as a channel through which new policies could lower pregnancy and dropout risk.

The statistical models treat pregnancy and dropout as a jointly determined bivariate-dependent variable, coded as a four-valued discrete dependent variable, rather than modeling one of these outcomes conditional on the other as in previous studies. Enlarging the choice set of the decisions being modeled—from binary models of pregnancy conditional on dropout or dropout decisions conditional on—provides sharper detail about influences on each of those outcomes. This methodology provides new answers to the following question: How much of the large unconditional difference between Hispanic and black girls' rates of pregnancy and dropout can be explained by neighborhood and school characteristics?

One important caveat is the role that different abortion rates by ethnicity could play in influencing the empirical findings we report. Hispanic and non-Hispanic girls appear to have different attitudes toward, access to, and perceived benefits and costs of abortion. The abortion and birth frequencies by ethnicity from Texas in Henshaw and Feivelson (2000) reflect some of these differences. We argue that the

identified pregnancies in our data, which exclude abortions and miscarriages, succeed at capturing a high-stakes outcome that affects girls' labor market prospects and economic wellbeing. We therefore believe that the identified pregnancy variable used in constructing our dependent variable is worthwhile studying in its own right.

The data, as filtered through the conditional probability models with full sets of neighborhood and school controls, indicate that Hispanic–black gaps largely disappear once girls' neighborhood and school environments are accounted for. The marginal effect of being Hispanic on girls' risk of pregnancy declines by slightly more than half after accounting for neighborhood and school characteristics. The marginal effect of being Black does not shrink by as much (and gets larger in some models reported). But because black girls' benchmark risk of pregnancy begins smaller (Model 1 in Tables 6 and 7), the risks in the fully conditional model are indistinguishable across Hispanics and blacks although both remain significantly higher than that of white girls.

This paper utilizes administrative public school and the US Census data to extract information that corroborates previous studies arguing that cultural differences explain some of the elevated pregnancy risks observed among Hispanics, although one must consider distinct cultural strands in subsets of the Hispanic population. Living at the border is among the most consistently important variables in terms of statistical significance and effect size across nearly all models reported. A poor Hispanic female attending school along the US–Mexico border has a substantially greater chance of being pregnant in combination with dropping out than poor Hispanic girls attending school in other parts of the state. In contrast, Hispanic girls who immigrated with their families have reduced risks of pregnancy, all else equal. Large differences among distinct subsets of the Hispanic girls in our cohort (e.g., immigrant vs. non-immigrant Hispanic, high vs. low-income Hispanics, and Hispanics participating in ESL vs. non-participants) are consistent with an explanation that cultural norms tied to immigrants' source countries, economic position, and linguistic environments function as important influences on attitudes toward fertility and academic achievement.

References

Afable-Munsuz, A., & Brindis, C. D. (2006). Acculturation and the sexual and reproductive health of Latino youth in the United States: A literature review. *Perspectives on Sexual and Reproductive Health, 38*(December), 208–219.

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70*(April), 87–107.

Astone, N. M., & Upchurch, D. M. (1994). Forming a family, leaving school early, and earning a GED: A racial and cohort comparison. *Journal of Marriage and the Family, 56*(3), 759–771.

Barro, S. M., & Andrew K. (1987). *Who drops out of high school? Findings from high school and beyond.* Report CS-87-397c. U.S. Department of Education/National Center for Education Statistics. Washington, DC: Superintendent of Documents, U.S. Government Printing Office.

Brewster, K. L., Cooksey, E. C., Guilkey, D. K., & Rindfuss, R. R. (1998). The changing impact of religion on the sexual and contraceptive behavior of adolescent women in the United States. *Journal of Marriage and the Family, 60*(May), 493–504.

Castillo, L. G., Perez, F. V., Castillo, R., & Ghosheh, M. R. (2010). Construction and initial validation of the Marianismo beliefs scale. *Counselling Psychology Quarterly, 23*, 163–175.

Chiswick, B. R. & Miller, P.W. (2002). Do enclaves matter in immigrant adjustment? Conference paper, the Annual Meeting of the American Economic Association, Atlanta, Georgia, January 2002. Revised November 2002.

Crane, J. (1991). The epidemic theory of Ghettos and neighborhood effects on dropping out and teenage childbearing. *The American Journal of Sociology, 96*(March), 1226–1259.

di Tommaso, L., & Weeks, M. (2000). *Decision structures and discrete choices: An application to labour market participation and fertility*. Cambridge Working Paper, Series 009. Retrieved April 15, 2004, from http://www.econ.cam.ac.uk/dae/repec/cam/pdf/wp0009.pdf.

Edin, K., & Kefalas, M. (2005). *Promises I can keep: Why poor women put motherhood before marriage*. Berkeley, CA: University of California Press.

Fernández, R. & Fogli, A. (2005). Culture: An empirical investigation of beliefs, work, and fertility. National Bureau of Economic Research, Working Paper 11268.

Galimbertti, P. (2004). *Teenage pregnancy and school dropouts in Texas: Effects on earnings, high school graduation and GED certification*. Ph.D. Dissertation, University of Texas-Dallas.

Ginther, D. K., & Sundström, M. (2010). *Does marriage lead to specialization? An evaluation of Swedish trends in adult earnings before and after marriage*. Working Paper Series 1/2010, Swedish Institute for Social Research.

Guttmacher Institute. (2002). Teen pregnancy: Trends and lessons learned. *The Guttmacher Report on Public Policy, 5*(1). Retrieved September 17, 2010, from http://www.guttmacher.org/pubs/tgr/05/1/gr050107.html.

Harris, R. J., Firestone, J. M., & Vega, W. A. (2005). The interaction of country of origin, acculturation, and gender role ideology on wife abuse. *Social Science Quarterly, 86*(June), 463–483.

Henshaw, S. K., & Feivelson, D. J. (2000). Teenage abortion and pregnancy statistics by state, 1996. *Family Planning Perspectives, 32*(6), 272–280.

Higginson, J. G. (1998). Competitive parenting: The culture of teen moms. *Journal of Marriage and Family, 60*(1), 135–149.

Hill, L. E., & Johnson, H. P. (2004). Fertility changes among immigrants: Generations, neighborhoods, and personal characteristics. *Social Science Quarterly, 85*(September), 811–826.

Jargowsky, P. A. (1997). *Poverty and place: Ghettos, Barrios and the American City*. New York: Russell Sage Foundation.

Jargowsky, P. A., & Bane, M. J. (1991). Ghetto poverty: Basic questions. In L. Lynn & M. McGeary (Eds.), *Inner-city poverty in the United States* (pp. 16–67). Washington, DC: National Academy Press.

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review, 30*(3), 420–437.

Jimerson, S. R., Anderson, G. E., & Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools, 39*(4), 441–457.

Kao, H. S., McHugh, M. L., & Travis, S. S. (2007). Psychometric tests of expectations of filial piety scale in a Mexican-American population". *Journal of Clinical Nursing*, *16*, 1460–1467.

Landale, N. S., Oropesa, R. S., & Bradatan, C. (2006). Hispanic families in the United States: Family structure and process in an era of family change. In M. Tienda & F. Mitchell (Eds.), *Hispanics and the future of America*. Washington, D.C.: The National Academies Press.

Leiske, J. (1993). Regional subcultures of the United States. *The Journal of Politics, 55*(4), 888–913.

Lofstrom, M. (2007). Why are Hispanic and African American dropout rates so high? *Williams Review, 2*(Fall), 91–122.

Lugo Steidel, A. G., & Contreras, J. M. (2003). A new familism scale for use with Latino populations. *Hispanic Journal of Behavioral Sciences, 25*, 312–330.

Manlove, J. (1998). The influence of high school dropout and school disengagement on the risk of school-age pregnancy. *Journal of Research on Adolescence, 8*(2), 187–220.

Manlove, R., & Franzetta, K. (2004). Contraceptive use and consistency in U.S. teenagers' most recent sexual relationships. *Perspectives on Sexual and Reproductive Health, 36*(6), 265–275.

Moore, K., Manlove, J., Glei, D., & Morrison, D. (1998). Nonmarital school-age motherhood: Family, individual and school characteristics. *Journal of Academic Research, 13*(4), 433–457.

Moore, K. A., & Waite, L. J. (1977). Early childbearing and educational attainment. *Planning Family Perspectives, 9*(5), 220–225.

O'Connor, M. L. (1998). Adolescents with close family relationships have reduced chances of engaging in risky behaviors. *Family Planning Perspectives, 30*(March/April), 97–98.

Oettinger, G. S. (1999). The effects of sex education on teen sexual activity and teen pregnancy. *Journal of Political Economy, 107*(3), 606–644.
Ogbu, J. U. (1992). Understanding cultural diversity and learnings. *Educational Researcher, 21*(November), 5–14.
Oropesa, R. S., Lichter, D. T., & Anderson, R. N. (1994). Marriage markets and the paradox of Mexican American nuptiality. *Journal of Marriage and the Family, 56*(4), 889–907.
Piña-Watson, B., Dornhecker, M., & Salinas, S. R. (2015). The impact of bicultural stress on Mexican American adolescents' depressive symptoms and suicidal ideation. *Hispanic Journal of Behavioral Sciences, 37*, 342–364.
Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology, 24*, 1–24.
Regnerus, M. D. (2007). *Forbidden fruit. Sex & religion in the lives of American teenagers*. Oxford: Oxford University Press.
Ribar, D. C. (1994). Teenage Fertility and High School Completion. *Review of Economics and Statistics, 76*(3), 413–424.
Rocca, C. H., Doherty, I., Padian, N. S., Hubbard, A. E., & Minnis, A. M. (2010). Pregnancy intentions and teenage pregnancy among Latinas: A mediation analysis. *Perspectives on Sexual and Reproductive Health, 42*(3), 186–196.
Rothe, E. M. (2005). Hispanic adolescents and their families: Sociocultural factors and treatment considerations. *Adolescent Psychiatry, 28*, 251–278.
Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *American Educational Research Journal, 32*(Autumn), 583–625.
Rumberger, R. W. (2001). *Who drops out of school and why*. In Report, Understanding Dropouts: Statistics, Strategies, and High-Stakes Testing. National Research Council, Committee on Educational Excellence and Testing Equity Workshop, School Completion in Standards-Based Reform: Facts and Strategies, Washington, D.C., 17–18 July 2000. Washington, D.C.: National Academy Press.
Rumberger, R. W., & Larson, K. A. (1998). Student mobility and the increased risk of high school dropout. *American Journal of Education, 107*(November), 1–35.
Santelli, J. S., Durberstein Lindberg, L., Finer, L. B., & Singh, S. (2007). Explaining recent declines in adolescent pregnancy in the United States: The contribution of abstinence and improved contraceptive use. *American Journal of Public Health, 97*(January), 150–156.
Sawhill, I. (2001). *What can be done to reduce Teen pregnancy and out-of-wedlock births?* CCF Briefs, No. 8. The Brookings Institution. Retrieved October 1, 2010, from http://www.brookings.edu/papers/2001/10childrenfamilies_sawhill.aspx.
Schuster, C. 2003. Latina Adolescent Health. Advocates for Youth www.advocatesforyouth.org.
Secada, W., Chavez-Chavez, R., Garcia, E., Munoz, C., Oakes, J., Santiago-Santiago, I., et al. (1998). *No More excuses*. The final report of the Hispanic Dropout Project. Reports-Evaluative (142). Department of Education, www.ed.gov/offices/OBEMLA/hdprepo.pdf.
Solorio, M. R., Yu, H., Brown, E. R., Becerra, L., & Gelberg, L. (2004). A Comparison of Hispanic and White Adolescent Females' Use of Family Planning Services in California. *Perspectives on Sexual and Reproductive Health*, *36*, 157–161.
Stephen, E. H., & Bean, F. D. (1992). Assimilation, disruption and the fertility of Mexican-origin women in the United States. *International Migration Review, 26*(Spring), 67–88.
Tanner, J. F., Jr., Raymond, M. A., & Ladd, S. D. (2009). Evaluating a community saturation model of abstinence education: An application of social marketing strategies. *Health Marketing Quarterly, 26*, 27–41.
Texas Secretary of State. (2005). What is a Colonia? Retrieved December 5, 2005, from www.sos.state.tx.us/border/colonias/what_colonia.shtml.
The Texas Education Agency (TEA). (1995). *1994–1995 Report on Public School Dropouts*. Austin, TX: The Texas Education Agency. Retrieved June 15, 2004, from www.tea.state.tx.us/research/dropout95/reasons.html.
The Texas Education Agency. (2001). *Secondary school completion and dropouts in Texas Public Schools 1999–2000*. Austin, TX: The Texas Education Agency. www.tea.state.tx.us.
Upchurch, D. M., Aneshensel, C. S., Sucoff, C. A., & Levy-Storms, L. (1999). Neighborhood and family contexts of adolescent sexual activity. *Journal of Marriage and the Family, 61*(4), 920–933.
Upchurch, D. M., & McCarthy, J. (1990). The timing of first birth and high school completion. *American Sociological Review, 55*(2), 224–234.

U.S. Department of Housing and Urban Development. (2000). What are Colonias. Retrieved December 5, 2005, from www.hud.gov/groups/frmwrkcoln/whatcol.cfm.

Warren, C., Santelli, J., Everett, S., Kann, L., Collins, J., Cassell, C., et al. (1998). Sexual behavior among U.S. high school students, 1990–1995. *Family Planning Perspectives, 30*(4), 170–173.

Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. Chicago: University of Chicago Press.