UCLA
On-Line Working Paper Series

Title
The Validity of Self-Reports of Incidents of Sexually Transmitted Diseases

Permalink
https://escholarship.org/uc/item/27m0t26m

Authors
Upchurch, Dawn M.
Mason, William M.

Publication Date
2002-10-01

eScholarship.org Powered by the California Digital Library
University of California

The Validity of Self-Reports of Incidents of Sexually Transmitted Diseases

Dawn M. Upchurch
William M. Mason

CCPR-008-02

October 2002

California Center for Population Research
On-Line Working Paper Series

THE VALIDITY OF SELF-REPORTS OF INCIDENTS OF SEXUALLY TRANSMITTED DISEASES

Dawn M. Upchurch, Ph.D. and William M. Mason, Ph.D.

This is joint work: The authors accord each other equal responsibility for the material presented herein. This research is supported by grant R01 HD36993 from the National Institute of Child Health and Human Development. The authors gratefully acknowledge the assistance of Yasamin Kusunoki and Maria Johnson Kriechbaum.

Dawn M. Upchurch is an Associate Professor in the Department of Community Health Sciences, UCLA School of Public Health, 650 Charles Young Drive South, Los Angeles, CA, 90095-1772, and a faculty affiliate of the California Center for Population Research, UCLA. Email: upchurch@ucla.edu.

William M. Mason is a Professor in the UCLA Department of Sociology, 264 Haines Hall, 375 Portola Plaza, Los Angeles, CA 90095-1551, and a faculty affiliate of the California Center for Population Research, UCLA. Email: masonwm@ucla.edu.

Address requests for reprints to Dawn M. Upchurch.

Summary

To assess the construct validity of STD self-reports, this study compares gender and race/ethnicity differentials in the National Longitudinal Study of Adolescent Health with corresponding differentials based on CDC surveillance records.

(32 words)

Key Words and Phrases

validity of STD self-reports

individual and contextual STD risk factors

National Longitudinal Survey of Adolescent Behavior

Addhealth

CDC surveillance records

comparison of Addhealth and CDC data

population based inference

STD rates among adolescents

gender and racial and ethnic differentials in STD rates among adolescents

limitations of clinic, survey, and surveillance research designs

Abstract

Background and Objectives. Population parameter estimates for behavioral and contextual risk factors of STDs in individuals are currently unavailable. Surveillance data systems and clinical studies inherently can not provide this information. Nationally representative sample surveys that use self-reports of STD history are potentially better suited for this purpose. It remains to be established that self-reports are valid.

Goal of this Study. Assess the validity of self-reports of STD history among school-going adolescents.

*Study Design***.** We compare gender and race/ethnicity differentials in self-report STD information in the mid-1990s National Longitudinal Survey of Adolescent Health (Add Health) with corresponding differentials obtained from CDC surveillance statistics.

Results. Add Health and CDC surveillance data yield similar results. Further Add Health analysis demonstrates that family structure, socioeconomic status, and neighborhood context contribute to STD presence at the individual level.

Conclusion. Self-reports of STD infection in a national sample of adolescents can be used to study behavioral and contextual risk factors.

(144 words)

Introduction

For several reasons, the adolescent subpopulation is epidemiologically significant for sexually transmitted disease (STD) surveillance, research, and intervention strategies. In contemporary society, adolescence and not young adulthood is the *first* period of sexual activity for a large percentage of any recent cohort. To the extent that the patterns of sexual behavior that play out over the life course are established during the teenage years, the adolescent age group is of particular salience. Relative to later stages of the life course, adolescents are among the most likely to engage in unprotected sex, have multiple sexual partners, and more generally to be sexually exploratory. One consequence is increased risk of contracting an STD (1,2). Annually, there are more than three million STD cases among teenagers, who account for one-quarter of all STD infections (3,4). Adolescent girls and teens who are members of racial or ethnic minority groups are disproportionately affected by STDs and their sequelae (5,6). Beyond this brief summary, further conclusions regarding the behavioral sources of variation in the distribution of STDs over the adolescent population are difficult to find. One reason for this is that methodological issues and different target populations balkanize the research literature.

Most national health surveys study adults only, and use self-reports to assess STD risk. Self-reports are clearly subject to errors in reporting relative to biomarker tests recorded by competent personnel as part of a research protocol. Clinical studies of adolescents solve this problem, but are limited in their generality even when they are comprehensive with respect to behavioral antecedents. CDC surveillance statistics are subject to selection bias, make use of differential measurement and recording regimes

across local jurisdictions and, like clinical studies, are limited with respect to their population generalizability. As a result, there is currently no one data source that can adequately serve as a reference standard for STD risk assessment in the adolescent population. Demonstration of the extent to which different research designs and types of data can be used to assess risk would constitute an important step toward research synthesis.

This study compares self-report STD information obtained from the National Longitudinal Study of Adolescent Health (Add Health) with CDC surveillance statistics. We ask whether Add Health self-reports are good enough to sustain meaningful analysis. If they are, then it should be possible to use self-reported STD infection experience to study behavioral and contextual antecedents. To effect the comparison we juxtapose statistical associations computed from the Add Health data with approximately corresponding associations based on the application of Census population base figures to CDC STD rates. Using CDC published reports we examine the gender and race/ethnic differences found in four specific STDs: chlamydia; gonorrhea; syphilis; and AIDS. We then assess the extent of similarity in findings across the four diseases and estimate a combined model for three of them. Next, we assess the extent to which the gender and race/ethnic differentials estimated from the self-report information in the Add Health survey replicate those calculated from the CDC/Census data. Having established a degree of comparability between the patterns exhibited in the Add Health data and those present in the CDC/Census data, we examine the influences of family background and neighborhood factors on self-reported STD incidence. This enrichment of the list of sociodemographic covariates has the potential not only for greater precision relative to

administrative process generated (CDC surveillance) data, but also for better understanding of gender and race/ethnicity differentials. Furthermore, the establishment of sociodemographic comparability suggests a valuable role for STD self-reports in the study of behavioral antecedents.

Measuring STD Status and Assessing Risk

Clinical studies, the public health surveillance system, and national health surveys are the most prominent sources of information for the study of STD status and risk and each of these data sources has unique strengths and limitations. Data obtained from clinical samples include biomarker results for STD outcomes, which are highly sensitive and specific, and thus most accurately assess STD status. They also provide important clinical information pertaining to diagnosis and assessment. Clinical samples, however, afford a highly limited basis for generalization to the population(s) from which the samples are drawn, despite the quality of the information.

Process-generated surveillance data also assess STD outcomes measured by biomarker tests, but there are recognized limitations (4). Although most reporting areas adhere to standardized case definitions (7), case reporting is not uniform. Statistics based on surveillance data are more selective of infections in members of minority groups, the economically disadvantaged, and those who attend publicly funded STD clinics (5,8,9). There is also state and local area variation in the completeness of reporting (4). In addition, surveillance data are limited with respect to the amount and kind of information that can be collected in public health, point of contact settings, which usually consists of a few demographic characteristics (e.g., gender, age, race/ethnicity, and location). Nevertheless, surveillance data can serve as a useful, if imperfect, standard for the

comparative assessment of STD information obtained from other data sources, such as national health surveys.

National health surveys solve some of the problems that clinical and surveillance designs cannot, such as implementation of study designs that can sustain finite population inference, but have their own limitations. National health surveys most often rely on self-reported STD histories obtained by interviewers, self-administered questionnaires, or CASI (computer-assisted-self-interview) or audio-CASI methods. Nevertheless, these national surveys have the advantage of also providing a comprehensive set of potential explanatory factors. Consequently, if self-reports from national surveys produce findings similar to those obtained from surveillance data, for covariates that are available in both kinds of data, this can be construed as evidence of the validity of self-reports for more extended analysis of the covariates of STD acquisition using survey data.

Sociodemographic Correlates of STD

Studies consistently demonstrate that race/ethnicity, gender, and age are associated with differences in STD risk, because of between-group differences in a complex set of factors that include sexual and protective practices, sexual networks, underlying disease prevalence, and biology (5,9-15). Among adolescents, blacks, Hispanics, and American Indians have higher STD rates than whites, and Asians have lower rates (4). Adolescent girls have higher STD rates than boys (4). Furthermore, gender and race/ethnicity can have interactive (joint) as well as additive effects. We examine both in the current analysis. Age is also an important marker of STD risk. Because adolescence is a period of rapid growth and development, a single year's

difference in age can reflect substantial changes in socioemotional development, knowledge, and experience.

Family and neighborhood environments are highly salient social contexts for adolescents. The family is a central institution in the formation of sexual attitudes, role models, and standards of sexual conduct, and it provides the proximal social and economic context for youth development (16,17). Family structure predicts numerous adolescent outcomes. Generally, adolescents living with both biological parents are identified as having the optimal opportunity for overall well-being (18,19). Income and parents' education are also associated with adolescent risk-taking behavior. Consequently, we expect teens living with both biological parents to have lower odds of contracting STDs than those who do not. More highly educated parents tend to have higher educational aspirations for their children than parents with less education, and thus are more likely to discourage their adolescents' sexual activity and encourage contraceptive use conditional on (presumably uncondoned) sexual activity (16,19,20). Hence there should be an inverse association between socioeconomic status and adolescent STD infection. Neighborhood socioeconomic status and race/ethnic heterogeneity affect neighborhood social capital and collective efficacy, especially as they pertain to norm setting, collective monitoring, and social cohesion. In economically disadvantaged neighborhoods, physical deterioration is associated with a breakdown of social relationships (21), which results in more high-risk behaviors among residents and increased rates of STDs (10,22,23). In the current study, we include several family characteristics and census tract measures that characterize the potentially important attributes of the adolescents' neighborhoods.

Methods

We use data from three sources: published STD and HIV/AIDS surveillance statistics (4,24); published intercensal population estimates (25); and the National Longitudinal Study of Adolescent Health—"Add Health" (26).

STD Surveillance and Census Data

To achieve maximum comparability with the Add Health data set, the published 1995 *rates* per 100,000[a], for ages 15-19, for chlamydia[b], gonorrhea, and syphilis (4) were used[c], as were the number of reported AIDS *cases* among 13-19 year olds[d] (24). More precisely, gender-by-race/ethnicity-specific information was obtained for all 10 groups (males and females for each of the following: non-Hispanic whites, non-Hispanic blacks, Hispanics, non-Hispanic Asian and Pacific Islanders, and American Indians (which includes Alaska Natives)). Race and ethnicity were not reported by nine states for the chlamydia statistics and by three states for gonorrhea statistics. Thus, our analysis is based on published gender-by-race/ethnicity-specific disease rates for the remaining states. The 1995 mid-year population estimates for 13-19 years olds were used to compute AIDS rates for each group (25). The STD regressions reported later are weighted using census-based race/ethnic composition.

National Longitudinal Study of Adolescent Health

The National Longitudinal Study of Adolescent Health was designed to assess the general as well as sexual and reproductive health status of adolescents in the United States.

Sample Design and Selection. The sampling frame consists of all U.S. high schools. The sample was stratified by region, size and type of place, school type, ethnic

mix, and size of the school. Eighty eligible high schools were selected with probability proportional to size. Younger students were sampled from one "feeder" school that included a seventh grade and sent graduates to one of the sampled high schools. Each participating school provided a student roster, which became the student-level sampling frame. From that listing, a baseline sample was drawn consisting of a core sample and several special oversamples. The core sample is a probability sample of 12,105 and is nationally representative of adolescents enrolled in grades 7-12 during the 1994-95 academic year. The oversamples include specific ethnic groups, disabled adolescents, and biologically related adolescents. The combined Wave I sample is 20,745. The data used for this research were drawn from the 20,745 cases that were interviewed at Wave I.

Because we are interested in population-based estimates, observations that do not have Wave I sample weights were dropped from our analytic sample. Also, because the observations in the sample are clustered (adolescents within households, neighborhoods, schools, and school "communities" (i.e., high school and feeder school)) our analytic data set includes only those observations for which we could establish this clustering. In addition, because many of the multiple sibling households are not truly representative of multiple sibling households (e.g., twins were oversampled), we randomly selected one teen in those households. Further drops for consistently poor data quality, missing census tract variables, and missing STD information, resulted in a final analytic sample of 16,517 adolescents.

Measurement and Variables

Response Variables. We created three binary STD variables based on self-reports elicited from a series of questions about sexual behavior, contraception, and STDs and

HIV/AIDS using audio-CASI techniques. Respondents were asked, "Have you ever been told by a doctor or nurse that you had …" for each of the following STDs: chlamydia; syphilis; gonorrhea; HIV or AIDS; genital herpes; genital warts; trichomoniasis; hepatitis B; and for females bacterial vaginosis and non-gonoccocal vaginitis (questionnaire items H1CO16A—J).[e] The first STD variable is coded $Y_1 = 1$ if an adolescent responded affirmatively to chlamydia, gonorrhea, or syphilis; and coded $Y_1 = 0$ otherwise. The second STD variable is coded $Y_2 = 1$ if an adolescent responded affirmatively to genital herpes, genital warts, trichomoniasis, hepatitis B, bacterial vaginosis, or non-gonoccocal vaginitis; and coded $Y_2 = 0$ otherwise. This variable corresponds to the nonreportable (i.e., not reported to the surveillance system, and therefore not to CDC) STDs that are available in the Add Health data. The third STD variable is coded as $Y_3 = 1$ if either $Y_1 = 1$ *or* $Y_2 = 1$, and coded $Y_3 = 0$ otherwise.[f] Although there is sufficient detail to construct pathogen-specific response variables, and it is in principle preferable to do so, the overall level of response is too low for this approach to be worthwhile. Apart from the likely downward bias in reporting, we suspect also that the pathogen-specific reports contain a large component of error.

Covariates. We make use of covariates at the individual, family, and census tract levels. The included individual level attributes are age, gender, and race/ethnicity. Age is measured in years and included as a linear term in a logistic regression. We created a race/ethnicity composite measure that gives priority to any mention of being Hispanic, with these categories: Hispanic; non-Hispanic white; non-Hispanic black; non-Hispanic American Indian; non-Hispanic Asian; and non-Hispanic Other. This polytomy enters

into logistic regressions as a set of dummy variables (with non-Hispanic white as the reference category).

Family background characteristics include family structure, resident parents' education, and household income. Family structure is categorized as: two biological parents; biological mother, stepfather; biological father, stepmother; biological mother only; biological father only; and other situations. Mother's and father's education are coded as years of school completed. For a resident parent whose education was not reported, the missing value was imputed using weighted mean education stratified by gender of the parent and ethnicity of the teen (parents' race/ethnicity was not available).[g] We also use the natural log of 1994 household income. Values for missing cases were imputed using OLS regression with mother's education, father's education, family structure, number of hours worked per week by the father and mother, and whether public assistance was received as covariates.

To reflect the social and demographic structural attributes of the neighborhood, we include four Census tract variables obtained from the 1990 Census and appended to the Add Health data set. They are: natural log of median household income; degree of racial dispersion[h] (28); residential stability—the proportion of individuals five years and older who have lived in the same household since 1985; and youth idleness—the proportion of youths ages 16-19 who are not employed, not in school, and not in the military.[i]

Analytic Strategy

CDC/Census Analysis. We begin by providing the published disease-specific STD rates for 15-19 year olds in 1995, for each gender and race/ethnicity group. Then, to

further assess the overall disease burden across groups, the number and proportion of STD cases for each group are computed. These STD counts are derived by the application of the population counts to the published STD rates, and are labeled "adjusted" because the published STD rates are not always national for each year-age-sex-race/ethnicity combination, as noted above. Next, as a step toward eventual comparison with the Add Health data, separate regressions are computed using the log rates of a specific STD, and applying weights provided by the 1995 U.S. population ages 15-19 (13-19 for AIDS). Based on observation of the race/ethnicity-specific gender ratios for each STD type, we tested several possible interaction models. There is a gender interaction among whites for gonorrhea and chlamydia, but not for syphilis. Whereas male AIDS rates generally far exceed those of females, for blacks the rates are about the same.

Finally, we constrain the gender and race/ethnicity coefficients to equality across the chlamydia, gonorrhea, and syphilis equations while allowing for STD-specific intercepts. For all CDC/Census regressions, the standard errors are ill-defined, since at best we have population figures and at worst the data provide the ingredients for a numerical essay. It is description we are after, not estimation per se. Furthermore, there is some uncertainty in the definition of the population because the rates are based on cases rather than individuals—yet another justification for not scrutinizing standard errors.

Add Health Analysis. To the extent possible, we first attempt to parallel the CDC/Census analysis using the Add Health data. We examine gender and race/ethnicity differences in STD history, using each of the three indicators of STD history described

earlier, and compare the Add Health results to the constrained regression based on the CDC/Census data. Then, based on these findings, we use our most inclusive STD measure (all STDs—Y_3) and extend the covariate list to include family background factors, age, and neighborhood characteristics. To allow for finite population inference, weighted logistic regression is used throughout, and standard errors are adjusted for clustering at the tract level using a specialization of the "Huber" correction[j] (29). We used Stata 7 software throughout (30).

Results

CDC/Census Analysis

Table 1 reproduces the 1995 STD rates per 100,000 for adolescents ages 15-19 (13-19 for AIDS) (4,24). The rates for chlamydia, gonorrhea, syphilis, and AIDS are specified for each gender by race/ethnic group. Three general patterns stand out. First, females are disproportionately affected by chlamydia, gonorrhea, and syphilis, as demonstrated by the female-to-male ratio of the rates. Second, among these STDs, black females have the highest rates of infection; in several instances their rates are double those of the next highest group. Third, with the exception of blacks, the gender ratios reverse for AIDS rates, which not only shows that males are disproportionately affected by AIDS, but also suggests the presence of a different behavioral etiology.

Table 1 here

Table 2 shows the number and proportion of adjusted STD cases and the 1995 mid-year population for adolescents ages 15-19. In 1995, there were 337,176 (adjusted) cases of STD among 18.2 million adolescents ages 15-19. The burden of STD infection among black females is even more evident in this table; black females account for 44

percent of all STDs, yet represent only 7.3 percent of the total population of 15-19 year olds. A disproportionate number of STD cases also occurs among black males who account for 17 percent of cases, but only 7.5 percent of the population. White females account for almost one-quarter of all STD cases and represent one-third of the 15-19 year old population. Lastly, Asian males and females account for lower proportions of total STD cases than their population figures would suggest.

Table 2 here

Table 3 is essentially a re-expression and a somewhat simplified structuring of the rates in Table 1, in log regression form. The specification captures the major differentials for each STD type. For chlamydia, gonorrhea, and syphilis, blacks have the highest rates. Females have higher rates than males. For syphilis and gonorrhea, the gender differential is greater for whites than for blacks. For AIDS the gender differential reverses, although black males and females have virtually identical rates. Again, blacks have higher rates than other groups. In sum, the regression specification captures the central features of variation as displayed in the "raw data" published by CDC.

Table 3 here

Because the overall gender and race/ethnicity patterns are similar for chlamydia, gonorrhea and syphilis, the regressions for these STDs were constrained to covariate coefficient equality while allowing for STD-specific intercepts.[k] The result is presented in Table 4. Overall, the generalized STD rate (based on the three pathogens) is lower for males than females; the gender differential is greatest among whites; blacks have the highest rates. The generalized rate for American Indians is next highest, followed by that

of Hispanics, then whites (the reference category), and finally Asians. Lastly, gonorrhea and chlamydia rates are higher than those for syphilis.

Table 4 here

In sum, our analysis of the 1995 CDC surveillance data reveals similar gender and race/ethnicity patterns for chlamydia, gonorrhea, and to a lesser extent, syphilis. In contrast, AIDS stands out as distinct. We now turn to the Add Health data.

Add Health Analysis

Table 5 provides the names and weighted and unweighted means of the variables used in the Add Health analysis. Except where indicated, weighted means are described in the following text. Of particular interest is the degree of comparability of the gender-by-race/ethnicity groups in the Add Health sample as compared to those in the Census population figures. As may be seen, the weighted distributions are quite similar to the Census estimates. Whites are about two-thirds of the sample, equally divided by gender. Blacks are about 16 percent of the sample, again equally divided by gender. Slightly over 12 percent of the sample is Hispanic; 3.4 percent is Asian and .8 percent is American Indian, all equally divided by gender. Note, however, that there is one additional race/ethnicity category ("other" race/ethnicity) not available from Census data.

Table 5 here

Over half of all adolescents live with both of their biological parents and another quarter live with their biological mother only. Ten percent of adolescents live in stepfamilies, and most of them live with their biological mother and stepfather. Less than four percent live with their biological father only and almost seven percent live in other family situations (most often with relatives). The next three rows show the distributions

for the three dependent variables used in the analysis. Using the CDC definition, 1.4 percent of the sample reported ever having one of these diseases; 1.2 percent of the sample reported ever having one of the non-reportable STDs asked about in the Add Health survey; and 2.7 percent of the sample reported ever having any type of STD (excluding HIV/AIDS).

The average age of respondents at the Wave I interview date was 15.4 years. Both mother's and father's education averaged 13 years of completed schooling. Mean household income at the time of the Wave I of the survey was $43,895.

The remaining variables pertain to the census tracts within which the adolescents in the sample reside; this information is derived from the 1990 Census. The average tract median household income was $29,782, and the average racial dispersion index was .258 (which indicates modest racial heterogeneity). Adolescents lived in tracts where, on average, half of the residents lived in the house they lived in five years earlier. In addition, adolescents lived in tracts where, on average, 12 percent of youth were not in the military, employed, or in school.

Table 6 presents logistic regressions for the three binary STD variables defined earlier, with gender and race/ethnicity as covariates. The first regression pertains to the STD definition (Y_1) that most closely resembles the definition implicit in the constrained CDC regression. Males are less likely than females to have had an STD, and the gender differential is largest for whites. Blacks are most likely to have had an STD. Rates for Hispanics cannot be differentiated from those for whites. The sample sizes for Asians, American Indians and Other race/ethnicities are small, and the absolute numbers of STD

cases for these groups are minute. For these reasons, and also because the "Other" category is heterogeneous, we do not interpret the results for these groups.

Table 6 here

The second column pertains to non-reportable STDs (Y_2). Compared to the reportable STD specification, the gender difference is much larger because two of the non-reportable STDs are specific to females and are fairly common. Nevertheless, the overall patterns for the two disease definitions are substantively very similar, as can be seen from the third regression, which uses the pooled (reportable and non-reportable—Y_3) STD measure. Males are significantly less likely than females to have any STD, with the gender differential largest for whites. Blacks are most likely to have had an STD. Again, rates for Hispanics cannot be distinguished from those for whites.

In a final logistic regression we use the pooled (Y_3) STD measure and expand the covariate list to include age, family, and neighborhood characteristics. Table 7 displays the results. The essential pattern of the gender and race/ethnicity contrasts is the same as that of the corresponding regression in Table 6.[1] As adolescents age, their likelihood of having had an STD significantly increases—a decline would signal measurement error. For those adolescents living in households where a mother (or mother figure) is present, maternal education is inversely related to the likelihood of ever having an STD. Compared to adolescents living with both biological parents, those living in stepfamilies composed of a biological mother and stepfather are more likely to have had an STD. With less statistical precision the same results hold for mother-only households, for which the contrast with two-biological-parent families is even larger. Household income is not significantly associated with having experienced an STD. Neither is father's

education. The only tract variable that is significant is youth idleness: as the percentage of "idleness" increases, the likelihood of having had an STD significantly increases. This result has face validity.

Table 7 here

Discussion

Our findings support the conclusion that self-reports of STD infection in the Add Health data set are a functional tool for the study of variability in STD risk in the adolescent population. We found similar gender and race/ethnicity differentials across the CDC/Census and Add Health data sources when we used a combined measure indicating the presence of any STD. Having established in this sense that the Add Health data were of sufficient quality, we extended the set of sociodemographic covariates and found that family background and neighborhood characteristics are also plausibly associated with STD risk. Thus, our conclusion rests not only on our comparison with the CDC/Census results, but also on the face validity of the directions of association for several covariates that could never be employed using surveillance data.

Although the Add Health survey includes representation of Asians, American Indians, and Other race/ethnicities, because of small sample sizes and small numbers of STD cases we caution against interpreting the contrasts for these groups. Those that *are* interpretable (white, black, Hispanic) remain stable when family background and neighborhood characteristics are included, which suggests that other factors drive these differentials. Specifically, group differences in sexual and protective practices, sexual networks, and biology may explain the variation. For example, although females tend to have fewer sexual partners and stricter standards regarding sexual involvement than

males (33), these practices do not uniformly lower risk of STD because of differences in partners' behaviors and sexual networks (5,33) and because female adolescents and teenagers are less likely to engage in protective behaviors due to gendered power differentials in sexual relationships (33,34). Adolescent females are also biologically more vulnerable than males because of cervical squamal-epithelial immaturity, and males are more efficient transmitters than females (5). In addition, racially and ethnically structured access to health care services, as well as variation in acculturation, social and political networks, ethnic-specific sexual norms and gender roles, availability of eligible male partners, and patterns of ethnic social networks may all contribute to differences in STD risk across race/ethnic groups (5,9,35,36). Many of these hypotheses have been tested using clinical and other purposive samples of adolescents. The findings presented here suggest that national health surveys of adolescents, such as the Add Health, also have utility in examining these possible explanations of group variation in STD risk.

We find that adolescents living with both biological parents are less likely to have had an STD than those living with a stepfather and biological mother or with just their biological mother (marginal significance). Some researchers suggest that the family disruption that is responsible for the formation of stepfamilies and single parent families is also responsible for the higher incidence of risk-taking behaviors among adolescents living in these family structures (37). Others posit that parental control and supervision vary across family types (38,39), while still others suggest that adults who have experienced divorce have more permissive attitudes about adolescent sexual behavior (17). We also find that higher maternal education is associated with a lower likelihood that adolescents will have positive STD histories. Better educated parents have higher

achievement expectations for their children and may also be able to provide greater parental control over their behaviors (16,19,20). Few national studies have closely examined the mechanisms by which family sociodemographic characteristics and parent-teen relationships influence STD risk. Again, our findings suggest that national surveys such as the Add Health may be useful for this purpose.

Lastly, we find an important neighborhood association with the likelihood of having had an STD. As the proportion of idle youth increases in a census tract, so does the likelihood that adolescents living there will have had an STD. Neighborhoods with high levels of idleness may be characterized as having a highly visible "youth culture," which is often at odds with the social norms and expectations of adults. In particular, there may be a breakdown in social relations such that effective monitoring and social control of young people in the neighborhood is undermined (21). Other studies have demonstrated that this leads to an increase in high-risk behaviors as well as increased rates of STD among residents (10,22,23). Our findings suggest that further investigation into the influences of multiple social contexts is warranted.

The research described here has several limitations. First, although we used the CDC surveillance data as a standard, this data source has its own problems, as previously discussed. In particular, there is variation in the coverage of reporting by state and local areas (4,5), overrepresentation of individuals who attend publicly-funded clinics (5,8,9), and race/ethnicity information is not reported by some states, which makes it difficult to provide valid national estimates of disease (4). Also, reportable STDs are a subset of all STDs, and many that are nonreportable are epidemiologically significant (5).

Second, our study is also limited by what is possible using the Add Health data. In particular, there are small numbers of cases for each STD type, which precludes pathogen-specific analysis. In addition, the information is self-reported, and subject to error. Although the Add Health survey used audio-CASI for the most sensitive portions of the interviews, it is still possible that some respondents misreported their diagnoses. In addition, both CDC surveillance and Add Health data are based on *diagnosed* cases of STD, and this underestimates true prevalence. Asymptomatic infection is common, particularly among females, and significantly contributes to the maintenance of disease in the population (5,6). Lastly, neither the CDC/Census nor the Add Health data provide relevant clinical information (e.g., duration and severity of symptoms and other pertinent health information). Despite these limitations, our results support the utility of self-reported STD information in the Add Health survey, and suggest that it is possible to use this information to study the behavioral and contextual antecedents of STDs in a nationally representative sample of adolescents.

Endnotes

[a] Note that these are *case* rates because a given individual is at risk of infection multiple times per year.

[b] The earliest year for which national surveillance data for chlamydia were collected is 1996. Because of the epidemiological significance of this pathogen, we included it in our analysis, using the 1996 rates of infection.

[c] The pattern of differentials observed for 1995 is fairly stable over the remainder of the 1990s, although there is some trending (27). The focus on mid-decade rates and differentials maximizes comparability with the Add Health data.

[d] CDC reports AIDS cases for 13-19 year olds, not 15-19 year olds. Disaggregation of the public data is not possible.

[e] This battery of questions was limited to the subsample of respondents who responded affirmatively to a question about intercourse (more specifically, heterosexual vaginal intercourse). The survey did not obtain STD histories from adolescents who engaged in other types of sexual activities.

[f] Cases of HIV/AIDS are thus excluded from all three measures. Because the epidemiological profile for HIV/AIDS differs from that of other STDs (in particular, it is more common among males in this age group), we eliminated HIV/AIDS from consideration in our analysis of the Add Health data. This resulted in a loss of five cases.

[g] Nonresident parents by virtue of family type were coded zero on education. Any constant would be valid; zero is convenient. Interpretation of contrasts between family types without a defined parent and family types that contain both parents requires post-estimation calculation.

[h] The racial dispersion index characterizing the degree of racial homogeneity in a census tract is based on the relative distributions of white, black, Asian, American Indian, and other. An index value of "0" means the tract is completely homogeneous (i.e., only one race/ethnic group); an index value of "1" means that the tract is equally divided across the racial groups.

[i] The selection of these four variables was based on theoretical considerations and empirical findings. Because many census tract variables measure similar aspects of neighborhood environment, we investigated several other variables prior to the selection of the four used here.

[j] As a check on the stability of our results, we also employed full information maximum likelihood fixed effects at the school "community" (high school plus feeder school) level, as well as conditional likelihood (i.e., fixed effects). The results did not change.

[k] This regression was computed as a "stacked" regression with dummies for STD type. Again, standard errors are moot.

[l] We sidestep the vexing question of the implicit standardization of logistic and probit regression coefficients: It suffices for our purposes to note that the gender and race/ethnic differentials observed without controls remain in the presence of the controls introduced in Table 7 (31,32).

References

1. Aral SO. Sexual behavior in sexually transmitted disease research: an overview. Sex Transm Dis 1994; 21:S59-S64.
2. Santelli JS, Brener ND, Lowry R, Bhatt A, Zabin LS. Multiple sexual partners among U.S. adolescents and young adults. Fam Plann Perspect 1998; 30:271-275.
3. Centers for Disease Control and Prevention. HIV/AIDS surveillance report, 1999; 11(2), 2000.
4. Division of STD Prevention. Sexually transmitted disease surveillance, 1999. In: U.S. Department of Health and Human Services, Public Health Service, eds. Atlanta: Center for Disease Control and Prevention, 2000.
5. Institute of Medicine, Committee on Prevention and Control of Sexually Transmitted Diseases. The Hidden Epidemic: Confronting Sexually Transmitted Diseases. In: Eng TR, Butler WT, eds. Washington, DC: Academy Press, 1997.
6. Institute of Medicine, Committee on HIV Prevention Strategies in the United States, Division of Health and Promotion and Disease Prevention. No Time to Lose: Getting More from HIV Prevention. Washington, DC: Academy Press, 2001.
7. Centers for Disease Control and Prevention. Case definitions for infectious conditions under public health surveillance. MMWR Morb Mortal Wkly Rep 1997; 46(RR-10):1-56.
8. Fox KK, Whittington WL, Levine WC, Moran JS, Zaidi AA, Nakashima AK. Gonorrhea in the United States, 1981-1996: demographic and geographic trends. Sex Transm Dis 1998; 25:386-393.

9. Laumann EO, Youm Y. Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: a network explanation. Sex Transm Dis 1999; 26:250-261.
10. Cohen D, Spear S, Scribner R, Kissinger P, Mason K, Wildgen J. "Broken windows" and the risk of gonorrhea. Am J Public Health 2000; 90:230-236.
11. Ellen JM, Aral SO, Madger LS. Do differences in sexual behaviors account for the racial/ethnic differences in adolescents' self-reported history of a sexually transmitted disease? Sex Transm Dis 1998; 25:125-129.
12. Ford K, Norris A. Sexual networks of African-American and Hispanic youth. Sex Transm Dis 1997; 24:327-333.
13. Miller HG, Cain VS, Rogers SM, Gribble JN, Turner CF. Correlates of sexually transmitted bacterial infections among U.S. women in 1995. Fam Plann Perspect 1999; 31:4-23.
14. Tanfer K, Cubbins LA, Billy JOG. Gender, race, class and self-reported sexually transmitted disease incidence. Fam Plann Perspect 1995; 27:196-292.
15. Zenilman JM, Ellish N, Fresia A, Glass G. The geography of sexual partnerships in Baltimore: applications of core theory dynamics using a geographic information system. Sex Transm Dis 1999; 26:75-81
16. Fox GL. The family's role in adolescent sexual behavior. In: Ooms T, ed. Teenage Pregnancy in a Family Context. Philadelphia: Temple University Press, 1981:73-130.
17. Thornton A. Camburn D. The influence of the family on premarital sexual attitudes and behavior. Demography 1987; 24:323-340.

18. Maccoby EE, Martin JA. Socialization in the context of the family: parent-child interaction. In: Mussen PH, ed. Handbook of Child Psychology. Volume 4. New York: Wiley, 1983:1-101.
19. Peterson GW, Leigh GK. The family and social competence in adolescence. In: Gullota T, Adams G, Montemayor R, eds. Advances in Adolescent Development: Social Competence. Newbury Park, CA: Sage, 1990:97-138.
20. Rogers KB. Parenting processes related to sexual risk-taking behaviors of adolescent males and females. Journal of Marriage and the Family 1999; 61:99-109.
21. Sampson RJ, Morenoff JD. Ecological perspectives on the neighborhood context of urban poverty: past and present. In: Brooks-Gunn J, Duncan GJ, Aber JL, eds. Neighborhood Poverty: Context and Consequences for Children. New York: Russell Sage, 1997:1-22.
22. Kilmarx PH, Zaidi AA, Thomas JC, et al. Sociodemographic factors and the variation in syphilis rates among U.S. counties, 1984 through 1993: an ecological analysis. Am J Public Health 1997; 87:1937-1943.
23. Ramirez-Valles J, Zimmerman MA, Newcomb MD. Sexual risk behavior among youth: modeling the influence of prosocial activities and socioeconomic factors. Journal of Health and Social Behavior 1998; 39:237-253.
24. Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 1995; 7 (2), 1995.
25. U.S. Census Bureau. PPL-91R U.S. Population Estimates by Age, Race, Sex, and Hispanic Origin: 1990-1997, 1998.

26. Bearman PS, Jones J, Udry JR. The National Longitudinal Study of Adolescent Health: Research Design. Available at: http://www.cpc.unc.edu/projects/addhealth/design.html.
27. Centers for Disease Control and Prevention. Tracking the hidden epidemic: Trends in STDs in the United States 2000, 2000. Available at: http://www.cdc.gov/nchstp/dstd/Reports_Publications/pub_and_rpts.htm.
28. Billy JOG, Wenzlow AT, Grady WR. User Documentation for the Add Health Contextual Database. Seattle: Battelle, Centers for Public Health Research and Evaluation, 1998.
29. Rogers W. Regression standard errors in clustered samples. Stata Technical Bulletin 1993; 13(STB-13):88-94.
30. StataCorp. Stata Statistical Software: Release 7.0. College Station, TX: Stata Corporation, 2001.
31. Allison PD. Introducing a disturbance into logit and probit regression models. Sociological Methods and Research 1987; 15:355-74.
32. Allison PD. Comparing logit and probit coefficients across groups. Sociological Methods and Research 1999; 28:186-208.
33. Oliver MB, Hyde JS. Gender differences in sexuality: a meta-analysis. Psychol Bull 1993; 114:29–51.
34. Laumann EO, Gagnon JH, Michael RT, Michaels S. The Social Organization of Sexuality: Sexual Practices in the United States. Chicago, IL: University of Chicago Press, 1994.

35. Marin BV, Tschann JM, Gomez CA, Keegles SM. Acculturation and gender differences in sexual attitudes and behaviors: Hispanic vs. non-Hispanic White unmarried adults. Am J Public Health 1993; 83:1759-1761.
36. Sabogal F, Perez-Stable EJ, Otero-Sabogal R, Hiatt RA. Gender, ethnic, and acculturation differences in sexual behaviors: Hispanic and non-Hispanic White adults. Hispanic Journal of Behavioral Sciences 1995; 17:139-159.
37. Wu LL, Martinson BC. Family Structure and the Risk of a Premarital Birth. American Sociological Review 1993; 58:210-232.
38. Hogan DP, Kitagawa EM. The impact of social status, family structure, and neighborhood on the fertility of black adolescents. American Journal of Sociology 1985; 90:825-855.
39. Whitbeck LB, Simons RM, Kao M. The effects of divorced mothers' dating behaviors and sexual attitudes on the sexual attitudes and behaviors of their adolescent children. Journal of Marriage and the Family 1994; 56:615-621.

Table 1. Chlamydia, gonorrhea and syphilis rates (per 100,000) for ages 15-19, and AIDS rates for ages 13-19, by gender and race/ethnicity, U.S., 1995

STD/Gender	White	Black	Hispanic	Asian/ Pacific Islander	American Indian/ Alaskan Native
Chlamydia					
Male	86	976	267	81	519
Female	1,065	6,556	2,235	675	3,658
Ratio(♀/♂)	**12.4**	**6.7**	**8.4**	**8.3**	**7.0**
Gonorrhea					
Male	45	3,235	195	35	141
Female	252	4,413	349	128	454
Ratio(♀/♂)	**5.6**	**1.4**	**1.8**	**3.7**	**3.2**
Syphilis					
Male	0.4	42	2	0	1
Female	2	81	3	1	7
Ratio(♀/♂)	**5.0**	**1.9**	**1.5**	**–**	**7.0**
AIDS					
Male	8	26	18	4	11
Female	2	29	8	1	1
Ratio(♀/♂)	**0.3**	**1.1**	**0.4**	**0.3**	**0.1**

Sources: Reference 4 (Tables 3B (Chlamydia), 12B (Gonorrhea), and 23B (Syphilis)) and reference 24 (Table 9 (AIDS)).
Note: The chlamydia rates are for 1996.

Table 2. Number and proportion of adjusted STD cases ages 15-19 in 1995 in the U.S., and 1995 mid-year U.S. population ages 15-19

	STD		Population	
	Counts	Proportions	Counts	Proportions
Male				
White	9,010	.0267	6,329,488	.3477
Black	58,674	.1740	1,368,051	.0752
Hispanic	6,105	.0180	1,249,462	.0687
Asian	395	.0012	324,838	.0178
American Indian	587	.0017	86,725	.0048
Female				
White	78,707	.2334	5,960,671	.3274
Black	147,729	.4381	1,332,118	.0732
Hispanic	29,924	.0887	1,151,957	.0633
Asian	2,535	.0075	314,855	.0173
American Indian	3,511	.0104	85,207	.0047
Totals	337,176	.9997	18,203,372	1.0021

Sources: STD counts are based on the STD rates in Table 1. U.S. population figures are from reference 25.
Notes: STD counts include AIDS cases for ages 13-19. Adjusted counts for chlamydia, gonorrhea and syphilis for ages 15-19 are derived by the application of the population counts to the STD rates in Table 1, and are labeled as "adjusted" because the rates reproduced in Table 1 are not always national for each year-age-sex-race/ethnicity combination in the original source. Asian includes Pacific Islander, and American Indian includes Alaskan Native. See text for further discussion.

Table 3. Log rate STD regressions for ages 15-19 (Chlamydia, gonorrhea, and Syphilis) and ages 13-19 (AIDS), weighted using 1995 U.S. population at ages 15-19 or 13-19 (AIDS).

	Chlamydia	Gonorrhea	Syphilis	AIDS
White male	−.4955	−1.1698	–	–
Black male	–	–	–	−1.4669
Male	−2.0192	−.5514	−1.2417	1.3745
White (reference)	–	–	–	–
Black	1.8758	2.9865	4.2309	2.7329
Hispanic	.6861	.3115	1.0262	1.1448
Asian	−.5064	−1.0543	−.4527	−.7018
American Indian	1.2676	.2794	1.2508	−.2138
Constant	6.9703	5.5278	.4527	.6223
R^2	.9995	.9963	.9838	.9855

Source: Tables 1 and 2.
Note: Asian includes Pacific Islander; American Indian includes Alaskan Native.

Table 4. Log rate regressions for chlamydia, gonorrhea, and syphilis with cross-equation equality constraints for gender and race/ethnicity coefficients and disease-specific intercepts

	Coefficients
White male	− .8372
Male	− 1.0762
White (reference)	–
Black	2.8874
Hispanic	.5283
Asian	− .7547
American Indian	.7894
Chlamydia	5.5433
Gonorrhea	4.7078
Syphilis (reference)	–
Constant	.9449
R^2	.9790

Source: Tables 1 and 2.
Notes: Asian includes Pacific Islander; American Indian includes Alaskan Native. See text for discussion of the constrained computation.

Table 5. Weighted and unweighted means for Add Health subsample

Variable	Weighted Means	Unweighted Means
Gender by race/ethnicity (dummies)		
White male	.339	.256
White Female	.326	.267
Black male	.079	.101
Black female	.080	.113
Hispanic male	.063	.086
Hispanic female	.061	.088
Asian male	.018	.037
Asian female	.016	.033
American Indian male	.005	.004
American Indian female	.003	.004
Other male	.006	.006
Other female	.005	.006
Family Structure (dummies)		
Two biological parents	.542	.538
Biological mother, stepfather	.085	.081
Biological father, stepmother	.020	.020
Biological mother only	.252	.259
Biological father only	.037	.037
Other situations	.065	.065
STD reportable, excluding HIV/AIDS	.014	.014
STD non-reportable	.012	.014
STD all except HIV/AIDS	.027	.028
Age at wave I	15.419	15.652
Mother's education	13.094	13.185
Father's education	13.355	13.398
Household income	42,954	43,345
(log) Household income	10.347	10.378
Tract median HH income	29,784	31,019
(log) Tract median HH income	10.215	10.250
Tract racial dispersion	.259	.321
Tract residential stability	.551	.553
Tract youth idleness	.117	.120

Notes: Means (proportions) for polytomies may not sum to 1.0 due to rounding. *N*=16,517 except for maternal and paternal education, for which the *N*s are appropriately reduced for adolescents in families lacking a mother or father.

Table 6. Weighted logistic STD regressions for gender and race/ethnicity, by different disease definitions, Add Health

Covariates	Y_1 Reportable STDs, except HIV/AIDS	Y_2 Non-reportable STDs	Y_3 All STDs except HIV/AIDS
White male	– .7874*	– .9640+	– .9161**
Male	– .3391+	– 1.6715***	– .8365***
White (reference)	–	–	–
Black	1.5333***	1.0906***	1.3448***
Hispanic	.1550	– .0772	.0226
Asian	– 1.5087*	– .2008	– .6982+
American Indian	.6086	1.5668**	1.1415*
Other	.6149	– 2.1343**	– .0267
Constant	– 4.4245***	– 4.0648***	– 3.5212***
Wald χ^2 (7 df)	128.15***	108.17***	200.47***

+ $p \leq .1$ * $p \leq .05$ ** $p \leq .01$ *** $p \leq .001$

Note: $N = 16{,}517$ for all regressions. Logit regressions are estimated with Add Health sample weights; standard errors are adjusted for clustering at the census tract level using the "Huber" adjustment in Stata.

Table 7. Weighted logistic STD (pooled measure) regression with gender, race/ethnicity, family, and census tract covariates

Covariates	Coefficients
White male	– .7928*
Gender	– 1.0739***
White (reference)	–
Black	1.1049***
Hispanic	– .3077
Asian	– .8697+
American Indian	1.3123*
Other	.0314
Age	.4736***
Family structure	
Two biological parents (reference)	–
Biological mother, stepfather	.5570*
Biological father, stepmother	.2853
One parent – biological mother	1.2309+
One parent – biological father	– .0363
Other situations	.2968
Household income (logged)	.0247
Mother's education	– .1204***
Father's education	.0488
Tract median HH income (logged)	.0054
Tract racial dispersion	– .2662
Tract residential stability	– .4384
Tract youth idleness	1.8077***
Constant	– 10.6367***
Wald χ^2 (20 df)	418.52***

$^{+}\ p \leq .1$ $^{*}\ p \leq .05$ $^{**}\ p \leq .01$ $^{***}\ p \leq .001$

Note: N = 16,517 for all regressions. Logit regressions are estimated with Add Health sample weights; standard errors are adjusted for clustering at the census tract level using the "Huber" adjustment in Stata.