Age-disparate Partnerships and Socioeconomic Status

A study on HIV risk for adolescent girls and young women in KwaZulu-Natal, South Africa

Beatrice Holgersson & Iina Ojala

Abstract: This thesis investigates the relationship between socioeconomic status and age-disparate partnerships for adolescent girls and young women aged 15-24 in KwaZulu-Natal. Age-disparate partnering is defined as engaging in a sexual partnership with a partner five years older or more. Separate measures of education level, employment status, and household income are used as indicators for socioeconomic status. The logistic regression model estimates the relationships. The findings imply an association between socioeconomic status and age-disparate partnerships. There is a significant negative relationship between education level and age-disparate partnerships. The same inverse pattern is suggested for household income, although non-significant and more modest. The probability for students engaging in age-disparate partnerships is lower compared to other employment categories. The probability of age-disparate partnering increase with age for all socioeconomic measures.

Key terms: Socioeconomic status, HIV**,** Age disparate partnership, Adolescent girls & young women, 90-90-90 treatment targets

Bachelor's thesis in Economics, 15 credits
Spring semester 2020
Supervisor: Dick Durevall

Department of Economics
School of Business, Economics and Law
University of Gothenburg

Table of contents

1. Introduction

The AIDS epidemic is supposed to be terminated by 2030 to advance the quality of people's lives. Despite progress, global HIV data shows that we are still far away from reaching that goal. The number of new HIV infections is still too high; since in 2018 about 1.7 million new HIV infections were reported (UNAIDS 2019). The 90-90-90 treatment targets were defined in 2013 to end the AIDS epidemic by 2030. If the targets are accomplished by 2020 the goal of putting a stop to the epidemic by 2030 is possible. The targets are aiming at that by 2020, 90 % of the people living with HIV know their status, 90 % of the people who are diagnosed with HIV receive antiretroviral therapy and that 90 % of the people who receive antiretroviral therapy have viral suppression (UNAIDS 2014).

Inequalities concerning who acquires HIV needs to be addressed to put a stop to the AIDS epidemic. Despite progress, adolescent girls and young women are still considered as an extra vulnerable group due to a high risk of contracting HIV. Adolescent girls and young women aged 15-24 only accounts for 11 % of the global population but represented 20 % of the new HIV infections among adults worldwide in 2015 (UNAIDS 2016). In 2018, 6 000 adolescent girls and young women aged 15-24 acquired HIV every week (UNAIDS 2019). High prevalence areas, vulnerable groups, and transmission patterns need to be pinpointed and be focused on to put a stop to the AIDS epidemic by 2030 (UNAIDS 2017).

The epidemic has struck the region of Eastern and Southern Africa the hardest. In 2018, 37.9 million people in the world lived with HIV, 20.6 million of these people lived in Eastern and southern Africa. Regarding the 90-90-90 targets, there is a gap identified between actual and desired values, since the values in 2018 measured 85%, 67%, and 58% (UNAIDS 2019). The grand number of new HIV infections amongst adolescent girls and young women alongside gender inequalities and an increasing number of sexual partners among men are identified as challenges for the region to reach the 90-90-90 targets (UNAIDS 2015).

South Africa is identified as one of the priority countries within the hard struck Eastern and Southern Africa region, with Durban and Johannesburg as critical cities for preventative actions (UNAIDS 2015). South Africa has the largest proportion of people living with HIV globally, and a skewed distribution between genders can be identified. In 2018, 7.7 million

people lived with HIV in South Africa; 4.7 million were women (aged 15+), and 2.8 million were men (aged 15+). Regarding new HIV-infections in 2018, 240 000 were newly infected, whereas 140 000 of them were women (aged 15+), and 86 000 were men (aged 15+). The desired measurements concerning the 90-90-90 targets were not fulfilled in 2018 as a distribution of 90%, 68 %, 87 % could be identified (UNAIDS 2019). The HIV prevalence in South Africa is unevenly distributed amongst the provinces. The province of KwaZulu-Natal experiences the highest HIV prevalence in all of South Africa at 18.1 % (HSRC 2019) and is even considered to be the epicenter of the epidemic as it probably experiences the highest HIV prevalence in the world (CAPRISA 2017).

Multiple studies try to understand the transmission patterns fueling the epidemic in the high prevalence setting of Kwazulu-Natal. Age-disparate partnerships, defined as engaging in a sexual partnership with a partner five years older or more (UNAIDS 2015), between young females and older men, is suggested to be a key factor within the transmission cycle fueling the epidemic in KwaZulu-Natal (de Oliveira et al. 2017). Although ambiguous results, several studies suggest that women aged 15-24 declaring age-disparate partnerships in any of their last three relationships run a higher risk of being HIV positive compared to women of the same age who reports age-similar partnerships (Maughan-Brown et al. 2018; Stoner et al. 2019). Evidence suggests that higher school attendance and staying in school decreases the engagement in age-disparate partnerships (Stoner et al. 2017).

The relationship between HIV and socioeconomic indicators have frequently been investigated throughout the epidemic. Evidence suggests a correlation between socioeconomic indicators and HIV incidence for women, implying that women with low socioeconomic status are the most vulnerable to HIV in the South African context (Hargreaves et al. 2007). Research suggests that educational attainments reduce HIV incidence (Bärnighausen, Hosegood, Timaeus & Newell 2007). However, it has been stated that the negative correlation between school attendance and HIV for young girls is due to unobserved differences between girls who stay in school and those who drop out (Durevall, Lindskog & George 2019). Most evidence suggests a negative association between relative wealth and HIV incidence (Bunyasi & Coetzee 2017), and research implies that no correlation exists between employment status and HIV incidence for women (Bunyasi & Coetzee 2017). However, unemployment and the economic insecurity it comes with are suggested to construct a key feature driving high HIV incidence rates for young women in

developing countries. Economic insecurity is implied to decrease young females' bargaining power and increase the vulnerability for young females to engage in sexual relations that can affect their health outcomes (Austin, Choi & Berndt 2017).

Earlier research leaves a gap on what distinguishes the young females engaging in age-disparate partnerships in comparison to females who do not, as well as a knowledge gap in the relationship between socioeconomic status and the behavior of engaging in age-disparate partnerships. Hence, this study aims to fill that gap by investigating the relationship between socioeconomic status and age-disparate partnerships for the vulnerable group of adolescent girls and young women in the high prevalence setting of KwaZulu-Natal. The research question and hypothesis are based on previously mentioned evidence, suggesting that age-disparate partnerships increase young females' HIV incidence and the presented mechanisms of how low socioeconomic measures increase HIV vulnerability. The question addressed is if a relationship exists between socioeconomic status and age-disparate partnerships among adolescent girls and young women aged 15-24 in KwaZulu-Natal. This study hypothesizes that a low socioeconomic status increases the probability of engagement in age-disparate partnerships.

The relationship between socioeconomic status and the probability of engaging in age-disparate partnerships is investigated by using the logistic regression model. The dataset used to conduct the analysis is from an area within KwaZulu-Natal. Separate measures of education level, employment status, and household income are used as indicators for socioeconomic status.

The purpose of this study is founded on the importance of understanding various characteristics of the vulnerable group of adolescent girls and young women and their participation within a transmission pattern. All in order to reach the goal of terminating the AIDS epidemic by 2030. Further understanding of the socioeconomic characteristics of the young females engaging in age-disparate partnerships can be valuable in order to create suitable preventative actions to put a stop to the transmission pattern fueling the epidemic in KwaZulu-Natal.

The remaining sections of the thesis proceed as follows. Section 2 presents the relevant theoretical framework. Section 3 presents a literary review over relevant previous research. Section 4 presents the data used to execute the analysis and descriptions of the variables. Section 5 presents the empirical method used to conduct the analysis. Section 6 presents the results and tables. Section 7 presents a discussion about the main findings and limitations. Section 8 presents the conclusion and suggestions for further research.

2. Theoretical framework

The following section presents two theories that have provided guidance in creating the empirical method and in interpreting the results of this study. The theories provide a further understanding of what drives people to invest in health as well as a greater awareness of the complexity of making a decision that might have an adverse health outcome.

2.1 Grossman's model of health capital

Grossman's model of health capital is a key model in health economics. The basic idea of the model is that everyone is born with an inherent health capital that depreciates as the person grows older but can increase through investments. As the health capital falls below a certain minimum level, death will happen. The timing of death depends on decisions about health investments with full knowledge of their impact. The model accounts for unhealthy behaviors explained as bad health investments that decrease the health capital. (Cawley & Rhum 2011).

A key concept of the model is that people desire good health for two reasons. The first reason, a greater capital stock will increase a person's productivity and yield a higher income. The second reason, people desire good health as it provides greater utility. The amount of health a person possesses is determined only by how a person allocates his/her time. Activities can either improve, maintain, or decrease the health capital stock which will lead to a certain outcome of healthy days (Folland, Goodman & Stano 2013).

The composition of health investments (*I)* is described in equation (1) and is composed of time spent on improving health (T_H) and consumption on health market inputs such as medicine *(M)*. Equation (2) describes the amount of health capital (H_t) in time period *t*, which is a function of the amount of health capital left from the last period $(H(1-\gamma)H_{t-1})$, time spent on improving health $(T_H)$and consumption on health market inputs such as medicine *(M)* (Folland, Goodman & Stano 2013).

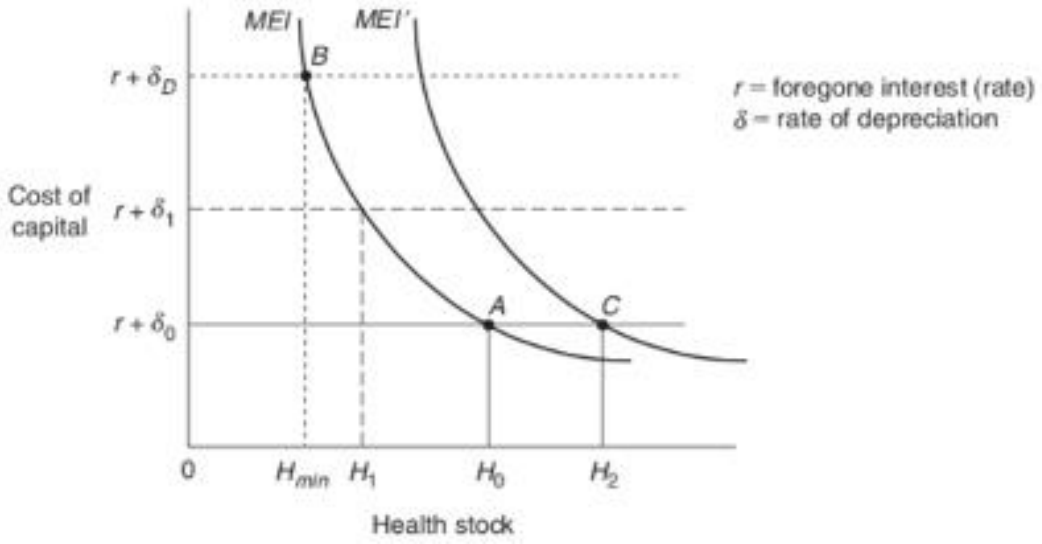
$$I = I(M\ T_H) \quad (1)$$

$$H_t = [H(1-\gamma)H_{t-1}, T_t^H, M_t] \quad (2)$$

The marginal efficiency of the investment curve (MEI) in Figure 1, describes the pattern of the rate of return on health investments, which indicates that as the amount of health investments increases the rate of return decreases. The slope is downward sloping due to the diminishing marginal returns in the production function for producing healthy days. The optimal health capital demanded (equilibrium) occurs when the cost of capital is equal to the marginal efficiency of investment curve, which occurs at point A and C in Figure 1. The cost of capital is the sum of the rate of depreciation due to age (δ) and the interest rate *(r)* (Folland, Goodman & Stano 2013).

Another key concept of the model is that the efficiency of health investments increases with the education level. The marginal product of inputs is raised by education, which reduces the number of inputs needed to create a large investment. So, people with higher education can experience a higher rate of return for the same cost of capital as a person with a lower level of education. As a product, more educated people will choose a higher optimal health capital stock than people who are less educated. In Figure 1, we can visualize a lower marginal efficiency curve for a less educated person (MEI) with an equilibrium at point A, and a higher curve for a more educated person (MEI´) with an equilibrium at point C. Point A and C show that a higher educated person demands a higher optimal health capital stock in comparison to a less educated person for the same cost of capital. (Folland, Goodman & Stano 2013).

Figure 1: Visualization of the optimal health capital stock and different MEI-curves due to different education levels



Source: Folland, Goodman & Stano (2013).

The Grossman model of health capital gives guidance in constructing our empirical model and interpreting the results. For example, an age-disparate partnership can be considered a bad health investment that could decrease the health capital due to increased HIV exposure. Education, which is one of the chosen socioeconomic indicators, could, in accordance with the model, lead to a higher health capital stock and more healthy days.

A problematic aspect of the model within the context of this study is that health investments are assumed to be done with full knowledge of their health impact. Even though research has stated that age-disparate partnerships increase the HIV risk for adolescent girls and young women (Maughan-Brown et al. 2018; Stoner et al. 2019), they themselves might not have the knowledge and be aware of the risk.

2.2 The theory of discounted utility

As discussed in the Grossman model, decisions concerning health have consequences in multiple time periods and can therefore be classified as intertemporal choices. Intertemporal choices are determined by time preference and a tradeoff exists between present and future

utility. The discount factor (δ)shows the degree of which a person discounts the future relative to the present. A low discount factor close to zero implies a high value on the present and discounting the future. A high discount factor close to one implies that a person does not discount the future. In order to analyze intertemporal choices, the theory of discounted utility can be applied with a focus on uncertainty and utility maximization. The following utility functions assumes rational choice and full information about risk (White & Dow 2015).

A lifetime utility model can be seen in equation (3). Lifetime utility (U) is a function of the consumption in each time period *(t =0,1,2,...,T),* u_0 is equal to the present utility, and $(\delta)$is equal to the discount factor, on which a person discounts future utility $(u_1, u_2, \ldots, u_T)$. The discount factor is constant in this model, in other words, a constant marginal rate of substitution exists (White & Dow 2015).

$$U = u_0 + \delta u_1 + \delta^2 u_2 + \delta^3 u_3 + \ldots \delta^T u_T \quad (3)$$

Equation (4) shows a model of expected utility E(U). Uncertainty is incorporated into the model, which captures the fact that people make decisions not only due to a certain outcome but also on the probability of an outcome. For example, making a choice about engaging in an age-disparate partnership entails consideration of the benefits of the partnership in relation to the uncertainty of contracting HIV. In other words, the probability of acquiring HIV and the probability of not acquiring HIV is taken into account. In equation (4), *p(s)* is equal to the probability of the outcome *s*, and *u(x|s)* denotes the instant utility consuming *x* given the outcome *s*. The instant utility of each outcome times its probability is equal to the expected utility (White & Dow 2015).

$$E(U) = p(1)\, u(x \mid 1) + p(2)\, u(x \mid 2) + \ldots + p(S)\, u(x \mid S) \quad (4)$$

$$= \sum_{S=1}^{S} p(s)\, u(x \mid s)$$

In order to capture how intertemporal choices are made, equation (3) and (4) are combined into equation (5). Equation (5) presents how a person makes intertemporal choices while maximizing expected lifetime utility, which can be done by maximizing the utility in every time period *(t = 0,1, …, T)*. The utility from the consumption of *x* is maximized in the present and future, with a discount factor on future utility. Each period entails uncertain choices, and the future is discounted by the factor δ(White & Dow 2015).

(5)

$$max_{xt} \sum_{S_0} p(s_0)\, u(x_0 \mid s_0) + \delta \sum_{S_1} p(s_1)\, u(x_1 \mid s_1) + \ldots + \delta^T \sum_{S_T} p(s_T)\, u(x_T \mid s_T)$$

$$= max_{xt} \sum_{t=0}^{T} \delta^t \sum_{St} p(st)\, u(x_t | s_t)$$

The central role of time preferences and tradeoffs between present and future utility in the context of health preventative actions has been discussed by Brougham & John (2007). Time preferences and utility tradeoffs are mentioned to have a larger impact on health decisions than the possible consequences of the decision itself. If the future is uncertain, the willingness to value utility in the future is reduced and the present utility is more valuable. It is emphasized that increased risk behavior can occur if the future is discounted for. Young adults often value the present more than the future compared to older individuals. The tradeoff between the future and the present is argued to be more influenced by personal values than lack of knowledge of consequences (Brougham & John 2007).

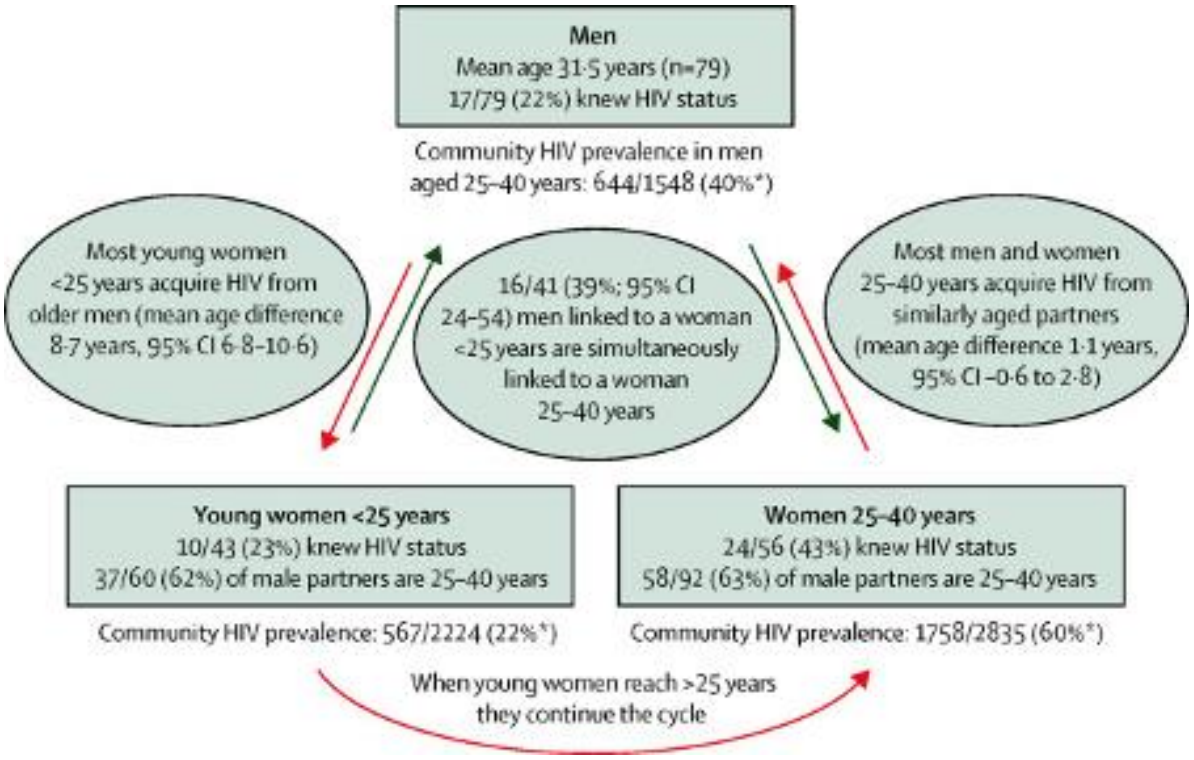
A discussion on time preferences and utility tradeoffs is essential to understand the choice of the age-disparate partnership due to its intertemporal nature. The adolescent girls and young women might not care about the long-term consequences the partnerships can expose them to if the future is uncertain. However, participating in age-disparate partnerships might yield higher utility despite the risk for HIV, which can steer the adolescent girls and young women into engaging in these relationships, although knowledge about the health consequences.

3. Literature review

Several studies have examined the transmission patterns and drivers of the HIV epidemic. The following review is a collection of relevant studies for a better understanding of the HIV epidemic in the province of KwaZulu-Natal, South Africa, with the main focus on young females, age-disparate partnerships, and socioeconomic measures.

Recent evidence has identified sexual networks fueling the epidemic in KwaZulu-Natal. The sexual networks power a transmission cycle that leads to a high number of adolescents' girls and young women aged 15-24 contracting HIV from men who are on average 8 years older (de Oliveira et al. 2017). The transmission networks consist of three interlinked groups; young women aged 15-24, men aged 25-40 and women aged 25-40. Dual linkages were identified within clusters and identified possible pairings, 39 % of the men who had linkages to the younger girls were also linked to the older group of women. The men who on average are in their 30`s presumably acquires the virus from women around the same age (usually their wifes), women between the ages of 25-40, which is the group with the highest prevalence of HIV in the area. Due to age-disparate partnerships and the identified dual linkages, men infect the young group of women. As the young cohort of women grow older, they fall into the category of women aged 25-40, creating the group with the highest prevalence. The new cohort of older women will then presumably, because of the fact of high prevalence, infect their sexual partners (usually their husbands), men in their 30's, who in their turn will infect the new cohort of young women. Consequently, age-disparate partnering becomes a key feature in the transmission cycle that generates the high HIV prevalence in KwaZulu-Natal. In order to reach the UNAIDS 90-90-90 treatment targets, preventions that pinpoints all three groups of the transmission cycle are crucial (de Oliveira et al. 2017). Figure 2 describes the sexual networks creating the transmission cycle.

Figure 2. Visualization of the sexual networks fueling the transmission cycle



Source: de Oliveira et al. (2017)

Several studies analyze the impact age-disparate partnerships have on adolescent girls and young women. For example, a previous research conducted in a rural and peri-urban area in KwaZulu-Natal, suggests a greater risk for females aged 15-24 to be HIV positive, if engaging in age-disparate partnering in any of their three most recent sexual partnerships, compared with females in the same age who reports age-similar partnerships (Maughan-Brown et al. 2018). Furthermore, another study conducted in an adjoining province to KwaZulu-Natal, agrees with the statement that age-disparate partnering increase the HIV risk for young females (Stoner et al. 2019). The evidence suggests that the risk of HIV infection is 12.6% higher if engaging in age-disparate partnerships than if not, and the findings concerning increased HIV risk still holds when controlling for risky sexual behavior, such as condom less sex as well as for transactional sex. Stoner et al. (2019) furthermore argues that age-disparate partnering increases young females' HIV risk, due to the fact that older men experience higher HIV prevalence than younger men. Consequently, age-disparate partnerships increase young females' exposure of the virus, compared with having sexual partnerships with males in their own age.

However, the evidence on the impact of age-disparate partnerships on HIV incidence for young females has been inconclusive within the province of KwaZulu-Natal. A study conducted in a deprived rural community in the province suggests, in contrast with previously mentioned research, no association between age-disparate partnerships and increased HIV risk for young women (Harling et al. 2014). Furthermore, Harling et al. (2014) mentions socioeconomic measures in relation to age-disparate partnerships, and how the results on HIV risk from age-disparate partnerships remain similar when controlling for socioeconomic factors such as education and household wealth. Nevertheless, Harling et al. (2014) mentions that the socioeconomic differences within the investigated community are not as pronounced as in other settings in Kwazulu-Natal, which might have led to the finding on the weak link between economic inequalities and age-disparate partnerships.

A small number of studies have focused on the relationship between socioeconomic indicators and engagement in age-disparate partnerships. However, one recent study in rural South Africa investigates who the adolescent girls and young women engaging in an age-disparate partnership are, with the main focus on the socioeconomic indicator of education (Stoner et al. 2017). The study suggests a relationship between school attendance and the likelihood of age-disparate partnerships. Stoner et al. (2017) argue that fewer days in school and dropping out of school increase the probability of age-disparate partnerships. Dropping out of school did not just have an association with the occurrence of age-disparate partnerships but also with more sexual partners; consequently, an association with higher-risk behavior. Stoner et al. (2017) imply that adolescent girls and young women who stay in school and have higher school attendance are more likely to have partners close to their age and being part of safer networks.

The relationship between HIV and socioeconomic measures have frequently been investigated throughout the epidemic. A population-based study conducted in a poor rural community in KwaZulu-Natal investigates the effect socioeconomic status has on HIV incidence (Bärnighausen, Hosegood, Timaeus & Newell 2007). Household wealth, expenditure, and educational attainments are used as indicators for socioeconomic status. The findings suggest that an increase in educational attainment reduces the HIV incidence while household expenditures do not have any effect at all. Evidence on household wealth suggests that middle-income households in terms of relative wealth experience the highest HIV

incidence. Bärnighausen, Hosegood, Timaeus & Newell (2007) emphasize that the most efficient way to reduce HIV incidence in the investigated rural setting of Kwazulu-Natal is to focus on increased educational attainments and not poverty reduction.

An additional study in rural South Africa suggests a correlation between socioeconomic measures and HIV incidence for women but not for men (Hargreaves et al. 2007). The study highlights that women with lower socioeconomic measures in rural settings are the most vulnerable to HIV within the South African context. Hargreaves et al. (2007) suggest a correlation between low education and higher HIV incidence for women. Furthermore, being a low educated woman is associated with more sexual partners and higher levels of infection relative to other groups. The evidence suggests a higher amount of risky behavior, such as less use of condoms within the group of low-income women and for the least educated women. In line with previously mentioned research, Hargreaves et al. (2007) emphasize the relationship between education level and HIV incidence, as the evidence suggests a negative correlation between levels of education and HIV incidence for women.

Research further investigates the frequently mentioned negative correlation between the socioeconomic measure of education and HIV incidence for women. A study in a rural community in KwaZulu-Natal investigates the causal effects behind the negative correlation between school attendance and HIV incidence for young women aged 15-24 (Durevall, Lindskog & George 2019). The findings suggest that the negative correlation is due to unobserved differences between the girls who stay in school and the ones who drop out. Durevall, Lindskog & George (2019) imply that the causal effect behind the negative correlation cannot be dismissed, even though it might be small. The authors provide examples of factors such as time preferences, family background, academic ability, and self-control that could explain the negative association between School attendance and HIV incidence. Durevall, Lindskog & George (2019) do not support policies that only focus on increased school attendance to reduce HIV incidence for the group of young women aged 15-25 in Kwazulu-Natal.

The relationship between socioeconomic status and HIV incidence for South African women is further investigated by comparing two different provinces; Western Cape and Free State (Bunyasi & Coetzee 2017). Socioeconomic status is measured through household wealth, years of formal education, and employment status. The associations differ by the province as

well as by measurement of socioeconomic status. Bunyasi and Coetzee (2017) suggest a negative association between relative wealth and HIV incidence in both provinces since women within the highest wealth quintile possess the lowest HIV incidence in both areas. However, the evidence is inconclusive concerning the wealth quintile that possesses the highest risk for HIV, given that the highest risk deviates between the lowest and the second-lowest wealth quintile depending on the province. The relationship between formal education and HIV incidence differ; a strong negative association is suggested for the province of Western Cape, as one extra year of education decrease HIV incidence by 10%. While in contradiction to earlier mentioned research, for example, by Hargreaves et al. (2007), education barely has any effect on HIV incidence in Free State. The evidence shows no significant correlation between employment status and HIV incidence in any of the investigated provinces (Bunyasi & Coetzee 2017).

Further research suggests that unemployment is a key feature driving the high HIV incidence rate for young women aged 15-24 in developing countries (Austin, Choi & Berndt 2017). Evidence indicates that economic insecurity is an essential element in the significant relationship between unemployment and HIV prevalence among young women. Austin, Choi & Berndt (2017) argue that increased economic insecurity contributes to young women's HIV incidence through reduced bargaining power and higher vulnerability of engaging in sexual relations that can affect their health outcomes, such as exchanging sex for economic security. However, Austin, Choi & Berndt (2017) emphasize that unemployment is not the only driver for the high levels of HIV incidence experienced by young women in developing countries.

4. Data

The dataset used in this study is the baseline survey of the HIV incidence provincial surveillance system study [hereinafter the HIPSS study]. The survey was conducted to measure HIV prevalence and incidence levels in the high prevalence setting of KwaZulu-Natal, South Africa, and consisted socioeconomic, demographic, behavioral, psychosocial, and HIV related questions (Kharsany et al. 2015). Through cross-sectional household surveys between June 11, 2014, and June 22, 2015, the data was gathered in the municipality of uMgungundlovu, in one rural and one peri-urban area. Households were selected on random, and one individual aged 15-49 within the selected household was likewise selected randomly. The random selection conducted total 9 812 individuals who were enrolled in the survey, 6 265 women and 3 547 men.

This study focuses on young females aged 15-24, which constitutes 2 224 individuals of the dataset. The HIPSS study states that the data is representative of the whole province of Kwazulu-Natal, despite that all observations derive from one municipality within the province (CAPRISA 2017). Hence, this study presumes the population to consist of young females aged 15-24 within the province of Kwazulu-Natal, and not merely within the municipality of uMgungundlovu where the data was gathered. The variables in this study derive from the dataset described above. The questionnaires for the cross-sectional surveys provides guidance in understanding the existing variables as well as in the process of constructing new variables (Kharsany et al. 2015).

To investigate the research question, a variable to measure engagement in age-disparate partnerships is created. The construction of the variable is based on the definition of age-disparate partnering earlier mentioned; engaging in a sexual partnership with a partner five years older or more (UNAIDS 2015). Most of the research on age-disparate partnerships have created a limitation to their research by only taking the three most recent sexual partners into account. But as the sample is rather small, it is crucial not to lose any observations, hence the decision of taking all the data on age difference with a sexual partner into account. There are two questions in the questionnaire that provides information about age of sexual partners. One question provides information on the age of sexual partner at first sex and the second on the age of the three most recent sexual partners. The variables connected to this data are

merged and creates the dependent variable, resulting in a binary variable that can capture the probability to engage in age-disparate partnerships. The variable takes on the value 1 if the female reports age-disparate partnership at first sex or with at least one of their three most recent sexual partners and the value 0 otherwise.

643 (28.91 %) females in the sample have engaged in age-disparate partnership at first sex or with the three most recent sexual partners and 1 581 (71.09 %) have not, giving the dependent variable a mean of 0.29 (Table 1).

The socioeconomic indicators of education level, employment status, and household income are separately measured and analyzed to capture the individuals' socioeconomic status within the sample. The variable capturing education level is an ordinal variable describing the young females' highest reported education, consisting of the following five different education levels: no schooling, primary school, incomplete secondary school, completed secondary school and tertiary education. Each level is represented by a dummy variable, taking on the value 1 if the education level is reported as highest education and the value 0 otherwise. The most common reported highest education within the sample are completed secondary (45 %) and incomplete secondary (45 %). Completed secondary is put as the reference category in the upcoming regressions.

The variable capturing the reported employment status of the young females is a nominal variable, consisting of the following employment categories: student, employed, unemployed and other employment status. Each employment status is represented by a dummy variable, which takes on the value 1 if the employment status is reported and the value 0 otherwise. For clarification, the dummy variable for the category of Other represents the categories of; housewife, retired, Ill, invalid or unable to work, or retrenched. The most common reported employment statuses within the sample are the categories of unemployed (43%) and student (41 %). The category of unemployed is put as a reference category in the upcoming regressions.

The dataset only consisted of income variables measuring the total income per household without accounting for the number of people living within the household. Since the household member count differs amongst the females in the sample, a new variable is created to capture the household income better. The new household income variable is an ordinal

variable that describes the estimated average income share per individual in each household [hereinafter household income]. The variable consists of the following income categories: no income, R1- R500, R501 - R2 500 and greater than R 2501. Each income category is represented by a dummy variable that takes on the value 1 if the income is fulfilled and the value 0 otherwise. 47 % of the females within the sample possess the household income of R501 - R2500, which makes it the most common income category within the sample. The household income category of R501 - R2500 is put as a reference category in the upcoming regressions.

The young female's engagement in age-disparate partnerships can be determined by many different factors and not simply by socioeconomic status. Hence, control variables are incorporated into the estimated model to obtain more accurate results. Due to the age interval of the population (15-24), it is crucial to control for age as presumably some of the females may never have engaged in sexual partnering, which may affect the probability of engaging in an age-disparate partnership. It can also be assumed that the young females possess different lifestyles and experiences due to age, as a young female aged 15 probably experience a different way of living in comparison to a young female aged 24, which can lead to different socioeconomic measures. The variable capturing the age of the females is an ordinal variable consisting of age-intervals between two years; 15-16, 17-18, 19-20, 21-22, 23-24. Each age-interval is represented by a dummy variable that takes on the value 1 if the age is fulfilled and the value 0 otherwise. The age-interval of 15-16 is put as a reference category in the upcoming regressions.

This study also controls for if the female lives with adult relatives or not. The choice of this control variable is based on evidence from previously mentioned research; on how economic insecurity can increase young females' vulnerability to engage in sexual relations that may have a negative impact on their health (Austin, Choi & Berndt 2017). Due to the fact that most females within the sample are unemployed, a young female living alone may be in a more economically insecure situation in comparison to a young female who lives with adult relatives, as the adult relatives assumedly can contribute to the economic situation of the household. Hence, if the female lives with adult relatives or not may have an effect on the probability to engage in age-disparate partnerships. Controlling for if the female lives with adult relatives can be of help in estimating the true effects from the socioeconomic indicators. The variable controlling for if the females still live at home or not is a dummy variable, that

takes on the value 1 if the young female lives with a parent, grandparent, aunt or uncle and takes on the value 0 otherwise. Table 1 illustrates the descriptive statistics of the variables.

Table 1. Variable names and descriptive statistics

Dependent variable	Type	Observations	Percentage (%)	Cumulative (%)
Age-disparate partnership	Dummy	2,224		
Yes		643	28.9	28.9
No		1,581	71.1	100.0
Independent variables	**Type**	**Observations**	**Percentage (%)**	**Cumulative (%)**
Education level	Ordinal	2,222		
No schooling		51	2.3	2.3
Primary		44	2.0	4.3
Incomplete secondary		994	44.7	49.0
Completed secondary		1,004	45.2	94.19
Tertiary		129	5.8	100.0
Employment status	Nominal	2,219		
Student		913	41.1	41.1
Employed		308	13.9	55.2
Unemployed		944	42.5	97.6
Other		54	2.4	100.0
Household income	Ordinal	2,024		
No income		289	14.3	14.3
R1-R500		687	33.9	48.2
R501-R2500		954	47.1	95.4
>R2500		94	4.6	100.0
Control variables	**Type**	**Observations**	**Percentage (%)**	**Cumulative (%)**
Age	Ordinal	2,224		
15-16 years		210	9.4	9.4
17-18 years		491	22.1	31.5
19-20 years		521	23.4	55.0
21-22 years		509	22.9	77.8
23-24 years		493	22.2	100.0
Lives with adult relatives	Dummy	2,224		
Yes		1,497	67.3	67.3
No		727	32.7	100.0

Note: This table presents variable names and descriptive statistics of variables of interest in this study.

5. Empirical Method

The choice of econometric model is based on the fact that the dependent variable is binary and will estimate the probability of Y taking on a certain value given the predictors. In this situation it makes sense to use a nonlinear model as the effects are nonlinear within the probability scale. Therefore, we use the logistic regression model which is a nonlinear model especially constructed for binary dependent variables. The model can handle the non-linearity of the true population regression function observed in equation (6), as well as ensure conditional probabilities within the interval of 0 – 1 for all values the predictors takes on. The logistic regression model described in equation (7) uses the cumulative standard logistic distribution function (*F*) in order to estimate the probability of the dependent variable taking on the value 1 given the independent variables (Stock & Watson 2015). In other words, within the context of this study, the logistic regression estimates the probability of engaging in age-disparate partnerships given the independent variables.

$$\mathrm{E}\,(\mathrm{Y}|\,\mathrm{X}) = \Pr(\mathrm{Y}{=}1|\,\mathrm{X}) \quad (6)$$

$$Pr\,(Y = 1|X_1\,, X, .., X_k) = \; F(\,\beta_0 \; + \; \beta_1 X_1 \; + \; \beta_2 X_2 + \ldots \beta_k X_k) \quad (7)$$

Equation (8) describes the estimated model, which is the estimated model for the main regression in the analysis. For clarification, *p is* equal to the probability of age-disparate partnership to occur, which makes $log(\frac{p}{1-p})$ equal to the log-odds of age-disparate partnership to occur. All socioeconomic variables as well as both control variables are incorporated into the estimated model and are denoted as follows: education level (X_1), employment status (X_2), household income (X_3), age (X_4) and if the female lives with adult relatives (X_5). Table A2 in appendix presents the estimated model.

$$log(\frac{p}{1-p}) = \beta_0 \; + \; \beta_1 X_1 \; + \; \beta_2 X_2 \; + \beta_3 X_3 \; + \beta_4 X_4 \; + \beta_5 X_5 \; (8)$$

The logistic function observed in equation (9) converts measurements within the log-odds scale (linear) into the probability scale (non-linear). In other words, the relationships between the age-disparate partnership and the predictors are measured by estimated probabilities using the logistic function. The logistic cumulative distribution function possesses a special functional form identified in terms of the exponential function, fitting the data to an S-shaped curve (sigmoid curve) instead of a straight line (Stock & Watson 2015).

$$p = \frac{1}{1+ e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k)}} \qquad (9)$$

where

$$0 < p < 1$$

As previously discussed in the data section, the effect of socioeconomic status measured through education level, employment status, and household income, on engagement in age-disparate partnerships might vary with age. One can assume that a female aged 15 may possess different experiences and opportunities compared to a female aged 24 regarding the socioeconomic measures. For example, a 24-year-old may have had more time to gain educational experience than a 15-year-old. Hence, to better understand the relationship between socioeconomic status and age-disparate partnerships, the analysis also consists of three logistic regressions run with a two-way interaction term between each socioeconomic indicator and age. The interaction terms are individually incorporated into the estimated model presented in equation (8), creating the following equations:

Equation (10), Two-way interaction term between education level & age $(X_1 * X_4)$:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_1 * X_4 \quad (10)$$

Equation (11), Two-way interaction term between employment status & age $(X_2 * X_4)$:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_2 * X_4 \quad (11)$$

Equation (12), Two-way interaction term between household income & age $(X_3 * X_4)$:

$$log(\frac{p}{l-p}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_2 * X_4 \quad (12)$$

One disadvantage of the logistic regression is that the coefficients measured in the log-odds scale are difficult to interpret. Hence, within the scope of this thesis, marginal effects and predicted probabilities of the logistic regression are used to interpret the model. It is also mentioned by Stock & Watson (2015) that predicted probabilities and differences in predicted probabilities are the best ways to interpret the logistic regression model. To clarify, this study reports the marginal effect at means, which models the differences in probabilities of engaging in age-disparate partnerships for one category relative to a reference group, keeping all other covariates at their means. The results of the marginal effects at means were compared with average marginal effects in order to check for the robustness of the findings. The results are very similar, and therefore not reported.

6. Results

This section presents the main findings of the study. The result of the main regression is estimated using Equation (8) and the possible interaction effects are estimated by estimating Equation (10) to (12).

6.1 Effects and predictions of the main regression

Table 2 shows marginal effects and predicted probabilities of the main regression that were estimated using Equation (8). In addition, Table 2 includes relative effects, which in this study is defined as the size of the marginal effect relative to the mean of age-disparate partnering.

Table 2. Predictions and effects of the main regression

Variable	Marginal effect	Relative effect	Predicted probability
Education level			
No schooling	0.148** (0.07)	0.513	0.404*** (0.07)
Primary	0.0768 (0.08)	0.266	0.333*** (0.08)
Incomplete secondary	0.0371 (0.02)	0.128	0.293*** (0.02)
Completed secondary	*(ref)*		0.256*** (0.02)
Tertiary	-0.0735** (0.04)	-0.254	0.182*** (0.03)
Employment status			
Student	-0.0893*** (0.02)	-0.309	0.222*** (0.02)
Employed	-0.00423 (0.03)	-0.015	0.307*** (0.03)
Unemployed	*(ref)*		0.311*** (0.02)
Other status	-0.00961 (0.07)	-0.033	0.302*** (0.07)
Household income			
No income	0.0277 (0.03)	0.096	0.287*** (0.03)
R1-R500	0.0265 (0.02)	0.092	0.286*** (0.02)
R501-R2,500	*(ref)*		0.259*** (0.02)
> R2,500	-0.00572 (0.05)	-0.020	0.254*** (0.05)
Age			
15-16 years	*(ref)*		0.105*** (0.02)
17-18 years	0.0774** (0.03)	0.268	0.182*** (0.02)
19-20 years	0.198*** (0.03)	0.684	0.303*** (0.02)

21-22 years	0.230*** (0.03)	0.795	0.335*** (0.02)
23-24 years	0.280*** (0.04)	0.969	0.385*** (0.02)
Lives with adult relatives			
Yes	-0.0567** (0.02)	-0.196	0.254*** (0.01)
No	*(ref)*		0.311*** (0.02)
Observations	2,022		2,022

*Note: The marginal effect is the discrete change from the reference group. In this study, the relative effect is defined as the size of the marginal effect relative to the mean of age-disparate partnering. Mean of age-disparate partnering is 0.289 (Table 1). Standard errors in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1. Coefficients of the model are found in Table A2 in the appendix.*

When looking at females' educational level in Table 2, an inverse relationship is identified between young females' educational level and their engagement in age-disparate partnerships. In other words, as the level of education increased, the probability of age-disparate partnering decreased. Females reporting no schooling experienced 15 percentage points ($p < 0.05$) higher probability to engage in age-disparate partnerships than the reference group (females reporting completed secondary school); and since the mean of engagement in age-disparate partnerships is 0.29 (Table 1), females reporting no schooling experienced 51% higher relative probability than females reporting completed secondary school. Furthermore, females reporting tertiary education had 7 percentage points ($p < 0.05$) lower probability to engage in age-disparate partnerships than the reference group (females reporting completed secondary school), implying that females reporting tertiary education experienced 25% lower relative probability than females who reports completed secondary school.

In contrast to education level, the females' employment status shows less difference. However, the results reveal lower probability to engage in age-disparate partnerships for female students, compared with females who were employed, unemployed or having other employment statuses. Students experienced 9 percentage points ($p < 0.01$) lower probability to engage in age-disparate partnerships than the reference group (unemployed females). This implies that the relative probability for students to engage in an age-disparate partnership is 31% lower than for unemployed females. Furthermore, engaging in age-disparate partnerships were almost similar among females who were employed, unemployed or having other employment statuses.

Furthermore, Table 2 shows an inverse relationship between females' monthly household income and the probability of engaging in age-disparate partnering. However, the relative effect is as good as absent between the no-income and the low-income level: the predicted margin was 0.12 percentage points lower between females with no household income and females with the income level R1-R500, which implies a 0.41% decreased probability of age-disparate partnering between the lowest income levels. Despite this similarity, the relative effects are systematic; the relative probability of age-disparate partnering for females with no household income was 10% higher than the probability of the reference group (females within income level R501-R2,500); females within the income level R1-R500 experienced 9% higher probability than the reference group; and females with a household income greater than R2,500 experienced a 2% lower probability than the reference group.

A strong, positive relationship between females' age and their probability of age-disparate partnering can be seen in Table 2. This relationship is crucial in the decision to further look at the interaction effects between age and the socioeconomic indicators, in purpose to distinguish the effects of the indicators depending on age. Furthermore, females who lived with adult relatives experienced 6 percentage points ($p < 0.05$) lower probability to engage in age-disparate partnerships than females who did not. This implies that females who lived with adult relatives experienced a 20% lower relative probability than females who did not.

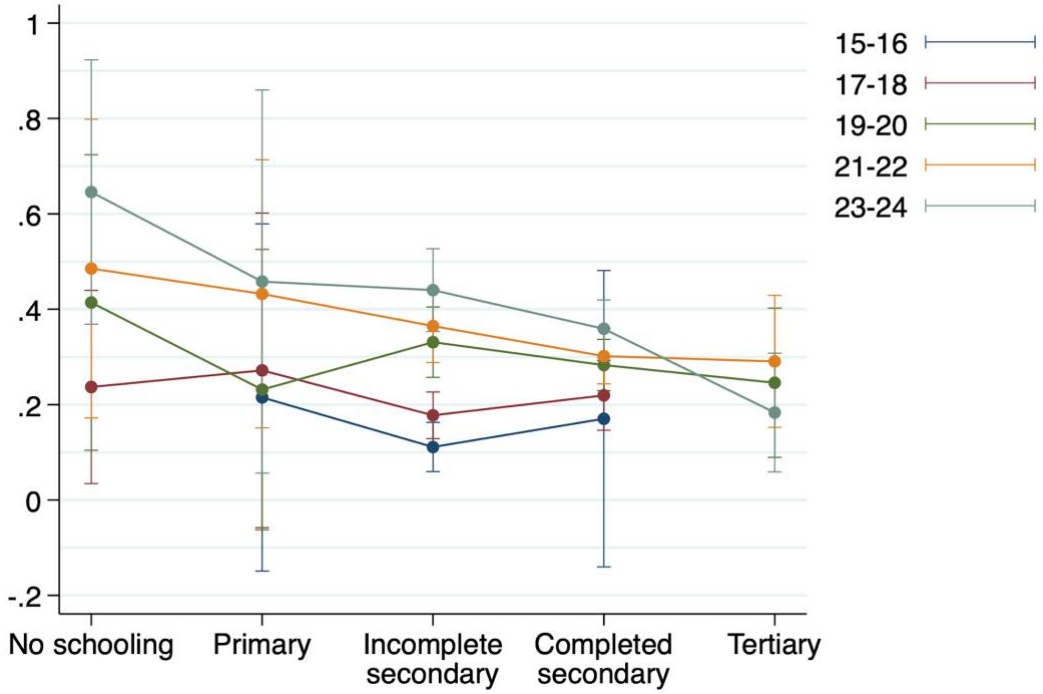
6.2 Interaction effects of socioeconomic status indicators and age

The following section presents the results of three regressions run with two-way interaction terms between age and the socioeconomic indicators presented in the following order; interaction term between education levels and age; interaction term between employment status and age; interaction term between household income and age. The graphs are based on Equations (10) to (12) described in the method section.

6.2.1 Interactions of education level and age

As previously stated, an inverse relationship between females' education level and age-disparate partnering could be seen. But it is important to have in mind, the different lifestyles depending on age, and how the effects of socioeconomic status indicators depend on age. Figure 3 shows the predicted probabilities of education level and age; and reports that for a given educational level, the probability of females entering into age-disparate partnerships is larger the older they are. Figure 3 is formed by the predicted probabilities in Table A3, which is found in appendix.

Figure 3. Interaction effects of education level and age



Note: This figure shows the interaction of age and education level using Equation 10, with 95% CIs. The y-axis measures predicted probability of age-disparate partnering, and the x-axis shows levels of education. The figure is formed by the predictions in Table A3 and can be found in the appendix.

Furthermore, females with no schooling experienced the largest difference in age-disparate partnering between the ages. The marginal effect of having no schooling was 41 percentage points higher ($p < 0.05$) for females aged 23-24 than for females aged 17-18, which is an

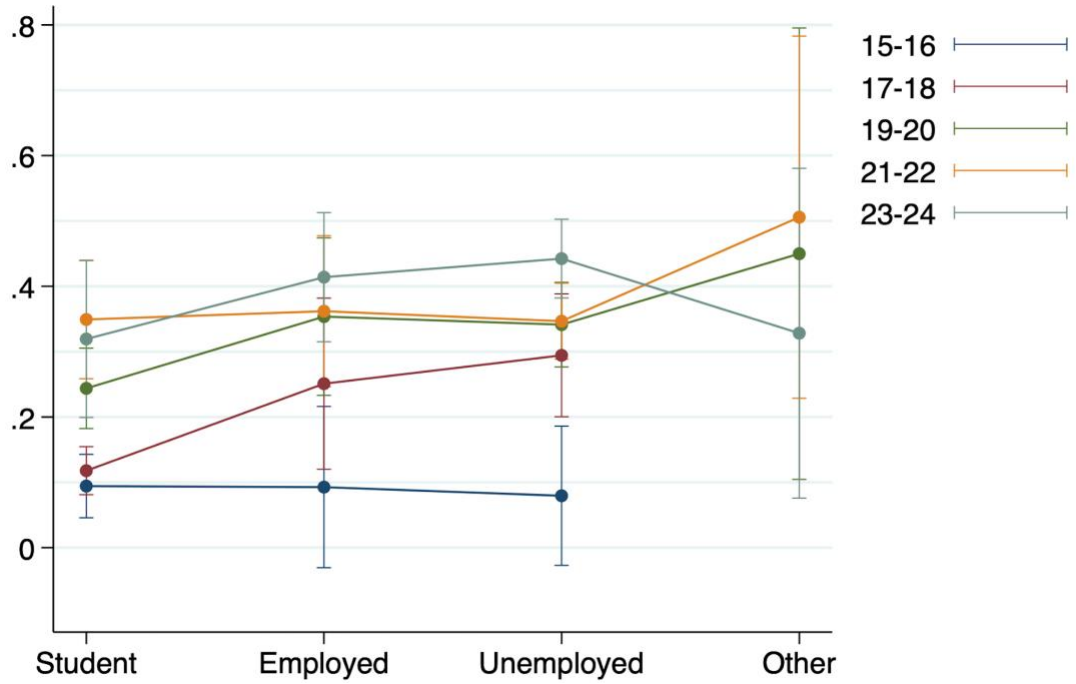
increase of 141% between 17-18 years old females and 23-24 years old females since the mean of age-disparate partnering is 0.29.

Moreover, females with tertiary education had the least difference. These differences imply that education affects the probability of age-disparate partnering more negatively the older the females are. For example, the effect of having tertiary education was 22 percentage points lower ($p < 0.01$) for females aged 23-24 than for females aged 19-20 which implies a 22% relative decrease between 19-20 years old females and 23-24 years old females.

6.2.2 Interactions of employment status and age

Figure 4 shows interaction effects of females' age and employment status, and reports as well as for education, that for a given employment status, the probability of females entering into age-disparate partnerships is larger the older they are. The figure is based on the predicted probabilities in Table A4, which is found in appendix.

Figure 4. Interaction effects of employment status and age



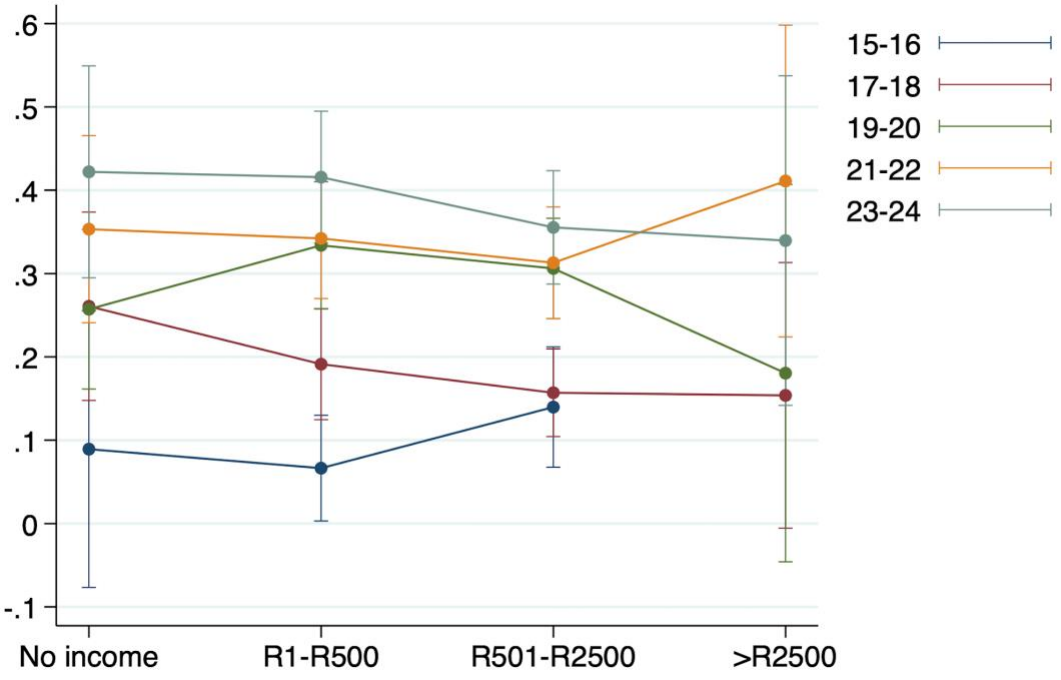
Note: This figure shows the interaction of age and employment status using Equation 11, with 95% CIs. The y-axis measures predicted probability of age-disparate partnering, and the x-axis shows employment status categories. The figure is formed by the predictions in Table A4 and can be found in the appendix.

The figure shows the largest difference among unemployed females who could not find any work. Though, the results of females aged 15-16 were not significant for unemployed and employed females (Table A4) because of few observations.

6.2.1 Interactions of household income and age

As explained earlier, an inverse relationship between females' household income and age-disparate partnering could be seen. Figure 5 shows the predicted probabilities of household levels and age, and similar to the previous socioeconomic indicators, the probability of age-disparate partnering is larger for females the older they are. The figure is formed by the predicted probabilities in Table A5 and is found in the appendix.

Figure 5. Interaction effects of age and household income



Note: This figure shows the interaction of age and modified household income levels using Equation 12, with 95% CIs. The y-axis measures predicted probability of age-disparate partnering, and the x-axis shows levels of household income. The figure is formed by the predictions in Table A5 and can be found in the appendix.

Largest differences of age-disparate partnering between the ages were found among females with no household income and females with a household income of R1 – R500. The marginal effect of having the income level R1 – R500 was 35 percentage points higher ($p < 0.05$) for females aged 23-24 than for females aged 15-16, which is an increase of 121% between 15-16 years old females and 23-24 years old females since the mean of age-disparate partnering is 0.29. Moreover, females with an income level of R501 – R2500 had least difference between the ages: the effect was 22 percentage points lower ($p <0.01$) for females aged 23-24 than for females aged 15-16 and implies a 89% relative decrease between females aged 15-16 and 23-24.

7. Discussion

This study investigates the relationship between age-disparate partnering and socioeconomic status for adolescent girls and young women aged 15-24 in KwaZulu-Natal. The findings suggest an association between socioeconomic status and age-disparate partnerships. The findings on the socioeconomic indicators of education and household income are in line with the hypothesis on how low socioeconomic status increases females' probability to engage in age-disparate partnerships.

The results suggest a significant negative correlation between females' education level and age-disparate partnering; as the level of education increases, the probability of age-disparate partnering decreases. As the mean of age-disparate partnering is 0.29, females with no schooling experienced 51% higher probability of age-disparate partnering than females who reported completed secondary school as highest education qualification; while females with tertiary education experienced 25% lower probability to age-disparate partnering than females who had completed secondary school. The employment categories of employed females and unemployed females barely differ in the predictions of engaging in age-disparate partnerships, as well as their effects are insignificant. However, females who were students had 31% lower probability to engage in age-disparate partnering than females who were unemployed. Similar to education, an inverse pattern is observed between monthly household income and age-disparate partnerships, although a moderate and non-significant pattern. For all socioeconomic measures, the interactions show larger probability for females to engage in age-disparate partnerships as older the females are. This means that the probability for age-disparate partnerships increase with age for all socioeconomic indicators.

The control variables provide significant information of who the females engaging in age-disparate partnerships are. The evidence suggests a strong, positive relationship between females' age and their probability to engage in age-disparate partnering. Furthermore, females who live with adult relatives experience 20% lower relative probability engage in age-disparate partnerships than females who does not.

The findings on the inverse relationship between education and age-disparate partnerships can be discussed further in relation to the key concept on how education increase the amount

of healthy days within Grossman´s health capital model explained by Folland, Goodman & Stano (2013). In the context of the model, age-disparate partnerships can be interpreted as a bad health investment due to increased HIV risk (Maughan-Brown et al. 2018; Stoner et al. 2019). The fact that education decrease the probability of age-disparate partnerships can therefore be seen as a decrease in the probability of making a bad health investment. Hence, not engaging in age-disparate partnerships may lead to a higher amount of healthy days, increased productivity, higher income as well as higher utility than if engaging in age-disparate partnerships.

The negative correlation identified for both education (significant) and household income (non-significant) can be discussed from the perspective of the discounted utility model described by White & Dow (2015). One can assume that education and household income may affect how the females in the sample perceive their future and therefore determinate their individual time preferences, hence, affect the tradeoff between utility in the present versus the future. The individual perspective on the future can determine the females' risk behavior and lead to different engagement in age-disparate partnerships. Further, the choice of engaging in age-disparate partnerships can be seen as an uncertain choice, to which the probability of attracting HIV versus not attracting HIV has to be taken into account. The utility from engaging in an age-disparate partnership may, despite the increased HIV risk, be higher relative to the utility of not engaging in these partnerships. In that sense engaging in age-disparate partnerships could be seen as a choice in line with utility maximization even when accounting for the increased HIV risk.

Even though the findings on education and household income can be incorporated into the theories mentioned above, it has to be acknowledged that both the health capital model and the discounted utility model assumes full knowledge of the health risk that comes with a decision (Folland, Goodman & Stano 2013; White & Dow 2015). It is important to discuss that it is not certain the young females in KwaZulu-Natal are aware of the increased HIV risk age-disparate partnerships supposedly expose them to. Hence, they may not be aware of the fact that age-disparate partnering could be a decision that decrease their amount of healthy days.

The evidence on the inverse relationship between education level and age-disparate partnerships is supported by earlier research from Stoner et al. (2017), who suggests that

lower school attendance and dropping out of school increase the likelihood of age-disparate partnerships. The significant inverse relationship between education and age-disparate partnerships, assuming that age-disparate partnerships lead to increased HIV risk, coincides with Hargreaves et al. (2007), who implies a correlation between low education level and increased HIV incidence. However, the findings on education with the assumption on increased HIV risk due to age-disparate partnerships, do not coincide with Bunyasi & Coetzee (2017) and their evidence in Free State, suggesting that education barely has any effect on HIV risk for young females.

The non-significant finding on the similar probabilities for employed and unemployed young females to engage in age-disparate partnerships, assuming that age-disparate partnerships increase HIV risk, is not supported by Austin, Choi & Berndt (2017); who suggests that unemployment leads to economic insecurity which is argued to increase vulnerability for females to engage in sexual relations that can affect their health. However, their evidence on economic insecurity supports the non-significant inverse relationship between income and age-disparate partnerships.

In terms of looking at engaging in age-disparate partnerships as risky behavior, it can be suggested that the non-significant inverse relationship between income and age-disparate partnerships is supported by Hargreaves et al. (2007), suggesting a higher amount of risky behavior within the group of low-income women. Assuming that age-disparate partnerships increase HIV risk for young females, the inverse pattern between income and age-disparate partnership is supported by Bunyasi and Coetzee (2017), who suggests a negative association between relative wealth and HIV incidence for females. However, the relationship between income and age-disparate partnerships, assuming increased HIV risk, is not supported by earlier findings from Bärnighausen, Hosegood, Timaeus & Newell (2007); who suggests that middle-income households in terms of relative wealth experience the highest HIV risk.
It is important to emphasize that the findings of this thesis show correlation and not causation. However, the findings of this study can be discussed from the vital evidence brought by Durevall, Lindskog & George (2019), on how differences between females who drop out of school and the ones who stay in school stand for some of the negative correlation between education and HIV incidence for young females. Furthermore, the findings in this study suggest that the girls who engage in age-disparate partnerships are different from those who do not. The young females who engage in age-disparate partnerships are in general

older, possess lower levels of education (significant), experience lower levels of household income (non-significant), and are, to a lesser extent, students than the young females who do not engage in age-disparate partnerships.

This study encourages preventative actions with a focus on increased educational attainments for young females in Kwazulu-Natal, in order to stop the transmission cycle fueling the epidemic, all towards ending the AIDS epidemic by 2030. The foregoing statement is based on the findings on education deriving from this study as well as the fact that education can be perceived as a policy variable. Governmental policies with a focus on increasing young females' educational attainments can in accordance with our findings decrease young females' participation in the risk behavior of age-disparate partnerships. Governmental policies and preventative actions focusing on decreasing young females' engagement in age-disparate partnerships through increased household income, may not be as effective based on the insignificant findings on household income in this study, as well as income policies might be more difficult to execute. However, the inverse pattern between income and age-disparate partnerships is not to be neglected. Furthermore, governmental policies with a focus on increasing education for young females may lead to a higher household income, as in accordance with the health capital model; higher levels of education leads to a higher demand of healthy days (higher health capital stock) which can increase productivity and income (Folland, Goodman & Stano (2013)

The small sample size of 2 224 individuals can create limitations concerning the reliability of the findings and may generate bias. The small sample can decrease the study's statistical power, and we may not be able to capture the actual effects of the socioeconomic indicators. The fact that none of the categories within the household income variable are significant limits the understanding of the effect income has on age-disparate partnerships. Although, it should be acknowledged that the non-significance can be due to mismeasurement in the creation of the variable. The fact that the age group of 15-16 almost is twice as small as the rest of the age groups can be a limitation for the reliability of the findings, especially for the results derived from the regressions where age is used in an interaction term and when comparing probabilities between age groups. Engaging in an age-disparate partnership may be determined by other factors that possess a correlation with the variables within the estimated model and can therefore create omitted variable bias, give misleading results and limit the findings. Consequently, a limitation to the findings is that few control variables are

incorporated into the model, which can result in this study not being able to capture the actual effects of the socioeconomic measures on age-disparate partnership. For example, knowledge about the possible increased HIV risk, levels of emotional or financial support, or experience of engaging in a sexual partnership could be factors that might have an effect on the probability of engaging in age-disparate partnerships.

8. Conclusion

This study aims to investigate the relationship between socioeconomic status and age-disparate partnerships for the vulnerable group of adolescent girls and young women aged 15-24, in the high prevalence setting of KwaZulu-Natal. The findings imply an association between socioeconomic status and age-disparate partnerships.

The results suggest a significant negative correlation between education level and age-disparate partnership; as the level of education increase, the probability of age-disparate partnering decrease. The employment categories of employed and unemployed do barely differ in regard to predicted probability of engaging in age-disparate partnerships as well as their effects are insignificant. However, females who are students possess a significant lower relative probability to engage in age-disparate partnerships than females who are unemployed. The same inverse pattern as mentioned for education is observed between household income and age-disparate partnerships, although a moderate and non-significant pattern. The probability for age-disparate partnership increase with age for all socioeconomic measures.

This study is a contribution to a greater understanding of the vulnerable group of adolescent girls and young women within the transmission cycle, fueling the epidemic in KwaZulu-Natal. Further research is needed to understand the group of adolescent girls and young women within this context in order to terminate the epidemic by 2030. Research should focus on the causal effects of the associations suggested in this study, which can be essential to put a stop to the transmission cycle. Further research is also suggested comparing rural and urban areas in the context of young females and age-disparate partnerships. Such research can be important in order to locate the transmission pattern further for more effective preventative actions. As this study focuses on one group within the transmission cycle, future research pinpointing all three groups participating in the transmission network may be crucial to put a proper stop to the cycle fueling the HIV epidemic in the area.

9. References

Austin, K., Choi, M., & Berndt, V. (2017). Trading sex for security: Unemployment and the unequal HIV burden among young women in developing nations. *SAGE Journals*: February 1, Volume: 32 Issue: 3, p 343-368. https://doi.org/10.1177/0268580917693172 [Accessed 2020-06-06]

Brougham, R & John, R. (2007). *Motivation of Health Behavior.* August, p 49-50. Editor: Paul W. O'Neal. New York: Nova Science Publishers Inc. https://books.google.se/books?redir_esc=y&hl=sv&id=E2twrZWJWhMC&q=Brougham%2C+R+%26+John%2C+R.#v=onepage&q&f=false [Accessed 2020-05-18]

Bunyasi EW & Coetzee DJ. (2017) Relationship between socioeconomic status and HIV infection: findings from a survey in the Free State and Western Cape Provinces of South Africa. *BMJ Journals:* November 20. doi: 10.1136/bmjopen-2017-016232 [Accessed 2020-06-05]

Bärnighausen, T., Hosegood, V., Timaeus, I & Newell, M-L. (2007). The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. *AIDS*: November, volume 21, p s29-238. doi: 10.1097/01.aids.0000300533.59483.95 [Accessed 2020-05-14]

CAPRISA. (2017). *HIPSS, HIV Incidence Provincial surveillance system project, Report on baseline findings.* May. South Africa; CAPRISA. [Accessed 2020-04-20]

Cawley, J & Rhum, C,J. (2011). Handbook of Health Economics; Chapter Three -The Economics of Risky Health Behaviour, volume 2, pages 95-199. https://doi.org/10.1016/B978-0-444-53592-4.00003-7 [Accessed 2020-05-15]

De Janvry, A. & Sadoulet, E. (2016). *Development Economics, Theory and practice*. New York: Taylor and Francis.

de Oliveira et al. (2017). Transmission networks and risk of HIV infection in KwaZulu-Natal, South Africa: a community-wide phylogenetic study. *The lancet. HIV*, January 1, Volume 4, Issue 1, p41-50. https://doi.org/10.1016/S2352-3018(16)30186-2. [Accessed 2020-04-15]

Durevall, D., Lindskog, A & George, G. (2019). Education and HIV incidence among young women in KwaZulu-Natal: An association but no evidence of a causal protective effect. PLOS ONE, March 4. https://doi.org/10.1371/journal.pone.0213056 [Accessed 2020-05-19].

Follan, S., Goodman, A & Stano, M. (2013). The Economics of Health and Health Care. 7th edition. P. 129-139. New Jersey:PEARSON. http://www.ssu.ac.ir/cms/fileadmin/user_upload/Daneshkadaha/dbehdasht/modiryat_khatamat_behdashti/kotob_farsi/The_economics_of_health_and_health_care-_Sherman_Folland.pdf[Accessed 2020-05-21]

Hargreaves, J., Bonnel, C., Morison, L., Kim, J., Phetla, G., Porter, J., Watts, C., Pronyk, P.(2007). Explaining continued high HIV prevalence in South Africa: socioeconomic factors, HIV incidence and sexual behaviour change among a rural cohort, 2001–2004. *AIDS*: November, Volume 21, Issue, p s39-s48. doi: 10.1097/01.aids.0000300534.97601.d6 doi: [Accessed 2020-05-14]

Harling, G., Newell, M.-L., Tanser, F., Kawachi, I., Subramanian, S.-V., Bärnighausen, T. (2014). Do Age-Disparate Relationships Derive HIV Incidence In Young Women? Evidence from a Population Cohort in Rural KwaZulu-Natal, South Africa. *JAIDS Journal of Acquired Immune Deficiency Syndromes*: August, Volume 66, issue 4, p 443-451. doi: 10.1097/QAI.0000000000000198 [Accessed 2020-06-05]

HSRC (2019). *South African National HIV Prevalence, Incidence, Behaviour and Communication Survey, 2017*. Cape Town, South Africa: Human Sciences Research Council. http://www.hsrc.ac.za/uploads/pageContent/10779/SABSSM%20V.pdf [Accessed 2020-04-26]

Kharsany et al. (2015). Strengthening HIV surveillance in the antiretroviral therapy era: rationale and design of a longitudinal study to monitor HIV prevalence and incidence in the uMgungundlovu District, KwaZulu-Natal, South Africa. *BMC Public Health* 15, 20

November, Article number: 1149. https://doi.org/10.1186/s12889-015-2179-2 [Accessed 2020-05-01]

Maughan-Brown, B., George, G., Beckett, S., Evans, M., Lewis, L., Cawood, C., Khanyile, D., & Kharsany, A. (2018). HIV Risk Among Adolescent Girls and Young Women in Age-Disparate Partnerships: Evidence From KwaZulu-Natal, South Africa. *JAIDS Journal of acquired immune deficiency syndrome*, June 1, Volume 78, Number 2, p 155-162.https://journals.lww.com/jaids/Fulltext/2018/06010/HIV_Risk_Among_Adolescent_Girls_and_Young_Women_in.4.aspx [Accessed 2020-04-10]

Maughan-Brown, B., Evans, M., & George, G. (2016). Sexual Behaviour of Men and Women within Age-Disparate Partnerships in South Africa: Implications for Young Women's HIV Risk. *PLOS ONE*, August 15.https://doi.org/10.1371/journal.pone.0159162 [Accessed 2020-04-18]

Stock, J. & Watson, M. (2015). *Introduction to Econometrics*. 3rd Edition. London: Pearson Education.

Stoner et al. (2019). Age-disparate partnerships and incident HIV infection in adolescent girls and young women in rural South Africa. *AIDS*: 02 January, Volume 33, Issue 1, p.83-91.doi: 10.1097/QAD.0000000000002037 [Accessed 2020-05-14]

Stoner et al. (2017). The effect of schooling on age-disparate relationships and number of sexual partners among young women in rural South Africa enrolled in HPTN 068. JAIDS *Journal of Acquired Immune Deficiency Syndromes*: December 15, Volume 76, Issue 5, pe107-e114.doi: 10.1097/QAI.0000000000001544 [Accessed 2020-05-14]

UNAIDS (2019). *UNAIDS DATA 2019*. Geneva, Switzerland: UNAIDS. https://www.unaids.org/sites/default/files/media_asset/2019-UNAIDS-data_en.pd [Accessed 2020-04-07]

UNAIDS (2017). *HIV Prevention 2020 Road Map - Accelerating HIV prevention to reduce new infections by 75%*. Geneva, Switzerland: UNAIDS.

https://www.unaids.org/sites/default/files/media_asset/hiv-prevention-2020-road-map_en.pdf [Accessed 2020-04-16]

UNAIDS (2016). *GLOBAL AIDS UPDATE 2016*. Geneva, Switzerland: UNAIDS. https://www.unaids.org/sites/default/files/media_asset/global-AIDS-update-2016_en.pdf [Accessed 2020-04-16]

UNAIDS (2015). UNAIDS TERMINOLOGY GUIDELINES. Geneva, Switzerland: UNAIDS. https://www.unaids.org/sites/default/files/media_asset/2015_terminology_guidelines_en.pdf [Accessed 2020-04-07]

UNAIDS (2015). *UNAIDS 2016–2021 Strategy - On the Fast-Track to end AIDS*. Geneva, Switzerland: UNAIDS. https://www.unaids.org/sites/default/files/media_asset/20151027_UNAIDS_PCB37_15_18_EN_rev1.pdf [Accessed 2020-04-17]

UNAIDS (2014). *90-90-90 An ambitious treatment target to help end the AIDS epidemic*.Geneva, Switzerland: UNAIDS. https://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf [Accessed 2020-04-07]

Vandormael, A., Akullian, A., Siedner, M., de Oliveira, T., Bärnighausen, T., Tanser, F., (2019). Declines in HIV incidence among men and women in a South African population-based cohort. *Nature communications* 10, 02 December, Article number: 5482 https://doi.org/10.1038/s41467-019-13473-y [Accessed 2020-04-19]

White, J & Dow, W. (2015). Behavioral Economics and Public Health. p.27-33. Editor: Roberto, C & Kawachi, I. New York: Oxford University Press Inc. [Accessed 2020-05-23]

Appendix

Table A1. Robust, linear regressions

Variable	Model 1a	Model 1b	Model 2a	Model 2b	Model 3a	Model 3b	Model 4a	Model 4b
Education level								
No schooling	0.139** (0.07)	0.143** (0.07)					0.139** (0.07)	0.141** (0.07)
Primary	0.0829 (0.07)	0.0710 (0.07)					0.0889 (0.08)	0.0820 (0.08)
Incomplete secondary	0.0375* (0.02)	0.0346 (0.02)					0.0426* (0.02)	0.0402* (0.02)
Completed secondary	*(ref)*	*(ref)*					*(ref)*	*(ref)*
Tertiary	-0.122*** (0.04)	-0.114*** (0.04)					-0.0902** (0.04)	-0.0856** (0.04)
Employment status								
Student			-0.104*** (0.02)	-0.0967*** (0.02)			-0.0914*** (0.03)	-0.0856*** (0.03)
Employed			-0.0197 (0.03)	-0.0288 (0.03)			0.00146 (0.03)	-0.00802 (0.03)
Unemployed			*(ref)*	*(ref)*			*(ref)*	*(ref)*
Other			-0.0281 (0.06)	-0.0536 (0.06)			0.0105 (0.07)	-0.00811 (0.07)
Household income								
No income					0.0494 (0.03)	0.0226 (0.03)	0.0474 (0.03)	0.0262 (0.03)
R1 – R500					0.0391* (0.02)	0.0400* (0.02)	0.0237 (0.02)	0.0244 (0.02)
R500 – R2,500					(ref)	(ref)	(ref)	(ref)
> R2,501					-0.0218 (0.05)	-0.0247 (0.05)	-0.00157 (0.05)	-0.00431 (0.05)
Age								
15 – 16 years	*(ref)*	*(ref)*	*(ref)*	*(ref)*	*(ref)*	*(ref)*	*(ref)*	*(ref)*
17 – 18 years	0.0952*** (0.03)	0.0843*** (0.03)	0.0753*** (0.03)	0.0668*** (0.03)	0.0749*** (0.03)	0.0675** (0.03)	0.0788*** (0.03)	0.0725** (0.03)
19 – 20 years	0.240*** (0.03)	0.221*** (0.03)	0.173*** (0.03)	0.160*** (0.03)	0.202*** (0.03)	0.187*** (0.03)	0.201*** (0.03)	0.190*** (0.03)
21 – 22 years	0.304*** (0.03)	0.276*** (0.03)	0.216*** (0.03)	0.198*** (0.03)	0.255*** (0.03)	0.233*** (0.03)	0.240*** (0.04)	0.225*** (0.04)
23 – 24 years	0.365*** (0.03)	0.336*** (0.03)	0.269*** (0.03)	0.251*** (0.03)	0.316*** (0.03)	0.291*** (0.03)	0.293*** (0.04)	0.277*** (0.04)
Lives with adult relatives								
Yes		-0.0769*** (0.02)		-0.0722*** (0.02)		-0.0771*** (0.02)		-0.0584** (0.02)
No		*(ref)*		*(ref)*		*(ref)*		*(ref)*
Constant	0.0471 (0.03)	0.119*** (0.04)	0.169*** (0.03)	0.229*** (0.03)	0.0795*** (0.02)	0.151*** (0.03)	0.110*** (0.04)	0.164*** (0.04)
Observations	2,222	2,222	2,219	2,219	2,024	2,024	2,022	2,022
R-squared	0.065	0.071	0.066	0.071	0.056	0.061	0.071	0.074

Note: Model 1a: linear regression with age-disparate partnership and education, with age as control variable. Model 1b: as Model 1a but with the control variable Lives With Adult Relatives included. Model 2a: linear regression with age-disparate relationship and employment status, with age control. Model 2b: as Model 2a but with the control variable Lives With Adult Relatives included. Model 3a: linear regression with age-disparate partnership and household income, with Age control. Model 3b: as Model but with the control variable Lives With Adult Relatives included. Model 4a: Multiple linear regression with age-disparate partnership, education level, employment status and household income, with age control. Model 4b: as Model 4a but with the control variable Lives With Adult Relatives included. Standard errors in parentheses.
**** p < 0.01, ** p < 0.05, * p < 0.1.*

Table A2. Logistic regression of the estimated model

Variable	Coefficient	Marginal effect	Predicted probability
Education level			
No schooling	0.679** (0.31)	0.148** (0.07)	0.404*** (0.07)
Primary	0.371 (0.36)	0.0768 (0.08)	0.333*** (0.08)
Incomplete secondary	0.187 (0.12)	0.0371 (0.02)	0.293*** (0.02)
Completed secondary	*(ref)*	*(ref)*	0.256*** (0.02)
Tertiary	-0.433* (0.24)	-0.0735** (0.04)	0.182*** (0.03)
Employment status			
Student	-0.460*** (0.13)	-0.0893*** (0.02)	0.222*** (0.02)
Employed	-0.0198 (0.15)	-0.00423 (0.03)	0.307*** (0.03)
Unemployed	*(ref)*	*(ref)*	0.311*** (0.02)
Other status	-0.0452 (0.34)	-0.00961 (0.07)	0.302*** (0.07)
Household income			
No income	0.140 (0.16)	0.0277 (0.03)	0.287*** (0.03)
R1-R500	0.134 (0.12)	0.0265 (0.02)	0.286*** (0.02)
R501-R2,500	*(ref)*	*(ref)*	0.259*** (0.02)
> R2,500	-0.0300 (0.26)	-0.00572 (0.05)	0.254*** (0.05)
Age			
15-16 years	*(ref)*	*(ref)*	0.105*** (0.02)
17-18 years	0.643** (0.28)	0.0774** (0.03)	0.182*** (0.02)
19-20 years	1.308*** (0.28)	0.198*** (0.03)	0.303*** (0.02)
21-22 years	1.457*** (0.28)	0.230*** (0.03)	0.335*** (0.02)
23-24 years	1.676*** (0.29)	0.280*** (0.04)	0.385*** (0.02)
Lives with adult relatives			
Yes	-0.280** (0.11)	-0.0567** (0.02)	0.254*** (0.01)
No	*(ref)*	*(ref)*	0.311*** (0.02)
Observations	2,022	2,022	2,022

Note: This table shows the results of the estimated model in equation (8) presented in the method section. Describes the coefficients, marginal effects at means and predicted probabilities of the logistic regression. Age-disparate partnership is the dependent variable. Standard errors in parentheses.

**** p < 0.01, ** p < 0.05, * p < 0.1.*

Table A3. Predicted probabilities of interactions of education and age

Age	15 – 16 years	17 – 18 years	19 – 20 years	21 – 22 years	23 – 24 years
Education level					
No schooling	*(not estimable)*	0.237** (0.10)	0.414*** (0.16)	0.485*** (0.16)	0.646*** (0.14)
Primary	0.215 (0.19)	0.272 (0.17)	0.232 (0.15)	0.432*** (0.14)	0.458** (0.20)
Incomplete secondary	0.111*** (0.03)	0.178*** (0.02)	0.331*** (0.04)	0.365*** (0.04)	0.440*** (0.04)
Completed secondary	0.170 (0.16)	0.219*** (0.04)	0.283*** (0.03)	0.302*** (0.03)	0.359*** (0.03)
Tertiary	*(not estimable)*	*(not estimable)*	0.246*** (0.08)	0.291*** (0.07)	0.183*** (0.06)
Observations	2,017	2,017	2,017	2,017	2,017

Note: The predicted probabilities in this table forms the graph in Figure 3 presented in the results section. Results with too few observations are not estimable. Standard errors in parentheses.
**** p < 0.01, ** p < 0.05, * p < 0.1.*

Table A4. Predicted probabilities of interactions of employment status and age

Age	15 – 16 years	17 – 18 years	19 – 20 years	21 – 22 years	23 – 24 years
Employment status					
Student	0.0943*** (0.02)	0.118*** (0.02)	0.244*** (0.03)	0.349*** (0.05)	0.319*** (0.06)
Employed	0.0927 (0.06)	0.251*** (0.07)	0.354*** (0.06)	0.362*** (0.06)	0.414*** (0.05)
Unemployed	0.0795 (0.05)	0.294*** (0.05)	0.341*** (0.03)	0.347*** (0.03)	0.442*** (0.03)
Other	*(not estimable)*	*(not estimable)*	0.450** (0.18)	0.506*** (0.14)	0.328** (0.13)
Observations	2,015	2,015	2,015	2,015	2,015

Note: The predicted probabilities in this table forms the graph in Figure 4 presented in the results section. Results with too few observations are not estimable. Standard errors in parentheses.
**** p < 0.01, ** p < 0.05, * p < 0.1.*

Table A5. Predicted probabilities of interactions of household income and age

	15 – 16 years	17 – 18 years	19 – 20 years	21 – 22 years	23 – 24 years
Household income					
No income	0.0893 (0.08)	0.261*** (0.06)	0.257*** (0.05)	0.353*** (0.06)	0.422*** (0.06)
R1-R500	0.0665** (0.03)	0.191*** (0.03)	0.334*** (0.04)	0.342*** (0.04)	0.416*** (0.04)
R501-R2500	0.140*** (0.04)	0.157*** (0.03)	0.306*** (0.03)	0.313*** (0.03)	0.355*** (0.03)
>R2500	*(not estimable)*	0.154* (0.08)	0.180 (0.12)	0.411*** (0.10)	0.340*** (0.10)
Observations	2,014	2,014	2,014	2,014	2,014

Note: The predicted probabilities in this table forms the graph in Figure 5 presented in the results section. Results with too few observations are not estimable. Standard errors in parentheses.
**** p < 0.01, ** p < 0.05, * p < 0.1.*