THE INFLUENCE OF SCHOOL ENROLLMENT AND ALCOHOL OUTLETS ON SEXUAL RISK AMONG RURAL SOUTH AFRICAN YOUNG WOMEN

Molly Sears Rosenberg

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology.

Chapel Hill
2014

Approved by:

Audrey Pettifor

Annelies Van Rie

Harsha Thirumurthy

Michael Emch

William Miller

ABSTRACT

Molly Sears Rosenberg: The Influence of School Enrollment and Alcohol Outlets
on Sexual Risk among Rural South African Young Women
(Under the direction of Audrey Pettifor)

Young women in South Africa are at extremely high risk for sexually transmitted infections and teen pregnancy. The identification of new intervention targets is critical to reduce the burden of these outcomes, yet place-based structural determinants of sexual risk have not been previously explored in this population. In Aim 1, we estimated the effect of school enrollment on teen pregnancy using longitudinal census data for 15,457 South African young women aged 12-18 years. A Cox proportional hazard model was constructed to compare the hazard of pregnancy between school enrollees and dropouts. Our findings suggested that young women who remained in school were at lower risk for teen pregnancy [aHR (95% CI): 0.57 (0.50, 0.65)]. In Aim 2, we estimated the association between visits to alcohol outlets and sexual risk using cross-sectional data from a sample of 2,533 South African young women. We also tested for interaction by alcohol consumption. Visiting alcohol outlets was associated with having more sex partners [aOR, one versus zero partners (95% CI): 1.51 (1.21, 1.88)], more unprotected sex acts [aOR, one versus zero acts (95% CI): 2.28 (1.52, 3.42)], higher levels of transactional sex [aOR (95% CI): 1.63 (1.03, 2.59)], and HSV-2 infection [aOR (95% CI): 1.30 (0.88, 1.91)]. Generally, the dual combination of exposure to alcohol outlets and alcohol consumption yielded stronger associations with the sexual risk outcomes than anticipated given the associations observed with each risk factor alone. In Aim 3, we estimated the association between number of alcohol outlets per village and prevalent HSV-2 infection using cross-sectional data from a sample of 2,174 young women living across 24 villages in rural South Africa. We used generalized estimating equations with log links to account for the clustered nature of the data. Young women who lived in villages with more alcohol outlets were more likely to be infected with HSV-2 [PR (95% CI): 1.08 (1.01, 1.15)]. Overall, the findings from all three aims suggest that place-based exposures may be important determinants of sexual risk among young women in South Africa.

In loving memory of Marcia Sears
who showed me that she could do anything
and told me I could, too.

ACKNOWLEDGMENTS

There were so many people who made this dissertation successful. I am especially indebted to my dissertation committee and the investment of their time, expertise, guidance, and patience. My dissertation research was supported by the Carolina Population Center, providing me not only with financial support, but also with a unique interdisciplinary research environment that inspired me to think differently. The administrative and emotional assistance of Nancy Colvin, Carmen Woody, and Jan Hendrickson-Smith have been invaluable. And, without the almost absurd levels of support from my fellow Epidemiology students, I almost certainly would still be stuck in the computer lab struggling with some SAS error codes.

For Aim 1, I used data originally collected by the Agincourt Health and Demographic Surveillance Site. These data were collected over an eleven year period and required an unbelievable amount of hard work and infrastructure development from which my research benefited. For Aims 2 and 3, I used data originally collected at baseline in the HIV Prevention Trial Network 068 Study. These high quality data were collected under a strict randomized controlled trial protocol requiring the investment of extensive research support staff time and energy. And, of course, I am deeply indebted to the research participants themselves, who volunteered their valuable time to make research like mine possible.

My family is my source of unconditional support. I would especially like to thank my parents for instilling in me a love of laughter and a love of learning (in that order). Brian Seavey did everything from reading my manuscript drafts to doing the dishes a disproportionate amount of the time. I wish I could grant him an honorary doctorate. Mae Seavey and her soon-to-be-born baby sister remind me not to take myself too seriously but also inspire me to do my part to make the world a healthier place.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACASI=audio-computer assisted self-interview

AIDS=acquired immunodeficiency syndrome

ANOVA= analysis of variance

AO=alcohol outlet

CI=confidence interval

GEE=generalized estimating equation

HDSS=health and demographic surveillance site

HIV=human immunodeficiency virus

HPTN=HIV Prevention Trials Network

HR=hazard ratio (aHR= adjusted hazard ratio)

HSV-2=herpes simplex virus type 2

ICC= intra-cluster coefficient

IQR= interquartile range

IRR= incidence rate ratio

LRT= likelihood ratio test

LTFU= loss to follow-up

MI= multiple imputation

MLM= multi-level modeling

OR=odds ratio (aOR= adjusted odds ratio)

PH= proportional hazards

PLACE= Priorities for Local AIDS Control Efforts

PR= prevalence ratio

PY= person-years

SD= standard deviation

SES=socio-economic status

STI= sexually transmitted infection

CHAPTER I: Specific Aims

This study was designed to examine how the places young South African women spend their time may influence their sexual risk. We hypothesized that spending time in structured and supervised places would be associated with safer behaviors and related outcomes, either by reducing opportunities to engage in unsafe behaviors or by putting youth in contact with safer friends and partners. Similarly, unstructured and unsupervised activity could provide opportunities to engage in unsafe behaviors or put youth in contact with unsafe friends and partners.

An important structured place where many young women spend time is school. We explored the influence of school enrollment on one specific outcome of risky sexual behavior – teen pregnancy. Using a longitudinal dataset from rural South Africa, we compared pregnancy rates at times when young women are enrolled in school to pregnancy rates at times when young women are not enrolled in school.

A potentially significant risky activity for this population is visiting alcohol outlets. Visits to alcohol outlets could allow for periods of unsupervised and unstructured time and provide environments where alcohol consumption and sexual risk behaviors may be promoted. To explore the association between frequency of visits to alcohol vendors and sexual risk, we performed a cross-sectional ancillary study to HPTN 068: *The Effects of Cash Transfer and Community Mobilization in Young South African Women Study* in rural South Africa.

Finally, characteristics of home communities may play a role in shaping the specific activities and, therefore, sexual risk of residents. The HPTN 068 study population was enrolled within a well-defined research site across 24 villages. We were specifically interested in exploring whether alcohol outlet accessibility within each village influenced sexual risk as measured by individual-level prevalent herpes simplex virus type 2 (HSV-2) infection.

Much previous research has focused on identifying individual sexual risk behaviors, but the structural and place-based characterization of sexual risk has been largely understudied, particularly within our target population. The results of these analyses provide insight into distal determinants of sexual risk outcomes as well as information that may be useful to better inform and target preventive interventions.

Aim 1: Estimate the effect of **school enrollment** on **incident teen pregnancy** among rural South African young women.

Hypothesis: Pregnancy rates will be lower among those enrolled in school compared to those not enrolled in school.

Methods: We constructed a cohort from the Agincourt health and socio-demographic surveillance site of nulliparous young women aged 12 to 18 followed between 2000 and 2012. In this longitudinal analysis, we built Cox proportional hazards models to estimate the hazard ratio comparing pregnancy rates among young women currently enrolled in school compared to those not currently enrolled in school. We treated this school enrollment exposure as time-varying and controlled for important time-varying and time-fixed confounders.

Aim 2: Estimate the association between **frequency of visits to alcohol outlets** and four sexual risk outcomes (**unprotected vaginal sex acts, number of sex partners,**

transactional sex, and **prevalent HSV-2 infection**) among rural South African young women.

Hypothesis: Increasing number of alcohol outlet visits will be associated with increased sexual risk behaviors and increased likelihood of prevalent HSV-2 infection.

Methods: In this cross-sectional analysis, we built logistic regression models to estimate the associations between number of alcohol outlet visits in the last six months and two dichotomous outcomes: transactional sex with most recent partner and prevalent HSV-2 infection. We built multinomial logistic regression models to estimate the association between number of tavern visits and two categorical outcomes: unprotected vaginal sex acts in the last three months (categorized as 0, 1, and 2+ acts) and number of sex partners in the last three months (categorized as 0, 1, and 2+ partners). The exposure and behavioral outcome variables were self-reported in the baseline questionnaire for all young women enrolled in HPTN 068. Biological testing for prevalent HSV-2 was also conducted for all participants at baseline.

Aim 3: Estimate the association between village-level **number of alcohol outlets** and individual-level **prevalent HSV-2 infection** among rural South African young women.

Hypothesis: We hypothesize that living in villages with more alcohol outlets will be associated with higher likelihood of prevalent HSV-2 infection.

Methods: Using a generalized estimating equation with log link, we modeled the association between number of alcohol outlets in home village and individual HSV-2 sero-status. Number of alcohol outlets per village were reported by key informants in a village asset mapping exercise associated with HPTN 068 baseline procedures in 2011. Biological testing for HSV-2 was conducted at baseline for each young woman enrolled in the study.

CHAPTER II: Background

A better understanding of the determinants of sexual risk is critical to informing new preventive interventions to reduce the burden of related health outcomes, including unintended pregnancy, HIV infection, and infection with other STIs. We hypothesize that an individual's routine activity pattern can influence her sexual risk. Below, we justify this hypothesis by: briefly summarizing the current public health burden of sexual risk outcomes with a focus on young women in South Africa; setting up the conceptual and theoretical framework for the association between routine activity patterns and sexual risk; and discussing the literature to-date on the associations between school enrollment, visits to alcohol outlets, alcohol outlet accessibility, and sexual risk.

1 Overview

The public health burden of HIV remains high throughout the world, particularly among young South African women. There are currently 34.0 million people living with HIV worldwide, according to most recent estimates, and the number has risen continuously since 1990.(1) South Africans face a burden from HIV/AIDS at a level largely unmatched in the rest of the world. Overall, the national HIV prevalence was estimated to be over 15% in 2011 with over 5.5 million people living with the disease.(2) Young women in South Africa are at exceptionally high risk for HIV infection. They are infected at a rate over three times that of young men, and, by age 21, nearly a third of young women are HIV positive.(3, 4) The questions remain: Why are these young women so strongly affected by HIV and what can be done to prevent the spread of the disease in this vulnerable population?

Similarly, there is a significant public health burden associated with teen pregnancy; this burden is particularly high among adolescents in sub-Saharan Africa. Teen childbearing has been associated with negative health and social outcomes for both teen mothers (maternal mortality, lower self-esteem, lower educational attainment, higher poverty levels, social stigma, and family-imposed sanctions) and their children (increased risk of preterm delivery, low birth weight, and infant mortality).(5-8) The negative consequences for children of teen mothers persist throughout their own childhood and early adulthood as they also tend to have higher poverty levels, attain lower educational levels, and are at higher risk of becoming teen parents themselves.(5) Globally, over 16 million births occur to mothers under the age of 20 every year, and nearly all (95%) occur in developing countries.(9) In South Africa, over 10% of all births are to teen mothers,(10) and, by the time they reach age 20, nearly one-third of all young women will have borne a child.(11) Successful preventive interventions for teen pregnancy, particularly in areas where it is most common, could potentially prevent a large burden of negative health outcomes.

Much previous research has been conducted to identify individual-level characteristics associated with sexual risk outcomes. In particular, increased risk of HIV has been associated with demographic factors such as age,(12) gender,(13) and socioeconomic status;(14) and behavioral factors such as early coital debut,(15) number of sex partners,(16) condom use,(16) and other STI diagnosis.(17) Becoming a teen mother has been associated with demographic factors such as race/ethnicity, large household size, and not living with biological father; and behavioral factors such as frequent sex and low contraceptive use.(7, 18, 19)

However, many of these individual-level characteristics are difficult to modify, or are non-modifiable, with little use in informing preventive interventions. Although more distal risk factors such as poverty and low education have also been associated with both HIV (14, 20, 21) and teen pregnancy,(22, 23) other structural and contextual risk factors, particularly those with the potential to inform interventions, have been explored much less frequently.

This study explored how routine activity patterns (school enrollment and frequency of visits to alcohol-serving establishments) and community characteristics (alcohol outlet accessibility) may influence sexual risk. We drew a distinction between activities that provide structure and supervision to adolescents (school enrollment) and activities that allow for periods of unstructured and unsupervised time (visits to alcohol outlets). In general, we hypothesized that structured and supervised activities would be associated with lower sexual risk and unstructured and unsupervised activities would be associated with greater sexual risk.

2 Conceptual and theoretical framework

Figure 2.1 is a conceptual diagram of the association between routine activities and sexual risk, representing the contextual factors that may influence routine activities and the potential mechanisms through which we hypothesize routine activities may influence sexual risk. The first column depicts the contextual factors we hypothesize influence activity participation. The propensity for an individual to participate in any specific activity is likely shaped by factors at the community-level (does the community provide opportunities to participate in the activity?), household-level (does the family encourage or discourage certain activities, either directly or because of characteristics associated with participation in the activity?), and individual-level (does the individual have certain characteristics that are associated with participation in the activity?).

The third column depicts our hypothesis that routine activities can influence sexual risk through three separate, though not mutually exclusive mechanisms. First**, participation in specific activities may improve one's socio-economic status (SES), which could offset some of the financial motivations to engage in risky sexual activity.** However, as the timeframe in which we might anticipate this kind of mechanism to influence sexual risk is likely long in duration and we are studying more temporally-proximate outcomes in a population of young women, this mechanism is not explored further here.

Second, **routine activities may influence sexual risk outcomes directly by providing periods of structured and supervised time during which sexual risk behavior is unlikely to occur, or by providing periods of unstructured and unsupervised time during which sexual risk behavior may be more likely to occur.** The justification for this proposed direct effect is grounded in a synthesis of Social Control Theory,(24, 25) and the routine activity/time-use perspective.(26-29) Both of these theories have been used primarily in criminology to explain delinquent behaviors, but their applicability can be extended to sexual behavior.(30) In general, Social Control Theory posits that individuals may choose to exhibit a wide range of behaviors, but this range can be limited by social forces. The routine activity perspective adds to this framework by proposing that activities involving unstructured socializing with peers (in the absence of authority figures) are the least limited by social forces, and allow time and opportunity for deviant behavior to occur.

Finally**, participation in specific activities may change one's social and sexual network.** If the co-participants are generally safer (e.g. younger, more educated) and become integrated into the network that influences a participant's sexual behavior, then one might expect decreased sexual risk for the participant. If the co-participants are generally riskier (e.g. older, less educated), then one might expect increased sexual risk for the participant. The justification for this proposed mechanism is grounded in Social Cognitive Theory.(31) This theory posits that people learn information and change their own behaviors by viewing others. Peers can be strongly influential providers of sexual information and sexual norms to adolescents; more traditional sources of general information (parents and schools) often do an inadequate job of frankly discussing sexual matters with adolescents.(32) These peer groups, thus, can be an important source of social support for positive behavior modeling or negative behavior modeling, depending on the norms of the group.

Here, we will briefly link our specific research aims to this conceptual framework; a literature review for each aim will be presented in subsequent sections below. In both Aims 1 and 2, we were interested in estimating the association between individual routine activities and sexual risk outcomes. In Aim 1, we investigated the association between the routine activity of school enrollment and the sexual risk outcome of teen pregnancy. In Aim 2, we investigated the association between frequency of visits to alcohol outlets and sexual risk, using three behavioral outcomes and one biological outcome. The conception and design of both of these aims was grounded in the theories listed above used to explain the potential mechanisms of the association between routine activities and sexual risk, as presented in Table 2.1.

In Aim 3, we investigated the association between community-level number of alcohol outlets and the individual-level sexual risk outcome of prevalent HSV-2 infection. The design and analysis of this aim was grounded in a Neighborhoods and Health theoretical framework.(33) This framework was developed to explore how neighborhood factors influence the health of residents, with emphasis on the idea that there exist neighborhood-level contextual factors (characteristics of the neighborhood itself, such as service availability or green spaces) that can influence individual health above and beyond the neighborhood-level compositional factors (aggregate individual-level factors such as median income and unemployment rate).(34)

3 Aim 1: School enrollment and teen pregnancy

An important structured place where many young women spend time is school. In South Africa, nearly all young people are enrolled in school for most of their school-aged years; by the time they reach secondary school age, the vast majority (94%) still remain in school.(35) However, school drop-out rates begin to rise after the mandatory enrollment age of 16.(36) The average school year in South Africa has 200 school days and students typically spend five to six hours a day at school.(37, 38) It is important to note, however, that the quality of education in South Africa is substandard with reports of low performance in literacy and mathematics, high failure rates, high drop-out rates, and underqualified teachers, particularly in poor, majority black schools.(39, 40) Nonetheless, schooling may provide large periods of structure and supervision in young adults' lives, as well as teach important life skills, provide sex education, and increase future opportunities.

Lower educational attainment is associated with higher sexual risk. Lower educational attainment has been associated with sexual risk behaviors such as early sexual debut, unprotected sex, and multiple partnerships throughout sub-Saharan Africa (41-43). Lower educational attainment has also been linked to teen childbearing.(41) Finally, lower educational attainment has been associated with HIV incidence (44) and HIV prevalence.(20, 42, 45) However, there is evidence that this relationship has changed over time, as, earlier in the epidemic, higher HIV risk was associated with higher education levels.(46-50)

Studies of educational attainment and sexual risk are compromised by incomplete exposure definition and weak study designs. The highest grade reached by an individual in her lifetime may provide a good measure of the education to which she has been exposed, the life skills she has learned, and/or the increased socio-economic status that comes from increased education.(51) However, measures of educational attainment do not incorporate the potential protective effects of structured time or safer networks that may result from school enrollment. This is particularly true in studies of relatively older populations who may be long out of school.(44) Further limiting the findings from the current research is that most of these studies were performed cross-sectionally (20, 41, 42, 45) so inferences about the directionality of the relationship between education and sexual risk are limited.

School enrollment status is also associated with sexual risk. Compared to educational attainment, school enrollment status can serve as a measure of the current level of exposure to the structured activity space and safer networks provided in schools. School dropout has been linked to sexual risk behaviors like multiple partnerships, partner age, unprotected sex, and transactional sex.(52, 53) School enrollment status is also associated with lower HIV prevalence, though no studies to date have linked enrollment status to HIV incidence.(52, 54) The relationship between school enrollment and teen pregnancy is more complex. Teen pregnancy among school enrollees leads to subsequent school dropout.(55-57) Conversely, school dropout among non-pregnant teens may lead to subsequent pregnancy. Results from two interventions designed to incentivize young women in sub-Saharan Africa to stay in school longer report lower teen pregnancy rates,(58, 59) supporting similar results from a single longitudinal study in the United States from 1988-1994.(60)

Due to several methodological limitations, the nature of the relationship between school enrollment and sexual risk remains unclear. First, as with educational attainment, the majority of studies linking school enrollment and sexual risk are cross-sectional in design (52-

56), limiting the ability to make inferences about the directionality of the observed associations. However, two of these studies did try to address this issue by using detailed calendars to record event histories.(55, 56) A more appropriate way to establish the directionality of the association between school enrollment and sexual risk is to analyze prospectively collected data. Relatedly, in order to establish the correct sequence of events, the timing of changes in school enrollment status and timing of sexual risk outcomes should be recorded finely. This is often difficult in cross-sectional studies relying on self-reported data, particularly those asking respondents to remember temporally-distant events. More objective measures of exposures and outcomes (such as school enrollment and delivery dates recorded in a yearly census) provide better information to understand this complex relationship. Responding to the methodological limitations and general sparseness of the current literature, Jukes, *et al* state that there is an "urgent need to strengthen the evidence base" on education and sexual risk.(51) Using a large, longitudinal, census-based dataset, this study addressed many of the above-mentioned limitations, adding to the small, yet growing body of literature on the association between school enrollment and sexual risk.

4 Aim 2: Frequency of visits to alcohol outlets and sexual risk behavior

Frequenting alcohol outlets may influence sexual risk behaviors. Visits to alcohol outlets (establishments where alcohol is sold) may allow for periods of unsupervised and unstructured time which we hypothesize is associated with increased sexual risk. Additionally, the alcohol consumption associated with visits to social venues like bars and taverns could directly influence sexual risk behaviors. We discuss both of these potential mechanisms below, but begin with the significance of using sexual behaviors as outcomes.

Sexual behaviors are important proximate determinants of sexual risk outcomes. Although we typically think of sexual risk outcomes as disease or unintended pregnancy, the behaviors that put young women at risk for these events can also be important to explore. In particular, becoming sexually active and having unprotected sex have a clear relationship with HIV, other STIs, and pregnancy as together they are the mechanism for sexual transmission of disease and conception. Unprotected sex has been strongly linked with STI infection (61-64) and pregnancy rates.(65) Young women in South Africa do not report particularly risky sexual behaviors relative to their counterparts in the United States, but in the context of heavy HIV/STI burden, any sexual activity can be risky.(66) Identifying factors associated with sexual risk behaviors in addition to biological outcomes is important to better understand the etiology of sexual risk and to develop potential prevention interventions.

Alcohol use and abuse is associated with sexual risk. The link between alcohol use and sexual risk is well-established in populations throughout the world.(67-71) This association has been observed regionally among populations in sub-Saharan Africa (72-74) and specifically in South Africa.(75, 76) These findings appear to hold across both genders, over a range of ages, and for both behavioral outcomes such as unprotected sex,(67, 69, 72) and biological outcomes such as HIV infection.(70, 71, 73) Additionally, these associations appear to be robust to different measures of alcohol use, holding among global exposure studies (‘What is the frequency and quantity of recent alcohol consumption?’), event-level exposure studies (‘Was the individual under the influence of alcohol during particular occasions of risky sex?’), and studies where exposure was randomized.(69, 77)

Biological evidence supports the claim for a causal relationship between alcohol use and sexual risk. Consumption of alcohol leads to disinhibition and impaired decision-making

(74, 78) and decreases awareness of social norms and risk-inhibiting cues.(79) Relatedly, alcohol use leads to a myopic focus on the present, where distant consequences of proximal activities are discounted.(79, 80) These disruptions to the normal social and individual constraints on risky activity are thought to lead to increased sexual risk.(81, 82) A parallel line of research indicates that periods of sexual arousal may also lead to impaired judgment and decision-making.(83) This negative effect is reinforced by the fact that alcohol users themselves report feelings of enhanced sexual appeal and reduced sexual control.(76) Alcohol consumption and sexual arousal may, therefore, work synergistically to heighten sexual risk taking. Finally, there is also evidence for a dose-response relationship between alcohol use and sexual risk as, generally, greater sexual risk is associated with increased quantity of alcohol consumption.(72)

Typical drinking patterns in South Africa are dangerous. Approximately 25% of men and 10% of women experience alcohol problem symptoms in South Africa.(84) Alcohol misuse is also common among South African young adults; nearly a quarter of all high school students report recent binge drinking activity.(84, 85) Within the WHO-based regional grouping for southern Africa, of which South Africa is a member, an estimated 55% of males and 33% of females drink alcohol. For each drinker in the region an estimated 16.6 liters of alcohol is consumed per year, placing it among the few areas in the world with the most detrimental drinking patterns.(86)

Alcohol use and misuse are associated with patronizing alcohol outlets. In South Africa, Morejele *et al.* have developed a conceptual model of the relationship between alcohol use and sexual risk behaviors informed by qualitative data.(87) They found that the environment in which drinking occurs is a moderating factor in the association between alcohol use and sexual risk. This proposed relationship is supported by quantitative data linking patronizing alcohol outlets to increased frequency and quantity of drinking in South Africa.(88)

Aside from the biological effects of alcohol consumption, the context of the places where alcohol is served may also influence sexual risk. The physical features of a drinking establishment may be conducive to unprotected sex occurring on premises. Qualitative evidence suggests that the music, dim lights, uni-sex toilets, and lack of condoms typical in bars create favorable environments for risky sexual activity.(87) Additionally, there is evidence that the network of people to whom an individual is exposed in drinking establishments may be riskier than in non-drinking environments. Sex partners are often met in these places, particularly for young women, and these partners are often older men looking for sex in exchange for money.(74, 89)

Risky social venues that sell alcohol have been identified using a community mapping methodology. Specific social venues have been identified as high-risk using the Priorities for Local AIDS Control Efforts (PLACE) methodology throughout the world.(90-99) This methodology identifies venues where people meet new partners using qualitative interviews and mapping exercises with key informants.(100) In South Africa, more than 85% of the places identified in this manner were alcohol-serving establishments.(90) These high-risk locations are appealing targets for interventions; however, interventions directed at PLACE-identified locations have had mixed results.(98, 99)

Direct evidence for the association between visits to alcohol outlets and sexual risk is sparse. Although the sexual risk profiles of participants recruited at alcohol outlets have been relatively well-characterized as risky (74, 93, 101-104), studies comparing the sexual risk profiles of those with different levels of alcohol outlet exposure are much fewer. Two cross-sectional studies in Zimbabwe report moderate associations between recent beerhall visits and prevalent HIV infection among men (105) and prevalent HIV infection and risky sexual behavior among both men and women.(106) A single cross-sectional study of an adult population in South Africa found that patronizing alcohol outlets was associated with increased unprotected vaginal sex, and an association with increased sexual risk index was significant even after controlling for alcohol consumption.(88) A final, cross-sectional study in India found an association between patronizing wine shops and prevalent STI infection (HIV, HSV-2, syphilis, chlamydia, or gonorrhea) among adult males.(107) However, interventions targeted in these high-risk venues have so far had weak or null results on sexual risk outcomes.(108, 109) Studies linking visits to specific locations to sexual risk in adolescents are rare, though one novel study used risk mapping and focus groups among secondary students in Zimbabwe to identify six high-risk area categories, including commercial centers where alcohol is served.(110)

The current literature on exposure to alcohol outlets and sexual risk is limited, particularly for young women in sub-Saharan Africa. Only four studies have empirically examined this association. Although all prior studies found significant results in adult populations, the association has not been confirmed in adolescent-specific or female-specific populations. Also, the South African and Indian studies used convenience sampling methods where many members of the study population were enrolled within alcohol outlets themselves which could cause bias if they were not representative of the general population.(88, 107) Finally, all four prior studies dichotomized the alcohol outlet exposure into patrons versus non-patrons; this dichotomy may mask interesting nuances in the relationship between amount of exposure to drinking establishments and sexual risk. Clearly, more research is needed to add to the small body of literature linking visits to alcohol outlets and sexual risk.

5 Aim 3: Alcohol outlet density and sexual risk

Characteristics of the communities in which young women live may play a large role in shaping their exposure to different risk environments. Communities that provide safe resources and opportunities for youth may directly or indirectly lower their sexual risk while communities that provide opportunities for youth to engage in unsupervised activities may directly or indirectly increase their sexual risk. We hypothesize that alcohol outlets provide opportunities for youth to engage in unsupervised activities and that living in communities with more alcohol outlets will be associated with increased sexual risk.

Specific community characteristics are associated with sexual risk, but modifiable community characteristics are understudied. Associations with increased sexual risk have been noted for multiple sociodemographic, behavioral and spatial factors, including: urban location and age/gender profiles;(111) urban status and ethnicity/unemployment profiles;(112) proximity to a city and HIV prevalence within a 25 kilometer buffer;(113) GDP, mobility, urbanization and drug use;(114) proximity to a primary road;(115, 116) proximity to a market and distance to a health clinic;(117) and literacy rates, unemployment, poverty, and urban residency.(118) However, most of the community-level exposures explored to-date are non-modifiable or difficult to modify. Alcohol outlet accessibility is attractive as a potential intervention target because it is relatively modifiable through strengthening or more strictly enforcing existing government regulation.(119)

Alcohol outlet accessibility is associated with alcohol consumption and alcohol-related problems. The association between increased alcohol outlet density and increased frequency and consumption of alcohol consumption is well-established.(120-123) Increased alcohol outlet density is also associated with a variety of alcohol-related problems, including automobile injuries (124) and violent crime.(125) Public policy and economic frameworks also support the link between alcohol outlet accessibility, alcohol consumption, and alcohol-related problems. These perspectives posit that people tend to consume less alcohol and therefore have fewer alcohol-related problems when the accessibility to alcohol is limited, either through higher legal purchasing age, increased prices, or decreased outlets.(119, 126, 127) It should be noted, however, that the research to date on alcohol outlet density is heavily skewed to populations in the developed world. The vast majority of studies were performed in American study populations and, to our knowledge, the exposure of alcohol outlet accessibility has been unstudied in South African study populations.

Community alcohol outlet accessibility may be associated with sexual risk as well. To our knowledge, there have been four studies examining the association between alcohol outlet accessibility and sexual risk.(128-131) Increasing alcohol outlet density was significantly associated with census-level gonorrhea rates in a cross-sectional study in Louisiana (129) and a time series analysis in California (128), and with self-reported STI infections in a cross-sectional study in Louisiana and California.(130) A final cross-sectional study in Namibia found that increasing alcohol outlet density was associated with increased HIV prevalence.(131) This evidence suggests that the alcohol environment in a neighborhood can influence sexual risk. However, in South Africa, researchers found that proximity of an alcohol outlet to home residence was inversely associated with increased sexual risk behavior among female patrons of alcohol outlets, indicating that the anonymity that comes from visiting more distant venues may have a disinhibiting effect.(132)

Current research on the association between alcohol outlet density and sexual risk is methodologically limited. Studies examining both exposures and outcomes at the community-level may be subject to the ecologic fallacy if inferences are made at the individual-level.(128, 129, 131) If studies, instead, examine the influence of community-level factors on individual-level outcomes, appropriate statistical methods like multilevel modeling or generalized estimating equations must be used to account for the fact that observations in the same community are likely non-independent.(130) Finally, the cross-sectional nature of nearly all of the prior studies also limits the ability to assess directionality of the association between exposure and outcome.

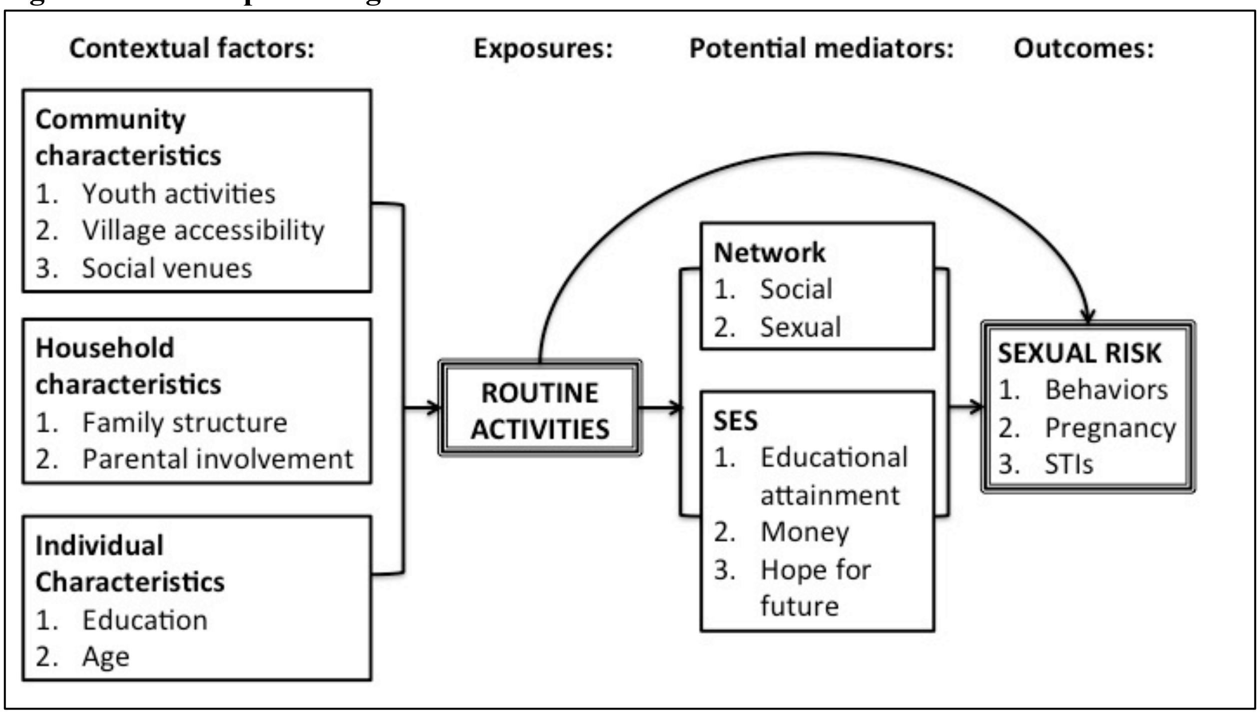
The evidence to-date on alcohol outlet density and sexual risk is incomplete. All but one of the four previous studies were conducted in the United States and all four were conducted in adult or general study populations. Replication of these results in youth-specific populations outside of the United States, such as the high-risk population of South African young women, would further extend the generalizability of the results. This is particularly pertinent as youth may have different alcohol outlet utilization patterns and alcohol-using behaviors than adults, so generalization from adult population studies to youth could be problematic.

6 Tables and Figures

Table 2.1: Theoretical support and exposure measures by study aim

Aim	Theoretical support	Exposure measure
1. School enrollment and teen pregnancy	1. Social cognitive theory 2. Social control theory	School enrollment status
2. Visits to alcohol outlets, sexual behavior and HSV-2	1. Social cognitive theory 2. Social control theory	Number of visits to taverns in last six months
3. Alcohol outlet density and HSV-2	Neighborhoods and Health	Number of alcohol outlets per village

Figure 2.1: Conceptual diagram of the influence of routine activities on sexual risk



CHAPTER III: Research Designs and Methods

1 Study Overview

Overall, the objective of this research project was to explore the association between routine activities on sexual risk among young women in rural South Africa. We approached this research question in three ways. In Aim 1, we used longitudinal data routinely collected from the Agincourt Health and socio-Demographic Surveillance Site to explore the influence of school enrollment on teen pregnancy. The second and third aims were an ancillary studies to HPTN 068 - The Effects of Cash Transfer and Community Mobilization in Young South African Women Study. The second aim explored the association between frequency of visits to alcohol outlets with sexual risk behaviors and HSV-2 infection. The third aim explored the association between village-level alcohol outlet accessibility and HSV-2 infection.

2 Study population and parent studies

To conduct the first aim of this study, we used previously collected data from the rural Bushbuckridge sub-district in the Mpumalanga province of South Africa. The Medical Research Council / Wits University Rural Public Health and Health Transitions Research Unit runs the Agincourt Health and socio-Demographic Surveillance Site (Agincourt HDSS) in this area and a complete census is performed on all the households yearly. The Agincourt HDSS is a multi-round, prospective, community study established in 1992; since then it has continually monitored all vital events occurring within the defined geographic area. As of 2011, the Agincourt HDSS had expanded to cover 90,000 people living in 16,000 households across 24 villages.(133)

Cohort profiles and details of the data collection methods have been described previously.(133, 134) Figure 3.1 displays a map of the study site location and the spatial distribution of study villages.

The second and third aims used baseline data originally collected by HPTN 068. HPTN 068 is a Phase III randomized trial currently being conducted within the Agincourt HDSS and surrounding area. The study aims are to determine whether providing cash transfers conditional on school enrollment reduces HIV risk in young women. Eligibility inclusion criteria at baseline were: female; age 13-20; enrolled in grades 8, 9, 10, or 11 at a school in the HDSS study site; intention to live in the study site until the end of follow-up; consent/assent to HIV and HSV-2 testing; ability to read sufficiently to self-administer a computer assisted interview; documentation to be able to open a bank account; and residence with a parent/guardian willing to consent to all study procedures. Eligibility exclusion criteria were: pregnancy at baseline; married at baseline; and no parent or legal guardian in the household. Importantly, young women who tested positive for HIV at baseline were not be excluded from study participation. A total of 2533 young women were enrolled in HPTN 068 and completed baseline procedures.

3 Aim 1: School enrollment and teenage pregnancy

3.1 Study design

To investigate the association between school enrollment and pregnancy, we conducted a study that was longitudinal in design with both the exposure and the outcome assessed and analyzed at the individual-level. We constructed a cohort from data originally collected by the Agincourt HDSS that included up to six years of follow-up for each participant. School enrollment exposure was included as a time-varying exposure and each participant was followed up retrospectively for incident first pregnancy.

3.2 Data collection and variable measurement

Cohort construction: We constructed the cohort using data from the Agincourt HDSS census. All 12-18 year-olds who had not yet experienced their first pregnancy and who lived in the study area between 2000-2012 contributed person-time to this study. Nulliparous young women who moved into the study site after their 12th birthday and nulliparous young women older than 12 in 2000 were treated as late entries. Young women who moved out of the study site or died during follow-up were censored at the time of move or death. All young women who did not experience their first pregnancy by the age of 18 were administratively censored on their 18th birthday.

Outcome assessment: The outcome we explored in Aim 1 was **incident first pregnancy**. All vital events that occur in the Agincourt HDSS study area, including births, deaths, and migrations, have been updated continuously in the database since 1992. In the yearly census update, fieldworkers administer a pregnancy module to collect information on all pregnancies that took place in the previous 12 months. Although this list of pregnancies includes those that did not result in live births, this outcome is self-reported, and miscarriages and abortions, particularly early losses, may be under-reported.

Date of delivery was recorded for all pregnancies. We calculated the estimated conception date for each pregnancy by subtracting 40 weeks (the average duration of pregnancy) from the date of delivery. The time to first pregnancy was calculated as the duration, in months, between a young woman's 12th birthday and the estimated conception date of her first pregnancy.

We restricted our outcome to first pregnancy as opposed to pregnancy of any order for several reasons. First, given the relatively young ages in our cohort, we did not expect to see many young women who would go on to have multiple pregnancies during follow-up. Second, we theorized that young women who had previously had a child would become incomparably different from those who had not, particularly with regards to their school enrollment. For example, young women who have had a child are less likely to go on and have further school enrollment exposure, even though previous pregnancy is not completely incompatible with future schooling, particularly in South Africa.(57)

Miscarriages, perinatal deaths, and abortions were all likely to be underreported in this sample. Unreported abortions could plausibly produce differential outcome misclassification with respect to school enrollment status. To address this, we compared the rate of abortions in the study sample to recent South African abortion rate estimates. As a sensitivity analysis, we explored whether unreported abortions could plausibly account for the observed association by estimating how many abortions would need to be unreported among school enrollees to nullify the observed association. If this total number would yield an abortion rate of improbable magnitude compared to national estimates, we considered our results robust to unreported abortions. Because they are not in direct control of the participant, unreported miscarriage and perinatal death rates were likely nondifferential with respect to school enrollment status.

Exposure assessment: The primary exposure we explored in Aim 1 was **school enrollment**. School enrollment was coded as a time-varying bivariate exposure equal to one when the participant was enrolled in school and zero when not enrolled in school. An education status module was administered in 1997, 2002, 2006, 2009, and 2012. At the time of each module, field workers recorded the highest education level each young woman had achieved and whether or not she was currently enrolled in school. Although these modules were only updated, on average, every three years, we could interpolate school enrollment at any given time by examining change in educational attainment between module years. For example, if a young woman reported current enrollment in Grade 8 in 2002 and current enrollment in Grade 12 in 2006, we inferred that she was continuously enrolled in school between 2002 and 2006. Table 3.1 shows other potentially more complex data configurations and the *a priori* coding decisions we made about school enrollment duration.

We performed three sensitivity analyses around these assumptions to evaluate the robustness of our results to potential school enrollment misclassification. First, to address the uncertainty around the sequence of events when exposure and outcome occurred around the same time, we created a restricted sample removing all observations with a pregnancy within one year before or after the estimated school dropout date. Second, to address the uncertain enrollment status of young women who may have temporarily dropped out of school or skipped grades, we created a restricted sample removing all observations with non-linear grade progression patterns. Third, we created a final restricted sample by combining both of the two restrictions above. We compared differences in results from the full sample and from each of these three restricted samples to assess the potential influence of misclassification of the school enrollment exposure.

Covariate assessment: We also explored the potential influence of several key covariates on the relationship between pregnancy and school enrollment. Specifically, we requested data on:

1) **Age**, calculated, in years, from birthdate. We hypothesized that increasing age would be associated with an increased likelihood of pregnancy and a decreased likelihood of school enrollment.
2) **Gender of household head**, defined as the gender (male or female) of the individual reported as the head of household. We hypothesized that living in a household headed by a female would be associated with decreased likelihood of school enrollment (directly

and through decreased parental monitoring and household SES) and increased likelihood of pregnancy (through decreased household SES).

3) **Household head employment**, defined as whether or not the household head reports employment (employment module administered in 2000, 2004, and 2008). We hypothesized that young women living in households headed by people with employment would be associated with increased likelihood of school enrollment (directly and through increased household SES) and associated with pregnancy (through increased household SES and decreased parental monitoring).
4) **Household head secondary education,** defined as whether or not the household head reports 12 years of educational attainment, as updated in the education module administered in 1997, 2002, 2006, 2009, and 2012. We hypothesized that young women living in households headed by people who completed secondary school would be associated with increased likelihood of school enrollment and decreased likelihood of pregnancy (through increased household SES and increased likelihood of employment).
5) **Household size**, defined as the total number of individuals reporting membership in the participant's household. We hypothesized that increasing household size would be associated with decreased school enrollment (directly and through decreased parental monitoring and household SES) and increased likelihood of pregnancy (through decreased household SES).
6) **Household socioeconomic status (SES)**, measured as a time-varying composite index of general SES based on household assets (SES module administered in 2001, 2003, 2005, and 2007). We hypothesized that increasing household SES would be associated with an increased likelihood of school enrollment and decreased likelihood of pregnancy.

7) We also explored the influence of **calendar year** on the relationship between school enrollment and pregnancy by creating an indicator for the year in which each participant turned 12 years old. As this study had a long follow-up period, it was possible that both teen pregnancy rates and school enrollment rates varied over time due to external factors.

The hypothesized relationships between the covariates, school enrollment and pregnancy are diagrammed in Figure 3.2.

3.3 Statistical approach

Using a Cox proportional hazards model, we compared the hazard of first pregnancy among those enrolled in school compared to those not enrolled in school. The origin for each participant began on her 12th birthday and the time scale was age. We partitioned the dataset so that young women who switched school enrollment status during follow-up contributed both exposed and unexposed person-time. Young women were censored at the time of their first pregnancy; if they moved, died, or were otherwise lost to follow-up; when they attained a grade 12 educational attainment; or when they reached 18 years of age.

We assessed for the potential confounding effects of time-fixed and time-varying covariates, using the directed acyclic graph to identify a minimally sufficient adjustment set. Age was adjusted for implicitly as age was the time scale in this analysis. Coding decisions for all covariates were based on likelihood ratio tests comparing a variety of different functional forms. We also considered the potential for effect measure modification by age. To do this, we categorized age into young (age 12 through 14 years) and old (age 15 through 17 years) and examined the stratified estimates to assess whether there were differences in the magnitude of effect across strata. To formally test for effect measure modification, we included an interaction term between the dichotomized age variable and the school enrollment exposure variable in the model. Interaction terms with Wald test statistic p-values of <0.10 were considered significant in which case stratified estimates were reported.

3.4 Power

We parameterized our *a priori* power calculation using the publically available ten percent sample of the full HDSS dataset. In this sample, a total of 2,097 young women between the ages of 12 to 18 contributed 6,308 person-years between 2000 and 2012. This number of person-years incorporates loss to follow up due to death or move outside of the study area, aging in and out of the cohort, and censoring after first pregnancy. We therefore assumed 63,080 person-years would be observed in the full cohort. Approximately 50% of the sample reported being enrolled in school at each of the time points; we assumed that 50% of the person years under study in the full sample will be exposed to school enrollment as well. A total of 152 first pregnancies (7.3%) were recorded in the ten percent sample; we assumed we would observe approximately 1,520 first pregnancies in the full sample.

Using an expected data approach for a simple Poisson model with a two-sided test with an alpha of 0.05, we calculated the 95% confidence limits across the range of potential protective hazard ratios from 0.50 to 0.90 (Table 3.2). All expected confidence intervals excluded the null and were precise with confidence limit ratios around 1.25. We maintained a statistical power greater than 80% for hazard ratios from 0.50 to 0.86, as confirmed in a power calculation using the same parameters estimating the expected power for all hazard ratios between 0.7 and 0.9 (Figure 3.3).

4 Aim 2: Alcohol outlet visits and sexual risk

4.1 Study design

To explore the association between alcohol outlet visits and sexual risk, we conducted a cross-sectional analysis using data collected in the baseline quantitative survey for HPTN 068. Information on the exposure (number of alcohol outlet visits) and outcomes (unprotected sex acts, number of sex partners, transactional sex, HSV-2 infection) were collected at the individual-level at the same point in time. All young women with baseline data collected in the parent study were included in the analysis.

4.2 Data collection and variable measurement

Outcome assessment: In Aim 2, we explored four indicators of sexual risk as outcomes. The three behavioral outcomes were self-reported in the HPTN 068 baseline questionnaire. To minimize the bias that may come from answering personal questions about sexual behavior to an interviewer, an ACASI (audio computer-assisted self-interviewing) component was incorporated into the HPTN 068 baseline questionnaire. The ACASI component allowed participants to privately read or listen to audiotaped questions and log their responses directly in the computer, without having direct interaction with the interviewer. To help prevent data entry errors, skip patterns and warnings for implausible or impossible responses were coded into the program. The biological outcome was also assessed in HPTN 068 baseline procedures. The outcomes of interest were:

1) **Unprotected sex acts**: This count variable was defined by subtracting the total number of condom-protected vaginal sex acts from the total number of vaginal sex acts self-reported in the last three months. In the full sample, those who reported no sex acts at all received a zero value for unprotected sex acts. We carefully examined the distribution of sex acts

to identify potential outliers and responses with repeated single digits that may have been due to unintentional double-striking in the ACASI software. We recoded probable double-strike responses based on corroborating information in the number protected sex act responses. Figure 3.4 shows the distribution of unprotected sex acts among those who reported at least one sex act (protected or not) in the last three months with responses greater than 15 truncated at 15. As there were very few young women with more than two unprotected sex acts, for analysis, we categorized unprotected sex acts into zero, one, and more than one sex acts. Table 3.3 shows the baseline frequencies of each category of unprotected sex acts.

2) **Number of sex partners:** This count variable was defined as the number of self-reported sex partners in the last three months. In the full sample, those who reported no sex partners ever received a zero value for this variable. We recoded probable double-strike responses based on corroborating information in the number of lifetime sex partner responses. Figure 3.5 shows the distribution of number of sex partners in the last three months. As there were very few young women with more than two sex partners, for analysis, we categorized sex partners into zero, one, and more than one partners. Table 3.3 shows the baseline frequencies of each category of number of sex partners.

3) **Transactional sex**: This dichotomous variable was constructed from responses to questions in the baseline questionnaire regarding whether the participant had received money or gifts from her most recent sex partner and whether she felt obligated to have sex in return. Table 3.3 shows the baseline frequency of transactional sex with most recent partner.

4) **Prevalent HSV-2 infection:** Testing for HSV-2 infection was performed at baseline using Kalon assays. Table 3.3 shows the baseline frequency of HSV-2 infection.

Exposure assessment: The exposure we assessed in Aim 2 was **frequency of alcohol outlet visits**. This count variable was defined as the number of self-reported visits to a tavern/shebeen in the last six months. We removed observations with suspected double strike responses and implausibly high responses. Figure 3.6 shows a histogram of the univariate distribution of frequency of alcohol outlet visits. We assessed the functional form of its relationship with each outcome variable considering different categorizations and transformations. For analysis, we categorized the exposure into three categories: No visits, 1-5 visits (fewer than one visit per month, on average), and 6+ visits (at least one visit per month, on average). Table 3.4 shows descriptive statistics for number of alcohol outlet visits and frequencies of each category. Overall, in the baseline sample, the mean number of alcohol outlet visits in the last six months was 1.2 with a standard deviation of 3.2. The values ranged from 0 to 81. The distribution was positively skewed with a large number of zero responses and the majority of responses ranging between zero and ten.

Covariate assessment: We explored the influence of several key covariates on the relationship between alcohol outlet visits and each sexual risk outcome. Specifically we examined:

1) **Age**, defined as age, in years, at the time of baseline interview. We hypothesized that increasing age would be associated with an increased number of alcohol outlet visits and increased likelihood of sexual risk.

2) **Education**, defined as the current grade in which the young woman was enrolled at baseline (an eligibility criterion for HPTN 068 was current school enrollment at baseline). We hypothesized that increasing education would be associated with a decreased number of alcohol outlet visits and decreased likelihood of sexual risk.
3) **Household size**, defined as the total number of individuals sharing a household with the young woman. We hypothesized that increasing household size would be associated with an increased number of alcohol outlet visits (through decreased parental monitoring) and increased likelihood of sexual risk.
4) **Primary caregiver relationship**, defined as how the young woman is related to her primary caregiver: daughter, sibling, niece, grandchild, other. We hypothesized that having a primary caregiver relationship other than daughter would be associated with an increased number of alcohol outlet visits and an increased likelihood of sexual risk. We hypothesized that having a primary caregiver who is a grandparent would influence sexual risk through household SES (increased likelihood that the household received the old age pension).
5) **Household SES**, defined as a measure of per capita monthly household consumption of food and durable goods. We hypothesized that an increased household SES would be associated with number of alcohol outlet visits though the direction of the association could be positive (increased SES leads to more expendable income to be spent at alcohol outlets) or negative (increased SES associated with decreased visits to alcohol outlets through decreased household size or increased parental monitoring). We hypothesized that an increased household SES would be associated with a decreased likelihood of sexual risk.

6) **Alcohol consumption frequency**, defined as the categorical response to the question "How often do you drink alcohol?" with seven possible responses ranging from "Never" to "Daily." For analysis, we dichotomized this variable separating those with any consumption from those with no consumption. We explored the potential for interaction between this variable and alcohol outlet visits in their relationship with each sexual risk outcome.

The hypothesized relationships between these key covariates, the exposure (visits to alcohol outlets), and the sexual risk outcomes are depicted in Figure 3.7.

4.3 Statistical approach

We used logistic regression models to estimate the association between alcohol outlet visits and the sexual risk outcomes of transactional sex and HSV-2 infection (Equation 4.1). This approach was appropriate given the dichotomous nature of these two outcomes.

Equation 4.1: $$P(Risk = 1) = \frac{1}{1+exp^{-\alpha+\beta_1(AO\_visit)+\beta_2 X_2}}$$

Risk: either of the dichotomous sexual risk outcomes – transactional sex or HSV-2

AO_visit: alcohol outlet visits

X: the matrix of the potential covariates in our final adjustment set

We used multinomial logistic regression models to estimate the assocation between alcohol outlet visits and the sexual risk outcomes of unprotected sex acts and number of sex partners.(135) This approach was appropriate given the categorical nature (0, 1, and 2+) of both these outcomes. We used multinomial as opposed to ordinal logistic regression because, although there is an inherent order to the categories, we did not wish to impose an assumption of proportional odds between all pairs of categories. To build the multinomial logistic regression model, we simultaneously modeled two expressions: 1. The natural log of the probability that the sexual risk outcome category was 1 versus zero (Equation 4.2) and 2. The natural log of the probability that the sexual risk outcome category was more than one versus zero (Equation 4.3).

Equation 4.2: $$ln\left[\frac{P(Risk=1)}{P(Risk=0)}\right] = \propto_1 + \beta_{11}(AO\_visit) + \beta_{12}X$$

Equation 4.3: $$ln\left[\frac{P(Risk=2+)}{P(Risk=0)}\right] = \propto_2 + \beta_{21}(AO\_visit) + \beta_{22}X$$

Risk: either of the categorical sexual risk outcomes – unprotected sex acts or number of sex partners

AO_visit: alcohol outlet visits

X: the matrix of the potential covariates in our final adjustment set

Solving these expressions yielded two separate beta parameters associated with alcohol outlet visits, one for each of the two comparisons we were making. Exponentiating the beta parameter derived from modeling Equation 4.2 yielded an approximation of the odds ratio comparing the odds of being in category one versus zero among those with exposure to alcohol outlets, compared to those without exposure to alcohol outlets. Exponentiating the beta parameter derived from modeling Equation 4.3 yielded an approximation of the odds ratio comparing the odds of being in category two versus zero among those with exposure to alcohol outlets, compared to those without exposure to alcohol outlets.

We assessed for the potential confounding effects of age, education, village, SES, household size, and caregiver relationship. First, we identified a minimally sufficient adjustment set from the directed acyclic graph (Figure 3.7). Each covariate's functional form with each sexual risk outcome was carefully examined by comparing different transformations and categorizations. Final coding decisions for each covariate were informed by likelihood ratio tests, separately for each outcome.

We also explored the potential for interaction between alcohol outlet visits and alcohol consumption frequency in their relationship with each sexual risk outcome. To do this, we dichotomized both the alcohol outlet visits and alcohol consumption variables into some versus no exposure levels. We then included an interaction term between visits and consumption in the model for each outcome. We used logistic regression models for the dichotomous transactional sex and HSV-2 infection outcomes. We used ordinal, as opposed to multinomia, logistic regression models for the categorical sex partner and unprotected sex act outcomes to maximize statistical power and because the unstratified multinomial results suggested an ordered relationship. We assessed the statistical significance of the interaction term in each model with likelihood ratio tests, considering any p-value under 0.2 to be statistically significant.

4.4 Power

We parameterized our power calculation for a simple logistic regression model estimating the odds ratio for the association between alcohol outlet visits (dichotomized as any versus no exposure) and transactional sex. We conservatively parameterized the power calculation with the transactional sex outcome as this is was the least common of the four outcomes we examined. We anticipated higher power to detect associations with the more common outcomes. We assumed a two-sided test for the statistical significance of the likelihood ratio chi-square statistic at alpha=0.05 and a total sample size of 2533. We assumed that 50% of the young women had any alcohol outlet exposure and that the proportion of young women who had transactional sex relationships with their most recent partners among those reporting no visits to taverns in the last six months was 3%, as calculated from the baseline survey data. Finally, we varied the expected odds ratios from 1.10 to 2.0 to cover a range of plausibly small effect estimates in the hypothesized direction.

Figure 3.8 shows the power curve over the range of odds ratios above the null. Overall, for all odds ratios greater than 1.76 we maintained a statistical power above 80% to estimate the effect of alcohol outlet visits on odds of transactional sex.

5 Aim 3: Alcohol outlet accessibility and sexual risk

5.1 Study design

To explore the influence of alcohol outlet accessibility (number of alcohol outlets per village) on sexual risk (prevalent HSV-2 infection), we performed a cross-sectional study using biological data collected at baseline in HPTN 068 and village asset data collected as part of a community survey performed as a complementary study to HPTN 068. Alcohol outlet number was measured and analyzed at the village-level while the HSV-2 outcome was measured and analyzed at the individual-level. The design and analysis of this aim was grounded in a multilevel framework because we included variables at both the individual- and village-level.

5.2 Data collection and variable measurement

Outcome assessment: The outcome of interest in Aim 3 was **prevalent HSV-2 infection.** There were a total of 123 prevalent HSV-2 infections (4.9%) identified in the study sample at baseline. HSV-2 infections are appropriate indicators of sexual risk as they are sexually transmitted and, unlike HIV, cannot be due to perinatal transmission. HSV-2 testing was performed at baseline for each young woman enrolled in HPTN 068 using Kalon™ HSV-2 gG2 ELISA.(136)

Exposure assessment: The exposure of interest for this aim was **number of alcohol outlets per village.** For each of 24 villages in which the community asset exercise was conducted, we calculated the total number of alcohol outlets by adding the number of taverns and number of bottle shops (liquor stores). The number of taverns and bottle shops per village were reported by consensus among several key informants from each village through a community mapping exercise undertaken in a HPTN 068 sub-study. Each young woman was assigned the alcohol outlet number value of her village of residence.

We used an absolute number instead of density per area measure because there was not a large variation in village area (interquartile range: 1.3 to 3.2 square kilometers). Further, we theorized that, in small areas like the villages in this region, absolute number was more important than density because all outlets were within reasonable walking distance to all inhabitants. The number of alcohol outlets per village ranged from zero to seven outlets with a median of four outlets (Table 3.5). The histogram for the individual distribution of alcohol outlet density is shown in Figure 3.9. The data were fairly uniformly distributed across the range of values, though there were slightly more data at moderate numbers than for the lowest and highest numbers.

Covariate assessment: We also collected information on several key covariates we considered in estimating the association between alcohol outlet number and HSV-2 infection. Specifically, we considered:

1) **Number of visits to alcohol outlets**, defined as the number of self-reported visits to a tavern/shebeen in the last six months, as recorded in the baseline survey. We hypothesized that this variable would act as a mediator between number of alcohol outlets and HSV-2 infection, though we did not explore the potential mediating relationship in this analysis.
2) **Alcohol consumption frequency**, defined as the categorical response to the question "How often do you drink alcohol?" with seven possible responses ranging from "Never" to "Daily." We hypothesized that this variable would be on a mediating pathway between number of alcohol outlets and HSV-2 infection, though we did not explore the potential mediating relationship in this analysis.
3) **Age**, defined as age in years at baseline. We hypothesized that increasing age would be associated with increased likelihood of HSV-2 infection but not directly associated with number of alcohol outlets.
4) **Household SES**, defined as a measure of per capita monthly household consumption of food and durable goods collected in the household baseline questionnaire asked of the primary caregiver of each study participant. We hypothesized that increasing household SES would be associated with lower likelihood of HSV-2 infection but not directly associated with number of alcohol outlets.
5) **Education**, defined as the current grade of school in which the young woman was enrolled at baseline. We hypothesized that increasing education would be associated with a lower likelihood of HSV-2 infection. We hypothesized that education would not be directly associated with number of alcohol outlets.

6) **Village population**, defined as the 2011 population estimate for each village, as derived from the Agincourt HDSS census. We hypothesized that increasing village population would be associated with increased number of alcohol outlets and increased likelihood of HSV-2 infection.
7) **Proportion of population employed**, defined as the total number of village residents reporting employment in the most recent labor status census module (2008) divided by the total population in each village. We hypothesized that proportion employed would be associated with increased number of alcohol outlets and decreased likelihood of HSV-2 infection.
8) **Proportion of population male**, defined as the total number of male residents in each village in the 2011 census update, divided by the total population in each village. We hypothesized that proportion male would be associated with increased number of alcohol outlets per village and increased likelihood of HSV-2 infection.

The hypothesized relationships between these covariates, the number of alcohol outlets per village and HSV-2 infection are depicted in the directed acyclic graph in Figure 3.10. All individual-level covariates (age, household SES, education) were only associated with the exposure and outcome through the pathways of alcohol outlet visits and alcohol consumption. The only potential confounders of the total association between village-level number of alcohol outlets and individual-level HSV-2 infection were village population size, proportion of population employed, and proportion of population male.

5.3 Statistical approach

As noted previously, the statistical analysis for Aim 3 was grounded in a *Neighborhoods and Health* theoretical framework. This framework was developed to explore the health effects of factors at multiple levels of organization.(33) Because the data in multilevel studies are structured in a way that the individual observations are clustered at a higher-level order of organization, it is important to statistically account for the non-independent nature of the data.(137-139) Although many multilevel studies use a multilevel modeling (MLM) statistical approach, in this aim we used generalized estimating equations (GEE) to account for the clustered nature of the data. Here, we were specifically interested in estimating an average effect of alcohol outlets on HSV-2 infection in the presence of potential clustering by village (GEE model strength) without estimating how additional parameters may vary by village (MLM strength).(140) A GEE model was preferred because it was appropriate for the level of complexity our research question dictated.

GEE models are at first constructed identically to generalized linear models. For dichotomous outcomes, like the HSV-2 infection outcome explored here, log links are appropriate, creating an equivalent to standard log linear regression (Equation 5.1).

Equation 5.1: $log[P(HSV2 = 1)] = \beta_0 + \beta_1 AO_1$

HSV2: HSV-2 outcome variable

AO: number of alcohol outlets exposure variable

The primary difference between standard logistic regression and GEE models involves differences in the underlying assumptions of correlation. With standard logistic regression models, there is an assumption of independence between all observations. With GEE models, this assumption is relaxed by specifying a working correlation structure to define how observations within groups are correlated. For this analysis, we specified an exchangeable correlation structure which was estimated so that every pair of observations within each group had the same correlation. This structure was appropriate given that the order of observations (young women) within each cluster (village) was arbitrary, and, therefore, the positions were exchangeable. The final GEE model produced a consistent and asymptotically normal estimate of the beta parameter associated with village number of alcohol outlets, and appropriately adjusted the standard error to account for the clustered nature of the data.(139, 141)

To control for confounding, we identified all unblocked confounding paths from the directed acyclic graph of the relationship between key covariates of interest, number of alcohol outlets per village, and HSV-2 infection (Figure 3.10). All identified confounding variables were coded as indicated by likelihood ratio tests comparing different functional forms of each variable's association with the HSV-2 outcome. No *a priori* potential effect measure modifiers were explored.

In this exploratory analysis, we were interested in establishing an estimate of the total effect of number of alcohol outlets on HSV-2 infection. A future line of analysis may be undertaken to further disentangle the contextual (direct) effects of alcohol outlet number on sexual risk from the individual (indirect) effects. To do so, we would conduct additional analyses exploring the potential mediation pathways through individual alcohol outlet visits and individual alcohol consumption. Any remaining direct effect of number of alcohol outlets on HSV-2 infection could potentially be attributed to contextual effects.

5.4 Power

To account for the inflated precision that came with the use of clustered data, we needed to adjust our projected sample size by a factor equal to the *design effect* to accurately estimate power. We calculated the *design effect* in this study to be **1.84** using equation 5.2. (142)

Equation 5.2: *Design effect =1+(n-1)ρ*

n: average participants per village = 85

ρ: the intracluster correlation coefficient, (ICC), which we estimated to be 0.01 based on the fact that ICC typically falls between 0.01 and 0.02 in human studies.(143)

We parameterized our power calculation for a simple logistic regression model estimating the odds ratio for a one-unit increase in number of alcohol outlets on the odds of HSV-2 infection. Given the design effect of 1.84, our effective sample size was decreased from 2533 to 1377. We further parameterized our power calculation assuming a total proportion of HSV-2 outcomes in the unexposed at 4.9%. We assumed that the alcohol outlet density variable was normally distributed with a mean of 4.2 and standard deviation of 2.0 as calculated from the baseline survey data**.** Finally, we varied the expected magnitude of effect from an odds ratio of 1.1 to 1.5. The results of the power calculation assuming a two-sided test for the statistical significance of the likelihood ratio chi-square statistic at alpha=0.05 is shown in Figure 3.11. Although, as expected, this aim is underpowered at odds ratios very close to the null, we maintain a statistical power above 80% for all odds ratios above 1.20.

6 Tables and Figures

Table 3.1: Example data configuration possibilities with school enrollment exposure coding decisions

	School enrollment data configuration possibilities					Exposure duration	Notes
	1997	*2002*	*2006*	*2009*	*2012*	*(2000-2012)*	
Enrolled?	No	No	No	No	No	**0 years**	
Attainment?	None	None	None	None	None		
Enrolled?	Yes	Yes	No	No	No	**4 years**	
Attainment?	Grade 5	Grade 10	Grade 12	Grade 12	Grade 12	**(2000-2004)**	
Enrolled?	Yes	Yes	No	No	No	**2 years**	
Attainment?	Grade 5	Grade 10	Grade 10	Grade 10	Grade 10	**(2000-2002)**	
Enrolled?	Missing	Yes	Missing	No	No	**4 years**	Infer over missing data in middle
Attainment?		Grade 8		Grade 10	Grade 10	**(2000-2004)**	
Enrolled?	Missing	Missing	Yes	No	No	**6 years**	Infer over missing data in beginning
Attainment?			Grade 10	Grade 10	Grade 10	**(2000-2006)**	
Enrolled?	Yes	Yes	Missing	Missing	Missing	**2+ years**	MI to impute educational attainment
Attainment?	Grade 3	Grade 8				**(2000-?)**	
Enrolled?	Yes	Yes	LTFU (2005)	LTFU	LTFU	**4 years**	Enrollment status extended to time of
Attainment?	Grade 3	Grade 8				**(2000-2004)**	LTFU
Enrolled?	Yes	No	LTFU (2005)	LTFU	LTFU	**1 year**	Enrollment status extended to time of
Attainment?	Grade 3	Grade 7				**(2000-2001)**	LTFU
Enrolled?	Yes	Yes	Yes	No	No	**6 years**	Assume continuous enrollment; flag for
Attainment?	Grade 3	Grade 8	Grade 10	Grade 10	Grade 10	**(2000-2006)**	temporary dropout or grade repeat

LTFU = loss to follow-up

Table 3.2: The expected confidence limits and statistical power across a range of protective hazard ratios for the association between school enrollment and pregnancy

Hazard Ratio	95% Confidence Interval
0.50	(0.45, 0.56)
0.60	(0.54, 0.67)
0.70	(0.63, 0.78)
0.80	(0.72, 0.89)
0.90	(0.81, 1.00)

Table 3.3: Frequency of four sexual risk outcomes in baseline sample

Outcome	N	(%)
Unprotected sex acts[a]		
0	2176	(92.3)
1	78	(3.3)
2+	103	(4.4)
Sex partners[a]		
0	1917	(76.3)
1	539	(21.4)
2+	58	(2.3)
Transactional sex[b]		
No	2451	(96.9)
Yes	79	(3.1)
HSV-2 infection		
No	2409	(95.1)
Yes	123	(4.9)

[a]In the last three months
[b]With most recent partner

Table 3.4: Descriptive statistics for number of tavern visits in the last six months

Continuous exposure	Mean	Median	SD	Range
# Visits to alcohol outlet	1.2	0	3.2	0-81
Categorical exposure	**N**	**%**		
No visits	1354	57.5		
1-5 visits	897	37.8		
6+ visits	124	5.2		

Table 3.5: Descriptive statistics for village alcohol outlets at individual-level

Variable	Mean	Median	SD	Range
# Taverns	3.6	4	1.6	0-6
# Bottle Shops	0.6	0	0.8	0-2
# Alcohol Outlets	4.2	4	2.0	0-7

Figure 3.1: Map of the study site location

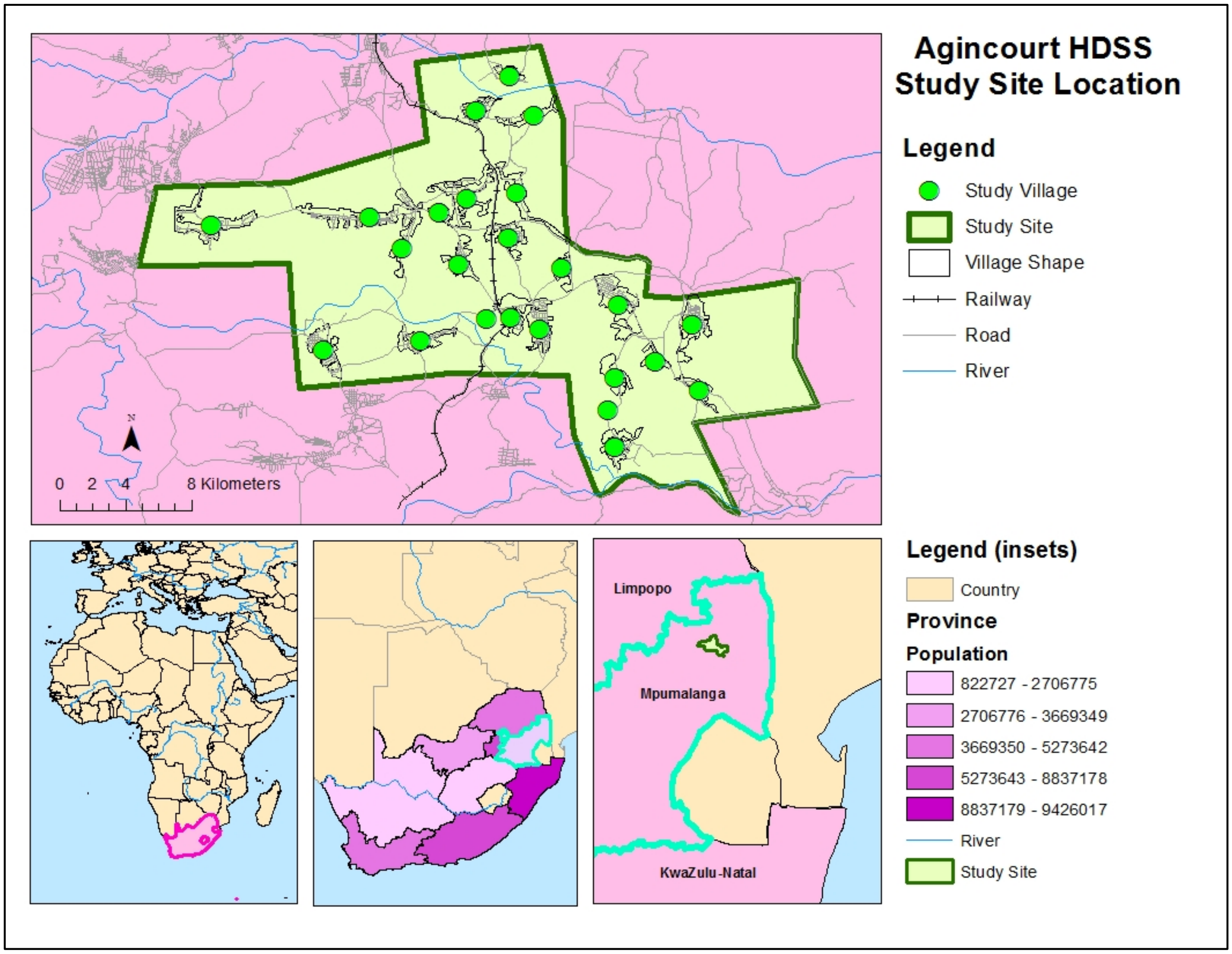


Figure 3.2: Directed acyclic graph depicting the hypothesized relationship between school enrollment and pregnancy

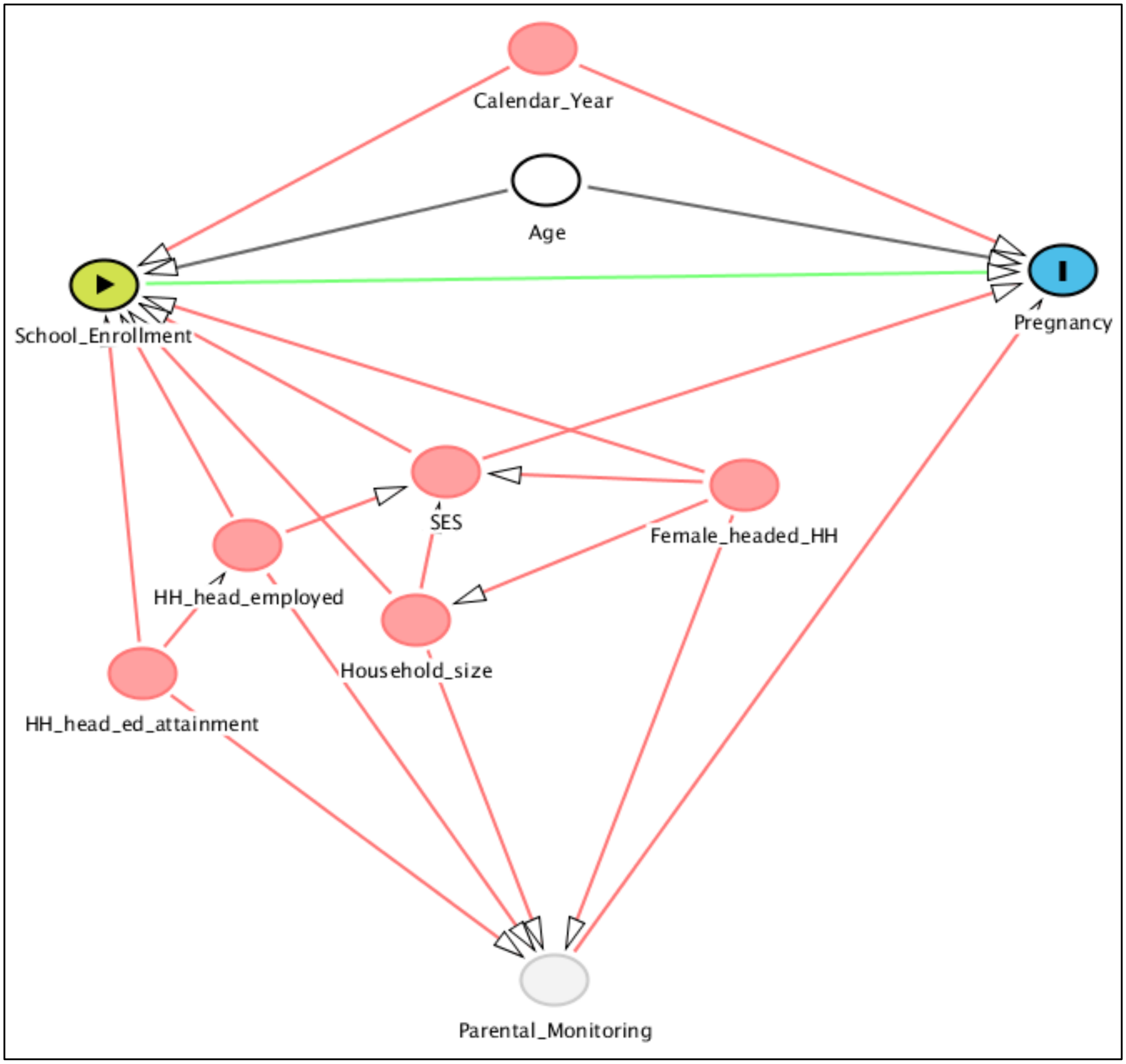


Figure 3.3: Expected statistical power to explore the association between school enrollment and teen pregnancy over a range of hazard ratios (0.7 to 0.9)

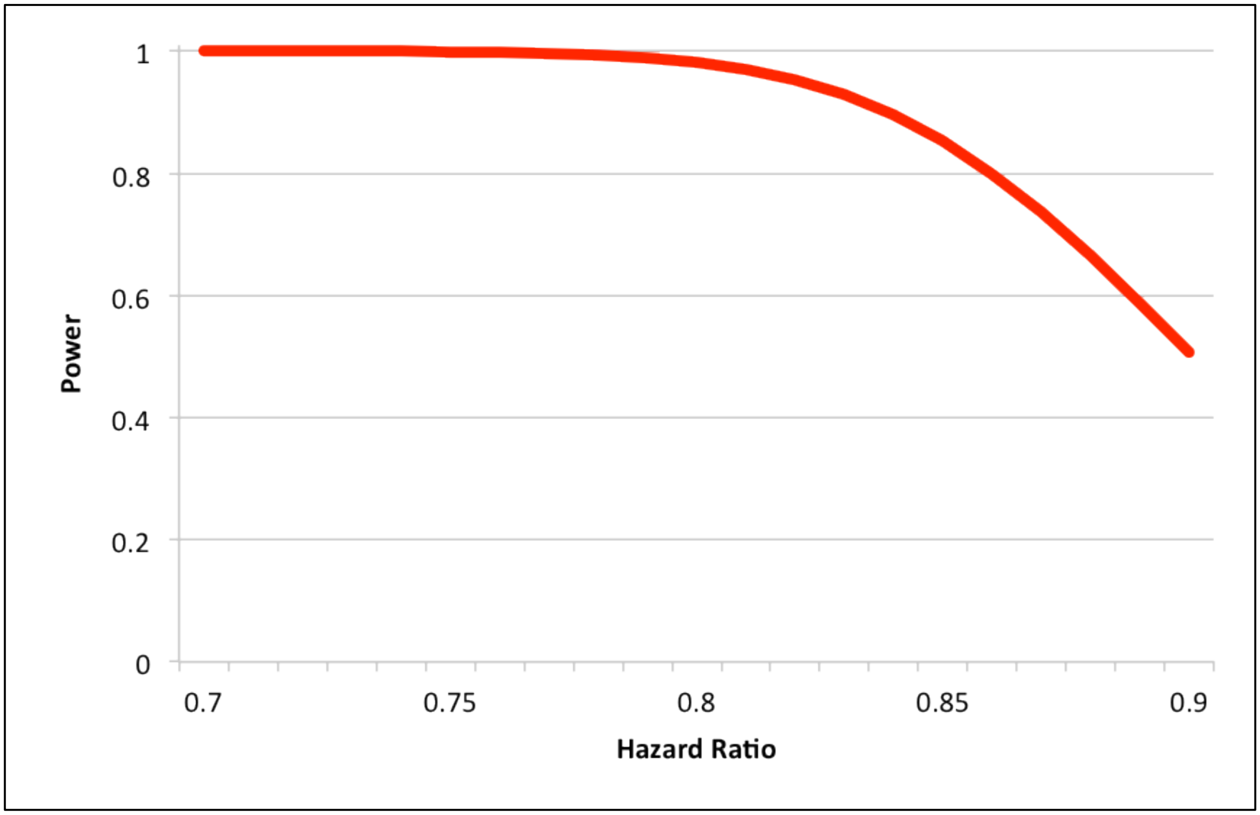


Figure 3.4: Histogram of unprotected sex acts in the last three months among respondents reporting any sex in last three months (n=527)

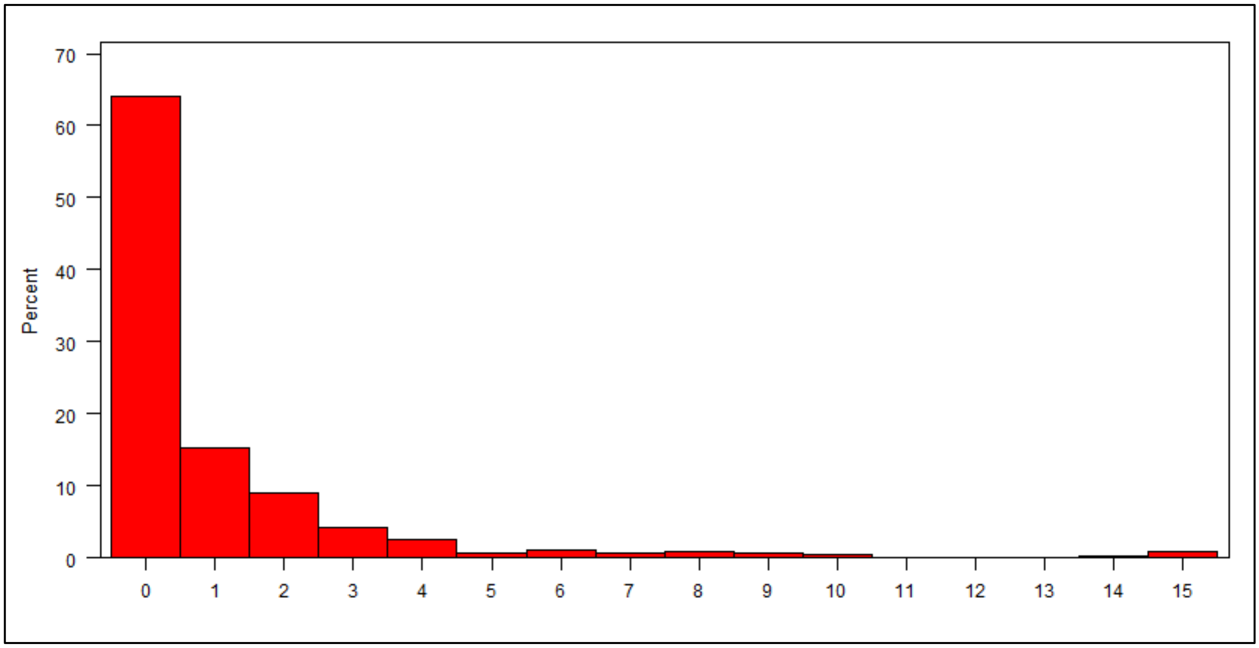


Figure 3.5: Histogram of number of sex partners in the last three months among all respondents at baseline (n=2533)

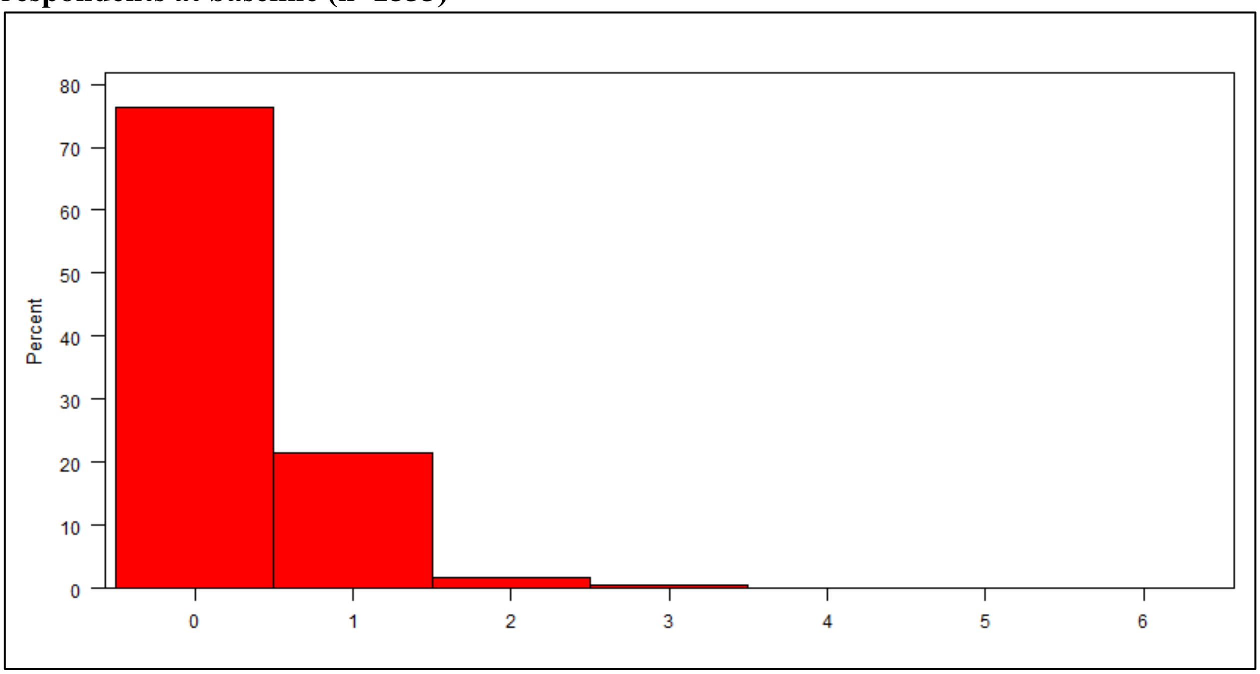


Figure 3.6: Histogram of number of alcohol outlet visits in the last six months in the full sample at baseline (n=2533)

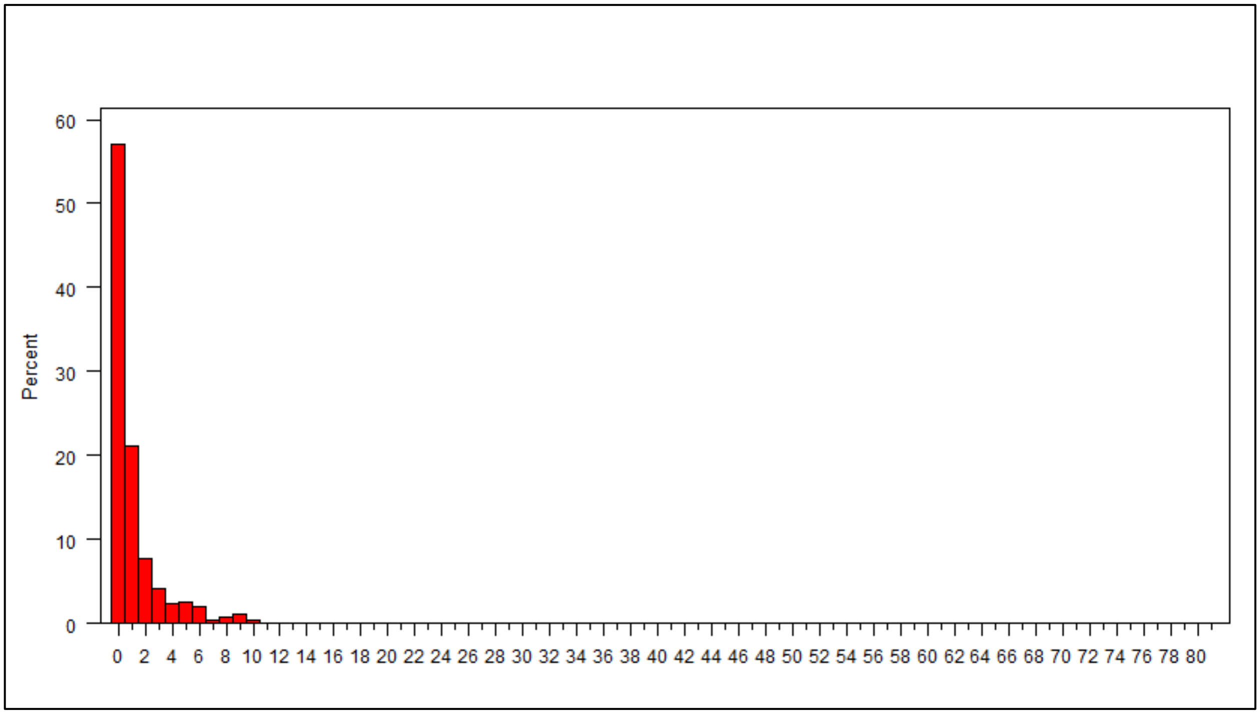


Figure 3.7: Directed acyclic graph depicting the hypothesized relationship between visits to alcohol outlets (AO) and sexual risk.

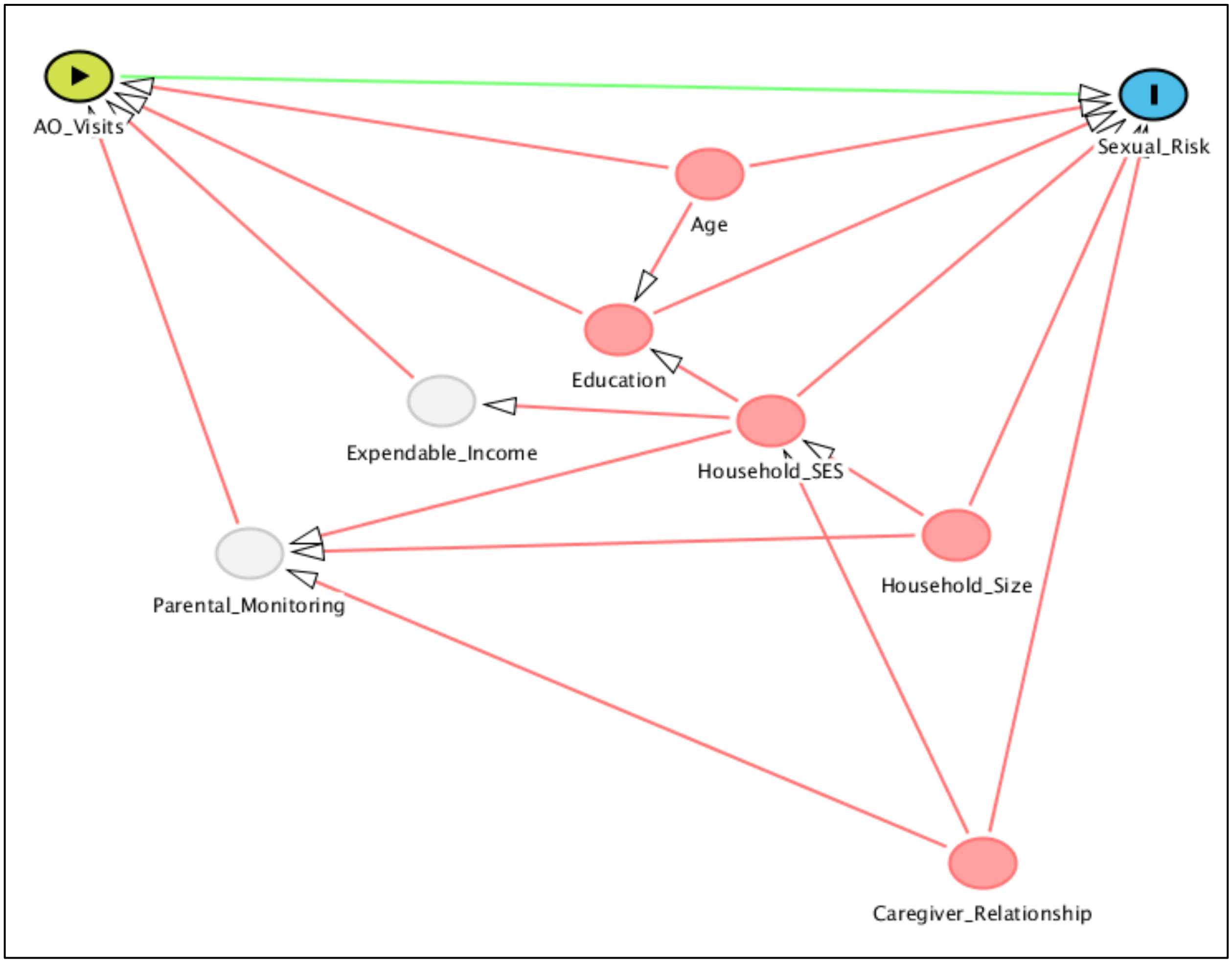


Figure 3.8: Expected statistical power to explore the association between exposure to alcohol outlet visits and transactional sex over a range of odds ratios from 1.1 to 2.0

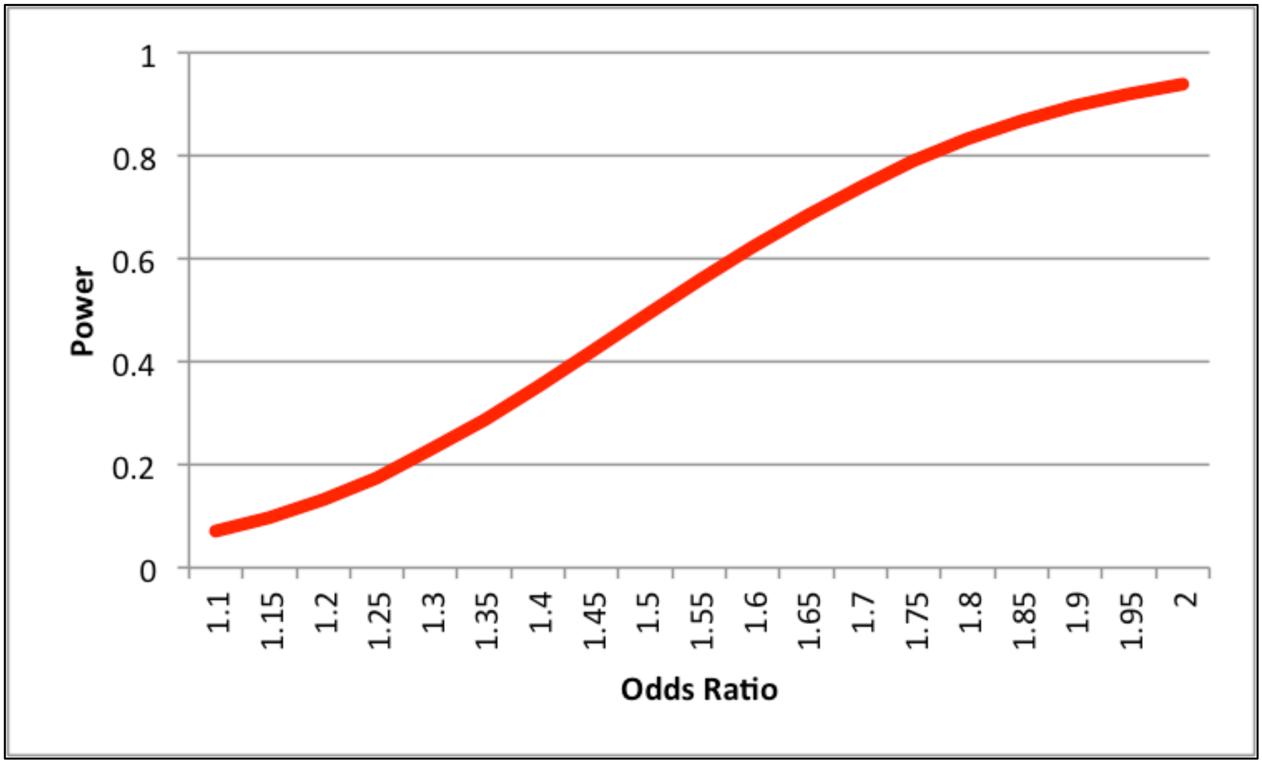


Figure 3.9: Histogram of the individual-level distribution of alcohol outlet number within home village

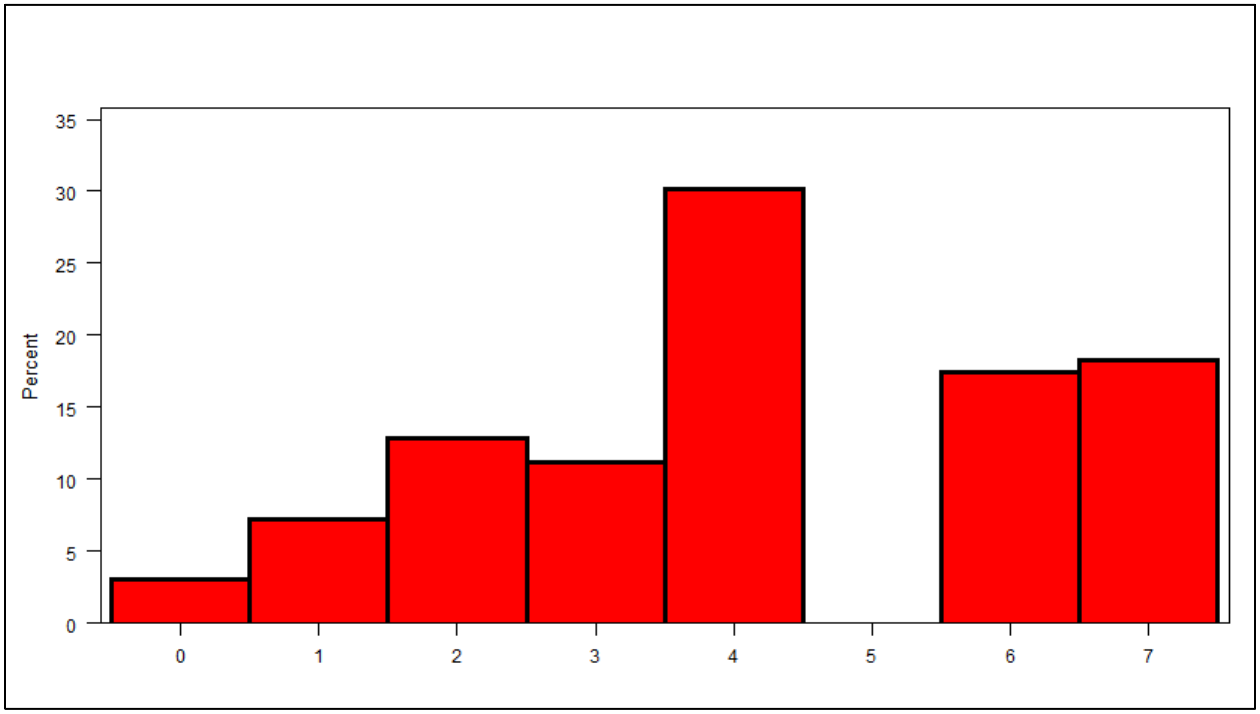


Figure 3.10: Directed acyclic graph depicting the hypothesized relationship between number of alcohol outlets per village and HSV-2 infection

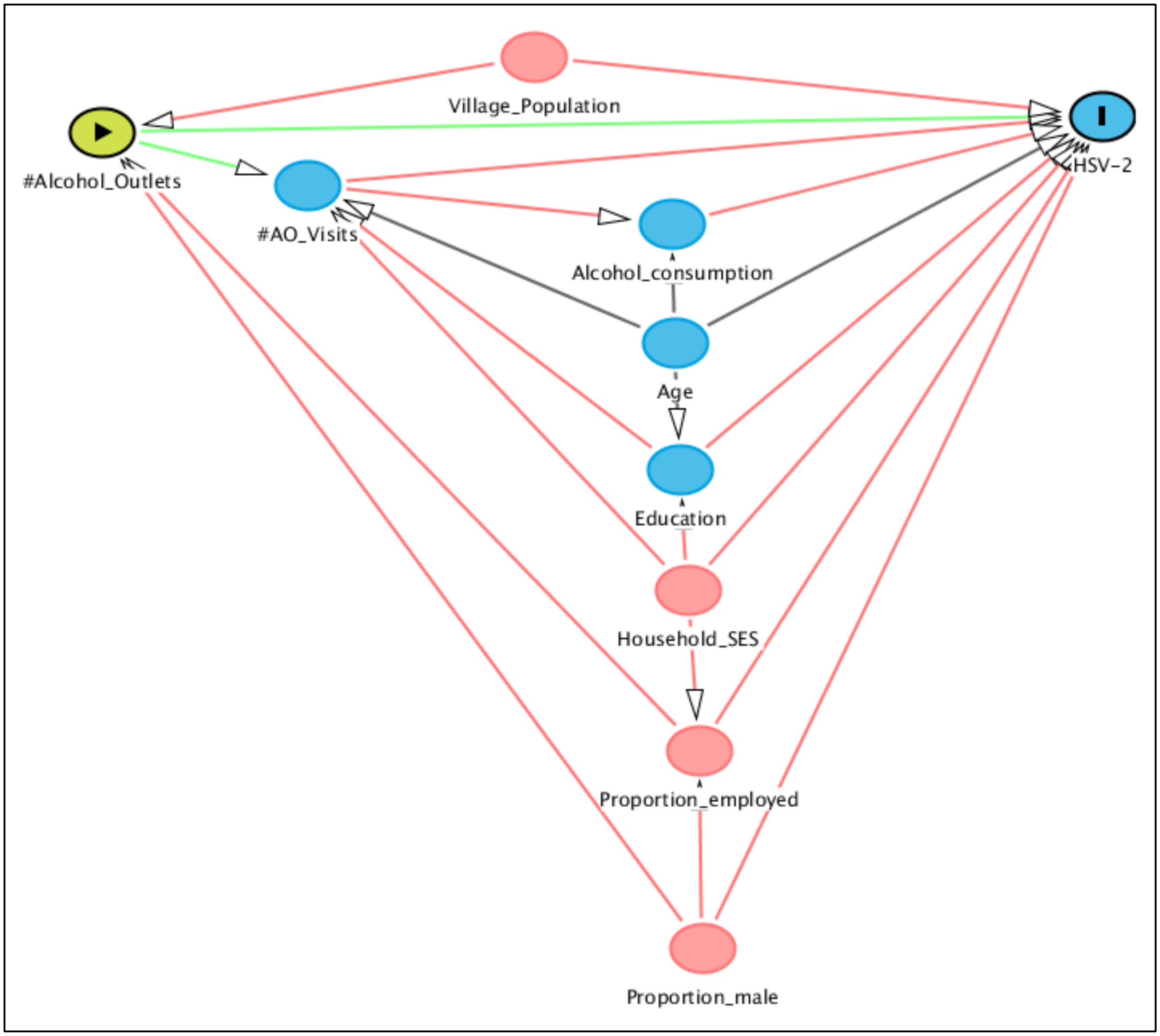
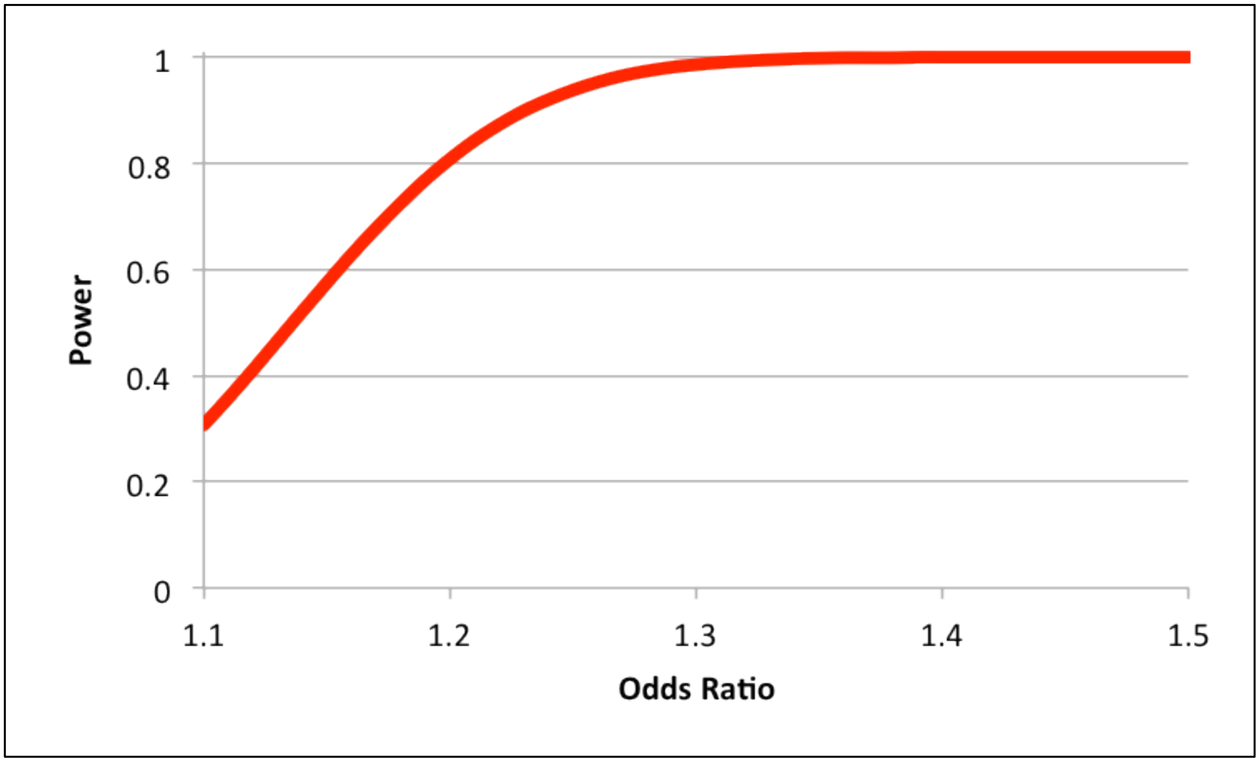


Figure 3.11: Expected statistical power to explore the association between number of alcohol outlets per village and HSV-2 infection across a range of odds ratios from 1.1 to 1.5



CHAPTER IV: Aim 1
Relationship between School Dropout and Teen Pregnancy
among Rural South African Young Women

1 Introduction

Teen pregnancy leads to negative health and social outcomes for both teen mothers and their children.(5-7) Over 16 million births occur to mothers under the age of 20 every year, and nearly all (95%) occur in developing countries.(9) In South Africa in particular, about one-third of young women have borne a child by the age of 20,(10, 11) and these births are often accompanied by social stigma and family-imposed sanctions.(8) Identifying protective factors against teen pregnancy, particularly in areas where it is most common, could inform interventions and prevent a large burden of negative health outcomes.

School is an important, structured place for young women and may protect against teen pregnancy.(51) In South Africa, nearly all young people are enrolled in school for most of the compulsory school-aged years (ages seven to 15 years);(35) however, school dropout rates begin to rise after the mandatory enrollment age.(36) As supported by *Social Control Theory*, attending school may provide large periods of structure and supervision in young adults' lives which could reduce opportunities for risky sexual activity.(24, 25) This potential relationship is supported by the links between low educational attainment (a measure of the highest grade reached by an individual) and teen childbearing,(41) sexual risk behaviors,(41-43) and HIV outcomes.(20, 42, 44, 45) However, though educational attainment provides a measure of the education to which the individual has been exposed in her lifetime, it does not establish a temporal relationship between exposure and outcome. Thus, inferences about the directionality of the relationship between educational attainment and sexual risk are limited.

An alternative education measure is school enrollment status. School dropout is associated with sexual risk behaviors including multiple partnerships, older partner age, unprotected sex, and transactional sex,(52, 53) and to higher HIV prevalence.(52, 54) The relationship between school enrollment and teen pregnancy is more complex. Teen pregnancy among school enrollees leads to subsequent school dropout.(55-57) Conversely, school dropout among non-pregnant teens may lead to subsequent pregnancy.(60) Young women who receive incentives to stay in school report fewer pregnancies,(58, 59) though the effects of school enrollment cannot be easily isolated from the effects of the incentive itself. Thus, the relationship between school enrollment and teen pregnancy remains unclear for two primary reasons: 1. few high-quality studies have been conducted, and 2. cross-sectional analyses of teen pregnancy or related sexual risk outcomes limit the ability to assess directionality of effect.(52-56) Using longitudinal census data from rural South Africa, in this paper we estimated the effect of school enrollment on incident teen pregnancy.

2 Methods

2.1 Study sample

We assembled a cohort of South African young women drawn from the Agincourt sub-district which is covered by a health and socio-demographic surveillance system (HDSS), located in a rural area of the Bushbuckridge municipality in Mpumalanga Province. The Agincourt HDSS is run by the Medical Research Council/Wits University Rural Public Health and Health Transitions Research Unit. This prospective community study has collected vital events data on all individuals living within the sub-district since 1992. The annual census update collects information on all births, deaths, and in- and out-migration, while other individual- and household-level data, such as educational attainment and household socio-economic status, are collected at less frequent, though regular, intervals.(133, 134) Community, household, and individual consent have been obtained for all Agincourt HDSS research since its inception. Ethics approval was obtained from the University of the Witwatersrand's Committee for Research on Human Subjects (updated # M110138; original # M960720) and the Mpumalanga Province Health Research and Ethics Committee. Research ethics approval for this secondary analysis was obtained from the Office of Human Research Ethics at the University of North Carolina – Chapel Hill (#13-2013).

We constructed the cohort by identifying all young women between the ages of 12-18 years who lived in the study area between 2000-2012. Each young woman began contributing person-time to the study on her 12th birthday. Young women who moved into the study site after their 12th birthday or whose 12th birthday occurred before January 1, 2000 were incorporated as late entries. Person-time for each young woman was censored when the first of the following events occurred: 1. administrative censoring date of December 31, 2011; 2. 18th birthday; 3. first report of educational attainment greater than or equal to 12 years; 4. loss to follow-up or death; or 5. conception date of first pregnancy. Young women with recorded pregnancies or high school graduation prior to the start of follow-up were excluded from this analysis. Young women with multiple records due to internal household movements within the study site were also excluded due to an inability to reliably link them to pregnancy outcomes.

2.2 Variables

The primary outcome in this analysis was *incident teen pregnancy*. In the yearly census update, fieldworkers administer a pregnancy outcome module to each recently pregnant woman to collect information on all pregnancies that took place in the previous 12 months, including the date of delivery. We calculated the estimated conception date for each pregnancy by subtracting 280 days (40 weeks) from the date of delivery. Conception dates for pregnancies recorded as abortions (n=9) were estimated based on the duration of pregnancy recorded in the census.

We restricted our outcome to first pregnancy for two reasons: 1) given the relatively young ages in our cohort, we did not expect to see many young women to go on to have second pregnancies during follow-up, particularly given that rural South African women tend to delay the timing of their second child;(144, 145) and 2) young women who have previously had a child may be less likely to go on and have further schooling, even though previous pregnancy does not preclude future schooling, particularly in South Africa.(55, 57)

The primary exposure in this analysis was *school enrollment status*. School enrollment was coded as a time-varying, binary exposure equal to one when the participant was enrolled in school and zero when not enrolled in school. An education status module was administered in 1997, 2002, 2006, 2009, and 2012. At the time of each module, field workers recorded the highest education level each young woman had achieved and whether or not she was currently enrolled in school. We recreated a continuous school enrollment history over the follow-up period for each young woman by inferring enrollment status at any given time by examining the change in educational attainment between module years. For example, if a young woman reported grade 7 attainment in 2006 and grade 10 attainment three years later in 2009, we assumed continuous enrollment for 2006, 2007, and 2008. Observations with illogical education data patterns (e.g. educational attainment decreases over time) were removed from analysis. Observations with non-linear, though plausible, education data patterns were flagged for temporary dropouts, grade repeats, or grade skips and incorporated into a sensitivity analysis.

We used a directed acyclic graph to identify a minimally sufficient adjustment set of potential confounders of the relationship between school enrollment and teen pregnancy. Specifically, we explored the potential confounding effects of: *age*, calculated in years from birthdate; *gender of household head*, defined as the gender (male or female) of the individual reported as the head of household; *household head employment status*, defined as whether or not the household head reports employment (updated in 2000, 2004, and 2008); *household head secondary education*, defined as whether or not the household head reports at least 12 years of educational attainment; *household size*, defined as the total number of individuals (including any temporary migrants) reporting membership in the participant's household at the time of interview; *household socioeconomic status (SES)*, measured as a composite index of general SES based on household assets (updated every other year starting in 2001);(146) and *calendar year*, an indicator for the year in which each participant turned 12 years old. We further considered age as a potential effect measure modifier by partitioning the person-time contributed by young women between the ages of 12-14 years and the ages of 15-17 years.

For observations with missing covariate information, we used multiple imputation to impute the missing values using the predictive ability of other observed time points of the missing covariates as well as other observed variables, including the outcome (pregnancy). We imputed 30 datasets to reduce the sampling variability using a Markov chain Monte Carlo method. We compared the distribution of covariates before and after imputation to assess comparability.

In a secondary analysis, we examined whether estimated conception dates of pregnancies occurred more frequently during times when school was in session versus school holidays. To create the time-varying *school session* variable, we used the exact dates of school holidays between 2000-2011 from official South African public school calendars published on-line.(147) All follow-up time that occurred during the active school year (typically mid-January through the beginning of December) was coded as exposed to 'school term.' All follow-up time that occurred during the summer holiday between school years was coded as exposed to 'school vacation.'

2.3 Statistical analyses

We used Cox proportional hazards (PH) models to compare the hazard of first pregnancy among those enrolled in school to those not enrolled in school. The origin for each participant began on her 12^{th} birthday with age as the time scale. We partitioned the dataset so that young women who switched school enrollment status during follow-up could contribute both exposed and unexposed person-time.

As age was the time scale of the Cox PH model, it was adjusted for implicitly in unadjusted and adjusted analyses. In all adjusted analyses, we coded originally continuous covariates as indicated by log-likelihood tests comparing different functional forms. Using this rule, calendar year was coded categorically with a cutpoint at before/after 2006, while household size and household SES were coded linearly. The binary structure of household head employment, gender, and educational attainment was maintained in adjusted analyses. Time-varying covariates (household size, household SES, household head employment status, and household head secondary education) were updated each time new data were provided.

We performed sensitivity analyses to assess whether our results were robust to potential misclassification of school enrollment exposure. First, to address the uncertainty around the sequence of events when estimated exposure and outcome occurred around the same time, we restricted the sample to exclude all person-time contributed by young women with estimated conception dates occurring within one year of school dropout. Next, to address the uncertain enrollment status of young women who reported non-linear grade progression between module years, we restricted the sample to remove those flagged as potential grade repeats and grade skips. As a final sensitivity analysis, we ran the statistical model on a restricted sample combining both of the restrictions mentioned above.

Effect measure modification by age was assessed by comparing the hazard ratio at young ages (among person-time contributed between the ages of 12 through 14) to that at older ages (among person-time contributed between the ages of 15 through 17). To calculate these age-specific hazard ratios, we added an interaction term between the dichotomous age variable and school enrollment variable to each model. Wald statistics of the interaction term were examined to inform our assessment of the difference between effect sizes at young and older ages.

To further explore whether time spent in school was protective against teen pregnancy, as a secondary analysis, we investigated the association between school year calendar time (summer vacation versus school session) and incident teen pregnancy among those enrolled in school. To do this, we pooled all the person-time contributed during the school year and all person-time contributed during the summer holidays. We then used a Poisson regression model to calculate the incidence rate ratio for teen pregnancy, comparing the incidence of conception during the school term to that occurring during school vacation.

3 Results

Overall, 22,661 young women between the ages of 12-18 lived in the study site for at least some period between January 1, 2000 and December 31, 2011 (Figure 4.1). We removed 512 young women who moved households within the study site during follow-up. A further 136 young women who graduated from high school before follow-up began and 1008 young women who experienced a pregnancy before follow-up began were also removed. Young women with previous pregnancies typically entered the cohort late with a mean age of 16.6 years. Also, 1462 and 4086 young women were removed for logically inconsistent education data and missing education data, respectively. Although the number of observations lost due to missing exposure data represented approximately 20% of all eligible young women, only 8% of total person-years were lost. Our final sample included 15,457 young women.

On average, at the beginning of follow-up, young women were about 13 years old and lived in a household with about seven other people (Table 4.1). Nearly half of the participants' households (44%) were headed by females, and, although over half (59%) of the household heads were employed, only a small fraction (11%) had graduated from high school. The household SES index was, on average, 2.3 (range: 0.3 – 3.7); however, baseline SES data were missing for over 35% of all observations. Covariate data for household size, household head employment and household head secondary education were missing for over 15% of the sample. After using multiply-imputed datasets to correct for missing covariate information, the distribution of the covariates was qualitatively similar to those in the non-imputed dataset.

A total of 2,140 first pregnancies occurred during the 48,271 person-years contributed (Table 4.2). The unadjusted pregnancy rate was much lower among young women enrolled in school (4.1 pregnancies per 100 person-years) compared to young women not enrolled in school (11.7 pregnancies per 100 person-years). Accordingly, in the full sample, the hazard of pregnancy was nearly 50% lower among young women enrolled in school compared to young women not enrolled in school [hazard ratio (HR) (95% confidence interval (CI)): 0.54 (0.48, 0.62)]. Adjustment for calendar year, household SES and size, and household head gender, employment, and secondary education did not substantially affect the association in the complete case analysis [adjusted HR (aHR) (95% CI): 0.56 (0.48, 0.65)] or with multiple imputation [multiple imputation-aHR (MI-aHR) (95% CI): 0.57 (0.50, 0.65)].

These results were reasonably robust to potential misclassification of school enrollment exposure. When observations with pregnancies occurring within one year of school dropout were removed (Restricted sample 1), the effect of school enrollment on teen pregnancy was still protective, though attenuated [MI-aHR (95% CI): 0.84 (0.71, 0.99)]. When all observations with non-linear grade progression were removed (Restricted sample 2), the results were qualitatively similar to those of the full sample [MI-aHR (95% CI): 0.54 (0.46, 0.64)]. Similar results were also found for Restricted sample 3, where the two previous restrictions were combined [MI-aHR (95% CI): 0.70 (0.57, 0.86)].

The effect of school enrollment on teen pregnancy differed by age group (Table 4.3). The protective effect of school enrollment was significantly stronger among participants between the ages of 12-14 years [MI-aHR (95% CI): 0.36 (0.25, 0.54)] compared to participants between the ages of 15-18 years [MI-aHR (95% CI): 0.59 (0.52, 0.68)]. Significant effect measure modification was observed in the unadjusted, complete-case adjusted, and multiply-imputed adjusted analyses.

The results of the Poisson model examining the association between school calendar year and teen pregnancy are shown in Table 4.4. A small association was observed among school enrollees. Conception dates for young women enrolled in school were about 10% less likely to occur during the school year compared to during summer vacation [IRR (95% CI): 0.90 (0.78, 1.04)].

4 Discussion

We examined the association between school enrollment and incident teen pregnancy in a large, longitudinal cohort of rural, South African young women. Overall, we found very high rates of teen pregnancy among school dropouts and the hazard of pregnancy was significantly lower during times of school enrollment compared to non-enrollment. A secondary analysis among school enrollees suggested that pregnancy rates may be lower during school term compared to summer holiday. Both the primary and secondary results suggest that schooling is protective against pregnancy and lend support to the hypothesis that the structured and supervised environment may contribute to the observed protective effect.

The use of prospectively collected census data allowed us to assess the temporal relationship between school enrollment and teen pregnancy. Most previous studies linking school enrollment to sexual risk outcomes had limited ability to make inferences about the directionality because they used a cross-sectional design.(52-56) Nonetheless, our observations are consistent with the protective associations observed in these prior studies and in a single longitudinal study from the United States over 20 years ago.(60) Our study extends the findings of a protective effect of school enrollment geographically to South Africa and temporally to the first decade of the 21st century.

School enrollment was slightly more protective against teen pregnancy among younger compared to older adolescents. Given these observations, we speculate that younger adolescents are at particularly high risk for teen pregnancy after dropout and that older dropouts may be given better alternatives to teen parenthood due to higher educational attainment. However, overall, the rates of pregnancy and school dropout were much lower in younger compared to older teens. Therefore, although the relative effect of school enrollment on teen pregnancy was greater among younger adolescents, the absolute effect was still greater among older adolescents.

School enrollment status was likely measured with some error. Enrollment data were collected with relatively long intervals between collection periods and misclassification may have occurred due to unreliable reporting from interviewees. We performed two analyses to assess the sensitivity of our results to exposure misclassification: 1. Removal of all observations with a pregnancy within one year of school dropout and 2. Removal of all observations with suspected unreliable reporting (non-linear grade progression). The findings of these analyses demonstrated that the results were reasonably robust to alternative exposure specifications.

Additionally, we assumed normal pregnancy durations of 40 weeks to calculate conception dates. If pregnancy and school dropout co-occurred with close temporal proximity, misspecification of the pregnancy duration could influence whether the pregnancy was classified as exposed or unexposed. It is plausible that young women who drop out of school are more likely to have preterm births and shorter pregnancy durations, perhaps due to lower socioeconomic status and decreased access to care. However, the first sensitivity analysis removed all observations with pregnancies within one year of dropout, effectively removing all observations that could have had school enrollment misclassification due to misspecified pregnancy duration. The fact that we still observed a protective, though attenuated, effect in this analysis indicates that pregnancy duration misspecification alone was not a likely explanation for our results.

The year before and after school dropout may be high-risk periods for teen pregnancy. The first sensitivity analysis removed all observations with a pregnancy during this potentially high-risk time. The effect estimate moved toward the null, though a significantly protective association was still observed. One explanation for the attenuation of the effect is that some young women may leave school midway through the year due to a pregnancy. This partial year of schooling would not be recorded by the census educational attainment measure and could be falsely inflating the observed effect estimate in the full sample. Alternatively, the first year after school dropout may be a particularly high-risk time for teen pregnancy and removal of these observations artificially attenuates the effect size in the restricted sample. Although the individual contributions of these two explanations cannot be directly assessed, both are likely to have had some impact.

The analysis employed here is likely the optimal approach to estimate the effect of school enrollment on teen pregnancy. Employing a randomized study design to answer this research question is not viable because the widely beneficial school enrollment exposure cannot be randomized. Novel study designs that randomize incentives to stay in school could be a good alternative,(148) but income effects of the incentive cannot be easily distinguished from the direct effects of school enrollment on sexual risk outcomes. To address some of the drawbacks of analyzing observational data, we adjusted for measured covariates to close all identified confounding paths between exposure and outcome. However, the possibility remains that the observed results were influenced by uncontrolled confounding.

Finally, pregnancies that did not end in a live birth were likely underrepresented in this analysis. Miscarriages, perinatal deaths occurring during the inter-censal period, and abortions were all likely to be underreported in the census. Abortions, in particular, were reported with a lower frequency than would be expected given the most recent abortion rates in South Africa.(149, 150) Unreported abortions could plausibly produce differential outcome misclassification with respect to school enrollment status. However, the magnitude of the difference in pregnancy rates between enrolled and non-enrolled young women suggest that the number of unreported abortions among enrollees would have to be improbably high to completely account for the observed effect. Because they are not in direct control of the participant, unreported miscarriage and perinatal death rates were likely nondifferential with respect to school enrollment status.

The development of better prevention interventions for South African young women will be critical to prevent the continued high burden of teen pregnancy in this vulnerable population. This study identified school enrollment as a protective factor for teen pregnancy, an outcome with significant negative health and social repercussions throughout the life course of the teen mother and child. Interventions designed to keep young women in school or provide recent dropouts with sexual health information may reduce the burden of teen pregnancy to yield better long-term health outcomes in this population.

Table 4.1: Baseline covariates of 15,457 young women in Agincourt, South Africa

	Prior to imputation		Post-imputation
Continuous covariates	**Mean (SD)**	**Missing**	**Mean**
Age	13.1 (1.7)	0	13.1
Calendar year at 12th birthday	2003.7 (4.6)	0	2003.7
Household SES	2.3 (0.5)	5495	2.2
Household size	6.8 (3.7)	4291	6.4
Binary covariates	**N (%)**	**Missing**	**%**
Female-headed household	6719 (43.9)	146	43.9
Household head employed	7321 (59.0)	3038	58.2
Household head secondary education	1276 (10.6)	3467	9.8

Table 4.2: Association between school enrollment and teen pregnancy among full and restricted samples of 15,457 young women in Agincourt, South Africa, 2000-2012

Model	Pregnancies	PY	Rate/ 100 PY	Unadjusted HR (95% CI)	p	Complete case aHR[a] (95% CI)[b]	p	MI aHR[c] (95% CI)[b]	p
Full sample *(n=15,457)*									
Not enrolled	261	2238	11.7	-	-	-	-	-	-
Enrolled	1879	46033	4.1	0.54 (0.48, 0.62)	<0.0001	0.56 (0.48, 0.65)	<0.0001	0.57 (0.50, 0.65)	<0.0001
Restricted sample 1: ***Remove pregnancies within 1 year of dropout*** *(n=15,172)*									
Not enrolled	159	2181	7.3	-	-	-	-	-	-
Enrolled	1724	45498	3.8	0.79 (0.67, 0.94)	0.01	0.82 (0.72, 1.02)	0.06	0.84 (0.71, 0.99)	0.03
Restricted sample 2: ***Linear grade progressors*** *(n=7,776)*									
Not enrolled	186	1780	10.4	-	-	-	-	-	-
Enrolled	576	17267	3.3	0.52 (0.44, 0.62)	<0.0001	0.50 (0.40, 0.62)	<0.0001	0.54 (0.46, 0.64)	<0.0001
Restricted sample 3: ***Restrictions 1 & 2 combined*** *(n=7,623)*									
Not enrolled	124	1746	7.1	-	-	-	-	-	-
Enrolled	498	17037	2.9	0.66 (0.54, 0.80)	<0.0001	0.61 (0.47, 0.79)	0.0002	0.70 (0.57, 0.86)	0.0006

[a] Complete case adjusted analysis restricted to person-time contributed with complete covariate information (n=10,797)

[b] Adjusted for calendar year of 12th birthday (coded categorically with cutpoints before/after 2006), gender of household head, educational attainment of household head (coded dichotomously at above/below secondary school graduation), employment of household head (coded dichotomously (yes/no) for reported employment at most recent census), household SES (coded linearly), and household size (coded linearly).

[c] Multiple imputation adjusted analysis uses full sample with missing covariates imputed using multiply imputed datasets (n=15,457)

PY=person-years; HR=hazard ratio; aHR=adjusted hazard ratio; CI=confidence interval; MI=multiple imputation

Table 4.3: Association between school enrollment and teen pregnancy, stratified by age, among a cohort of 15,457 young women in Agincourt, South Africa, 2000-2012

Model	Pregnancies	PY[a]	Rate/ 100 PY	Unadjusted HR (95% CI)	p	Complete case aHR[a] (95% CI)[b]	p	MI aHR[c] (95% CI)[b]	p
Ages 12-14 *(n=12,889)*									
Not enrolled	29	734	4.0	-	-	-	-	-	-
Enrolled	297	26510	1.1	0.34 (0.23, 0.50)	<0.0001	0.31 (0.19, 0.48)	<0.0001	0.36 (0.25, 0.54)	<0.0001
Ages 15-17 *(n=11,099)*									
Not enrolled	232	1496	15.5	-	-	-	-	-	-
Enrolled	1583	19262	8.2	0.58 (0.50, 0.67)	<0.0001	0.59 (0.50, 0.70)	<0.0001	0.59 (0.52, 0.68)	<0.0001
Interaction term statistics[d]				**χ^2 statistic**	**p**	**χ^2 statistic**	**p**	**t-statistic**	**p**
				6.73	0.01	7.38	0.007	-2.60	0.02

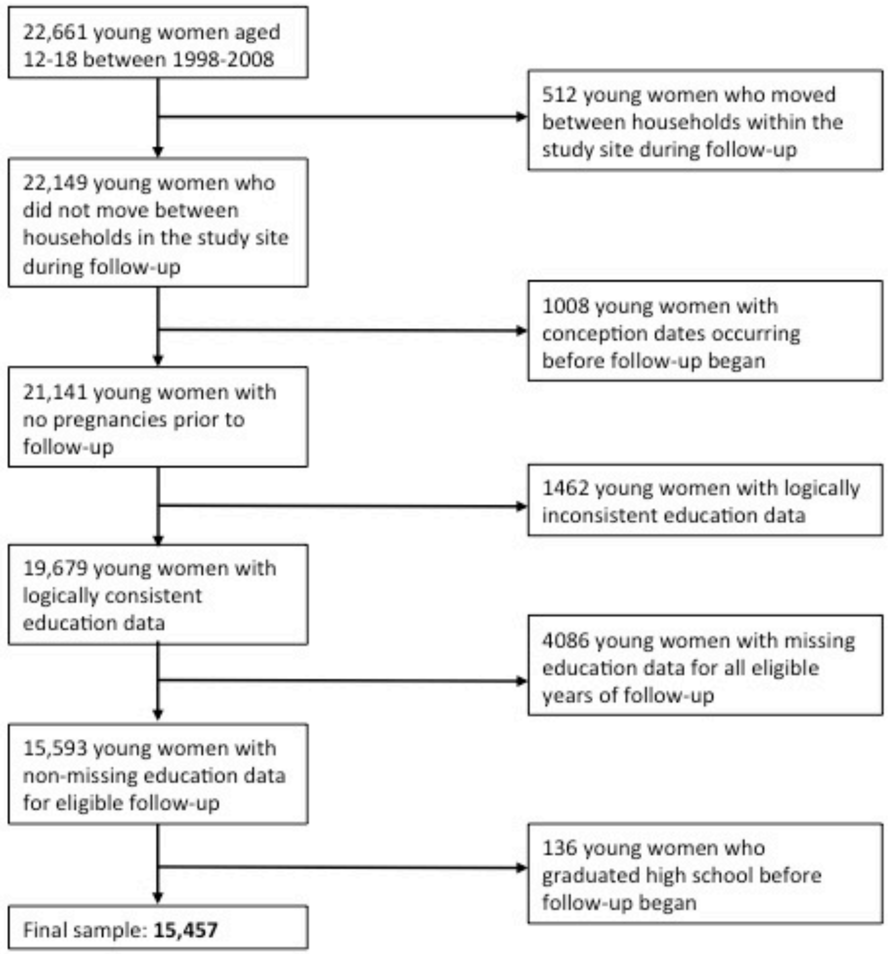
[a] Complete case adjusted analysis restricted to person-time contributed with complete covariate information
[b] Adjusted for calendar year of 12th birthday (coded categorically with cutpoints before/after 2006), gender of household head, educational attainment of household head (coded dichotomously at above/below secondary school graduation), employment of household head (coded dichotomously (yes/no) for reported employment at most recent census), household SES (coded linearly), and household size (coded linearly).
[c] Multiple imputation adjusted analysis uses full sample with missing covariates imputed using multiply imputed datasets
[d] χ^2- and t-statistics testing the null hypothesis that the beta parameter for the interaction term between age and school enrollment status is zero
PY=person-years; HR=hazard ratio; aHR=adjusted hazard ratio; CI=confidence interval; MI=multiple imputation

Table 4.4: Association between school holidays and teen pregnancy among 14,759 young women while enrolled in school in Agincourt, South Africa, 2000-2012

Calendar period	Pregnancies	PY	Rate/100 PY	IRR (95% CI)	p
School vacation	220	4924	4.5	-	-
School term	1653	41114	4.0	0.90 (0.78, 1.04)	0.1

PY=person-years; IRR=Incidence rate ratio; CI=confidence interval

Figure 4.1: Flowchart of cohort construction of young women, aged 12-18, in Agincourt, South Africa



CHAPTER V: Aim 2

Visits to Alcohol Outlets, Sexual Risk Behavior, and HSV-2 Infection among Female Adolescents in South Africa

1 Introduction

Frequenting alcohol outlets (establishments where alcohol is sold and consumed) may influence sexual risk. Alcohol use and abuse is associated with increased sexual risk throughout the world,(67-71) regionally among populations in sub-Saharan Africa (72-74), and specifically in the Republic of South Africa.(75, 76) Additionally, characteristics of the outlets themselves (e.g. music, dim lights, lack of condoms) (87) and the network of people who typically frequent outlets (e.g. older men willing to exchange money for sex) (74, 89) may create favorable environments for risky sexual activity. Relatedly, those inclined to risky sexual activity may visit alcohol outlets with the intention of engaging in risky behaviors.

Although the sexual risk profiles of individuals recruited at alcohol outlets have been characterized as risky,(74, 93, 101-104) the relative difference in sexual risk between those with and without alcohol outlet exposure is less clear. Adults who patronize alcohol outlets tend to engage in riskier sexual behaviors than those who do not.(88, 105-107) However, this association has not been studied in adolescent-specific or female-specific populations.

Young women in South Africa have moderately high alcohol consumption (151, 152) and are at exceptionally high risk for negative sexual risk outcomes like HIV infection (3, 4) and teen pregnancy.(10, 11) Identifying novel risk factors and intervention targets is critical to reduce the burden of these outcomes in this vulnerable population. Alcohol outlets are appealing as targets for sexual risk reduction because access can be modified through regulation and they can serve as locations in which to deliver prevention interventions.(90, 153) In this paper, we aim to examine the association between alcohol outlets and sexual risk in a population-based sample of rural South African young women.

2 Methods

2.1 Population

To explore the association between alcohol outlet visits and sexual risk, we conducted a cross-sectional analysis using data collected at baseline in HPTN 068. HPTN 068 is a Phase III randomized trial in the rural Bushbuckridge sub-district in the Mpumalanga province of South Africa where the Medical Research Council/Wits University Rural Public Health and Health Transitions Unit has run a health and socio-demographic surveillance system since 1992.(133) This study has the aim to determine whether cash transfers conditional on school attendance reduce HIV risk in young women and enrolled a total of 2533 female adolescents, aged 13 to 20 years, currently enrolled in school, and not currently pregnant or married. Ethical approval for the parent study and secondary analysis was provided by the Office of Human Research Ethics at the University of North Carolina-Chapel Hill (#10-1868; #13-2013). Additional approval for the parent study was provided by the Mpumalanga Province Health Research and Ethics Committee.

2.2 Variables

All variables were collected in HPTN 068 baseline biological testing and surveys. To minimize the bias that may come from providing sensitive information to an interviewer, an ACASI (audio computer-assisted interviewing) component was incorporated into the survey. The ACASI component allowed participants to privately read and listen to audiotaped questions and log their responses in a computer, without interacting with an interviewer.

The exposure, *alcohol outlet visits,* was a count variable in response to the question: "How many times in the past six months have you been to a tavern/shebeen?" We examined the distribution and removed observations with implausibly high responses (n=1) and responses with repeated single digits (i.e. 11, 22, 33) that were likely a result of unintentional double-striking in ACASI (n=26). We categorized the exposure with dichotomous cutpoints (0 visits versus ≥1 visits) and categorical cutpoints to separate those with no exposure (0 visits), low exposure (1-5 visits - on average, fewer than one visit per month), and high exposure (≥6 visits – on average, one or more visits per month).

We examined four sexual risk outcomes. *Sex partners* was the reported number of sex partners in the last three months. *Unprotected sex acts* was constructed by subtracting the total number of condom-protected vaginal sex acts from the total number of vaginal sex acts over the last three months. For both count variables, those with no prior sexual activity received a zero value. We recoded two sex partner responses likely due to double striking based on corroborating information in the reported lifetime sex partner number and seven sex act responses likely due to double striking based on corroborating information in the reported number of protected sex acts. For all analyses, we categorized sex partners and unprotected sex acts each into three categories: zero, one, and greater than one, as there were few responses greater than two for either variable. *Transactional sex* was a dichotomous variable constructed from responses to questions regarding whether the participant had received money or gifts from her most recent sex partner and whether she felt obligated to have sex in return. Testing for *prevalent herpes simplex virus 2 (HSV-2) infection* was performed at baseline using Kalon™ HSV-2 gG2 ELISA (Kalon Biological, Ltd., Surrey, United Kingdom).(136)

We also explored the influence of several key covariates. Specifically, we examined *age*, in years, at baseline; *education,* the grade in which the young woman was enrolled at baseline; *household size,* the total number of people sharing a household with the participant; *primary caregiver relationship*, the relation between the young woman and her primary caregiver: daughter, sibling, niece, grandchild, other; and *household socio-economic status (SES)*, a log-transformed measure of monthly household expenditures, per capita. We also calculated a dichotomous variable, *age for grade,* flagging young women who were older than they should be had they progressed linearly through school (e.g. above age 14 in grade 8). Finally, *alcohol consumption frequency* was defined categorically in response to the question: "How often do you drink alcohol?" with six responses ranging from "Never" to "More than once per week."

2.3 Statistical analysis

We used logistic regression models to estimate the association between alcohol outlet visits and the dichotomous transactional sex and HSV-2 outcomes. We used multinomial logistic regression models to estimate the assocation between alcohol outlet visits and the categorized sex partner and unprotected sex act outcomes. To assess whether the observed results were driven by the lack of sexual risk outcomes among those not yet sexually active, we also ran each model in a restricted sample of those who had experienced sexual debut.

To control for potential confounding, we identified a minimally sufficient adjustment set from a directed acyclic graph. We then assessed the functional form for each covariate with each outcome separately and coded them as suggested by likelihood ratio tests. To test for interaction by alcohol consumption, we first dichotomized both the alcohol outlet visits and alcohol consumption variables into "some" versus "no" exposure levels. We then included an interaction term between visits and consumption in each model. We used ordinal logistic regression models for the categorical sex partner and unprotected sex act outcomes to maximize statistical power and because the unstratified results suggested an ordered relationship. We compared the effect estimates among those with each risk factor alone to the effect estimate among those with both risk factors. We assessed the statistical significance of the interaction term in each model with likelihood ratio tests. Because interaction tests are typically underpowered, we considered p-values under 0.2 to be statistically significant.

3 Results

Overall, 2,533 young women were enrolled in HPTN 068 and completed baseline procedures (Figure 5.1). Only young women with complete exposure, outcome, and covariate data were included in the analyses. After removing observations with missing data, a total of 2,348, 2,348, 2,364, and 2,366 young women remained in the analytical samples for sex partners, unprotected sex acts, transactional sex, and HSV-2, respectively.

Over half (57%) of the young women with complete exposure information reported no alcohol outlet exposure in the last six months, over one-third (38%) reported low levels of exposure, and 5% reported high levels of exposure (Table 5.1). The average age of study participants (15.5 years), household size (6.2 people), and household SES (natural log of per capita expenditures: 5.2) did not vary significantly by level of alcohol outlet exposure. Most young women (74%) reported a mother or father as primary caregiver; these relationships were similar across levels of exposure. Current grade enrollment was significantly different across exposure frequencies. Those with higher exposure levels were more likely to be enrolled in lower grades (χ^2 p-value: 0.003) and more likely to be older than the expected age for their grade level (χ^2 p-value: 0.004).

About one quarter of the young women reported sexual debut prior to baseline (27%). A similar proportion reported any sex partners (24%) and few reported any unprotected sex acts (8%) in the last three months. Just over 3% of young women reported transactional sex with their most recent partner and approximately 5% tested positive for HSV-2. All outcomes qualitatively appeared to increase with increasing exposure levels. As expected, alcohol consumption appeared to be higher with increasing alcohol outlet exposure (χ^2 p-value: <0.0001); however, of the young women who reported visiting alcohol outlets, 87% reported no alcohol consumption.

Visiting alcohol outlets was positively associated with each sexual risk outcome (Table 5.2). Those reporting any alcohol outlet visits were more likely to report one versus zero [adjusted odds ratio (aOR) (95% confidence interval (CI)): 1.51 (1.21, 1.88)], and two or more versus zero sex partners [aOR (95%CI): 2.27 (1.29, 3.97)], compared to those with no visits. Similarly, those reporting any visits were more likely to report one versus zero [aOR (95% CI): 2.28 (1.52, 3.42)], and two or more versus zero unprotected sex acts [aOR (95% CI): 2.33 (1.53, 3.56)], compared to those with no visits. Alcohol outlet visits were also associated with increased transactional sex with most recent partner [aOR (95% CI): 1.63 (1.03, 2.59)] and prevalent HSV-2 infection [aOR (95% CI): 1.30 (0.88, 1.91)]. Generally, adjustment for age, grade, primary caregiver, household SES, and household size did not alter effect estimates appreciably from the unadjusted estimates. It is important to note, for rarer outcomes (transactional sex, HSV-2, and the highest category for sex partners and unprotected sex acts), the results were imprecise, with confidence limit ratios above 2.0, and the CI around the HSV-2 estimate included the null.

A dose-response relationship was not observed between alcohol outlet visits and any of the sexual risk outcomes. Using the more finely categorized alcohol outlet visit variable (0, 1-5, ≥6 visits) tended to produce effect estimates of similar magnitude for both low and high exposure. As a representative example, the odds ratios for transactional sex at both low [aOR (95%CI)=1.63 (1.01, 2.62)] and high [aOR (95%CI)=1.66 (0.66, 4.14)] exposure levels were qualitatively indistinguishable. Due to small numbers, results were imprecise for rarer outcomes and at the highest exposure level.

Visiting alcohol outlets was also positively associated with each sexual risk outcome in the restricted sample of 672 young women who had experienced sexual debut (Table 5.3). The results for each outcome were generally smaller but qualitatively similar in magnitude as those in the full sample. However, with the reduced sample size, estimates were less precise. For example, among the sexually active, those reporting any alcohol outlet visits were still more likely to report transactional sex, compared to those with no visits [aOR (95% CI): 1.43 (0.86, 2.36)].

Generally, alcohol outlet visits and alcohol consumption interacted to produce strong associations with the sexual risk outcomes (Table 5.4). The associations for each risk factor alone were predominantly positive, but small and not statistically significant. However, those who reported both alcohol outlet visits and alcohol consumption, compared to those with neither visits nor consumption, were strongly associated with a higher number of sex partners [aOR (95%CI): 5.20 (3.54, 7.63)], a higher number of unprotected sex acts [aOR (95%CI): 4.39 (2.65, 7.28)], transactional sex [aOR (95%CI): 2.87 (1.38, 5.98)], and HSV-2 infection [aOR (95%CI): 2.44 (1.29, 4.59)]. Statistical evidence for improved model fit from the addition of the interaction term ranged from weak to strong and was observed for all outcomes except transactional sex.

4 Discussion

We found that frequenting alcohol outlets was associated with increased sexual risk in South African female adolescents, especially when they consume alcohol. Young women who visited alcohol outlets reported more sex partners, more unprotected sex acts, and higher levels of transactional sex; and were more likely to have a prevalent HSV-2 infection, than young women who did not. The combination of exposure to both alcohol consumption and alcohol outlet visits was especially important - young women who reported both risk factors were much more likely to have experienced all four sexual risk outcomes, compared to those with neither risk factor.

This analysis provides important methodological improvements to and extends the generalizability of the current literature linking alcohol outlet visits to sexual risk.(88, 105-107) To our knowledge, this is the first study to explore and confirm this association among adolescents, in particular adolescent women from sub-Saharan Africa, a vulnerable and high-risk group. Also, this population was randomly sampled from within a health and socio-demographic surveillance site, yielding improved validity over studies using convenience sampling from within alcohol outlets.(88, 107)

We also provide the first evidence that there is little or no dose-response relationship between alcohol outlet visits and sexual risk. All previous studies dichotomized the alcohol outlet exposure, potentially masking interesting variations in the relationship with sexual risk. However, we found that even those with alcohol outlet exposure levels fewer than six visits over six months had similarly elevated sexual risk outcomes as those with higher levels of alcohol outlet exposure. It is important to note that the level of exposure to alcohol outlets is generally low in this population: less than half of the entire sample reported any visits to alcohol outlets; among them, most reported fewer than six visits over a six month period. However, this is a population of minors who, legally, should not be visiting alcohol outlets, and we found that even adolescents at young ages were at risk for exposure. Even given this constraint, low levels of alcohol outlet exposure were robustly associated with indicators of sexual risk.

We also found that, with the exception of transactional sex, young women who visited alcohol outlets and consumed alcohol had stronger associations with all sexual risk outcomes than anticipated given the associations with each risk factor alone. Although the alcohol consumption and alcohol outlet visits, as recorded, may not necessarily have occurred at the same time, it is plausible to speculate that these doubly-exposed young women were consuming alcohol within the drinking establishments themselves. Consumption of alcohol leads to disinhibition, impaired decision-making, and feelings of reduced sexual control.(74, 76, 78-80) These disruptions to the normal constraints on risky activity lead to increased sexual risk.(81, 82) In conjunction, characteristics typical of alcohol outlets (limited supervision, music, dim lights, unisex toilets, lack of condoms) may directly create favorable environments for risky sexual activity.(87) Sex partners are also often met in these places, particularly for young women, and these partners are often older men willing to exchange money for sex.(74, 89) This combination of risk disinhibition from alcohol consumption with the risk opportunities presented within alcohol outlets may explain the heightened sexual risk observed among the doubly-exposed.

The observed associations could plausibly be driven by young women who had not yet experienced sexual debut. Our primary analysis included young women with and without prior sexual debut. Those who were not sexually active, and therefore precluded from experiencing any of the sexual risk outcomes, were less likely to visit alcohol outlets. However, the results from the sub-analysis restricted to those having experienced sexual debut do not suggest this is the case. Visiting alcohol outlets appeared to have similar associations with each outcome among young women with sexual experience and among the full sample of young women.

It is also possible that we observe an association between alcohol outlet visits and sexual risk because young women inclined to sexual risk are also inclined to visit alcohol outlets, or because they visit alcohol outlets in order to meet like-minded partners or transactional sex partners. We attempted to minimize these possibilities by controlling for a set of covariates with hypothesized relationships to both exposure and outcome. However, the possibility remains that the observed association may be due to uncontrolled confounding.

The cross-sectional nature of the data does not allow us to assess the directionality of the observed association. In particular, we cannot say when the HSV-2 outcome occurred in relation to the alcohol outlet exposure. However, as this was a young cohort (mean age: 15.5), we expect that the dates of sexual debut and, therefore, earliest possible HSV-2 infection occurred relatively recently. Moreover, the information on sex partners, unprotected sex acts, and alcohol outlet exposure were collected with reference to the same three- to six-month time frame. The transactional sex outcome was restricted to refer to the most recent partner, so the timing was likely similar to the six-month exposure window as well.

There were also several factors that could have led to data error. First, participants may have had difficulty remembering precise counts of activities, such as the exact number of sex partners, sex acts, or alcohol outlet visits that occurred over three and six months. Second, an unintended consequence of the ACASI data collection method is the potential for measurement error. The data suggested that some participants entered unintended responses. For this reason, we carefully examined the distributions of alcohol outlet visits, number of sex partners, and number of sex acts, and removed implausibly high responses and suspected double strikes.

5 Conclusion

In the context of the high HIV/STI burden among South African female adolescents, identifying new risk factors and appropriate interventions for sexual risk is critical. This study suggests that young women who frequent alcohol outlets and consume alcohol have heightened sexual risk compared to those who do not. Consequently, alcohol outlets could be important places to reach high-risk adolescent women with sexual health interventions. Future studies that establish the directionality of the association will be able to inform whether sexual risk in adolescent women could be lowered by introducing interventions or policies to reduce their exposure to alcohol outlets.

Table 5.1: Demographic profile and sexual risk outcomes of 2533 young women, by frequency of alcohol outlet visits in the last six months

	Total (n=2533)	No visits (n=1354)	1-5 visits (n=897)	≥6 visits (n=124)	p-value
Continuous variables	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*ANOVA Test*
Age	15.5 (1.7)	15.5 (1.6)	15.6 (1.7)	15.5 (1.8)	0.8
Household size	6.2 (2.6)	6.2 (2.7)	6.1 (2.5)	6.0 (2.7)	0.6
Household SES[a]	5.2 (0.8)	5.2 (0.8)	5.2 (0.7)	5.2 (0.9)	0.8
Categorical variables	*N (%)*	*N (%)*	*N (%)*	*N (%)*	*χ^2 Test*
Education					
Grade 8	640 (25.3)	331 (24.5)	233 (26.0)	39 (31.5)	0.003
Grade 9	685 (27.0)	334 (24.7)	253 (28.2)	45 (36.3)	
Grade 10	696 (27.5)	393 (29.0)	236 (26.3)	26 (21.0)	
Grade 11	512 (20.2)	296 (21.9)	175 (19.5)	14 (11.3)	
Age for grade					
Expected age or younger	1790 (70.7)	995 (73.5)	604 (67.3)	82 (66.1)	0.004
Older than expected age	743 (29.3)	359 (26.5)	293 (32.7)	42 (33.9)	
Primary caregiver					
Mother/Father	1870 (73.8)	985 (72.7)	680 (75.8)	91 (73.4)	0.8
Grandparent	359 (14.2)	199 (14.7)	116 (12.9)	16 (12.9)	
Sibling	169 (6.7)	93 (6.9)	56 (6.2)	11 (8.9)	
Aunt/Uncle	104 (4.1)	57 (4.2)	34 (3.8)	6 (4.8)	
Other	26 (1.0)	16 (1.2)	10 (1.1)	0 (0)	
Alcohol consumption					
Never	2301 (91.1)	1273 (94.2)	786 (87.7)	99 (79.8)	<0.0001
< once a month	91 (3.6)	36 (2.7)	44 (4.9)	6 (4.8)	
Once a month	77 (3.1)	25 (1.9)	39 (4.4)	11 (8.9)	
2-3 times a month	21 (0.8)	9 (0.7)	8 (0.9)	3 (2.4)	
Once a week	21 (0.8)	5 (0.4)	13 (1.5)	1 (0.8)	
> once a week	15 (0.6)	4 (0.3)	6 (0.7)	4 (3.2)	
Sexual debut					
No	1851 (73.4)	1013 (75.1)	630 (70.5)	86 (69.4)	0.03
Yes	672 (26.6)	336 (24.9)	264 (29.5)	38 (30.7)	
Sex partners[b]					
0	1917 (76.3)	1061 (79.0)	643 (72.0)	90 (72.6)	0.001
1	539 (21.4)	260 (19.4)	220 (24.6)	31 (25.0)	
2+	58 (2.3)	22 (1.6)	30 (3.4)	3 (2.4)	
Unprotected sex acts[b]					
0	2176 (92.3)	1265 (94.4)	799 (89.5)	112 (90.3)	0.0002
1	78 (3.3)	36 (2.7)	39 (4.4)	3 (2.4)	
2+	103 (4.4)	39 (2.9)	55 (6.2)	9 (7.3)	
Transactional sex[c]					
No	2451 (96.9)	1319 (97.5)	859 (95.8)	118 (95.2)	0.05
Yes	79 (3.1)	34 (2.5)	38 (4.2)	6 (4.8)	
HSV-2 infection					
No	2409 (95.1)	1296 (95.7)	846 (94.4)	117 (94.4)	0.3
Yes	123 (4.9)	58 (4.3)	50 (5.6)	7 (5.7)	

[a]Household socio-economic status measured as natural log of per capita expenditures
[b]In the last three months
[c]With most recent partner

Table 5.2: The association between frequency of alcohol outlet visits in the last 6 months and behavioral and biologic sexual risk outcomes, among 2533 female adolescents

	Sex partners[a,b]		*Unprotected sex acts*[c,d]		*Transactional sex*[e,f]	*HSV-2 infection*[g,h]
Model	**1 vs. 0 partners OR (95% CI)**	**2+ vs. 0 partners OR (95% CI)**	**1 vs. 0 acts OR (95% CI)**	**2+ vs. 0 acts OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**
Unadjusted						
0 visits	1	1	1	1	1	1
≥1 visits	1.40 (1.15, 1.70)	2.17 (1.26, 3.75)	1.62 (1.03, 2.55)	1.69 (1.06, 2.70)	1.75 (1.11, 2.75)	1.32 (0.91, 1.92)
Adjusted						
0 visits	1	1	1	1	1	1
≥1 visits	1.51 (1.21, 1.88)	2.27 (1.29, 3.97)	2.28 (1.52, 3.42)	2.33 (1.53, 3.56)	1.63 (1.03, 2.59)	1.30 (0.88, 1.91)
Unadjusted						
0 visits	1	1	1	1	1	1
1-5 visits	1.40 (1.14, 1.71)	2.25 (1.29, 3.93)	1.72 (1.08, 2.72)	2.23 (1.47, 3.40)	1.72 (1.07, 2.75)	1.32 (0.90, 1.95)
≥6 visits	1.41 (0.91, 2.16)	1.61 (0.47, 5.47)	0.94 (0.29, 3.11)	2.61 (1.23, 5.52)	1.97 (0.81, 4.79)	1.34 (0.60, 3.00)
Adjusted						
0 visits	1	1	1	1	1	1
1-5 visits	1.50 (1.19, 1.88)	2.37 (1.34, 4.20)	1.78 (1.10, 2.85)	2.27 (1.47, 3.50)	1.63 (1.01, 2.62)	1.29 (0.86, 1.93)
≥6 visits	1.59 (0.97, 2.60)	1.56 (0.44, 5.54)	1.04 (0.30, 3.53)	2.88 (1.29, 6.39)	1.66 (0.66, 4.14)	1.37 (0.59, 3.20)

[a]Multinomial logistic regression model modeling the association between alcohol outlet visits and number of sex partners in the last three months, categorized as 0, 1, and 2+ partners

[b]Adjusted estimates are adjusted for age (coded with a quadratic term), current grade enrollment (coded with disjoint indicators for each grade), primary caregiver (coded dichotomously as parent versus non-parent), household size (coded linearly), and household SES (coded as deciles with a linear trend).

[c]Multinomial logistic regression model modeling the association between alcohol outlet visits and number of unprotected sex acts in the last three months, categorized as 0, 1, and 2+ acts

[d]Adjusted estimates are adjusted for age (coded linearly), current grade enrollment (coded with disjoint indicators for each grade), primary caregiver (coded dichotomously as parent versus non-parent), household size (coded linearly), and household SES (coded as deciles with a linear trend).

[e]Logistic regression model modeling the association between alcohol outlet visits and transactional sex with most recent sex partner

[f]Adjusted estimates are adjusted for age (coded with a quadratic term), current grade enrollment (coded linearly), primary caregiver (coded dichotomously as parent versus non-parent), household size (coded linearly), and household SES (coded as deciles with a linear trend).

[g]Logistic regression model modeling the association between alcohol outlet visits and prevalent HSV-2 infection

[h]Adjusted estimates are adjusted for age (coded linearly), current grade enrollment (coded linearly), primary caregiver (coded dichotomously as parent versus non-parent), household size (coded linearly), and household SES (coded with a quadratic term).

OR=odds ratio; CI=confidence interval

Table 5.3. The association between frequency of alcohol outlet visits in the last 6 months and behavioral and biologic sexual risk outcomes, among 672 female adolescents who experienced sexual debut prior to interview

	Sex partners[a,b]		*Unprotected sex acts*[c,d]		*Transactional sex*[e,f]	*HSV-2 infection*[g,h]
Model	**1 vs. 0 partners OR (95% CI)**	**2+ vs. 0 partners OR (95% CI)**	**1 vs. 0 acts OR (95% CI)**	**2+ vs. 0 acts OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**
Unadjusted						
0 visits	1	1	1	1	1	1
≥1 visits	2.19 (1.29, 3.73)	3.20 (1.53, 6.72)	1.51 (0.93, 2.46)	2.19 (1.40, 3.41)	1.46 (0.89, 2.40)	1.28 (0.79, 2.06)
Adjusted						
0 visits	1	1	1	1	1	1
≥1 visits	2.08 (1.21, 3.58)	2.97 (1.39, 6.34)	1.47 (0.90, 2.41)	2.09 (1.33, 3.29)	1.43 (0.86, 2.36)	1.17 (0.71, 1.94)

[a]Multinomial logistic regression model modeling the association between alcohol outlet visits and number of sex partners in the last three months, categorized as 0, 1, and 2+ partners
[b]Adjusted estimates are adjusted for age (coded with a quadratic term), current grade enrollment (coded with disjoint indicators for each grade), primary caregiver (coded dichotomously as parent versus non-parent), household size (coded linearly), and household SES (coded as deciles with a linear trend).
[c]Multinomial logistic regression model modeling the association between alcohol outlet visits and number of unprotected sex acts in the last three months, categorized as 0, 1, and 2+ acts
[d]Adjusted estimates are adjusted for age (coded linearly), current grade enrollment (coded with disjoint indicators for each grade), primary caregiver (coded dichotomously as parent versus non-parent), household size (coded linearly), and household SES (coded as deciles with a linear trend).
[c]Logistic regression model modeling the association between alcohol outlet visits and transactional sex with most recent sex partner
[f]Adjusted estimates are adjusted for age (coded with a quadratic term), current grade enrollment (coded linearly), primary caregiver (coded dichotomously as parent versus non-parent), household size (coded linearly), and household SES (coded as deciles with a linear trend).
[g]Logistic regression model modeling the association between alcohol outlet visits and prevalent HSV-2 infection
[h]Adjusted estimates are adjusted for age (coded linearly), current grade enrollment (coded linearly), primary caregiver (coded dichotomously as parent versus non-parent), household size (coded linearly), and household SES (coded with a quadratic term).

OR=odds ratio; CI=confidence interval

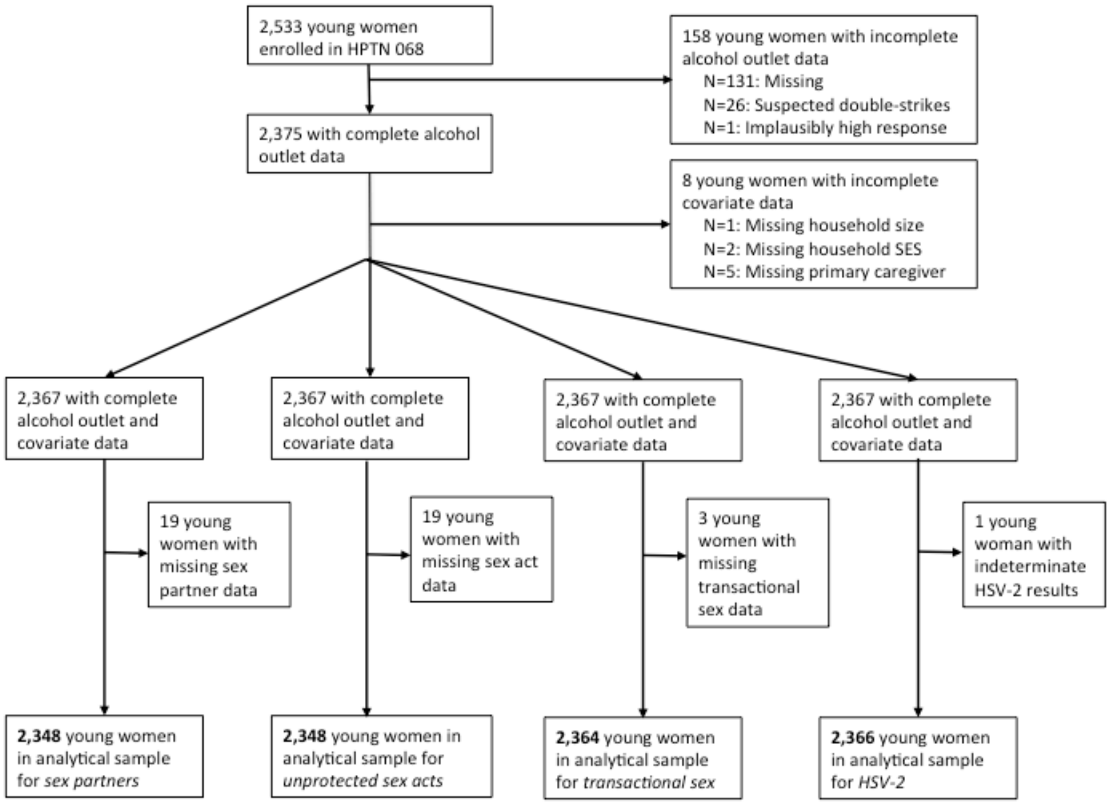
Table 5.4: Interaction between alcohol outlet visits and alcohol consumption on sexual risk outcomes, among 2533 female adolescents

		Sex partners[b,c]		*Unprotected sex acts*[c,d]		*Transactional sex*[c,e]		*HSV-2 infection*[c,f]	
AO visits	**Alcohol consumption**	**aOR (95% CI)**		**aOR (95% CI)**		**aOR (95% CI)**		**aOR (95% CI)**	
no	no	1		1		1		1	
yes	no	1.24 (0.99, 1.55)		1.72 (1.22, 2.50)		1.53 (0.91, 2.57)		1.10 (0.72, 1.70)	
no	yes	1.34 (0.77, 2.32)		1.36 (0.58, 3.22)		2.37 (0.88, 6.43)		0.98 (0.34, 2.87)	
yes	yes	5.20 (3.54, 7.63)		4.39 (2.65, 7.28)		2.87 (1.38, 5.98)		2.44 (1.29, 4.59)	
LRT[a] statistics		**X^2**	**p**	**X^2**	**p**	**X^2**	**p**	**X^2**	**p**
		3.1	0.0008	1.6	0.2	0.14	0.7	1.74	0.2

[a]LRT=Likelihood ratio test for whether the addition of the interaction term provides a significantly better fitting model compared to a model without the interaction term
[b]Ordinal logistic regression model estimating the association between each risk factor and number of sex partners categorized at 0, 1, and 2+ partners.
[c]Each adjusted estimate was adjusted for age, current grade enrollment, primary caregiver, household size, and household SES. All covariates were coded as noted above for each respective outcome.
[d]Ordinal logistic regression model estimating the association between each risk factor and number of unprotected sex acts categorized at 0, 1, and 2+ acts.
[e]Logistic regression model estimating the association between each risk factor and transactional sex with most recent partner.
[f]Logistic regression model estimating the association between each risk factor and prevalent HSV-2 infection.

AO=alcohol outlet; aOR=adjusted odds ratio; CI=confidence interval

Figure 5.1: Flowchart of study sample construction for each of four sexual risk outcomes



CHAPTER VI: Aim 3

Relationship between Community-level Alcohol Outlets and Individual-level HSV-2 Infection among Young Women in South Africa

1 Introduction

Exposure to alcohol outlets (places where alcohol is sold and consumed) may influence sexual risk. At the individual level, alcohol outlets facilitate alcohol consumption which increases sexual risk behaviors, such as unprotected sex,(67, 69, 72) and biological outcomes, including HIV infection.(70, 71, 73) Similarly, characteristics of alcohol outlets (e.g. music, dim lights, unisex bathrooms) (87) and the network of potential sex partners who typically frequent them (74, 89) may provide environments with heightened potential for risky activity. At the community level, the presence of alcohol outlets may influence or reflect community norms around acceptable behaviors, yielding heightened risk for all community members, whether they patronize the establishments or not.

Community-level alcohol outlet accessibility is typically quantified with a community density measure, but can include other measures such as hours/days of sale and price. Alcohol outlet accessibility is associated with increased alcohol consumption,(120-123) and with increased population-level (128, 129, 131) and individual-level (130) sexually transmitted infections. However, the relationship between alcohol outlet accessibility and sexual risk has been underexplored in areas outside of the United States and entirely unexplored in adolescent-specific populations. Adolescents may have different alcohol outlet utilization patterns and alcohol-using behaviors than adults, so generalization from adult population studies to adolescents could be problematic.

Young women in South Africa are at exceptionally high risk for sexually transmitted infections (STIs).(3) New targets for interventions will be crucial to reduce STI transmission. Alcohol outlet accessibility is attractive as a potential intervention target because it is relatively modifiable through strengthening or more strictly enforcing existing government regulation.(119, 126, 127) In this study, we aim to provide the first exploration of the association between alcohol outlet accessibility and sexual risk among adolescents in the developing world, using a population-based sample of young women in rural South Africa.

2 Methods

2.1 Population

To explore the relationship between alcohol outlet accessibility and sexual risk, we performed a cross-sectional study using data originally collected in the HPTN 068 HIV prevention trial. HPTN 068 is a Phase III randomized controlled trial in rural Mpumalanga province, South Africa within the Agincourt Health and Demographic Surveillance Site (HDSS). School-attending young women aged 13-20 were enrolled in the study between March 2011 and December 2012 (n=2533). Biological HSV-2 data were collected during routine baseline procedures completed by each young woman enrolled in the study. Village asset data were collected as part of a community mapping exercise completed by two key informants from each village in the study (n=24) administered in 2011.

2.2 Key measures

Alcohol outlet accessibility was quantified as the *number of alcohol outlets per village*, defined as the combined number of taverns and bottles shops within each village. For analysis, we considered the alcohol outlet measure in two ways: 1. categorized into No, Low (one to four outlets), and High (above four outlets) categories, informed by visual inspection of the data and model fit statistics., and 2. ordinally.

We used an absolute number instead of density per area measure because village area exhibited minimal variation (interquartile range: 1.3 to 3.2 square kilometers). Further, we theorize that, in small areas like the villages in the study region, absolute number is more important than density because all outlets were within reasonable walking distance to all inhabitants. We used a community- as opposed to individual-level exposure because we were interested in capturing contextual as well as individual effects.

From a directed acyclic graph, we identified a minimally sufficient adjustment set of three potential village-level confounding variables. Each variable was derived from the Agincourt HDSS census dataset. *Total population* was defined by the 2011 estimated population size of each village. *Proportion male* was defined as the number of male residents divided by the total population of each village in 2011. *Proportion employed* was defined by proportion of village residents reporting employment in the most recent labor survey in 2008. Final coding decisions were informed by comparing 'Quasi-likelihood under the Independence model Criterion' fit statistics for a variety of categorizations and transformations to best reflect each covariate's association with HSV-2 prevalence.

The sexual risk outcome measure we used was individual-level prevalent *HSV-2 infection.* Testing for HSV-2 was performed at baseline using Kalon™ HSV-2 gG2 ELISA (Kalon Biological, Ltd., Surrey, United Kingdom).(136)

Statistical modeling

To estimate the association between number of alcohol outlets per village and HSV-2 infection, we used generalized estimating equations (GEE) with log links. The GEE models corrected the standard errors of our estimates for the clustered nature of the data. The log link was appropriate for the dichotomous HSV-2 outcome. To preliminarily explore trends in HSV-2 prevalence across different levels of alcohol outlet exposure, we treated the alcohol outlet number exposure categorically. In our final analysis, we treated the alcohol outlet number exposure ordinally, as indicated by the categorical results. All analyses were performed in SAS statistical software, v9.1.2 (Cary, NC).

3 Results

Overall, 2,533 young women were enrolled in HPTN 068 and 2,174 (86%) lived in one of the 24 villages mapped in the community survey. The young women ranged in age from 13 to 21 years with a median age of 15 years (Table 6.1). The average household spent 164 Rands ($15USD) on each resident each month. The sample was relatively evenly distributed with respect to enrollment across four grades: grade 8 (22.4%), grade 9 (26.8%), grade 10 (26.7%) and grade 11 (20.9%). None of these individual-level covariates varied significantly by number of alcohol outlets per village of residence. The overall prevalence of HSV-2 infection was 5% (n=108).

The population of the 24 villages ranged from 171 to 9836. The number of males per village ranged from 37% to 50% and the number of employed residents per village ranged from 12% to 21% of the total population. The number of alcohol outlets per village ranged from zero to seven and was positively correlated with population size (r=0.88), proportion male (r=0.42), and proportion employed (r=0.29).

Young women who lived in villages with more alcohol outlets were more likely to have prevalent HSV-2 infections (Table 6.2). The estimated prevalence of HSV-2 increased from young women living in villages with no alcohol outlets [prevalence% (95% CI): 1.4 (0.2, 12.1)], to those living in villages with low numbers of alcohol outlets [prevalence% (95% CI): 4.5 (3.7, 5.5)], to those living in villages with high numbers of alcohol outlets [prevalence% (95% CI): 6.3 (5.6, 7.1)]. The prevalence estimate for the zero exposure level was imprecise, likely due to the small sample size and single HSV-2 infection in this category.

Treating the alcohol outlet exposure ordinally, for every one unit increase in number of alcohol outlets per village, there was a corresponding 8% increase in prevalent HSV-2 infections [prevalence ratio (PR) (95% CI): 1.08 (1.01, 1.51)]. The point estimate changed minimally after adjustment for village population size, proportion of village population male, and proportion of village population employed [PR (95% CI): 1.10 (0.94, 1.29)]; however, this adjusted estimate was notably less precise. Visually, the ordinal model provided a reasonable representation of the data compared to the categorical model (Figure 6.1).

4 Discussion

Overall, we found that living in villages with higher numbers of alcohol outlets was associated with increased HSV-2 prevalence among young women in rural South Africa. This finding extends the generalizability of the associations found in previous studies to adolescents in the developing world.(128-131) Furthermore, the use of individual-level HSV-2 biological data strengthens inference over prior ecological studies (128, 129, 131) and studies with self-reported STI outcomes.(130)

Due to the cross-sectional nature of this analysis, certainty about the temporality of the relationship between alcohol outlets and sexual risk is not possible. However, we theorize that, although individual businesses can open and close, at the village-level, the absolute number likely remained relatively constant over recent history. Conversely, although we used a prevalence measure, since this was a young cohort, those with HSV-2 infection were likely to have acquired the infection recently. Under these assumptions the alcohol outlet exposure likely predates the HSV-2 outcome, though the findings should be interpreted with caution.

The observational nature of the data limited our ability to assess a causal relationship between alcohol outlets and sexual risk. Self-selection of where one lives could theoretically be influenced by alcohol outlet accessibility. However, adolescents are presumably less likely to make household decisions as they share households with parents or other adult caregivers. Also, other village characteristics may vary with alcohol outlet accessibility and influence sexual risk. Although the association between alcohol outlets and HSV-2 remained after adjusting for village population size, proportion employed, and proportion male, the possibility remains that confounding by other unmeasured village-level characteristics influenced the results.

This study found that young women who live in communities with more accessible alcohol environments have higher HSV-2 prevalence than those who do not. These results suggest that high-risk young women may be found in higher proportions within alcohol-accessible communities and that sexual health interventions may be appropriately targeted at communities based on their alcohol environment characteristics. Future studies examining potential mediation of the observed associations by individual-level alcohol outlet patronage and alcohol consumption could help disentangle individual from contextual effects. A better understanding of how alcohol outlets are associated with sexual risk could inform recommendations about whether reducing access to outlets could reduce sexual risk, or whether the number of outlets is more likely a reflection of community norms. Future studies incorporating more information about the individual alcohol outlets (more finely specified geographic locations, general characteristics of the outlets) could also help identify what types of alcohol outlets might be potential targets for sexual risk reduction.

5 Tables and Figures

Table 6.1: Demographic characteristics of cross-sectional sample of 2,174 South African young women across 24 villages, by number of alcohol outlets in village of residence

	Total population N=2174 Villages=24	**No AO N=67 Villages=4**	**Low AO N=1333 Villages=16**	**High AO N=774 Villages=4**	
Individual-level variables	*Median (IQR)*	*Median (IQR)*	*Median (IQR)*	*Median (IQR)*	*ANOVA p* [c]
Age	15 (14-17)	16 (15-17)	15 (14-17)	16 (14-17)	0.1
Household SES[a]	164 (110-236)	174 (104-239)	167 (111-237)	161 (108-233)	0.9
	N (%)	*N (%)*	*N (%)*	*N (%)*	χ^2 *p* [c]
Grade					0.2
8	557 (25.6)	15 (22.4)	350 (26.3)	192 (24.8)	
9	583 (26.8)	18 (26.9)	362 (27.2)	203 (26.2)	
10	580 (26.7)	12 (17.9)	358 (26.9)	210 (27.1)	
11	454 (20.9)	22 (32.8)	263 (19.7)	169 (21.8)	
Village-level variables	*# Villages*	*# Villages*	*# Villages*	*# Villages*	*r (p)* [d]
Total population[b]					0.88 (<0.001)
Quartile 1 (<2258)	6	4	2	0	
Quartile 2 (2258-3543)	6	0	6	0	
Quartile 3 (3544-5352)	6	0	6	0	
Quartile 4 (>5353)	6	0	2	4	
Proportion male[b]					0.42 (0.04)
At or above median (≥48.3%)	10	0	8	2	
Below median (<48.3%)	14	4	8	2	
Proportion employed[b]					0.29 (0.2)
Quartile 1 (<18.2%)	9	4	4	1	
Quartile 2 (18.2-18.4%)	4	0	1	3	
Quartile 3 (18.4-19.0%)	5	0	5	0	
Quartile 4 (≥19.0%)	6	0	6	0	

[a]Household SES measured as per capita household expenditures, in South African Rands

[b]Quartiles for population size were computed based on village-level data. Median for proportion male and quartiles for proportion employed were computed based on individual-level data, as used in the final statistical models. Total population size was coded linearly in the final statistical models.

[c]p-values for continuous covariates are for ANOVA tests, p-values for categorical covariates are for chi-square tests. Note that the statistical significance of these relationships is likely inflated as these tests are not adjusted for the clustered nature of the data. For this test, the household SES variable was log-transformed to meet assumption of normality.

[d]Pearson correlation coefficient and p-value for the village-level correlation between each covariate (treated linearly) and number of alcohol outlets (treated ordinally)

AO=alcohol outlets; IQR=interquartile range; SES=socio-economic status; ANOVA=analysis of variance between groups test; r=Pearson correlation coefficient

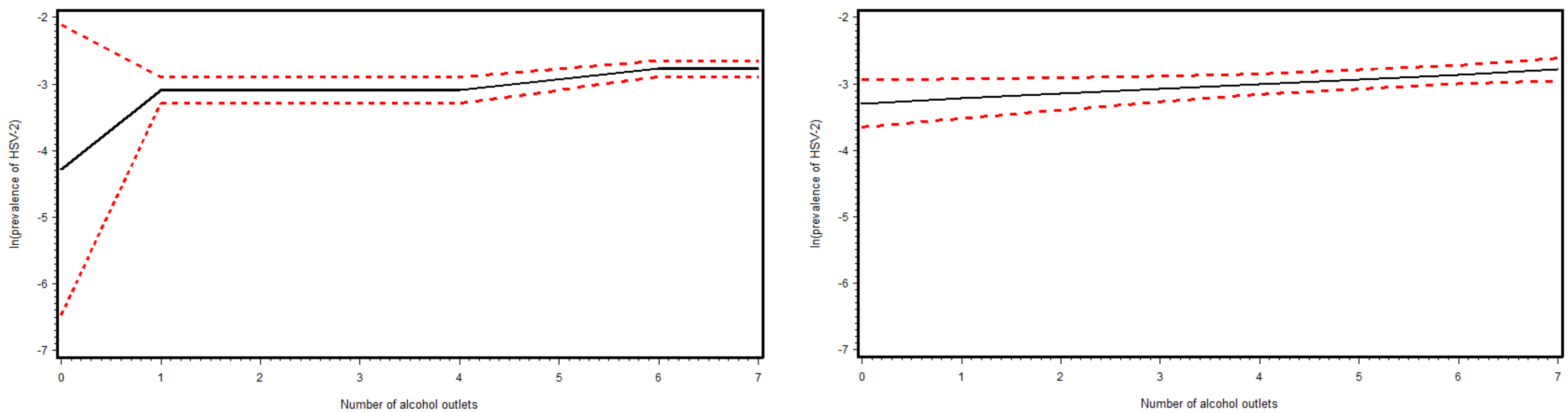
Table 6.2: Association between number of alcohol outlets in home village and prevalent HSV-2 infection among 2,174 South African young women

# AO per village	N	# Villages	# HSV-2+	HSV-2 prevalence (95% CI)[a]
0	67	4	1	1.4 (0.2, 12.1)
1 to 4	1333	16	61	4.5 (3.7, 5.5)
5+	774	4	46	6.3 (5.6, 7.1)
Model description				**PR (95% CI)**
Unadjusted GEE model, ordinal AO exposure				1.08 (1.01, 1.15)
Adjusted GEE model, ordinal AO exposure[b]				1.10 (0.94, 1.29)

[a]Results from an unadjusted GEE model with log link, with categorical AO exposure, per 100 young women
[b]Adjusted for village population size (coded linearly), proportion of village population male (coded dichotomously with cutpoint at median – 48.3%), proportion of village population employed (coded categorically in quartiles)

AO=alcohol outlets; PR=prevalence ratio; CI=confidence interval; GEE=generalized estimating equation

Figure 6.1: Graphical representations of the association between number of alcohol outlets in home village and prevalent HSV-2 infection among 2,174 South African young women. Solid black line indicates HSV-2 prevalence estimates across varying number of alcohol outlets per village. Dashed red line indicates 95% confidence intervals around the prevalence estimates



(a) Number of alcohol outlets treated categorically at 0, 1, and more than one outlet per village in an unadjusted generalized estimating equation (GEE) model with a log link.

(b) Number of alcohol outlets treated ordinally in an unadjusted GEE model with a log link.

.

CHAPTER VII: Discussion

Young women in South Africa are at high risk for negative sexual health outcomes like sexually transmitted infections (3) and teen pregnancy.(10) Novel risk factors and potential intervention targets need to be identified to reduce the future public health burden of these outcomes. Although much previous research has been performed identifying individual behavioral and demographic risk factors for sexual risk outcomes, structural and place-based factors have been underexplored.

This dissertation's key theme revolves around the intersection of place and health, with specific aims driven by the overarching question: 'Can where one spends time influence sexual risk?' We sought to address this question by focusing on two specific place-based exposures important to the study population of adolescent females in South Africa: school (Aim 1) and alcohol outlets (Aims 2-3). We further divided the alcohol outlet exposure into two separate, but related place-based research questions: 1. 'Does visiting alcohol outlets increase sexual risk?' (Aim 2), and 2. 'Does living in communities with more alcohol outlets increase sexual risk?' (Aim 3). We hypothesized that places with structured and supervised environments (school) would be associated with decreased sexual risk outcomes. We hypothesized that places with unstructured and unsupervised environments (alcohol outlets) would be associated with increased sexual risk outcomes. Alcohol outlets were further hypothesized to be associated with increased sexual risk outcomes because outlets facilitate alcohol consumption, the networks of potential partners within outlets may be riskier, and characteristics of the outlets may encourage risky activity.

1 Summary of findings

Overall, we found support for our hypothesis that place and routine activities were important predictors of sexual risk. The direction of the association, as hypothesized, was dependent on the type of activity. School, classified as providing a structured and supervised environment, was found to be a protective place, while alcohol outlets, classified as providing an unstructured and unsupervised environment, were found to be risky, at both the individual- and community-level.

In Aim 1, our primary analysis explored the effect of school enrollment on incident teen pregnancy, using data from a large, longitudinal cohort with over 11 years of follow-up. Overall, the teen pregnancy rate in this cohort was high (4.4 pregnancies per 100 person-years) and was much lower among young women enrolled in school (4.1 pregnancies per 100 person-years) compared to young women not enrolled in school (11.7 pregnancies per 100 person-years). After adjusting for age and other important demographic characteristics hypothesized to be associated with both school enrollment and teen pregnancy, we found that the hazard of pregnancy was significantly lower during times of school enrollment compared to non-enrollment [HR (95% CI): 0.57 (0.50, 0.65)]. This finding was robust to several alternate exposure specifications. To further explore the potential protective effects of physically being in school, our secondary analysis explored whether dates of conception were less likely to occur during school term than school holidays, among school enrollees. Here, we also observed a protective association [IRR (95% CI): 0.90 (0.78, 1.04)], though small in magnitude. Both our primary and secondary results support the overall hypothesis that the structured and supervised environment of school is protective against teen pregnancy.

In Aim 2, we explored the association between visits to alcohol outlets and four sexual risk outcomes (number of unprotected sex acts, number of sex partners, transactional sex, and HSV-2 infection) in a cross-sectional analysis of data from a cohort of young women in South Africa. We found that frequenting alcohol outlets was associated with having more unprotected sex acts [aOR, one versus zero acts (95% CI): 2.28 (1.52, 3.42)], more sex partners [aOR, one versus zero partners (95% CI): 1.51 (1.21, 1.88)], higher levels of transactional sex [aOR (95% CI): 1.63 (1.03, 2.59)], and prevalent HSV-2 infection [aOR (95% CI): 1.30 (0.88, 1.91)]. We also found that dual exposure to alcohol outlets and alcohol consumption was especially important with respect to the outcomes. Compared to those with neither exposure, young women who visited alcohol outlets and consumed alcohol reported a higher number of sex partners [aOR (95%CI): 5.20 (3.54, 7.63)], reported a higher number of unprotected sex acts [aOR (95%CI): 4.39 (2.65, 7.28)], reported more transactional sex [aOR (95%CI): 2.87 (1.38, 5.98)], and were more likely to be HSV-2 positive [aOR (95%CI): 2.44 (1.29, 4.59)]. Although the cross-sectional nature of the analysis yielded results that should be interpreted cautiously with respect to causality, overall, the results support the hypothesis that alcohol outlets provide risky environments that may lead to negative sexual risk outcomes.

In Aim 3, we explored the cross-sectional association between the number of alcohol outlets in a community and prevalent HSV-2 infection in the same cohort of young South African women. Overall, we found that young women who lived in communities with more alcohol outlets were more likely to have a prevalent HSV-2 infection [PR (95% CI): 1.08 (1.01, 1.51)]. Several community-level characteristics were positively correlated with number of alcohol outlets: population size, proportion of population male, and proportion of population employed. After adjusting for these potential confounders, the point estimate remained largely unchanged, though the estimate was notably less precise [PR (95% CI): 1.10 (0.94, 1.29)]. Again, due to the cross-sectional nature of these data, certainty about the directionality of the association is not possible. However, the results support the hypothesis that living in communities with more accessible alcohol environments can lead to sexual risk.

2 Contributions

This dissertation contributes importantly to the epidemiological literature on adolescent sexual risk for several reasons. All three aims were designed to explore the potential influence of place and routine activities on sexual risk, an understudied set of structural exposures. Although the hypothesis that routine activities would influence sexual risk is well-supported theoretically, these potential exposures have been largely ignored in the epidemiological literature to date.

Another important aspect of these place-based exposures is their inherent link to practical interventions. Unlike some demographic or individual-level factors that are inherently non-modifiable or difficult to modify (e.g. age and socio-economic status), the associations observed with school enrollment and alcohol outlets could conceivably be extended into the development of analogous potentially preventive interventions.

In Aim 1, we found that school enrollment was protective against teen pregnancy. Interventions designed to keep young women in school or providing recent dropouts with sexual health information could, therefore, have potential to reduce teen pregnancy incidence. For example, programs offering incentives conditional on school attendance could plausibly prevent pregnancies among those who remain enrolled in school due to the incentive.(58, 59, 148)

In Aims 2 and 3, we found that exposure to alcohol outlets at both the individual- and community-level was associated with negative sexual risk outcomes. Interventions that target safe sex messages and condom distribution in alcohol outlets or in communities with high numbers of alcohol outlets could, therefore, effectively reach high-risk young women.(93, 153) These findings also suggest that interventions to decrease accessibility of alcohol outlets for adolescents may be effective at reducing sexual risk, though future studies will first need to confirm the directionality of the cross-sectional associations we observed.

We conducted each of our study aims within a study population of young women living in rural South Africa. Our research questions were particularly pertinent to this study population because, as mentioned previously, they are at extraordinarily high risk for negative sexual risk outcomes such as HIV and teen pregnancy. However, the associations between school enrollment and teen pregnancy and between alcohol outlet exposure (at either the individual- or community-level) and sexual risk have not been explored previously among South African young women. The single published study examining the effect of school enrollment on teen pregnancy was performed in the United States using data from over 20 years ago.(60) Our study extends the findings of a protective effect of school enrollment to South Africa and to the first decade of the 21st century. Of the four studies examining the association between alcohol outlet visits and sexual risk, all were conducted within adult study populations; the association had not previously been studied in adolescent-specific or female-specific populations.(88, 105-107) Similarly, the relationship between alcohol outlet accessibility and sexual risk has been underexplored in areas outside of the United States and entirely unexplored in adolescent-specific populations.(128-131) Results from Aims 2 and 3 provided the first evidence of an association between individual- and community-level alcohol outlet exposure and sexual risk among adolescent females in South Africa.

However, given the specific study population used in this dissertation, the results may not be generalizable outside of young women in rural South Africa. The associations we observed in each aim were generally consistent in both direction and magnitude with the associations reported in prior studies conducted in American populations and within adult, as opposed to adolescent, populations. (60, 88, 105-107, 128-131) This consistency suggests that the associations we observed may not be specific to our study population. However, future studies are needed to confirm the associations among young men and with adolescents in other parts of the world.

3 Strengths and limitations

3.1 Data sources

The datasets employed in this study were appropriate to explore the associations between place/routine activities and sexual risk. All three aims of this dissertation made use of existing, rich datasets that had well-measured indicators of routine activity exposures and sexual risk outcomes. In Aim 1, to examine the relationship between school enrollment and teen pregnancy, we used information from the Agincourt HDSS census, which regularly updated information on school enrollment status of every resident and collected data on every pregnancy that occurred within the study area during follow-up of our study. In Aims 2-3, to examine the association between alcohol outlets and sexual risk, we used information from baseline data collection activities for the HPTN 068 HIV Prevention Trial. This rich dataset included measures of self-reported alcohol outlet visit frequency and self-reported sexual risk behaviors (Aim 2). Biological data on prevalent HSV-2 were collected and tested under high quality standards

(Aims 2 and 3). Community-level alcohol outlets were enumerated by key informants in a community mapping exercise (Aim 3). Use of these existing datasets with complete exposure and outcome data allowed us to efficiently answer our research questions with relatively little investment of time or money.

However, given that this dissertation made use of existing data, we relied on information that was previously collected for purposes other than our study aims. Potential for bias and measurement error were, therefore, of significant concern. Specifically, in Aim 1, the school enrollment data were collected with long intervals (three to four years) between updates, leading to the possibility for exposure misclassification. To address this potential for measurement error, we performed several sensitivity analyses to explore the robustness of our findings to different exposure specifications. From these results, we concluded that exposure misclassification alone was not a likely explanation for our results.

In Aim 2, the alcohol outlet exposure and behavioral outcomes were self-reported by the young women. Given the sensitive nature of these questions, the respondents potentially felt influenced to give socially desirable responses. For this reason, a computer-based interviewing system (ACASI) was used instead of a traditional interviewer; however, the possibility remains that social desirability bias influenced our results.

Finally, in Aim 3, key informants from each community provided information on a variety of community assets, including number of bottle shops and number of taverns/shebeens. Misinformation from key informants could have led to inaccurate reporting in the number of alcohol outlets. As the alcohol outlet exposure was analyzed at the community-level, potential inaccuracies in key informant reporting for a single village would influence not just a single observation, but the exposure level for each young woman living in the community. However, as the key informants were residents of the communities about which they were reporting and the communities themselves were relatively small with few outlets, large inaccuracies in the number of alcohol outlets were not expected.

3.2 Study design

The longitudinal design of Aim 1 was advantageous for several reasons. First, it allowed us to meaningfully incorporate time into the analysis. This added information allowed us to assess *when* the pregnancy occurred as opposed to simply knowing whether it happened or not. With an outcome like teen pregnancy, the timing of the pregnancy in adolescence may have implications for how disruptive it may be to the life course of the adolescent and how prepared she may be for motherhood. Second, the longitudinal design allowed us to confidently assess the temporality between exposure and outcome. Temporality assessment is an important element of causal inference for any exposure and outcome, but particularly so in assessing whether school enrollment influences pregnancy given that the reverse relationship is also likely to hold (i.e. teen pregnancy while enrolled in school likely influences school dropout decisions).(55-57)

Aims 2 and 3 were both cross-sectional in design. Though cross-sectional analyses for these aims were appropriate given the structure of the data we employed, the inferences one can make from analyzing cross-sectional data are limited. Two general weaknesses of cross-sectional studies are the uncertainty regarding the direction of association between exposure and outcome and the potential bias induced using prevalent instead of incident outcomes. These concerns were mitigated for Aim 2 for the following reasons:

1) This analysis aimed to establish whether or not there was an association between visits to alcohol outlets and sexual risk. The investment in future longitudinal studies to establish

the temporal relationship between alcohol outlets and sexual risk may now be worthwhile, given that a significant cross-sectional association was found.

2) We hypothesized that visits to alcohol outlets would influence sexual behavior. Although, given the cross-sectional analysis, we cannot assess the directionality of the relationship between exposure and outcome, the knowledge that groups of high-risk young women gather in alcohol outlets is useful regardless of the directionality of the association. Preventive interventions with efficient place-based recruitment of potential participants or place-based delivery of information/materials could be performed on this evidence alone.

3) At baseline, this was a cohort of relatively young women (mean age: 15.6), most of whom had not yet experienced sexual debut (proportion with sexual debut: 26.6%) or only very recently sexually debuted (median age at debut: 16). Therefore, we assume that the durations of the prevalent sexual behavior outcomes and prevalent HSV-2 infection were fairly short and that they reasonably approximated incident outcomes, minimizing the potential for prevalence-incidence bias. Also, the nature of the outcomes we explored (sexual behavior and HSV-2 infection) does not naturally lend itself to some of the major problems with using prevalent versus incident outcomes. As mortality is not immediately associated with the outcomes we explored, it is unlikely that there were young women who did not survive the short duration between the time that they first experience a risky sexual behavior or became infected with HSV-2 and the baseline data collection.

The concerns about cross-sectional designs were mitigated for Aim 3 for the following reasons:

1) As in Aim 2, establishing a causal relationship between the community alcohol outlet accessibility and HSV-2 infection was not an objective of the study. The aim of this analysis was to assess whether the community exposure was associated with sexual risk outcome. The investment in future longitudinal studies to establish the temporal relationship between community alcohol accessibility and sexual risk may now be worthwhile, given that a significant cross-sectional association was found.
2) We assumed that the number alcohol outlets per community were relatively constant over recent history as the opening or closing of a business is a rather long process and the influence of changes in a single outlet's status was not likely to influence the overall measure in a village strongly. However, the sexual transmission of HSV-2 was assumed to be relatively recent in this cohort of young women, given their young ages. Under these assumptions the alcohol outlet exposure likely predated the HSV-2 outcome.
3) As with Aim 2, the nature of the HSV-2 outcome we explored here does not naturally lend itself to some of the major problems with using prevalent versus incident outcomes. As mortality is not directly associated with HSV-2, it is unlikely that there were young women who did not survive the short duration between the time they became infected with HSV-2 and baseline data collection.

4 Conclusion

In response to the overarching question guiding this dissertation, 'Can where one spends time influence sexual risk?,' our findings all indicate 'Yes.' We found that school enrollment was protective against teen pregnancy, that visits to alcohol outlets were associated with multiple sexual risk outcomes, and that living in communities with more alcohol outlets was associated with HSV-2 infection. The findings from these studies further the understanding of the etiology of sexual risk outcomes like sexually transmitted infections and teen pregnancy and provide insight into how sexual behaviors are learned and transmitted. In conjunction with future research, our findings could inform the development of preventive interventions to target place-based exposures to reduce sexual risk outcomes. Specific interventions that may be effective at reducing sexual risk based on our results include interventions to retain young women in school and interventions to reduce adolescent access to alcohol outlets. The development of better prevention interventions for South African young women is critical to prevent the continued spread of sexually transmitted infections and high burden of teenage pregnancy in this vulnerable population.

REFERENCES

1 UNAIDS. UNAIDS report on the global AIDS epidemic; 2012.

2 UNAIDS. South Africa country overview. Geneva: UNAIDS; 2011.

3 Pettifor AE, Rees HV, Kleinschmidt I, et al. Young people's sexual health in South Africa: HIV prevalence and sexual behaviors from a nationally representative household survey. AIDS. 2005 Sep 23;**19**(14):1525-34.

4 UNAIDS. Global AIDS Response Progress Report, 2012: Republic of South Africa; 2012.

5 Ruedinger E, Cox JE. Adolescent childbearing: consequences and interventions. Curr Opin Pediatr. 2012 Aug;**24**(4):446-52.

6 Black AY, Fleming NA, Rome ES. Pregnancy in adolescents. Adolesc Med State Art Rev. 2012 Apr;**23**(1):123-38, xi.

7 Pinzon JL, Jones VF. Care of adolescent parents and their children. Pediatrics. 2012 Dec;**130**(6):e1743-56.

8 Zwang J, Garenne M. Social context of premarital fertility in rural South-Africa. Afr J Reprod Health. 2008 Aug;**12**(2):98-110.

9 WHO. Fact Sheet: Adolescent Pregnancy; 2012.

10 United Nations Statistics Division. Live births by age of mother and sex of child, general and age-specific fertility rates: latest available year, 2000–2009; 2000-2009.

11 Republic of South Africa. South Africa Demographic and Health Survey; 2003.

12 Serwadda D, Wawer MJ, Musgrave SD, Sewankambo NK, Kaplan JE, Gray RH. HIV risk factors in three geographic strata of rural Rakai District, Uganda. AIDS. 1992 Sep;**6**(9):983-9.

13 UNAIDS. Women out loud: How women living with HIV will help the world end AIDS; 2012.

14 Wojcicki JM. Socioeconomic status as a risk factor for HIV infection in women in East, Central and Southern Africa: a systematic review. J Biosoc Sci. 2005 Jan;**37**(1):1-36.

15 Pettifor AE, van der Straten A, Dunbar MS, Shiboski SC, Padian NS. Early age of first sex: a risk factor for HIV infection among women in Zimbabwe. AIDS. 2004 Jul 2;**18**(10):1435-42.

16 Parazzini F, Cavalieri D'oro L, Naldi L, et al. Number of sexual partners, condom use and risk of human immunodeficiency virus infection. Int J Epidemiol. 1995 Dec;**24**(6):1197-203.

17 Fleming DT, Wasserheit JN. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. Sex Transm Infect. 1999 Feb;**75**(1):3-17.

18 Vundule C, Maforah F, Jewkes R, Jordaan E. Risk factors for teenage pregnancy among sexually active black adolescents in Cape Town. A case control study. S Afr Med J. 2001 Jan;**91**(1):73-80.

19 Lavin C, Cox JE. Teen pregnancy prevention: current perspectives. Curr Opin Pediatr. 2012 Aug;**24**(4):462-9.

20 Pettifor AE, Levandowski BA, MacPhail C, Padian NS, Cohen MS, Rees HV. Keep them in school: the importance of education as a protective factor against HIV infection among young South African women. Int J Epidemiol. 2008 Dec;**37**(6):1266-73.

21 Lopman B, Lewis J, Nyamukapa C, Mushati P, Chandiwana S, Gregson S. HIV incidence and poverty in Manicaland, Zimbabwe: is HIV becoming a disease of the poor? AIDS. 2007 Nov;**21 Suppl 7**:S57-66.

22 McHunu G, Peltzer K, Tutshana B, Seutlwadi L. Adolescent pregnancy and associated factors in South African youth. Afr Health Sci. 2012 Dec;**12**(4):426-34.

23 Blum R, Mmari K. Risk and protective factors affecting adolescent reproductive health in developing countries: WHO; 2004.

24 Hirschi T, Stark R. Hellfire and delinquency. Social Problems. 1969:202-13.

25 Hirschi T. Causes of delinquency: Transaction publishers; 2002.

26 Cohen LE, Felson M. Social change and crime rate trends: A routine activity approach. American sociological review. 1979:588-608.

27 Felson M, Gottfredson M. Social indicators of adolescent activities near peers and parents. Journal of Marriage and the Family. 1984:709-14.

28 Hawdon JE. Daily Routines and Crime Using Routine Activities as Measures of Hirschi's Involvement. Youth & Society. 1999;**30**(4):395-415.

29 Osgood DW, Wilson JK, O'Malley PM, Bachman JG, Johnston LD. Routine activities and individual deviant behavior. American Sociological Review. 1996:635-55.

30 Barnes GM, Hoffman JH, Welte JW, Farrell MP, Dintcheff BA. Adolescents' time use: Effects on substance use, delinquency and sexual activity. Journal of Youth and Adolescence. 2007;**36**(5):697-710.

31 Bandura A. Social foundations of thought and action: A social cognitive theory. 1986.

32 Bandura A. Social cognitive theory and exercise of control over HIV infection. Preventing AIDS: Theories and methods of behavioral interventions. 1994:25-59.

33 Kawachi I, Berkman L, (Eds.). Neighborhoods and Health. New York, NY: Oxford University Press; 2003.

34 Kawachi I, Subramanian SV, Almeida-Filho N. A glossary for health inequalities. J Epidemiol Community Health. 2002 Sep;**56**(9):647-52.

35 UNESCO Institute for Statistics. Enrolment ratios by ISCED level, South Africa; 2009.

36 Strassburg S, Meny-Gibert S, Russell B. Left unfinished: Temporary absence and drop-out from South African schools. Johannesburg, South Africa: Social Surveys Africa; 2010.

37 South African Ministry of Basic Education. School calendars for public schools; 2010.

38 Wittenberg M. The school day in South Africa. Southern Africa Labour and Development Research Unit, University of Cape Town: SALDRU/CSSR Working Papers; 2005.

39 Modisaotsile BM. The failing standard of basic education in South Africa. Policy Brief. 2012;**72**.

40 Mbuyisi M. South Africa: A fateful decision; 2009.

41 Doyle AM, Mavedzenge SN, Plummer ML, Ross DA. The sexual behaviour of adolescents in sub-Saharan Africa: patterns and trends from national surveys. Trop Med Int Health. 2012 Jul;**17**(7):796-807.

42 Hargreaves JR, Slaymaker E, Fearon E, Howe LD. Changes over time in sexual behaviour among young people with different levels of educational attainment in Tanzania. J Int AIDS Soc. 2012;**15 Suppl 1**:1-7.

43 Zuilkowski SS, Jukes MC. The impact of education on sexual behavior in sub-Saharan Africa: a review of the evidence. AIDS Care. 2012;**24**(5):562-76.

44 Barnighausen T, Hosegood V, Timaeus IM, Newell ML. The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. AIDS. 2007 Nov;**21 Suppl 7**:S29-38.

45 Bradley H, Bedada A, Brahmbhatt H, Kidanu A, Gillespie D, Tsui A. Educational attainment and HIV status among Ethiopian voluntary counseling and testing clients. AIDS Behav. 2007 Sep;**11**(5):736-42.

46 Hargreaves JR, Glynn JR. Educational attainment and HIV-1 infection in developing countries: a systematic review. Trop Med Int Health. 2002 Jun;**7**(6):489-98.

47 Hargreaves JR, Bonell CP, Boler T, et al. Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-Saharan Africa. AIDS. 2008 Jan 30;**22**(3):403-14.

48 de Walque D, Nakiyingi-Miiro JS, Busingye J, Whitworth JA. Changing association between schooling levels and HIV-1 infection over 11 years in a rural population cohort in south-west Uganda. Trop Med Int Health. 2005 Oct;**10**(10):993-1001.

49 Michelo C, Sandoy IF, Fylkesnes K. Marked HIV prevalence declines in higher educated young people: evidence from population-based surveys (1995-2003) in Zambia. AIDS. 2006 Apr 24;**20**(7):1031-8.

50 Johnson LF, Dorrington RE, Bradshaw D, du Plessis H, Makubalo L. The effect of educational attainment and other factors on HIV risk in South African women: results from antenatal surveillance, 2000-2005. AIDS. 2009 Jul 31;**23**(12):1583-8.

51 Jukes M, Simmons S, Bundy D. Education and vulnerability: the role of schools in protecting young women and girls from HIV in southern Africa. AIDS. 2008 Dec;**22 Suppl 4**:S41-56.

52 Hargreaves JR, Morison LA, Kim JC, et al. The association between school attendance, HIV infection and sexual behaviour among young people in rural South Africa. J Epidemiol Community Health. 2008 Feb;**62**(2):113-9.

53 Stroeken K, Remes P, De Koker P, Michielsen K, Van Vossole A, Temmerman M. HIV among out-of-school youth in Eastern and Southern Africa: a review. AIDS Care. 2012;**24**(2):186-94.

54 Gavin L, Galavotti C, Dube H, et al. Factors associated with HIV infection in adolescent females in Zimbabwe. J Adolesc Health. 2006 Oct;**39**(4):596 e11-8.

55 Grant MJ, Hallman KK. Pregnancy-related school dropout and prior school performance in KwaZulu-Natal, South Africa. Stud Fam Plann. 2008 Dec;**39**(4):369-82.

56 Clark S, Mathur R. Dating, sex, and schooling in urban Kenya. Stud Fam Plann. 2012 Sep;**43**(3):161-74.

57 Marteleto L, Lam D, Ranchhod V. Sexual behavior, pregnancy, and schooling among young people in urban South Africa. Studies in Family Planning. 2008;**39**(4):351-68.

58 Duflo E, Dupas P, Kremer M, Sinei S. Education and HIV/AIDS prevention: evidence from a randomized evaluation in Western Kenya. World Bank Policy Research Working Paper. 2006(4024).

59 Baird S, Chirwa E, McIntosh C, Ozler B. The short-term impacts of a schooling conditional cash transfer program on the sexual behavior of young women. Health Econ. 2010 Sep;**19 Suppl**:55-68.

60 Manlove J. The influence of high school dropout and school disengagement on the risk of school-age pregnancy. J Res Adolesc. 1998;**8**(2):187-220.

61 Warner L, Newman DR, Austin HD, et al. Condom effectiveness for reducing transmission of gonorrhea and chlamydia: the importance of assessing partner infection status. Am J Epidemiol. 2004 Feb 1;**159**(3):242-51.

62 Niccolai LM, Rowhani-Rahbar A, Jenkins H, Green S, Dunne DW. Condom effectiveness for prevention of Chlamydia trachomatis infection. Sex Transm Infect. 2005 Aug;**81**(4):323-5.

63 Shlay JC, McClung MW, Patnaik JL, Douglas JM, Jr. Comparison of sexually transmitted disease prevalence by reported level of condom use among patients attending an urban sexually transmitted disease clinic. Sex Transm Dis. 2004 Mar;**31**(3):154-60.

64 Celentano DD. Editorial response: it's all in the measurement: consistent condom use is effective in preventing sexually transmitted infections. Sex Transm Dis. 2004 Mar;**31**(3):161-2.

65 Minnis AM, van der Straten A, Gerdts C, Padian NS. A comparison of four condom-use measures in predicting pregnancy, cervical STI and HIV incidence among Zimbabwean women. Sex Transm Infect. 2010 Jun;**86**(3):231-5.

66 Pettifor AE, Levandowski BA, Macphail C, et al. A tale of two countries: rethinking sexual risk for HIV among young people in South Africa and the United States. J Adolesc Health. 2011 Sep;**49**(3):237-43 e1.

67 Sales JM, Brown JL, Vissman AT, DiClemente RJ. The association between alcohol use and sexual risk behaviors among African American women across three developmental periods: a review. Curr Drug Abuse Rev. 2012 Jun;**5**(2):117-28.

68 Li Q, Li X, Stanton B. Alcohol use and sexual risk behaviors and outcomes in China: a literature review. AIDS Behav. 2010 Dec;**14**(6):1227-36.

69 Rehm J, Shield KD, Joharchi N, Shuper PA. Alcohol consumption and the intention to engage in unprotected sex: systematic review and meta-analysis of experimental studies. Addiction. 2012 Jan;**107**(1):51-9.

70 Fisher JC, Bang H, Kapiga SH. The association between HIV infection and alcohol use: a systematic review and meta-analysis of African studies. Sex Transm Dis. 2007 Nov;**34**(11):856-63.

71 Cook RL, Clark DB. Is there an association between alcohol consumption and sexually transmitted diseases? A systematic review. Sexually Transmitted Diseases. 2005;**32**(3):156-64.

72 Kalichman SC, Simbayi LC, Kaufman M, Cain D, Jooste S. Alcohol use and sexual risks for HIV/AIDS in sub-Saharan Africa: systematic review of empirical findings. Prev Sci. 2007 Jun;**8**(2):141-51.

73 Pithey A, Parry C. Descriptive systematic review of sub-Saharan African studies on the association between alcohol use and HIV infection. SAHARA J (Journal of Social Aspects of HIV/AIDS Research Alliance). 2009;**6**(4).

74 Mataure P, McFarland W, Fritz K, et al. Alcohol use and high-risk sexual behavior among adolescents and young adults in Harare, Zimbabwe. AIDS and Behavior. 2002;**6**(3):211-9.

75 Chersich MF, Rees HV. Vulnerability of women in southern Africa to infection with HIV: biological determinants and priority health sector interventions. AIDS. 2008 Dec;**22 Suppl 4**:S27-40.

76 Morojele NK, Kachieng'a MA, Nkoko MA, et al. Perceived effects of alcohol use on sexual encounters among adults in South Africa. African Journal of Drug and Alcohol Studies. 2004;**3**:1-20.

77 Kalichman SC, Grebler T. Reducing numbers of sex partners: do we really need special interventions for sexual concurrency? AIDS Behav. 2010 Oct;**14**(5):987-90.

78 Pihl RO, Peterson JB, Lau MA. A biosocial model of the alcohol-aggression relationship. Journal of Studies on Alcohol and Drugs. 1993(11):128.

79 Steele CM, Josephs RA. Alcohol myopia: Its prized and dangerous effects. American Psychologist. 1990;**45**(8):921.

80 Graham K, West P, Wells S. Evaluating theories of alcohol-related aggression using observations of young adults in bars. Addiction. 2000;**95**(6):847-63.

81 Zawacki T. Effects of alcohol on women's risky sexual decision making during social interactions in the laboratory. Psychology of Women Quarterly. 2011;**35**(1):107-18.

82 Davis KC. The influence of alcohol expectancies and intoxication on men's aggressive unprotected sexual intentions. Experimental and clinical psychopharmacology. 2010;**18**(5):418.

83 Ariely D, Loewenstein G. The heat of the moment: The effect of sexual arousal on sexual decision making. Journal of Behavioral Decision Making. 2006;**19**(2):87-98.

84 Parry CD. South Africa: alcohol today. Addiction. 2005;**100**(4):426-9.

85 Parry CD, Myers B, Morojele NK, et al. Trends in adolescent alcohol and other drug use: findings from three sentinel sites in South Africa (1997–2001). Journal of Adolescence. 2004;**27**(4):429-40.

86 Rehm J, rgen u, Rehn N, et al. The global distribution of average volume of alcohol consumption and patterns of drinking. European addiction research. 2003;**9**(4):147-56.

87 Morojele NK, Kachieng'a MA, Mokoko E, et al. Alcohol use and sexual behaviour among risky drinkers and bar and shebeen patrons in Gauteng province, South Africa. Social science & medicine. 2006;**62**(1):217-27.

88 Cain D, Pare V, Kalichman SC, et al. HIV risks associated with patronizing alcohol serving establishments in South African Townships, Cape Town. Prev Sci. 2012 Dec;**13**(6):627-34.

89 Watt MH, Aunon FM, Skinner D, Sikkema KJ, Kalichman SC, Pieterse D. "Because he has bought for her, he wants to sleep with her": alcohol as a currency for sexual exchange in South African drinking venues. Soc Sci Med. 2012 Apr;**74**(7):1005-12.

90 Weir SS, Pailman C, Mahlalela X, Coetzee N, Meidany F, Boerma JT. From people to places: focusing AIDS prevention efforts where it matters most. AIDS. 2003 Apr 11;**17**(6):895-903.

91 Wohl DA, Khan MR, Tisdale C, et al. Locating the places people meet new sexual partners in a southern US city to inform HIV/STI prevention and testing efforts. AIDS Behav. 2011 Feb;**15**(2):283-91.

92 Weir SS, Merli MG, Li J, et al. A comparison of respondent-driven and venue-based sampling of female sex workers in Liuzhou, China. Sex Transm Infect. 2012 Dec;**88 Suppl 2**:i95-101.

93 Singh K, Sambisa W, Munyati S, et al. Targeting HIV interventions for adolescent girls and young women in Southern Africa: use of the PLACE methodology in Hwange District, Zimbabwe. AIDS Behav. 2010 Feb;**14**(1):200-8.

94 Speizer IS, Beauvais H, Gomez AM, Outlaw TF, Roussel B. Using multiple sampling approaches to measure sexual risk-taking among young people in Haiti: programmatic implications. Stud Fam Plann. 2009 Dec;**40**(4):277-88.

95 Tate J, Singh K, Ndubani P, Kamwanga J, Buckner B. Measurement of HIV prevention indicators: a comparison of the PLACE method and a household survey in Zambia. AIDS Behav. 2010 Feb;**14**(1):209-17.

96 Khan MR, Rasolofomanana JR, McClamroch KJ, et al. High-risk sexual behavior at social venues in Madagascar. Sex Transm Dis. 2008 Aug;**35**(8):738-45.

97 Mwandagalirwa K, Jackson EF, McClamroch K, Bollinger R, Ryder RW, Weir SS. Local differences in human immunodeficiency virus prevalence: a comparison of social venue patrons, antenatal patients, and sexually transmitted infection patients in eastern kinshasa. Sex Transm Dis. 2009 Jul;**36**(7):406-12.

98 Figueroa JP, Weir SS, Byfield L, Hall A, Cummings SM, Suchindran CM. The challenge of promoting safe sex at sites where persons meet new sex partners in Jamaica: results of the Kingston PLACE randomized controlled trial. Trop Med Int Health. 2010 Aug;**15**(8):945-54.

99 Weir SS, Tate JE, Zhusupov B, Boerma JT. Where the action is: monitoring local trends in sexual behaviour. Sex Transm Infect. 2004 Dec;**80 Suppl 2**:ii63-8.

100 Weir SS, Morroni C, Coetzee N, Spencer J, Boerma JT. A pilot study of a rapid assessment method to identify places for AIDS prevention in Cape Town, South Africa. Sex Transm Infect. 2002 Apr;**78 Suppl 1**:i106-13.

101 Sivaram S, Srikrishnan AK, Latkin C, et al. Male alcohol use and unprotected sex with non-regular partners: evidence from wine shops in Chennai, India. Drug Alcohol Depend. 2008 Apr 1;**94**(1-3):133-41.

102 Parks KA, Hsieh YP, Collins RL, Levonyan-Radloff K, King LP. Predictors of risky sexual behavior with new and regular partners in a sample of women bar drinkers. J Stud Alcohol Drugs. 2009 Mar;**70**(2):197-205.

103 Kalichman SC, Simbayi LC, Vermaak R, Jooste S, Cain D. HIV/AIDS risks among men and women who drink at informal alcohol serving establishments (Shebeens) in Cape Town, South Africa. Prev Sci. 2008 Mar;**9**(1):55-62.

104 Fritz KE, Woelk GB, Bassett MT, et al. The association between alcohol use, sexual risk behavior, and HIV infection among men attending beerhalls in Harare, Zimbabwe. AIDS and Behavior. 2002;**6**(3):221-8.

105 Bassett MT, McFarland WC, Ray S, et al. Risk factors for HIV infection at enrollment in an urban male factory cohort in Harare, Zimbabwe. JAIDS Journal of Acquired Immune Deficiency Syndromes. 1996;**13**(3):287-93.

106 Lewis JJ, Garnett GP, Mhlanga S, Nyamukapa CA, Donnelly CA, Gregson S. Beer halls as a focus for HIV prevention activities in rural Zimbabwe. Sex Transm Dis. 2005 Jun;**32**(6):364-9.

107 Go VF, Solomon S, Srikrishnan AK, et al. HIV rates and risk behaviors are low in the general population of men in Southern India but high in alcohol venues: results from 2 probability surveys. J Acquir Immune Defic Syndr. 2007 Dec 1;**46**(4):491-7.

108 Kalichman SC, Simbayi LC, Vermaak R, et al. Randomized trial of a community-based alcohol-related HIV risk-reduction intervention for men and women in Cape Town South Africa. Ann Behav Med. 2008 Dec;**36**(3):270-9.

109 Fritz K, McFarland W, Wyrod R, et al. Evaluation of a peer network-based sexual risk reduction intervention for men in beer halls in Zimbabwe: results from a randomized controlled trial. AIDS Behav. 2011 Nov;**15**(8):1732-44.

110 Power R, Langhaug L, Cowan F. "But there are no snakes in the wood": risk mapping as an outcome measure in evaluating complex interventions. Sex Transm Infect. 2007 Jun;**83**(3):232-6.

111 Kandala NB, Ji C, Cappuccio PF, Stones RW. The epidemiology of HIV infection in Zambia. AIDS Care. 2008 Aug;**20**(7):812-9.

112 Kleinschmidt I, Pettifor A, Morris N, MacPhail C, Rees H. Geographic distribution of human immunodeficiency virus in South Africa. Am J Trop Med Hyg. 2007 Dec;**77**(6):1163-9.

113 Messina JP, Emch M, Muwonga J, et al. Spatial and socio-behavioral patterns of HIV prevalence in the Democratic Republic of Congo. Soc Sci Med. 2010 Oct;**71**(8):1428-35.

114 Moran D, Jordaan JA. HIV/AIDS in Russia: determinants of regional prevalence. Int J Health Geogr. 2007;**6**:22.

115 Tanser F, Lesueur D, Solarsh G, Wilkinson D. HIV heterogeneity and proximity of homestead to roads in rural South Africa: an exploration using a geographical information system. Trop Med Int Health. 2000 Jan;**5**(1):40-6.

116 Barnighausen T, Tanser F, Gqwede Z, Mbizana C, Herbst K, Newell ML. High HIV incidence in a community with high HIV prevalence in rural South Africa: findings from a prospective population-based study. AIDS. 2008 Jan 2;**22**(1):139-44.

117 Gabrysch S, Edwards T, Glynn JR. The role of context: neighbourhood characteristics strongly influence HIV risk in young women in Ndola, Zambia. Trop Med Int Health. 2008 Feb;**13**(2):162-70.

118 Moise I, Kalipeni E. Applications of geospatial analysis to surveillance data: a spatial examination of HIV/AIDS prevalence in Zambia. GeoJournal. 2012 2012/08/01;**77**(4):525-40.

119 Room R, Babor T, Rehm J. Alcohol and public health. Lancet. 2005 Feb 5-11;**365**(9458):519-30.

120 Campbell CA, Hahn RA, Elder R, et al. The effectiveness of limiting alcohol outlet density as a means of reducing excessive alcohol consumption and alcohol-related harms. Am J Prev Med. 2009 Dec;**37**(6):556-69.

121 Popova S, Giesbrecht N, Bekmuradov D, Patra J. Hours and days of sale and density of alcohol outlets: impacts on alcohol consumption and damage: a systematic review. Alcohol Alcohol. 2009 Sep-Oct;**44**(5):500-16.

122 Livingston M, Chikritzhs T, Room R. Changing the density of alcohol outlets to reduce alcohol-related problems. Drug Alcohol Rev. 2007 Sep;**26**(5):557-66.

123 Bryden A, Roberts B, McKee M, Petticrew M. A systematic review of the influence on alcohol use of community level availability and marketing of alcohol. Health Place. 2012 Mar;**18**(2):349-57.

124 Grube JW, Stewart K. Preventing impaired driving using alcohol policy. Traffic Inj Prev. 2004 Sep;**5**(3):199-207.

125 Alaniz ML. Alcohol availability and targeted advertising in racial/ethnic minority communities. Alcohol Health Res World. 1998;**22**(4):286-9.

126 Cook PJ, Moore MJ. The economics of alcohol abuse and alcohol-control policies. Health Aff (Millwood). 2002 Mar-Apr;**21**(2):120-33.

127 Babor TF, Aguirre-Molina M, Marlatt GA, Clayton R. Managing alcohol problems and risky drinking. Am J Health Promot. 1999 Nov-Dec;**14**(2):98-103.

128 Cohen DA, Ghosh-Dastidar B, Scribner R, et al. Alcohol outlets, gonorrhea, and the Los Angeles civil unrest: a longitudinal analysis. Soc Sci Med. 2006 Jun;**62**(12):3062-71.

129 Scribner RA, Cohen DA, Farley TA. A geographic relation between alcohol availability and gonorrhea rates. Sex Transm Dis. 1998 Nov;**25**(10):544-8.

130 Theall KP, Scribner R, Cohen D, et al. The neighborhood alcohol environment and alcohol-related morbidity. Alcohol Alcohol. 2009 Sep-Oct;**44**(5):491-9.

131 Nichols BE, Nkalamo D, Whitcomb BW. Density of drinking establishments and HIV prevalence in a migrant town in Namibia. AIDS Behav. 2012 May;**16**(4):829-34.

132 Eaton LA, Kalichman SC, Pitpitan EV, et al. The relationship between attending alcohol serving venues nearby versus distant to one's residence and sexual risk taking in a South African township. J Behav Med. 2013 Feb 13.

133 Kahn K, Collinson MA, Gomez-Olive FX, et al. Profile: Agincourt Health and Socio-demographic Surveillance System. Int J Epidemiol. 2012 Aug;**41**(4):988-1001.

134 Kahn K, Tollman SM, Collinson MA, et al. Research into health, population and social transitions in rural South Africa: data and methods of the Agincourt Health and Demographic Surveillance System. Scand J Public Health Suppl. 2007 Aug;**69**:8-20.

135 Kleinbaum DG, Klein M. Ordinal logistic regression. Logistic Regression: Springer; 2010. p. 463-88.

136 Delany-Moretlwe S, Jentsch U, Weiss H, et al. Comparison of focus HerpesSelect and Kalon HSV-2 gG2 ELISA serological assays to detect herpes simplex virus type 2 antibodies in a South African population. Sexually transmitted infections. 2010;**86**(1):46-50.

137 Subramanian S, Jones K, Duncan C. Multilevel methods for public health research. In: Kawachi I, Berkman L, editors. Neighborhoods and Health. New York, NY: Oxford University Press; 2003.

138 Diez Roux AV, Aiello AE. Multilevel analysis of infectious diseases. J Infect Dis. 2005 Feb 1;**191 Suppl 1**:S25-33.

139 Hanley JA, Negassa A, Forrester JE. Statistical analysis of correlated data using generalized estimating equations: an orientation. American journal of epidemiology. 2003;**157**(4):364-75.

140 Carlin JB, Wolfe R, Brown CH, Gelman A. A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. Biostatistics. 2001;**2**(4):397-416.

141 Kleinbaum DG, Klein M. Logistic Regresion for Correlated Data: GEE. Logistic Regression: A Self-Learning Text. 2002:327-75.

142 Todd J, Carpenter L, Li X, Nakiyingi J, Gray R, Hayes R. The effects of alternative study designs on the power of community randomized trials: evidence from three studies of human immunodeficiency virus prevention in East Africa. Int J Epidemiol. 2003 Oct;**32**(5):755-62.

143 Killip S, Mahfoud Z, Pearce K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. Ann Fam Med. 2004 May-Jun;**2**(3):204-8.

144 Garenne M, Tollman S, Kahn K. Premarital fertility in rural South Africa: a challenge to existing population policy. Stud Fam Plann. 2000 Mar;**31**(1):47-54.

145 Williams J, Ibisomi L, Sartorius B, et al. Convergence in fertility of South Africans and Mozambicans in rural South Africa, 1993-2009. Glob Health Action. 2013;**6**:19236.

146 Collinson M, Gerritsen A, Clark S, Kahn K, Tollman S. Migration and Socio-Economic change in rural South Africa, 2000–2007. In: Collinson M, Adazu K, White M, Findley S, editors. The Dynamics of Migration, Health and Livelihoods: INDEPTH Network Perspectives. Farnham, England: Ashgate Publishing Limited; 2009.

147 Republic of South Africa: Government Communication and Information System. South Africa Government Information: School Calendar. 2013 [cited 2013; Available from: http://www.info.gov.za/aboutsa/schoolcal.htm

148 HIV Prevention Trials Network. HPTN 068: Effects of Cash Transfer for the Prevention of HIV in Young South African Women. 2011 [cited 2012 November 1, 2012]; Available from: http://www.hptn.org/research_studies/hptn068.asp

149 Health Systems Trust. Terminations of Pregnancy by South African Province. 2013 [cited 2013; Available from: http://indicators.hst.org.za/healthstats/47/data

150 Statistics South Africa. Recorded Live Births in 2010; 2011.

151 Meghdadpour S, Curtis S, Pettifor A, MacPhail C. Factors associated with substance use among orphaned and non-orphaned youth in South Africa. Journal of Adolescence. 2012;**35**(5):1329-40.

152 Reddy S, James S, Sewpaul R, et al. Umthente Uhlana Usamila - The South African Youth Risk Behavior Survey 2008. Cape Town; 2010.

153 Kalichman S. Social and Structural HIV Prevention in Alcohol-Serving Establishments. Alcohol research & health: the journal of the National Institute on Alcohol Abuse and Alcoholism. 2009;**33**(3):184-94.