Socioeconomic Status, Socioeconomic Context and Sexually Transmitted Infections

GUY HARLING

A Dissertation Submitted to the Faculty of
The Harvard School of Public Health
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Science
in the Department of Social and Behavioral Sciences
Boston, Massachusetts.

May 2013

Author's Copyright Statement

_____

Signature of copyright holder

_____

Date

This dissertation has been read and approved by:

_____

Dr. Ichiro Kawachi

_____

Dr. SV Subramanian

_____

Dr. Till Bärnighausen

Acknowledgements

It has been a long road that has led me to the end of this dissertation. Thankfully I have had many, many friends and companions on the road with whom to share it; without you it would have been a cold and lonely path.

First, I must thank my doctoral committee for their willingness to support me in my endeavours, even when they did not directly relate to their own fields of enquiry. Ichiro Kawachi has been a wonderful academic advisor, both in the broad strokes of choosing topics to pursue, and in the nitty-gritty of line-edits. His research is the reason that I came to HSPH in the first place, and I am very glad of my decision. Subu Subramanian has been an invaluable source of statistical and practical advice, and a kind ear when, not-infrequently, problems arose. Till Bärnighausen has helped show me how I might bring together my economics and epidemiology training to answer important questions clearly. He has also been unstinting in his willingness to help me with my analyses at any hour of the day or night. I am also extremely grateful for the research opportunity extended to me by the members of the Africa Centre for Health and Population Studies in Mtubatuba, KwaZulu-Natal, and the warmth of their greeting when I visited (I thank the Committee on African Studies at Harvard for funding this visit).

Within HSPH I must thank all the staff, faculty and students who have made my time here the learning experience that it has been. Amongst the many staff I have had to pleasure to work with, I would particularly like to mention: Amy Cohen who saw the

potential in me to wrangle statistical problems (as well as all those who worked with me at the Instructional Computing Facility over the past four years); Andy Eisenman, Leah Kane and all those in Student Services who kept me on an even keel when I was working in Student Government, and when I was being a real student; and Elizabeth Solomon, Monika Szperka and Noreen Loughran who smoothed my path administratively through the department. I am also immensely grateful to all those at the Population and Development Center who supported and encouraged my work. Amongst the faculty, my time working for Marcia Castro, Günter Fink, Marc Lipsitch, Marcello Pagano and Laura White has taught me much about research and teaching; I am very grateful for their willingness to share their time and expertise with me.

There have been so many friends amongst the student body over the past five years, who have made learning a group endeavour and even sometimes a joy. I would like to single out a couple of people in particular however: Liesbeth Smit, whose intensity of outlook and focus on enjoying every moment kept my early years from ever being dull; and Heather Lanthorn, whose warmth, openness and willingness to debate any point with a smile helped my dissertation years seem less meandering than they actually were. I also offer thanks for having had the opportunity to meet Walter Guerrero, quite possibly the most genuine person I have ever known, and an inspiration to us all to always question everything, and then to do something about it. Rest in Peace, Bala.

Outside of the HSPH bubble, I would like to thank Katherine Howitt, Teresa Lilly, Ashley and Joldosh Pollock, and Tara Rooney for being the bedrocks of my time in Boston – you have listened to more of my complaints and concerns than anyone has a right to expect, and encouraged me to experience life in New England more deeply than I ever would have otherwise. Thank you.

I reserve my deepest gratitude, however, for my long-suffering family – Alan, Lesley and Isla. I didn't get where I am today without their patience and willingness to let me travel my own path, and to support me in the strange places to which said path has led. To them, I promise that this is the very last time I will graduate. At least for the next couple of decades.

And finally, I would like to dedicate this dissertation to the memory of Robin James Adam, my grandfather, whose example of scholarly study was such a powerful force in leading me down (or is it up?) the academic path.

Signature of degree candidate

Date

Table of Contents

List of Figures

List of Tables

Introduction

The explicit study of the social epidemiology of infectious diseases is a recent phenomenon (Cohen et al., 2007). However, since infection requires (either directly or via an intermediary) the interaction of an infectious and a susceptible individual, the distribution of any infectious disease necessarily reflects social processes. Those seeking to explain infectious disease outcomes have thus long relied either explicitly or implicitly on social explanations for disease risk (Brandt, 1985; Dubos & Dubos, 1987 [1952]; Krieger, 2011).

In addition to its social nature, infectious disease epidemiology is also inherently complex, since the required interaction between agents ensure non-independence of outcomes within and between populations (Halloran & Struchiner, 1995). This is accentuated by the reality that people do not mix with one-another at random, but instead often assortatively – whether by geography, social status or age we tend to associate with those who are more like us than the average. As a result, initial differences in risk are accentuated through feedback processes. The subset of infectious disease outcomes that relate to sexual transmission tend to be particularly strongly patterned by social factors, since humans tend to be highly non-random in whom they partner with sexually, and how they behave within these partnerships.

The causal pathways connecting social determinants, particularly group-level determinants, to health outcomes are necessarily lengthy – passing as they do either

through behavioural or psychological mediators on their way to having biological effects on outcomes. The length, and multiplicity, of such pathways makes it particularly likely that there will be interactions between risk factors. This in turn implies a high likelihood that risk factors will vary in their impact within and across populations and contexts – i.e. effect modification will be frequent. In the context of infectious disease and especially STIs, this goes double.

Several conceptual frameworks have been proposed that link social determinants to STIs (Bärnighausen & Tanser, 2009; Boerma & Weir, 2005; Gorbach & Holmes, 2003; Poundstone et al., 2004). Between them they propose a dizzying array of potential factors and interactions; common themes include a multiplicity of levels of aggregation and proximity to the ultimate outcome. No one study can consider them all, so careful choices need to be made in selecting which factors to examine and at which level to consider them.

The three papers in this dissertation each consider how risk for STI acquisition varies according to social factors that pattern sexual behaviour. They focus on income, income inequality and age-disparities in sexual relationships, and why they might or might not be important in specific social contexts.

The first two papers both examine whether there are economic patterns to STI risk amongst young people within the United States, over and above the well-described racial/ethnic differentials in risk. The first paper finds evidence for a negative

association between income and STI diagnosis that is stronger in the non-White population. The second paper provides a framework for evaluating which of three different mechanisms – absolute deprivation, structural inequality and relative deprivation – may be responsible for an apparent association between income inequality and health. The analysis suggests that increased absolute deprivation and structural inequality were both independently associated with STI diagnosis, and that being poor in an unequal community imposes a small additional risk. The relationship between inequality and STI risk was however confounded by race-ethnicity.

The third paper considers the role of a prominent asymmetry in context of the African HIV epidemic – that of age. The premise of this study was that age-disparities within relationships have been hypothesized to increase the risk of HIV infection amongst young women for both biological and social reasons. Biologically, older men are more likely to be infected with HIV than their younger neighbours; socially, age disparities are associated with economic and power asymmetries, and thus risky behaviours. Our analysis shows that age disparities were not associated within increased risk for HIV acquisition for young women in a rural KwaZulu-Natal context.

Since infectious disease risks are never independent across individuals, and there is such a range of risk factors at so many levels, attempts to attribute infections to

specific single causes are unlikely to be successful or helpful. Rather, it is crucial to understand the context within which infections are occurring (Aral et al., 2005), and to apply a range of theories to this context-specific knowledge (Garnett, 2007). The contexts for social determinants of STIs in the United States and sub-Saharan Africa are very different.

In the United States, while economic circumstances have long been considered predictors of STI risk (Krieger et al., 2003; Morton et al., 1979), the stratifying role of race/ethnicity in the realm of sexual behaviour has dominated research. Risk of acquisition is markedly higher amongst African-American and other Black populations for both bacterial and viral STIs (Miller et al., 2004; Prejean et al., 2011). It remains far higher even allowing for one's own sexual risk behaviours (Hallfors et al., 2007). This risk differential has been attributed to both assortative mixing by race/ethnicity, and a range of social factors affecting sexual mixing patterns within African-American populations (Adimora & Schoenbach, 2005; Laumann & Youm, 1999). It is within this context that one must interpret the limited associations we find between the economic characteristics of both individuals and their environment, and STI risk.

In the poor, rural South African setting of the other paper in this dissertation, where cumulative HIV incidence rates are over 70% by age 50 (Bärnighausen et al., 2008), age asymmetries do not appear to be generating infections for young women. HIV in Africa has consistently surprised those who expected risk factors from other settings and health outcomes to transfer neatly to this epidemic. Examples have

included the effectiveness of treating STIs to prevent HIV (Grosskurth et al., 2000), and the relationship between socioeconomic status and HIV prevalence (Mishra et al., 2007; Wojcicki, 2005). In both cases, a full understanding of how risks differed by time and place required an understanding of what has been described as the "dynamic topology of sexually transmitted disease epidemics" (Wasserheit & Aral, 1996). In our case, both biological and social factors are likely to play a role in explaining our null finding: the high background risk of HIV infection, regardless of partner's age, and the intertwined employment and marital dynamics of the setting may each have acted to dampen the differential risk of infection for age-disparate couples.

While these studies are clearly only small steps forward, each provides an opportunity to explore the social determinants of STIs in a particular context. Each study suggests further possible research into the more proximate determinants which might be mediating the observed findings (or their absence where expected), and each could be usefully replicated in other settings to examine which contextual factors are important in determining which causal mechanisms do or do not act.

Ultimately, however, the most important role of these papers and their findings may be in pointing to the centrality of considering the social context in understanding how individual-level social factors are associated with STI risk. A firm knowledge of context allows us to make sense of, and even anticipate and build into our studies analysis of, broad social processes likely to modify the effect of even the social determinants of health.

References

Adimora AA & Schoenbach VJ. Social Context, Sexual Networks, and Racial Disparities in Rates of Sexually Transmitted Infections. *Journal of Infectious Diseases*, 2005; 191(Suppl 1): S115-22.

Aral SO, Padian NS & Holmes KK. Advances in Multilevel Approaches to Understanding the Epidemiology and Prevention of Sexually Transmitted Infections and HIV: An Overview. *Journal of Infectious Diseases*, 2005; 191 Suppl(Suppl 1): S1-6.

Bärnighausen T & Tanser F. Rethinking the Role of the Local Community in HIV Epidemic Spread in Sub-Saharan Africa: A Proximate-Determinants Approach. *HIV Therapy*, 2009; 3(5): 435-45.

Bärnighausen T, Tanser F, Gqwede Z, Mbizana C, Herbst K & Newell M-L. High HIV Incidence in a Community with High HIV Prevalence in Rural South Africa: Findings from a Prospective Population-Based Study. *AIDS*, 2008; 22(1): 139-44.

Boerma JT & Weir SS. Integrating Demographic and Epidemiological Approaches to Research on HIV/AIDS: The Proximate-Determinants Framework. *Journal of Infectious Diseases*, 2005; 191(Suppl 1): S61-7.

Brandt AM. *No Magic Bullet: A Social History of Venereal Disease in the United States since 1880*; 1985. New York, NY: Oxford University Press.

Cohen JM, Wilson ML & Aiello AE. Analysis of Social Epidemiology Research on Infectious Diseases: Historical Patterns and Future Opportunities. *Journal of Epidemiology and Community Health*, 2007; 61(12): 1021-7.

Dubos R & Dubos J. *The White Plague: Tuberculosis, Man, and Society*; 1987 [1952]. Rutgers University Press.

Garnett GP. Theory Is Critical in Understanding the Risks of Acquiring HIV. *Sexually Transmitted Diseases*, 2007; 34(10): 737-8.

Gorbach PM & Holmes KK. Transmission of STIs/HIV at the Partnership Level: Beyond Individual-Level Analyses. *Journal of Urban Health*, 2003; 80(4 Suppl 3): iii15-25.

Grosskurth H, Gray R, Hayes R, Mabey D & Wawer M. Control of Sexually Transmitted Diseases for HIV-1 Prevention: Understanding the Implications of the Mwanza and Rakai Trials. *Lancet*, 2000; 355(9219): 1981-7.

Hallfors DD, Iritani BJ, Miller WC & Bauer DJ. Sexual and Drug Behavior Patterns and HIV and STD Racial Disparities: The Need for New Directions. *American Journal of Public Health*, 2007; 97(1): 125-32.

Halloran ME & Struchiner CJ. Causal Inference in Infectious Diseases. *Epidemiology*, 1995; 6(2): 142-51.

Krieger N. *Epidemiology and the People's Health: Theory and Context*; 2011. New York, NY: Oxford University Press.

Krieger N, Waterman PD, Chen JT, Soobader M-J & Subramanian SV. Monitoring Socioeconomic Inequalities in Sexually Transmitted Infections, Tuberculosis, and Violence: Geocoding and Choice of Area-Based Socioeconomic Measures--the Public Health Disparities Geocoding Project (US). *Public Health Reports*, 2003; 118(3): 240-60.

Laumann EO & Youm Y. Racial/Ethnic Group Differences in the Prevalence of Sexually Transmitted Diseases in the United States: A Network Explanation. *Sexually Transmitted Diseases*, 1999; 26(5): 250-61.

Miller WC, Ford C, Morris M, Handcock MS, Schmitz JL, Hobbs MM et al. Prevalence of Chlamydial and Gonococcal Infections among Young Adults in the United States. *JAMA*, 2004; 291(18): 2229-36.

Mishra V, Assche SB, Greener R, Vaessen M, Hong R, Ghys PD et al. HIV Infection Does Not Disproportionately Affect the Poorer in Sub-Saharan Africa. *AIDS*, 2007; 21 Suppl 7: S17-28.

Morton WE, Horton HB & Baker HW. Effects of Socioeconomic Status on Incidences of Three Sexually Transmitted Diseases. *Sexually Transmitted Diseases*, 1979; 6(3): 206-10.

Poundstone KE, Strathdee SA & Celentano DD. The Social Epidemiology of Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome. *Epidemiologic Reviews*, 2004; 26: 22-35.

Prejean J, Song R, Hernandez A, Ziebell R, Green T, Walker F et al. Estimated HIV Incidence in the United States, 2006-2009. *PLoS ONE*, 2011; 6(8): e17502.

Wasserheit JN & Aral SO. The Dynamic Topology of Sexually Transmitted Disease Epidemics: Implications for Prevention Strategies. *Journal of Infectious Diseases*, 1996; 174 Suppl 2: S201-13.

Wojcicki JM. Socioeconomic Status as a Risk Factor for HIV Infection in Women in East, Central and Southern Africa: A Systematic Review. *Journal of Biosocial Science*, 2005; 37(1): 1-36.

Socioeconomic disparities in sexually transmitted infections among young adults in the United States: examining the interaction between income and race/ethnicity

Guy Harling, SV Subramanian, Till Bärnighausen, Ichiro Kawachi

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

Abstract

Background: There is considerable evidence of racial/ethnic patterning of sexually transmitted infection (STI) risk in the United States. There is also evidence that poorer persons are at increased STI risk. Evidence regarding the interaction of race/ethnicity and income is limited, particularly nationally at the individual level.

Methods: We examined the pattern of socioeconomic gradients in STI infection amongst young people in a nationwide US study, and determined how these gradients varied by race/ethnicity. We estimated the cumulative diagnosis prevalence of Chlamydia, Gonorrhoea or Trichomoniasis (via self-report or laboratory confirmation) for young adults (ages 18-26) Hispanics and non-Hispanic Whites, Blacks, and Others across income quintiles in the Add Health dataset. We ran regression models to evaluate these relationships adjusting for individual- and school-level covariates.

Results: STI diagnosis was independently associated with both racial/ethnic identity and with low income, although the racial/ethnic disparities were much larger than income-based ones. A negative gradient of STI risk with increasing income was present within all racial/ethnic categories, but was stronger for non-Whites.

Conclusions: Both economic and racial/ethnic factors should be considered in deciding how to target STI prevention efforts in the United States. Particular focus may be warranted for poor, racial/ethnic minority women.

Introduction

It has long been recognized that sexually transmitted infection (STI) risk in the United States (US) is strongly patterned by race/ethnicity, with much the highest rates amongst African-Americans (Centers for Disease Control and Prevention, 2011; McQuillan et al., 2004; Miller et al., 2004; Miller et al., 2005; Nakashima et al., 1996). Black men have higher-risk sexual behaviors than White men, however even within strata of sexual and substance-use behavior there remain large differences in STI rates between race/ethnicities, particularly for Black individuals (Hallfors et al., 2007).

We expect STI risk to be associated with income, since lower income is associated with less access to preventative information and healthcare, and increased use of sex for economic purposes and as a psychosocial coping mechanism (Hogben & Leichliter, 2008). Past studies of income and STI in the US have found mixed results. Ecological studies have found a positive correlation between STI rates and area-level socioeconomic status (SES) (Dolan & Delcher, 2008; Krieger et al., 2003; Springer et al., 2010). Two nationally representative studies have collected individual-level information on STI infection and income. Among adults in the National Health and Nutrition Examination Survey (NHANES), Chlamydia was associated with poverty (Datta et al., 2007). In the National Longitudinal Study of Adolescent Health (Add Health), higher household income did not predict bacterial STI risk in adolescence, while by early adulthood, Add Health respondents'

childhood experiences of low income were only crudely associated with increased STI risk (Buffardi et al., 2008).

Race/ethnicity is an important predictor of income in U.S. society; if race/ethnicity confounds the association between income and STIs, income may be a mediator or moderator of the race/ethnicity-STI relationship (LaVeist, 2005). The former would reflect a pathway from racial/ethnic identity through income that acts similarly for all racial/ethnic groups; the latter a pathway that acts differentially by race/ethnicity. To fully explore how income interacts with race/ethnicity with respect to STIs, and in particular how income predicts STI risk within racial/ethnic groups, it is necessary to examine race/ethnicity and income jointly.

Previous analysis has highlighted the interplay of incarceration, social and sexual network segregation and impoverished circumstances that places some racial/ethnic groups, particularly African-Americans, at far higher risk of infection than the rest of the population (Adimora & Schoenbach, 2005; Aral, 1999; Laumann & Youm, 1999). Analysis of within-race/ethnicity risk gradients has been less well explored. One study, of Gonorrhoea risk in California by area-level poverty, found that although race/ethnicity was the strongest predictor of risk, gradients existed within all racial/ethnic groups; the gradient was steepest for Whites and shallowest for Hispanics (Springer et al., 2010). Another study, of the relationship between education and STI risk in Add Health, found steeper gradients amongst White women for self-reported diagnosis, and steeper gradients amongst Black women for laboratory-confirmed STI (Annang et al., 2010). Finally, a study of poverty and STI

risk in Add Health found a non-significant trend among Black adolescent men (Newbern et al., 2004), but no clear gradient for Whites or Black women.

None of these existing studies provide national evidence regarding income gradients in STI diagnosis within racial/ethnic categories using individual-level data, with the exception of Newbern and colleagues (Newbern et al., 2004), who focus on school-aged respondents. We extend their analysis to cover the period up to young adulthood, to determine how adolescent economic circumstances predict STI risk during individuals' most high-risk years.

Materials and Methods

This analysis used Waves I to III of the Add Health survey, which has followed a nationwide cohort since their adolescence in the mid-1990s ; understanding sexual behavior and health was one of its primary design interests (Harris et al., 2009). A sample of 80 US high schools (plus 52 of these schools' largest feeder schools) was selected to represent US schools with respect to region of country, urbanicity, school size, school type, and ethnicity. Wave I (1994-95) surveyed a sample of all enrolled students in grades 7 through 12 at home. Wave II (1996) re-surveyed those who had been in grades 7 through 11 at Wave I. Wave III (2001-02; ages 18-26) sought to locate and interview all those surveyed at home in Wave I.

The base study population for this analysis comprised all respondents who were interviewed at Waves I and III, provided information on their age and sex, and were

affiliated with one of the 132 core schools. We then excluded respondents whose parents did not provide information on family income or household size. Ethical approval for the Add Health study was obtained from the Institutional Review Board (IRB) at the University of North Carolina, Chapel Hill. This analysis was exempted by the Harvard School of Public Health IRB as a secondary analysis of existing data.

Measures

The primary outcome for this study was a binary measure reflecting whether a respondent had self-reported or laboratory-confirmed *Chlamydia trachomatis*, *Neisseria gonorrhoeae* or *Trichomonas vaginalis* at either Wave II or III. At Wave III, respondents were asked to provide a urine sample testing; detailed descriptions of the testing methods and evidence of their sensitivity and specificity are available elsewhere (The Add Health Biomarker Team). Also at Wave III, respondents were asked whether a health professional had, within the past 12 months, told them that they were infected with each of these STIs. At Wave II respondents were asked whether they had been diagnosed since Wave I and at Wave I they were asked if they had ever been diagnosed.

Income was based on parental reports at Wave I of 1994 total pre-tax household income (in $1000 increments, top-coded at $999,000; no income data was collected at Wave II). Household incomes were equivalized by dividing them by the square root of the number of individuals in the household – an approach adopted by the "Luxembourg Income Study" which accounts for economies of scale arising from

some household consumption being non-rivalrous in consumption, i.e. use by one member does not diminish the amount available for others (Atkinson et al., 1995). Incomes were categorized into quintiles, using the highest quintile as the reference category, to allow for the detection of non-linearities in STI diagnosis gradients. We classified race/ethnicity into four categories based on respondents' self-report of Hispanic ethnicity and their primary racial identification: White non-Hispanic, Black non-Hispanic, any Hispanic, and Other non-Hispanic (hereafter "White", "Black", "Hispanic" and "Other").

Additional covariates considered as potential confounders of the relationships between race/ethnicity, income and STIs included respondents' age (in years) and sex at Wave I and school urbanicity (urban, suburban or rural), regional location (West, Midwest, South, North-East) and type (public or private).

Analytic approach

We calculated cumulative risk proportions for each combination of race/ethnicity and income quintile, and their adjusted Wilson score 95% approximate binomial confidence intervals (CI). All statistical tests were two-sided at $\alpha=0.05$ and regression analyses were conducted in SAS version 9.3 (SAS Institute; Cary, NC). Multivariable analysis was conducted using logistic regression models, using survey procedures which allow for clustering at the school level, and sampling weights which adjust for non-response and the unequal probability of selection. We initially established the relationship between racial/ethnic category and STI risk, adjusting

for age, sex and school-level covariates. We then added income quintile as a covariate to assess the degree to which income mediated the race/ethnicity-STI relationship, and finally included interactions of race/ethnicity and income quintile to assess effect modification. We also ran models stratified by sex and considering each STI separately.

We conducted three sensitivity analyses: first, given a low likelihood of reverse causation, we included as cases those individuals reporting an STI diagnosis at any age prior to Wave I; second, we restricted our sample to respondents who were interviewed at all three waves; and third, since the impact of family income might be expected to exert its greatest effect while students were in school, we restricted our sample to those interviewed at Waves I and II, and used self-reported STI diagnosis at either wave as our outcome.

Results

10,791 respondents were interviewed at both Waves I and III, were affiliated with a core school, provided information on age and sex, had parents who reported household size and income, and either answered questions relating to STI history at Waves II and III or provided a valid urine sample for STI testing at Wave III. Age or sex information was missing for 13 Wave III respondents, a further 82 were not from core schools, 3,594 more had no household income information and 423 others lacked sampling weights.

Respondents were almost all aged between 13 and 18 at baseline with an even gender split (Table 1.1). The sample was more Black and Hispanic than the general US population, and the schools which they attended, reflecting Add Health's intentional oversampling of minorities. Median equivalized per-capita income was $22,660 (95%CI $20,972 - $24,348). Respondents falling in the poorer quintiles of the sample were more likely to be Black or Hispanic and less likely to be White. They were also more likely to come from the South and to attend urban or rural, as opposed to suburban, schools.

Across Waves II and III, prevalence of either a recent diagnosis of, or positive test for, at least one STI was 9.2%. The most common diagnosis was of Chlamydia (6.7%), followed by Trichomoniasis (2.6%) and then Gonorrhoea (1.5%). Diagnosis risk was highest for Blacks (26.1%), followed by Hispanics (10.6%), Others (9.3%) and finally Whites (5.4%). The risk of diagnosis fell as income increased, from 14.7% in the poorest quintile to 5.2% in the richest quintile. This gradient was observed for all four racial/ethnic groups, although the patterns were not strictly monotonic in every instance (Figure 1.1, values in Supplementary Table 1.4). Bivariate regression analysis confirmed that Whites were at significantly lower risk of STI diagnosis than all other groups, and all income quintiles at significantly higher risk compared with the richest quintile (Supplementary Table 1.5).

In multivariable models (Table 1.2, complete results in Supplementary Table 1.6), Hispanics and Others had approximately double the odds of STI diagnosis compared with Whites, while Blacks had more than six times the odds. The addition of income

reduced the race/ethnicity differentials marginally; however income had an independent association with STIs, with the poorest quintile having 83% increased, and the middle three quintiles having roughly 50% increased, odds compared to the richest quintile. When we interacted race/ethnicity and income, the racial/ethnic differences in the highest income quintile changed little for Blacks and Hispanics. Income gradients were steepest amongst Others, followed by Hispanics and Blacks and flattest for Whites. As a result, disparities between Whites and Others were most pronounced amongst the poorest (Supplementary Table 1.7).

When we stratified the analysis by sex, there were no clear gradients in income for White men or women (Table 1.3; Supplementary Table 1.7 and Supplementary Table 1.8 present the same results with direct comparisons within income groups and within racial/ethnic groups respectively). Amongst all other groups, income gradients were steeper for women than for men. The strongest gradient existed for Black women, amongst whom the two poorest quintiles had over 2.5 times the odds of STI diagnosis compared to the richest quintile (OR: 2.68, 95%CI: 1.48-4.85 and OR: 2.70, 95%CI: 1.37-5.30). The average diagnosis ratio between Blacks and all others was wider for women than for men; this reflected White-Black disparities which were larger amongst poor women which than men, but became roughly equal by the highest income quintile.

Analyzing each STI outcome separately, the greatest racial/ethnic disparity existed for Gonorrhoea, reflecting a particularly large gap between Blacks and all other groups. Income gradients were visible for Chlamydia and Gonorrhoea; the gradient

for Trichomoniasis was shallow. Within race/ethnicities, significant differences existed between richest and poorest quintiles for all groups for Chlamydia diagnosis; for other infections small numbers of diagnoses led to unstable estimates and gradients, although a gradient was notable by its absence for Trichomoniasis amongst Whites. The three sensitivity analyses had limited impact on the key findings (Supplementary Table 1.9).

Discussion

This study provides the first analysis of income gradients in STIs within race/ethnicity groups using a national US sample of individuals including young adults. In line with existing studies, including a previous report using this dataset (Miller et al., 2004), we find large differentials in STI risk across racial/ethnic groups – over a sixfold increase in the odds of either physician or laboratory report for Blacks compared to Whites. This study moves beyond prior analyses in finding that this racial/ethnic disparity continues and perhaps strengthens into young adulthood. We find that these racial/ethnic disparities are only weakly related to income – adding income to a model containing measures of race/ethnicity reduced the point estimates on the latter by between 5 and 15%. Income is nonetheless an independent predictor of STI risk. This is consistent with existing race/ethnicity-adjusted ecological analyses linking area-level poverty and Gonorrhoea rates in California (Springer et al., 2010). In contrast, previous studies of STI diagnosis in

Add Health have suggested little relationship with parental income using single-wave outcomes (Buffardi et al., 2008; Newbern et al., 2004), although parental and own education levels were predictive (Annang et al., 2010; Ford & Browning, 2011). Pooling diagnoses across waves may have increased our power to detect an association.

When we allow for income and race/ethnicity to interact in our models, we find some evidence for effect-modification, reflected in the better fit of the interaction model as measured by the Akaike Information Criteria (Table 1.2). Our analyses show that income affected STI diagnosis probability less for Whites than for others, and that racial/ethnic disparities were least pronounced amongst the rich. Moreover, stratification by sex led to income gradients amongst Whites disappearing entirely, suggesting that income is only related to STI diagnosis amongst Whites insofar as it reflects gender differences in income. Our finding of steeper risk gradients amongst Blacks is congruent with existing studies of SES, race/ethnicity and STIs in Add Health: past research has found maternal education and occupation at Wave I, and own education at Wave III, to be associated with STIs amongst Blacks but not Whites (Annang et al., 2010; Newbern et al., 2004).

Stratification by gender also shows a stronger income gradient for women than for men amongst Blacks and Hispanics. Combined with the finding of greater disparities by income within non-White groups, this result highlights that gender, race/ethnicity and income interact to place poor Black women at particularly increased risk of STIs.

There has been significant research describing how sexual networks, and hence sexual risks, are heavily structured by race/ethnicity, and how this leads to racial/ethnic disparities in STI rates (Aral, 1999; Laumann & Youm, 1999). Whilst being poor, female and African-American are all independently risk factors for STIs (Sharpe et al., 2012), our finding of a stronger income gradient for Black women is somewhat surprising. This is because African-Americans have relatively low risk-homophily – i.e. women who are otherwise low-risk tend to have higher-risk partners, due to a range of factors (including racial/ethnic homophily, imbalanced sex ratios – due to higher male mortality rates – and extremely high incarceration rates) limiting their choice of sexual partners (Adimora & Schoenbach, 2005). Such disassortative mixing should theoretically lead to less variation in STI risk across the income gradient. A possible explanation is that being "poor" in this study does not have the same meaning for all racial/ethnic groups. Median net worth within the bottom income quintile in the 2000 census was $24,000 for Whites but less than $100 for Blacks (LaVeist, 2005). Income quintiles may therefore not reflect the same socioeconomic circumstances for each race/ethnicity, and thus steeper gradients for Blacks may reflect the greater depth of their poverty. (We note that this does not explain why Black women have a steeper income-risk gradient than Black men.)

We find variation in income gradients by race/ethnicity. One potential explanation of this finding relates to spatial concentration. STIs with low population prevalence (e.g. Syphilis, Gonorrhoea) tend to be most concentrated by geography (Kerani et al., 2005), by race/ethnicity (Centers for Disease Control and Prevention, 2011), and by

income (Miller et al., 2004). Within a single STI, concentration of infection also appears to be highest amongst African-Americans (Springer et al., 2010), and to be highest in areas where Blacks experience certain dimensions of geographic and economic segregation (Biello et al., 2012). These neighborhoods exhibit high levels both of prevalent STIs and social risk factors (such as drug use rates and high-risk sexual norms) likely to increase risky sexual behavior (Jennings et al., 2010; Jennings et al., 2012). Our observed steeper income gradients amongst racial/ethnic minorities might then reflect the higher likelihood of poor minority individuals living in these areas of concentrated high STI risk – compared to poor Whites. Such an argument is congruent with a concentration of STI risk amongst poor minority individuals. Given existing evidence that sexual risk behaviors do not explain racial/ethnic disparities in STI rates (Hallfors et al., 2007), it might also be of interest to explore whether they explain the income gradients seen in this study.

Our use of the Add Health dataset provides some notable strengths. The prospective, longitudinal nature of the dataset should limit concerns regarding the temporal direction of any associations, especially since attrition is relatively low and does not appear to greatly affect prevalence estimates (Chantala et al.). Using multiple waves of outcome data additionally raises our power to detect effects. Furthermore, the study's national coverage allows us to draw nationwide conclusions. Additionally, our use of both laboratory testing (avoiding bias arising from variation in healthcare access) and audio-computer-assisted-interview self-report data (ensuring prior but treated cases are captured) strengthens our approach and has the added benefit

that social desirability bias in self-reports should be limited by respondents' knowledge that they are also being laboratory-tested.

Nevertheless, there are also a number of potential limitations to our analysis. First, we rely on school (and residential) location at a single time-point, which may have resulted in misclassification of context across waves. Second, although our STI outcomes combine self-reported clinical diagnosis and laboratory testing, there is a five-year gap in respondents' self-report (from Wave II up to one year prior to Wave III) which is troubling if diagnostic patterns (by race/ethnicity or income) differ systematically across respondents' lifetimes. This concern is somewhat allayed by the fact that Add Health effectively comprises six cohorts (since enrolment covered six grades), and thus all ages between 14 and 26 are covered by both self-report and laboratory testing. Consequently, these gradients would need to vary systematically both by age of respondent and by birth cohort in order to generate bias. We have no hypothesis as to why this form of systematic variation might exist; however were it the case then our results would not be generalizable to other birth cohorts.

Third, as is common in survey-based analyses, many individuals do not have income data (24.3%), which may have led to selection bias. Those with missing income data were significantly more likely to be non-White, however they were not more likely to have a positive test result across the whole sample, or within racial/ethnic groups (Supplementary Table 1.10). Finally, it is important to note the context of this study – US youth in the 1990s and early 2000s – when Chlamydia rates were rising and

Gonorrhoea (and perhaps Trichomoniasis) rates were falling (Aral et al., 2007), when extrapolating results elsewhere.

Our study provides evidence that, while racial/ethnic differentials are significantly larger than income differentials in STI rates nationwide, both factors are independent predictors of increased risk. The Centers for Disease Control and Prevention are committed to integrating consideration of social determinants of health into STI prevention program design (Centers for Disease Control and Prevention, 2010). Prevention efforts for STIs in the United States often focus on African-American populations (Crepaz et al., 2009; Henny et al., 2012), reflecting their very high infection rates at all income levels. Our analysis highlights that there may be added benefit in targeting interventions to assist the poorest within other racial/ethnic groups, particularly other minorities, given their independently higher risk of STIs. This does not imply singling such individuals out for targeted prevention messages, but rather the importance of providing interventions relevant to such individuals, including consideration of structural interventions that lower such individuals' vulnerability to high-risk behaviours, partners and settings (Purcell & Mccree, 2009).

References

Adimora AA & Schoenbach VJ. Social Context, Sexual Networks, and Racial Disparities in Rates of Sexually Transmitted Infections. *Journal of Infectious Diseases*, 2005; 191(Suppl 1): S115-22.

Annang L, Walsemann KM, Maitra D & Kerr JC. Does Education Matter? Examining Racial Differences in the Association between Education and STI Diagnosis among Black and White Young Adult Females in the U.S. *Public Health Reports*, 2010; 125(Suppl 4): 110-21.

Aral SO. Sexual Network Patterns as Determinants of STD Rates: Paradigm Shift in the Behavioral Epidemiology of STDs Made Visible. *Sexually Transmitted Diseases*, 1999; 26(5): 262-4.

Aral SO, Fenton KA & Holmes KK. Sexually Transmitted Diseases in the USA: Temporal Trends. *Sexually Transmitted Infections*, 2007; 83(4): 257-66.

Atkinson AB, Rainwater L & Smeeding TM. *Income Distribution in OECD Countries: Evidence from the Luxembourg Income Study*; 1995. Paris: Organisation for Economic Co-operation and Development.

Biello KB, Kershaw T, Nelson R, Hogben M, Ickovics J & Niccolai L. Racial Residential Segregation and Rates of Gonorrhea in the United States, 2003-2007. *American Journal of Public Health*, 2012; 102(7): 1370-7.

Buffardi AL, Thomas KK, Holmes KK & Manhart LE. Moving Upstream: Ecosocial and Psychosocial Correlates of Sexually Transmitted Infections among Young Adults in the United States. *American Journal of Public Health*, 2008; 98(6): 1128-36.

Centers for Disease Control and Prevention. *Establishing a Holistic Framework to Reduce Inequities in HIV, Viral Hepatitis, STDs, and Tuberculosis in the United States*. 2010. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2010*. 2011. Atlanta, GA: U.S. Department of Health and Human Services.

Chantala K, Kalsbeek WD & Andraca E. *Non-Response in Wave III of the Add Health Study*. n.d. Available from: http://www.cpc.unc.edu/projects/addhealth/data/guides/W3nonres.pdf. [Last Accessed 1 October 2012].

Crepaz N, Marshall KJ, AuPont LW, Jacobs ED, Mizuno Y, Kay LS et al. The Efficacy of HIV/STI Behavioral Interventions for African American Females in the United States: A Meta-Analysis. *American Journal of Public Health*, 2009; 99(11): 2069-78.

Datta SD, Sternberg M, Johnson RE, Berman S, Papp JR, Mcquillan G et al. Gonorrhea and Chlamydia in the United States among Persons 14 to 39 Years of Age, 1999 to 2002. *Annals of Internal Medicine*, 2007; 147(2): 89-96.

Dolan C & Delcher C. Monitoring Health Inequities and Planning in Virginia: Poverty, Human Immunodeficiency Virus, and Sexually Transmitted Infections. *Sexually Transmitted Diseases*, 2008; 35(12): 981-4.

Ford JL & Browning CR. Neighborhood Social Disorganization and the Acquisition of Trichomoniasis among Young Adults in the United States. *American Journal of Public Health*, 2011; 101(9): 1696-703.

Hallfors DD, Iritani BJ, Miller WC & Bauer DJ. Sexual and Drug Behavior Patterns and HIV and STD Racial Disparities: The Need for New Directions. *American Journal of Public Health*, 2007; 97(1): 125-32.

Harris KM, Halpern CT, Whitsel E, Hussey J, Tabor J, Entzel P et al. *The National Longitudinal Study of Adolescent Health: Research Design*. 2009. Available from: http://www.cpc.unc.edu/projects/addhealth/design. [Last Accessed 1 October 2012].

Henny KD, Crepaz N, Lyles CM, Marshall KJ, AuPont LW, Jacobs ED et al. Efficacy of HIV/STI Behavioral Interventions for Heterosexual African American Men in

the United States: A Meta-Analysis. *AIDS and Behavior*, 2012; 16(5): 1092-114.

Hogben M & Leichliter JS. Social Determinants and Sexually Transmitted Disease Disparities. *Sexually Transmitted Diseases*, 2008; 35(12 Suppl): S13-8.

Jennings JM, Taylor R, Iannacchione VG, Rogers SM, Chung SE, Huettner S et al. The Available Pool of Sex Partners and Risk for a Current Bacterial Sexually Transmitted Infection. *Annals of Epidemiology*, 2010; 20(7): 532-8.

Jennings JM, Taylor RB, Salhi RA, Furr-Holden CD & Ellen JM. Neighborhood Drug Markets: A Risk Environment for Bacterial Sexually Transmitted Infections among Urban Youth. *Social Science and Medicine*, 2012; 74(8): 1240-50.

Kerani RP, Handcock MS, Handsfield HH & Holmes KK. Comparative Geographic Concentrations of 4 Sexually Transmitted Infections. *American Journal of Public Health*, 2005; 95(2): 324-30.

Krieger N, Waterman PD, Chen JT, Soobader M-J & Subramanian SV. Monitoring Socioeconomic Inequalities in Sexually Transmitted Infections, Tuberculosis, and Violence: Geocoding and Choice of Area-Based Socioeconomic Measures--the Public Health Disparities Geocoding Project (US). *Public Health Reports*, 2003; 118(3): 240-60.

Laumann EO & Youm Y. Racial/Ethnic Group Differences in the Prevalence of Sexually Transmitted Diseases in the United States: A Network Explanation. *Sexually Transmitted Diseases*, 1999; 26(5): 250-61.

LaVeist TA. Disentangling Race and Socioeconomic Status: A Key to Understanding Health Inequalities. *Journal of Urban Health*, 2005; 82(2 Suppl 3): iii26-34.

Mcquillan GM, Kruszon-Moran D, Kottiri BJ, Curtin LR, Lucas JW & Kington RS. Racial and Ethnic Differences in the Seroprevalence of 6 Infectious Diseases in the United States: Data from NHANES III, 1988-1994. *American Journal of Public Health*, 2004; 94(11): 1952-8.

Miller WC, Ford C, Morris M, Handcock MS, Schmitz JL, Hobbs MM et al. Prevalence of Chlamydial and Gonococcal Infections among Young Adults in the United States. *JAMA*, 2004; 291(18): 2229-36.

Miller WC, Swygard H, Hobbs MM, Ford C, Handcock MS, Morris M et al. The Prevalence of Trichomoniasis in Young Adults in the United States. *Sexually Transmitted Diseases*, 2005; 32(10): 593-8.

Nakashima AK, Rolfs RT, Flock ML, Kilmarx P & Greenspan JR. Epidemiology of Syphilis in the United States, 1941--1993. *Sexually Transmitted Diseases*, 1996; 23(1): 16-23.

Newbern EC, Miller WC, Schoenbach VJ & Kaufman JS. Family Socioeconomic Status and Self-Reported Sexually Transmitted Diseases among Black and White American Adolescents. *Sexually Transmitted Diseases*, 2004; 31(9): 533-41.

Purcell DW & Mccree DH. Recommendations from a Research Consultation to Address Intervention Strategies for HIV/AIDS Prevention Focused on African Americans. *American Journal of Public Health*, 2009; 99(11): 1937-40.

Sharpe TT, Voute C, Rose MA, Cleveland J, Dean HD & Fenton K. Social Determinants of HIV/AIDS and Sexually Transmitted Diseases among Black Women: Implications for Health Equity. *Journal of Women's Health*, 2012; 21(3): 249-54.

Springer YP, Samuel MC & Bolan G. Socioeconomic Gradients in Sexually Transmitted Diseases: A Geographic Information System-Based Analysis of Poverty, Race/Ethnicity, and Gonorrhea Rates in California, 2004-2006. *American Journal of Public Health*, 2010; 100(6): 1060-7.

The Add Health Biomarker Team. *Biomarkers in Wave III of the Add Health Study*. Available from: http://www.cpc.unc.edu/projects/addhealth/data/guides/biomark.pdf. [Last Accessed 1 October 2012].

Figure 1.1: Cumulative STI diagnosis rates (% and 95% confidence intervals), by race/ethnicity and income quintile, Waves II and III of Add Health

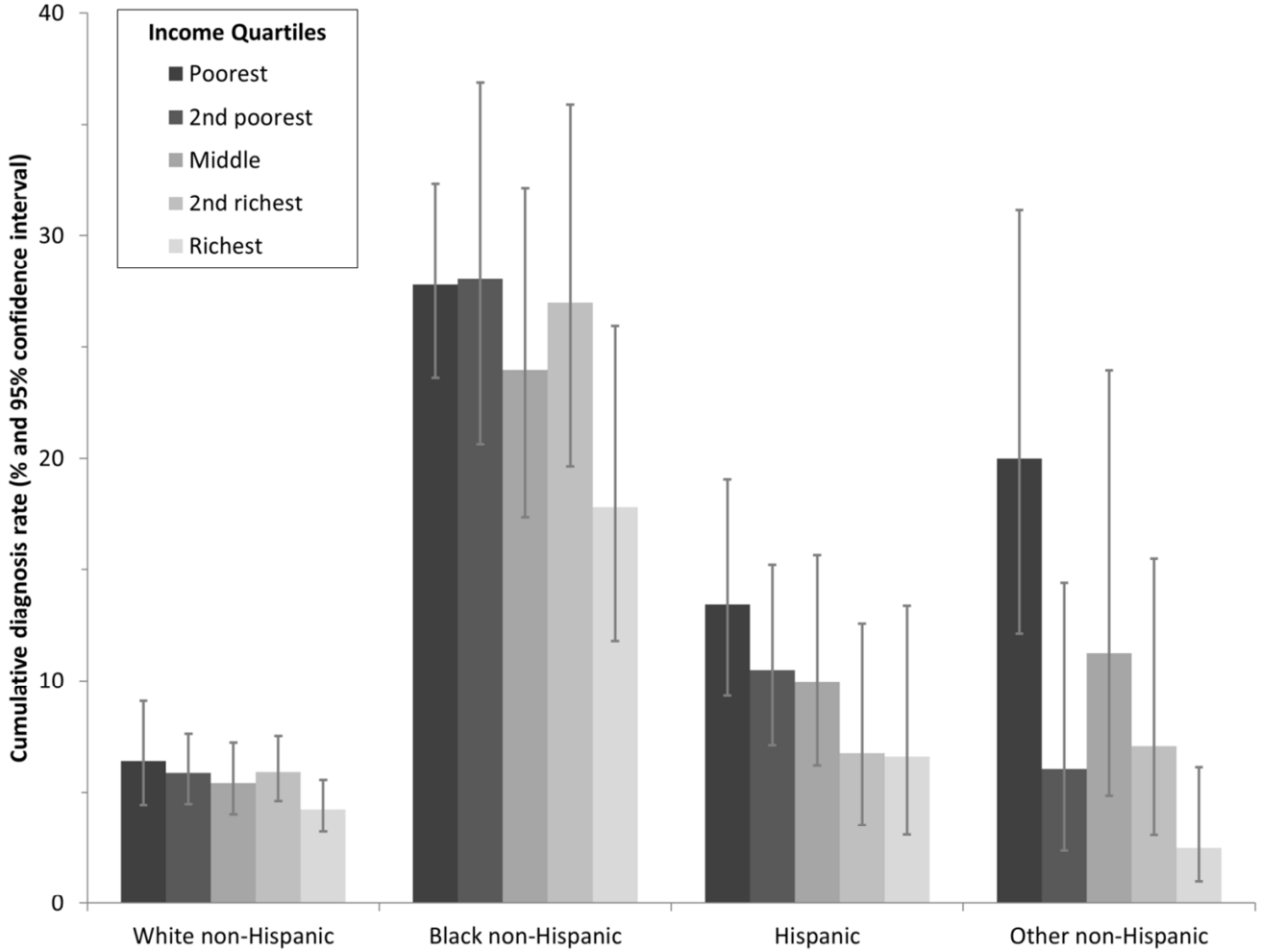

Table 1.1: Descriptive statistics by income quintile of a sample of respondents from Waves I to III of Add Health

	Poorest		2nd poorest		Middle		2nd richest		Richest		Total	
No. of respondents	2,107		2,154		2,195		2,119		2,216		10,791	
No. of STI diagnoses	14.7%	(1.4%)	10.6%	(1.2%)	8.3%	(0.9%)	7.4%	(0.7%)	5.2%	(0.6%)	9.2%	(0.6%)
Individual Race/Ethnicity												
White non-Hispanic	45.8%	(4.5%)	62.5%	(3.7%)	74.4%	(3.0%)	82.7%	(2.1%)	83.4%	(2.0%)	70.0%	(2.9%)
Black non-Hispanic	29.1%	(4.2%)	18.6%	(2.7%)	11.7%	(2.0%)	6.5%	(1.1%)	6.9%	(1.2%)	14.4%	(2.1%)
Hispanic	20.5%	(3.7%)	13.5%	(2.3%)	9.4%	(1.4%)	7.2%	(1.2%)	5.1%	(0.7%)	11.0%	(1.7%)
Other non-Hispanic	4.6%	(1.1%)	5.4%	(1.3%)	4.4%	(1.1%)	3.6%	(0.8%)	4.6%	(1.0%)	4.5%	(0.8%)
Sex												
Male	48.6%	(1.7%)	49.8%	(1.4%)	46.7%	(1.3%)	49.4%	(1.9%)	49.3%	(1.4%)	48.7%	(0.7%)
Female	51.4%	(1.7%)	50.2%	(1.4%)	53.3%	(1.3%)	50.6%	(1.9%)	50.7%	(1.4%)	51.3%	(0.7%)
Age at baseline												
<14	18.9%	(2.5%)	20.5%	(2.4%)	21.6%	(2.9%)	21.0%	(2.8%)	18.8%	(2.7%)	20.1%	(2.2%)
14	20.1%	(2.4%)	18.0%	(1.8%)	15.7%	(1.4%)	15.9%	(1.6%)	16.1%	(1.7%)	17.1%	(1.4%)
15	17.5%	(1.2%)	17.2%	(1.3%)	18.0%	(1.3%)	16.7%	(1.2%)	17.2%	(1.6%)	17.4%	(0.9%)
16	15.8%	(1.6%)	15.6%	(1.4%)	17.0%	(1.5%)	17.7%	(1.5%)	16.0%	(1.4%)	16.4%	(1.1%)
17	15.2%	(1.7%)	14.4%	(1.4%)	15.2%	(1.2%)	16.5%	(1.5%)	19.6%	(1.6%)	16.2%	(1.1%)
>17	12.5%	(1.7%)	14.4%	(1.5%)	12.5%	(1.3%)	12.2%	(1.1%)	12.3%	(1.1%)	12.8%	(1.0%)
Urbanicity												
Urban	33.1%	(5.6%)	28.3%	(4.6%)	23.7%	(4.0%)	21.3%	(3.7%)	22.6%	(4.5%)	25.7%	(3.9%)
Suburban	45.6%	(5.8%)	53.6%	(5.2%)	58.8%	(5.5%)	65.2%	(5.0%)	67.2%	(5.6%)	58.2%	(4.8%)
Rural	21.2%	(5.3%)	18.0%	(4.3%)	17.5%	(5.5%)	13.4%	(4.3%)	10.2%	(4.1%)	16.0%	(4.2%)
Region												
West	12.6%	(3.5%)	14.1%	(3.1%)	15.7%	(3.3%)	17.7%	(3.6%)	22.1%	(4.8%)	16.5%	(3.2%)
Midwest	28.9%	(5.7%)	31.3%	(5.2%)	35.0%	(5.9%)	31.5%	(5.6%)	33.3%	(6.8%)	32.0%	(5.1%)
South	47.6%	(5.9%)	41.6%	(5.0%)	35.4%	(4.8%)	34.5%	(4.9%)	27.5%	(4.7%)	37.1%	(4.5%)
Northeast	11.0%	(3.2%)	13.0%	(3.3%)	13.9%	(3.3%)	16.4%	(3.8%)	17.1%	(4.2%)	14.3%	(3.1%)
Type of school												
Public	98.1%	(1.0%)	95.0%	(1.9%)	93.2%	(2.3%)	91.8%	(2.9%)	88.4%	(3.8%)	93.3%	(2.0%)
Private	1.9%	(1.0%)	5.0%	(1.9%)	6.8%	(2.3%)	8.2%	(2.9%)	11.6%	(3.8%)	6.7%	(2.0%)
Per capita family income, equivalized scale $ Mean (95% confidence interval)	4,972	(4,783 - 5,160)	12,139	(12,009 - 12,270)	18,674	(18,543 - 18,805)	26,425	(26,255 - 26,595)	49,428	(46,068 - 52,787)	22,660	(20,972 - 24,348)

Percentages (and standard errors) are based on data weighted for non-random sampling and non-response.

Table 1.2: Multivariable regressions considering race/ethnicity and income as predictors of STI diagnosis at Waves II or III of Add Health

	Model 1: Race		Model 2: Race & Income		Model 3: Race & Income interaction†	
Individual Race/Ethnicity						
White non-Hispanic	1.00		1.00		1.00	
Black non-Hispanic	6.48	[5.16 - 8.14]	5.88	[4.68 - 7.37]	5.08	[2.93 - 8.80]
Hispanic	2.28	[1.64 - 3.17]	2.04	[1.47 - 2.83]	1.73	[0.74 - 4.07]
Other non-Hispanic	2.02	[1.28 - 3.19]	1.89	[1.21 - 2.96]	0.66	[0.30 - 1.48]
Per capita family income						
Poorest quintile			1.83	[1.38 - 2.42]		
2nd quintile			1.57	[1.15 - 2.16]		
Middle quintile			1.44	[1.05 - 1.99]		
3rd quintile			1.48	[1.10 - 2.00]		
Richest quintile			1.00			
Per capita family income for White non-Hispanics						
Poorest quintile					1.54	[1.00 - 2.37]
2nd quintile					1.40	[0.93 - 2.10]
Middle quintile					1.30	[0.86 - 1.98]
3rd quintile					1.43	[0.98 - 2.09]
Richest quintile					1.00	
Per capita family income for Black non-Hispanics						
Poorest quintile					1.83	[1.10 - 3.05]
2nd quintile					1.86	[0.93 - 3.70]
Middle quintile					1.52	[0.82 - 2.81]
3rd quintile					1.77	[0.97 - 3.22]
Richest quintile					1.00	
Per capita family income for non-Hispanics						
Poorest quintile					2.25	[0.97 - 5.22]
2nd quintile					1.69	[0.67 - 4.28]
Middle quintile					1.58	[0.65 - 3.84]
3rd quintile					1.05	[0.38 - 2.89]
Richest quintile					1.00	
Per capita family income for Other non-Hispanics						
Poorest quintile					8.97	[3.22 - 24.98]
2nd quintile					2.43	[0.85 - 6.94]
Middle quintile					4.95	[1.46 - 16.79]
3rd quintile					2.77	[0.85 - 9.08]
Richest quintile					1.00	
Akaike Information Criteria (AIC)	9,604,887		9,563,401		9,535,430	

N is 10,791 individuals from 132 schools for all regressions. Values are odds ratios and [95% confidence intervals].

All models shown are also adjusted for individual age in years and sex, and for school region, urbanicity and public/private school type.

† In Model 3 the top set of odds ratios are comparisons across Race/Ethnicity within the highest income quintile; the lower four sets of odds ratios are comparisons across Income quintiles within racial/ethnic categories.

Table 1.3: Multivariable logistic regressions for race/ethnicity, income and STI diagnosis in Add Health: sub-group analyses

	Primary analysis		Male		Female		Chlamydia		Gonorrhoea		Trichomoniasis	
White non-Hispanic												
Poorest quintile	1.54	[1.00 - 2.37]	0.71	[0.31 - 1.59]	1.08	[0.64 - 1.81]	1.79	[1.11 - 2.88]	2.35	[0.51 - 10.90]	0.99	[0.48 - 2.05]
2nd poorest quintile	1.40	[0.93 - 2.10]	0.96	[0.48 - 1.89]	1.07	[0.66 - 1.72]	1.46	[0.92 - 2.32]	4.33	[1.00 - 18.63]	1.27	[0.60 - 2.72]
Middle quintile	1.30	[0.86 - 1.98]	0.90	[0.49 - 1.66]	0.82	[0.46 - 1.45]	1.31	[0.82 - 2.08]	1.12	[0.25 - 4.97]	0.92	[0.42 - 2.04]
2nd richest quintile	1.43	[0.98 - 2.09]	0.91	[0.51 - 1.62]	1.07	[0.69 - 1.67]	1.42	[0.91 - 2.21]	2.37	[0.59 - 9.56]	0.85	[0.38 - 1.93]
Richest quintile	1.00		1.00		1.00		1.00		1.00		1.00	
Black non-Hispanic												
Poorest quintile	9.32	[6.70 - 12.96]	4.86	[2.85 - 8.28]	7.18	[4.75 - 10.84]	8.94	[5.87 - 13.63]	23.35	[7.07 - 77.17]	5.40	[3.20 - 9.09]
2nd poorest quintile	9.45	[5.77 - 15.48]	5.58	[2.53 - 12.31]	7.23	[4.42 - 11.83]	9.07	[5.28 - 15.58]	26.58	[7.08 - 99.78]	5.49	[2.89 - 10.44]
Middle quintile	7.71	[4.66 - 12.75]	5.85	[3.20 - 10.67]	5.35	[2.75 - 10.39]	7.52	[4.34 - 13.01]	21.38	[6.02 - 75.93]	6.94	[3.27 - 14.72]
2nd richest quintile	9.00	[5.71 - 14.21]	6.58	[3.36 - 12.91]	5.76	[2.72 - 12.19]	7.82	[4.55 - 13.44]	28.18	[5.77 - 137.70]	8.30	[3.60 - 19.15]
Richest quintile	5.08	[2.93 - 8.80]	3.17	[1.67 - 6.02]	2.68	[1.37 - 5.24]	4.74	[2.37 - 9.46]	14.57	[3.76 - 56.44]	3.34	[1.50 - 7.42]
Hispanic												
Poorest quintile	3.90	[2.24 - 6.76]	1.88	[0.87 - 4.05]	2.63	[1.26 - 5.47]	4.29	[2.28 - 8.09]	7.06	[1.48 - 33.66]	2.10	[0.99 - 4.48]
2nd poorest quintile	2.93	[1.72 - 5.01]	0.79	[0.27 - 2.32]	2.47	[1.17 - 5.20]	3.49	[1.96 - 6.23]	4.65	[0.98 - 22.00]	1.61	[0.42 - 6.25]
Middle quintile	2.73	[1.57 - 4.76]	3.14	[1.47 - 6.72]	1.55	[0.70 - 3.44]	3.14	[1.64 - 6.01]	1.19	[0.15 - 9.72]	1.30	[0.45 - 3.76]
2nd richest quintile	1.81	[0.85 - 3.87]	1.28	[0.54 - 3.04]	0.64	[0.22 - 1.84]	1.68	[0.64 - 4.37]	2.81	[0.32 - 24.51]	2.00	[0.74 - 5.39]
Richest quintile	1.73	[0.74 - 4.07]	1.06	[0.23 - 4.96]	1.18	[0.39 - 3.60]	1.26	[0.47 - 3.42]	-		2.24	[0.57 - 8.89]
Other non-Hispanic												
Poorest quintile	5.93	[2.90 - 12.14]	11.34	[4.41 - 29.18]	2.13	[0.62 - 7.27]	5.15	[2.02 - 13.10]	17.74	[2.17 - 144.87]	3.36	[1.00 - 11.34]
2nd poorest quintile	1.61	[0.57 - 4.53]	0.24	[0.04 - 1.50]	1.51	[0.58 - 3.91]	1.23	[0.32 - 4.66]	-		1.97	[0.37 - 10.53]
Middle quintile	3.27	[1.22 - 8.77]	1.50	[0.25 - 9.11]	3.08	[0.96 - 9.83]	2.69	[0.76 - 9.53]	11.72	[1.53 - 90.03]	1.63	[0.34 - 7.94]
2nd richest quintile	1.84	[0.72 - 4.71]	0.46	[0.14 - 1.55]	2.63	[0.84 - 8.21]	2.33	[0.87 - 6.26]	-		0.27	[0.04 - 1.89]
Richest quintile	0.66	[0.30 - 1.48]	0.30	[0.07 - 1.28]	0.68	[0.19 - 2.43]	0.62	[0.20 - 1.92]	-		0.61	[0.16 - 2.26]
No. of individuals	10,791		5,156		5,635		10,791		10,791		10,791	

Values are odds ratios and [95% confidence intervals]. All models are also adjusted for individual age in years and sex, and for school region, urbanicity and public/private school type.
All odds ratios are for comparisons with White non-Hispanics in the richest income quintile. Cells containing en-dashes represent categories with insufficient numbers of cases to allow estimation.

Supplementary Table 1.4: Cumulative STI diagnosis rates (% and 95% confidence intervals), by race/ethnicity and income quintile, Waves II and III of Add Health

	Poorest quintile		2nd poorest quintile		Middle quintile		2nd richest quintile		Richest quintile		Total	
White non-Hispanic	6.39	(4.45 - 9.08)	5.86	(4.48 - 7.64)	5.42	(4.04 - 7.25)	5.91	(4.62 - 7.53)	4.25	(3.24 - 5.55)	5.44	(4.78 - 6.19)
Black non-Hispanic	27.78	(23.62 - 32.36)	28.04	(20.64 - 36.87)	23.98	(17.35 - 32.16)	27.00	(19.63 - 35.89)	17.80	(11.80 - 25.96)	26.13	(22.95 - 29.58)
Hispanic	13.45	(9.33 - 19.02)	10.50	(7.12 - 15.21)	9.96	(6.20 - 15.63)	6.79	(3.56 - 12.57)	6.59	(3.11 - 13.41)	10.63	(8.51 - 13.20)
Other non-Hispanic	19.99	(12.12 - 31.16)	6.03	(2.39 - 14.41)	11.25	(4.85 - 23.95)	7.08	(3.08 - 15.48)	2.50	(1.00 - 6.11)	9.29	(6.43 - 13.26)
Total	14.69	(12.17 - 17.62)	10.63	(8.42 - 13.33)	8.28	(6.63 - 10.30)	7.38	(6.17 - 8.81)	5.23	(4.21 - 6.48)	9.18	(7.99 - 10.52)

95% confidence intervals are calculated based on binomial proportions using the adjusted Wilson score method.

Supplementary Table 1.5: Bivariate associations between independent predictors and STI diagnosis at Wave II or III of Add Health

	Odds Ratio	95% Confidence Interval
Individual Race/Ethnicity		
White non-Hispanic	1.00	
Black non-Hispanic	6.01	[5.14 - 7.02]
Hispanic	2.21	[1.74 - 2.81]
Other non-Hispanic	2.17	[1.60 - 2.94]
Per capita family income		
Poorest quintile	2.22	[1.76 - 2.80]
2nd poorest quintile	2.03	[1.61 - 2.55]
Middle quintile	1.41	[1.11 - 1.78]
2nd richest quintile	1.27	[0.99 - 1.62]
Richest quintile	1.00	
Sex		
Male	0.70	[0.62 - 0.80]
Female	1.00	
Age at baseline		
<14	0.99	[0.79 - 1.25]
14	1.02	[0.81 - 1.27]
15	0.92	[0.74 - 1.14]
16	1.00	
17	0.71	[0.56 - 0.89]
>17	0.72	[0.55 - 0.95]
Urbanicity		
Urban	1.00	
Suburban	0.74	[0.57 - 0.96]
Rural	0.62	[0.46 - 0.82]
Region		
West	0.67	[0.49 - 0.91]
Midwest	0.74	[0.57 - 0.97]
South	1.00	
Northeast	0.87	[0.60 - 1.26]
Type of school		
Public	1.00	
Private	0.90	[0.64 - 1.27]

N is 10,791 individuals from 132 schools for all regressions.

Supplementary Table 1.6: Multivariable logistic regressions for race/ethnicity and income and STI diagnosis in Add Health: Primary analysis including covariate values (odds ratios and 95% confidence intervals)

	Race only		Race and Income		Race & Income interaction	
No. of individuals	10,791		10,791		10,791	
Akaike Information Criteria	9,604,887		9,563,401		9,535,430	
Individual Race/Ethnicity						
White non-Hispanic	1.00		1.00		1.00	
Black non-Hispanic	6.48	[5.16 - 8.14]	5.88	[4.68 - 7.37]	5.08	[2.93 - 8.80]
Hispanic	2.28	[1.64 - 3.17]	2.04	[1.47 - 2.83]	1.73	[0.74 - 4.07]
Other non-Hispanic	2.02	[1.28 - 3.19]	1.89	[1.21 - 2.96]	0.66	[0.30 - 1.48]
Per capita family income						
Poorest quintile			1.83	[1.38 - 2.42]		
2nd poorest quintile			1.57	[1.15 - 2.16]		
Middle quintile			1.44	[1.05 - 1.99]		
2nd richest quintile			1.48	[1.10 - 2.00]		
Richest quintile			1.00			
Sex						
Male	0.62	[0.53 - 0.73]	0.62	[0.52 - 0.73]	0.62	[0.53 - 0.73]
Female	1.00		1.00		1.00	
Age at baseline						
<14	0.93	[0.68 - 1.27]	0.92	[0.68 - 1.26]	0.92	[0.67 - 1.26]
14	1.12	[0.79 - 1.61]	1.11	[0.78 - 1.58]	1.09	[0.77 - 1.53]
15	1.09	[0.79 - 1.50]	1.10	[0.80 - 1.52]	1.10	[0.80 - 1.52]
16	1.00		1.00		1.00	
17	0.91	[0.64 - 1.31]	0.93	[0.65 - 1.33]	0.92	[0.65 - 1.32]
>17	0.81	[0.60 - 1.10]	0.82	[0.60 - 1.11]	0.81	[0.60 - 1.11]
Urbanicity						
Urban	1.00		1.00		1.00	
Suburban	1.00	[0.76 - 1.30]	1.02	[0.79 - 1.32]	1.03	[0.79 - 1.34]
Rural	1.01	[0.75 - 1.37]	0.98	[0.73 - 1.31]	0.99	[0.74 - 1.33]
Region						
West	0.90	[0.71 - 1.14]	0.95	[0.76 - 1.20]	0.96	[0.77 - 1.20]
Midwest	1.24	[0.95 - 1.63]	1.27	[0.98 - 1.64]	1.27	[0.98 - 1.64]
South	1.00		1.00		1.00	
Northeast	1.02	[0.74 - 1.40]	1.05	[0.77 - 1.44]	1.07	[0.78 - 1.46]
Type of school						
Public	1.00		1.00		1.00	
Private	0.98	[0.73 - 1.31]	1.07	[0.81 - 1.40]	1.09	[0.83 - 1.42]
Per capita family income for White NH						
Poorest quintile					1.54	[1.00 - 2.37]
2nd poorest quintile					1.40	[0.93 - 2.10]
Middle quintile					1.30	[0.86 - 1.98]
2nd richest quintile					1.43	[0.98 - 2.09]
Richest quintile					1.00	
Per capita family income for Black NH						
Poorest quintile					1.83	[1.10 - 3.05]
2nd poorest quintile					1.86	[0.93 - 3.70]
Middle quintile					1.52	[0.82 - 2.81]
2nd richest quintile					1.77	[0.97 - 3.22]
Richest quintile					1.00	
Per capita family income for Hispanic						
Poorest quintile					2.25	[0.97 - 5.22]
2nd poorest quintile					1.69	[0.67 - 4.28]
Middle quintile					1.58	[0.65 - 3.84]
2nd richest quintile					1.05	[0.38 - 2.89]
Richest quintile					1.00	
Per capita family income for Other NH						
Poorest quintile					8.97	[3.22 - 24.98]
2nd poorest quintile					2.43	[0.85 - 6.94]
Middle quintile					4.95	[1.46 - 16.79]
2nd richest quintile					2.77	[0.85 - 9.08]
Richest quintile					1.00	

Supplementary Table 1.7: Multivariable logistic regressions for race/ethnicity, income and STI diagnosis in Add Health: within-income comparisons

	Primary analysis		Male		Female		Chlamydia		Gonorrhoea		Trichomoniasis	
White non-Hispanic												
Poorest quintile	1.54	[1.00 - 2.37]	0.71	[0.31 - 1.59]	1.08	[0.64 - 1.81]	1.79	[1.11 - 2.88]	2.35	[0.51 - 10.90]	0.99	[0.48 - 2.05]
2nd poorest quintile	1.40	[0.93 - 2.10]	0.96	[0.48 - 1.89]	1.07	[0.66 - 1.72]	1.46	[0.92 - 2.32]	4.33	[1.00 - 18.63]	1.27	[0.60 - 2.72]
Middle quintile	1.30	[0.86 - 1.98]	0.90	[0.49 - 1.66]	0.82	[0.46 - 1.45]	1.31	[0.82 - 2.08]	1.12	[0.25 - 4.97]	0.92	[0.42 - 2.04]
2nd richest quintile	1.43	[0.98 - 2.09]	0.91	[0.51 - 1.62]	1.07	[0.69 - 1.67]	1.42	[0.91 - 2.21]	2.37	[0.59 - 9.56]	0.85	[0.38 - 1.93]
Richest quintile	1.00		1.00		1.00		1.00		1.00		1.00	
Poorest quintile												
White non-Hispanic	1.00		1.00		1.00		1.00		1.00		1.00	
Black non-Hispanic	6.05	[3.96 - 9.24]	6.87	[2.88 - 16.41]	6.68	[3.80 - 11.72]	5.01	[3.04 - 8.27]	9.92	[3.78 - 26.08]	5.47	[2.73 - 10.95]
Hispanic	2.53	[1.32 - 4.84]	2.66	[0.92 - 7.69]	2.44	[1.02 - 5.86]	2.40	[1.15 - 5.05]	3.00	[0.75 - 12.00]	2.13	[0.87 - 5.21]
Other non-Hispanic	3.85	[1.75 - 8.47]	16.04	[4.71 - 54.61]	1.98	[0.57 - 6.89]	2.88	[1.07 - 7.78]	7.54	[0.99 - 57.29]	3.41	[0.96 - 12.06]
2nd poorest quintile												
White non-Hispanic	1.00		1.00		1.00		1.00		1.00		1.00	
Black non-Hispanic	6.75	[3.98 - 11.44]	5.84	[2.09 - 16.33]	6.79	[3.79 - 12.15]	6.22	[3.51 - 11.01]	6.15	[2.15 - 17.59]	4.31	[2.00 - 9.28]
Hispanic	2.10	[1.24 - 3.55]	0.82	[0.25 - 2.72]	2.32	[1.06 - 5.07]	2.39	[1.37 - 4.18]	1.07	[0.31 - 3.69]	1.27	[0.32 - 5.04]
Other non-Hispanic	1.15	[0.41 - 3.25]	0.25	[0.04 - 1.70]	1.41	[0.49 - 4.04]	0.84	[0.22 - 3.25]	0.00	[0.00 - 0.00]	1.55	[0.29 - 8.38]
Middle quintile												
White non-Hispanic	1.00		1.00		1.00		1.00		1.00		1.00	
Black non-Hispanic	5.92	[3.56 - 9.86]	6.46	[3.16 - 13.22]	6.51	[2.95 - 14.41]	5.75	[3.11 - 10.63]	19.03	[6.54 - 55.33]	7.54	[3.51 - 16.17]
Hispanic	2.10	[1.11 - 3.96]	3.47	[1.42 - 8.46]	1.89	[0.74 - 4.81]	2.40	[1.18 - 4.87]	1.06	[0.14 - 7.84]	1.41	[0.40 - 4.99]
Other non-Hispanic	2.52	[0.94 - 6.73]	1.65	[0.25 - 10.83]	3.75	[1.10 - 12.79]	2.06	[0.57 - 7.38]	-		1.77	[0.35 - 9.06]
2nd richest quintile												
White non-Hispanic	1.00		1.00		1.00		1.00		1.00		1.00	
Black non-Hispanic	6.30	[3.82 - 10.39]	7.23	[3.35 - 15.60]	5.36	[2.59 - 11.09]	5.52	[3.11 - 9.82]	11.89	[3.45 - 41.04]	9.72	[3.73 - 25.33]
Hispanic	1.27	[0.61 - 2.65]	1.41	[0.52 - 3.83]	0.60	[0.20 - 1.78]	1.18	[0.46 - 3.06]	1.19	[0.16 - 8.91]	2.34	[0.82 - 6.62]
Other non-Hispanic	1.28	[0.53 - 3.13]	0.50	[0.14 - 1.77]	2.44	[0.82 - 7.25]	1.64	[0.65 - 4.17]	-		0.31	[0.04 - 2.28]
Richest quintile												
White non-Hispanic	1.00		1.00		1.00		1.00		1.00		1.00	
Black non-Hispanic	5.08	[2.93 - 8.80]	3.17	[1.67 - 6.02]	2.68	[1.37 - 5.24]	4.74	[2.37 - 9.46]	14.57	[3.76 - 56.44]	3.34	[1.50 - 7.42]
Hispanic	1.73	[0.74 - 4.07]	1.06	[0.23 - 4.96]	1.18	[0.39 - 3.60]	1.26	[0.47 - 3.42]	-		2.24	[0.57 - 8.89]
Other non-Hispanic	0.66	[0.30 - 1.48]	0.30	[0.07 - 1.28]	0.68	[0.19 - 2.43]	0.62	[0.20 - 1.92]	-		0.61	[0.16 - 2.26]
No. of individuals	10,791		5,156		5,635		10,791		10,791		10,791	

Values are odds ratios and [95% confidence intervals]. All models are also adjusted for individual age in years and sex, and for school region, urbanicity and public/private school type.
The top set of odds ratios are comparisons across Income quintiles within White non-Hispanics; the lower five sets of odds ratios are comparisons across Race/Ethnicity within income quintiles.
Cells containing en-dashes represent categories with insufficient numbers of cases to allow estimation.

Supplementary Table 1.8: Multivariable logistic regressions for race/ethnicity, income and STI diagnosis in Add Health: within-race/ethnicity comparisons

	Primary analysis		Male		Female		Chlamydia		Gonorrhoea		Trichomoniasis	
Individual Race/Ethnicity												
White non-Hispanic	1.00		1.00		1.00		1.00		1.00		1.00	
Black non-Hispanic	5.08	[2.93 - 8.80]	3.17	[1.67 - 6.02]	2.68	[1.37 - 5.24]	4.74	[2.37 - 9.46]	14.57	[3.76 - 56.44]	3.34	[1.50 - 7.42]
Hispanic	1.73	[0.74 - 4.07]	1.06	[0.23 - 4.96]	1.18	[0.39 - 3.60]	1.26	[0.47 - 3.42]	2.81	[0.32 - 24.51]	2.24	[0.57 - 8.89]
Other non-Hispanic	0.66	[0.30 - 1.48]	0.30	[0.07 - 1.28]	0.68	[0.19 - 2.43]	0.62	[0.20 - 1.92]	11.72	[1.53 - 90.03]	0.61	[0.16 - 2.26]
Per capita family income												
White non-Hispanic												
Poorest quintile	1.54	[1.00 - 2.37]	0.71	[0.31 - 1.59]	1.08	[0.64 - 1.81]	1.79	[1.11 - 2.88]	2.35	[0.51 - 10.90]	0.99	[0.48 - 2.05]
2nd poorest quintile	1.40	[0.93 - 2.10]	0.96	[0.48 - 1.89]	1.07	[0.66 - 1.72]	1.46	[0.92 - 2.32]	4.33	[1.00 - 18.63]	1.27	[0.60 - 2.72]
Middle quintile	1.30	[0.86 - 1.98]	0.90	[0.49 - 1.66]	0.82	[0.46 - 1.45]	1.31	[0.82 - 2.08]	1.12	[0.25 - 4.97]	0.92	[0.42 - 2.04]
2nd richest quintile	1.43	[0.98 - 2.09]	0.91	[0.51 - 1.62]	1.07	[0.69 - 1.67]	1.42	[0.91 - 2.21]	2.37	[0.59 - 9.56]	0.85	[0.38 - 1.93]
Richest quintile	1.00		1.00		1.00		1.00		1.00		1.00	
Black non-Hispanic												
Poorest quintile	1.83	[1.10 - 3.05]	1.53	[0.74 - 3.18]	2.68	[1.48 - 4.85]	1.89	[1.02 - 3.49]	1.60	[0.79 - 3.25]	1.62	[0.84 - 3.12]
2nd poorest quintile	1.86	[0.93 - 3.70]	1.76	[0.66 - 4.71]	2.70	[1.37 - 5.30]	1.91	[0.86 - 4.24]	1.82	[0.75 - 4.43]	1.64	[0.70 - 3.84]
Middle quintile	1.52	[0.82 - 2.81]	1.85	[0.82 - 4.15]	1.99	[0.87 - 4.60]	1.59	[0.75 - 3.35]	1.47	[0.58 - 3.72]	2.08	[0.89 - 4.83]
2nd richest quintile	1.77	[0.97 - 3.22]	2.08	[0.82 - 5.29]	2.15	[0.88 - 5.24]	1.65	[0.83 - 3.28]	1.93	[0.60 - 6.19]	2.49	[0.97 - 6.35]
Richest quintile	1.00		1.00		1.00		1.00		1.00		1.00	
Hispanic												
Poorest quintile	2.25	[0.97 - 5.22]	1.77	[0.30 - 10.38]	2.23	[0.68 - 7.26]	3.40	[1.20 - 9.64]	2.51	[0.28 - 22.45]	0.94	[0.23 - 3.79]
2nd poorest quintile	1.69	[0.67 - 4.28]	0.74	[0.11 - 5.00]	2.10	[0.52 - 8.39]	2.76	[0.95 - 8.07]	1.65	[0.49 - 5.62]	0.72	[0.14 - 3.76]
Middle quintile	1.58	[0.65 - 3.84]	2.95	[0.62 - 14.08]	1.32	[0.35 - 5.02]	2.48	[0.79 - 7.77]	0.43	[0.03 - 5.83]	0.58	[0.10 - 3.37]
2nd richest quintile	1.05	[0.38 - 2.89]	1.20	[0.26 - 5.50]	0.54	[0.12 - 2.53]	1.33	[0.34 - 5.19]	1.00		0.89	[0.24 - 3.25]
Richest quintile	1.00		1.00		1.00		1.00		-		1.00	
Other non-Hispanic												
Poorest quintile	8.97	[3.22 - 24.98]	38.42	[7.29 - 202.51]	3.11	[0.54 - 18.02]	8.32	[2.05 - 33.78]	1.51	[0.15 - 15.32]	5.52	[1.05 - 28.94]
2nd poorest quintile	2.43	[0.85 - 6.94]	0.81	[0.07 - 8.78]	2.20	[0.70 - 6.93]	1.98	[0.63 - 6.29]	-		3.24	[0.46 - 22.64]
Middle quintile	4.95	[1.46 - 16.79]	5.07	[0.53 - 47.99]	4.50	[0.69 - 29.21]	4.34	[0.93 - 20.23]	1.00		2.68	[0.36 - 19.90]
2nd richest quintile	2.77	[0.85 - 9.08]	1.56	[0.31 - 7.85]	3.84	[0.57 - 25.84]	3.76	[0.82 - 17.27]	-		0.44	[0.05 - 4.32]
Richest quintile	1.00		1.00		1.00		1.00		-		1.00	
No. of individuals	10,791		5,156		5,635		10,791		10,791		10,791	

Values are odds ratios and [95% confidence intervals]. All models are also adjusted for individual age in years and sex, and for school region, urbanicity and public/private school type.
The top set of odds ratios are comparisons across Race/Ethnicity within the highest income quintile; the lower four sets of odds ratios are comparisons across Income quintiles within racial/ethnic categories.
Cells containing en-dashes represent categories with insufficient numbers of cases to allow estimation.

Supplementary Table 1.9: Multivariable logistic regressions for race/ethnicity and income and STI diagnosis in Add Health: robustness checks

	Primary analysis		Added Wave I outcomes		Restricted to those responding at all three waves		Restricted to data from Waves I & II only	
White non-Hispanic								
Poorest quintile	1.54	[1.00 - 2.37]	1.52	[1.03 - 2.24]	1.44	[0.92 - 2.25]	1.50	[0.60 - 3.78]
2nd poorest quintile	1.40	[0.93 - 2.10]	1.37	[0.95 - 1.99]	1.48	[0.94 - 2.31]	1.81	[0.80 - 4.10]
Middle quintile	1.30	[0.86 - 1.98]	1.29	[0.88 - 1.90]	1.37	[0.90 - 2.08]	0.78	[0.18 - 3.35]
2nd richest quintile	1.43	[0.98 - 2.09]	1.45	[1.01 - 2.10]	1.55	[1.03 - 2.32]	1.75	[0.79 - 3.89]
Richest quintile	1.00		1.00		1.00		1.00	
Black non-Hispanic								
Poorest quintile	9.32	[6.70 - 12.96]	9.27	[6.60 - 13.03]	9.27	[6.38 - 13.48]	11.05	[5.21 - 23.43]
2nd poorest quintile	9.45	[5.77 - 15.48]	9.49	[5.92 - 15.20]	10.34	[6.29 - 16.98]	5.82	[2.47 - 13.75]
Middle quintile	7.71	[4.66 - 12.75]	8.05	[4.70 - 13.78]	9.40	[5.33 - 16.57]	6.69	[2.58 - 17.37]
2nd richest quintile	9.00	[5.71 - 14.21]	8.86	[5.78 - 13.60]	9.29	[5.68 - 15.19]	6.12	[2.42 - 15.48]
Richest quintile	5.08	[2.93 - 8.80]	4.93	[2.90 - 8.37]	4.29	[2.44 - 7.56]	6.25	[1.81 - 21.60]
Hispanic								
Poorest quintile	3.90	[2.24 - 6.76]	3.74	[2.25 - 6.21]	3.75	[2.08 - 6.74]	2.54	[0.94 - 6.82]
2nd poorest quintile	2.93	[1.72 - 5.01]	2.63	[1.55 - 4.48]	2.90	[1.60 - 5.23]	1.93	[0.64 - 5.79]
Middle quintile	2.73	[1.57 - 4.76]	2.56	[1.49 - 4.42]	3.43	[1.87 - 6.29]	6.59	[2.64 - 16.47]
2nd richest quintile	1.81	[0.85 - 3.87]	1.69	[0.82 - 3.50]	1.67	[0.78 - 3.57]	1.97	[0.54 - 7.20]
Richest quintile	1.73	[0.74 - 4.07]	2.41	[1.11 - 5.25]	1.88	[0.74 - 4.76]	1.46	[0.34 - 6.33]
Other non-Hispanic								
Poorest quintile	5.93	[2.90 - 12.14]	5.33	[2.57 - 11.08]	7.08	[3.35 - 14.94]	8.58	[2.82 - 26.13]
2nd poorest quintile	1.61	[0.57 - 4.53]	1.43	[0.52 - 3.93]	1.07	[0.40 - 2.91]	0.10	[0.01 - 1.03]
Middle quintile	3.27	[1.22 - 8.77]	2.99	[1.11 - 8.08]	3.79	[1.36 - 10.62]	1.06	[0.18 - 6.40]
2nd richest quintile	1.84	[0.72 - 4.71]	1.66	[0.65 - 4.21]	2.46	[0.92 - 6.58]	3.07	[0.56 - 16.68]
Richest quintile	0.66	[0.30 - 1.48]	1.00	[0.41 - 2.43]	0.59	[0.19 - 1.82]	-	
No. of individuals	10,791		10,791		8,390		10,370	

Values are odds ratios and [95% confidence intervals]. All models are also adjusted for individual age in years and sex, and for school region, urbanicity and public/private school type. All odds ratios are for comparisons with White non-Hispanics in the richest income quintile. Cells containing en-dashes represent categories with insufficient numbers of cases to allow estimation.

Supplementary Table 1.10: A comparison of Add Health respondents at Wave II or III with Missing and Non-Missing family incomes at Wave I

	Poorest		2nd poorest		Middle		2nd richest		Richest		All non-Missing		Missing		χ^2 value	p-value
No. of respondents	2,107		2,154		2,195		2,119		2,216		10,791		3,517			
Individual																
Race/Ethnicity																
White non-Hispanic	45.8%	(4.5%)	62.5%	(3.7%)	74.4%	(3.0%)	82.7%	(2.1%)	83.4%	(2.0%)	70.0%	(2.9%)	56.9%	(3.8%)		
Black non-Hispanic	29.1%	(4.2%)	18.6%	(2.7%)	11.7%	(2.0%)	6.5%	(1.1%)	6.9%	(1.2%)	14.4%	(2.1%)	20.3%	(3.1%)		
Hispanic	20.5%	(3.7%)	13.5%	(2.3%)	9.4%	(1.4%)	7.2%	(1.2%)	5.1%	(0.7%)	11.0%	(1.7%)	14.6%	(2.5%)		
Other non-Hispanic	4.6%	(1.1%)	5.4%	(1.3%)	4.4%	(1.1%)	3.6%	(0.8%)	4.6%	(1.0%)	4.5%	(0.8%)	8.2%	(1.5%)	44.60	<.0001
Sex																
Male	48.6%	(1.7%)	49.8%	(1.4%)	46.7%	(1.3%)	49.4%	(1.9%)	49.3%	(1.4%)	48.7%	(0.7%)	50.6%	(1.2%)		
Female	51.4%	(1.7%)	50.2%	(1.4%)	53.3%	(1.3%)	50.6%	(1.9%)	50.7%	(1.4%)	51.3%	(0.7%)	49.4%	(1.2%)	1.74	0.187
Age at baseline																
<14	18.9%	(2.5%)	20.5%	(2.4%)	21.6%	(2.9%)	21.0%	(2.8%)	18.8%	(2.7%)	20.1%	(2.2%)	14.2%	(1.9%)		
14	20.1%	(2.4%)	18.0%	(1.8%)	15.7%	(1.4%)	15.9%	(1.6%)	16.1%	(1.7%)	17.1%	(1.4%)	13.5%	(1.4%)		
15	17.5%	(1.2%)	17.2%	(1.3%)	18.0%	(1.3%)	16.7%	(1.2%)	17.2%	(1.6%)	17.4%	(0.9%)	14.8%	(0.9%)		
16	15.8%	(1.6%)	15.6%	(1.4%)	17.0%	(1.5%)	17.7%	(1.5%)	16.0%	(1.4%)	16.4%	(1.1%)	16.5%	(1.2%)		
17	15.2%	(1.7%)	14.4%	(1.4%)	15.2%	(1.2%)	16.5%	(1.5%)	19.6%	(1.6%)	16.2%	(1.1%)	17.8%	(1.3%)		
>17	12.5%	(1.7%)	14.4%	(1.5%)	12.5%	(1.3%)	12.2%	(1.1%)	12.3%	(1.1%)	12.8%	(1.0%)	23.2%	(1.6%)	114.78	<.0001
Urbanicity																
Urban	33.1%	(5.6%)	28.3%	(4.6%)	23.7%	(4.0%)	21.3%	(3.7%)	22.6%	(4.5%)	25.7%	(3.9%)	30.4%	(4.8%)		
Suburban	45.6%	(5.8%)	53.6%	(5.2%)	58.8%	(5.5%)	65.2%	(5.0%)	67.2%	(5.6%)	58.2%	(4.8%)	55.7%	(5.2%)		
Rural	21.2%	(5.3%)	18.0%	(4.3%)	17.5%	(5.5%)	13.4%	(4.3%)	10.2%	(4.1%)	16.0%	(4.2%)	13.9%	(3.7%)	5.06	0.080
Region																
West	12.6%	(3.5%)	14.1%	(3.1%)	15.7%	(3.3%)	17.7%	(3.6%)	22.1%	(4.8%)	16.5%	(3.2%)	17.0%	(3.8%)		
Midwest	28.9%	(5.7%)	31.3%	(5.2%)	35.0%	(5.9%)	31.5%	(5.6%)	33.3%	(6.8%)	32.0%	(5.1%)	24.9%	(4.7%)		
South	47.6%	(5.9%)	41.6%	(5.0%)	35.4%	(4.8%)	34.5%	(4.9%)	27.5%	(4.7%)	37.1%	(4.5%)	45.9%	(5.3%)		
Northeast	23.1%	(3.2%)	31.3%	(3.3%)	39.4%	(3.3%)	47.5%	(3.8%)	62.1%	(4.2%)	14.3%	(3.1%)	26.7%	(3.2%)	13.37	0.004
Type of school																
Public	98.1%	(1.0%)	95.0%	(1.9%)	93.2%	(2.3%)	91.8%	(2.9%)	88.4%	(3.8%)	93.3%	(2.0%)	93.4%	(2.2%)		
Private	1.9%	(1.0%)	5.0%	(1.9%)	6.8%	(2.3%)	8.2%	(2.9%)	11.6%	(3.8%)	6.7%	(2.0%)	6.6%	(2.2%)	0.00	0.946
Proportion testing positive for any STI																
All respondents	14.7%	(1.4%)	10.6%	(1.2%)	8.3%	(0.9%)	7.4%	(0.7%)	5.2%	(0.6%)	9.2%	(0.6%)	10.8%	(1.0%)	3.55	0.059
White non-Hispanic	6.4%	(1.2%)	5.9%	(0.8%)	5.4%	(0.8%)	5.9%	(0.7%)	4.3%	(0.6%)	5.4%	(0.4%)	5.5%	(0.7%)	0.41	0.524
Black non-Hispanic	27.8%	(2.2%)	28%	(4.1%)	24%	(3.8%)	27%	(4.1%)	17.8%	(3.6%)	26.1%	(1.7%)	28.4%	(2.5%)	0.15	0.697
Hispanic	13.5%	(2.4%)	10.5%	(2.0%)	10%	(2.4%)	6.8%	(2.2%)	6.6%	(2.5%)	10.6%	(1.2%)	8.3%	(1.6%)	3.32	0.069
Other non-Hispanic	20%	(4.8%)	6%	(2.8%)	11.3%	(4.7%)	7.1%	(3.0%)	2.5%	(0.9%)	9.3%	(1.7%)	8.7%	(2.3%)	0.87	0.352

Percentages (and Standard Errors) are based on data weighted for non-random sampling and non-response.
The χ^2 tests have k-1 degrees of freedom and are for comparisons of all respondents with non-missing values to those with missing values.

Evaluating alternative explanations for an association between income inequality and sexually transmitted infections in the United States

Guy Harling, SV Subramanian, Till Bärnighausen, Ichiro Kawachi

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

Abstract

Three causal processes have been proposed to explain associations between group income inequality and individual health outcomes. We present a novel conceptual and analytic framework for the quantitative evaluation of these pathways, assessing the contribution of: (i) absolute deprivation – affecting the poor in all settings – using family income; (ii) structural inequality – affecting all those in unequal settings – using the Gini coefficient; and (iii) relative deprivation – affecting the poor in unequal settings – using the Yitzhaki index. We conceptualize relative deprivation as the interaction of the other two measures. We test our model using hierarchical models of 11,183 individuals in the National Longitudinal Study of Adolescent Health (Add Health) to examine school-level inequality and sexually transmitted infections (STI) – self-reported or laboratory-confirmed Chlamydia, Gonorrhoea or Trichomoniasis. Results suggest that increased poverty and inequality were both independently associated with STI diagnosis, and that being poor in an unequal community imposes a small additional risk.

Background

Theoretical explanations for a relationship between inequality and health

There is a growing body of evidence that higher levels of economic inequality both across and within countries are associated with worse health outcomes (Kondo et al., 2009; Wilkinson & Pickett, 2006). Multiple causal processes relating economic inequality to health have been proposed (Kawachi, 2000), but there is considerable debate as to whether these relationships are truly causal, and if so how the mechanisms might vary by health outcome (Deaton, 2003; Gravelle, 1998). The causal mechanisms which might explain an empirical association between inequality and health have been divided into three broad categories (Leigh et al., 2009; Nilsson, 2009).

First, the absolute deprivation hypothesis (ADH) posits that inequality is associated with ill-health for the poorest through its effect on the distribution of income in a community. This arises because increased inequality implies (ceteris paribus) more individuals living at lower levels of income, due to the redistribution of resources from those on the lower rungs of society to the better-off. Given the empirically observed concave relationship between resources and health – i.e. positive but with diminishing marginal returns – increased inequality at any given mean level of income leads to lower average health in a community, since the disadvantaged will lose more than the advantaged will gain (Leigh et al., 2009).

Second, the inequality hypothesis (IH) focuses on the idea that the structure of unequal societies harms the health of everyone within them. Higher heterogeneity in circumstances often causes weaker social bonds and less cohesion; this may have several effects. First, it can lead to lower levels of public good provision, either due to low cohesion limiting group efforts to secure such goods, or because community members have less in common thus lowering the likelihood of a majority supporting the provision of any given good (Kawachi & Berkman, 2000; Leigh et al., 2009). Furthermore, increasing inequality in society shifts the holder of the median unit of power or income up the socioeconomic distribution. As a result, such a person will be less inclined to support the provision of public goods – since they will have less in common with the average member of society and more resources to obtain such services privately – which is likely to lead to reduced overall provision.

Heterogeneity in economic circumstances may reduce interaction between community members. This can affect health by reducing the diffusion of healthy behaviours by limiting informal social control of unhealthy behaviours or by generating distrust leading increased anxiety or depression – each of which can lead to poorer physiological and behavioural outcomes (Kawachi & Berkman, 2000; Kubzansky & Kawachi, 2000). For example, if lower social bonds lead to increased propensity to commit crime against others, this might lower physical mobility within neighbourhoods, increase stress levels and of course increase violence-related ill-health.

More community heterogeneity may also, paradoxically, ensure increased mixing by socioeconomic status (SES) compared to more homogenous communities – by increasing the proportion of unlike people who are socially or geographically proximate. This may lead to increased diffusion of behaviours or infections, if disease prevalence or health behaviours differ systematically by SES. The direction of effect in such a situation is ambiguous, since those with poorer health might impact the healthier, or vice versa. In the case of infectious diseases, mixing of heterogeneous risk groups has been shown theoretically to lead to slower disease spread but ultimately higher total prevalence (Doherty et al., 2006; Garnett & Anderson, 1996).

Finally, the relative deprivation hypothesis (RDH) suggests that inequality affects the worst-off in unequal communities uniquely, by placing them at an increased distance from the standard of living of their relevant reference group – i.e. others living in social or geographical proximity (Spriggs et al., 2009). This reference group provides an expectation for normative living standards and behaviours (Runciman, 1966; Webber, 2007). When the worse-off are unable to achieve this standard of living due to limited resources, the resulting stress and shame may lead to worse health through either psychosocial or behavioural pathways (Kondo et al., 2008).

It is notable that each hypothesis above makes a different prediction regarding who should be affected by higher/lower levels of deprivation (see Figure 2.1). If the ADH holds we expect the poor to be at increased risk of ill-health relative to the rich, regardless of the level of community inequality. If the IH holds we expect those

living in more unequal communities to be at increased risk for disease acquisition, relative to those living in more equal communities. Finally, if the RDH holds we expect poorer members of more unequal communities to be at increased risk, relative to both the richer members of unequal communities and to poorer individuals in more equal communities. It is important to note that some, all, or none of these three hypotheses may hold for any given context or any disease outcome.

Linking income inequality to sexually transmitted infections

Theoretically, SES can pattern sexually transmitted infection (STI) risk in multiple ways. These include affecting whom one selects as a partner, and thus how likely the partner is to be infectious, how many partners one has and the actions an individual takes within a relationship (Bärnighausen & Tanser, 2009; Boerma & Weir, 2005; Poundstone et al., 2004). Empirically, sexual behaviour and STI rates have been shown to be associated with socioeconomic conditions at both the individual and group level (Eaton et al., 2003; Parker et al., 2000; Rotheram-Borus et al., 2009).

Bacterial STIs and HIV have long been empirically associated with poverty in the United States (US), particularly amongst women (Zierler & Krieger, 1997); this effect is intensified in the African-American community (Adimora & Schoenbach, 2002). Furthermore, a range of group-level measures of SES – including neighbourhood poverty, general disadvantage/deprivation, low social capital, low

collective efficacy and social disorganization – have been found to be associated with higher-risk sexual behaviours and STIs in North America and Europe (Bauermeister et al., 2011; Browning et al., 2004; Cohen et al., 2000; Dupéré et al., 2008; Henderson et al., 2008; Holtgrave & Crosby, 2004; Krieger et al., 2003; Monteiro et al., 2005; Ramirez-Valles et al., 1998; Roche & Leventhal, 2009).

Research on the relationship between economic inequality and STIs in the US has been limited. To date the only two analyses have been ecological: one of neighbourhood-level inequality in Massachusetts and Rhode Island (Krieger et al., 2003); the other of state-level inequality nationally (Holtgrave & Crosby, 2004). Both studies found inequality to be positively associated with reported STI rates. Internationally, income inequality has been found to be associated with HIV in men in India (Perkins et al., 2009), but not in Malawi (Feldacker et al., 2010).

It is possible to envisage pathways specifically from inequality to STIs for each of the three categories of mechanism. In line with the ADH, we might expect a link between income and STIs to arise either because income levels determine with whom one partners (for example, because it affects where one can afford to live and socialize), or because of the level of sexual health education and care one is able to access varies by income. If both these mechanisms act in concert, differential distribution of knowledge and healthcare resources across social strata would lead to the reinforcement of high STI rates amongst the poorest. IH mechanisms may

include reduced provision of sexual health services – perhaps because there is inadequate funding to support their provision – or changes in sexual mixing patterns, e.g. increased mixing between high and low SES groups in more unequal communities in the context of an existing negative SES gradient in STI rates leading to higher STI rates for the richest. RDH mechanisms linking relative deprivation to STIs seem most likely to arise through behavioural mechanisms, for example if the perceived gap between individuals and their better-off neighbours lead to increased use of alcohol and other substances, leading in turn to more risky sexual behaviour.

In the present study we sought to test out a method for examining how the three above mechanisms might explain any association between economic inequality and STI acquisition risk amongst young adults in the US. To do this we propose a novel approach linking commonly used economic measures to specific theoretical causal mechanisms.

Data and Methods

We conducted a secondary data analysis using Waves I to III of the National Longitudinal Survey of Adolescent Health (Add Health). Add Health is a nationwide survey which sampled adolescents from 80 US high schools and 52 of these schools' largest feeder schools (Harris et al., 2009). Schools were selected so as to ensure coverage across regions, levels of urbanicity, school sizes and types, and race/ethnicity. Wave I (1994-95) surveyed a sample of all enrolled students in

grades 7 through 12 at home, Wave II (1996) re-surveyed those who had been in grades 7 through 11 at Wave I, and Wave III (2001-02; ages 18-26) re-interviewed all Wave I respondents. Understanding sexual behaviour and health was one of the primary interests in the design of Add Health (Resnick et al., 1997).

The study sample for this analysis comprised all respondents who were at minimum interviewed at Waves I and III, who provided information on their age and sex, whose parents provided information on family income, household size and parental education, and who were affiliated with one of the 132 core Add Health schools at Wave I interview. Ethical approval for the original Add Health study and each subsequent wave was obtained from the Institutional Review Board (IRB) at the University of North Carolina, Chapel Hill. This analysis was granted an exemption by the IRB of the Harvard School of Public Health since it was a secondary analysis of existing data.

Measures

The primary outcome for this study was a binary measure of whether respondents had any self-reported or laboratory-confirmed STI at Wave II or Wave III. At Wave III, respondents were asked to provide a urine sample for *Chlamydia trachomatis*, *Neisseria Gonorrhoeae* and *Trichomonas vaginalis* testing. Also at Wave III, respondents were asked whether a health professional had within the past 12 months told them that they were infected with each of these STIs. At Wave II they

were asked whether they had been diagnosed since Wave I and at Wave I they were asked if they had ever been diagnosed.

Three economic exposure measures were used for this analysis, each reflecting one of the three hypothesised effects of income inequality on STI acquisition risk. All were based on parental reports of 1994 total pre-tax family income at Wave I (in $1000 increments, top-coded at $999,000). To arrive at a per-capita equivalent figure that reflected household economies of scale, the family figure was divided by the square root of the number of individuals in the household – the "Luxembourg Income Study Scale" approach (Atkinson et al., 1995).

Absolute deprivation for the ADH was measured using each respondent's family per-capita equivalent income. Structural inequality for the IH was measured using the Gini coefficient of per-capita equivalent income for all respondents within the same school, reflecting overall inequality within the respondent's community. The Gini coefficient is a measure of the gap between each person's income and that of each other person in their community, standardized to a value between 0 – everyone has the same income – and 1 – one person has all the income (Cowell, 2011).

Relative deprivation for the RDH was measured using the Yitzhaki index (Yitzhaki, 1979), reflecting the difference between an individual's circumstances and the normative level in their community. The Yitzhaki index defines relative deprivation for an individual as the sum of the income gaps between them and all those ranked

above them in their reference group, normalized by the size of the reference group. In this analysis, this meant summing the differences in income between the respondent and all other respondents in the same school with a higher per-capita income, and then dividing this total by the number of respondents in the school.

We considered as potential confounders: respondents' sex, age (in years) at Wave I, primary self-reported racial/ethnic identification (White non-Hispanic, Black non-Hispanic, Hispanic, and all others; hereafter: White, Black, Hispanic, Other); their parent's highest level of education (less than High School completion, High School completion or GED, some tertiary, completed college, any postgraduate); and their school's urbanicity (urban, suburban or rural), regional location (West, Midwest, South, North-East) and type (public or private).

Analytic methodology

Analyses were conducted using two-level hierarchical models where each individual *i* was nested within the school *j* which they were attending at Wave I. Bivariate relationships between the outcomes and each independent variable were examined using logistic regression to generate odds ratios (OR) and 95% confidence intervals (CI). After running an empty model, we re-established the bivariate relationship between income inequality and the primary outcome. Next we added absolute income and relative deprivation variables to quantify how much of the inequality-STI relationship might be attributable to each mechanism. We then added covariates for own age and sex, parental education and school-level urbanicity, region and

public/private type. Finally we added own race/ethnicity to the model, which led to final model of the form:

$$y_{ij} \sim Binomial\left(n_{ij}, p_{ij}\right)$$

$$ln\left[\frac{p_{ij}}{1-p_{ij}}\right] = \beta_{0j} + \beta_2 Income_{ij} + \beta_3 Yitzhaki_{ij} + \beta_4 Age_{ij} + \beta_5 Sex_{ij} + \beta_6 ParentalEducation_{ij} + \beta_7 Race/Ethnicity_{ij}$$

$$\beta_{0j} = \beta_0 + \beta_1 Gini_j + \beta_8 Urbanicity_j + \beta_9 Region_j + \beta_{10} SchoolType_j + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2) \qquad e_{ij} z_{ij}\,, z_{ij} = \sqrt{\frac{\hat{p}_{ij}\left(1-\hat{p}_{ij}\right)}{n_{ij}}}, \sigma_e^2 = {\pi^2}/{3}$$

Analyses were conducted using SAS version 9.1.3 (SAS Institute; Cary, NC), using the NLMixed procedure for hierarchical models. All statistical tests were two-sided at α=0.05. The reference category for each economic variable was the quintile we expected *a priori* to be lowest risk (low inequality, high absolute deprivation, low relative deprivation); for covariates we used either the most common category or the one expected to be lowest risk.

As extensions we reran the models: (i) stratified by sex; (ii) stratified by race/ethnicity; and (iii) considering each STI separately as an outcome. We also conducted three robustness analyses: first, given a low likelihood of reverse

causation, we included as cases any individuals who reported having been diagnosed with an STI at any date prior to their Wave I interview; second, in case those individuals responding at Wave II systematically differed from the overall sample, we removed those respondents who were not interviewed at all three waves; and third, since the impact of school might be expected to exert its strongest effect during adolescence, we restricted the analysis to self-reported STI diagnosis at Waves I or II.

Results

14,808 respondents were interviewed at both Waves I and III, were affiliated with one of the 132 core schools and provided information on age and sex. Family income or household size information was missing on 3,596 of these respondents and 2,052 were missing parental interview information on education (all but 31 of these last were also missing information on income). Of the remaining 11,185 respondents, a further two respondents declined to respond to STI history questions and did not provide a valid urine sample. The analytic sample size was thus 11,183.

Descriptive statistics for all variables for the primary analysis are provided in Table 2.1. Respondents were almost all aged between 13 and 18 at baseline (3.8% were aged 11 or 12, 1.4% aged 19-21) with a roughly even gender split. More than 55% were White, approximately 20% were Black and around 15% Hispanic; respondents were markedly more White and less Black and Hispanic than the nation as a whole

and the schools from which they were drawn – reflecting the purposive oversampling of racial/ethnic minorities in Add Health.

Almost 90% of respondents had at least one parent who had completed high school, and one-third had a parent who had completed a college degree. Median adjusted per-capita income was a little below $20,000 with a wide range. The median family had an average per-capita income gap from themselves to families above them at the same school of almost $6,400 and the median school Gini coefficient in the sample was 33.7, somewhat lower than the national value based on raw household income data. Figures for males and females were very similar to one-another. Across Waves II and III, 10.5% of respondents either reported a recent diagnosis of, or tested positive at Wave III for, at least one of the STIs of interest. The most common diagnosis was of Chlamydia (7.8%), followed by Trichomoniasis (3.0%) and then Gonorrhoea (1.6%).

All three measures of economic wellbeing were inversely associated with STI diagnosis in bivariate analysis (first column, Table 2.2). Individuals in the most unequal quintile had 2.5 times the odds of an STI diagnosis compared to those in the least unequal quintile, and there is significantly increased risk for those in the three most unequal quintiles. Similarly, the odds of STI infection for the poorest quintile was double that of the richest, and those who were most relatively deprived had 50% increased odds of STI compared to the least relatively deprived. (Supplementary Table 2.5 provides details of the bivariate associations between STIs and covariates.)

Adding income to the bivariate income inequality model (Model 2, Table 2.2) lead to an attenuation of the inequality relationship by approximately one-third but had little impact on the income-STI relationship. Further adding the Yitzhaki index to this model (Model 3, Table 2.2) had little impact on the inequality relationship; relative deprivation (in a model already adjusted for income and inequality) was associated with lower STI risk. This latter result amounts to a finding that both being poor and living in an unequal community was somewhat less risky than the combination of individual income and inequality effects would suggest (i.e. negative multiplicative interaction). Relationships between economic variables and STIs were essentially unchanged by adjustment for baseline age and sex, parental education and school characteristics (Model 4, Table 2.2). However, adjustment for individual race/ethnicity greatly attenuated the inequality relationship (Model 5, Table 2.2).

Stratifying the analysis by sex showed that the gradient in inequality was slightly steeper for women than for men, both in bivariate analysis and once income and relative deprivation were added to the model (Table 2.3). Stratification by race/ethnicity suggested that most of the unadjusted association between income inequality and STI risk is to be found amongst Hispanics and Others, although these associations had limited power due to smaller sample sizes and case counts (Table 2.4). Low income appeared to be associated with STI risk in all groups except Hispanics, and uniquely in this analysis Others exhibit a large, but non-significant,

positive adjusted effect for relative deprivation – suggesting positive multiplicative interaction between income and inequality in this group.

Modelling each STI outcome separately suggested that income inequality was most strongly associated with higher rates of Trichomoniasis, although power to detect effects was low (Supplementary Table 2.7). The first two robustness checks – adding Wave I outcomes and excluding individuals missing Wave II interviews – had a negligible effect on inequality and STI relationship (Supplementary Table 2.8). However, restricting the analysis to Waves I and II exposures and outcomes changed the results somewhat: although the number of cases was considerably lower, income inequality was more strongly associated with STIs, even in models containing race/ethnicity (2nd most unequal quintile: OR: 1.65, 95%CI: 1.01-2.69; most unequal quintile: OR: 1.48, 95%CI: 0.88-2.49). It was also notable that age was a strong predictor in this adolescent sample and that racial/ethnic differentials were attenuated.

Discussion

This study proposes a novel analytic framework for understanding the relationship between income inequality and health, and applies it to data on STI incidence amongst adolescents and young adults in the US. Our framework explicitly connects three sets of theoretical inequality-related causal mechanisms (absolute income deprivation; structural inequality; relative deprivation) to three economic measures

(personal income; community Gini coefficient; personal Yitzhaki index) to empirically test which sections of the population are affected by inequality (the poor; all in unequal settings; the poor in unequal settings). We find our proposed framework is feasible for examining STI diagnosis in the setting of school-based communities in the US.

Examining inequality

In bivariate analysis we find more income inequality to be strongly associated with higher risk of STI diagnosis amongst US youth. This is consistent with two previous ecological studies of income inequality and STIs in the US, both of which found strong positive associations between group-level inequality and group-level STI rates (Holtgrave & Crosby, 2004; Krieger et al., 2003).

Two arguments have been proposed to suggest that any observed inequality-health relationship does not reflect the structural effect of inequality on health, but rather relates to other economic factors. First, the ADH suggests that individuals' incomes confound the relationship. In this study, adding income to a model of inequality and STI risk led to a partial attenuation of the bivariate inequality association, but a significant relationship remained. Our findings suggest that while some of the inequality association may be explicable by income, and while income is independently associated with STI risk in these data, those living in more unequal settings have increased risk of STI, adjusting for income level. Second, the RDH suggests that relative deprivation may mediate the effect of inequality on health, by

generating stress and thus affecting physiology or behaviour. This does not appear to be the case in these data: although in bivariate analysis relative deprivation is positively associated with STI risk, adding the Yitzhaki index to the bivariate Gini-STI model has almost no impact on the inequality-STI model (Supplementary Table 2.9).

These analyses suggest that the crude relationship seen between the Gini coefficient and STIs partly reflect associations arising from the ADH and IH, but not the RDH. While such findings are preliminary, they might point us towards further investigation of inequality-STI pathways that relate to absolute deprivation and structural inequality. For example, the ADH suggests that those with fewer resources will have more risky sexual behaviours and potentially less health knowledge, while the IH suggests that more unequal communities will have higher levels of mixing across socioeconomic strata, or have lower levels of preventative care provision. These are testable hypotheses. The limited finding with regards to the RDH suggests that social comparison factors may play a relatively small role in determining risk for STIs amongst young US adults and, in combination with the finding of a robust income effect, suggests that material concerns might play a stronger role than psychosocial ones in this setting.

Once all three economic variables are included in the same model (Model 3, Table 2.2) it becomes clear that the poorest individuals in the most unequal schools were at increased risk for STI acquisition. Less clear in the table, however, is that those living with high levels of relative deprivation are probably at even higher levels of

risk than those who were either just poor or just living in unequal communities, even though relative deprivation appears to be protective (adjusted OR comparing highest to lowest Yitzhaki quintile: 0.70, 95%CI: 0.51 - 0.96).

This paradox arises because those living with high levels of relative deprivation are mechanistically likely to also be poor and living in unequal settings, since in order to be relatively deprived they must by definition have richer schoolmates whose family incomes is considerably greater than their own. This can be seen visually in Figure 2.1, which highlights that in order to compare someone who is relatively deprived to someone who is not, we must also adjust for absolute income and group income inequality: relative deprivation is conceptually similar to an interaction term for income and income inequality. When we allow for this, for example by comparing someone who is in the least deprived quintile for all three economic measures (i.e. richest, least relatively deprived, attending the most egalitarian school) to someone in the most deprived quintile for all measures, we find they are at 3.60 (95% CI: 2.48-5.22) times the odds of STI diagnosis – worse than those who are just poor (OR: 2.58, 95% CI: 1.88-3.56) or only attending unequal schools (OR: 1.94, 95% CI: 1.47-2.69). This difference is significant when comparing rich and poor within unequal settings (OR: 1.81, 95 % CI: 1.42-2.30), but not when comparing the poor located in more vs. less equal communities (OR: 1.39; 95% CI: 0.87-2.23). This implication of this is that while poverty matters for everyone, inequality only has a significant effect on risk amongst richer quintiles.

While the interaction of inequality and income can be included mechanically, the use of a measure of relative deprivation, such as the Yitzhaki index, provides two benefits. First, the Yitzhaki index has a clear interpretation as the average economic distance of an individual from those above him or herself in a community. This should help in translating findings to community and policy arenas. Second, in an analysis which uses more nuanced measures of income and inequality (i.e. not continuous or binary), such as this one, the use of a Yitzhaki index reduces the number of model terms (in this case we use 12 – four for each economic model; an interaction model requires 24), increasing the power to detect effects. When we ran models including fully interacted income and inequality quintiles the results were unstable; however they gave qualitatively similar results, with the largest odds ratios for those who were simultaneously in the poorest and most unequal quintiles (not shown).

The role of race/ethnicity

The addition of an individual's race/ethnicity as a potential confounder to our analysis consistently reduced the relationship between inequality and STI risk to negligible levels (OR comparing most to least unequal quintile in the bivariate model with race/ethnicity added: 1.12, 95%CI: 0.86-1.45; Supplementary Table 2.9). This was less markedly the case for models that included only infections diagnosed prior to leaving school (Supplementary Table 2.8). This null finding is in line with other research on school-level income inequality in the Add Health dataset – previous studies have found changes in inequality to be weakly linked to change in

racial test-score gaps (Campbell et al., 2008), and depressive symptoms to be (negatively) associated with school income equality only in models excluding individual-level covariates (Goodman et al., 2003).

The role of race/ethnicity as a potential confounder of an observed crude relationship between inequality and health has been extensively debated in the literature (Deaton & Lubotsky, 2003; Lynch et al., 2003; Mellor & Milyo, 2001; Subramanian & Kawachi, 2003). African-Americans are at very substantially elevated risk for STI acquisition in the US, as a result of assortative partnering by race/ethnicity and relatively disassortative partnering by behavioural risk (Adimora & Schoenbach, 2005; Aral, 1999). There is not strong evidence that Blacks in the US live in more unequal (as opposed to poorer or more segregated) communities, however our findings suggest that race/ethnicity may indeed confound the inequality and STI relationship in the US. We note that in this dataset, adding school-level racial/ethnic composition had almost no effect on the results reported, over and above the role of individual-level race/ethnicity (not shown).

Subgroup analyses

While inequality was not associated with STIs across the whole population after adjusting for race/ethnicity, notably stronger associations were seen: (i) for women, compared to men; (ii) for Hispanics and Others, compared to White and Blacks; (iii) for Trichomoniasis, compared to Chlamydia and Gonorrhoea; and (iv) for Waves I and II, compared to Wave III.

Since these are post-hoc findings, any interpretation of them should be cautious. Two of these findings should be considered jointly - Trichomoniasis was the most feminized of the three infections (Supplementary Table 2.7), reinforcing the suggestion that inequality plays a stronger role for women than men. The stronger result for women stands in contrast to a Indian study of income inequality and HIV, where inequality was positively associated with prevalent infection in men, but not in women (Perkins et al., 2009). Epidemiological studies for Trichomoniasis are limited to date, with six general population surveys reported worldwide as of 2008 (Johnston & Mabey, 2008), of which the only US study used the Add Health dataset (Miller & Zenilman, 2005). Careful analysis connecting the epidemiology of specific STIs and social dynamics is needed to understand why inequality might be particularly risky for specific racial/ethnic/gender strata. These factors are likely to vary by geography and time period (Aral et al., 2006), and strong conclusions will require intentional sampling to ensure sufficient power to make comparisons between subgroups.

The stronger association between inequality and STIs at Waves I and II than once Wave III outcomes were included is unsurprising for several reasons. First, school-level measures of community economic circumstances are likely to be more relevant while respondents are still in school - all respondents were still of school age at Wave II, since respondents in grade 12 at Wave I were excluded. This is especially likely for causal explanations of STI risk based on social comparisons or partnership-mixing. Second, any time-invariant community-based measure is likely

to become less valid over time: geographic mobility will lead to ever-increasing levels of misclassification regarding their relevant community and Wave II outcomes were measured only one year after family income data collection. Third, community economic factors may truly play a stronger role amongst younger individuals, and such factors may decline in importance as young people pass into their twenties. Alternatively, community economic factors may play a stronger role in determining risk for STI diagnosis than for actual acquisition – since only at Wave III do we have a laboratory-confirmed diagnosis. In order to disentangle these possible explanations we would need to collect more complete measures of geographic location, individual and community SES.

Strengths and Limitations

There are a number of potential threats to the strength of any inference drawn from this analysis. First, this study makes the working assumption that socioeconomic comparisons and effects at the school level are germane to the sexual behaviour and health of young adults. This may not be correct for two reasons. The first concern is a static one: that schools may not be the appropriate community forms for some or all community effects – e.g. if peer groups are formed across educational or administrative boundaries. This concern is allayed partially by earlier work suggesting that the aggregational unit used did not significantly affect ecological relationships between SES and STIs in two New England states (Krieger, 2002). Furthermore, a very high proportion of romantic relationships are formed within schools in the Add Health study (Raley & Sullivan, 2010). Additionally, past research

has found that neighbourhood factors influence adolescent sexual behaviour through their influence on the nature of school environments, rather than directly (Teitler & Weiss, 2000).

A second, and perhaps more worrying, concern is dynamic: mobility in the United States is considerable, particularly in early adulthood, and thus residential and scholastic location at Wave I may be a poor proxy for community over the subsequent six years. An important extension to this study would therefore be to gather information on geographic location of respondents over time, and link these locations to time-specific measures of socioeconomic structure.

Second, the measures of STI outcomes found in Add Health are not as comprehensive as we might wish: respondents report any STIs diagnosed up to Wave II and within one year of their Wave III interview, and laboratory testing captures any unresolved STI at Wave III. This misses any diagnosed and treated cases arising in the gap between Waves II and III. While we have no clear prior hypothesis as to why the pattern of diagnosis would differ systematically when comparing the intervening period to the years immediately prior to Waves II and III, it remains a possible source of bias. This concern is lessened since at least some respondents in the survey were of every age from 12 to 24 across the two waves, due to the multiple cohort structure of Add Health, so no specific age-range is missed (although sample sizes are insufficient to allow for a cohort-specific analysis).

Third, although we have considered multiple inequality measures and STIs in this study, it may be that our results are specific to the measures of SES and STIs considered. Further research looking at other measures of SES – for example, subjective measures of inequality (Atkinson, 1970) – will be important in identifying exactly what forms of social stratification affect disease risk. Furthermore, it was not within the scope of this paper to analyse how or if these associations are mediated by specific social processes, but future work could usefully consider whether those pathways outlined above do in fact mediate the inequality-STI relationships seen.

Fourth, as is common in survey-based analyses, there is considerable missingness of data relating to SES – 13.9% in the case of education and 24.3% for income. STI risk is not systematically different by missingness (Supplementary Table 2.10), and there is no clear reason to believe that STI risk within levels of SES variables is likely to differ depending on whether parents did or did not report their SES, however this remains an untestable assumption.

Finally, generalizing the results of this study should be done with care. Add Health is intrinsically rooted in the time, place and population from which the data – especially the outcomes – were collected. The 1990s in the United States was a period during which Gonorrhoea rates declined significantly, Trichomoniasis may have declined somewhat, and Chlamydia rates rose (Aral et al., 2007); 15-24 year olds have the highest rates of STIs of any age group in the country (Centers for Disease Control and Prevention, 2010). Whether any results found will translate to

situations where STI rates are different, or to age groups where infection rates are lower, is not clear.

Despite these limitations, this study has a number of strengths. First, the analysis is based on a prospectively interviewed, longitudinal cohort, which should limit concerns regarding the temporality of any effects seen. While attrition is a generic concern in cohort studies, it was relatively low in Add Health and analysis suggests it has had limited effect on estimates (Chantala et al.). Second, this sample is nationally representative of these six cohorts of students, allowing generalization to the US population at these ages. Third, we are able to combine laboratory STI testing (which allows us to avoid biases arising from non-random variation in healthcare access and testing) and self-report data (which allows us to move beyond currently prevalent infection to include treated cases). Additionally, any social desirability bias in reporting should be limited both by the audio-computer-assisted interview method used and by the knowledge that respondents are being laboratory-tested at the same time.

Conclusions

This study has focused on testing alternative hypotheses regarding how economic inequality affects health. Our analysis has both methodological implications regarding the importance of jointly considering multiple realisations of socioeconomic status, and practical implications regarding which factors might affect which young adults and put them at risk of STIs in the United States. Each of

these can be furthered by additional research. First, we believe that the joint use of income, inequality and relative deprivation measures is a potentially fruitful one for partitioning the effects of different sets of causal mechanisms on health. Second, preliminary findings regarding socioeconomic mechanisms can help direct analyses of the mediating processes themselves. In the example of STI risk in the US, next steps could include analyses of sexual partnership patterns by socioeconomic status, and how behaviours vary according to income levels or disparities.

In providing a feasible analytic framework for operationalizing theoretical mechanisms long discussed in the inequality and health literature, this study should encourage further careful analysis of the range of possible pathways connecting socioeconomic factors to health outcomes. Our approach also provides a correspondence between the causal mechanisms relating to absolute deprivation, structural inequality and relative deprivation and the segments of the population in which we should expect to see worse health outcomes if the respective mechanisms are at play. This should both assist those observing worse health in certain groups (e.g. the poor in unequal communities) in their search for putative causes (stress caused by social exclusion), and those finding empirical relationships in data (e.g. an association between the Yitzhaki index and STI risk) in pinpointing which parts of which communities are most likely affected (again, the poor in unequal communities).

References

Adimora AA & Schoenbach VJ. Contextual Factors and the Black-White Disparity in Heterosexual HIV Transmission. *Epidemiology*, 2002; 13(6): 707-12.

Adimora AA & Schoenbach VJ. Social Context, Sexual Networks, and Racial Disparities in Rates of Sexually Transmitted Infections. *Journal of Infectious Diseases*, 2005; 191(Suppl 1): S115-22.

Aral SO. Sexual Network Patterns as Determinants of STD Rates: Paradigm Shift in the Behavioral Epidemiology of STDs Made Visible. *Sexually Transmitted Diseases*, 1999; 26(5): 262-4.

Aral SO, Fenton KA & Holmes KK. Sexually Transmitted Diseases in the USA: Temporal Trends. *Sexually Transmitted Infections*, 2007; 83(4): 257-66.

Aral SO, O'Leary A & Baker C. Sexually Transmitted Infections and HIV in the Southern United States: An Overview. *Sexually Transmitted Diseases*, 2006; 33(7 Suppl): S1-5.

Atkinson AB. On the Measurement of Inequality. *Journal of Economic Theory*, 1970; 2: 244-63.

Atkinson AB, Rainwater L & Smeeding TM. *Income Distribution in OECD Countries: Evidence from the Luxembourg Income Study*; 1995. Paris: Organisation for Economic Co-operation and Development.

Bärnighausen T & Tanser F. Rethinking the Role of the Local Community in HIV Epidemic Spread in Sub-Saharan Africa: A Proximate-Determinants Approach. *HIV Therapy*, 2009; 3(5): 435-45.

Bauermeister JA, Zimmerman MA & Caldwell CH. Neighborhood Disadvantage and Changes in Condom Use among African American Adolescents. *Journal of Urban Health*, 2011; 88(1): 66-83.

Boerma JT & Weir SS. Integrating Demographic and Epidemiological Approaches to Research on HIV/AIDS: The Proximate-Determinants Framework. *Journal of Infectious Diseases*, 2005; 191(Suppl 1): S61-7.

Browning CR, Leventhal T & Brooks-Gunn J. Neighborhood Context and Racial Differences in Early Adolescent Sexual Activity. *Demography*, 2004; 41(4): 697-720.

Campbell ME, Haveman R, Wildhagen T & Wolfe BL. Income Inequality and Racial Gaps in Test Scores. In: Magnuson KA & Waldfogel J (Eds). *Steady Gains and Stalled Progress: Inequality and the Black-white Test Score Gap*; 2008. New York: Russell Sage Foundation.

Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2009*. 2010. Atlanta, GA: Centers for Disease Control and Prevention.

Chantala K, Kalsbeek WD & Andraca E. *Non-Response in Wave III of the Add Health Study*. n.d. Available from: http://www.cpc.unc.edu/projects/addhealth/data/guides/W3nonres.pdf. [Last Accessed 1 October 2012].

Cohen DA, Spear S, Scribner R, Kissinger P, Mason K & Wildgen J. "Broken Windows" and the Risk of Gonorrhea. *American Journal of Public Health*, 2000; 90(2): 230-6.

Cowell FA. *Measuring Inequality*; 2011. Oxford University Press.

Deaton A. Health, Inequality, and Economic Development. *Journal of Economic Literature*, 2003; 41(1): 113-58.

Deaton A & Lubotsky D. Mortality, Inequality and Race in American Cities and States. *Social Science and Medicine*, 2003; 56(6): 1139-53.

Doherty IA, Shiboski S, Ellen JM, Adimora AA & Padian NS. Sexual Bridging Socially and over Time: A Simulation Model Exploring the Relative Effects of Mixing and Concurrency on Viral Sexually Transmitted Infection Transmission. *Sexually Transmitted Diseases*, 2006; 33(6): 368-73.

Dupéré V, Lacourse E, Willms JD, Leventhal T & Tremblay RE. Neighborhood Poverty and Early Transition to Sexual Activity in Young Adolescents: A

Developmental Ecological Approach. *Child Development*, 2008; 79(5): 1463-76.

Eaton L, Flisher AJ & Aarø LE. Unsafe Sexual Behaviour in South African Youth. *Social Science and Medicine*, 2003; 56(1): 149-65.

Feldacker C, Emch M & Ennett S. The Who and Where of HIV in Rural Malawi: Exploring the Effects of Person and Place on Individual HIV Status. *Health and Place*, 2010; 16(5): 996-1006.

Garnett GP & Anderson RM. Sexually Transmitted Diseases and Sexual Behavior: Insights from Mathematical Models. *Journal of Infectious Diseases*, 1996; 174(Suppl 2): S150–61.

Goodman E, Huang B, Wade TJ & Kahn RS. A Multilevel Analysis of the Relation of Socioeconomic Status to Adolescent Depressive Symptoms: Does School Context Matter? *Journal of Pediatrics*, 2003; 143(4): 451-6.

Gravelle H. How Much of the Relation between Population Mortality and Unequal Distribution of Income Is a Statistical Artefact? *British Medical Journal*, 1998; 316(7128): 382-5.

Harris KM, Halpern CT, Whitsel E, Hussey J, Tabor J, Entzel P et al. *The National Longitudinal Study of Adolescent Health: Research Design*. 2009. Available from: http://www.cpc.unc.edu/projects/addhealth/design. [Last Accessed 1 October 2012].

Henderson M, Butcher I, Wight D, Williamson L & Raab G. What Explains between-School Differences in Rates of Sexual Experience? *BMC Public Health*, 2008; 8: 53-.

Holtgrave DR & Crosby RA. Social Determinants of Tuberculosis Case Rates in the United States. *American Journal of Preventive Medicine*, 2004; 26(2): 159-62.

Johnston VJ & Mabey DC. Global Epidemiology and Control of Trichomonas Vaginalis. *Current Opinion in Infectious Diseases*, 2008; 21: 56-64.

Kawachi I. Income Inequality and Health. In: Berkman LF & Kawachi I (Eds). *Social Epidemiology*; 2000. New York: Oxford University Press.

Kawachi I & Berkman LF. Social Integration, Social Capital, and Health. In: Berkman LF & Kawachi I (Eds). *Social Epidemiology*; 2000. New York: Oxford University Press.

Kondo N, Kawachi I, Subramanian SV, Takeda Y & Yamagata Z. Do Social Comparisons Explain the Association between Income Inequality and Health?: Relative Deprivation and Perceived Health among Male and Female Japanese Individuals. *Social Science and Medicine*, 2008; 67(6): 982-7.

Kondo N, Sembajwe G, Kawachi I, Van Dam RM, Subramanian SV & Yamagata Z. Income Inequality, Mortality, and Self Rated Health: Meta-Analysis of Multilevel Studies. *British Medical Journal*, 2009; 339: b4471-b.

Krieger N. Geocoding and Monitoring of US Socioeconomic Inequalities in Mortality and Cancer Incidence: Does the Choice of Area-Based Measure and Geographic Level Matter?: The Public Health Disparities Geocoding Project. *American Journal of Epidemiology*, 2002; 156(5): 471-82.

Krieger N, Waterman PD, Chen JT, Soobader M-J & Subramanian SV. Monitoring Socioeconomic Inequalities in Sexually Transmitted Infections, Tuberculosis, and Violence: Geocoding and Choice of Area-Based Socioeconomic Measures--the Public Health Disparities Geocoding Project (US). *Public Health Reports*, 2003; 118(3): 240-60.

Kubzansky L & Kawachi I. Affective States and Health. In: Berkman LF & Kawachi I (Eds). *Social Epidemiology*; 2000. New York: Oxford Univeristy Press.

Leigh A, Jencks C & Smeeding TM. Health and Economic Inequality. In: Salverda W, Nolan B & Smeeding TM (Eds). *The Oxford Handbook of Economic Inequality*, 384-405; 2009. New York, NY: Oxford University Press.

Lynch J, Harper S & Davey Smith G. Commentary: Plugging Leaks and Repelling Boarders--Where to Next for the Ss Income Inequality? *International Journal of Epidemiology*, 2003; 32(6): 1029-36.

Mellor JM & Milyo J. Reexamining the Evidence of an Ecological Association between Income Inequality and Health. *Journal of Health Politics, Policy and Law*, 2001; 26: 487-522.

Miller WC & Zenilman JM. Epidemiology of Chlamydial Infection, Gonorrhea, and Trichomoniasis in the United States - 2005. *Infectious Disease Clinics of North America*, 2005; 19: 281-96.

Monteiro EF, Lacey CJN & Merrick D. The Interrelation of Demographic and Geospatial Risk Factors between Four Common Sexually Transmitted Diseases. *Sexually Transmitted Infections*, 2005; 81(1): 41-6.

Nilsson T. *Inequality, Globalization and Health*. 2009. Unpublished Thesis. Lund University.

Parker RG, Easton D & Klein CH. Structural Barriers and Facilitators in HIV Prevention: A Review of International Research. *AIDS*, 2000; 14(Suppl 1): S22-32.

Perkins JM, Khan KT & Subramanian SV. Patterns and Distribution of HIV among Adult Men and Women in India. *PLoS ONE*, 2009; 4(5): e5648-e.

Poundstone KE, Strathdee SA & Celentano DD. The Social Epidemiology of Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome. *Epidemiologic Reviews*, 2004; 26: 22-35.

Raley RK & Sullivan MK. Social-Contextual Influences on Adolescent Romantic Involvement: The Constraints of Being a Numerical Minority. *Sociological Spectrum*, 2010; 30(1): 65-89.

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(3): 237-53.

Resnick MD, Bearman PS, Blum RW, Bauman KE, Harris KM, Jones J et al. Protecting Adolescents from Harm: Findings from the National Longitudinal Study on Adolescent Health. *JAMA*, 1997; 278(10): 823-32.

Roche KM & Leventhal T. Beyond Neighborhood Poverty: Family Management, Neighborhood Disorder, and Adolescents' Early Sexual Onset. *Journal of Family Psychology*, 2009; 23(6): 819-27.

Rotheram-Borus MJ, Swendeman D & Chovnick G. The Past, Present, and Future of HIV Prevention: Integrating Behavioral, Biomedical, and Structural Intervention Strategies for the Next Generation of HIV Prevention. *Annual Review of Clinical Psychology*, 2009; 5: 143-67.

Runciman WG. *Relative Deprivation and Social Justice*; 1966. London: Routledge & Kegan Paul.

Spriggs AL, Halpern CT, Herring AH & Schoenbach VJ. Family and School Socioeconomic Disadvantage: Interactive Influences on Adolescent Dating Violence Victimization. *Social Science and Medicine*, 2009; 68(11): 1956-65.

Subramanian SV & Kawachi I. The Association between State Income Inequality and Worse Health Is Not Confounded by Race. *International Journal of Epidemiology*, 2003; 32(6): 1022-8.

Teitler JO & Weiss CC. Effects of Neighborhood and School Environments on Transitions to First Sexual Intercourse. *Sociology of Education*, 2000; 73(2): 112-32.

Webber C. Revaluating Relative Deprivation Theory. *Theoretical Criminology*, 2007; 11(1): 97-120.

Wilkinson RG & Pickett KE. Income Inequality and Population Health: A Review and Explanation of the Evidence. *Social Science and Medicine*, 2006; 62(7): 1768-84.

Yitzhaki S. Relative Deprivation and the Gini Coefficient. *Quarterly Journal of Economics*, 1979; 7(4): 321-4.

Zierler S & Krieger N. Reframing Women's Risk: Social Inequalities and HIV Infection. *Annual Review of Public Health*, 1997; 18(70): 401-36.

Figure 2.1: Conceptual map of economic disadvantages

	Non-poor Family	Poor Family
Equal Community		Absolute deprivation
Unequal Community	Inequality	Absolute deprivation Inequality Relative deprivation

Table 2.1: Univariate descriptive statistics of a sample of respondents from Waves I to III of Add Health

	Entire sample		Male		Female	
	N	**%**	**N**	**%**	**N**	**%**
Number of respondents	11,183		5,350		5,833	
Self-reported or lab-confirmed STI at Waves II or III	1,170	10.5	455	8.5	715	12.3
Age at baseline						
<14	1,823	16.3	801	15.0	1,022	17.5
14	1,671	14.9	771	14.4	900	15.4
15	2,085	18.6	984	18.4	1,101	18.9
16	2,184	19.5	1,066	19.9	1,118	19.2
17	2,054	18.4	1,037	19.4	1,017	17.4
>17	1,366	12.2	691	12.9	675	11.6
Sex						
Male	5,833	52.2				
Female	5,350	47.8				
Parental Education						
< High School graduate/GED	1,250	11.2	582	10.9	668	11.5
High School graduate/GED	2,693	24.1	1,251	23.4	1,442	24.7
Some college	3,460	30.9	1,692	31.6	1,768	30.3
Completed 4-year college	2,051	18.3	1,003	18.7	1,048	18.0
Any post-graduate	1,729	15.5	822	15.4	907	15.5
Urbanicity of school						
Urban	3,148	28.1	1,458	27.3	1,690	29.0
Suburban	5,994	53.6	2,888	54.0	3,106	53.2
Rural	2,041	18.3	1,004	18.8	1,037	17.8
Region of country						
West	2,630	23.5	1,277	23.9	1,353	23.2
Midwest	2,957	26.4	1,389	26.0	1,568	26.9
South	4,085	36.5	1,978	37.0	2,107	36.1
Northeast	1,511	13.5	706	13.2	805	13.8
Type of school						
Public	10,344	92.5	4,938	92.3	5,406	92.7
Private	839	7.5	412	7.7	427	7.3
Individual Race/Ethnicity						
White non-Hispanic	6,448	57.7	3,082	57.6	3,366	57.7
Black non-Hispanic	2,261	20.2	1,014	19.0	1,247	21.4
Hispanic	1,656	14.8	831	15.5	825	14.1
Other non-Hispanic	818	7.3	423	7.9	395	6.8
	Median	**IQR**	**Median**	**IQR**	**Median**	**IQR**
School Gini coefficient	33.7	[29.5 - 38.6]	33.7	[29.5 - 38.6]	33.7	[29.5 - 38.7]
Per capita family income ($†)	18,475	[10,500 - 28,868]	18,898	[10,614 - 28,868]	18,031	[10,392 - 28,868]
School Yitzhaki index ($†)	6,387	[3,126 - 10,833]	6,265	[3,085 - 10,833]	6,456	[3,168 - 10,833]

STI: Diagnosis of Chlamydia, Gonorrhea or Trichomoniasis. GED: General Educational Development tests. IQR: Inter-quintile range.

† All income-based measures use an equivalence scale such that these figures are family income divided by the square root of the number of family members.

Table 2.2: Multivariable regressions of income inequality as a predictor of STI diagnosis at Waves II or III of Add Health

	Bivariate Models		Model 2		Model 3		Model 4 (adjusted†)		Model 5 (adjusted†)	
School Gini coefficient										
Most equal quintile	1.00		1.00		1.00		1.00		1.00	
2nd most equal quintile	1.06	[0.76 - 1.49]	1.05	[0.77 - 1.44]	1.06	[0.79 - 1.44]	1.19	[0.83 - 1.70]	0.96	[0.74 - 1.23]
Middle quintile	1.84	[1.32 - 2.57]	1.69	[1.23 - 2.31]	1.61	[1.19 - 2.18]	1.63	[1.16 - 2.29]	0.89	[0.68 - 1.18]
2nd least equal quintile	2.24	[1.65 - 3.04]	1.98	[1.46 - 2.67]	1.87	[1.39 - 2.52]	1.82	[1.35 - 2.47]	1.14	[0.89 - 1.46]
Least equal quintile	2.50	[1.82 - 3.45]	2.12	[1.55 - 2.89]	1.99	[1.47 - 2.70]	2.00	[1.44 - 2.78]	1.06	[0.81 - 1.38]
Per-capita family income										
Poorest quintile	2.04	[1.64 - 2.54]	1.94	[1.56 - 2.41]	2.58	[1.88 - 3.56]	2.42	[1.68 - 3.47]	1.64	[1.18 - 2.27]
2nd poorest quintile	1.57	[1.26 - 1.96]	1.53	[1.22 - 1.91]	1.89	[1.43 - 2.52]	1.76	[1.29 - 2.40]	1.39	[1.04 - 1.86]
Middle quintile	1.38	[1.10 - 1.72]	1.37	[1.09 - 1.71]	1.60	[1.23 - 2.07]	1.49	[1.13 - 1.97]	1.29	[0.99 - 1.69]
2nd richest quintile	1.32	[1.05 - 1.66]	1.33	[1.06 - 1.67]	1.43	[1.13 - 1.81]	1.35	[1.06 - 1.73]	1.28	[1.00 - 1.63]
Richest quintile	1.00		1.00		1.00		1.00		1.00	
Family Yitzhaki index										
Least relatively deprived quintile	1.00				1.00		1.00		1.00	
2nd least deprived quintile	1.10	[0.90 - 1.35]			0.86	[0.69 - 1.08]	0.86	[0.69 - 1.08]	0.93	[0.74 - 1.16]
Middle quintile	1.26	[1.03 - 1.53]			0.83	[0.65 - 1.06]	0.80	[0.62 - 1.03]	0.89	[0.69 - 1.13]
2nd most deprived quintile	1.28	[1.04 - 1.57]			0.73	[0.55 - 0.95]	0.71	[0.53 - 0.94]	0.81	[0.62 - 1.05]
Most relatively deprived quintile	1.49	[1.21 - 1.85]			0.70	[0.51 - 0.96]	0.67	[0.48 - 0.94]	0.78	[0.58 - 1.04]
Individual Race/Ethnicity										
White non-Hispanic									1.00	
Black non-Hispanic									4.93	[4.14 - 5.88]
Hispanic									1.84	[1.45 - 2.35]
Other non-Hispanic									1.98	[1.48 - 2.64]
Akaike Information Criteria (AIC)			7,247.7		7,249.6		7,203.9		6,892.2	
Intraclass correlation (ICC)			0.073		0.065		0.055		0.009	

All models are two-level hierarchical models of 11,183 individuals nested in 132 schools. Figures are odds ratios with 95% confidence intervals in brackets.

† Models 4 and 5 are adjusted for individual baseline age, sex and parental education and for school-level urbanicity, region and public/private type. Full results for these models are provided in Supplementary Table 2.6.

Table 2.3: Multivariable regressions for income inequality and STI diagnosis in Add Health: sub-group analyses by sex

	Male						Female					
School Gini coefficient												
Most equal quintile	1.00		1.00		1.00		1.00		1.00		1.00	
2nd most equal quintile	1.37	[0.88 - 2.14]	1.35	[0.89 - 2.05]	0.99	[0.69 - 1.41]	1.27	[0.83 - 1.93]	1.22	[0.83 - 1.79]	0.99	[0.72 - 1.36]
Middle quintile	1.46	[0.92 - 2.33]	1.34	[0.87 - 2.08]	0.76	[0.52 - 1.10]	1.61	[1.08 - 2.40]	1.50	[1.01 - 2.21]	0.97	[0.68 - 1.36]
2nd least equal quintile	1.85	[1.19 - 2.88]	1.74	[1.16 - 2.63]	1.12	[0.79 - 1.58]	1.99	[1.36 - 2.92]	1.81	[1.26 - 2.61]	1.19	[0.87 - 1.63]
Least equal quintile	2.03	[1.26 - 3.26]	1.90	[1.21 - 2.98]	0.90	[0.62 - 1.32]	2.44	[1.60 - 3.71]	2.18	[1.46 - 3.26]	1.14	[0.81 - 1.61]
Per-capita family income												
Poorest quintile			2.54	[1.53 - 4.23]	1.28	[0.77 - 2.14]			2.86	[1.81 - 4.53]	1.88	[1.22 - 2.89]
2nd poorest quintile			1.56	[0.99 - 2.44]	1.02	[0.64 - 1.61]			2.25	[1.51 - 3.36]	1.72	[1.17 - 2.52]
Middle quintile			1.50	[1.00 - 2.24]	1.15	[0.76 - 1.75]			1.67	[1.16 - 2.40]	1.40	[0.98 - 2.00]
2nd richest quintile			1.19	[0.82 - 1.71]	1.15	[0.79 - 1.67]			1.55	[1.11 - 2.15]	1.47	[1.06 - 2.04]
Richest quintile			1.00		1.00				1.00		1.00	
Family Yitzhaki index												
Least relatively deprived quintile			1.00		1.00				1.00		1.00	
2nd least deprived quintile			0.69	[0.49 - 0.98]	0.93	[0.64 - 1.35]			0.97	[0.71 - 1.30]	1.04	[0.77 - 1.41]
Middle quintile			0.70	[0.48 - 1.01]	0.87	[0.58 - 1.32]			0.83	[0.60 - 1.15]	0.92	[0.67 - 1.26]
2nd most deprived quintile			0.52	[0.34 - 0.79]	0.82	[0.53 - 1.27]			0.77	[0.53 - 1.10]	0.90	[0.63 - 1.27]
Most relatively deprived quintile			0.47	[0.29 - 0.76]	0.71	[0.44 - 1.14]			0.72	[0.47 - 1.09]	0.86	[0.59 - 1.25]
Individual Race/Ethnicity												
White non-Hispanic					1.00						1.00	
Black non-Hispanic					5.10	[3.95 - 6.59]					5.07	[4.03 - 6.36]
Hispanic					1.87	[1.31 - 2.66]					1.80	[1.32 - 2.47]
Other non-Hispanic					1.36	[0.85 - 2.19]					2.51	[1.75 - 3.60]
No. of individuals (level 1)	5,350		5,350		5,350		5,833		5,833		5,833	
No. of schools (level 2)	132		132		132		132		132		132	
Akaike Information Criteria (AIC)	3,038.0		3,036.3		2,907.0		4,214.8		4,202.8		4,003.9	
Intraclass correlation (ICC)	0.068		0.046		-		0.068		0.052		0.009	

All models are two-level hierarchical models, and also adjust for individual baseline age, sex and parental education and for school-level urbanicity, region and public/private type.
Figures are odds ratios with 95% confidence intervals in brackets.

Table 2.4: Multivariable regressions for income inequality and STI diagnosis in Add Health: sub-group analyses by race/ethnicity

	White non-Hispanic		Black non-Hispanic		Hispanic		Other non-Hispanic	
School Gini coefficient	1.00		1.00		1.00		1.00	
Most equal quintile	0.92	[0.67 - 1.25]	0.92	[0.53 - 1.63]	1.10	[0.48 - 2.51]	1.82	[0.56 - 5.92]
2nd most equal quintile	0.91	[0.63 - 1.32]	0.71	[0.40 - 1.26]	1.54	[0.71 - 3.35]	1.87	[0.62 - 5.67]
Middle quintile	0.99	[0.71 - 1.38]	0.99	[0.57 - 1.70]	1.92	[0.91 - 4.02]	2.04	[0.72 - 5.78]
2nd least equal quintile	1.15	[0.76 - 1.73]	0.94	[0.55 - 1.60]	1.25	[0.53 - 2.93]	3.46	[0.97 - 12.4]
Least equal quintile								
Per-capita family income	1.68	[0.98 - 2.90]	1.62	[0.95 - 2.77]	1.03	[0.40 - 2.60]	1.87	[0.53 - 6.57]
Poorest quintile	1.53	[0.96 - 2.42]	1.37	[0.85 - 2.22]	1.05	[0.45 - 2.43]	0.79	[0.25 - 2.48]
2nd poorest quintile	1.29	[0.85 - 1.94]	1.16	[0.74 - 1.82]	1.23	[0.55 - 2.73]	1.19	[0.42 - 3.33]
Middle quintile	1.31	[0.91 - 1.88]	1.22	[0.81 - 1.86]	1.32	[0.62 - 2.81]	1.02	[0.38 - 2.72]
2nd richest quintile	1.00		1.00		1.00		1.00	
Richest quintile								
Family Yitzhaki index	1.00		1.00		1.00		1.00	
Least relatively deprived quintile	0.80	[0.55 - 1.16]	0.97	[0.67 - 1.39]	0.99	[0.54 - 1.82]	1.38	[0.52 - 3.67]
2nd least deprived quintile	0.92	[0.62 - 1.37]	0.81	[0.55 - 1.20]	0.99	[0.51 - 1.92]	1.52	[0.55 - 4.22]
Middle quintile	0.78	[0.51 - 1.19]	0.71	[0.46 - 1.10]	1.17	[0.58 - 2.37]	1.77	[0.56 - 5.60]
2nd most deprived quintile	0.58	[0.36 - 0.93]	0.83	[0.52 - 1.34]	1.07	[0.48 - 2.41]	2.15	[0.62 - 7.38]
No. of individuals (level 1)	6,448		2,261		1,656		818	
No. of schools (level 2)	126		097		107		99	
Akaike Information Criteria (AIC)	2,793.6		2,531.7		1,129.4		535.6	
Intraclass correlation (ICC)	-		0.009		0.006		-	

All models are two-level hierarchical models, and also adjust for individual baseline age, sex and parental education and for school-level urbanicity, region and public/private type. Figures are odds ratios with 95% confidence intervals in brackets.

Supplementary Table 2.5: Bivariate relationships between covariates and self-reported or laboratory-confirmed STI at Wave II or III of Add Health

	Entire sample		Male		Female	
School Gini coefficient						
Most equal quintile	1.00		1.00		1.00	
2nd most equal quintile	1.06	[0.76 - 1.49]	1.27	[0.78 - 2.02]	1.28	[0.87 - 1.94]
Middle quintile	1.84	[1.32 - 2.57]	1.69	[1.07 - 2.66]	1.48	[1.01 - 2.18]
2nd least equal quintile	2.24	[1.65 - 3.04]	2.02	[1.29 - 3.18]	2.23	[1.51 - 3.29]
Least equal quintile	2.50	[1.82 - 3.45]	2.02	[1.28 - 3.18]	2.43	[1.59 - 3.72]
Per-capita family income						
Poorest quintile	2.04	[1.64 - 2.54]	1.73	[1.25 - 2.40]	2.61	[1.94 - 3.50]
2nd poorest quintile	1.57	[1.26 - 1.96]	1.19	[0.85 - 1.67]	2.08	[1.55 - 2.79]
Middle quintile	1.38	[1.10 - 1.72]	1.22	[0.88 - 1.71]	1.60	[1.18 - 2.16]
2nd richest quintile	1.32	[1.05 - 1.66]	1.10	[0.78 - 1.56]	1.56	[1.15 - 2.11]
Richest quintile	1.00		1.00		1.00	
Family Yitzhaki index						
Least relatively deprived quintile	1.00		1.00		1.00	
2nd least deprived quintile	1.10	[0.90 - 1.35]	0.91	[0.66 - 1.24]	1.29	[0.98 - 1.70]
Middle quintile	1.26	[1.03 - 1.53]	1.08	[0.80 - 1.47]	1.38	[1.06 - 1.81]
2nd most deprived quintile	1.28	[1.04 - 1.57]	0.99	[0.72 - 1.36]	1.48	[1.13 - 1.94]
Most relatively deprived quintile	1.49	[1.21 - 1.85]	1.09	[0.79 - 1.51]	1.76	[1.32 - 2.35]
Age at baseline						
<14	0.77	[0.58 - 1.02]	0.48	[0.33 - 0.71]	0.99	[0.72 - 1.34]
14	0.93	[0.72 - 1.21]	0.69	[0.49 - 0.98]	1.15	[0.86 - 1.53]
15	1.00	[0.82 - 1.23]	0.83	[0.61 - 1.12]	1.10	[0.85 - 1.43]
16	1.00					
17	0.76	[0.62 - 0.94]	0.67	[0.49 - 0.91]	0.84	[0.64 - 1.11]
>17	0.74	[0.58 - 0.94]	0.56	[0.39 - 0.81]	0.90	[0.66 - 1.22]
Sex						
Male vs. Female	0.66	[0.58 - 0.75]				
Parental Education						
< High School graduate/GED	1.11	[0.90 - 1.37]	0.91	[0.65 - 1.28]	1.32	[1.01 - 1.72]
High School graduate/GED	1.00		1.00		1.00	
Some college	0.90	[0.74 - 1.09]	0.76	[0.56 - 1.03]	1.01	[0.78 - 1.30]
Completed 4-year college	0.58	[0.46 - 0.73]	0.57	[0.40 - 0.82]	0.57	[0.41 - 0.78]
Any post-graduate	1.38	[1.25 - 1.51]	1.42	[1.20 - 1.69]	1.50	[1.22 - 1.84]
Urbanicity of school						
Urban	0.73	[0.51 - 1.05]	0.75	[0.45 - 1.25]	0.76	[0.50 - 1.14]
Suburban	1.00		1.00		1.00	
Rural	0.76	[0.55 - 1.04]	0.68	[0.48 - 0.99]	0.91	[0.68 - 1.22]
Region of country						
West	0.59	[0.45 - 0.78]	0.53	[0.37 - 0.75]	0.84	[0.51 - 1.40]
Midwest	0.78	[0.56 - 1.08]	0.53	[0.36 - 0.78]	0.82	[0.55 - 1.23]
South	1.00		1.00		1.00	
Northeast	1.35	[1.23 - 1.48]	1.30	[1.10 - 1.53]	1.45	[1.14 - 1.84]
Type of school						
Private vs. Public	1.41	[1.28 - 1.55]	1.42	[1.20 - 1.67]	1.60	[1.29 - 1.98]
Individual Race/Ethnicity						
White non-Hispanic	1.00					
Black non-Hispanic	5.53	[4.70 - 6.50]	3.89	[1.00 - 1.00]	5.66	[4.58 - 6.99]
Hispanic	2.05	[1.66 - 2.54]	1.76	[1.00 - 1.00]	2.07	[1.56 - 2.74]
Other non-Hispanic	1.91	[1.44 - 2.52]	1.26	[1.00 - 1.00]	2.42	[1.72 - 3.42]

STI: Diagnosis of Chlamydia, Gonorrhea or Trichomoniasis. GED: General Educational Development tests.

Supplementary Table 2.6: Multivariable regressions of income inequality and STI diagnosis: Full covariate results

	I		II		III		IV		V	
School Gini coefficient										
Most equal quintile	1.00		1.00		1.00		1.00		1.00	
2nd most equal quintile	1.06	[0.76 - 1.49]	1.05	[0.77 - 1.44]	1.06	[0.79 - 1.44]	1.19	[0.83 - 1.70]	0.96	[0.74 - 1.23]
Middle quintile	1.84	[1.32 - 2.57]	1.69	[1.23 - 2.31]	1.61	[1.19 - 2.18]	1.63	[1.16 - 2.29]	0.89	[0.68 - 1.18]
2nd least equal quintile	2.24	[1.65 - 3.04]	1.98	[1.46 - 2.67]	1.87	[1.39 - 2.52]	1.82	[1.35 - 2.47]	1.14	[0.89 - 1.46]
Least equal quintile	2.50	[1.82 - 3.45]	2.12	[1.55 - 2.89]	1.99	[1.47 - 2.70]	2.00	[1.44 - 2.78]	1.06	[0.81 - 1.38]
Per-capita family income										
Poorest quintile			1.94	[1.56 - 2.41]	2.58	[1.88 - 3.56]	2.42	[1.68 - 3.47]	1.64	[1.18 - 2.27]
2nd poorest quintile			1.53	[1.22 - 1.91]	1.89	[1.43 - 2.52]	1.76	[1.29 - 2.40]	1.39	[1.04 - 1.86]
Middle quintile			1.37	[1.09 - 1.71]	1.60	[1.23 - 2.07]	1.49	[1.13 - 1.97]	1.29	[0.99 - 1.69]
2nd richest quintile			1.33	[1.06 - 1.67]	1.43	[1.13 - 1.81]	1.35	[1.06 - 1.73]	1.28	[1.00 - 1.63]
Richest quintile			1.00		1.00		1.00		1.00	
Family Yitzhaki index										
Least relatively deprived quintile					1.00		1.00		1.00	
2nd least deprived quintile					0.86	[0.69 - 1.08]	0.86	[0.69 - 1.08]	0.93	[0.74 - 1.16]
Middle quintile					0.83	[0.65 - 1.06]	0.80	[0.62 - 1.03]	0.89	[0.69 - 1.13]
2nd most deprived quintile					0.73	[0.55 - 0.95]	0.71	[0.53 - 0.94]	0.81	[0.62 - 1.05]
Most relatively deprived quintile					0.70	[0.51 - 0.96]	0.67	[0.48 - 0.94]	0.78	[0.58 - 1.04]
Age at baseline										
<14							0.79	[0.62 - 1.00]	0.79	[0.63 - 0.99]
14							0.95	[0.76 - 1.19]	0.95	[0.77 - 1.19]
15							1.01	[0.83 - 1.23]	0.98	[0.81 - 1.20]
16							1.00		1.00	
17							0.77	[0.63 - 0.95]	0.77	[0.62 - 0.95]
>17							0.74	[0.58 - 0.93]	0.74	[0.58 - 0.94]
Sex										
Male vs. Female							0.67	[0.59 - 0.76]	0.67	[0.59 - 0.77]
Parental Education										
< High School graduate/GED							0.98	[0.79 - 1.21]	1.05	[0.84 - 1.30]
High School graduate/GED							1.00		1.00	
Some college							0.99	[0.83 - 1.17]	0.95	[0.80 - 1.13]
Completed 4-year college							1.02	[0.83 - 1.25]	0.98	[0.80 - 1.21]
Any post-graduate							0.71	[0.55 - 0.91]	0.69	[0.53 - 0.89]
Urbanicity of school										
Urban							0.75	[0.57 - 0.99]	0.89	[0.70 - 1.14]
Suburban							1.00		1.00	
Rural							1.02	[0.82 - 1.27]	1.01	[0.85 - 1.21]
Region of country										
West							0.84	[0.63 - 1.12]	0.90	[0.72 - 1.13]
Midwest							0.88	[0.68 - 1.14]	1.13	[0.92 - 1.38]
South							1.00		1.00	
Northeast							0.65	[0.47 - 0.90]	0.86	[0.67 - 1.10]
Type of school										
Private vs. Public							1.25	[0.78 - 1.99]	1.12	[0.82 - 1.54]
Individual Race/Ethnicity										
White non-Hispanic									1.00	
Black non-Hispanic									4.93	[4.14 - 5.88]
Hispanic									1.84	[1.45 - 2.35]
Other non-Hispanic									1.98	[1.48 - 2.64]
Akaike Information Criteria (AIC)	7,278.5		7,247.7		7,249.6		7,203.9		6,892.2	
Intraclass correlation (ICC)	0.084		0.073		0.065		0.055		0.009	

All models are two-level hierarchical models of 11,183 individuals nested in 132 schools.
Figures are odds ratios with 95% confidence intervals in brackets.

Supplementary Table 2.7: Multivariable regressions for income inequality and STI diagnosis in Add Health: secondary outcomes of specific STIs

	Chlamydia				Gonorrhoea				Trichomoniasis			
School Gini coefficient												
Most equal quintile	1.00		1.00		1.00		1.00		1.00		1.00	
2nd most equal quintile	1.23	[0.87 - 1.75]	1.03	[0.79 - 1.33]	1.38	[0.72 - 2.64]	0.94	[0.52 - 1.70]	1.46	[0.83 - 2.58]	1.10	[0.66 - 1.85]
Middle quintile	1.31	[0.94 - 1.83]	0.93	[0.70 - 1.22]	1.49	[0.71 - 3.14]	0.79	[0.42 - 1.51]	1.59	[0.92 - 2.77]	1.03	[0.59 - 1.79]
2nd least equal quintile	1.76	[1.29 - 2.42]	1.19	[0.92 - 1.54]	1.83	[0.98 - 3.42]	0.99	[0.54 - 1.80]	1.62	[0.94 - 2.80]	1.08	[0.64 - 1.81]
Least equal quintile	1.80	[1.26 - 2.56]	0.98	[0.74 - 1.31]	1.65	[0.82 - 3.30]	0.68	[0.35 - 1.31]	2.46	[1.42 - 4.25]	1.31	[0.77 - 2.23]
Per-capita family income												
Poorest quintile	2.65	[1.77 - 3.96]	1.76	[1.24 - 2.51]	3.36	[1.46 - 7.72]	1.70	[0.77 - 3.75]	2.95	[1.60 - 5.45]	1.77	[0.96 - 3.25]
2nd poorest quintile	1.93	[1.37 - 2.73]	1.51	[1.09 - 2.07]	2.34	[1.12 - 4.90]	1.58	[0.77 - 3.21]	2.12	[1.24 - 3.62]	1.54	[0.90 - 2.62]
Middle quintile	1.59	[1.17 - 2.17]	1.37	[1.02 - 1.84]	1.39	[0.69 - 2.81]	1.08	[0.54 - 2.15]	1.57	[0.96 - 2.55]	1.29	[0.80 - 2.10]
2nd richest quintile	1.36	[1.03 - 1.80]	1.29	[0.98 - 1.69]	1.22	[0.63 - 2.35]	1.11	[0.58 - 2.12]	1.34	[0.86 - 2.08]	1.25	[0.80 - 1.93]
Richest quintile	1.00		1.00		1.00		1.00		1.00		1.00	
Family Yitzhaki index												
Least relatively deprived quintile	1.00		1.00		1.00		1.00		1.00		1.00	
2nd least deprived quintile	0.87	[0.67 - 1.12]	0.93	[0.73 - 1.20]	1.08	[0.61 - 1.90]	1.24	[0.71 - 2.18]	0.70	[0.47 - 1.04]	0.77	[0.52 - 1.15]
Middle quintile	0.79	[0.59 - 1.05]	0.87	[0.67 - 1.14]	0.95	[0.52 - 1.76]	1.16	[0.63 - 2.11]	0.55	[0.35 - 0.86]	0.64	[0.41 - 1.00]
2nd most deprived quintile	0.69	[0.50 - 0.95]	0.79	[0.59 - 1.06]	0.78	[0.40 - 1.54]	1.06	[0.56 - 2.01]	0.50	[0.31 - 0.82]	0.62	[0.38 - 1.01]
Most relatively deprived quintile	0.63	[0.43 - 0.91]	0.75	[0.55 - 1.01]	0.48	[0.22 - 1.04]	0.67	[0.33 - 1.36]	0.55	[0.32 - 0.95]	0.72	[0.42 - 1.24]
Age at baseline												
<14	0.88	[0.68 - 1.15]	0.83	[0.65 - 1.06]	0.99	[0.59 - 1.67]	1.00	[0.61 - 1.65]	0.62	[0.40 - 0.95]	0.61	[0.40 - 0.93]
14	1.09	[0.85 - 1.40]	1.04	[0.82 - 1.32]	0.90	[0.53 - 1.52]	0.91	[0.54 - 1.51]	0.82	[0.55 - 1.22]	0.81	[0.55 - 1.20]
15	1.08	[0.86 - 1.35]	1.03	[0.82 - 1.29]	1.04	[0.65 - 1.68]	1.03	[0.64 - 1.65]	0.94	[0.67 - 1.33]	0.91	[0.65 - 1.29]
16	1.00		1.00		1.00		1.00		1.00		1.00	
17	0.77	[0.61 - 0.98]	0.77	[0.61 - 0.98]	0.79	[0.48 - 1.31]	0.81	[0.49 - 1.36]	0.76	[0.53 - 1.10]	0.77	[0.53 - 1.11]
>17	0.74	[0.57 - 0.97]	0.75	[0.57 - 0.99]	0.51	[0.26 - 0.99]	0.52	[0.27 - 1.01]	0.95	[0.65 - 1.40]	0.96	[0.65 - 1.42]
Sex												
Male vs. Female	0.68	[0.58 - 0.78]	0.68	[0.59 - 0.79]	1.13	[0.83 - 1.53]	1.18	[0.87 - 1.61]	0.56	[0.45 - 0.71]	0.57	[0.45 - 0.72]
Parental Education												
< High School graduate/GED	0.92	[0.72 - 1.18]	0.97	[0.76 - 1.25]	1.02	[0.63 - 1.64]	1.19	[0.73 - 1.92]	1.01	[0.70 - 1.46]	1.07	[0.73 - 1.55]
High School graduate/GED	1.00		1.00		1.00		1.00		1.00		1.00	
Some college	1.05	[0.86 - 1.27]	1.01	[0.83 - 1.23]	0.85	[0.57 - 1.27]	0.80	[0.54 - 1.20]	0.88	[0.65 - 1.19]	0.84	[0.62 - 1.13]
Completed 4-year college	1.12	[0.89 - 1.40]	1.09	[0.87 - 1.38]	0.63	[0.36 - 1.09]	0.58	[0.34 - 1.01]	0.91	[0.63 - 1.32]	0.87	[0.60 - 1.27]
Any post-graduate	0.78	[0.59 - 1.05]	0.76	[0.57 - 1.01]	0.79	[0.43 - 1.46]	0.71	[0.39 - 1.31]	0.83	[0.54 - 1.30]	0.81	[0.52 - 1.26]
Individual Race/Ethnicity												
White non-Hispanic			1.00				1.00				1.00	
Black non-Hispanic			4.52	[3.73 - 5.48]			7.34	[4.74 - 11.36]			4.80	[3.50 - 6.60]
Hispanic			1.94	[1.49 - 2.51]			1.63	[0.84 - 3.14]			1.69	[1.07 - 2.67]
Other non-Hispanic			1.87	[1.36 - 2.58]			0.91	[0.31 - 2.67]			1.89	[1.08 - 3.30]
Akaike Information Criteria (AIC)	5,927.9		5,707.8		1,766.7		1,664.7		2,931.4		2,833.8	
Intraclass correlation (ICC)	0.038		-		0.060		-		0.101		0.063	

All models are two-level hierarchical models of 11,183 individuals nested in 132 schools, and also adjust for school-level urbanicity, region and public/private type. Figures are odds ratios with 95% confidence intervals in brackets.

Supplementary Table 2.8: Multivariable regressions for income inequality and STI diagnosis in Add Health: robustness checks

	Check 1: Add Wave I outcomes				Check 2: Require presence at Waves II & III				Check 3: Use only Wave I and Wave II outcomes			
School Gini coefficient												
Most equal quintile	1.00		1.00		1.00		1.00		1.00		1.00	
2nd most equal quintile	1.18	[0.84 - 1.68]	0.97	[0.76 - 1.25]	1.22	[0.89 - 1.69]	0.95	[0.72 - 1.24]	1.33	[0.77 - 2.30]	1.08	[0.65 - 1.79]
Middle quintile	1.55	[1.10 - 2.20]	0.89	[0.68 - 1.16]	1.37	[0.98 - 1.93]	0.80	[0.59 - 1.08]	1.30	[0.76 - 2.23]	0.93	[0.55 - 1.59]
2nd least equal quintile	1.85	[1.37 - 2.51]	1.15	[0.90 - 1.47]	1.74	[1.28 - 2.38]	1.15	[0.88 - 1.50]	2.29	[1.36 - 3.83]	1.65	[1.01 - 2.69]
Least equal quintile	1.97	[1.42 - 2.74]	1.05	[0.80 - 1.37]	1.93	[1.36 - 2.73]	0.98	[0.73 - 1.31]	2.40	[1.38 - 4.14]	1.48	[0.88 - 2.49]
Per-capita family income												
Poorest quintile	2.42	[1.71 - 3.42]	1.66	[1.20 - 2.29]	2.54	[1.72 - 3.75]	1.56	[1.08 - 2.24]	3.03	[1.63 - 5.62]	1.95	[1.05 - 3.62]
2nd poorest quintile	1.71	[1.26 - 2.31]	1.36	[1.02 - 1.81]	1.83	[1.30 - 2.57]	1.35	[0.97 - 1.86]	1.67	[0.95 - 2.94]	1.29	[0.73 - 2.27]
Middle quintile	1.49	[1.15 - 1.95]	1.30	[1.00 - 1.68]	1.61	[1.19 - 2.18]	1.33	[0.99 - 1.80]	1.26	[0.74 - 2.14]	1.06	[0.62 - 1.79]
2nd richest quintile	1.38	[1.09 - 1.74]	1.30	[1.03 - 1.65]	1.48	[1.12 - 1.94]	1.38	[1.05 - 1.82]	1.33	[0.83 - 2.12]	1.26	[0.79 - 2.02]
Richest quintile	1.00		1.00		1.00		1.00		1.00		1.00	
Family Yitzhaki index												
Least relatively deprived quintile	1.00		1.00		1.00		1.00		1.00		1.00	
2nd least deprived quintile	0.83	[0.67 - 1.04]	0.90	[0.72 - 1.12]	0.87	[0.68 - 1.12]	0.94	[0.73 - 1.21]	0.65	[0.41 - 1.01]	0.69	[0.44 - 1.08]
Middle quintile	0.80	[0.62 - 1.01]	0.88	[0.69 - 1.11]	0.85	[0.65 - 1.12]	0.97	[0.74 - 1.27]	0.83	[0.53 - 1.31]	0.95	[0.60 - 1.49]
2nd most deprived quintile	0.69	[0.53 - 0.91]	0.79	[0.61 - 1.03]	0.71	[0.52 - 0.97]	0.88	[0.65 - 1.18]	0.58	[0.35 - 0.97]	0.75	[0.45 - 1.24]
Most relatively deprived quintile	0.65	[0.47 - 0.89]	0.74	[0.56 - 0.99]	0.62	[0.43 - 0.88]	0.81	[0.58 - 1.11]	0.52	[0.29 - 0.91]	0.69	[0.40 - 1.20]
Baseline age												
< 14	0.73	[0.57 - 0.92]	0.72	[0.57 - 0.90]	0.74	[0.57 - 0.95]	0.74	[0.58 - 0.94]	0.22	[0.12 - 0.40]	0.22	[0.12 - 0.39]
14	0.90	[0.72 - 1.12]	0.89	[0.72 - 1.10]	0.99	[0.78 - 1.25]	0.97	[0.77 - 1.22]	0.51	[0.33 - 0.79]	0.49	[0.32 - 0.77]
15	1.00	[0.82 - 1.21]	0.97	[0.80 - 1.18]	1.03	[0.84 - 1.28]	1.00	[0.80 - 1.24]	0.70	[0.48 - 1.01]	0.69	[0.48 - 0.99]
16	1.00		1.00		1.00		1.00		1.00		1.00	
17	0.84	[0.69 - 1.02]	0.84	[0.69 - 1.03]	0.86	[0.68 - 1.09]	0.86	[0.68 - 1.09]	1.56	[1.13 - 2.16]	1.58	[1.14 - 2.19]
> 17	0.93	[0.75 - 1.15]	0.94	[0.75 - 1.17]	1.03	[0.76 - 1.40]	1.03	[0.76 - 1.41]	1.67	[1.11 - 2.51]	1.70	[1.13 - 2.57]
Individual Race/Ethnicity												
White non-Hispanic			1.00				1.00				1.00	
Black non-Hispanic			4.95	[4.17 - 5.87]			4.78	[3.93 - 5.81]			3.44	[2.47 - 4.78]
Hispanic			1.81	[1.44 - 2.28]			1.93	[1.47 - 2.52]			1.56	[1.01 - 2.42]
Other non-Hispanic			1.82	[1.37 - 2.42]			2.03	[1.47 - 2.80]			0.89	[0.46 - 1.69]
No. of individuals (level 1)	11,183		11,183		8,754		8,754		10,932		10,932	
No. of schools (level 2)	132		132		132		132		132		132	
Akaike Information Criteria (AIC)	7,633.2		7,298.9		5,745.7		5,507.3		2,584.2		2,528.2	
Intraclass correlation (ICC)	0.057		0.012		0.044		0.009		0.052		0.021	

All models are two-level hierarchical models, and also adjust for individual baseline age, sex and parental education and for school-level urbanicity, region and public/private type.
Figures are odds ratios with 95% confidence intervals in brackets.

Supplementary Table 2.9: Multivariable regressions for income inequality and STI diagnosis in Add Health: unadjusted models of economic measures

	Bivariate models		Bivariate models with race/ethnicity†		Multivariable models III		IV		V		VI	
School Gini coefficient												
Most equal quintile	1.00		1.00		1.00		1.00				1.00	
2nd most equal quintile	1.06	[0.76 - 1.49]	0.97	[0.75 - 1.26]	1.05	[0.77 - 1.44]	1.01	[0.71 - 1.44]			1.06	[0.79 - 1.44]
Middle quintile	1.84	[1.32 - 2.57]	0.94	[0.72 - 1.23]	1.69	[1.23 - 2.31]	1.88	[1.33 - 2.65]			1.61	[1.19 - 2.18]
2nd least equal quintile	2.24	[1.65 - 3.04]	1.24	[0.97 - 1.58]	1.98	[1.46 - 2.67]	2.22	[1.63 - 3.03]			1.87	[1.39 - 2.52]
Least equal quintile	2.50	[1.82 - 3.45]	1.12	[0.86 - 1.45]	2.12	[1.55 - 2.89]	2.44	[1.74 - 3.43]			1.99	[1.47 - 2.70]
Per-capita family income												
Poorest quintile	2.04	[1.64 - 2.54]	1.60	[1.29 - 1.99]	1.94	[1.56 - 2.41]			3.18	[2.27 - 4.46]	2.58	[1.88 - 3.56]
2nd poorest quintile	1.57	[1.26 - 1.96]	1.38	[1.10 - 1.72]	1.53	[1.22 - 1.91]			2.18	[1.62 - 2.93]	1.89	[1.43 - 2.52]
Middle quintile	1.38	[1.10 - 1.72]	1.30	[1.04 - 1.62]	1.37	[1.09 - 1.71]			1.73	[1.33 - 2.26]	1.60	[1.23 - 2.07]
2nd richest quintile	1.32	[1.05 - 1.66]	1.31	[1.04 - 1.64]	1.33	[1.06 - 1.67]			1.48	[1.16 - 1.88]	1.43	[1.13 - 1.81]
Richest quintile	1.00		1.00		1.00				1.00		1.00	
Family Yitzhaki index												
Least relatively deprived quintile	1.00		1.00				1.00		1.00		1.00	
2nd least deprived quintile	1.10	[0.90 - 1.35]	1.09	[0.88 - 1.34]			1.10	[0.89 - 1.35]	0.83	[0.66 - 1.04]	0.86	[0.69 - 1.08]
Middle quintile	1.26	[1.03 - 1.53]	1.17	[0.95 - 1.43]			1.24	[1.01 - 1.51]	0.78	[0.60 - 1.00]	0.83	[0.65 - 1.06]
2nd most deprived quintile	1.28	[1.04 - 1.57]	1.13	[0.92 - 1.39]			1.23	[1.00 - 1.51]	0.66	[0.49 - 0.87]	0.73	[0.55 - 0.95]
Most relatively deprived quintile	1.49	[1.21 - 1.85]	1.17	[0.95 - 1.45]			1.42	[1.15 - 1.76]	0.60	[0.43 - 0.83]	0.70	[0.51 - 0.96]
Akaike Information Criteria (AIC)					7,247.7		7,274.3		7,273.8		7,249.6	
Intraclass correlation (ICC)					0.073		0.091		0.068		0.065	

All models are two-level hierarchical models of 11,183 individuals nested in 132 schools. † Coefficient values for race/ethnicity not shown.
Figures are odds ratios with 95% confidence intervals in brackets.

Supplementary Table 2.10: A comparison of Add Health respondents at Wave II or III with Missing and Non-Missing family incomes at Wave I

	Most equal	2nd most equal	Middle	2nd least equal	Least equal	All non-missing	Missing	χ^2 test	p-value
No. of respondents	2,417	2,355	2,171	2,284	1,956	11,183	3,625		
Amongst the Missing	518	620	968	632	887				
Proportion testing positive for any STI	6.7%	8.5%	10.3%	13.4%	14.3%	10.5%			
Amongst the Missing	6.0%	8.9%	9.4%	15.8%	16.0%		11.6%	3.43	0.064
School Gini coefficient									
Most equal quintile						21.6%	14.3%		
2nd most equal quintile						21.1%	17.1%		
Middle quintile						19.4%	26.7%		
2nd least equal quintile						20.4%	17.4%		
Least equal quintile						17.5%	24.5%	245.89	<.0001
Per-capita family income									
Poorest quintile	9.4%	14.5%	17.4%	27.2%	31.7%	19.8%	0.3%		
2nd poorest quintile	17.4%	21.6%	20.1%	21.8%	19.1%	20.0%	0.2%		
Middle quintile	23.9%	23.8%	22.0%	18.4%	11.9%	20.4%	0.1%		
2nd richest quintile	28.7%	23.5%	19.0%	17.9%	11.0%	19.6%	0.1%		
Richest quintile	28.6%	21.9%	18.5%	16.9%	14.0%	20.3%	0.1%	9.44	0.051
Missing							99.1%		
Family Yitzhaki index									
Least relatively deprived quintile	26.1%	21.1%	19.7%	20.5%	12.6%	20.0%	0.1%		
2nd least deprived quintile	24.7%	22.9%	19.2%	19.6%	13.7%	20.1%	0.1%		
Middle quintile	19.1%	21.7%	20.3%	21.1%	17.9%	20.0%	0.1%		
2nd most deprived quintile	19.8%	19.2%	20.2%	20.6%	20.2%	20.0%	0.2%		
Most relatively deprived quintile	18.4%	20.4%	17.8%	20.4%	23.1%	20.0%	0.3%		
Missing							99.1%	9.07	0.059
Individual Race/Ethnicity									
White non-Hispanic	31.6%	26.4%	13.6%	17.6%	10.8%	57.7%	41.7%		
Black non-Hispanic	5.0%	14.2%	22.1%	26.9%	31.8%	20.2%	24.4%		
Hispanic	10.0%	13.2%	29.7%	17.9%	29.2%	14.8%	19.8%		
Other non-Hispanic	12.6%	13.6%	37.3%	29.5%	7.1%	7.3%	14.1%	335.83	<.0001
Sex									
Male	21.6%	20.9%	18.8%	20.6%	18.1%	52.2%	54.4%		
Female	21.7%	21.2%	20.1%	20.2%	16.8%	47.8%	45.6%	5.38	0.020
Age at baseline									
<14	26.4%	23.1%	8.2%	33.2%	9.1%	16.3%	11.5%		
14	27.7%	21.0%	9.7%	28.7%	13.0%	14.9%	11.3%		
15	19.8%	23.4%	19.4%	17.2%	20.2%	18.6%	15.5%		
16	19.2%	18.9%	26.1%	16.2%	19.6%	19.5%	19.3%		
17	18.8%	20.1%	26.0%	14.7%	20.5%	18.4%	20.5%		
>17	18.6%	19.9%	25.8%	13.5%	22.2%	12.2%	21.9%	265.90	<.0001
Highest parental education									
< High School graduate/GED	6.6%	13.4%	25.5%	22.0%	32.5%	11.2%	9.3%		
High School graduate/GED	18.6%	23.4%	16.3%	23.5%	18.2%	24.1%	11.5%		
Some college	24.5%	21.1%	19.6%	21.4%	13.5%	30.9%	10.7%		
Completed 4-year college	26.6%	22.3%	19.9%	16.9%	14.2%	18.3%	6.6%		
Any post-graduate	25.4%	21.6%	18.7%	16.8%	17.5%	15.5%	5.4%	152.29	<.0001
Missing							56.6%		
Urbanicity									
Urban	15.2%	14.2%	14.7%	27.6%	28.4%	28.2%	29.3%		
Suburban	25.4%	18.5%	27.1%	19.5%	9.5%	53.6%	56.1%		
Rural	20.4%	39.3%	4.1%	11.9%	24.3%	18.3%	14.6%	25.16	<.0001
Region									
West	22.9%	18.3%	36.4%	21.1%	1.3%	23.5%	29.2%		
Midwest	22.4%	36.1%	23.5%	10.9%	7.1%	26.4%	19.6%		
South	15.5%	13.5%	10.5%	27.0%	33.5%	36.5%	40.2%		
Northeast	34.7%	16.9%	5.8%	20.1%	22.6%	13.5%	11.1%	109.67	<.0001
Type of school									
Public	18.8%	21.8%	21.0%	21.1%	17.3%	92.5%	93.3%		
Private	56.1%	12.3%	0.0%	11.8%	19.8%	7.5%	6.7%	2.58	0.108

Do “sugar daddies” drive HIV incidence in young women? Evidence from a population cohort in rural KwaZulu-Natal, South Africa

Guy Harling, Marie-Louise Newell, Frank Tanser, Ichiro Kawachi, SV Subramanian, Till Bärnighausen.

Acknowledgement: This research uses data collected at the Africa Centre for Health and Population Studies and would not have been possible without the kind contributions of all respondents and the support of many staff at the Centre, for which the authors are extremely grateful.

Abstract

Background: Older sexual partners are considered a major risk factor for HIV for young women in sub-Saharan Africa. Ethnographic investigations have shown age-disparities between sexual partners to be associated with power imbalances and differential sexual behaviours. Numerous public health campaigns have been conducted to discourage young women from relationships with older men, often framed as initiatives against "sugar daddy" relationships. However, quantitative evidence on the effect of sex-partner age-disparity on HIV acquisition in women has been largely lacking.

Methods: We studied whether sex partner age-disparity was associated with future HIV acquisition risk in a population-based, open cohort of 2,444 15-29 year old women in rural KwaZulu-Natal, South Africa (2003-2012).

Findings: 458 HIV seroconversions occurred over 5,913 person-years of follow-up (incidence rate: 7·75 per 100 person-years). The age-disparity of women's most recent sexual partner at each interview round was not associated with HIV acquisition when measured continuously (hazard ratio [HR] for a one-year increase in partner's age: 1·00, 95% CI 0·97-1·03), or when measured categorically (man ≥5 years older: HR 0·98, 95% CI 0·81-1·20; man ≥10 years older: HR 0·98, 95% CI 0·67-1·43). These results were robust to the inclusion of known socio-demographic and behavioural risk factors for HIV infection and did not vary significantly by women's age, marital status, education attainment, or household wealth.

Interpretation: In this rural setting in KwaZulu-Natal, age-disparity within a sexual partnership did not predict HIV acquisition amongst young women. In particular, "sugar daddy" relationships did not seem to play a major role in driving the epidemic. Investments in "sugar daddy" campaigns are unlikely to be a cost-effective use of HIV-prevention resources.

Introduction

In recent years, there has been much concern in sub-Saharan Africa regarding "sugar daddy" relationships. Definitions of "sugar daddies" vary widely, but there is general agreement that they involve non-marital relationships between young women and substantially older men, often also displaying social or economic disparities and thus either financial or in-kind resource transfer (Hope, 2007; Leclerc-Madlala, 2008; Luke, 2003, 2005; Wyrod et al., 2011). "Sugar daddy" relationships are therefore some subset of age-disparate ones. Most current National Strategic HIV Plans in countries with generalized epidemics name age-disparate relationships as a driver of the HIV epidemic (see Table 3.1). The World Health Organization considers sexual partnerships between young women and substantially older men an important contributor to young people's vulnerability for HIV (WHO, 2004).

As a result, numerous public health campaigns have been aimed at "sugar daddies". These have aimed to educate young women and older men, to increase economic opportunities for women, and to change cultural norms by stigmatizing such relationships(Hope, 2007). Population Services International (PSI), an international Non-Governmental Organization, has run health messaging programmes targeting age-disparate relationships in at least eight sub-Saharan countries (PSI). Similarly, in 2012, the KwaZulu-Natal (KZN) provincial Department of Health in South Africa began a campaign under the title '"Sugar Daddies" Destroy Lives', aiming to "create a

taboo against cross-generational sex" with 14-25 year old women in order to reduce HIV infection (Kwa-Zulu Natal Department of Health, 2012).

Since the 1990s, theoretical mathematical models have investigated the potential impact of age-mixing in relationships on population HIV dynamics, and on individual infection risk (Garnett & Anderson, 1993; Hallett et al., 2007). In the context of dynamic populations, these models predict that age-mixing increases the level of infection in those with little sexual history such as the young, and plays an important role in propagating the epidemic to the next generation (Garnett & Anderson, 1996).

At the relationship level, there are two main reasons why we might expect to see an association between "sugar daddies" and HIV incidence (Hope, 2007). First, HIV infection rates rise rapidly from youth through middle age amongst men in sub-Saharan Africa (UNAIDS, 2012). As a result, young women in an age-disparate relationship are, ceteris paribus, at higher risk of HIV exposure than if they were in a relationship with someone their own age.

Second, sexual behaviours may vary with the level of age-disparity in sexual relationships; specifically, fewer preventative measures may be taken. This risky behaviour may arise because the man perceives younger women to be less likely to be infected with HIV or because the economic and social power differentials generated by the age-disparity may make it difficult for women to negotiate safe sex with their older partners (Wingood & DiClemente, 2000; Woolf & Maisto, 2008). For

example, there is some evidence that men are less likely to use condoms in sexual relationships with younger women (Bankole et al., 2007; Luke, 2005).

Based on the theoretical and empirical literature, relationships between young women and older men, including "sugar daddy" relationships, are a plausible driver of the HIV epidemic in sub-Saharan Africa. However, while many countries in the region are already investing HIV prevention resources in campaigns to reduce the prevalence of "sugar-daddy" and age-disparate relationships (Table 3.1), quantitative evidence for a causal relationship remains very limited.

The hypothesis that age-disparate sexual partnerships raise the risk of HIV is indirectly supported by evidence that sub-Saharan Africa has the largest average relationship age gaps (Wellings et al., 2006), and the highest HIV rates (UNAIDS, 2012), in the world. Population-level variation in age-disparity and HIV rates also appears to positively co-vary within sub-Saharan Africa, with higher age-disparities and HIV rates in the South of the sub-continent compared to the East (Chapman et al., 2010). Individual-level empirical evidence for the relationship in Africa remains scarce. Cross-sectional analyses in Rakai, Uganda (1994-1998), Manicaland, Zimbabwe (1998-2000), and nationwide in South Africa (2003) found positive associations between having an older partner and HIV prevalence (Gregson et al., 2002; Kelly et al., 2003; Pettifor et al., 2005).

Unlike in longitudinal studies, however, in cross-sectional studies the exposure does not precede the outcome and it is thus impossible to rule out many alternative

explanations for an association between age-disparity and HIV infection. For example, women could preferentially seek out relationships with older men *after* they have become HIV-infected – for instance, because older men are better able to support them economically (Hope, 2007; Leclerc-Madlala, 2008); and HIV-infected women who have older partners may live longer because they are economically better off and thus have better access to antiretroviral treatment.

Given the difficulties in identifying causal effects with cross-sectional studies, a longitudinal analysis of the "sugar daddy" hypothesis would represent a significant improvement on existing evidence, and could provide strong evidence for a causal relationship. We therefore test the hypothesis that "sugar daddy" and age-disparate relationships increase the risk of HIV infection amongst young women, using one of Africa's largest HIV incidence cohorts, which is located in a rural community in KZN where HIV incidence in young women has been very high over the past decade (Bärnighausen et al., 2009).

Methods

We conducted survival analysis using data from population-based, longitudinal surveillance conducted by the Africa Centre for Health and Population Studies (hereafter Africa Centre) in a predominantly rural community in the uMkhanyakude district of KZN. The district is one of the most deprived in the country and is characterized by high levels of circular migration and HIV infection. The Africa

Centre has been collecting household demographic data since 2000 (Tanser et al., 2008). In addition, since 2003 adults have annually been invited to participate in anonymised HIV testing, when they are also asked questions relating to their sexual history and behaviours over the past 12 months. These questions are asked face-to-face by fieldworkers recruited from the local community (Bärnighausen et al., 2009). HIV test results are linked anonymously to other information in the database.

Data were available from January 2003 until June 2012. Inclusion criteria for our analysis were that respondents were female and during this period: (i) were aged between 15 and 30; (ii) were HIV seronegative at first participation in the HIV surveillance and had at least one more valid HIV test result recorded; and (iii) participated at least once in the General Health surveillance questionnaire which elicits information on sexual behaviour, including sexual partners' ages. Individuals entered the cohort at the date of their first report of a sexual partnership, and were right-censored at their thirtieth birthday or on the date of their most recent seronegative HIV test prior to 30 June 2012. Person-time during which the respondent indicated that they had had no sexual partners was not included in the study.

The primary outcome was HIV seroconversion; we assumed the date of HIV seroconversion to be midway between the date of an individual's last negative and first positive HIV test. The exposure of interest was the age-disparity in the respondent's most recently reported sexual partnership (when no information on the age of sexual partners was provided at an interview round, the previous age-

disparity was presumed to still be the relevant one). For our main analysis individuals could not enter the cohort until they had reported the age of at least one partner.

The literature provides many competing definitions of "sugar daddy" relationships: the level of age-disparity required ranges from five to twenty years, the level of resource transfer is rarely defined, and there is sometimes an additional requirement that the woman be aged under 20 or under 25. Given this range of definitions, we used several different functional forms of age-disparity in sexual relationships. We first used age-disparity as a continuous variable, reflecting both a theoretical orientation towards a continuum of risk and the empirical finding that there was no sharp change in risk at any particular cut-off. We additionally conducted analyses using step functions to capture age-disparities that are larger than 5, larger than 10, and larger than 20 years, corresponding to common definitions of "sugar daddy" and age-disparate relationships (Hope, 2007).

We considered as potential socio-demographic time-varying confounders: current completed education (none or primary, 0-7 years; secondary, 8-12 years; tertiary, >12 years); household wealth (quintiles of the first component identified by principal-components analysis of 28 household assets, toilet type, and sources of water, electricity, and energy); and marital status (never married, engaged, married, previously married). We considered as potential behavioural confounders or mediators both age at sexual debut and three time-varying measures of sexual behaviour in the past 12 months: number of partners (0-1 vs >1); any casual partner

(yes vs. no); and lowest level of condom use with any partner (never, sometimes, always).

We used Cox proportional hazards models, verifying the proportional-hazards assumption using the Schoenfeld residuals from each regression. The primary model included the woman's age (centred at age 15) and their relationship age-disparity. Based on tests of functional form (see Supplementary Table 3.4), we included linear, quadratic and cubic terms in age, but only a linear term in age-disparity. Our model also included indicator variables for the year of observation. We then considered whether any effect of age-disparity varied by women's age in three categories (15-19, 20-24, and 25-29 years old) using interaction terms. Finally, we added socio-demographic and sexual behaviour covariates to the model. We reran our analyses after multiple imputation of missing variables in the dataset (see Supplementary Table 3.5).

Ethical approval for Africa Centre surveillance was granted by the Biomedical Research Ethics Committee, University of KwaZulu-Natal. Informed consent is required separately for the main and sexual behaviour questionnaires, and for HIV sero-testing. This analysis was exempted from ethical review by the Harvard School of Public Health Institutional Review Board due to its use of anonymised secondary data.

Results

Between January 2003 and June 2012, 2,444 women contributed 5,913 years of person-time that met the inclusion criteria and had full covariate information. Baseline characteristics of the respondents are provided in Table 3.2, divided into three five-year age cohorts. The young women were predominantly educated to secondary level; relatively few lived in the richest wealth quintile of households. The great majority had never been married, although 22% were engaged by ages 25-29. Median age at sexual debut was 17 years. Very few women reported multiple partners or a casual partner in the past year – the latter declined strongly with age.

The median sexual partnership for women involved a man who was three years older than her, ranging from 10 years younger to 47 years older, with 922 (37·7%) respondents reporting having had a partner five or more years older than themselves at any point in the study period, and 222 (9·1%) reporting one 10 or more years older.

During follow-up 458 HIV seroconversions were observed. The overall incidence rate was 7·75 per 100 person-years (95% confidence interval (CI): 7·07-8·49). The incidence rate per 100 person-years rose from 7·79 (95% CI: 6·59-9·22) amongst those aged under 20 to 8·63 (95% CI: 7·63-9·77) for those aged 20-24, before dropping to 5·63 (95% CI: 4·46-7·11) for those aged 25-29. A crude comparison of the age-disparity of each woman's most recent sexual partner at baseline and their

subsequent risk of seroconverting whilst under observation (Figure 3.1), suggests no obvious correlation.

In survival analysis containing only respondent's age and relational age-disparity (Model 1, Table 3.3), there was no significant relationship between age-disparity and HIV acquisition (Hazard ratio (HR) for a one-year increase in partner's age: 1.00, 95% CI 0.97-1.03). This result appeared to vary only very slightly – and non-significantly – by five-year categories of respondent age (Model 2). The addition of socioeconomic covariates (Model 3) and behavioural covariates (Model 4) had almost no effect on the result.

Analyses using categorical measures of age-disparity found similar results (man ≥5 years older: HR 0·98, 95% CI 0·81-1·20; man ≥10 years older: HR 0·98, 95% CI 0·67-1·43; man≥20 years older: HR 0·61, 95% CI 0·15-2·46). After multiple imputation of missing values, our results remained essentially unchanged (Supplementary Table 3.6). Interactions of age-disparity with woman's age did not change our findings (Supplementary Table 3.7). No significant effect-modification of the relationship between age-disparities and HIV infection was seen for marital status or educational attainment (Supplementary Table 3.8). For wealth, there may be an effect such that those in the richest and poorest quintiles are protected by greater age-disparities while those in the middle quintile are placed at greater risk by them, but this was only significant for a binary measure of age-disparity with a cut-point of ≥5 years.

Discussion

We tested whether age-disparities, including those associated with common definitions of "sugar daddy" relationships, affect HIV incidence in young women using data from one of Africa's largest population-based HIV incidence cohorts. In a context where HIV prevalence and incidence are very high, we find no evidence to suggest that having an older male partner increases the risk of acquiring HIV in young women. We obtain this finding despite the fact that age-disparate relationships in this community are common. Over the study period, more than one-third of young women had a partner who was at least five years older than themselves – a similar level to that reported nationally for South Africa (32·6%) (Pettifor et al., 2005) – and almost one in ten had a partner who was at least ten years older – comparable to the level seen in rural Uganda (16·3%) (Kelly et al., 2003). Our results are robust to the functional form of age-disparity and to the inclusion of a wide range of potentially confounding covariates in the regression analysis.

There are several plausible explanations for the absence of significant associations between "sugar daddy" (and other age-disparate) relationships and HIV acquisition risk in this community. First, young women in "sugar daddy" relationships may be more careful in selecting their partners so as to offset their risk. There is evidence that while men have more control over partnership sexual practices, including condom use, women have greater control over partnership formation and dissolution (Luke, 2003, 2005). This can be seen in the effectiveness of interventions

aimed at changing young women's partners using information or financial incentives (Baird et al., 2012; Dupas, 2011). In particular, it could explain our findings, if young women increasingly selected low-risk partners as the age-disparity in their sexual relationships increased. If young women are aware that older partners are likely to be more risky, they may use knowledge drawn from their social networks to identify lower-risk partners from amongst the pool of older men available to them. Such as strategy may indeed be successful in rural communities such as the one studied here, where dense social networks may allow such identification (Reynolds et al., 2013).

Second, age-disparities are likely to be more weakly linked to economic disparities in this setting than elsewhere, reducing or eliminating the infection risk that has been hypothesized to exist within age-disparate relationships due to steep differentials in sexual negotiation power (Leclerc-Madlala, 2008; Luke, 2003). Socioeconomic differentials in this poor rural community are far less pronounced than in many other settings, in particular urban areas, limiting the potential for resource transfers from older men to younger women. In fact, the majority of men living in this community were unemployed over the observation period (Ardington et al., 2009).

Our findings regarding effect-modification by education and wealth variables support the absence of an effect of socio-economic inequalities on the relationship between age-disparities and HIV in this setting. Were the "sugar daddy" hypothesis an important risk factor, we would expect the effects of age-disparities to be

greatest for those women who are least educated and poorest, and thus in greatest need of material support and least able to refuse older men's advances. Our finding of no effect-modification by educational attainment, and reduced risk of HIV related to age-disparities for the poorest women, suggests that "sugar daddies" are unlikely to be playing a substantial role in propagating HIV in this setting.

It is notable that partnership age-disparities in this cohort are highest for marriages and lowest for casual relationships (Ott et al., 2011), reflecting how long it takes men to save sufficient amounts of money to pay a "lobola" or bride price (Hosegood et al., 2009). Since age-disparate married relationships are not considered "sugar daddy" relationships, their comparatively high prevalence in this community could hide a true "sugar daddy" effect. However, comparing the age-disparity effect on HIV acquisition risk in women who have never been married or engaged to those who have, we detect neither significant nor large differences, further strengthening the conclusion that "sugar daddies" do not drive HIV incidence in this community.

This study has several strengths, notably the longitudinal nature of the data, collected over almost a decade, and the rich set of covariates, allowing us to rule out many confounding and reverse causation relationships. Furthermore, the data constitute one of the largest HIV incidence cohorts in young women in Africa, with 458 directly observed seroconversions over almost 6,000 person-years, providing very high power to detect significant effects.

We also note some limitations. As with any long-term community-based study, the cohort suffered from attrition and non-response. However, non-response was limited and our results did not change after accounting for data missingness through multiple imputation. Furthermore, caution is needed in generalizing our findings. Age-disparate relationships may place young women at particularly high risk when they need economic support for survival (Hope, 2007; Leclerc-Madlala, 2008). Such survival needs may well be greater in other settings: while this setting is one of the poorest in South Africa, it is considerably wealthier than many other high-prevalence regions of Africa. Additionally, this community may contain relatively few older, rich men who can act as "sugar daddies", since the community is relatively homogeneous in economic circumstances, and the impact of the HIV-related mortality over the past 20 years has substantially reduced the number of older men (Bor et al., 2013). Finally, as noted above, the dense social networks in this area may allow women to differentiate higher and lower risk older men; in more urban or less settled areas such differentiation may not be possible, leading to increased risk from age-disparate relationships.

That age-disparate sexual relations and "sugar daddies" are a primary driver of HIV incidence for young women in sub-Saharan Africa has often been taken as fact, despite limited evidence. Our analysis suggests that in a typical rural South African community, sexual relationship age-disparities are not a strong predictor of HIV risk for young women. Campaigns warning women about the risks of sexual partnerships with older men may well provide social benefits, particularly if they

reduce unwanted pregnancies and increase sexual risk awareness. However, investing in this area specifically to reduce HIV infection rates is unlikely to be an optimal use of scarce HIV prevention resources. Further research into how the relationship between relational age-disparities and HIV risk may vary by geographic and social context is needed to justify the continuation of such campaigns.

References

Ardington C, Case A & Hosegood V. Labor Supply Responses to Large Social Transfers: Longitudinal Evidence from South Africa. *American Economic Journal. Applied Economics*, 2009; 1: 22–48.

Baird SJ, Garfein RS, McIntosh CT & Ozler B. Effect of a Cash Transfer Programme for Schooling on Prevalence of HIV and Herpes Simplex Type 2 in Malawi: A Cluster Randomised Trial. *Lancet*, 2012; 379(9823): 1320–9.

Bankole A, Biddlecom A, Guiella G, Singh S & Zulu E. Sexual Behavior, Knowledge and Information Sources of Very Young Adolescents in Four Sub-Saharan African Countries. *African Journal of Reproductive Health*, 2007; 11(3): 28–43.

Bärnighausen T, Tanser F & Newell M-L. Lack of a Decline in HIV Incidence in a Rural Community with High HIV Prevalence in South Africa, 2003-2007. *AIDS Research and Human Retroviruses*, 2009; 25(4): 405–9.

Bor J, Herbst AJ, Newell M-L & Bärnighausen T. Increases in Adult Life Expectancy in Rural South Africa: Valuing the Scale-up of HIV Treatment. *Science*, 2013; 339(6122): 961–5.

Chapman R, White RG, Shafer LA, Pettifor AE, Mugurungi O, Ross D et al. Do Behavioural Differences Help to Explain Variations in HIV Prevalence in Adolescents in Sub-Saharan Africa? *Tropical Medicine and International Health*, 2010; 15(5): 554–66.

Dupas P. Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya. *American Economic Journal. Applied Economics*, 2011; 3(1): 1–34.

Garnett GP & Anderson RM. Factors Controlling the Spread of HIV in Heterosexual Communities in Developing Countries: Patterns of Mixing between Different Age and Sexual Activity Classes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 1993; 342(1300): 137–59.

Garnett GP & Anderson RM. Sexually Transmitted Diseases and Sexual Behavior: Insights from Mathematical Models. *Journal of Infectious Diseases*, 1996; 174(Suppl 2): S150–61.

Gregson S, Nyamukapa CA, Garnett GP, Mason PR, Zhuwau T, Caraël M et al. Sexual Mixing Patterns and Sex-Differentials in Teenage Exposure to HIV Infection in Rural Zimbabwe. *Lancet*, 2002; 359(9321): 1896–903.

Hallett TB, Gregson S, Lewis JJC, Lopman BA & Garnett GP. Behaviour Change in Generalised HIV Epidemics: Impact of Reducing Cross-Generational Sex and Delaying Age at Sexual Debut. *Sexually Transmitted Infections*, 2007; 83(Suppl 1): i50–4.

Hope R. *Addressing Cross-Generational Sex: A Desk Review of Research and Programs*. 2007. Washington, DC: Population Reference Bureau.

Hosegood V, McGrath N & Moultrie T. Dispensing with Marriage: Marital and Partnership Trends in Rural Kwazulu-Natal, South Africa 2000-2006. *Demographic Research*, 2009; 20: 279–312.

Kelly RJ, Gray RH, Sewankambo NK, Serwadda D, Wabwire-Mangen F, Lutalo T et al. Age Differences in Sexual Partners and Risk of HIV-1 Infection in Rural Uganda. *Journal of Acquired Immune Deficiency Syndromes*, 2003; 32(4): 446–51.

Kwa-Zulu Natal Department of Health. *Sugar Daddy Campaign*. 2012. Available from: http://www.kznhealth.gov.za/sugardaddy.htm. [Last Accessed March 13 2013].

Leclerc-Madlala S. Age-Disparate and Intergenerational Sex in Southern Africa: The Dynamics of Hypervulnerability. *AIDS*, 2008; 22(Suppl 4): S17–25.

Luke N. Age and Economic Asymmetries in the Sexual Relationships of Adolescent Girls in Sub-Saharan Africa. *Studies in Family Planning*, 2003; 34(2): 67–86.

Luke N. Confronting the 'Sugar Daddy' Stereotype: Age and Economic Asymmetries and Risky Sexual Behavior in Urban Kenya. *International Family Planning Perspectives*, 2005; 31(1): 6–14.

Ott MQ, Bärnighausen T, Tanser F, Lurie MN & Newell M-L. Age-Gaps in Sexual Partnerships: Seeing Beyond 'Sugar Daddies'. *AIDS*, 2011; 25(6): 861–3.

Pettifor AE, Rees HV, Kleinschmidt I, Steffenson AE, MacPhail C, Hlongwa-Madikizela L et al. Young People's Sexual Health in South Africa: HIV Prevalence and Sexual Behaviors from a Nationally Representative Household Survey. *AIDS*, 2005; 19(14): 1525–34.

PSI. *Cross-Generational Sex*. Available from: http://www.psi.org/our-work/healthy-lives/interventions/cross-generational-sex. [Last Accessed March 13 2013].

Reynolds L, Cousins T, Newell ML & Imrie J. The Social Dynamics of Consent and Refusal in HIV Surveillance in Rural South Africa. *Social Science and Medicine*, 2013; 77: 118–25.

Tanser F, Hosegood V, Bärnighausen T, Herbst K, Nyirenda M, Muhwava W et al. Cohort Profile: Africa Centre Demographic Information System (ACDIS) and Population-Based HIV Survey. *International Journal of Epidemiology*, 2008; 37(5): 956–62.

UNAIDS. *2012 Report on the Global AIDS Epidemic*. 2012. Geneva, Switzerland: Joint United Nations Programme on HIV/AIDS.

Wellings K, Collumbien M, Slaymaker E, Singh S, Hodges Z, Patel D et al. Sexual Behaviour in Context: A Global Perspective. *Lancet*, 2006; 368(9548): 1706–28.

WHO. *National AIDS Programmes: A Guide to Indicators for Monitoring and Evaluating National HIV/AIDS Prevention Programmes for Young People*. 2004. Geneva, Switzerland: World Health Organization.

Wingood GM & DiClemente RJ. Application of the Theory of Gender and Power to Examine HIV-Related Exposures, Risk Factors, and Effective Interventions for Women. *Health Education and Behavior*, 2000; 27(5): 539–65.

Woolf SE & Maisto SA. Gender Differences in Condom Use Behavior? The Role of Power and Partner-Type. *Sex Roles*, 2008; 9-10(58): 689–701.

Wyrod R, Fritz K, Woelk G, Jain S, Kellogg T, Chirowodza A et al. Beyond Sugar Daddies: Intergenerational Sex and AIDS in Urban Zimbabwe. *AIDS and Behavior*, 2011; 15(6): 1275–82.

Figure 3.1: Age-disparity between female respondent and most recent male sexual partner

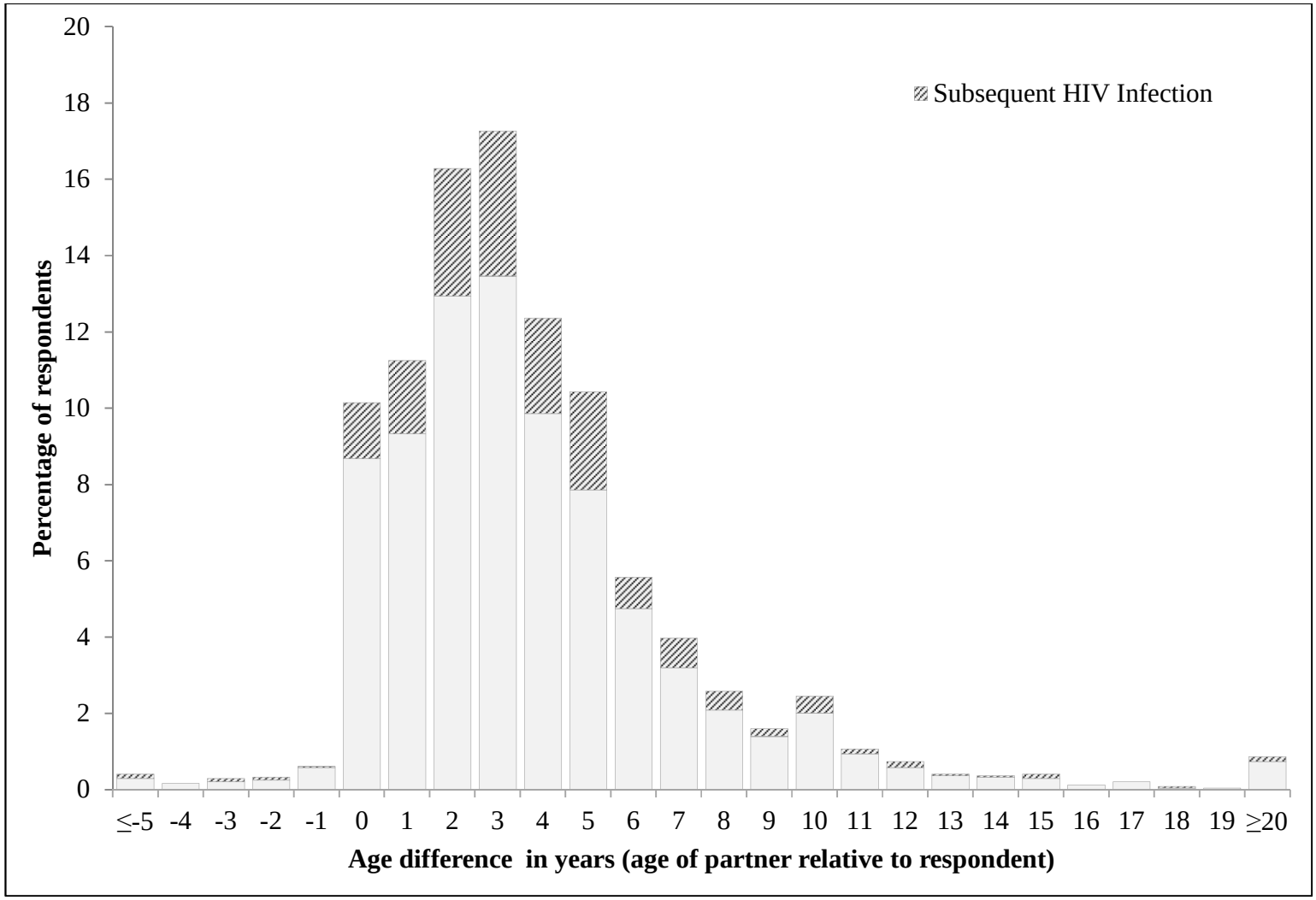

Table 3.1: National Strategic Plans containing references to "Sugar Daddy" or age-disparate sexual relationships

	% HIV prevalence (ages 15-49)*	NSP period	Age-disparity or "sugar daddy" cited as risk factor	Justification given	Policies mandated	Explicit reduction target set
Swaziland	26·0	2009-14	Yes	Prevalence of inter-generational sex, especially amongst out-of-school young women	Mandated, no specific policy	
Botswana	23·4	2010-16	Yes	Older men have longer sexual history; Young women having intergenerational sex have lower condom use	Mandated, no specific policy	
Lesotho	23·3	2012-16	Yes	None	Yes†	Increase BCC coverage†
South Africa	17·3	2012-16	Yes	Intergenerational relationships increase risk of HIV exposure	Yes	
Zimbabwe	14·9	2011-15	Yes	Inter-generational sex is a factor that makes women and girls more vulnerable	Mandated, no specific policy	
Namibia	13·4	2011-16	Yes	"Sugar daddy" phenomenon well-known in Namibia. Inter-generational sex associated with STIs, multiple and concurrent partnerships and introduction of HIV into younger cohort	Yes †	
Zambia	12·5	2011-15	Yes	-	Yes	
Mozambique	11·3	2010-14	Yes	Sexual relations between individuals from different generations associated with transactional and less-safe sex		
Malawi	10·0	2009-13†	Yes	-	Yes	
Uganda	7·2	2011/12 -14/15	Yes	Prevalence of cross-generational sex (15-19 year-old women with men at least 10 years older) †	Yes	Reduce cross-generational sex by 50% by 2015†
Kenya	6·2	2008/09 -12/13	No	Transactional relationships may be important but "sugar daddy" relationships are possibly less frequent than generally thought		
Tanzania	5·8	2008-12	Yes	Prevalence of cross-generational sex, which is linked to transactional sex and multiple concurrent partnerships	Yes	Reduce cross-generational sex by 50% for teenage girls by 2025
Cameroon	4·6	2011-15	Yes	-	No	
Central African Republic	4·6	2006-10	No			
Nigeria	3·7	2010-15	Yes	-	Yes	
Togo	3·4	2012-15	Yes	Prevalence of intergenerational relationships is increasing, and is a source of infections among young women	No	
Congo	3·3	2009-13	No			
Chad	3·1	2007-11	Yes	-	No	
Côte d'Ivoire	3·0	2011-15	Yes†	Young people in sexual relationships with older partners lack negotiating power	Mandated, no specific policy	80% of youth adopt lower risk sexual behaviours
Rwanda	2·9	2009-12	Yes	Disparity in HIV prevalence amongst 20-24 year old men and women attributed to cross-generational sex	Yes	

Based on the most recent NSP available for countries with adults HIV prevalence over 2% in 2011. No NSPs published since 2002 were available for Angola, Equatorial Guinea, Gabon or Guinea-Bissau.

BCC: Behaviour change communication. NSP: National Strategic Plan. * HIV prevalence for 2011 from UNAIDS Report on the global AIDS epidemic 2012.(UNAIDS, 2012). † From an accompanying National Prevention Strategy rather than National Strategic Plan.

Table 3.2: Baseline characteristics of the study sample of 15-29 year old women

	All	15-19	20-24	25-29
Sample size	2,444	1,112	982	350
Number of subsequent seroconversions	458	136	251	71
Age at baseline	20 (18 to 23)	18 (17 to 19)	21 (20 to 23)	27 (25 to 28)
Partner age-disparity in most recent relationship	3 (2 to 5)	3 (2 to 5)	3 (1 to 5)	4 (2 to 6)
Highest educational attainment				
None or Primary (0-7 years)	10	14	4	11
Secondary (8-12 years)	86·5	84·7	92·0	77·1
Tertiary	3·8	1·4	3·8	11·7
Household wealth quintile				
Lowest	19·5	19·1	19·5	20·9
2nd lowest	25·7	23·9	26·6	29·1
Middle	24·9	24·4	25·7	24·6
2nd highest	18·0	19·6	17·2	15·1
Highest	11·9	13·0	11·1	10·3
Marital status				
Never Married	87·6	93·5	88·4	66·3
Engaged	10·1	6·2	10·2	22·0
Married	2·3	0·3	1·4	11·4
Divorced/Separated/Widowed	0·0	0·0	0·0	0·3
Age at sexual debut	17 (16 to 18)	16 (15 to 17)	18 (16 to 19)	18 (16 to 19)
Multiple partners in past 12 months	1·3	1·2	1·8	0·3
Casual partner in past 12 months	4·3	5·8	3·4	2·0
Lowest condom use level in relationships in past 12 months				
Never	49·3	45·1	50·7	59·1
Sometimes	27·1	25·0	29·0	28·3
Always	23·6	29·9	20·3	12·6

Figures for categorical data are percentages; figures for continuous data are medians and (Interquartile ranges)

Table 3.3: Multivariable Cox proportional hazards models of HIV acquisition (hazard ratios and 95% confidence intervals)

	Model 1		Model 2†		Model 3		Model 4	
Age-disparity (one year increase in partner's age)	1·00	(0·97 - 1·03)			1·00	(0·98 - 1·03)	1·00	(0·98 - 1·03)
Respondent aged 15-19			1·03	(0·98 - 1·07)				
Respondent aged 20-24			0·96	(0·90 - 1·01)				
Respondent aged 25-29			0·98	(0·92 - 1·05)				
Age of respondent (centred at 15 years old)*								
Age	1·18	(1·07 - 1·29)	1·09	(0·95 - 1·24)	1·19	(1·08 - 1·31)	1·21	(1·09 - 1·33)
Age squared	0·73	(0·67 - 0·80)	0·73	(0·67 - 0·81)	0·73	(0·67 - 0·81)	0·74	(0·67 - 0·81)
Age cubed	1·11	(1·05 - 1·17)	1·13	(1·06 - 1·21)	1·11	(1·05 - 1·17)	1·10	(1·05 - 1·17)
Highest educational attainment								
None or Primary (0-7 years)					1·36	(0·98 - 1·89)	1·29	(0·93 - 1·81)
Secondary (8-12 years)					1·00		1·00	
Tertiary					0·81	(0·50 - 1·32)	0·82	(0·51 - 1·33)
Household wealth quintile								
Lowest					1·38	(0·95 - 2·00)	1·38	(0·94 - 2·01)
2nd lowest					1·34	(0·94 - 1·91)	1·36	(0·95 - 1·94)
Middle					1·54	(1·09 - 2·18)	1·56	(1·10 - 2·22)
2nd highest					1·36	(0·94 - 1·95)	1·36	(0·94 - 1·96)
Highest					1·00		1·00	
Current marital status of respondent								
Never Married					1·00		1·00	
Engaged					0·96	(0·69 - 1·32)	0·97	(0·71 - 1·34)
Married					0·10	(0·01 - 0·72)	0·10	(0·01 - 0·73)
Divorced/Separated/Widowed					13·41	(1·84 - 98·0)	12·34	(1·68 - 90·7)
Age at sexual debut (one-year increment)							0·98	(0·92 - 1·04)
Any casual partner in past 12 months							1·12	(0·72 - 1·74)
Multiple partners in past 12 months							2·10	(1·16 - 3·81)
Lowest condom use level in relationships in past 12 months								
Never							1·00	
Sometimes							0·89	(0·71 - 1·11)
Always							1·08	(0·86 - 1·37)
Akaike Information Criteria	6,482		6,484		6,472		6,474	

For all models, n=2,444, time at risk = 5,913 person-years and there were 458 seroconversions. All models contain indicator variables for year of observation (not shown).

* The coefficient on Age squared represents a 10-unit change in this variable; the coefficient on Age cubed represents a 100-unit change in this variable.

† This model also contains indicator variables for age categories. A joint test for equality on the three age by age-disparity interaction terms was not statistically significant (χ^2 = 2·26, p-value: 0·324).

Supplementary Table 3.4: Cox proportional hazards models for appropriate functional form (hazard ratios and 95% confidence intervals)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Respondent's age (in years, centred at age 15)							
Linear	0·94 (0·91 - 0·97)	1·20 (1·09 - 1·33)	1·18 (1·07 - 1·29)	1·11 (0·95 - 1·30)	1·18 (1·07 - 1·29)	1·18 (1·07 - 1·29)	1·18 (1·07 - 1·29)
Quadratic†		0·98 (0·97 - 0·99)	0·97 (0·96 - 0·98)	0·97 (0·96 - 0·98)	0·97 (0·96 - 0·98)	0·97 (0·96 - 0·98)	0·97 (0·96 - 0·98)
Cubic†			2·81 (1·65 - 4·78)	10·94 (0·51 - 234·5)	2·81 (1·65 - 4·78)	2·81 (1·65 - 4·78)	2·81 (1·65 - 4·78)
Quartic†				0·44 (0·07 - 2·74)			
Most recent partner's age-disparity (in years, centred at 0)							
Linear					1·00 (0·97 - 1·03)	1·01 (0·96 - 1·06)	1·00 (0·94 - 1·07)
Quadratic†						1·00 (1·00 - 1·00)	1·00 (0·99 - 1·01)
Cubic†							0·95 (0·73 - 1·23)
Akaike Information Criteria	6,510	6,490	6,478	6,480	6,480	6,482	6,484

For all models, n=2,444, time at risk = 5,913 person-years and there were 458 seroconversions. All models contain indicator variables for year of observation (not shown).

† Coefficients for squared, cubic and quartic terms represent 10, 100 and 1000 unit change in the respective variables.

Supplementary Table 3.5: Descriptive statistics for multiply imputed datasets

	Multiply imputed dataset	Complete case dataset	Missingness (%)	Change (%) in incidence rate**
All women aged 15-29 years old				
Time at risk (person-years)	7,722	5,913	23.4	
Number of respondents at baseline	3,358	2,444	27.2	
Number of seroconversions	584	458		
Incidence rate (per 100 person-years)	7·56	7·75		-2·5
95% confidence interval	(6·97 - 8·20)	(7·07 - 8·49)		
15-19 years old				
Time at risk (person-years)	2,599	1,746	32·8	
Number of respondents at baseline	1,770	1,112	37·2	
Number of seroconversions	222	136		
Incidence rate (per 100 person-years)	8·54	7·79		8·8
95% confidence interval	(7·49 - 9·74)	(6·59 - 9·22)		
20-24 years old				
Time at risk (person-years)	3,668	2,907	20·7	
Number of respondents at baseline	1,128	982	12·9	
Number of seroconversions	285	251		
Incidence rate (per 100 person-years)	7·77	8·63		-11·1
95% confidence interval	(6·92 - 8·73)	(7·63 - 9·77)		
25-29 years old				
Time at risk (person-years)	1,455	1,260	13·4	
Number of respondents at baseline	460	350	23·9	
Number of seroconversions	77	71		
Incidence rate (per 100 person-years)	5·29	5·63		-6·4
95% confidence interval	(4·23 - 6·62)	(4·46 - 7·11)		

	Incidence rate for all valid responses	Missingness (%)*	
Age disparity of partner in most recent relationship	7·69	17·6	-1·7
Highest educational attainment	7·56	-	0·0
Household wealth quintile	7·38	4·9	2·4
Marital status	7·50	0·4	0·8
Age at sexual debut	7·59	0·7	-0·4
Multiple partners in past 12 months	7·50	5·0	0·8
Casual partner in past 12 months	7·56	-	0·0
Lowest condom use level in relationships in past 12 months	7·49	0·9	0·9

* Missingness for covariates is the amount of person-time with missing values prior to imputation for each variable, as a proportion of the 7,722 years of person-time in the Multiply Imputed dataset.
** Change in the incidence rate comparing all person-time in the imputed dataset to all person-time in the complete-case dataset (in the top half of the table this is the complete-case dataset for all independent variables; in the bottom half it is the complete-case dataset for each individual variable).

The dataset for multiple imputation consisted of all person-time for which there was a valid HIV test result at the beginning and end of the observation period, where the respondent had participated in the General Health module by the beginning of the period, and where we were unable to exclude the possibility that they had yet to reach their sexual debut (i.e. women who reported never having had sex, or had never reported any sexual partner up to the beginning of the period).

Missing data were imputed 20 times using chained imputations via the MI Impute command in SAS. The chained imputation process used all variables included in analyses in this paper, plus other variables that might be expected to be correlated with sexual and socio-demographic variables: urban/suburban/rural household; distance to nearest health clinic, major road, minor road, primary and secondary school; Isigodi (traditional Zulu area) of household; currently economically active (yes/no); and year of visit.

Supplementary Table 3.6: Regression results from multiply imputed dataset (hazard ratios and 95% confidence intervals)

	Model 1		Model 2†		Model 3		Model 4	
Age-disparity (one year increase in partner's age)	0·96	(0·85 - 1·08)			0·98	(0·86 - 1·11)	0·97	(0·86 - 1·11)
Respondent aged 15-19			1·06	(0·86 - 1·29)				
Respondent aged 20-24			0·88	(0·73 - 1·05)				
Respondent aged 25-29			0·99	(0·76 - 1·28)				
Age of respondent (centred at 15 years old)*								
Age	0·85	(0·67 - 1·07)	0·84	(0·66 - 1·06)	0·87	(0·68 - 1·10)	0·88	(0·69 - 1·12)
Age squared	1·01	(0·97 - 1·05)	1·01	(0·97 - 1·05)	1·01	(0·97 - 1·05)	1·01	(0·97 - 1·05)
Age cubed	0·48	(0·07 - 3·16)	0·55	(0·07 - 3·87)	0·49	(0·06 - 3·60)	0·56	(0·07 - 4·24)
Highest educational attainment								
None or Primary (0-7 years)					1·17	(0·86 - 1·58)	1·16	(0·85 - 1·59)
Secondary (8-12 years)					1·00		1·00	
Tertiary					0·74	(0·46 - 1·18)	0·77	(0·48 - 1·23)
Household wealth quintile								
Lowest					1·39	(0·98 - 1·97)	1·40	(0·98 - 1·99)
2nd lowest					1·31	(0·94 - 1·82)	1·33	(0·94 - 1·85)
Middle					1·46	(1·05 - 2·02)	1·49	(1·06 - 2·08)
2nd highest					1·34	(0·95 - 1·88)	1·39	(0·98 - 1·95)
Highest					1·00		1·00	
Current marital status of respondent								
Never Married					1·00		1·00	
Engaged					0·98	(0·73 - 1·31)	0·90	(0·66 - 1·22)
Married					0·18	(0·04 - 0·74)	0·19	(0·04 - 0·78)
Divorced/Separated/Widowed					1·76	(0·24 - 12·6)	2·93	(0·40 - 21·3)
Age at sexual debut (one-year increment)							1·01	(0·95 - 1·06)
Any casual partner in past 12 months							1·11	(0·89 - 1·37)
Multiple partners in past 12 months							2·30	(1·31 - 4·02)
Lowest condom use level in relationships in past 12 months								
Never							1·00	
Sometimes							0·90	(0·72 - 1·10)
Always							1·11	(0·89 - 1·37)
Sample size	7,722		7,722		7,317		7,122	
Person-time at risk	2,972		2,972		2,870		2,807	
Number of subsequent seroconversions	584		584		537		526	

All models contain indicator variables for year of observation (not shown).

* The coefficient on the variable "age squared" represents a 10-unit change in this variable; the coefficient on the variable "age cubed" represents a 100-unit change in this variable.

† This model also contains indicator variables for age categories. A joint test for equality on the three age by age-disparity interaction terms was not statistically significant (χ^2 = 1·90, p-value: 0·386).

Supplementary Table 3.7: Cox proportional hazards models for sensitivity analyses of results using binary age-disparity categories (hazard ratios and 95% confidence intervals)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Age of respondent						
Linear	1.18 (1.07 - 1.29)	1.18 (1.07 - 1.30)	1.18 (1.07 - 1.29)	1.19 (1.08 - 1.30)	1.18 (1.07 - 1.29)	1.18 (1.07 - 1.29)
Quadratic	0.73 (0.67 - 0.80)	0.73 (0.67 - 0.80)	0.73 (0.67 - 0.80)	0.73 (0.67 - 0.80)	0.73 (0.67 - 0.80)	0.73 (0.67 - 0.81)
Cubic	1.11 (1.05 - 1.17)	1.11 (1.05 - 1.17)	1.11 (1.05 - 1.17)	1.10 (1.05 - 1.17)	1.11 (1.05 - 1.17)	1.11 (1.05 - 1.17)
Age-disparity ≥5 years	0.98 (0.81 - 1.20)					
Respondent aged 15-19		1.00 (0.71 - 1.42)				
Respondent aged 20-24		0.98 (0.75 - 1.28)				
Respondent aged 25-29		0.96 (0.60 - 1.52)				
Age-disparity ≥10 years			0.98 (0.67 - 1.43)			
Respondent aged 15-19				1.38 (0.68 - 2.81)		
Respondent aged 20-24				0.77 (0.42 - 1.41)		
Respondent aged 25-29				1.05 (0.54 - 2.04)		
Age-disparity ≥20 years					0.61 (0.15 - 2.46)	
Respondent aged 15-19						1.69 (0.24 - 12.10)
Respondent aged 20-24						0.52 (0.07 - 3.70)
Respondent aged 25-29						- (0.00 - 0.00)
Akaike Information Criteria	6,474	6,480	6,474	6,478	6,476	6,478

For all models, n=2,444, time at risk = 5,913 person-years and there were 458 seroconversions.
All models contain indicator variables for year of observation (not shown).

Supplementary Table 3.8: Cox proportional hazards models containing interactions of socio-demographic variables and age-disparity measures (hazard ratios and 95% confidence intervals)

Functional form of age-disparity	Continuous (one-year increase in partner's age)	Binary, partner ≥5 years older	Binary, partner ≥10 years older	Binary, partner ≥20 years older
Current marital status				
Never married	1.01	1.09	1.21	0.29
	(0.93 - 1.09)	(0.57 - 2.08)	(0.40 - 3.69)	(0.02 - 4.77)
Engaged, married or formerly married	1.00	0.93	0.87	1.59
	(0.92 - 1.08)	(0.51 - 1.72)	(0.31 - 2.47)	(0.22 - 11.73)
χ^2 test of equality	0.02	0.06	0.09	0.56
p-value	0.90	0.81	0.76	0.45
Highest educational attainment				
None or Primary (0-7 years)	0.99	1.14	0.98	
	(0.92 - 1.07)	(0.62 - 2.11)	(0.38 - 2.54)	
Secondary (8-12 years)	1.01	1.03	1.11	1.01
	(0.98 - 1.04)	(0.83 - 1.28)	(0.73 - 1.69)	(0.25 - 4.08)
Tertiary	0.84	0.42		
	(0.70 - 1.01)	(0.14 - 1.27)		
χ^2 test of equality	3.54	2.63	0.05	-
p-value	0.17	0.27	0.82	
Household wealth quintile				
Lowest	0.97	0.58	0.75	
	(0.90 - 1.04)	(0.34 - 1.00)	(0.30 - 1.86)	
2nd lowest	0.99	1.04	1.35	1.28
	(0.93 - 1.06)	(0.70 - 1.54)	(0.68 - 2.68)	(0.18 - 9.30)
Middle	1.02	1.56	1.32	1.29
	(0.98 - 1.07)	(1.09 - 2.23)	(0.69 - 2.53)	(0.18 - 9.28)
2nd highest	0.99	1.01	0.83	
	(0.92 - 1.07)	(0.63 - 1.60)	(0.30 - 2.28)	
Highest	0.87	0.52	0.39	
	(0.77 - 0.98)	(0.24 - 1.12)	(0.05 - 2.86)	
χ^2 test of equality	5.77	12.74	2.66	0.00
p-value	0.22	0.01	0.62	1.00

This table contains summary results for 12 separate regression models. For all models, n=2,444, time at risk = 5,913 person-years and there were 458 seroconversions. All models contain indicator variables for year of observation, age of respondent (linear, quadratic and cubic terms centred at age 15) and all socio-demographic and behavioural covariates from the main analysis.

Empty cells in this table reflect categories in which there were so few observed individuals with age-disparate partners that they had to be removed in order for models to converge.

The χ^2 tests have k-1 degrees of freedom, where k is the number of categories being compared.