Age-Period-Cohort and Educational Attainment Effects on HIV Prevalence in Zambian Pregnant Women, 1994 through 2011

By

Webster Kasongo

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Epidemiology

August, 2013

Nashville, Tennessee

Approved

Sten H. Vermund, MD, PhD

Bryan Shepherd, PhD

Mary Lou Lindegren, MD, MPH

Marie Griffin, MD, MPH

DEDICATION

To my late father, Robert Chishala Kasongo, my mother, Luisa Mwape Kasongo, and my beloved wife Ireen, and my sons, Terry and Timothy, and my brothers and sister.

ACKNOWLEDGEMENT

My PhD training in Epidemiology would not have been possible without the financial support from Dr. Sten Vermund's AIDS International Training and Research Program grant number D43 TW001035-15. I am especially indebted to Dr.Sten Vermund for hosting me at the Vanderbilt Institute for Global Health (VIGH) and providing financial and academic support as well as his mentorship. I am grateful to VIGH management team and Holly Cassel for support during the training, and support and understanding when my father and my brother passed on, events that were so overpowering, and could have derailed my training.

Dr.Vermund, my academic mentor and chair of my dissertation committee was very instrumental during my didactic course training, proposal writing and final dissertation writing. I thank him for his time and invaluable support provided to me, both directly and via his contacts. I could not have asked for a better dissertation committee than I had, and benefited greatly from the acute feedback each one of them provided: Dr. Bryan Shepherd, Dr. Mary Lou Lindegren and Dr. Marie Griffin, and I' am grateful. I thank Meridith Blevins for her encouragement and support in learning R computing and programming language, and also for her acute insight, and Dr. Doug Heimburger for his supervision during my stay at VIGH when Dr.Vermund was away from the VIGH. I am grateful to Jennifer Lynn St Clair for always finding time for me to meet with Dr. Vermund, and to Anne Neubecker, Megan Johnson Pask, and Clay Wilson for their various supports.

Further, I thank Dr. Katherine Hartmann and her team for summoning the courage to start a new Epidemiology PhD Program at Vanderbilt, and giving me the opportunity to be in the first group. I am grateful to Spencer Toye the Program Manager for Graduate studies in Epidemiology for her coordination of program.

I am appreciative of my late father's care and support in my earlier training in Zambia, and of my mother for braving my absence from Zambia. My wife, Ireen, and my two sons Terry and Timothy provided inspirational support and I am deeply indebted. I am also grateful to all the pregnant women from whom the data were collected. Last but not the least, I am grateful to Dr.Modest Mulenga and Dr.Chanda Mulenga from Tropical Diseases Research Centre management for their support, and the Zambian Ministry of Health for granting permission to antennal attendees HIV sentinel surveillance data for my work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIDS	Acquired immunodeficiency syndrome
ANC	Antenatal Care
ANC-HIV-SS	ANC-based HIV sentinel surveillance
APC	Age-period-Cohort
BED-CEIA	BED capture enzyme immune assay (i.e., B,E,D are HIV subtypes)
cART	Combinational antiretroviral therapy
CCREM	Cross-Classified Random Effect Model
CDC	Centers for Diseases Control and Prevention
CI	Confidence Interval
CL	Confidence Level
CRF	Circulating recombinant form
DAG	Directed Acyclic Graph
DHS	Demographic Health Survey
ELISA	Enzyme Linked Immunosorbent Assay
EMBASE	Excerpta Medica Database
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
GRZ	Government of Republic of Zambia
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HIV-1&2	HIV strain type-1 & HIV strain type-2
ICC	Intra-class correlation
ID	Identify Number
IQR	Interquartile Range
IRB	Institutional Review Board

LMIC	Low and Middle Income Countries
LRT	Likelihood Ratio Test
MeSH	Medical Sub-Heading
MSM	Men who have sex with men
NIHHSTP	National Center for HIV/AIDS Viral Hepatitis and Tuberculosis Prevention
OR	Odds Ratio
PBS	Population-based Survey
PICOT	Population, Intervention, Comparator, Outcome and Time-frame
PMTCT	Prevention of mother to child HIV transmission
PUBMED	Publication Medicals
RCS	Restricted Cubic Splines
RE	Random Effect
SSA	sub-Saharan Africa
STIs	Sexually Transmitted Infections
TDRC	Tropical Diseases Research Centre
UNAIDS	United Nations Program on AIDS
USA	United States of America
UTH	University Teaching Hospital
VIGH	Vanderbilt Institute for Global Health
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

1.1. Human immunodeficiency virus (HIV) infections burden among pregnant women in sub-Saharan Africa

HIV burden among pregnant women is a critical public health concern globally, but most profound in sub-Saharan Africa (SSA). An estimated 1.5 million pregnant women were living with HIV in 2011.[1] Without effective interventions for prevention of mother to child HIV transmission (PMTCT), nearly half of HIV infected pregnant women in SSA are likely to pass HIV to their babies during pregnancy, childbirth and after birth through breastfeeding.[2-7] For example, an estimated 370,000 new HIV infections occurred in children in 2009, largely through MTCT, and mostly in poorly resourced settings.[8, 9]

Sub-Saharan Africa (SSA) comprises 48 of 54 countries in Africa, and is home to approximately 900 million people, 13% of the global human population but in 2011 accounted for 69% of the 34 million people living with the human immunodeficiency virus (HIV), the virus that causes acquired immunodeficiency virus (AIDS).[10, 11] Additionally, the estimated number of people newly infected with HIV in 2011 was fewer by 700,000, compared to number of newly HIV infected persons in 2001, suggesting a net decline in global HIV incidence.[10] HIV-1, the most virulent of the two known HIV types, is also the most widespread in SSA.[4, 8] HIV-2 although similarly transmitted as HIV-1, and causes acquired immunodeficiency disease syndrome (AIDS) as HIV-1 does

is less virulent and far less widespread, endemic in west Africa but presenting only very rarely in southern Africa.[12]

1.1.1. HIV infection burden heaviest in women than men in SSA

HIV infection burden in SSA is heavier in women than men.[4, 8, 13] One half of the estimated 7000 new HIV infections that occur per day in SSA are in women, with 41% HIV infections occurring in young people in the 15 to 24 age group.[4, 8] By the end of 2009, 12 million women and 8 million men were living with HIV in SSA.[4, 14] Within SSA, HIV is dominantly transmitted via unprotected sexual intercourse. Although less widespread and less accurately documented, HIV transmissions among men who have sex with men (MSM) have also reported in SSA.[15-19]

1.1.2. Decline in new HIV infections but increased HIV burden in 2011 globally

The United Nations Program on AIDS (UNAIDS) reported a 19% drop in the number of people newly infected with HIV between 1990 (3.1 million) and 2011 (2.5 million).[10] Innumerous prevention interventions implemented over time have contributed to the drop in the estimated number of new HIV infections globally, but HIV burden which is function of HIV incidence and HIV-related mortality, has increased over four-fold from 8 million in 1990 to 34.2 million in 2011.[1, 4, 8].

1.1.3. Fewer HIV-related deaths occurred in 2011

Worldwide, an estimated 70 million people have been infected with HIV and 35 million people have died from AIDS-related conditions since the beginning of the HIV

epidemic.[20] Encouragingly, possibly because of the expanded access to effective combinational antiretroviral therapy (cART), the number of people who died from HIV-related complications in 2011 was 600,000 fewer than 1.8 million AIDS-related deaths estimated in 2005. According to UNAIDS, the heightened HIV burden in 2011 might be largely due an increased access to combination antiretroviral therapy (cART), with resultant improvement in quality of life and survival of people living with HIV.[10]

1.1.4. SSA accounted for the highest number of new HIV infections in 2011

Consistent with reports in prior years, compared to other regions globally, SSA has been disproportionately impacted by the HIV epidemic, accounting for 1.8 million out of 2.5 million (72%) HIV newly infected globally in 2011.[10] By the end of 2010, most regions of SSA had HIV prevalence greater than 1% among adults 15 to 59 years, and consequently qualified as experiencing "generalized HIV epidemics", a World Health Organization (WHO) classification.[21, 22] Unrelenting multipronged HIV prevention efforts in SSA have contributed to the considerable success noted, especially in recent years. For example, 13 countries among the 25 countries worldwide that recorded more than a 50% reduction in HIV prevalence are located in SSA. [10, 23]

1.2. Marked geographic heterogeneity in HIV burden globally and within SSA

Within SSA, HIV burden is heaviest in southern SSA, where the HIV prevalence is greater than 10% in most countries.[8, 10] The differential burden in HIV infections between and within countries globally may reflect the uneven distribution of risk factors for HIV infection in different communities.[10, 24] Population HIV prevalence is

influenced by myriad factors including individual-level, biologic and community-level factors. Additionally, certain people in some population (e.g., adolescent girls) are vulnerable to HIV infections because of they are invariably exposed to socioeconomic, cultural and behavioral factors that synergistically raises the risk of HIV infection.[25]

1.3. Increased burden in HIV infections in 2011 may be linked to improved survival of HIV-infected persons on cART and high rate of new HIV infections in SSA

HIV-related mortality depletes community HIV prevalence, but access to effective cART improves survival of HIV-infected persons and increases the HIV prevalence pool.[26-28] An estimated 1.8 million people died from HIV-related illnesses in 2010, but estimates indicate that improved access to effective cART averted an estimated 2.5 million HIV-related deaths in low and middle income countries (LMIC).[8]

Continued growth of the HIV burden in a community may be due to sustained occurrence of new HIV infections, immigration of HIV-infected persons into a given venue, and reduction in number of HIV-related death due to access to cART.[10] Based on the “treatment as prevention concept”, researchers have argued that HIV transmission rates may be lowered among adults via effective chemotherapy (cART) in a similar manner that chemotherapeutic prevention intervention limit HIV transmission from mother to child, thus cART may slow the growth of the HIV epidemic (if behavioral risk factors remain constant).[29]

1.4. HIV preventive interventions and HIV epidemic maturation effects likely to contribute to the falling number of new HIV infections

Undoubtedly, HIV preventive and treatment interventions implemented over the years have played key roles in driving the number of people newly HIV-infected downwards in recent years.[29] However, maturation effects of the HIV epidemic have possibly contributed to the noted downward trends in HIV incidence. Specifically, as HIV epidemic mature and become more widespread, more people are awakened to the risk factors for HIV infection, and may adopt less risky sexual behavior.[30]

With increasing knowledge on the routes and risk factors for HIV infection, the number of people who shun risk sexual behavior (i.e., unprotected sexual behavior) has increased along with the maturation of the HIV epidemic. Also, there can be saturation phenomena, such as the high prevalence seen in highest risk persons, with early high death rates; prevalence can decline merely as a function of these saturation dynamics.[31, 32]

HIV preventive interventions are crafted to curb the spread of HIV infections on the basis factors identified as being associated with increased odds of HIV infections. Admittedly, several factors that are complexly interrelated, and overarches individual-level factors, community-level and structural level factors drive the spread of HIV infection.[25] Highlighting the connectedness of risk factors for HIV infections, Vermund and Hayes (2013) emphasized that dampening the HIV epidemic require concerted and multipronged interventions.[29] To design interventions that are appropriate to local environment require adequate understanding of the HIV epidemic

dynamics in affected communities, and hence the importance of monitoring trends in HIV incidence and prevalence.[20, 29, 33]

1.5. Distribution of prevalent HIV infections in Zambia

Zambia has a generalized HIV epidemic (W.H.O definition of HIV prevalence > 1% in the general population). Within a few years after official report of the first AIDS case in Zambia in 1985, HIV/AIDS had emerged as a prominent public health problem. For example, by 1994, the estimated HIV prevalence among pregnant women in Livingstone, an urban area setting, had spiked to 32%. Between 900,000 and 1,100,000 people out of 13.2 million people were living with HIV in Zambia in 2011, an HIV burden nearly as great as the estimated 1.3 million people living with HIV infection in the United States of America, a country with 24 times the number of persons (314 million in 2012).[10, 34, 35]

1.6. Information gaps on the HIV epidemic

1.6.1. HIV prevalence in Zambia varies geographically and by socio-demographic factors

Zambia is among countries (e.g., Namibia, Botswana, Swaziland, Lesotho, Mozambique, Malawi, Zimbabwe, and South Africa) in SSA adversely affected by the HIV epidemic, with 14.3% of adults' aged 15 to 49 years living with HIV in 2007.[1, 10, 23, 36] The national weighted estimate of HIV prevalence (14.3%) conceals existing regional variation in HIV prevalence (e.g., 7% in Northern and Northwestern provinces, and 21% in Lusaka Province).[36] The uneven burden of HIV infection across

geographical regions may be indicative of differential distribution of factors that predispose to HIV infection (e.g., educational attainment, residence, age of first sex, prevalence of unprotected sex, and age).[37] Investigation of trends in HIV incidence and prevalence by selected risk factors and/or predisposing factors .(e.g., age, marital status, education) may unveil critical information to better understanding of the HIV epidemic.[38]

1.6.2. Few studies have examined non-linear patterns in HIV prevalence trends in Zambia

Most studies conducted to examine HIV prevalence trends in SSA have largely assumed a linear decline of HIV prevalence over time (e.g., Kenya, Tanzania, Uganda, South Africa and Zambia) and have not explored the possibility of non-linear trends.[24, 39-43] Nevertheless, non-linear HIV prevalence trends may exist and, in fact, have been documented in some dramatic examples, as with Uganda's decline and recent rise in background prevalence.[44] Fewer studies have attempted to capture non-linearity in HIV prevalence trends.[24, 45, 46] Because of challenges inherent in conducting trends analyses using longitudinal studies, repeated cross-sectionally collected data have been used in most countries to inspect trends in HIV prevalence.[45-47]

1.6.3. Divergent findings on the association between educational attainment and HIV in SSA

Divergent findings have been published regarding the association between educational attainment and prevalent HIV infection. For example a number of studies conducted in SSA in earlier years (i.e., 1980s and 1990s) of the HIV epidemic reported

higher odds of prevalent HIV infections among people with higher educational attainment than the odds of prevalent HIV infection among people with lower educational attainment.[48, 49] Studies conducted over different stages of the HIV epidemic on the association between educational attainment and HIV infection have yielded mixed (i.e., negative, null and positive).[45, 46, 50-56]

The observed divergent findings may be explained by a number of reasons including underpowered studies, methodological limitations of prior studies (i.e., cross-sectionally designed studies), and the varied definition of low of educational attainment over time, from study to study. Educational attainment effects on risk of HIV infection are on a continuum, and do not conform to a rigid cut-points introduced when educational attainment categories are formed (e.g., primary and secondary). Divergent findings may reflect a changing relationship between educational attainment and prevalent HIV infection over the years (i.e., year-education interaction).[57, 58]

1.6.4. Limited data on the simultaneous effects of age, period, and birth cohort on trends in HIV prevalence

Although age, period, and birth cohort effects may influence HIV prevalence trends, fewer studies have examined simultaneously age, period, and cohort effects on HIV prevalence. Thorough understanding of age, period, and birth cohort effects on HIV prevalence may yield key information on birth cohorts that are severely affected by HIV epidemic, which might guide targeting of HIV preventive and treatment interventions.[59-61]

Prior studies have described HIV prevalence by age and time periods but less limited information exist on the independent association of age, period, and birth cohort with HIV prevalence trends in Zambia.[43, 45, 46] Simultaneous inspection of age, period, and birth cohort's effects on HIV prevalence in Zambia may reveal patterns in HIV trends that might have eluded prior research on HIV prevalence trends.

Pregnant woman's age is an important factor in understanding the dynamics of HIV incidence and prevalence.[62] Age effects represent person-level variations, and reflect physiological changes as well as cumulative lifetime and social experiences.[63-65] Period effects captures external factor influences such as social, cultural, economic, or physical environment that may induce changes HIV incidence and/or prevalence.[59, 63] For example, period-specific behavior patterns and lifestyles that apply across all age groups in the population may qualify to exert period effect.[66-69]

Birth cohort's effects are unique to persons who were born around the same period. Persons in the same birth cohort period are more likely to have shared similar experiences (e.g., factors that predispose or prevent acquisition of HIV infection) than persons in a different birth cohort.[59, 63] For example, pregnant women in a younger birth cohort (e.g., 1990-1996) may have similar attitudes toward unprotected sex, concurrent partnerships compared to women in an older birth cohort (e.g., 1965-1969).[59] Age, period, and birth cohort effects can interact, as with the so-called sexual revolution in the late 1960s in North America and Western Europe where young people in an early baby boomer birth cohort were faced with rapidly changing social norms in a brief time period.[70] Persons may have differential receptiveness of HIV preventive interventions across birth cohorts (e.g., condom acceptability and consistent use).

CHAPTER 2

SPECIFIC AIMS AND SIGNIFICANCE

My dissertation work was anchored on three specific aims: specific aim 1 focused on a meta-analysis of literature on association of educational attainment, and specific aim 2 concentrated on understanding HIV prevalence trends among pregnant women using antenatal care based HIV sentinel surveillance (ANC-HIV-SS) data collected between 1994 and 2011 in Zambia.

2.1. Specific aim 1

To conduct a meta-analysis of peer-reviewed research literature on the association between educational attainment and prevalent HIV infection among pregnant women in SSA.

2.1.1. Research question

Among pregnant women in SSA, is higher educational attainment associated with increased odds of being HIV infected?

2.1.2. Hypothesis

The odds of being HIV-infected are lower among pregnant women with higher education (i.e., $\geq$ primary school education).

2.2. Specific aim 2

To examine trends in the prevalence of HIV by age, educational attainment, urban or rural residence, and parity among pregnant women aged 15 – 44 years attending ANC clinics used for the Zambia ANC-HIV-SS in 1994, 1998, 2004, 2006, 2008, and 2011.

2.2.1. Hypothesis 1

Linear models of trends in HIV prevalence among pregnant women in Zambia suggest a decline over time. The decline in HIV prevalence is not consistently linear, and I hypothesized that non-linear models will reveal significant recent increase in HIV prevalence. Because HIV infected people are living longer due to cART treatment, and as the benefit of cART spread, fear associated with AIDS may dissipate, increasing participation in risk sexual behavior. The Ugandan HIV prevalence and incidence upsurge is a classic example.[71]

2.2.2. Hypothesis 2

Age, period and birth cohort effect do not affect trends in HIV prevalence among pregnant women in Zambia between 1998 and 2011.

2.2.3. Hypothesis 3

Higher educational attainment is associated with reduced likelihood of being HIV seropositive among pregnant women attending ANC clinics in Zambia.

2.3. Significance of the study

This study seeks to augment prior research on HIV prevalence trends among pregnant women in Zambia, by using non-linear regression models to explore trends in HIV prevalence. Linear models used in prior studies to examine trends in HIV prevalence may not capture non-linear trends.[24, 42, 43] To explore non-linear trends in HIV prevalence among pregnant women, I used restricted cubic splines (RCS) functions to relax the linearity assumption between survey year and log-odds of HIV prevalence, resting on the assumption that decline in HIV prevalence trends may not be linear in some sites or overall.

There is limited information on age, period, and birth cohort effects on HIV prevalence because fewer studies have examined simultaneously the distinctive influence of age, period, and cohort effects on HIV prevalence in Zambia, and globally.[66, 68, 72-75] Houweling et al (1999) examined age, period and cohort effects on HIV incidence trends among drug users in France, and highlighted that age-period-cohort (APC) analyses may disentangle age, period and cohort effects, and may together with information from other sources (e.g., population characteristics and public health response), provide an elaborate description (i.e., identification of birth cohorts with plateauing HIV burden) of the growth and direction of the HIV epidemic.[72, 76, 77] Rosinska et al (2011) applied APC analyses to HIV surveillance data collected in Poland.[75] Guided by the method by Yang and Land (2006) for assessing age, period and birth cohort effects, the current analyses were conducted to investigate age, period, and birth cohort effects on HIV prevalence trends using ANC-HIV-SS data collected in seven rounds of the Zambian ANC-SS between 1994 and 2011.

Divergent findings on the association between educational attainment and prevalent HIV infection have been reported, yet fewer studies have been primarily set up to evaluate the relationship. Educational attainment is a key component of the social determinants of health, and is included as part of the human development index (i.e., education, and income, health), a measure used to rank countries in tiers of human development. The key role of educational attainment in health and economic outcomes is also reflected in its use as a target for Millennium Development Goal 2 that seeks to achieve universal primary school education by 2015.[48, 78, 79] Against the intuitive expectation of greater risk of infectious diseases among the poor, illiterate, and less educated, research findings from studies conducted SSA have reported higher odds of prevalent HIV infections among more educated persons, particularly in earlier years of the HIV epidemic (i.e., 1980s and early 1990s).[42, 80] Hargreaves and Glynn (2002) highlighted in their systematic review that 20th century studies that had examined the association between educational attainment and prevalent HIV infections reported disparate findings, but a subsequent systematic review by Hargreaves et al. (2008) focused on SSA revealed what appeared to be a waning relationship.[51, 81]

A better understanding of the association between educational attainment and HIV infection among pregnant women may be a critical step in packaging and targeting HIV prevention and treatment interventions, and identifying groups at higher risk of HIV. [24, 42, 43, 82, 83] To provide a synthesis of available literature on association between educational attainment and prevalent HIV infection, I conducted a meta-analysis focused on studies conducted among pregnant women in SSA. Further, I examined the association between educational attainment and prevalent HIV infection using ANC-HIV-SS data

collected between 1994 and 2008 in Zambia. Findings from these investigations may be complementary and might provide key information for a better understanding the HIV epidemic among pregnant women.

CHAPTER 3

REVIEW OF THE LITERATURE

3.1. Overview of HIV epidemic among women in SSA

The HIV epidemic in SSA affects nearly people from all social and demographic groups, but the impact and extent of the distribution of HIV infections across subgroups of the population is different, with the heaviest burden among women. Buve et al. (2001) reported findings based on multi-site cross-sectional survey (1998/1999) conducted in sexually active men and women in regions with low (i.e., Cotonou, Benin and Yaoundé, Cameroon) and high (Nairobi, Kenya and Ndola, Zambia) burden of HIV infections, and highlighted that sexually active women aged 15 to 19 years were six times more likely to be HIV-infected compared to men aged 15 to 19 years.[84, 85] Several studies conducted in SSA have corroborated the heightened odds of prevalent HIV infection among young women compared to young men.[4, 37, 86-89]

3.2. What factors drive the HIV epidemic in SSA?

3.2.1. Biological factors specifically among women

Prior research focused on treatment, vaccine development, prevention control measures for HIV, and on the risk factors for HIV infection in SSA have yielded speculative and plausible explanations.[90] However, widely accepted consensus recognize the interplay of biological, social, cultural, behavioral factors and contextual factors in the spread of HIV infections.[25] For example, a larger surface area of the

vagina than the penis is exposed during unprotected heterosexual intercourse and infectious seminal fluids are retained longer in women than men post-coitus: circumstances that may heighten risk of HIV acquisition in women. [91, 92] Further, young women who have cervical ectopy, a condition characterized by extension of delicate cells that normally occur inside the cervix to the surface of the cervix (i.e., consequently susceptible to damage during penetrative-sex) have heightened risk of HIV infection during unprotected penetrative sex trauma.[92-94] Additionally, risk of sexual encounter with an HIV-infected person is higher in a community with higher background HIV prevalence as in most SSA settings than in a community with lower HIV prevalence.[13, 95]

3.2.2. Circulating HIV clade C is the most virulent

HIV-1 has three major groups (i.e., M, N and O), and among the nine genetically distinct HIV-1 clades within group M (A, B, C, D, E, F, G, H, J and K, circulating recombinant forms [CRF]), clade C, the major clade circulating in southern SSA, is regarded the most virulent, and possibly a contributing factor to the rapid spread of HIV in SSA.[96, 97] Even though there are geographic variations in the distribution of genes associated with susceptibility to HIV, the documented differences in the distribution of protective human genetic markers for HIV are inadequate to explain the global disparity in HIV incidence and prevalence.[98, 99]

3.2.3. Multiple factors acting together and singly drive the HIV epidemic in SSA

High rate of sexual mixing and low condom usage contribute to the growing HIV burden in SSA.[20, 90] Some of the cultural rites practiced in Africa possibly contributed to the spread of HIV infection. For example, although less commonly practiced in recent times, post-partner's death cleansing rite, a cultural norm in some parts of Africa, mandated sexual intercourse between the surviving partner and the deceased partner's relative.[100] On average, women have limited ability to fully utilize the ABC (abstinence, be faithful and Condom use) paradigm because of their dependence on men for social and economic sustenance: ensuing power imbalance lowers women's abilities to negotiate safer sex, including hindering women's academic and economic progression. [101, 102]

3.2.4. Factors that drive HIV incidence and prevalence are complex and interrelated

The model proposed by Poundstone et al. (2004) highlighted how complexly related individual sexual behaviors, environmental structural, cultural, demographic and socioeconomic factors drive the HIV epidemic in the population.[25, 103, 104] To explain factors that drive the HIV epidemic, Boemer and Wier (2005) adapted the proximate determinant model (PDM) from fertility studies, and categorized factors that drive HIV spread in three groups: (1) underlying (e.g., political, geographic, social, economic, demographic and cultural); (2) proximate (e.g., concurrency, condom use and cART use); and (3) biological factors (e.g., exposure to at risk population, circumcision). Boemer and Wier (2005) avered that proximate factors connects underlying factors to biological factors.[95]

Even though the PDM is attractive, measurement of proximate factors (e.g., consistency of condom use) is challenged by validity concerns, and often imprecise, and consequently limits utility of the PDM.[95] Cognizant of the important role of community-level factors in the dynamics of HIV epidemic, Bärnighausen & Tanser (2009) updated the PDM to include community-level factors.[105] Additionally, Vermund et al. (2009) stressed the need for concerted biomedical and behavioral research efforts to counter the challenges presented by HIV diversity.[20, 90]

3.2.5. Key risk factors for heterosexual HIV transmission in SSA

Table 3.1 presents selected factors related to the risk of HIV infection in SSA that have been examined in prior studies.[4, 20, 90] Documented prominent factors associated with increased risk of HIV infection include commercial sex, concurrent partnerships, co-infection with bacterial and viral STIs (e.g., human simplex virus type 2 [HSV-2]), unprotected sexual intercourse, and lack of male circumcision. [106-112]. The risk-increasing effect of sexually transmitted infections (STIs), has been documented to be most profound in early stages of HIV epidemic when HIV transmission is mostly from the core high-risk group, but wanes as the HIV epidemic becomes generalized in late stages. [113-116] Widespread prevalence of factors that predispose to HIV infection may sustain further spread of HIV infections in a community.[110, 117, 118]

Table 3.1. Factors associated with HIV infection in sub-Saharan Africa based on selected peer-literature review

Selected factors deemed to associated with HIV in SSA	
Socio-demographic factors	• Age • Education attainment (variously defined) • Higher income level • Sex • Marital status • Widowhood
Biological factors	• Acute HIV infection • Circumcision • cART use • Past or current diagnosis STI
Behavioral factors	• Unprotected sexual intercourse • Commercial sex • Multiple lifetime or concurrent partners • Consistent condom use during sex • Assortative sexual mixing in the community • Age of partners and Wide between-partner age difference (>5 years) • Condom use • Alcohol abuse • Frequent absence from home for many days
Other factors	• Residence (urban or rural) and High HIV prevalence in local community • Low mean education level in neighborhood • Remarriage, and duration of marriage, and Gender-related social and economic inequalities and Poverty and lower social status

3.3. Association of education and HIV infection (Literature review for specific aim 1 and hypothesis 3 in specific aim 2)

Research conducted in earlier stages of the HIV epidemic (i.e., 1980 and 1990s) show greater odds of prevalent HIV infection among educated, affluent, and mobile people than less educated, poor and less mobile people.[119-121] Prior research, mostly cross-sectionally designed, have yielded divergent findings on the association between educational attainment and HIV infection (e.g., null, negative, and positive).[50, 51, 81] To clarify the relationship between education and HIV infection, Zuilkowski et al (2011) emphasized the role of intention and control as key proximate determinants of sexual risk behavior based on the cognitive theory model.[58]

On average educated people are more likely to have a higher income, cognitive ability and/or self-efficacy, and are more likely to have greater control over their intentions, and consequently over their behavior than less educated people.[53] Some empirical studies reviewed have indicated more frequent unsafe sexual behavior (e.g., multiple concurrent and lifetime sex partnerships) among educated compared to less educated people.[122, 123] Cross-sectional comparison of risky sexual behavior among men with and without secondary school education in Cameroon found that men with ≥ secondary school were more likely to have had unprotected sex (i.e., non-use of a condom during sexual intercourse), (OR =4.17, 95% CI: 2.65, 6.25), and had more lifetime sexual partners (OR=2.59, 95% CI: 2.02, 3.31). However, a subsequent meta-analysis study concluded that men with higher education who practiced risky sexual behavior were more likely to use a condom during risky sexual escapades compared with men with lower education.(OR=3.1, 95% CI: (2.48, 3.77).[124]

3.3.1. Educational attainment as an indicator of socioeconomic status (SES)

High educational attainment often positively correlates with high socioeconomic status (SES) but may not be a precise marker of SES given the multidimensional nature of SES (i.e., education, occupation and income that is defined as social and economic standing of a person in a community).[125, 126]. Prior research conducted in western countries has linked educational attainment and SES to better health, social and economic benefits. Despite its use as a proxy for SES, educational attainment may not be a perfect marker in some settings, specifically in developing countries.[48, 126-128]

Findings on the association between SES and HIV infection varies across HIV epidemic settings and gender in SSA. [67, 129, 130] For example, in earlier years of the HIV epidemic (i.e., 1980s) greater opportunities for travel away from home for many days, disposable income and ability to buy sex, and multiple and concurrent sexual partnerships common among well-educated and high-SES persons have been advanced as possible contributors to the noted elevated odds of prevalent HIV infection among educated and high-SES subgroup.[56] Impoverished women than men may adopt risky sexual behavior that increases risk of HIV infection. Therefore, a vicious influence of individual, community, cultural and social factors drives the HIV epidemic.[48, 49, 126, 129, 131]

3.3.2. Profound decline in HIV prevalence among 15 to 24 year-olds in SSA

HIV prevalence trend analyses in the 15 to 24 year-olds have revealed profound decline among educated compared less educated persons in most parts of SSA, but a concomitant rise in HIV burden has been noted among least educated people.[46, 55]

Researchers have suggested the differential implementation HIV prevention and treatment interventions as well as decline due HIV epidemic maturation, have contributed to the observed decline in HIV prevalence. The fall in HIV prevalence may be explained by the assuming differential response to preventive intervention with educated people tending to respond more favorably than do less educated people (e.g., adopting less risky sexual behavior). Many researchers have reasoned that educated people are more likely to have greater information on HIV risk factors, and also on modes of transmission as well as preventive means (e.g., consistent condom use).[81] Based on prior research, self-efficacy appears to be key factor in adoption of safer sexual behavior in both educated and uneducated persons, but reports indicate that educated people have a greater propensity to adopt safer sexual behavior.[20, 132-135] Figure 3.1 shows the hypothesized relationship between educational attainment and HIV infection.

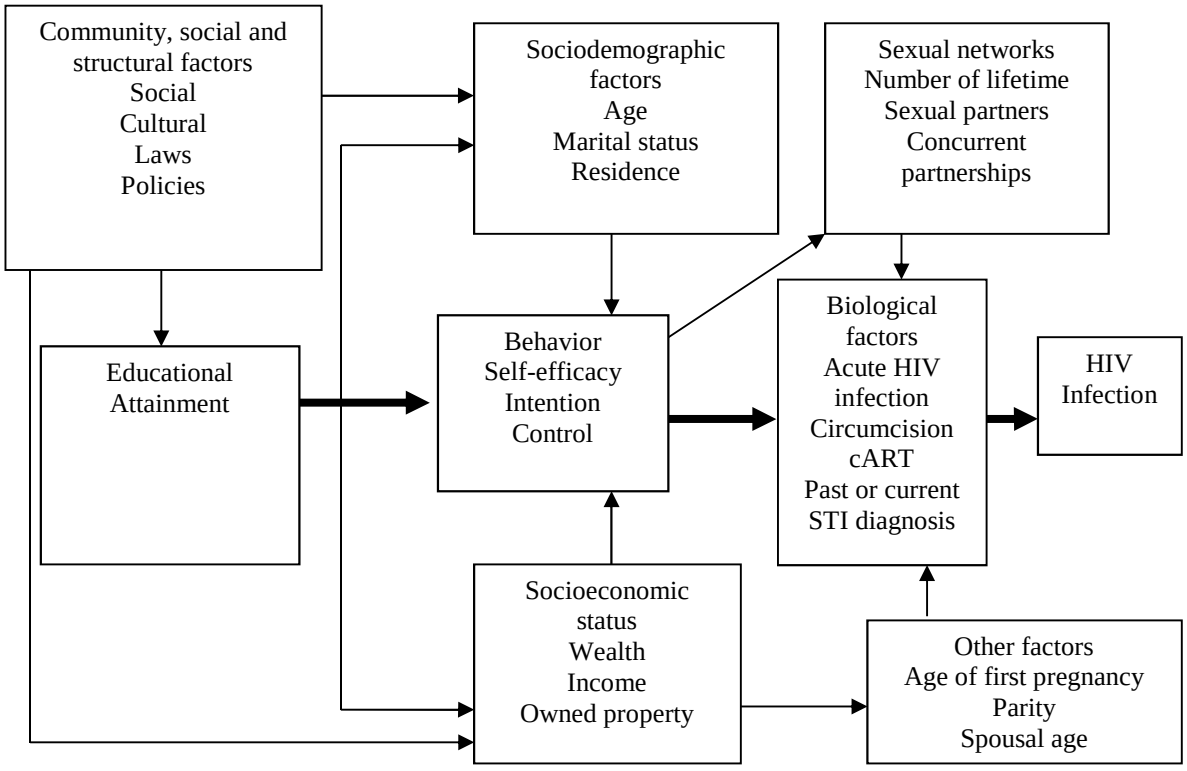

Figure 3.1. Hypothesized relationship for the relationship between educational attainment and HIV highlighting the multifactorial (i.e., individual, community and structural factors) and complex interrelationships that drive the HIV incidence and prevalence. Source: adapted from Poundstone et al (2004), Jukes eta l (2001, Vermund et al (2009) and Boerma & Weir (2005)

3.4. Methodological limitations and variation of the association over time may explain divergent findings

A number of factors may plausibly explain the divergent findings (i.e., negative, null, and positive association) on the association between educational attainment and prevalent HIV infection in SSA (e.g., Tanzania, Uganda, Zambia, and Zimbabwe).[35, 39, 40, 54, 117, 118, 136-140] First, real variation in the association study calendar year and educational attainment as the epidemic has progressed may to some extent explain the divergence in the association between educational attainment and prevalent HIV infection at different stages of the HIV epidemic.[20] Second, methodological weaknesses in the design and analysis of some of the prior studies (e.g., underpowered

studies; inconsistent definition of education attainment as a study factor in the analysis; and use of cross-sectional data).[20] Third, diverse characteristics of studied populations; background community HIV prevalence; and imprecise variable measurements individually and collectively may explain the divergent findings.[58, 121, 139, 141]

3.5. Hypothesized relationship between educational attainment and prevalent HIV infection using a "Nebulous" directed acyclic graph (DAG)

The association between educational attainment and prevalent HIV infection may be confounded by number demographic factors (e.g., factors from childhood through adolescents to adulthood).[142] For example, becoming an orphan at a younger age may disadvantage education advancement as exemplified in Figure 3.2.

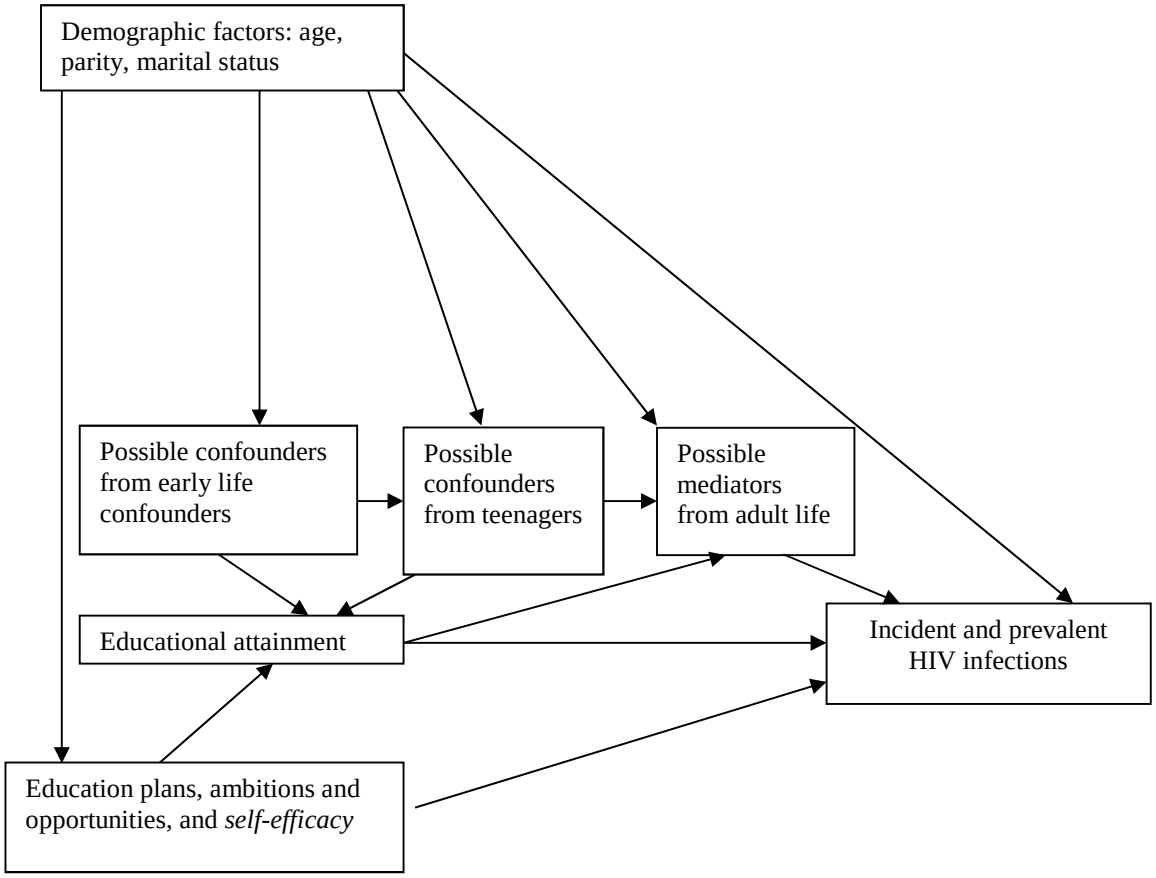

Figure 3.2. Directed acyclic graph representation of the hypothesized relationship between educational attainment and prevalent HIV infection adapted from Cohen et al (2013)[142] As suggested by Zuilkowski et al (2012) self-efficacy is an important factor in explaining the relationship between educational attainment and HIV.[58] Higher educational attainment without self-efficacy in adopting safer sexual behavior might not tend to be protective.

3.6. Higher odds of prevalent HIV infections among educated than less educated people in earlier years of the HIV epidemic (i.e., 1990s)

Hargreaves and Glynn (2002) systematic review of 27 studies on the association between educational attainment and HIV infection in developing countries by revealed divergent findings: negative, null and positive association.[52] Focusing on studies from Africa, the odds of prevalent HIV infections were higher among educated than among less educated people.[41, 51, 138, 139, 143] Hargreaves et al (2008) summarized evidence from studies conducted between 1997 and 2003 that examined the association

between educational attainment and prevalent HIV infections, and the association seemed to have waned in post-1996 studies that tended to be mostly null or negative or null association with less frequency of studies that reported positive association.[36, 50, 81, 116].

3.7. Young female school dropouts at increased risk of contracting HIV

Adolescents in SSA who drop out of school bear increased risk of HIV infection compared to those who remain in school.[81, 94, 144] [145] Several studies have reported a correlation between low educational attainment and sex debut at a younger age (< 15 years), early marriage, high fertility, high number of lifetime sexual partners, and alcohol use.[145] Lower odds of HIV infections have been documented among women with ≥ 7 years of formal education compared to women with <7 years.[24, 39, 46, 55, 146]

3.8. Ethical and logistical constraints limit use of randomized and prospective studies to examine education-HIV relationship

The often used design to assess the association between educational attainment and HIV infection are cross-sectional studies. Consequently, causal inference is limited because of inherent limitations of these cross-sectionally designed studies (e.g., failure to establish temporality). Use of prospective observational studies for examining the education-HIV relationship is limited by logistical and financial challenges. Further, randomized control studies may not be used because ethical concerns that invariably arise if subjects are randomly assigned education groupings.[58] Continued exploration of the

relationship between education attainment and prevalent HIV infection is merited, and would obviate transposition of dated associations to current contexts.

3.9. Peer-reviewed literature on the association between educational attainment and HIV in Zambia

Fylkesnes et al (1997). Fylkesnes et al (1997) examined HIV prevalence by selected self-reported sociodemographic factors among pregnant women, and described HIV prevalence trends based cross-sectional data sourced from ANC-based HIV surveillance program conducted in 1990, 1992, 1993 and 1994.[147] Eligible pregnant women recruited via non-probability sampling strategy that captured pregnant women who sought ANC care at specific health centers used for ANC-based HIV surveillance. Among women aged 25-44 years, pregnant women ≥10 years of education were 3.1 times more likely HIV-infected (OR=3.1, 95% CI:, 1.59, 3.79) than pregnant women with <5 years of education.[41]

Two approaches were applied to define educational attainment. First, Fylkesnes et al (1997) created five educational attainment categories: ≥4; 5 to 6; 7-8; 9-10; and ≥10 years of education. The second approach of educational attainment categorization by Fylkesnes et al (1997) was to analyses conducted in pregnant women aged 15 to 19 years: ≥4; 5 to 6; 7 and ≥8 years of education.[41]

Fylkesnes et al (1997) invariably assumed constant educational attainment effects within categories that resulted from categorizing a continuous variable (i.e., number of education years). This assumption may be questionable if there is profound variability of the continuous educational attainment effects.[148] Furthermore, justifying how a the

substantive difference between a pregnant woman who had six years of schooling and another pregnant women who completed five years of schooling may be difficult where these pregnant women belong to two different categories based on a cutoff (e.g., educational attainment <5.5). Even with extant guidelines on categorization of continuous variable based on subject matter information, information is inevitably lost when a continuous variable is categorized. To avoid loss of statistical efficiency that may arise from categorization, educational attainment may be examined as a continuous variable.[149-151]

Fylkesnes et al (1998). Based on cross-sectional ANC-HIV-SS data collected via non-probability sampling and population-based data (PBS) captured via random cluster sampling from Chelstone (Lusaka Province) and Kapiri Mposhi (Central Province), Fylkesnes et al (1998) examined the HIV prevalence educational attainment and reported immaterial differences in HIV prevalence estimates based on the ANC sample and the population-based data.[39] HIV prevalence tended to increase by educational attainment among 25 to 39 year-olds.

Further assessment revealed lower proportion of women who self-reported educational attainment ≥ 8 schooling years among the ANC-based sample (44%) compared to the proportion of women in the PBS sample with ≥ 8 schooling year completed (73%).[39] The lower proportion of pregnant women with ≥ 8 years of formal education may be suggestive of presence of selection bias of young and less educated women who are more likely to be pregnant younger ages.[39]

Michelo eta l (2006a). Michelo et al (2006) examined the association between educational attainment and HIV infection in men and women aged 15 to 57 years using data sourced from a population-based cross-sectional surveys conducted in Kapiri-Mposhi (Central province) and Chelstone (Lusaka province) in 1993, 1999, and 2003. HIV serostatus were established via screening for HIV specific antibodies in saliva specimens collected from participants. Socio-demographic data were self-reported via structured questionnaire. Pregnant women were grouped according to the following educational attainment categories: 0-7; 8-10; and ≥10 schooling years).[152] Here as in Fylkesnes et al (1998), categorization of continuous educational attainment possibly led to loss of information in the continuous variable.

Michelo et al (2006) using 2003 data reported that men in urban areas aged 15 to 19 years and who had reported ≥ 10 years of formal schooling were 80% less likely to be HIV-infected compared to men urban areas who self-reported having completed ≤7 years formal education (OR=0.20, 95% CI: 0.05, 0.73). Women in urban areas with more than ≥ 10 years formal education were 67% less likely to be HIV infected than women with <8 years of formal education (OR=0.33, 95% CI: 0.15, 0.72). Women in rural areas with ≥ 10 years of schooling had lower but not significant odds of prevalent HIV infections (OR=0.77, 95% CI: 0.28, 2.1).

Among men in rural areas aged 24 to 49 years, those men with ≥ 10 schooling-years were 57% less likely to have prevalent HIV infection relative to men who had < 8 schooling years. Among 24 to 49 year-olds in rural areas, the odds of prevalent HIV infection were higher among women with ≥ 10 years formal education compared with women with fewer than 8 years (OR =2.31, 95% CI, 1.13, 4.47). The population-based

sampling of participants is strength for this study but finding might not be generalizable to other regions of Zambia because of the possible differences in socioeconomic, cultural and social factors in Lusaka and Kapiri Mposhi compared with other areas in Zambia. For example, Kapiri Mposhi is at an intersection of major road networks connecting urban towns: therefore may not represent typical rural set-ups in Zambia.

Sandøy et al (2006) Sandøy et al (2006) examined the association between self-reported educational attainment and prevalent HIV infection using ANC-HIV-SS data collected in 1994, 1998 and 2002, and reported higher odds of prevalent HIV infections among women who had >10 schooling years compared to women who <5 schooling years (OR=1.45, 95% CI: 1.27, 1.66 in urban sites and OR=3.03, 95% CI: 2.47,3.72 in rural sites).[45] Sandøy et al (2006) reported significant statistical multiplicative statistical interactions: (1) educational attainment and residence among pregnant women aged 25 to 49 years; (2) educational attainment and age; and (3) educational attainment and survey-year. Survival bias is likely because prevalent cases were used all the reviewed studies, especially analyses restricted to 25 to 49 year-olds.

3.9.1. Summary of the literature review on the association between educational attainment and HIV

Literature review of findings from observational studies conducted in SSA revealed divergent findings regarding the association between educational attainment and prevalent HIV infection. However, the association tended towards non-significant protective association in later than earlier years (1980s and early 1990s) of the HIV epidemic in SSA. [52]. Cutoff points used to educational attainment categories were inconsistent raising the possibility of misclassification of pregnant women.[51, 81]

Most of the studies that used educational attainment as study factors categorized the continuous form (i.e., number of school years completed) or captured educational attainment as a categorical variable (e.g., no education, primary and secondary).[24, 40, 42, 43, 46, 81, 141, 146, 152, 153] Analysis of educational attainment in its continuous form obviates the use of subjective cutoff points to create categories. Further, variable cutoff points used for categorization limits objective comparison of study findings.

Studies that used ANC-HIV-SS data collected from multiple sentinel sites in different geographic location did not account for possible intra-site clustering among pregnant women. Substantial intra-site clustering of pregnant women may not affect parameter estimates but will harm inference: smaller standard errors arising from clustering. Selection bias is a potential threat to validity when ANC-HIV-SS data is used because women who become pregnant may be different from those who do not become pregnant. Consequently, external validity is limited. Further, selection bias may arise in PBS if the response rate is low.

The cross-sectional designed of the studies reviewed that examined the association between educational attainment and HIV infection limits causal inference. Because sociodemographic and/or behavioral data were captured via self-report, information bias may be eminent: recall bias. Residual confounding may plague the reported association due imprecise measurements of variable used as potential confounders, and because some key variables were not measured and therefore not controlled for in the analyses. The main strength of all the studies was the use of serologically confirmed HIV serostatus.[24, 46, 146, 153]

Examination of the association between educational attainment and HIV infection among men and women older than 25 years may be less trustworthy because of the greater chance of survival bias influence. Impressively, most study reports conducted stratified the analysis: 15 to 24 and 25 to 44 year-olds. UNAIDS has recommended the use of the number of prevalent HIV infections in the 15 to 24 year olds as a proxy for the number of new HIV infection based on the assumption that men and women in the 15 to 24 years old are more likely to have had recent sexual activity onset.[24, 46, 146, 153]

3.9.2. Literature review for specific aim #2 and hypothesis #1 and Hypothesis #2: Trend in HIV prevalence in Zambia

3.9.3. Epidemiology of HIV in Zambia

Zambia, shown in Figure 3.3, is 752,612 square kilometers (the size of Texas in the US).[19] Provinces with large expanse of urban areas tend to be densely populated: 62.6 and 100.4 persons per square kilometers for Copperbelt and Lusaka provinces respectively.[154, 155] The urban population in Zambia (39% of 13.1 million) in 2010 appears to be concentrated along what is referred as the “line of rail” traversing Southern, Copperbelt, and Lusaka province. Towns dotted along the “line of rail” are relatively more urbanized and characterized with relative high commercial activity.[154, 155]

Estimates from ANC-based HIV surveillance system set in 1990 for monitoring HIV prevalence trends among pregnant women indicate that by 1998, HIV prevalence among pregnant women had increased beyond 25% in most of the sentinel sites located in urban areas.[39-41, 154-156] The first country-wide PBS via Demographic and Health Survey (DHS) in 2001 placed HIV prevalence at 14.3%. Subsequent DHS in 2007

revealed a non-significant decline of HIV prevalence in adults 15 to 49 years old, from 15.6% in 2001 to 14.3% in 2007.[36] In 2009, an estimated 1.1 million people aged 15 to 49 years were living with HIV in Zambia.[4, 13]

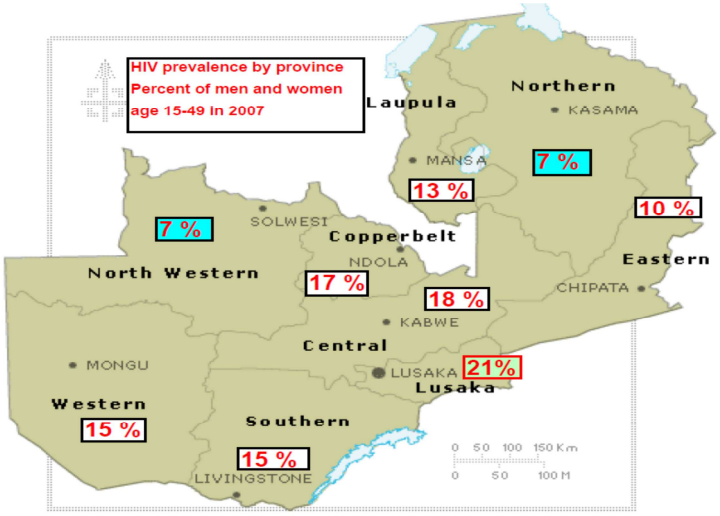

Figure 3.3. Map of Zambia showing HIV prevalence distribution by province based on the 2007 DHS: HIV prevalence ranged from 7% Northern/Northwestern Provinces to 21% in Lusaka Province.

Fewer studies have attempted to directly measure HIV incidence data, and most countries have depended on mathematical model-based estimation of HIV incidence in the 15 to 24 year-olds generated by UNAIDS.[19, 157] UNAIDS estimated that 76,000 of new HIV infections occurred in Zambia in 2009.[4] This is 50% higher than the estimated incident cases in the USA in 2009, even though Zambia has less than 5% of the US population.[158] Compared to 2001, new HIV infection in Zambia dropped by 58%.[14]

3.9.4. High HIV burden female sex workers and prisoners in Zambia

Because of high risk of exposure to HIV, HIV burden in female sex worker and prisoner is higher than in the population. Buve et al (1991) reported that 69% of the 319 female sex workers (i.e., 1998 and 1999) were living with HIV, and a parallel study by Kamanga et al (2005) reported that 65% of the 283 female sex workers were HIV-infected.[34, 84, 85, 159, 160] Zulu et al (2006) reported 33% HIV-infected persons among 641 MSM, although MSM-related HIV infections are a lesser contributor to the HIV pandemic in Zambia.[34] Simooya et al (2001) surveyed 1566 prisoners in Kamfinsa prison in Kitwe, Mukobeko prison in Kabwe and Solwezi prison in Solwezi in 1998-1999, and placed HIV prevalence at 27%.[161] HIV prevalence among pregnant women in refugee camps (i.e., harboring people from Democratic Republic of Congo [DRC] and Angola) in Zambia ranged from 2.4% to 3.9% in 2006, paralleling HIV prevalence rates of the Angola and DRC.[157].

3.9.5. HIV incidences rates from population sub-group has limited generalizability

A number of studies have estimated HIV incidence in Zambia but estimates may be plagued by external validity concerns. Hira et al (1997) placed HIV incidence rate among pregnant women in a cohort of discordant couples at 87 per 1000 couple-years. [162] Stephenson et al (2007) studied a cohort of couples in Lusaka, and calculated that HIV incidence was 93 per 1000 person-years in this cohort.[163] Celum et al (2008) and Kapina et al (2009) estimated HIV incidence of 45 per 1000 person-years among HIV-2 seropositive women and HIV incidence of 26 per 1000 among 239 women in Lusaka.[164, 165] Heffron et al (2011) studied 731 HIV negative migrant workers, and estimated that HIV incidence was 4.1 per 1000 person-months.[165, 166] Despite the

importance of HIV incidence estimates for mapping out prevention strategies, the external validity of generated HIV incidence estimates is limited given the studied populations may not represent the general population (i.e., studied maybe plagued by selection bias).

3.9.6. Overview of studies on trends in HIV prevalence in Zambia

Fylkesnes et al (1997). Fylkesnes eta l (1997) used ANC-SS cross-sectional data to report trends in HIV prevalence among pregnant women from 1990 through 1994.[41] Fewer and less geographically spread ANC-SS sentinel sites (10 to 12) were used in the pre-1994 surveys (10 to 12); data were collected from 27 sites in 1994. Self-reported socio-demographic and reproductive data were collected via structured questionnaire, and HIV serostatus determined by serologically via anonymous and unlinked testing.[41] Fylkesnes eta l (1997) reported that HIV prevalence in urban sites increased from 27% in 1992 to 35% in 1994 in Chilenje; declined from 24.5% in 1990 to 21.7% in 1994 in Kalingalinga, but stabilized at high prevalence in Chelstone (25%) and Matero (28%). HIV prevalence in rural sites increased from 11.4% in 1990 to 14.6% in 1994 in Kashikishi in Luapula Province.[41] Furthermore, there was a decline in HIV prevalence in pregnant women aged 15 to 19 years from 27.5% in 1993 to 22.5% in 1994.[41]

Limitations were noted in the study. First, HIV prevalence trend analyses were limited to 10 sites that had data for at least two survey rounds between 1990 and 1994.[41] Second, time period examined (1990 to 1994) and the between-survey time durations (i.e., one-year), were not long enough to enable meaningful analysis of trends in HIV prevalence.[167] Third, earlier ANC-HIV-SS rounds (e.g., 1990) lacked data on

pregnant woman's age: consequently, age-adjusted HIV prevalence trends were not examined for 1990, 1991 and 1992 data.[41]

Sandøy et al (2006). Sandøy et al (2006) examined HIV prevalence trends in pregnant women using ANC-HIV-SS data cross-sectionally collected in 1994, 1998, and 2002[45] Between 1994 and 2002, HIV prevalence declined by 27% (i.e., 28.5% to 21.8%) in urban and by 11% (i.e., 11.4% to 10.1%) in rural sentinel sites in the 15 to 24 age group.[45] HIV prevalence remained stable in some sites, increased in other sites, and was less clear in some sites. The decline in HIV prevalence was profound among pregnant women > 10 years of schooling (i.e., 35% in 1994 to 22% in 2002). [45]

Michelo et al (2006b). Michelo et al (2006) examined trends in HIV prevalence by educational attainment in urban and rural area of Lusaka and Kapiri Mposhi (north of Lusaka in northern Central Province near the border of the Copperbelt Province) using population-based data collected cross-sectionally in 1995, 1999 and 2003.[152] Participants in the surveys were recruited via stratified random cluster-sampling based on census mapping as sampling frame.[46]

HIV prevalence declined among women age 15 to 24 declined (21.2% in 1995, 16.1% in 1999 and 8.5% in 2003) in urban areas with >10 schooling years. Further HIV prevalence declined among men during 1995 through 2003 period (30.2% to 11.7% among urban men and 18.1% to 15.3% in rural men). Similar declining trends were observed among women (34.3% to 17.5% in urban women and 29.7% to 17.3% in rural women. HIV prevalence declined among women in urban area with ≥ 11 schooling year during this period (45.6% (1995), 39.9% (1999 and 29% (2003), but stable prevalence,

albeit at high level in urban women with ≤7 schooling years (27.3% (1995), 26.7% (1999) and 31% (2003).

Study limitations: categorization of continuous data often lead to loss of information and loss of sensitivity of the analysis. Further, Mantel-Haenszel chi-square applied for examining linear trends may not pick out non-linear trends in HIV prevalence. Broad categories used as adjustment covariates in the multivariable logistic regression may limit examination of non-linear age effects within categories, and possible source of residual confounding (15 to 24 and 25 to 44). Further, self-reported data is subject to information bias (i.e., sociodemographic and behavioral data). Possible clustering was accounted for in the analysis.

Stringer et al (2008). Stringer et al (2008) based their analysis of HIV prevalence trends on data from ANC, PMTCT program, and cord blood HIV surveillance cross-sectional data derived from 24 obstetrical health centers in Lusaka, Zambia collected between 2002 through 2006. An estimated 23% (54,853) of the specimens screened were HIV seropositive. Significant decline in HIV prevalence noted (24.5% in 2002 to 21.4% in 2006).[168] Overall decline in HIV prevalence declined across all age groups (i.e., ≤17, 18-19, 20-24, 25-29, 30-34 and ≥35), but greatest decline were noted among pregnant women aged ≤ 17 years, by 37%, from 12.1% to 7.7%. HIV prevalence declined significantly in 11 out of 24 sites (p-value <0.05).[168] Generalized linear mixed model (GLMM) were applied to account for possible intra-site site clustering given data were collected from multiple health centers (i.e., health center was modeled as random effect).

Limitations of the study: First, Cochrane-Armitage test used for assessing linearity in HIV prevalence at specific site may not detect non-linear trends. Second, multiple data sources were used but investigators did not assess the impact of the different data sources (i.e., PMTCT data and cord blood surveillance data). Third, the likely differential willingness to participate in PMTCT program could a source of bias because characteristics of women who agree to an HIV test in the PMTCT program may be different from characteristics of women who refuse an HIV test. Stringer et al reported acceptance rates ranging 71% in the early years of PMTCT to 94% in later years, consistent with other studies that have reported substantial refusal rate in early years of PMTCT.[168, 169]

Kayeyi et al (2012). Kayeyi et al (2012) conducted HIV prevalence trend analyses based on cross-sectionally collected ANC-HIV-SS (1994 through 2008) and DHS (2002 and 2007) in the 15 to 24 year-olds.[24] Sociodemographic data were self-reported and HIV serostatus were serologically confirmed. Site-specific HIV prevalence trend analyses in 15 to 24 year-olds were conducted in 12 urban sentinel sites and 10 rural sentinel sites that had complete data for the six survey periods (1994, 1998, 2002, 2004, 2006, and 2008).

Analyses for HIV prevalence trends by educational attainment (0-4 years; 5 to 7 years; 8 to 9 years; and ≥10 year) were performed for the period 1994 through 2008.[24] HIV prevalence among educated pregnant women (≥ 5 years) declined in rural sites (11.4% to 6.4%) and urban sites (27.4% to 15.5%). Non-significant decline in HIV prevalence occurred in two sites in urban areas (i.e. Kalingalinga and Matero) and rural sites (e.g., Minga, Isoka and Ibenga). Between 1994 to 2008, statistically significant

decline in HIV prevalence among pregnant women were observed in 10 urban and 4 rural sites (ANC-HIV-SS data). Trend analyses from 2001/2002 and 2007 DHS data indicated an increase in HIV prevalence increased in urban men (3.7% to 5.0%), a drop HIV prevalence in rural men (3.1% to 2.9%); urban women (15.2% to 12.5%); and rural women (7.8% to 6.4%). Trends in HIV prevalence in Minga and Kalingalinga) appeared to be non-linear.[24]

Limitations of the study: Kayeyi et al (2008) used the Chi-square linear-by-linear trends test to assess trends in HIV prevalence over the considered years (1994 to 2008), which is more sensitive to linear than non-linear trends. Scores assignment to survey years in Chi-square linear-by-linear may be problematic because it does not consider the distance between survey years and the assigned scores may influence trend analysis. It appears the Chi-square linear-by-linear trend applied by Kayeyi et al (2012) failed to pick out non-linear trends in HIV prevalence in Minga and Kalingalinga sentinel sites.[24]

3.10. Summary of the literature review on HIV prevalence trends in Zambia

In general, both PBS and ANC-SS-based HIV prevalence estimates indicated a decline in HIV prevalence over the considered periods. Most studies conducted to examine trends HIV prevalence often assumed a linear decline in HIV prevalence. Trends in HIV prevalence may differ by region or site and may not be consistently linear. Fewer studies have explored the possibility of non-linear trends. Furthermore, statistical methods applied in the reviewed studies conducted to examine HIV prevalence trends had limited power detect non-linear trends, e.g., chi-square linear by linear Mantel

Haenszel trend test.[24] Several factors may influence HIV incidence and prevalence trends in a given population (e.g., changes in the prevalence of risk factors, treatment options, prevention interventions, and age-period-cohort effects). Fewer studies have examined contemporaneous age, period and birth cohort effects on HIV prevalence. Studies that have used ANC-HIV-SS have not accounted for possible within-site clustering in data.[170]

CHAPTER 4

DATA COLLECTION METHODS FOR ANC-HIV-SS: PRIMARY DATA COLLECTION FOR DATA USED IN THE PHD DISSERTATION

4.1. Main data sources for HIV prevalence data: PBS-DHS-HIV prevalence estimates and ANC-HIV-SS prevalence estimates

The ANC-HIV-SS and the PBS-DHS are the chief sources of HIV prevalence data in Zambia, as in many countries in SSA.[13, 36, 157] ANC-HIV-SS data are cross-sectionally sourced via convenient samples of pregnant women at participating health centers but the DHS data are cross-sectionally collected using two-stage sampling that apply proportional to size random sampling technique based on a sampling frame from the national census tract.[36] The DHS-based HIV prevalence estimates are superior to ANC-HIV-SS-based HIV prevalence estimates because, in addition to using more sound sampling strategy that minimize selection bias, DHS-based sample generates HIV prevalence estimates for both men and women.[36]

4.1.1. ANC-HIV-SS data used for my dissertation

To answer the research questions posed in my Ph.D. dissertation proposal, I used the ANC-HIV-SS data obtained collected between 1994 and 2011. ANC-HIV-SS has been conducted consistently in 22 sites since 1994, and in 24 sites since 2002 and HIV prevalence estimates from repeated cross-sectional surveys have been used to investigate trends in HIV prevalence.[147] Pre-1994 ANC-SS surveys had limited geographic coverage because the survey was conducted in 10 sentinel sites which were located in mostly urban areas.[36] Non-probability sampling strategy was used for recruiting

pregnant women in all the seven rounds of ANC-HIV-SS (i.e., 1994, 1998, 2002, 2004, 2006, 2008, and 2011).[157]

4.1.2. Difference between ANC and DHS HIV prevalence surveys

Table 4.1. Comparison of methodology of the Zambia ANC-SS and DHS used in HIV prevalence estimation

Survey name	Survey study areas	Biologic specimen	Sampling method	Responses rates	Possible biases
ANC-based HIV sentinel surveillance	Health facility-based survey	Serum/plasma prepared from whole blood [unlinked and anonymous]	Convenient sampling (non-probability)	Almost all women attending ANC are captured	Selection biases
Population-based DHS	Household based survey	Dried blood samples (identity stripped-off)	proportional to size random sampling based on the national census tract	76% in 2001 and 77% in 2007 [Response rates to interview were 95% and 94%]	Selection biases if response rate is low

4.1.3. PBS are regarded gold standard for estimating HIV prevalence in generalized epidemic settings

Table 4.1 compares PBS-DHS and ANC-HIV-SS methods for estimation of HIV prevalence. PBS-based HIV prevalence estimates may be regarded as gold standard provided the survey that generated the HIV prevalence estimates was not threatened by low participation rate or affected by methodological constraints (e.g., incomplete sampling frame).[36, 171] DHS are household-based PBS that have collected information on population health, nutrition and fertility in >300 surveys in 90 countries

worldwide. PBS-DHS based HIV prevalence estimates use sound statistically robust sampling strategy for recruiting survey participants, and are a trustworthy strategy provided response rates for the survey are sufficiently high.[36, 155, 171, 172] Selection bias may threaten the external validity of DHS-based HIV prevalence estimates if characteristics of those who agreed and those who did not agree to participate in the survey are different, or if key respondents are unavailable at the time of the survey (men migrating for work, for example).[36, 171, 172]

4.1.4. Two data points of HIV prevalence data may not provide reliable HIV prevalence trend analyses

HIV prevalence trends are better described when more than two time points of HIV prevalence data exist. Without doubt, DHS-HIV prevalence estimates in 2001 and 2007 DHS have provided useful information on the HIV burden, but may be inadequate to describe HIV prevalence given there are only two data points as of June 2013. However, DHS-based HIV prevalence trend analyses in Zambia have been corroborated by ANC-HIV-SS based HIV prevalence trend analyses.[24, 36, 40]

4.1.5. Identical HIV prevalence estimates in ANC-HIV-SS and PBS-DHS HIV prevalence estimates in Zambia

Some of the studies conducted in SSA indicated that ANC-based HIV prevalence estimates may either underestimate or overestimate community HIV prevalence, but Dzekedzeke et al (2006) compared 2001-2002 PBS-DHS based HIV prevalence estimates and 2002 ANC-HIV-SS based HIV prevalence estimates, and reported nearly congruent prevalence estimates in the 15 to 49 year-olds. Earlier in the HIV epidemic, Fylkesnes et

al (2001) had reported that ANC-HIV-SS based data among pregnant 15 to 19 year olds tended to higher than PBS based HIV prevalence estimates.[40] On the other hand, ANC-based HIV prevalence in ≥25 years tend to underestimated population HIV prevalence because of diminished fertility rates with increasing age.[40] Based on the extant literature, one may argue that whether PBS based HIV prevalence estimates and ANC-HIV-SS based HIV prevalence estimates are comparable may depend of the settings and stage of the HIV epidemic.

4.1.6. ANC-HIV-SS based HIV prevalence estimates must be interpreted in the context of their inherent biases

Notwithstanding the biases that accompany ANC-based HIV surveillance methods (e.g., potential for selection bias and exclusion of non-pregnant women and men), ANC-based HIV prevalence estimates have provided key data for understanding, assessing and monitoring magnitude of the HIV epidemic in Zambia.[24, 41, 157] Because ANC-based HIV prevalence estimates may be biased (e.g., possible selection bias of young pregnant women), ANC-based HIV prevalence estimates may not approximate general population HIV prevalence estimates.

4.1.7. HIV incidence rate is a preferable measure of progression of HIV epidemic but less used because of measurement challenges

Examination of trends in HIV incidence, although more informative than HIV prevalence trend inspection, is hampered by logistical, technical, and financial challenges that arise in direct measurement of HIV incidence.[173] Challenges encountered include a lack of a simple and reliable cost effective assay for detecting recent HIV infections

(though progress is being made in this) and the high costs and complexities of cohort study design of large sample of negative persons to be followed over long time interval.[174, 175] In fact, cohorts can give distorted incidence estimates due to the, Hawthorne effect, a circumstance in which persons may change their behavior because they know that they are in a study.[176-179] Because HIV prevalence is much easier to measure compared to HIV incidence, most countries in SSA use HIV prevalence estimates in adolescents and young adults ages 15-24 years old to approximate the number of people newly infected with HIV over time.[10, 24]

The onset of sexual activity in the 15 to 24 year olds is assumed recent, and prevalent HIV infections in the 15 to 24 year-olds are invariably regarded as recent infections, and consequently used as proxies for the number of persons newly acquired HIV infection.[4] This HIV in cadence estimation strategy disregard incident HIV infection in ≥ 25 year-olds. Additionally, 15 to 24 year-olds based HIV incidence approximations are less influenced by AIDS-related mortality, that may impact ≥25 year-olds based HIV incidence estimates. The relative ease of measurement of HIV prevalence than HIV incidence, encourages reporting of HIV prevalence in most surveys in SSA.[180, 181]

4.2. Study design and study population

4.2.1. Data collection and management for the ANC-SS for HIV in Zambia

My PhD dissertation relied on secondary analysis of repeated cross-sectional survey data collected from 82,086 pregnant women aged 15 to 44 years who participated in the ANC-HIV-SS (i.e., 1994, 1998, 2002, 2004, 2006, 2008, and 2011). Details of the study design and data collection procedures used in ANC-HIV-SS program have been described previously.[24, 41, 157, 182] Briefly, ANC-HIV-SS is a series of cross-sectional surveys done every 2 to 4 years, focused on estimating and monitoring HIV prevalence trends among pregnant women seeking antenatal care in Zambia: 1994, 1998, 2002, 2004, 2006, 2008, and 2011. The number of consistently used sentinel sites have only varied slightly over time: 22 sites from 1994 to 2002 and 24 sentinel sites from 2004 to 2011.

4.2.2. Study population and inclusion criteria

Pregnant women were eligible for recruitment if they made the first contact with antenatal care clinic for the current pregnancy during 4-month survey period at a specific health center designated as a sentinel site for ANC-HIV-SS.

4.2.3. Target sample size

For ANC-HIV-SS rounds conducted between 1994 and 2008, each of the sentinel sites was expected to recruit at least 500 pregnant women, based on an expected HIV prevalence of 20% and desired precision of 0.35% at 95% confidence level. Most sites attained the target sample size, except sites located in sparsely populated areas (e.g.,

Ibenga in Copperbelt Province). Urban sentinel sites in Lusaka and Ndola located in densely populated areas were assigned larger expected target sample size of at least 800 pregnant women per site. However, a protocol change in 2011 mandated the recruitment of a minimum of only 360 pregnant women per site.[182] Table 4.2 presents summary of pregnant women recruited in ANC-HIV-SS between 1994 and 2011.

Table 4.2. Summary of number of pregnant women recruited in ANC-HIV-SS in Zambia between 1994 and 2011

Number of women recruited in Zambia Antenatal Clinic Surveillance for HIV and syphilis for specific years		
Year	Sample Size	Number of participating sentinel sites
1994	11592	27
1998	12,017	22
2002	13,111	24
2004	12,404	24
2006	13,260	24
2008	13,370	24
2011	8881	24

4.2.4. Criteria for site selection for the ANC-HIV-SS

To achieve countrywide geographic coverage, at least two health centers in each of the nine provinces were used as sentinel sites for ANC-HIV-SS, an urban site (i.e., situated in the headquarter town of the province) and rural sentinel site (Table 4.3). Urban-located sentinel sites were conveniently selected, and represent urban settings within provinces, whereas rural-located sentinel sites were randomly selected from health centers within each province.[183] Further guiding principle for site a site to qualify as a sentinel site was the site capacity to recruit the targeted number of pregnant women (~500) within the 4-month survey period.

4.2.5. Response rate of the survey

Resting on the premise that all pregnant women who seek antenatal care provided a venous blood specimen for routine screening of syphilis on their first antenatal care visit (i.e., Ministry of Health care package for pregnant women in Zambia), it was assumed that all eligible pregnant women provided a blood sample, part of which was used for ANC-HIV-SS reporting. Following non-probability convenient sampling strategy, eligible pregnant women were recruited in a consecutive manner until targeted sample size (i.e., 500) of pregnant women per site was attained and/or when 4-month survey period elapsed.

Table 4.3. Sentinel sites used for data collection for ANC-based HIV and syphilis sentinel surveillance program: 1994 to 2011.

	Province	Sentinel sites	
		Rural	Urban
1	Central	Kapiri-Mposhi, Serenje	Kabwe
2	Copperbelt	Ibenga	Ndola
3	Eastern	Minga	Chipata
4	Luapula	Nchelenge, Kasaba	Mansa
5	Lusaka	Luangwa	Chelstone, Chilengi, Kalingalinga, Matero
6	Northern	Isoka	Kasama
7	Northwestern	Mukinge, Kabompo	Solwezi
8	Southern	Macha	Livingstone
9	Western	Kalabo	Mongu

4.2.6. Sociodemographic variable collected via questionnaire [i.e. 1994 to 2008] and abstracted from pregnant woman routine card [i.e. 2011]

Study nurses trained on the ANC-HIV-SS protocol identified, recruited and interviewed eligible pregnant women in a chronological manner. Questionnaire data (e.g., age and education) and blood specimen for HIV serostatus determination were collected on the first visit to the antenatal clinic for routine care of the current pregnancy (i.e., 1994 through 2008). ANC-HIV-SS data for 2011 were abstracted from routine antenatal medical record card of each eligible pregnant woman, consequently limiting

data collection to variables that are collected routinely (e.g., age, number of children birth by pregnant woman).

4.3. Serological HIV testing of plasma/serum specimens for ANC-HIV-SS

The blood specimen from each eligible pregnant woman was divided into two containers: one container bore the name of the pregnant woman, and was used for routine reporting of syphilis, and the other container was marked with a distinctive survey identify number (ID), and was used for survey reporting. Plasma/serum specimens collected in all ANC-HIV-SS rounds were tested for presence of HIV specific antibodies according to W.H.O guidelines on anonymous and unlinked HIV antibody screening.[157, 182, 184-186]

The HIV-1 test assays used across the years were not consistent but an identical three-stage (i.e., screening, confirmatory, and tie breaking) survey-specific HIV testing algorithm was used in all seven survey rounds to assure trustworthiness of HIV test results.[157] First, on-site HIV screening of serum/plasma specimens by site laboratory technician using rapid HIV antibody test (e.g., Determine^{TM} HIV in 2004, 2006 and 2008). Second, plasma/serum for the survey were frozen and transported to Tropical Diseases Research Centre (TDRC) and University Teaching Hospital (UTH) for confirmatory and quality control HIV testing. [157]

4.3.1. Criteria for determining HIV serostatus of survey specimen

To limit misclassification errors and assure trustworthiness and validity of the HIV serostatus of survey specimens, a pre-specified proportion of plasma/serum specimens

(e.g., 10% in 2004, 2006, 2008, and 2011 surveys) classified as HIV-seronegative specimens at the site HIV screening were further tested according to the pre-specified survey quality control HIV testing protocol, described in details in prior reports.[24, 40, 56, 182, 183]

Confirmatory HIV testing was performed on all survey specimens that tested positive for HIV specific antibodies (i.e., using rapid HIV test) during the site-screening. The specimen was considered seropositive for HIV specific antibodies if both the site-based screening HIV test result and the reference laboratory confirmatory test HIV result indicated presence of HIV specific antibodies (i.e., positive). Specimens classified as HIV seronegative during site-based HIV testing, and not selected into the 10% quality control testing sample were reported as HIV seronegative. Where site-based HIV screening test result and reference-laboratory HIV confirmatory test result were discrepant, a different test assay (i.e., tie-breaker test such as a Western blot) was performed and tie-breaker test result reported as final. [40, 56, 157, 182, 183] Table 4.4 summarizes the various commercial HIV test assay used in ANC-HIV-SS between 1994 and 2011. Table 4.4 presents the HIV test assay used in the seven ANC-HIV-SS rounds between 1994 and 2011.

4.3.2. Quality control HIV testing to limit misclassification of serostatus

Table 4.4. Summary of HIV assays used for HIV screening, confirmatory and quality control and tie-breaker testing for specific sentinel surveillance rounds

Survey year	Screening	Confirmatory	Tie-breaker
1994	Capillus HIV-1 & 2 test (Cambridge Biotechnology, Galway, Ireland)	Wellcozyme HIV Recombinant HIV-1 (Murex, Johannesburg, South Africa)	Bionor HIV-1& 2 (BIONOR, AS, Skien, Norway)
1998	Capillus HIV-1 & 2 test (Cambridge Biotechnology, Galway, Ireland)	Wellcozyme HIV Recombinant HIV-1 (Murex, Johannesburg, South Africa)/ (Murex Diagnostics LtD., UK)	Bionor HIV-1& 2 (BIONOR, AS, Skien, Norway)
2002	Capillus HIV-1 & 2 test (Cambridge Biotechnology, Galway, Ireland)	Wellcozyme HIV Recombinant HIV-1 (Murex, Johannesburg, South Africa)	Bionor HIV-1& 2 (BIONOR, AS, Skien, Norway)
2004	Abbott Determine® HIV	ELISA HIV-1 & 2 (Murex)	Bionor HIV-1/HIV-2
2006	Abbott Determine® HIV	ELISA HIV-1 & 2 (Murex)	Bionor HIV-1/HIV-2
2008	Abbott Determine® HIV	ELISA HIV-1&2 (Murex)	Bionor HIV-1/HIV-2
2011	Vironostika® anti-HIV plus	Enzygonst® HIV integral	Western blot 2.2 (MP Diagnostics)

Summary of HIV assays used for HIV screening, confirmatory and quality control and tie-breaker testing for specific sentinel surveillance rounds. Only specimens with discrepant screening and confirmatory or quality control HIV test results were tested with Tie-breaker HIV test. 5-10% of HIV-seronegative specimens were subjected to quality control testing. Tie-breaker is used to report final HIV serostatus when the screening and confirmatory test results are discrepant. The sensitivities and specificities of the assays were the over-riding characteristics in determining which assays to use for surveillance data in each of the seven survey rounds. Details of HIV testing algorithms may be obtained from survey protocols, and prior publications.[24, 40, 56, 182, 183]

4.4. Data management for ANC-based surveillance in Zambia

To assure data quality, a nurse-supervisor designated at each sentinel site monitored data collection and recording procedures (i.e., review recorded data, identify errors and inconsistencies, and implement measures to avert recurrence), and also liaised with the reference centers, TDRC for the northern zone and UTH for southern zone. TDRC and UTH staff conducted periodic supervisory visits to monitor survey progress

and data quality, and address operational concerns. Filled questionnaires and laboratory data records (i.e., HIV and syphilis) were transported to TDRC for centralized data entry in an EPI-INFO database (Center for Disease Control and Prevention, GA, USA) at the end of the survey period.

4.4.1. HIV test results and sociodemographic data were double entered at TDRC

ANC-HIV-SS data were centrally entered at TDRC. Two data databases (i.e., one for sociodemographic information and another for HIV serostatus data) were created using the most current version of EPI INFO software during each survey round.[157, 187] Data were double entered to curb typographic errors, and the two databases subsequently merged via the distinctive ID numbers. The sociodemographic and serological databases were merged, and standard data procedures (e.g., data cleaning, cross-checks) applied by the study teams to cross-check for completeness and consistency at all stages of data management (i.e., before, during, and data set merging).[157, 182]

4.4.2. Preparation of data for secondary data analyses for the dissertation

My dissertation analyses relied on ANC-HIV-SS data collected between 1994 and 2011 (i.e., repeated cross-sectional surveys in i.e., 1994, 1998, 2002, 2004, 2006, 2008 and 2011). Each survey round had a specific data set: therefore seven data sets were merged to facilitate my planned analyses (e.g., HIV prevalence trends between 1994 and 2011). Because I had planned to use regression models with survey year as exposure variable and HIV prevalence as an outcome variable, I created a new continuous variable

(i.e., survey year) in the merged data set. All the analyses were restricted to pregnant women aged 15 to 44 years to ensure comparability across survey year (i.e., 2004 and 2006 data were restricted between age 15 and 44). Standard data cleaning procedures (i.e., consistency, completeness and plausibility checks) were conducted prior to performing the planned analyses.

Briefly, the following procedure was followed in merging the seven data sets (i.e., 1994, 1998, 2002, 2006, 2008 and 2011). First, I identified relevant variables for the analysis. Second, I identically named variables containing same type of information (e.g., age named as M_AGE in all data sets. Third, I created a new variable for survey calendar year. Fourth, I merged data sets that bore similar variable names. Data management were performed using R-statistical software version 3.0, which been saved for future updating, reference and auditing.[188]

4.4.3. Missing data in the ANC-SS (1994 through 2011)

The merged data set had two types of missing values: (1) data missing because pregnant woman did not provide information and (2) data missing because questions were not asked in a particular survey year (e.g., spousal age in 1998 and 2004, and educational attainment in 2011).[189-191] Figure 4.1 shows the follow chart of how analytic samples of ANC-HIV-SS data were created.

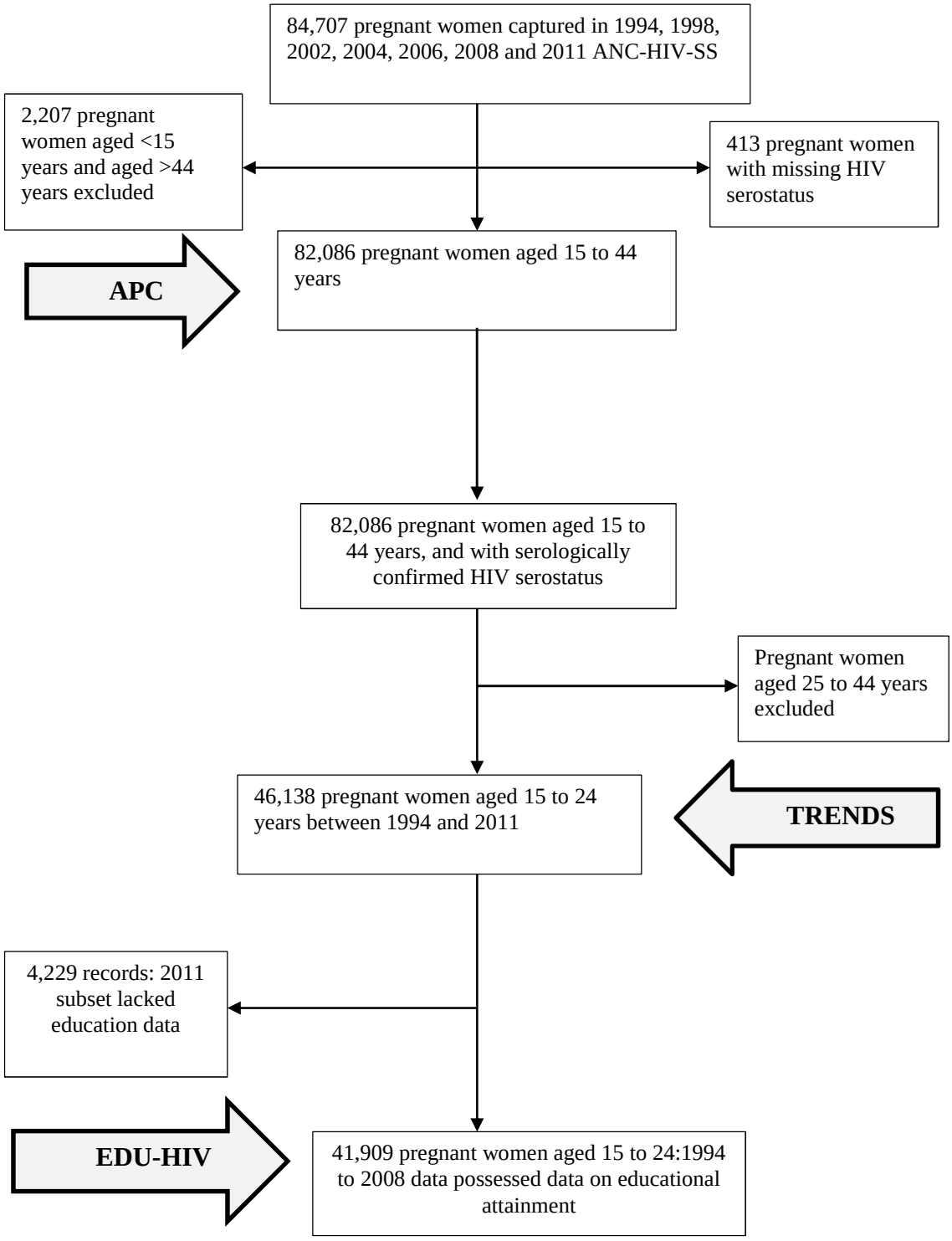

Figure 4.1. Flowchart of eligibility criteria of the analytic sample of pregnant women from seven ANC-HIV-SS in Zambia for answering the research question for my PhD dissertation. The grey arrow shows the records used for a specific analysis: APC for age-period-cohort analysis; TRENDS for trends in HIV prevalence and EDU-HIV for the association between educational attainment and HIV prevalence.

4.5. Key variables used in analyses for PhD dissertation

The analyses focused on variables that were collected in the all the seven rounds of the ANC-HIV-SS conducted between 1994 and 2011 (e.g., age, location of site and the number of children birthed by a woman).

4.6. Primary outcome variable: HIV serostatus

The outcome variable was HIV serostatus (i.e., dichotomously defined as HIV seropositive (i.e., presence of HIV-specific antibodies) and HIV seronegative (i.e., non-detection of HIV specific antibodies).

4.7. Dependent variables

1.1.1 Age

Self-reported pregnant woman's age captured as full years lived at the time of the first visit to the ANC clinic for the current pregnancy.

4.7.1. Survey year

Calendar survey year coded as a continuous variable corresponding to the year in which the survey was conducted.

4.7.2. Parity

Parity was defined as the self-reported number of children birthed by pregnant woman: three groups were defined: no child, one child and ≥2 children.

4.7.3. Residence

Pregnant women's area of residence (i.e., urban or rural) was classified according to the urban-rural classification of site location areas by the Government of the Republic of Zambia (GRZ). Pregnant women recruited at a specific site were assumed to have come from the catchment areas for the participating health center.

4.7.4. Educational attainment

Educational attainment was measured as the schooling years completed by the pregnant woman (1994 through 2008 rounds of ANC-HIV-SS). Educational attainment values were truncated at 17 schooling years and educational attainment >17 schooling years coded as 17 schooling years completed. Education-HIV association analyses were restricted to earlier survey data that included educational attainment data.[192]

4.7.5. Human subject considerations

In-country ethics committee approved primary data collection for ANC-HIV-SS as a public health surveillance system. Vanderbilt University Institutional Review Board approved the secondary data analysis for the PhD dissertation.[157]

4.8. Components of the PhD dissertation work

My dissertation work had four components guided by two specific aims. First, I examined the association between educational attainment and prevalent HIV infection among pregnant women via a meta-analysis of peer-reviewed literature based on data collected in SSA. Second, I described trends in HIV prevalence using repeated cross-

sectionally collected ANC-HIV-SS data from Zambia (i.e., 1994 through 2011). Third, I assessed age, period and birth cohort effects on HIV prevalence between 1994 and 2011. Fourth, I investigated the association between educational attainment and HIV between 1994 and 2008 using ANC-HIV-SS data. Data analysis were performed using R version 3.0 statistical Program (R foundation, available at http://www.r-project.org), and part of the statistical analyses for the fourth component were conducted using STATA version 12.0 (StataCorp LP, College Station, TX, USA).[188, 193]

CHAPTER 5

ASSOCIATION BETWEEN EDUCATIONAL ATTAINMENT AND HIV INFECTION AMONG PREGNANT WOMEN IN SUB-SAHARAN AFRICAN: META-ANALYSIS BASED ON OBSERVATIONAL STUDIES

5.1. Background

5.1.1. Serious HIV burden in sub-Saharan Africa

Human immunodeficiency virus (HIV) , the virus that causes acquired immunodeficiency syndrome (AIDS), remains a serious public health concern worldwide.[10] Between 1990 and 2011, the number of people living with HIV increased fourfold from 8 million to 34 million.[4, 8, 10] The HIV burden is severest in sub-Saharan Africa (SSA) (i.e., accounting for 23.5 million (69%) of global HIV infections in 2011) where for more than 25 years, the yearly estimates of incident HIV infections and HIV burden have been highest.[10]

Of the 2.7 million new HIV infections estimated to have occurred worldwide by the end of 2011, an estimated 70% (1.8 million) occurred in SSA. The HIV epidemic in most countries in SSA is generalized (HIV prevalence beyond 1% in the general population), and HIV burden varies widely between and within countries.[10] For example, HIV prevalence in most of Central and West Africa is comparatively low, though high compared to countries outside Africa (i.e., ranging from 0.9% in Senegal in 2009 to 5.3% in Cameroon in 2009) but lower compared to burden in most southern SSA countries. HIV prevalence among adults aged ≥15 years in most southern SSA countries is > 10% and represents the region of the world's with the most intense transmission.[4, 23]

5.1.2. Risk factors for HIV infections are diverse and interrelated

To understand the dynamics of the HIV epidemic in SSA, several factors (e.g., urban-rural residence, age, educational attainment, socioeconomic status, sex, spouse age, marital status, and parity, young age at first sexual intercourse, spousal characteristics, multiple lifetime partners and unprotected heterosexual intercourse) have been investigated as possible risk factors for the growth of the HIV epidemic culminating in both convincing and divergent findings.[40, 56, 118, 130] Specifically, educational attainment is a key factor in several health and economic outcomes, and its association with prevalent HIV infections has been examined broadly in a number of epidemiologic studies, but the findings have been divergent: negative association, no association to positive association.[48, 52, 80, 81, 141, 194-197].

5.1.3. Education's plausible protective HIV effect is inconsistent in the literature

Whereas evidence from chronic diseases research in western countries (e.g., diabetes, asthma) has consistently found low education, low literacy and low health knowledge to be associated with poor health outcomes, findings from studies that examined the association between educational attainment and HIV infection have yielded mixed findings.[126] Hargreaves & Glynn (2002)'s systematic review of the literature on the association between educational attainment and prevalent HIV infection based on 27 studies comprising 27 studies published by 2001, and reported a non-significant association between educational attainment and HIV infection in seven studies.[139] Further, most studies conducted in SSA in earlier years of the epidemic (i.e., 1980s and 1990s) found a positive association between high educational attainment and HIV infection.[52, 80, 81, 196] On the other hand, some studies had found significant

protective association between education attainment and HIV infection among young girls in Zimbabwe [138], men and women in Uganda[143], and women in Cameron and Benin[139].

A subsequent systematic review by Hargreaves et al (2008) focused on studies conducted in SSA published between 1987 and 2003 advanced that odds of HIV prevalence were lower among educated people than among less educated people. The protective effect of higher educational attainment were prominent in later years of the HIV epidemic, post 1990s.[81] The protective effect of education is highly anticipated, but review of literature published 1985 and 2012 indicate an ally of studies that have reported mixed findings on the association between educational attainment and prevalent HIV infection.[50, 52, 81, 183, 196] Encouragingly, recent reports have indicated marked decline in HIV prevalence among educated people, specifically in urban areas of most countries in SSA with historically high HIV burden (e.g., Zambia).[24, 36, 52, 196]

Resting on the assumption that educated people are more likely to adequately process information about risk factors for HIV infection, and may be more willing to respond to preventive interventions in a positive manner than would the less educated people, HIV prevalence might be expected to decline more steeply among educated persons than less educated people.[40, 50, 56] For example, the demographic and health survey (DHS), a country-wide cross-sectional study conducted in Zambia in 2007 revealed an increasing prevalence of HIV with increasing level of education among women (10.8%), primary; (15.8%) secondary; (17.4%) and post-secondary (21.3%).[29, 36, 51, 140]

5.2. Overview of the educational attainment and HIV infection relationship

Because the association between educational attainment and prevalent HIV infection seem to vary over time, continued assessment the association between educational attainment and risk of HIV infection might be a key step in guiding packaging and targeting HIV prevention and treatment interventions, and identifying groups at higher risk of HIV. [24, 42, 43]. For example, higher educational attainment has been linked to better treatment outcome among people living with HIV and receiving combinational antiretroviral therapy (cART).[82, 198] Recent studies have suggested that favorable treatment outcomes among HIV-infected people receiving cART are more likely among educated than among less educated people. Therefore, the association between educational attainment and prevalent HIV infection is likely to be plagued by survival bias, especially if people aged >25 are used in the analyses. Consequently, the relationship between educational attainment and prevalent HIV infection may include survival bias, unless incident cases data are used for the analysis. [82, 83, 196]

5.2.1. Meta-analysis of research findings on the association between educational attainment and prevalent HIV infection

This study was crafted to conduct a meta-analysis of the more recent peer-reviewed literature regarding the association between educational attainment and prevalent HIV infection, based on data from studies conducted between 2000 and 2012 in SSA. No meta-analysis has been conducted to association between educational attainment and prevalent HIV infection.

Because of anticipated differences in the study procedures, populations and data analysis methods, heterogeneity of eligible studies for planned meta-analysis was

anticipated. DerSimonian-Laird random effect model was applied to account for possible heterogeneity concerns across the eligible studies in the meta-analysis to investigate whether greater than primary school education attainment was associated with increased odds of prevalent HIV infection.

5.2.2. Restricting meta-analysis to studies conducted among pregnant women in SSA may limit heterogeneity

Lumping studies conducted in diverse populations' together increases diversity and, perhaps, generalizability, and may be handled with a pre-specified subgroup analysis. However, pre-specified inclusion criteria focused on a specific population can increase the likelihood of finding more valid and focused answers to specific research questions than when looser inclusion criteria are applied.[199-201] To limit the variability and diversity of the study populations from whence data for the current meta-analysis were drawn, I restricted eligible studies to those conducted in pregnant women.

Consequently, studies included in the current meta-analysis were drawn from subpopulation that is highly generalizable to all women of child-bearing age and also minimized the variability of estimates compared to the variability that would arise if studies were drawn from more diverse populations.[201] My research question was to examine whether ≥ primary school education among pregnant women in SSA was associated with increased odds of prevalent HIV infection.

5.3. Methods

5.3.1. Overview of search criteria

The development of the search criteria to identify studies that examined the association between educational attainment and prevalent HIV infection among women in SSA was guided by the PICOTS framework.[201-203] The PICOT framework directs researchers to pre-specify the population, intervention or exposure, comparison groups, time frame, and research designs of studies that would be eligible for the planned meta-analysis.[201-203] The present meta-analysis focused on observational studies indexed in the electronic databases of scientific literature that were conducted between January 2000 and December 2012. Randomized studies are rare in this field of research due ethical and practical challenges.[58, 140, 201]

5.3.2. Study outcome, exposure variable and measure of association

The primary exposure variable for the current meta-analysis was the pregnant woman's self-reported educational attainment. Lower educational attainment was defined as less than primary school educational attainment. To be eligible for inclusion in the current meta-analysis, the study's definition of lower educational attainment needed to match closely with the inclusion criteria definition of lower educational attainment (i.e., <primary educational attainment). The exposure category was defined as ≥ primary educational attainment. The outcome variable was defined as serologically confirmed HIV serostatus (i.e., HIV seronegative or seropositive) based on any standard, validated assay (e.g., rapid test, ELISA, and/or Western blot).

5.3.3. Measure of association for the meta-analysis

The measure of association for the current meta-analysis was the odds ratio. The odds ratios and 95% confidence intervals (CI) were estimated from counts and proportions of prevalent HIV infections for studies that did not report odds ratio but reported prevalence ratios or counts and proportion.[201, 204]

5.3.4. Study site locations, study design and time frame for eligible literature

The present meta-analysis was restricted to observational studies (cross-sectional, case-control or cohort studies) of which > 95% were cross-sectional studies conducted in any of the 47 countries in SSA as defined by the Library of Congress (http://www.loc.gov/rr/amed/guide/afr-countrylist.html, accessed June 9, 2013). We also considered the newly formed country South Sudan as part of SSA, though it is not listed on the Library of Congress web site. Peer-reviewed articles indexed in electronic databases (see below) between January 1, 2000 and December 30, 2012.

5.4. Literature search strategy

To minimize bias that could have arisen from omission of germane studies, a comprehensive and systematic literature search was conducted to identify eligible studies in the following databases: MEDLINE via PubMed, Academic Search Premier, PsycINFO, Embasse, Web of Science, African Journal Online and Africa Index Medicus.[200] With the help of the librarian at Eskind Library [Ms. Marcia Epelbaum], search criteria that maximized the yield of potentially eligible studies were developed. Pre-specified search criteria were applied to enhance consistent, reliable, rigorous, and reproducible retrieval of information.[200, 205, 206]

5.4.1. Literature search and search terms

Systematic searches, without language restrictions were performed between February 2012 and 30th March 2013 to identify studies that met the pre-defined inclusion criteria from the following databases: PubMed/MEDLINE, Web of Science, Africa Journal Online and Embasse. To accomplish a comprehensive retrieval of literature on the association between educational attainment and prevalent HIV infections, medical subject headings (MeSH) were used in crafting search strategies. For example, six search criteria were applied for harvesting articles indexed in MEDLINE and the other databases. Search criteria comprised medical subject heading, subject headings, and key words relevant to the current research question. Because articles identified from other databases were subsets of articles indexed in MEDLINE, an overview of search criteria applied in MEDLINE database to identify eligible articles are presented in the following sections.

5.4.2. First search criteria

Two hundred and thirty (230) articles were identified from MEDLINE using the first search: ((("Educational Status"[MeSH]) AND "HIV Infections"[MeSH]) AND "Africa South of the Sahara"[MeSH].

5.4.3. Second search criteria

Thirty nine articles were identified from MEDLINE using the second search: ((("Educational Status"[MeSH]) AND ("HIV"[MeSH] OR "HIV Infections"[MeSH])) AND "Africa South of the Sahara"[MeSH]) AND pregnancy [MeSH].

5.4.4. Third search criteria

The third search criteria harvested 1397 articles from MEDLINE using third search criteria: ("educational status"[MeSH Terms] OR ("educational"[All Fields] AND "status"[All Fields]) OR "educational status"[All Fields] OR ("educational"[All Fields] AND "attainment"[All Fields]) OR "educational attainment"[All Fields]) AND ("hiv"[MeSH Terms] OR "hiv"[All Fields])

5.4.5. Fourth search strategy

The fourth search criteria yielded 59 articles from MEDLINE using the fourth search criteria: ((("educational status"[MeSH Terms] OR ("educational"[All Fields] AND "status"[All Fields]) OR "educational status"[All Fields] OR ("educational"[All Fields] AND "attainment"[All Fields]) OR "educational attainment"[All Fields]) AND ("hiv infections"[MeSH Terms] OR ("hiv"[All Fields] AND "infections"[All Fields]) OR "hiv infections"[All Fields])) AND ("pregnancy"[MeSH Terms] OR "pregnancy"[All Fields])) AND ("africa"[MeSH Terms] OR "africa"[All Fields])

5.4.6. Fifth search strategy

Based on the fifth search criteria, 331 articles were identified from MEDLINE: (("educational status"[MeSH Terms] OR ("educational"[All Fields] AND "status"[All Fields]) OR "educational status"[All Fields] OR ("educational"[All Fields] AND "attainment"[All Fields]) OR "educational attainment"[All Fields]) AND ("hiv infections"[MeSH Terms] OR ("hiv"[All Fields] AND "infections"[All Fields]) OR "hiv infections"[All Fields])) AND ("africa south of the sahara"[MeSH Terms] OR ("africa"[All Fields] AND "south"[All Fields] AND "sahara"[All Fields]) OR "africa

south of the sahara"[All Fields] OR ("sub"[All Fields] AND "saharan"[All Fields] AND "africa"[All Fields]) OR "sub saharan africa"[All Fields])

5.4.7. Sixth search strategy

This search strategy returned 353 articles from MEDLINE: ("Educational Status"[MeSH] OR "educational attainment"[tiab] OR education*[tiab] OR "literacy"[tiab]) AND ("HIV"[MeSH] OR "HIV Infections"[MeSH] OR HIV[tiab]) AND ("Africa South of the Sahara"[MeSH] OR Africa[tiab]) AND ("pregnancy"[MeSH Terms] OR pregnancy[tiab])

5.4.8. Search for related articles in databases

To capture studies that could missed by the "MeSH" terms alone; the "related study" feature in PubMed was used to identify more articles from four highly relevant articles (i.e., Hargreaves and Glynn (2002), Hargreaves et al (2008), Sandoy et al (2006) and Johnson et al (2009).[52, 81, 141] For example using "related article" feature in PubMed based on the study by Johnson et al (2009) and Hargreaves et al (2008) yielded a further 803 articles related to Johnson et al (2009) and 69 articles related to Hargreaves et al (2010) that were then reviewed for eligibility for inclusion into my study.[141, 146]) The first, second and corresponding author's names were also used to identify more articles. Additionally, published systematic reviews, meta-analysis and editorials were searched to identify potentially eligible articles.[50, 52, 81]

5.4.9. Cross-references and specialized journal searches

To identify studies that might have been missed during the electronic searches, bibliographic sections of all eligible articles were examined to limit the possibility of selection bias that could arise if important articles were left out of the study. Further searches of the literature were conducted in HIV/AIDS-specific journals (e.g., *AIDS, AIDS Care, JAIDS*).[207].

5.5. Grey literature

The so-called "grey literature" may provide further information and limit publication bias, but my study was restricted to peer-reviewed articles because of validity concerns from reports that were not peer-reviewed.[200, 206, 208]

5.6. Exclusion criteria

Studies that did not meet the inclusion criteria were excluded. First, because conduct of meta-analysis requires estimates of odds ratios and standard errors, studies that did not report odds ratios and did not provide enough raw data to facilitate estimation of the odds ratios and standard errors for the an association between educational attainment and HIV were excluded. Second, I screened for studies that adjusted for covariates regarded as intermediates in the relationship between educational attainment and prevalent HIV infection because adjustment for a variable assumed to be on the causal pathway may adjust away the association of interest. Third, I excluded studies whose definition of educational attainment did not fit closely with the meta-analysis definitions of exposure and outcome variables. Fourth, I excluded studies that based their analyses on data collected before year 2000 but published between 2000 and 2012.

5.6.1. Screening articles for eligibility

Potentially eligible articles harvested using the pre-specified search criteria were screened according to the pre-specified inclusion criteria for the current meta-analytic study. First, all the titles of citations were read to judge their relevance to the research question. Second, I read all potentially eligible abstracts, and subsequently retrieved full articles for further reading and inspection for relevant data. Eligible articles were assigned a distinctive identity numbers for systematic and convenient tracking.

5.6.2. Data coding, and abstraction

I extracted relevant data from eligible studies. The following data were abstracted: name of first author; year of publication of the study; study design; primary exposure level; outcome measure; publication year of the study; country; measure of association and 95% CI; number of HIV-infected woman among lower and higher educated women to aid calculation of the odds ratio and standard errors and year in which the study was conducted. Further data on sampling strategy (e.g., random or non-probability), study size, and study year were abstracted.

5.6.3. Data management and statistical analysis

The bibliographic data of eligible reports were filed in an EndNote library. Data relevant to the current meta-analysis were entered into a standard electronic data form in a data entry screen in Epi Info™ 3.5.1, a public domain statistical software developed by Centers for Disease Control and Prevention.[187] The study data set created in Epi Info 3.5.1 was imported in R-statistical and computing software for subsequent data analysis.[188]

First author's name and corresponding author's names; publication year of the article and study site were used to screen for duplicate articles among potentially eligible studies. The current report with permission from my advisor is based on data I singly extracted, and this has been noted as a limitation. Cognizant of the biases that may arise when data are extracted by one person, plans are underway for an independent person, yet to be identified to abstract 20% of the randomly selected eligible articles. The reporting format for the current meta-analysis was guided by the PRIMA guidelines (Preferred Item Reporting for systematic review and Meta-analysis).

5.6.4. Assessment of methodological quality

Findings of a meta-analysis are trustworthy to the extent that primary studies included in the analytic sample are validity. Eligible studies were examined for possible biases (i.e., study design, data collection procedure and statistical methods), although in general the risk of bias was high given the observational nature of the eligible studies. Study design features that could have compromised the primary studies included: clarity of definition of educational attainment; magnitude of the sample size; clarity of the inclusion and exclusion criteria; ascertainment of study variables; selection of study participants (i.e., sampling approaches whether convenient or probability samples or probability sampling); clarity of multivariable modeling process (e.g., excluding studies that adjusted for variables on causal pathway.[204]

Eligible studies for the current meta-analysis were examined by one person [Webster Kasongo]. Further review of the study reports by a second person will be conducted prior to peer-reviewed publication. Because the definition of the primary

exposure and outcome variable for the current meta-analysis were fairly straightforward, and the need to consult a second opinion did not arise, my advisor has approved my presentation of these findings for my doctoral dissertation, given time constraints.[209] This is not to disregard enhancement in the validity of meta-analysis findings derived from cross-checking by a second reviewer.

5.6.5. Statistical analyses

Data analyses were conducted using Metafor package in R software version 3.0 and to a limited extent (i.e., trim and fill analysis) Stata™ software version 12.1.[188, 193, 210] Data were converted into stata version 12.1 SE format via STAT-transfer software. Data cleaning, consistency and completeness checks were performed using R software version 3.0.[188, 193, 210] All analyses were conducted with a 95% confidence level.

The distribution of sample sizes of eligible were examined and range computed.[211] Log odds ratios (OR) and corresponding standard errors at 95% confidence level were computed to enable the meta-analyses.[211]

5.6.6. Computation of log odd ratios and 95% CI

The measure of association for the current meta-analysis was the odds ratio. Therefore, log odds ratios and standard errors at 95% confidence level were computed from odds ratios and 95% CIs. Where the odds ratio was not directly reported, log odds ratios and corresponding standard errors at 95% confidence levels were computed from cell counts standard formulas shown below. Briefly, the odds ratio were computed by dividing the odds of being HIV-infected among pregnant woman who reported

attainment of greater than primary school education by the odds of being HIV-infected among pregnant women who reported educational attainment that was less than primary school.[204, 211] The standard error of the log odds ratio was computed as the square root of the sum of the inverses of cell counts as shown in Equation 1.[204, 211]

Table 5.1 shows the positioning of the cell counts for computing log-odds ratio and standard errors where “a”, “b”, “c”, and “d” are cell counts that were used in equation 1

Table 5.1. Counts of pregnant women by HIV serostatus and educational attainment category

	HIV seropositive	HIV seronegative	
>Primary education	a	b	
≤Primary education	c	d	
Total			

$$\text{OR} = \left(\frac{a*d}{b*c}\right) \text{ and Standard Error of Log OR} = \sqrt{\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right)}$$

5.6.7. Log odd ratio and standard error calculation

Log-odds ratio was computed via natural logarithm of odds ratio value. Standard errors of log-odds ratios for each report were computed using the equation indicated below programmed in statistical and computing software R-version 3.0.[204, 211]

$$\text{Standard Error of Log OR} = \left(\frac{\text{logOR}_{\text{Upper confidence limit}} - \text{logOR}}{1.96}\right)$$

To detect outlying log-odds ratios and standard error of log-odds ratio that represent a source of undesirable influence on the estimates from the meta-analysis, I used box plots.

5.7. Meta-analysis conducted using log odds ratios and their standard errors

Because the studies were drawn from different settings, the random effects model based on the DerSimonian and Laird method was used to conduct the meta-analysis of odds ratios from eligible studies. [212-214] The random effects model enables an investigator to account for between-study variability in the measure of association of the relationship between educational attainment and prevalent HIV infection.[201] The studies included in the meta-analysis were by no means indistinguishable but are comparable enough to be combined. Use of the random effects model enabled accounting for within and between study variability (i.e., statistical heterogeneity among studies). [206]. Sensitivity analyses were performed to assess the robustness of the meta-analyses with and without outliers for log-odds ratios and corresponding standard errors

5.7.1. Justification for using random effect model for the current meta-analysis

Meta-analysis studies are often plagued with between-study heterogeneity to varying degrees because “primary studies” included in the meta-analysis differ in the subject recruitment procedures, laboratory or serological methods for HIV diagnosis, overall study quality and trustworthiness. Because within-study measurements tend to be correlated beyond what would be expected for between-studies measurements, a need arise to account for possible heterogeneity, which may harm inference if present.[215] Rather than considering heterogeneity as a potential problem that may hamper meta-analyses, use of the random effect model to perform a meta-analyses was pre-specified given the anticipated heterogeneity.[201]

5.7.2. I-square, Q-statistic, and tau-square to assess between-study heterogeneity

Three measures of heterogeneity were estimated to assess between-study variability among eligible studies. First, I evaluated between-study variability among studies that were eligible studies for the meta-analysis by computing the I-squared statistic. I-squared is the ratio of true heterogeneity to total observed variation. Second, high I-squared statistical value might be substantively less meaningful if an estimated tau-squared value were very small, therefore I computed tau-squared to also assess between-study variance. Third, the Q-statistic, a descriptive statistic that tests the null hypothesis that studies share a common effect size was reported. Collectively, three measures of parameters were applied to assess the extent of heterogeneity (I-squared, Q-statistic and tau-squared) and whether the detected heterogeneity was substantively meaningful.

5.7.3. Visual assessment of publication and small study size bias: funnel plot

To assess potential publication bias and/or overly influential small studies, standard error of log odds ratios for each study were plotted against log-odds ratios and odd ratios to generate funnel plots.[201] Asymmetric funnel plots signaled possible but not definite presence of publication bias and/or small sample bias.[201] To evaluate potential publication bias, Egger's tests were performed. Publication bias and small study bias were said to be present if the p-value for the Egger's test was < 0.05. [216]

5.7.4. Trim and fill method for assessment of small study bias

To assess whether publication and small study bias influenced the estimates from meta-analysis, trim and fill analysis strategy for assessment of bias was applied to the analytic sample of eligible studies. Trim and fill analyses generates a funnel plot that

includes estimates from iterative non-parametric method synonymous with conducting sensitivity analysis on the basis of the assumption that studies with the most extreme measures of association were left out of the analysis, and are subsequently imputed and included in the trim and fill analysis.

5.7.5. Meta-regression to assess variation of the association between educational attainment and HIV infection

Further motivation for the current meta-analysis was to assess whether the association between educational attainment and prevalent HIV infection has changed over the course of the HIV epidemic. Consequently, study year was applied as an explanatory variable in a meta-regression to evaluate whether the association between educational attainment and HIV infection has varied using data collected in SSA between 1st January, 2000 and 31st December, 2012.[5, 201, 210]

5.8. Results

5.8.1. Eighteen studies met the pre-specified inclusion criteria

Figure 5.1 outlines the process followed in identifying eligible citations for the current meta-analysis: from screening of abstract titles to selection for inclusion as well as specific reasons for inclusion or exclusion of citations. A total of 3020 studies were generated by the search criteria, 53 were identified as potentially eligible but only 8 study reports met the pre-specified inclusion criteria for the current meta-analysis. Twenty-nine reports were excluded because the studies were conducted in non-pregnant population. A further 6 reports were dropped from the 24 reports that met the inclusion criteria because the data were collected prior to 2000, leaving 18 eligible studies for inclusion into the meta-analysis (Figure 5.1).

5.8.2. Eligible studies were dominantly were cross-sectional surveys

Seventeen studies (94.4%) included in the meta-analysis collected data cross-sectionally, and one study collected data via a case-control design. Only one study report among the 18 studies considered for the meta-analysis was specifically crafted to examine the association educational attainment and prevalent HIV infection.[141] The search criteria did not impose a language restriction but all studies returned from the search were reported in English.

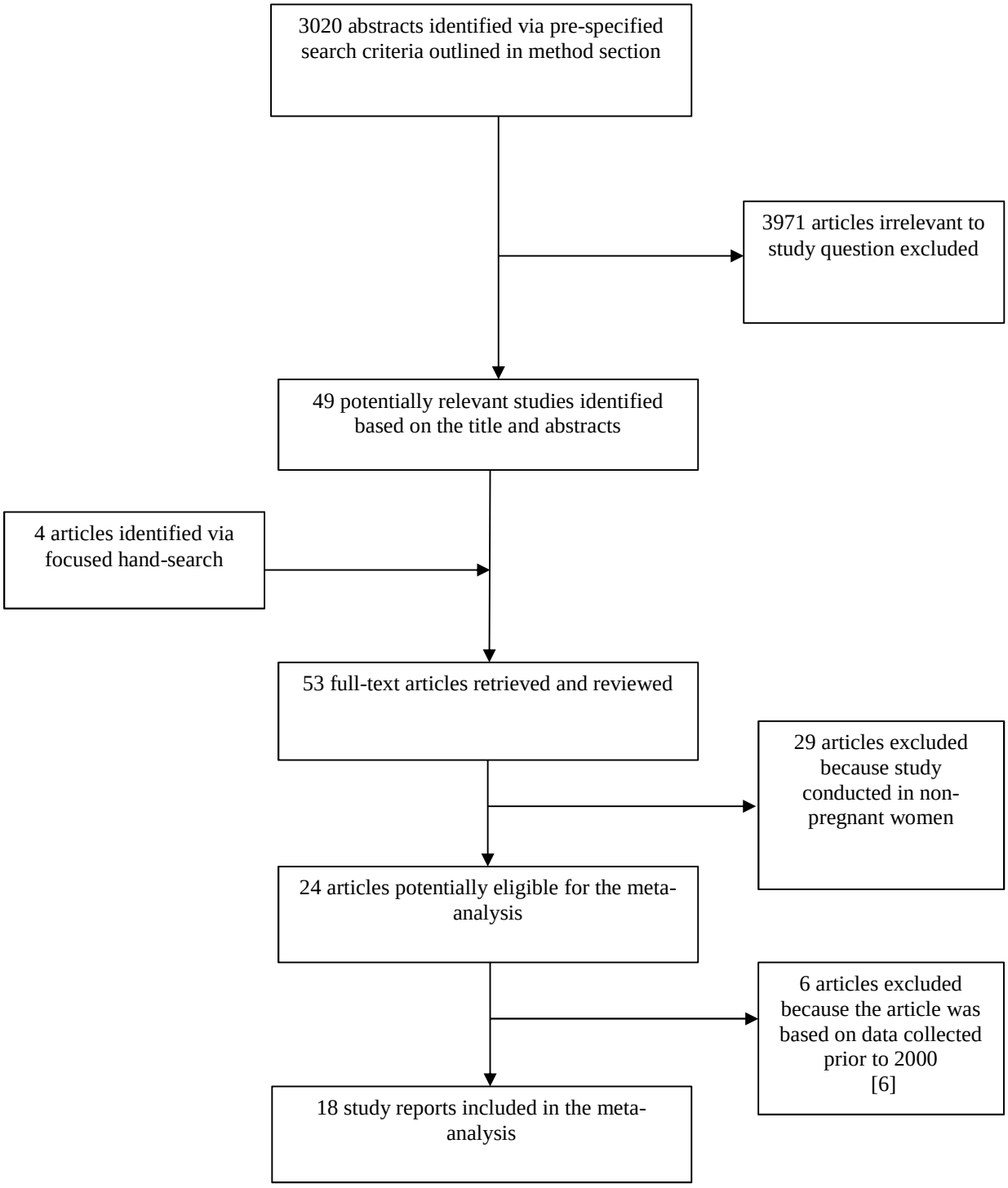

Figure 5.1. Flow diagram of the screening procedure for citations that examined the association between educational attainment and prevalent HIV infection that were identified for the meta-analysis: studies were published between 2000 and 2012. Indicated in the above figure the number of citations screened, identified, excluded, and ultimately incorporated in the meta-analysis.

5.8.3. Box plot of log odds ratio and corresponding standard error identified an outlier

To assess the extent to which outliers among log-odds ratio and among standard error of log-odds ratio might influence the estimate of the measure of association from a meta-analysis, the distribution of log odds ratios and the corresponding standard error of the log odds ratios were assessed using box plots (Figure 5.2).[217] The standard error for the log-odds ratio for Yahya-Malima et al. (2006) qualified as a possible outlier.[218, 219].

Extreme value or possible outliers can be winsorized to limit influence of extreme values on parameter estimates and consequently moderate the effect of possibly spurious outliers.[220] Although some meta-analysis experts recommend winsoring to mitigate the influence of extreme values, I chose not to winsorize outliers in the present meta-analysis analysis because of the inherent assumptions in random effect models meta-analysis that the eligible studies were sampled from a population of studies in which the measure of association characterized by variability in measure of association. Therefore, one might argue that outliers maybe representing unique studies: inherent variability in the relationship across studies may yield outliers. However, a sensitivity analysis of meta-analysis with (OR=1.18, 95% CI: 0.93, 1.50) and without (OR=1.21, 95% CI: 0.95, 1.53) an outlier.[220]

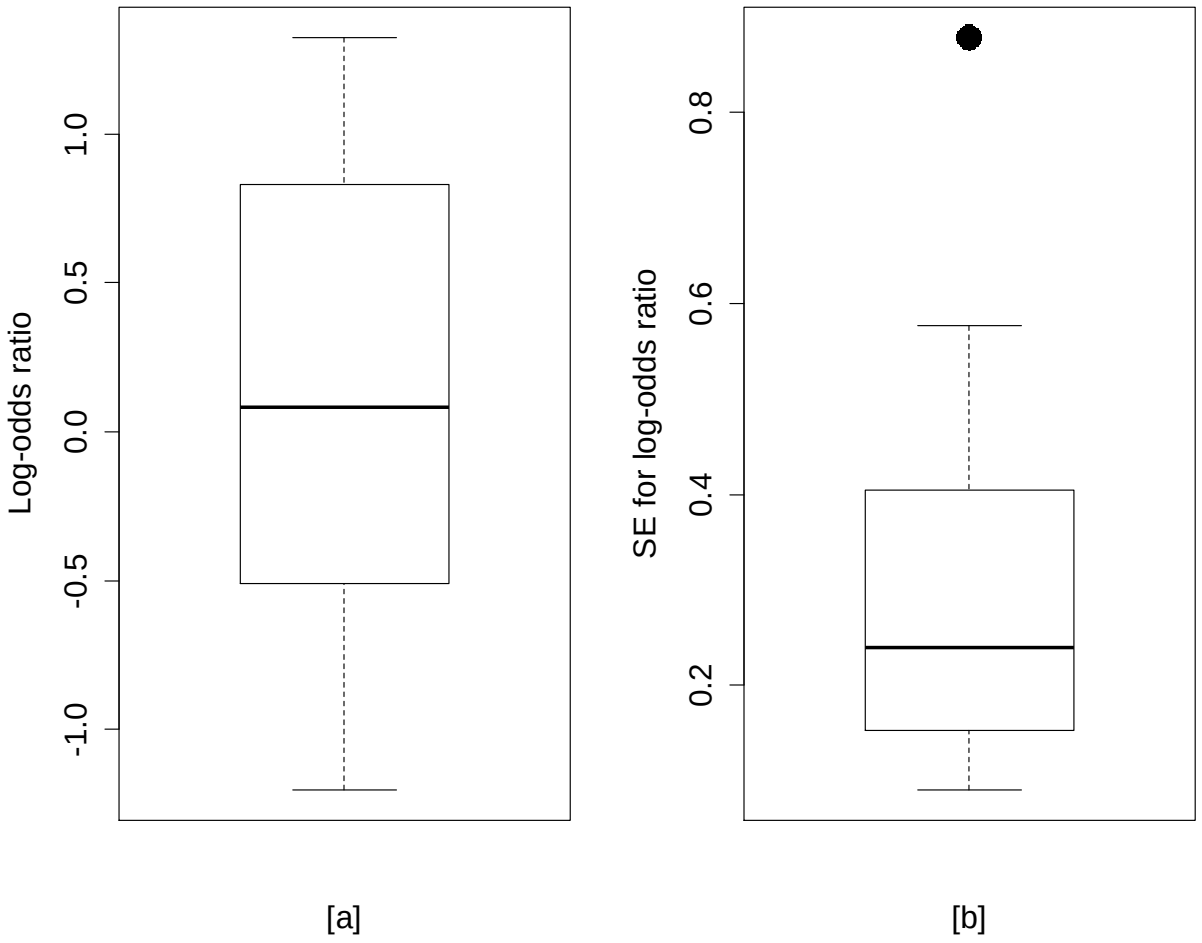

Figure 5.2. Distribution of the log odds ratios and corresponding standard errors of the log odds ratios among 18 studies published between 2000 and 2012 included in the meta-analysis of the association between educational attainment and prevalent HIV infection. Standard error for log-odds ratio for association for the Yahya-Malima et al (2006) study seemed to be a possible outlier as represented by a filled circle Fig 6b. The thick horizotal bar indicate the median values whereas the horizontal bars on the extremities of the box represent minimum value and maximum values respectively. The upper end and lower end of the boxes represent the 75^{th} percentile and 25^{th} percentile respectively.

5.9. DerSimonian-Laird random effects model meta-analysis

5.9.1. Overall estimate for the relationship between greater than primary school educational attainment and HIV infection

There was no significant association between educational attainment and HIV infection based on meta-analysis of odds ratios and 95% CI from 18 observational studies conducted in SSA between 2000 and 2012. The pooled estimate of the odds ratio and 95% CI computed using DerSimonian-Laird random effect model method indicated a non-significant positive association (OR=1.18, 95% CI: 0.93, 1.50) between educational attainment and prevalent HIV infection.

5.10. Evidence of heterogeneity detected

5.10.1. I-squared revealed presence between-study variability

The I-squared measures the proportion of the total variability in the measure of association that can be attributed to the heterogeneity across studies in the measure of association than chance; I found this to be 84.4% with a lower and upper confidence limit of 78.4% and 96.1% respectively for the 18 studies. The estimated value of the I-squared estimated is suggestive of considerable heterogeneity between-studies in odds ratios among studies that were included in the current meta-analysis.

5.10.2. Q-statistic revealed presence between-study variability

Based on the DerSimonian-Laid random effect model meta-analysis, the Q-statistic was 109.1 with 17 degrees of freedom and a corresponding p-value of less than 0.001 at 95% confidence level. Because the p-value corresponding to the estimated Q-statistics

with 17 degrees of freedom is <0.001, there seems to be evidence of between-study variability in odds ratio than would be expected by chance (P-value>0.001).

5.10.3. Tau-squared statistic as a measure heterogeneity

The tau-squared is the between-study variance. Based on the 18 studies included in the present meta-analysis tau-squared was estimated at 0.18 with a lower bound of 0.12 and an upper bound of 0.83. Noteworthy is the fact that the I-squared, even though high (I-squared =84.4 %) may be substantively less meaningful where the tau-squared value is very low (as it was here, equaling 0.18). Given that the three measures (i.e., I-squared, Q-statistic, and tau-squared) presence of heterogeneity between studies and therefore use of random effects model for the present meta-analysis is justified. To explain the observed variability in the odd ratio between studies would require further assessment.

5.11. Forest plot of the result of the meta-analysis

To visually represent the measure of association (i.e., odds ratios and 95% confidence interval) in my meta-analysis, I generated a forest plot (Figure 5.3).

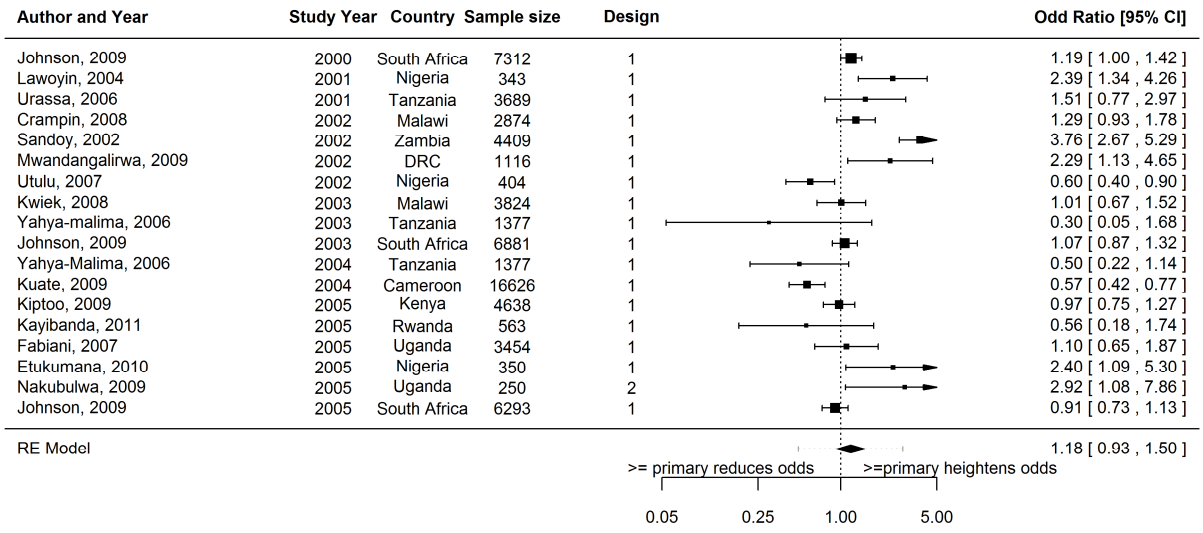

Figure 5.3. DerSimonian and Laird random effects model meta-analysis of odds ratios based on 18 studies that examined the relationship between educational attainment and prevalent HIV infection. Horizontal bars in the forest plot represent 95% confidence interval and the diamond shaped object at the bottom of the plot represents the pooled summary odds ratio. The square shaped objects on the horizontal bars denote the odds ratio of respective studies. Study year indicates the year the study was conducted: For example, Johnson, 2009 indicates paper was published in 2009 but the data included are from 2003. Design of the study is indicated under design column: "1" denotes cross-sectional and "2" denote non-cross-sectional study.

5.12. Funnel plot for publication and/or small study bias assessment

Figure 5.4 shows funnel plot as a visual assessment of publication bias based on the 18 studies included in the meta-analysis. The funnel plot is asymmetric and is suggestive, but not confirmative of small study and/or publication bias.

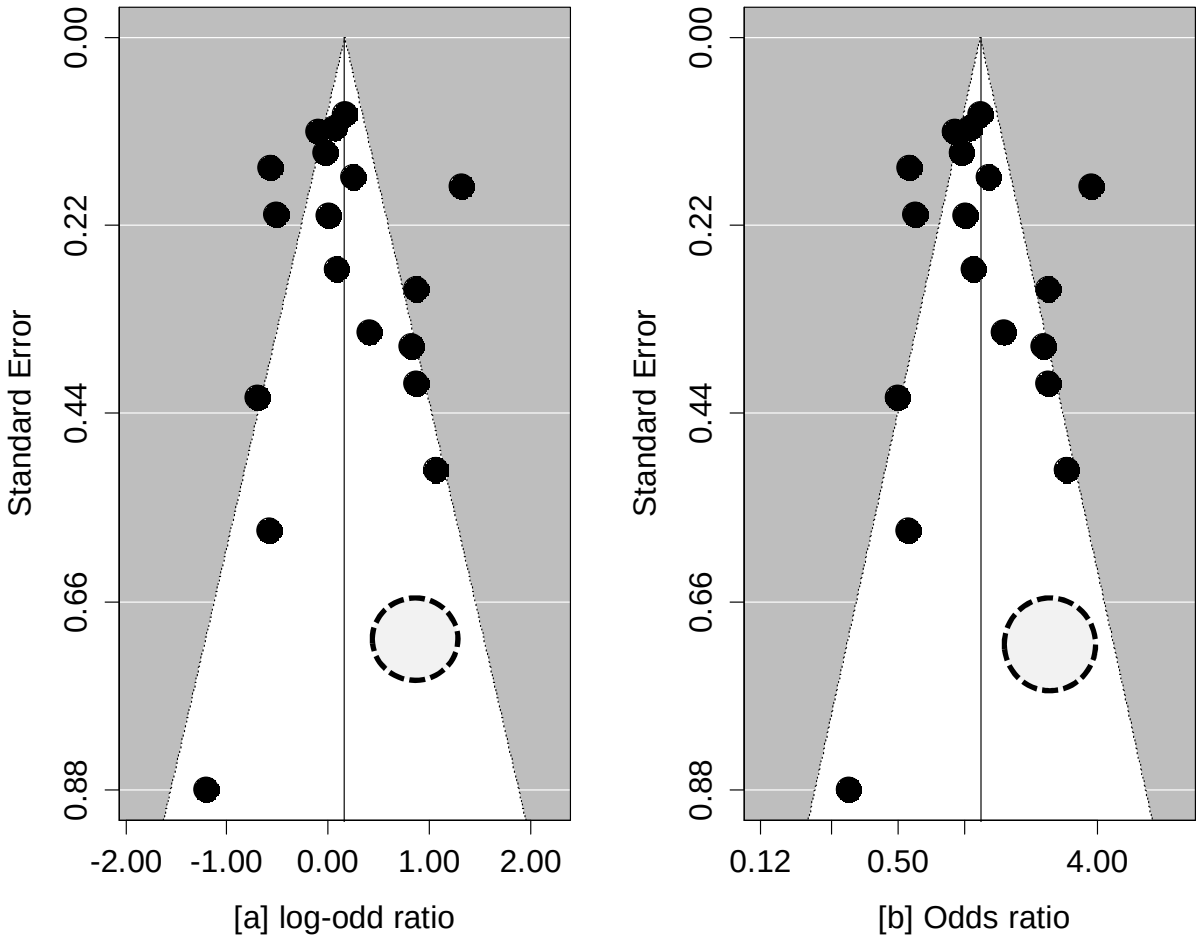

Figure 5.4. Funnel plot for visual assessment of presence of small study and/or publication bias among 18 studies conducted in SSA that were included in the meta-analysis for the association between educational attainment and prevalent HIV infection. Plots of values of standard errors versus log-odds ratios (Fig. 4a) and versus odds ratios (Fig.4b) and both funnel plots appear to be slightly asymmetric. The large circles with broken lines highlight areas likely to be sources of asymmetry. Caution: funnel plot interpretation is subjective because it visually based.

5.12.1. Egger's test for assessment of small study and publication bias

Further examination of the evidence of asymmetry as an indication of publication and/or small study bias by Egger's test (p-value =0.65) revealed a statistically non-

significant result, thus implying a lack of evidence for presence of publication and/or small study size bias

5.12.2. Sensitivity analysis using meta-trim suggests that small studies with extreme values were suppressed

As part of the sensitivity analyses to assess the influence of studies with extreme values on the estimated pooled odds ratio, trim and fill approach was used, and two studies imputed as shown by the open circles in the funnel plot in Figure 5.5. The p-value from the trim and fill analysis suggested significant heterogeneity across studies was detected (p-value =0.001).[201, 221] Figure 5.5 shows two open circles that represent studies which were imputed via iterative procedure of the trim and fill method. The open circles are mirror images of the extreme studies. [201, 221]

Based on the sensitivity analysis using trim and fill random effect model, the relationship between educational attainment and prevalent HIV infection was not statistically significant (OR=1.25, 95% CI: 0.99, 1.58). The estimates (OR=1.25, 95% CI: 0.99, 1.58) with data based on filled data (i.e., with three studies) were comparable to estimates (OR=1.18, 95% CI: 0.83, 1.50) without filled-in studies.

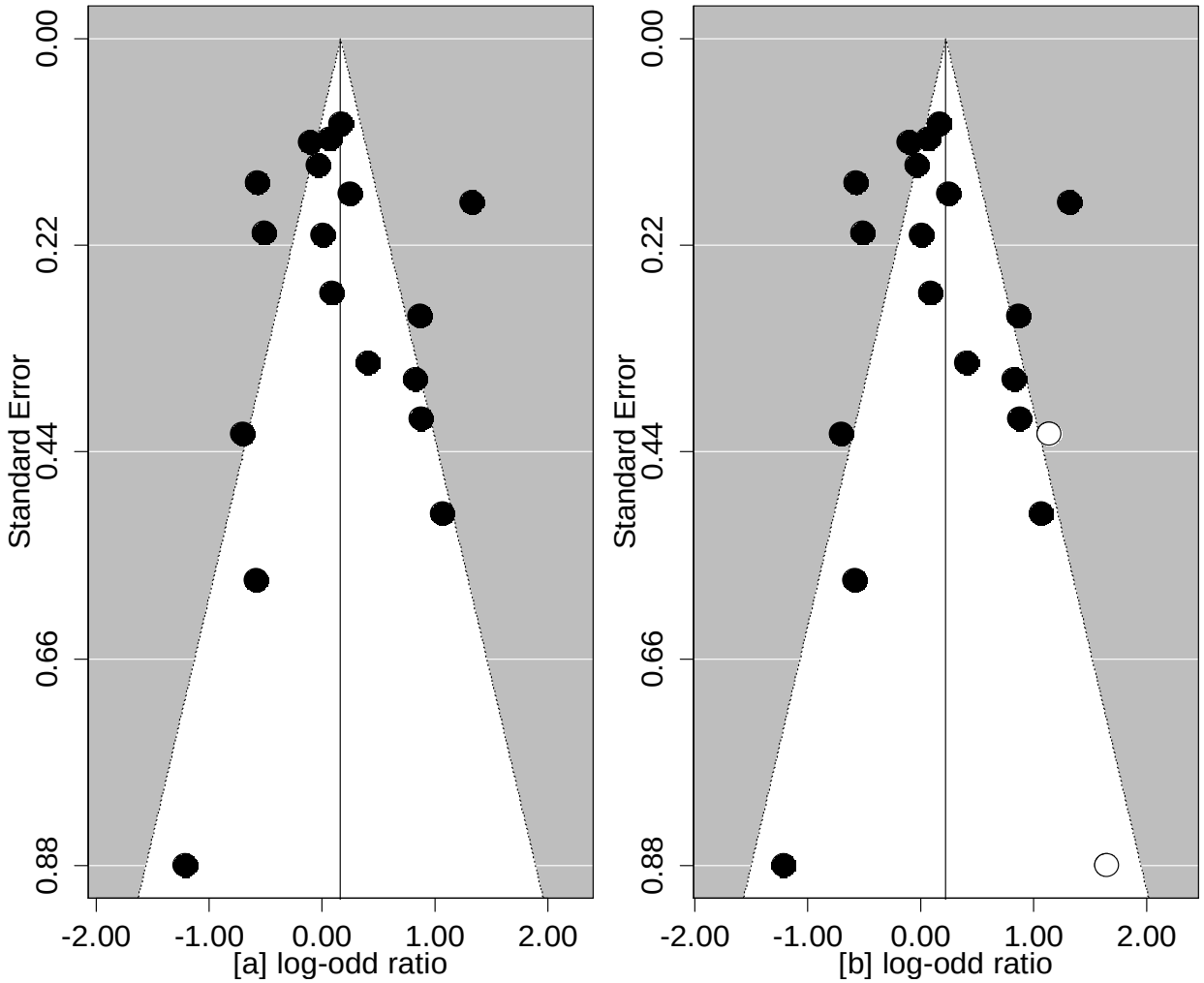

Figure 5.5. Display of trim and fill plot for evaluating the robustness of the funnel plot for assessing evidence of presence of small study size and/or publication bias. Figure 5b represent the funnel plot based on the trim and fill sensitivity analyses. The trim and fill method assumes that studies with the most extreme measures of association are suppressed. Two studies were imputed in the trim and fill sensitivity analyses to correct the asymmetry, and are represented by the open circles. The analyses were based on 18 studies conducted in SSA between 2000 and 2012 that examined the association between educational attainment and prevalent HIV infection.

5.13. Cumulative meta-analysis to examine evidence between 2000 and 2012

The association between educational attainment and prevalent HIV infection changed from positive association to positive but null for studies conducted between 2000 and 2006 but published between 2000 and 2012.[222]

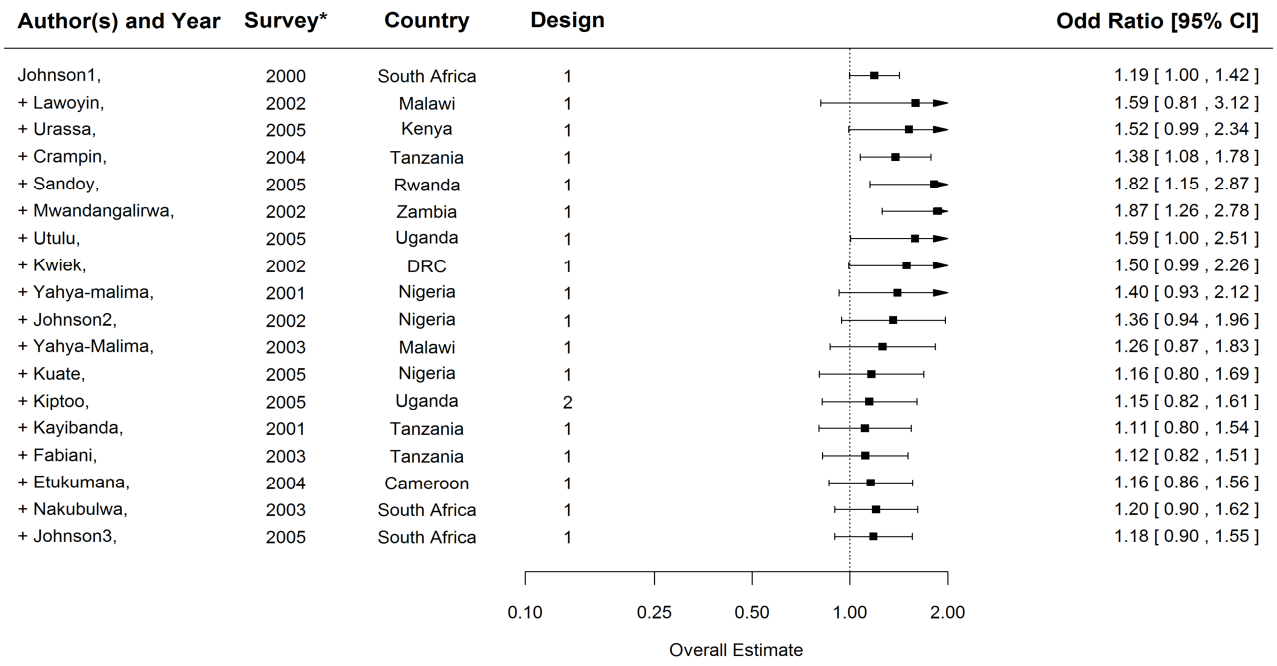

Figure 5.6. Cumulative meta-analysis forest plot based on 18 observational studies conducted in SSA between 2000 and 2006 but published between 2000 and 2012 that examined the relationship between educational attainment and prevalent HIV infection. Odds ratio > 1.0 indicated increased odds of prevalent HIV infection among pregnant women ≥primary school education. For cumulative meta-analysis, analyses proceeds chronologically, at each stage adding one study and computing the overall odds ratio and the analysis terminates with most recent study.

5.13.1. Meta-regression with survey year as the predictor variable

Meta-regression was conducted with survey year as the explanatory variable and the estimated amount of residual heterogeneity (i.e., tau-squared) was 0.87. The lower bound was residual heterogeneity was 0.65 and upper bound at 1.52. This meta-regression assessed whether association between educational attainment and prevalent HIV infection varied by the year in which the survey was conducted over the course of

maturation of the epidemic in SSA. Using meta-regression with study year as explanatory variable accounted for an estimated 15.8% of the total amount of heterogeneity in the odds ratio.

5.13.2. Association between educational attainment and prevalent HIV infection

The pooled odd ratio estimate for the association between educational attainment and prevalent HIV infection was 1.18 (95% CI: 0.93, 1.50) indicating a non-significant modestly positive association among pregnant women in sub-Saharan Africa. Although non-significant pregnant women with greater than primary school education were 18% more likely be HIV seropositive than were pregnant women with less than primary school education. The finding could be a chance finding.

The change in the estimated I-squared estimated from meta-regression was little compared to I-squared without using study years as explanatory variable. Figure 5.7 is based on meta-regression with survey years as predictor. The estimated I-squared was 87.40% with lower bound of 75.0% and upper bound of 95.7 %. The p-value <0.001 for residual heterogeneity suggest that other factors not considered in the current analysis may be swaying the relationship between educational attainment and prevalent HIV infection. Based on the current data, the year when the study was conducted does not seem to have considerable influence on the observed heterogeneity. Between 2000 and 2012 and based on estimates from the meta-regression one year change year (Figure 5.7) when study was conducted implies a change of -0.0944 on the log-odds ratio (OR=0.91, 95% CI: 0.77, 1.07).

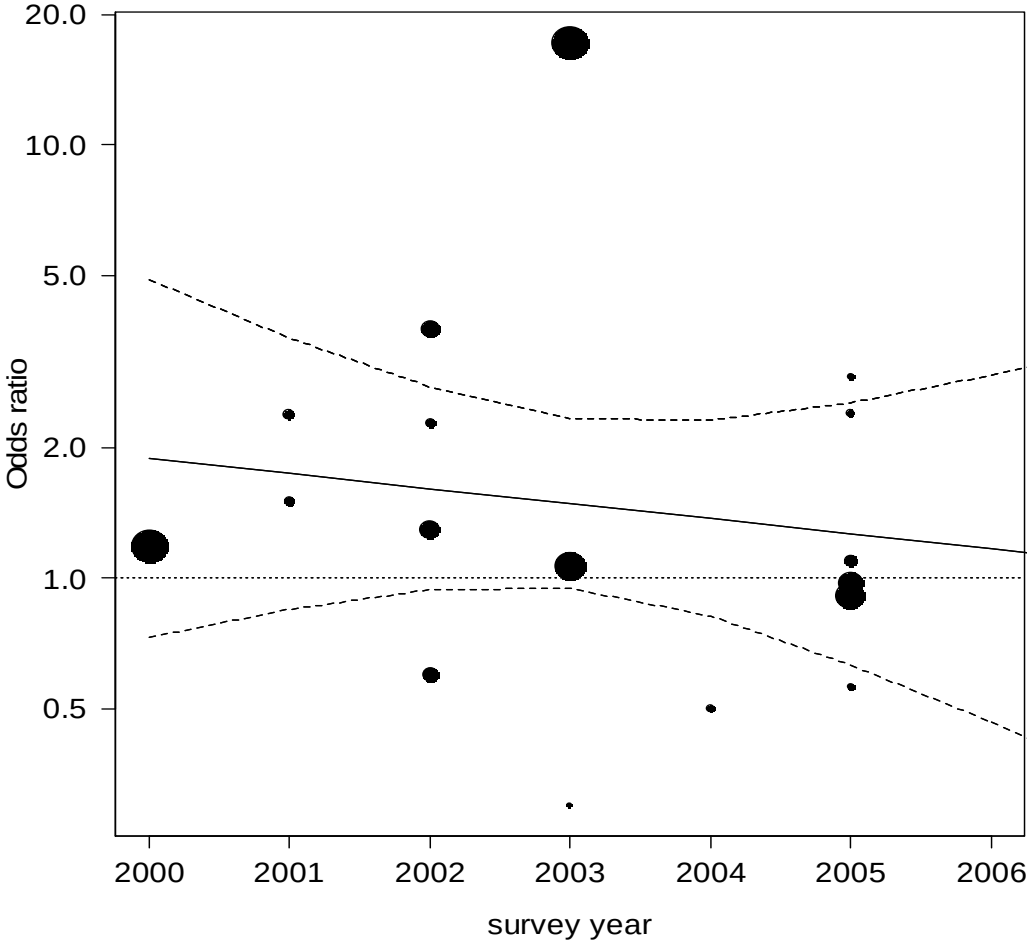

Figure 5.7. Odds ratio of prevalent HIV infection as a function of year in which the study was conducted for the period between 2000 and 2006, corresponding to period in which studies were conducted: analysis included 18 studies published between 2000 and 2012. Each filled circle represents odds ratio from a specific study, and their size are proportional to the inverse of the respective standard error reported for each study (i.e., study weight for the analysis).

5.14. Discussion

The current meta-analysis was conducted to examine ≥primary educational attainment was associated with increased odds of prevalent HIV infection.[212-214] Based on the DerSimonian-Laird random effect model meta-analysis of 18 peer-reviewed observational studies conducted in SSA, there was no significant association between educational attainment and prevalent HIV infection (OR=1.18 (95% CI: 0.93, 1.50).

Further analysis via cumulative meta-analysis to examine how the evidence on the association between educational attainment and prevalent HIV infection accumulated. The forest plot in Figure 5.6 indicate overall non-significant association between educational and prevalent HIV infection during the 2000s.[146, 223] The findings are consistent with the hypothesis of a waning relationship between educational attainment and prevalent HIV infection by Hargreaves et al (2008) as the HIV epidemic progressed.[223] The meta-regression seems suggest that study or survey year may not explain away all the between-study heterogeneity. The meta-regression also revealed that the association between educational attainment and prevalent HIV infection tended to be modestly protective as the number of years progressed away from 2000, the reference year (OR=0.91 (95% CI: 0.77, 1.07).

The funnel plots in Figure 5.4 were slightly asymmetric and therefore suggestive of presence of small study and publication bias, although the Egger's test (p-value =0.72) failed to suggest any substantial publication or small study bias. However, the sensitivity analysis using trim and fill method revealed possible suppression of two studies with extreme values of odd ratios as shown by open circles in Figure 5.5.

Although the Egger's regression symmetry test did not detect publication bias, it is important to note that the Egger's test may have limited power to detect publication bias because the probability of the test to wrongly reject the null hypothesis increases with increasing sample size of the meta-analysis study. To assess the presence of asymmetry, the Egger's test examines the null hypothesis of no small study effect by searching for a straight line relationship via linear regression of log-odds ratio effect estimates on standard errors weighted by the reciprocal of the variance of log-odds ratio. Digression of the estimated intercept of the linear regression from zero suggests presence funnel plot asymmetry but do not confirm publication bias.[216, 224] Small sample size of a meta-analysis may invariably translate into limited power of Egger's test for presence of small study bias.[225]

The findings of this meta-analysis that there is no significant association between educational and prevalent HIV infection among pregnant women are consistent with the hypothesis by Hargreaves et al (2008) that with increasing intensity of HIV preventive interventions, any early association between educational and HIV infection is likely to wane in a more established HIV epidemic.[81] Cognizant of the fact that the strength of evidence emanating from cross-sectionally collected data may not be used for causal inference, restraint should be exercised in drawing definitive conclusions; data from prospective studies are scarce and randomized trials are not feasible. The collection of prospective data is often challenged by ethical and practical cost concerns. Therefore, it is wise to acknowledge the limitations inherent in observational study designs, particularly the influence of unrecognized/uncontrolled confounders and reporting bias.

For the current meta-analysis, the pooled odds ratio seems plausible (OR=1.18 (95% CI: 0.93, 1.50). For example, in earlier years of the HIV epidemic in the 1990s, most studies found increased odds of prevalent HIV infections among persons who reported higher educational attainment. People with high educational attainment and higher wealth quintile were assumed to have greater travel opportunities and more sexual mixing, for example than less educated and poor people. But as the epidemic spread, more educated persons may have internalized key HIV prevention messages better than less educated persons, and the difference in sexual risk behaviors may have lessened across the educational spectrum.[46, 81].

In addition, cART for eligible HIV-infected people has become widespread in SSA, with implication that number of people surviving longer with HIV is likely to increase. Therefore, careful assessment of the relationship educational attainment and risk of HIV infection is required to avoid using estimates that are affected by survival bias. Examination of the education-HIV relationship in the 15 to 24 year-olds is attractive because the 15 to 24 year-olds are less influenced by survival bias. Because some of the estimates were computed using data that included pregnant with age ranging from 15 to 49 years, the estimated pooled estimates might be influenced by survival bias.[62, 204]

On average, people who have higher educational attainment may be more likely to understand treatment and adherence instructions that accompany combination therapies for HIV infected persons than people who have less education.[226] Monge et al (2012) reported that despite comparable accessibility to HAART, persons with low educational attainment were at comparatively increased risk of poor treatment outcomes. Consequently, survival will be greater among educated people than less educated, and the

association between prevalent HIV infection and prevalent HIV infection will be contaminated by survival bias, and therefore might not reflect educational attainment as a risk factor for HIV infection.

The slightly asymmetric shape of the funnel plots in Figure 5.4 suggests presence of publication and/or small-size study effects. Asymmetry in the funnel plot may be caused by many factors including but not limited to poor methodological quality and overestimation of measures of association in small studies. Although both the Egger's and Begg's statistics did not detect presence of publication bias and/or small study, the non-detection of publication bias and/or small study bias might be due to limited power to detect bias. Because the Egger's test may have limited power to detect publication bias, caution is required in interpreting the result. Small study bias may be a consequence of several factors including publication bias, differences in methodological quality and true heterogeneity in the measure of association may also yield asymmetry in the funnel plot as well as influence Egger's test.[200, 201, 216]

5.14.1. Limitations

Meta-analyses findings are reliable and convincing to the extent that data from primary studies are trustworthy and resounding. Validity concerns are possible given cross-sectional design of the primary studies included in the meta-analysis (17 out of 18 studies were cross-sectionally designed). Thus, causal or non-causal inference cannot be made based on the findings in this study because temporal sequence information of exposure and outcome lacks in cross-sectional study designs. Furthermore, the findings from this meta-analysis may not be generalized beyond the studied population of

pregnant women. While the level of confidence in the findings from cross-sectional studies is less than the level of confidence from randomized control trials, such work is not feasible in an experimental context.

Data on possible confounding and moderating variables were not collected in most studies and therefore could not be studied to try to explain the between-study variability observed. Further research is needed to explain identify or explain the sources of heterogeneity detected by meta-analysis. It is possible that the varying quality and quantity of education may contribute to the varying measure of association. Further, different studies used different set of confounders for adjustment, and therefore, there may be various degrees of residual confounding in the estimates from the studies.

The plot of standard error versus odds ratio in a funnel plot was suggestive of small size study bias and/or publication bias (i.e., asymmetry of funnel plot), contrary to both Egger's and Begg's tests (p-values >0.05). Given that grey literature and studies were harvested from MEDLINE that search criteria was restricted to PubMed, there might be some degree of publication bias. Because of differences in the definition of exposure variable, some studies were excluded from the meta-analysis. The Egger's test for assessing publication bias has limited power to detect small study and/or publication bias given our small sample size of 18 studies. The meta-analysis focused on reports reported in English and peer-reviewed article. Thus, the likelihood of publication bias exists because of we found no non-English articles, we did not search the so-called grey literature or unpublished studies. Studies published in English may be different from studies published in non-English languages.

Meta-analysis guidelines recommend that data are retrieved and reviewed by at least two persons so that inclusion and exclusion criteria decisions and data entry accuracy can be cross-checked. The studies included in the meta-analysis were retrieved and entered in Epi-Info version 3.5.4 by only one person, and were not cross-checked. The abstracted data will be counterchecked by a second person before submission for publication.

The number and types of variables adjusted for in multivariable models varied across studies but none of the studies included adjusted for intermediate variables (i.e., variables that might be on causal pathway that could lead to underestimation of the measure of association). Of concern is the possibility non-differential misclassification of the exposure variable in the lower primary and higher educational category because the differences the cutpoints for classifying primary school education were not completely consistent. However, I used a reasonable ability to estimate this as a criterion for study inclusion. Further, stratified analysis were not performed to assess the effect geographical region (i.e., proxy for stage and extent of the HIV epidemic, and expansion and availability of effective cART).[227] The small number of studies limited my ability to conduct stratified analyses, and also, the analyses were not pre-specified.

5.14.2. Strengths of the study

My meta-analysis had a number of strengths that differentiated the study from prior systematic reviews. The current analysis focused on the year in which the data were collected, and not the publication year, thus attributing the estimates to specific years in which the data were collected rather than the year of publication year of the report.

Therefore, study year was applied for describing the cumulative evidence in the cumulative meta-analysis. The year of the survey is more realistic way of describing the accumulation of evidence because some studies are only published several years after the studies were completed (i.e., study published in 2009, but data collected between 2000 and 2005).

All the studies included in the present meta-analysis used HIV serostatus data that were objectively confirmed using serological methods, and therefore not subject to self-report biases. Further, methods used for studies conducted among pregnant women were to a large extent similar, although in discrete geographic areas. Admittedly, there are study power and precision gains that are inherent in pooling observational studies, but the benefits comes with measurement errors, selection bias and confounding that may be stereotypic of observation studies included in the meta-analysis.

The base populations for the primary studies included in the meta-analysis comprised pregnant women; therefore a consistent population was used. Admittedly, the social, economic and environmental factors may differ according to settings of study areas. To take into consideration the varied settings and data methodological approaches used in specific studies included the meta-analysis, the random effect model based on the DerSimonian-Laird method was applied to compute the overall odds ratio that accounted for the between-study and within-study variance. The observational studies that were meta-analyzed conducted among pregnant women in SSA, thus limiting the generalizability of study findings to non-pregnant women.

5.14.3. Conclusion

The meta-analysis of observational studies that examined the association between educational attainment and prevalent HIV infection in SSA countries with generalized HIV epidemic indicates significant association. Limited information presented by primary studies did not allow for a full exploration of observed between-study heterogeneity. The observed findings must be cautiously interpreted as the computed odds ratio may be contaminated by survival bias, especially for studies that included women who had long-term HIV infection. Most importantly, the meta-analysis emphasizes the need for assessing the association between educational attainment and incident HIV infection rather than prevalent HIV infections. Educational attainment may be associated with longer survival of HIV-infected persons, specific to understanding the relationship between educational attainment and risk of HIV infection, future studies should focus on examining the association between educational attainment and HIV incidence, as well as examining the relations between literacy and HIV incidence

Table 5.2. Characteristics of studies included in the meta-analysis examining the association between educational attainment and prevalent HIV infection

Characteristics of the studies included in the meta-analysis: Table 8							
Study author and reference number		Publication year	Site	Sample size	Study design	Selected adjustment variable	Comments
1	Sandøy et al. [46]	2002	Zambia	Missing	1	Age, marital status, and parity	Limitations Continuous variables were categorized Non-probability sampling Limited generalizability as study restricted pregnant women Strengths Large sample size Country-wide coverage
2	Lawoyin et al.[223]	2004	Nigeria	343	1	Age, marital status, religion and other variables[223]	Limitations Limited generalizability as study restricted pregnant women. Basis of selection of covariates not noted. Strengths Compared estimates among pregnant women with estimates among population-based sample
3	Yahya-Malima et al[219]	2006	Tanzania	1296	1	Age, residence, marital status, education [categorized], number of partner, age at first pregnancy. [219]	Limitations Inclusion of the covariates in the MVM† l based on p-value cut off <0.25, and dropped non-significant variables from the final model

Characteristics of the studies included in the meta-analysis: Table 8							
Study author and reference number		Publication year	Site	Sample size	Study design	Selected adjustment variable	Comments
							Several variables in addition to listed in this table were included in the MVM[†] raising the possibility of over-adjustment Survival bias: analysis based on pregnant women ≥ 20 years at the time of the survey <u>Strengths</u> Large sample size
4	Fabiani et al[228]	2006	Uganda	3454	1	Age, residence, parity, mobility, marital status and occupation	<u>Limitations</u> Limited generalizability because study was limited pregnant women Large sample size [strife-stricken study areas] Excluded adjustment for occupation which may be on the causal pathway in some models. Strengths Used hierarchical structure to account for hierarchical structure of the data
5	Utulu et al[229]	2007	Nigeria	404	1	Age, marital status and behavioral factors [229]	<u>Limitations</u> Behavioral factors may mediate the relationship between educational attainment and prevalent HIV infection

Characteristics of the studies included in the meta-analysis: Table 8							
Study author and reference number		Publication year	Site	Sample size	Study design	Selected adjustment variable	Comments
							Survey response rate not indicated: since consent was sought <u>Strengths</u> Collected sexual behavior information
6	Kwiek et al [218]	2008	Malawi	3824	1	Residence, tribe, marital status, employment, gravidity and age[218]	<u>Limitation</u> Continuous variable [i.e., age, educational attainment were categorized] <u>Strengths</u> Large sample size Covariates for the MVM† apriori specified
7	Crampin et al [230]	2008	Malawi	2874	1	Age, parity, age of partner, prior use of contraceptive,	<u>Limitation</u> Some behavioral factors adjusted for may mediate the

Characteristics of the studies included in the meta-analysis: Table 8							
Study author and reference number		Publication year	Site	Sample size	Study design	Selected adjustment variable	Comments
						residence[230]	relationship of interest Validity of measurement of behavioral factors Choosy reporting of behavior preferred by health worker may cause bias <u>Strengths</u> Fairly large sample size
8	Mwandangilirwa et al [231]	2009	DRC	Missing	1	Age, marital status, employment and other sexual behavior characteristics	<u>Limitations</u> Selection bias and limited generalizability. Large sample size used. <u>Strengths</u> Large sample size Examined risk factor HIV prevalence in different settings [e.g., Antenatal care clinic, Commercial sex worker]
9	Kiptoo et al [232]	2009	Kenya	4638	1	none[232]	<u>Limitations</u> Sub-study of the main study that examined drug resistance Lack of MVM Possibility of selection bias <u>Strengths</u> Large sample size

Characteristics of the studies included in the meta-analysis: Table 8							
Study author and reference number		Publication year	Site	Sample size	Study design	Selected adjustment variable	Comments
9	Etukumana et al.[233]	2010	Nigeria	350	1	Age, marital status, occupation and other behavioral factors[233]	Limitations Selection bias and limited generalizability. Small size compared to other studies reported. Strengths Modest sample size used and outcome serologically confirmed. MVM performed
10	Kayibanda et al [234]	2011	Rwanda	563	1	NA	Limitations Possible selection bias due to changes in the population composition as a result of genocide: mortality, changes in the population dynamics Strengths Modest sample size used and outcome serologically confirmed. Local language used for capturing data
	Urassa et al.[235]	2006	Tanzania	3689	1	Age, marital status, occupation, clinic, and sexual behavior variables[235]	Limitations Odds ratio for the education-HIV computed from cell counts and method of HIV serostatus ascertainment not indicated [self-report or serological]

Characteristics of the studies included in the meta-analysis: Table 8							
Study author and reference number		Publication year	Site	Sample size	Study design	Selected adjustment variable	Comments
							Strengths Participation was voluntary Variability in the sample was increased by enrolling 25 pregnant women per day.
11	Ntanganira et al.[236]	2008	Rwanda	600	1	NA	Limitations Strengths MVM used
12	Nakubulwa et al [237]	2009	Uganda		1	Unadjusted and less rigorous statistical analysis[237]	Limitations The association between educational attainment and prevalent HIV infection unadjusted for potential confounders Strengths Matching of cases and control [HIV positive and HIV negative pregnant women]
13	Johnson et al.[141]	2009	South Africa	99,153 pregnant women between	1	MVM and adjusted partner age difference, parity, syphilis and race[205]	Limitations Cross-sectional design and selection bias of pregnant women

Characteristics of the studies included in the meta-analysis: Table 8							
Study author and reference number		Publication year	Site	Sample size	Study design	Selected adjustment variable	Comments
				2000 and 2005			Strengths Stratified estimates for education-HIV relationship for 15 to 24 year-olds and 25 to 44 year-olds
14	Kaute et al [238]	2009	Cameroon	16626	1	Age, marital status, years of schooling, and year of data collection [indicator variable][238]	Limitations Cross-sectional design and selection bias of pregnant women Strength Rigorous multilevel analysis accounting for possible clustering of the data collected from multiple sites; High participation rate [97%]
† MVM-Multivariable model							

CHAPTER 6

HIV PREVALENCE TRENDS AMONG PREGNANT WOMEN IN ZAMBIA USING ANC-HIV-SS SURVEILLANCE DATA, 1994 THROUGH 2011

6.1. Background

6.1.1. Decline in new HIV infection but increased HIV burden in 2010

Human immunodeficiency virus (HIV) burden among pregnant women is critical public health concern globally, but is most profound in sub-Saharan Africa (SSA). Of the two known types of HIV, the most virulent and widespread in SSA is HIV-1.[4, 8] HIV-2 also causes acquired immunodeficiency disease syndrome (AIDS) as HIV-1 does but is less virulent and far less widespread, endemic in west Africa but presenting only very rarely in southern Africa.[12]

6.2. The HIV epidemic is heaviest in sub-Saharan Africa

The United Nations Program on AIDS (UNAIDS) estimated that 2.5 million people were newly infected with HIV in 2011 compared to 3.1 million people that were newly infected with HIV in 1990.[10] The estimated global HIV burden increased over four-fold from 8 million in 1990 to 34.2 million in 2011.[1, 4, 8]. An estimated 70 million people have been infected with HIV globally since the beginning of the HIV epidemic, and 35 million people have died from AIDS-related conditions.[20] Thirteen percent of human population (800,000,000) reside in SSA , the region where 69% of the 34.2 million of the people living with HIV globally resided in 2011.[10, 11, 239] UNAIDS

reasoned part of the explanation for the burgeoning HIV burden globally in 2011 could be linked to the cumulative effects of the high rates of new HIV infections, particularly in SSA, and also could be a result of improved survival of HIV-infected people who were receiving life-prolonging antiretroviral therapy (cART).

6.3. Unprotected heterosexual intercourse is the main route of HIV transmission

Factors regarded as drivers of the HIV epidemic tend to be region-specific, and tend to be aligned common mode of HIV transmission within a region (e.g., needle sharing during injection drug use in Eastern Europe, Central Asia, USA and Latin America; commercial sex work in Southern Asia and Africa).[4] Prominent routes of HIV transmission in adults include injection drug use, blood transfusion, and unprotected sexual intercourse with an infected partner in marriage, cohabiting partners, casual, or commercial sex.[4, 17, 240-242]. An estimated 85% of the HIV infections transmitted globally are spread via unprotected heterosexual intercourse, the predominant route of HIV transmission in SSA. [3]

6.3.1. Zambia has a generalized HIV epidemic

Zambia located in southern SSA and populated with an estimated 13 million people, has a generalized HIV epidemic (i.e., > 1% of HIV prevalence in the general population, World Health Organization definition). The first official report of AIDS in Zambia was in 1984.[10, 23, 34, 36, 180] The gold standard for reporting national HIV prevalence is the population-based Demographic and Health Survey (DHS) which placed HIV prevalence among 15 to 49 year-olds at 15.6% in 2001 and 14.3% in 2007,

representing non-statistically significant decline.[36] DHS of 2001 and 2007 also highlighted profound geographic variation in HIV prevalence across the nine provinces of Zambia: HIV infection burden was higher in urban than rural areas (e.g., Lusaka Province [22%] and Northern Province [7%] respectively). In 2011, the estimated number of people living with HIV infection in Zambia ranged from 900,000 and 1,100,000.[23, 34]

6.3.2. HIV incidence and prevalence data are key for monitoring the HIV epidemic

Like most countries faced with the HIV epidemic in SSA, the key data sources for monitoring national trends in HIV prevalence in Zambia are antenatal clinic attendees based HIV-sentinel surveillance (ANC-HIV-SS) and the population-based surveys (PBS) HIV prevalence estimates generated as part of DHS. DHS-based HIV prevalence estimates have only been generated at two time points (i.e., 2001 and 2007) compared to ANC-based HIV prevalence estimates at seven time points (i.e., 1994, 1998, 2002, 2004, 2006, 2008, and 2011) since 1994.[19, 36] HIV prevalence trend analyses using two time points' data (i.e., 2001 and 2007 DHS HIV prevalence) may be informative to a limited extent without consideration of other data sources for HIV prevalence trend assessment. In Zambia ANC-HIV-SS based HIV prevalence estimates complement and extend population-based DHS-based HIV prevalence estimates in monitoring trends in HIV prevalence.[24]

6.3.3. ANC-HIV-SS based HIV prevalence estimates may be subject to selection bias

ANC-based HIV prevalence estimates may be biased (e.g., possible selection bias of young pregnant women) and may not be regarded as proxy HIV prevalence estimates for the general population. Part of the explanations for differences in the ANC-based HIV prevalence estimates and population-based HIV prevalence estimates could due to the dissimilarities in the sample selection criteria (i.e., subject to selection and refusal bias) and characteristics of study sample.[171] The sampling strategies employed in ANC based HIV surveillance and PBS DHS-based HIV surveillance differ (i.e., convenient sample of pregnant women and probability sampling techniques respectively).[36, 171] PBS-based HIV prevalence surveys, although regarded as "gold standard for estimating HIV prevalence in general population", may provide biased HIV prevalence estimates if participation rates are low. However, prior epidemiologic research have revealed identical HIV prevalence estimates based on ANC-HIV-SS data and PBS-DHS data in Zambia as explained in Chapter 4 of this dissertation.

6.4. HIV incidence is a preferred measure for tracking HIV epidemic

HIV prevalence estimates, although commonly used by most countries for monitoring the HIV epidemic may not provide adequate information for monitoring changes in the HIV epidemic and focusing HIV prevention and treatment interventions as might HIV incidence. Methodologically robust longitudinal studies are the optimal design for estimating HIV incidence, but their implementation is not feasible in most settings because of logistical and technical challenges encountered. Longitudinal cohort studies require follow-up of a large sample of HIV seronegative persons to estimate valid

HIV incidence, identify new HIV infections, and their costs are prohibitive for most low and middle income countries (LMIC).

Because HIV risk behavior counseling is an imperative for longitudinal studies in HIV research, study participants may adopt less risky sexual behaviors, consequently influencing the risk of HIV infection among participants in the study population. Additionally, the Hawthorne effect, a circumstance in which persons may change their behavior because they know that they are in a study, longitudinal cohorts can give distorted incidence estimates.[176-179] HIV incidence measurement using cross-sectionally-collected biological marker are still under-development and not widely acceptable.[174, 175, 243-258] Although other laboratory-based methods with great potential have been proposed there are too expensive to implement on wider scale in most resource-challenged settings.[259]

6.4.1. UNAIDS recommend using number of prevalent HIV infections in 15 to 24 year-olds to approximate the number of new HIV infections

Most SSA countries including Zambia use the cross-sectionally estimated number of prevalent HIV infections in 15 to 24 year-olds to approximate the number of new HIV infections, a UNAIDS recommendation.[10] The recommendation rest on the assumption that mortality rate and migration rate are lower in the 15 to 24 year-olds than mortality rate in the 25 to 44 year-olds. Further premise for using 15 to 24 year-olds HIV prevalence estimates as an estimation of HIV incidence is the relative recency of viral acquisition whose risk only dates back to their coital debut.

6.5. Gap in knowledge

6.5.1. Few studies have examined non-linear patterns in HIV prevalence trends

Epidemiologic studies that have reported declining HIV prevalence (e.g., Kenya, Tanzania, Uganda, South Africa, and Zambia) relied on linear models for examination of HIV prevalence despite the known inadequacy of linear models to capture non-linear trends.[24, 42, 43] Although the overall, HIV prevalence might be linear, certain geographic regions may be experiencing nonlinear trends in HIV prevalence.[24, 39-43] Non-linear HIV prevalence trends may exist and, in fact, have been documented in some dramatic examples, as with Uganda's decline and recent rise in background prevalence.[44]

Review of selected epidemiologic literature from SSA revealed that the statistical methods used in some of the studies to investigate linear trends in HIV prevalence may be not be adequate to detect non-linear trends in HIV prevalence.[24, 45-47] Fewer studies, despite using data collected from multiple sites accounted for possible intra-site clustering. For example the possibility of intra-site clustering within sentinel sites is highly likely but prior studies using data collected from multiple sites did not account for within-site clustering. Further, prior studies have mostly used linear assumption to examine site-specific trends in HIV prevalence in Zambia.[24]

6.6. Specific aim

To examine trends in the prevalence of HIV by selected covariates (e.g., parity and residence) among pregnant women aged 15 to 44 years attending antenatal clinics as

sentinel sites for the Zambia ANC-SS for HIV and syphilis in 1994, 1998, 2004, 2006, 2008, and 2011.

6.6.1. Hypothesis # 1

Linear models of trends in HIV prevalence among pregnant women in Zambia suggest a decline over time. The decline in HIV prevalence is not consistently linear, and we hypothesize that non-linear models will reveal significant recent increase in HIV prevalence. Because HIV infected people are living longer due to cART treatment, and as the benefit of cART spread, fear associated with AIDS may dissipate, increasing participation in risk behavior. The Ugandan HIV prevalence and incidence decline followed by a recent upsurge is a classic example.[71]

6.6.2. Justification for the study

HIV incidence and prevalence trends data are key data to the implementation of prevention and treatment interventions for HIV. Therefore, monitoring HIV prevalence, especially in the 15 to 24 year-olds provide key information on the impact of HIV preventive interventions.[260] The sub-goal milestone for goal six (i.e., combat HIV/AIDS, malaria and other diseases) of the Millennium Development Goals (MDG) is improvement of information, surveillance, and monitoring and operation research systems by 2015. HIV prevalence trends assessment using ANC-based HIV surveillance data (i.e., 1994 to 2011) would provide key information for monitoring HIV epidemic, and also contribute towards improvement of HIV epidemiological information base for the HIV epidemic in Zambia.[260]

Although ANC-HIV-SS based HIV prevalence estimates may not be as reliable as DHS PBS-based HIV prevalence estimates, ANC-HIV-SS data enable estimation of HIV prevalence estimates at a level closer to the specific communities (i.e., at the district level), unlike DHS-based HIV prevalence estimates that are aggregated at provincial and nation level. Further, because ANC-based HIV prevalence surveys (i.e., every 2 to 4 years) are more frequently conducted compared to DHS-based PBS HIV prevalence surveys (i.e., no more than every five-years). HIV prevalence estimates based on ANC-HIV-SS data may provide early warning on the direction of the HIV epidemic. The ANC-HIV-SS based HIV prevalence estimates and PBS-DHS HIV prevalence are complementary in providing coherent and comprehensive data on the HIV epidemic for implementation of preventive and treatment interventions.

6.7. Methods

6.7.1. Overview of data collection methods for the ANC-HIV-SS

The ANC-HIV-SS program in Zambia is a Ministry of Health initiative implemented by the Tropical Diseases Research Centre (TDRC) in Ndola and the University Teaching Hospital (UTH) Virology Unit in Lusaka, designed to monitor trends in HIV prevalence among pregnant women attending selected sentinel sites. Seven rounds of ANC-HIV-SS data have been collected between 1994 and 2011(i.e., 1994, 1998, 2002, 2004, 2006, 2008, and 2011), and were used for the current analyses. Details of the ANC-HIV-SS were described in Chapter 4. Briefly, pre-1994 ANC-HIV-SS rounds had limited geographic coverage (i.e., conducted in only 10 sites, and largely in urban areas). The number of sentinel sites was expanded to 27 in 1994, and

impressively 22 sentinel sites have been consistently used for data collection between 1994 and 2011, and 24 sentinel sites have been consistently used for data capturing since 2002. Serenje in Central Province and Luangwa in Lusaka Province, were designated as sentinel sites in 2002, bringing the total number of sentinel sites to 24. The current analyses included data from 22 sites 22 sites since 1994 and 24 sites since 2002.

6.7.2. Study design and study population

Cross-sectional survey design was applied to collect ANC-HIV-SS data in all the seven rounds. Pregnant women who sought ANC at health centers that serve as sentinel sites for ANC-HIV-SS program constituted the study population. Different independent samples of pregnant women were recruited in each of the seven rounds of the ANC-HIV-SS, although some women might have participated in more than ANC-HIV surveillance round. Data were de-identified in all survey rounds and questions were not asked to suss whether a particular pregnant woman had participated in any of the prior surveys: thus is was impossible to identify pregnant women had participated in more than one survey round.

6.7.3. Inclusion criteria and sampling strategy

Pregnant women who sought antenatal care for the current pregnancy for the first time during the four-month survey period were eligible for the study. Pregnant women were chronologically recruited at the sentinel site (i.e., health center designated as sentinel site for HIV surveillance), non-probability convenience sampling strategy (i.e.,

non-random) that recruited nearly all eligible pregnant who presented within the four-month survey period

6.7.4. Sentinel site selection criteria

Geographic coverage for sentinel sites rests on the assumption that each site recruited pregnant women from the catchment area of the sentinel site. Urban-located sentinel sites were conveniently selected to achieve country-wide geographic coverage, whereas rural-located sentinel sites were randomly selected within each of the nine provinces.[183] Further considerations for selecting health centers as sentinel sites included the capacity of sentinel site to recruit the target number of pregnant women (~500) within the survey period (~four months) was confirmed and sites unable to accommodate this were excluded.

6.7.5. Response rate of the survey

The ANC-HIV-SS protocol mandated recruitment of all eligible pregnant women who made their first visit for antenatal care for the current pregnancy during the four-month survey period.[157] Because all pregnant women who seek antenatal care provide a venous blood sample for routine screening of syphilis (i.e., Ministry of Health care package for pregnant women in Zambia), I assumed that nearly all pregnant women provided a blood sample, part of which was used for ANC-HIV surveillance reporting.[182]

6.7.6. Specific sample size calculation

For ANC-HIV-SS rounds conducted between 1994 and 2008, each of the sentinel sites was expected to recruit at least 500 pregnant women, based on an expected HIV prevalence of 20% and desired precision of 0.35% at 95% confidence level. Most sites attained the target sample size, except sites located in sparsely populated areas (e.g., Ibenga in Copperbelt Province). Because the urban sites in Lusaka and Ndola are located in densely populated areas, the target sample size for these sites were increased to at least 800 pregnant women per site. However, a protocol change in 2011 mandated the recruitment of a minimum of only 360 pregnant women per site.[182]

6.8. Data collection

6.8.1. Sociodemographic data in pre-2011 surveys captured via questionnaire

Sociodemographic and pregnant woman's birth histories (i.e., number of children birthed by the pregnant woman) were captured via a nurse-administered questionnaire. The questionnaires was administered on the first antenatal care clinic visit for the current pregnancy for the eligible pregnant woman. During 1994-2008 period, minor changes were made to the questionnaire question-content (i.e., questions dropped or modified between 1994 and 2011) but most questions were worded in a similar manner across survey years. Consequently, data for some variables collected during 1994 through 2011 were missing in some years (e.g., educational attainment in 2011). The new survey protocol used in 2011 mandated abstraction of data from (e.g., age) from routine antenatal clinic care card. Data that are routinely collected data (i.e., data collected as part

of the routine care for pregnant women) were not captured in 2011 (e.g., educational attainment, marital status and spousal).

6.8.2. HIV serostatus determined using unlinked anonymous HIV testing strategy

Plasma/serum from venous the blood specimen provided by each of the pregnant woman was screened for the presence of HIV specific antibodies following unlinked anonymous HIV testing strategy, as recommended by the WHO.[185] Unlinked anonymous HIV testing of survey specimens precluded linkage of HIV serostatus results to identifier of pregnant women who provided the survey blood sample. To further enhance and assure unlinked anonymous HIV testing principles, questionnaire data (i.e., socioeconomic data and birth history variables) and serological data (i.e., HIV serostatus) were only linked at the data analysis stage via a distinctive identify number assigned to the pregnant women at recruitment. Consequently, ANC-HIV-SS data for the ANC-based HIV surveillance were completely anonymised

6.8.3. Unlinked anonymous HIV testing for survey plasma/serum specimens

On their first antenatal care clinic visit for the current pregnancy, pregnant women provided a venous blood specimen for routine syphilis screening. The blood specimens from each of the pregnant women were divided into two containers, one container bore the name of the pregnant woman, and was used for routine reporting of syphilis, and the other container was marked with a distinctive survey identify number (ID). The specimens in ID marked container were used for unlinked anonymous HIV testing for ANC-based HIV surveillance reporting. The guidelines on anonymous and unlinked HIV

antibody screening by the WHO were followed for HIV testing in all survey years.[182, 184, 185]

6.8.4. Three stage survey specific HIV testing algorithm

The final HIV serostatus of survey specimens in all the seven rounds of ANC-HIV-SS were based on a three-stage survey-specific serological HIV testing protocol that comprised HIV screening, confirmatory testing and tie-breaking test for discrepant test results. Site-based screening for HIV-specific antibodies using a rapid HIV test, and reference laboratories (TDRC and UTH Virology) confirmatory HIV testing (i.e., enzyme linked immunosorbent assay) of all plasma/serum specimens with HIV specific antibodies (i.e., positive HIV test result) based HIV screening test conducted at the sites. The tie-breaker test, requiring HIV testing using an independent HIV assay (i.e., Western Blot assay, Western Blot 2.2 (MP Diagnostics™) was conducted on specimens with discrepant HIV screening test results to report final HIV serostatus.[182] Details of HIV testing algorithms were provided in Chapter 4.

6.8.5. Quality control HIV testing to limit misclassification of serostatus

To limit misclassification errors and assure reliability and validity of the HIV serostatus of survey specimens, a pre-specified proportion of plasma/serum specimens (e.g., 10% in 2004, 2006, 2008, and 2011 surveys) classified as HIV-seronegative specimens at the site HIV screening were further tested according to the pre-specified survey quality control HIV testing protocol, described in details in prior reports.[24, 40, 56, 182, 183] Specific details of sensitivities and specificities of commercial brands of

HIV test assays used in the seven surveys have been published previously. In 2011survey round, Vironostika® anti-HIV plus, Enzygonst® HIV integral, and Western blot 2.2 (MP Diagnostics) were used for screening, confirmatory and discrepant HIV test result resolution.[24, 40, 56, 182, 183]

6.8.6. Criteria for determining HIV serostatus of survey specimen

Survey specimens were classified as HIV seronegative or seronegative depending on whether HIV specific antibodies were detected or not detected. HIV seropositive serostatus was reported if both the site-based screening HIV test result and the reference laboratory confirmatory test HIV result were positive. Where a survey specimen was classified as HIV seronegative during site-based HIV testing, and the specimen not selected in the 10% quality control testing sample, the specimen was classified as HIV seronegative. For specimens where the site-based HIV screening test result and reference-laboratory HIV confirmatory test result were discrepant, a different test assay (i.e., tie-breaker test such as a Western blot) was performed and result reported as final HIV serostatus.[40, 56, 157, 182, 183]

6.9. Data management for analysis of trends in HIV prevalence, 1994-2011

Data management for specific survey round was explained in Chapter 4. The seven data sets (i.e., i.e., 1994, 1998, 2002, 2004, 2006, 2008 and 2011) were merged to facilitate assessment of trends in HIV prevalence trends between 1994 and 2011. Because regression models used to assess trends in HIV prevalence required a variable representing survey years, I created new variable that represented the calendar year in the

merged data set. The analyses were restricted to pregnant women aged 15 to 44 years. Data management and analyses were performed using R-statistical software version 3.0.[188]

6.9.1. The 15 to 24 year-olds were used to assess trends in HIV prevalence

The current HIV prevalence trend analyses were examined using data from pregnant women aged 15 to 24 years surveyed between 1994 and 2011. UNAIDS asserts that HIV prevalence trends in the 15 to 24 year-olds are less influenced by AIDS-related mortality compared to 25 to 44 year-olds (thus may not be used to approximate trends in the number of new HIV infections). Some of the current approaches to direct measurement of HIV incidence are less widely accepted; therefore my HIV prevalence trends analysis, as prior analysis by other investigators, focused on the 15 to 24 year-olds.

6.10. Variables included in the analyses

The current analyses focused on variables (e.g., age, parity, year of survey, residence) which were captured in all seven rounds of the ANC-HIV-SS conducted between 1994 and 2011 except educational attainment which was not collected in 2011 survey

6.10.1. HIV serostatus

The outcome variable was HIV serostatus, defined as HIV seropositive (i.e., detection of HIV-specific antibodies) and HIV seronegative (i.e., non-detection of HIV specific antibodies).

6.11. Primary exposure variable

6.11.1. Survey year

The primary exposure variable was the survey calendar year: year in which the ANC-HIV-SS round was conducted, and represented as a continuous variable. Survey year was centered on 1994 achieved by subtracting 1994 from each of the survey calendar year.

6.12. Other covariates

Four variables were assessed in the current analyses. First, parity, defined as self-reported number of children birthed by pregnant woman. Pregnant women were grouped into three categories of parity: no child, one child and two or more children. Second, residence was defined according to the site location (i.e., according to the urban-rural classification of areas by the Government of the Republic of Zambia) as pregnant women were regarded as having been recruited from the site catchment area. Third, Educational attainment was measured as the number of schooling years completed by the pregnant woman, and only captured in 1994 through 2008 rounds of ANC-HIV-SS. Because fewer pregnant women reported educational attainment beyond 17 years, maximum number of schooling years in the study sample was assumed to be 17 schooling years. Consequently, educational attainment greater than 17 schooling years was coded as 17 schooling years.

Additionally, to enable assessment of trends in HIV prevalence within categories that corresponded to the education system in Zambia, pregnant women were grouped as follows: lower primary (0 to 4 years); upper primary (5 to 7 years); junior secondary (8 to

9 year); incomplete senior secondary (10 to 11); and complete secondary school and higher (12 to 17 years). Educational attainment was modeled as a continuous in all regression-based analyses.

6.12.1. Descriptive statistics were conducted to characterize the study sample

Median and interquartile range IQR) were computed to describe the distribution of continuous variables (e.g., age and number of school years completed), because compared to the mean, the median as a measure of central tendency is less sensitive to outliers. To describe distribution of categorical variables, counts and percent frequencies were computed. Cognizant of the harmful influence of a large proportion of missing data on parameter estimates, where data are not missing completely at random, the extent of missing data was explored, and missing data reported as counts and percent frequencies.

6.12.2. Wilson's score method was used to compute 95% confidence interval (CI) for HIV prevalence

The 95% CI for HIV prevalence estimates were calculated using the Wilson's score method. While the 95% CI could have been computed by the Wald methods as is commonly practiced, my analyses benefited from the less conservative and but tighter 95% CI generated by Wilson's method. Further, Wilson's method estimated confidence intervals have better coverage probability and provide consistent approximation of the nominal confidence intervals for proportions.[261] On the other hand, Wald method may generate erroneous confidence intervals with poor coverage probability, specifically in small samples, and when the proportion is close to zero or one.[261]

6.13. Generalized linear mixed model (GLMM) applied to examine the relationship between survey calendar year and prevalent HIV infection

The relationship between survey calendar year and prevalent HIV infection was modeled using GLMM adjusted for age (i.e., age fitted as RCS function), and sentinel site fitted as a random component to account for possible intra-site clustering. Intra-site clustering may threaten validity of inference (i.e., erroneous standard error of model parameter estimates, invalid confidence interval and p-values) because data were collected from several sentinel sites (i.e., 22 to 24 sentinel sites).Within-site clustering might be present in the data because pregnant women who sought antenatal care from the sentinel site might have been more similar than pregnant women who sought antenatal care from another sentinel site. Stringer et al (2008) used a similar strategy to describe trends in HIV prevalence in health centers in Lusaka, Zambia.[168, 238]

6.13.1. Logit link function for GLMM because outcome was dichotomously defined: HIV seronegative or HIV seropositive

To describe overall trends using GLMM, I assumed that the HIV prevalence could be explained by a set of fixed effects and a set of random effects, and also that the HIV seroprevalence follows a Bernoulli distribution. Because HIV seroprevalence followed a Bernoulli distribution (HIV seropositive or seronegative), the logit link function was applied in the GLMM. The lme4 library in R program version 3.0, a statistical analysis and computing program R version was used for fitting GLMM.[188, 262] Laplacian approximation was used to estimate parameters in GLMM because there is no simple closed-form solution for estimation of the likelihood function in GLMM when the outcome is dichotomous.

6.13.2. The likelihood ratio tests (LRT) was applied to detect significant HIV prevalence trends by specific covariates

To assess presence of trends in HIV prevalence by specific covariates, cross-product terms between survey calendar year and specific covariate were created (e.g., time*residence). Briefly, two nested models with main effects covariates were fit to the same data, with and without cross product interaction term, and the difference in log likelihood values evaluated via the LRT. The LRT rests on the theory that the difference in log-likelihood of a pair of nested models fitted to the same data is assumed to follow an approximate Chi-distribution. For all LRT conducted for the current study, p-value <0.05 were interpreted as presence of statistical multiplicative interaction, and justified stratified analyses.

6.13.3. Trends in HIV prevalence by pregnant woman's age were assessed via a multiplicative statistical interaction between calendar time of survey and age

To examine whether HIV prevalence trends differed by age, a cross-product term between age and survey calendar year was created, and LRT used to evaluate the presence of statistical multiplicative interaction. The survey calendar year was centered by subtracting 1994 from each of the seven survey calendar years. Consequently, the values for the centered-survey year for the survey calendars years 1994, 1998, 2002, 2004, 2006, 2008 and 2011 were modeled as 0, 4, 8, 10, 12, 14 and 17 respectively.

Centered-survey year was fit as primary exposure variable and HIV serostatus as outcome variable in a GLMM where the random component was sentinel site. Two nested GLMM models with main effects covariates, and with and without cross-product term were fit to assess the presence of trends in HIV prevalence by age. The first GLMM

comprised only main covariates (i.e., age and survey year). The second GLMM contained main effects covariates and the cross-product term (i.e., age*centered survey year). The log-likelihoods of the two nested models fitted to the same data were compared using the LRT. A p-value less than 0.05 suggested presence of statistical multiplicative interaction in HIV prevalence trends between survey year and pregnant women's age.

6.13.4. Trends in HIV prevalence by urban and rural residence were evaluated via cross-product between residence and centered survey year

Cognizant of the prior literature that the burden of HIV infections among sexually active persons is higher in urban than in rural areas of SSA, I evaluated the presence of meaningful trends in HIV prevalence by residence via LRT for two nested GLMM fitted to the same data, the first GLMM with main covariates (urban and survey time) only, and the second GLMM another with main covariates and the cross-product interaction term between survey calendar year and survey calendar time (urban*centered survey year).

6.13.5. LRT test used to assess linearity assumption for the relationship between age and prevalent HIV infection

To evaluate whether the relationship between a pregnant woman's age and prevalent HIV infection was best captured by a linear function of a pregnant woman's age, two nested GLMMs were fit with sentinel sites fit as a random component to capture intra-site clustering effects in the relationship between survey year and prevalent HIV infection. In the first GLMM, it was assumed that age was linearly related to log-odds of prevalent HIV infections, and in the second GLMM, it was assumed that age was

nonlinearly (i.e., continuous variable fit using RCS function) related to log-odds of prevalent HIV infection.[149, 263]

LRT was applied to assess whether the GLMM in which age was assumed to be nonlinearly related to log-odds of prevalent HIV infection provided meaningfully better description of the relationship above and beyond the GLMM model that assumed a linear relationship between age and log-odds of prevalent HIV infection.[149, 264, 265] Linearity assumption was tenable if the p-value associated with LRT was >0.05, and nonlinearity was tenable if LRT p-value was <0.05. Therefore, a significant LRT p-value implied that the relationship between age and log-odds of prevalent HIV infections might be non-constant over the observed range of age, and would be captured adequately by fitting age a RCS function.[149, 263, 266]

6.13.6. Statistical multiplicative interaction between sentinel site and survey year was evaluated using cross-product term via LRT of nested models

Assessment of the multiplicative interaction between calendar year and site location was motivated by prior reports that have indicated differential burden of HIV infections across site in different geographic areas. To objectively rationalize stratified assessment of HIV prevalence trends by sentinel sites, LRT test was applied to assess whether trends in HIV prevalence differed by sentinel sites via evaluation of statistical multiplicative interaction between calendar survey year and sentinel site. Survey calendar year was centered by subtracting 1994 from respective survey calendar years. Two nested GLMM fitted to the same data with main effects covariates (i.e., site and centered survey year), with and without cross-product term (i.e., site*survey year) were fit to detect meaningful statistical multiplicative interaction between site and survey year using LRT. A p-value <

0.05 indicated presence of meaningful statistical multiplicative interaction between survey calendar time and sentinel site.

6.13.7. Within-site HIV prevalence trends were evaluated using generalized linear models (GLM)

HIV prevalence trends in all the 24 sentinel sites were evaluated using GLM with centered survey calendar year as primary exposure variable and HIV serostatus as outcome variable. As with GLMM an analysis, the logit link function was applied because the outcome variable, HIV serostatus, was dichotomously defined (HIV seronegative or HIV seropositive). The primary exposure variable, survey calendar years, was centered by subtracting 1994 from each survey year, and centered survey year was fit as RCS function with pre-specified knots at 10^{th}, 50^{th} and 90^{th} percentile.

6.13.8. Complete case analysis was used to estimate regression model parameters

Proportions of missing data on variables (e.g., residence, age, survey year and HIV serostatus) considered for the analysis in assessing trends in HIV prevalence were not substantial, representing < 5% missing data. Consequently, the HIV prevalence trends analyses were based on complete case analysis, although the assumption of missing completely at random required for complete case analysis cannot be confirmed. However, given that the sample size used for the analysis was large, the computed proportion of observations with missing data might not substantively considerable.

6.14. Results

6.14.1. Description of study sample

The merged data set contained records for 82,561 pregnant women aged 15 to 44 years recruited in the 1994, 1998, 2002, 2004, 2006, 2008 and 2011 rounds of ANC-based HIV surveillance. The number of pregnant women recruited for ANC-based HIV surveillance between 1994 and 2011 ranged from 8881 in 2011 to 13,298 in 2008.

6.14.2. Nearly all records (99.4%) among 15 to 44 year-olds had serologically confirmed HIV serostatus

Out of the 82561 records of pregnant women aged 15 to 44 years in the merged data set, 82,086 (99.4%) had serologically confirmed HIV serostatus data. The current analysis was restricted to 82086 pregnant women aged 15 to 44 years who had serologically confirmed HIV serostatus result. Among the pregnant women aged 15-44 years recruited during 1994 to 2011 period, 54.4% (44683/82,086) were aged below 25 years, and constituted the sub-sample used for investigating trends in HIV prevalence (i.e., approximating the number new HIV infections). Among HIV seropositive pregnant women, 46.5% (7071/15,505) were aged 15 to 24 years, 46.4% (7195/15,505) were aged 25 to 34 years, and 8.0% (1239/15,505) were aged 35 to 44 years. Of the 82086 pregnant women, 57.7% (47,400/66,581) were recruited from sites located in sites located in urban areas. Among HIV seropositive women, 11,600 (74.8%) were recruited in urban sites. Among HIV seronegative pregnant women, 35,800 (53.8%) were recruited from sites in urban areas. Table 6.1 presents the descriptive characteristics of pregnant women aged 15 to 44 years recruited between 1994 and 2011. Table 6.2 provides year-specific

distribution of pregnant women characteristics for survey years 1994, 1998, 2002, 2004, 2006, 2006, 2008 and 2011

6.14.3. Higher proportion of pregnant women had two or more children

Parity was defined according to the number of children birthed by a pregnant woman. Overall, 28.9% (23,208/80,443) of the pregnant women were nulliparous (i.e., no child), 22.1% (17,779/80,443) had one child, and 49.0% (39,456/80,443) two or more children. Comparison of pregnant women by HIV serostatus revealed that there were differences in distribution of parity as indicated in Table 1 (p-value=0.001). For example, among HIV seronegative pregnant women, 30.1% (19,626/65,113) had no child compared to 23.4% (3582/15,330) HIV seropositive pregnant women who did not report having a child.

6.14.4. Slight increase in the median age of pregnant women who participated in the ANC-HIV surveillance

The median age for HIV seronegative women was 23 years (IQR= 20 to 29 years), and among seropositive pregnant, the median age was 25 years (IQR= 22 to 29 years). Assuming that the highest number of schooling years completed by pregnant women in the study sample was 17, the overall median educational attainment was 7 schooling years and IQR was 5 to 9 years. Educational attainment data were missing for all the pregnant recruited in the 2011 ANC-based HIV surveillance. The median age of pregnant women who participated in the ANC-HIV surveillance in Zambia between 1994 and 2011 seems to have increased slightly as indicated in the box plot in Figure 6.1.

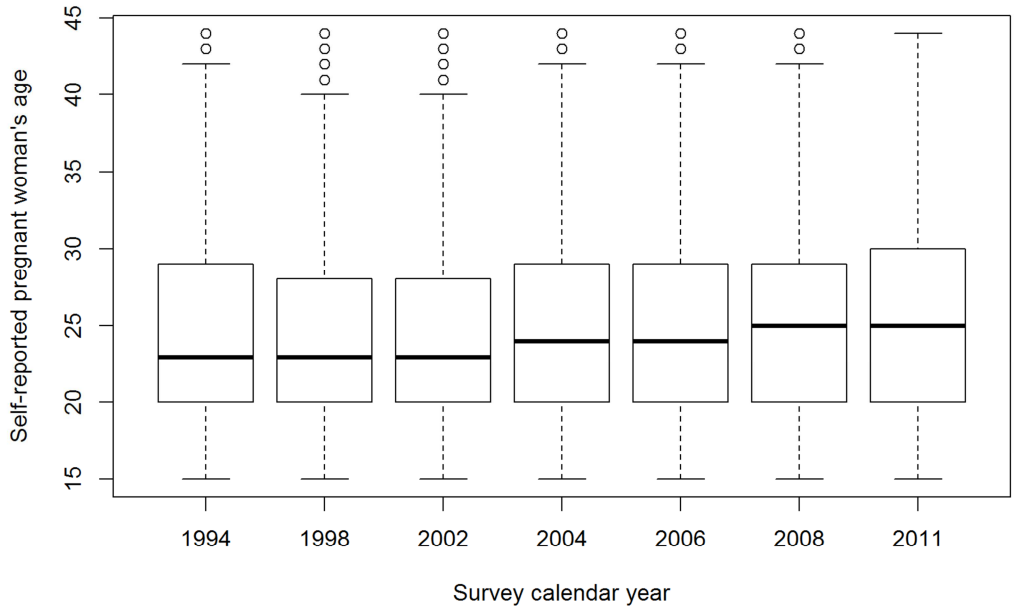

Figure 6.1. Year-specific distribution of self-reported age among pregnant women aged 15 to 44 years who participated in ANC-based HIV sentinel surveillance conducted between 1994 and 2011 in Zambia. The thick horizontal lines in the boxes indicate median age whereas upper and lower horizontal boundaries of the box indicate upper and lower interquartile respectively. The open circles located beyond 1.5 times that IQR indicate observations that may be outliers.

6.15. Trends in prevalent HIV infections varied by age

Whether HIV prevalence trends vary according to age was assessed by a cross-product interaction term between year (i.e., centered year on 1994) and self-reported pregnant woman's age. Two nested GLMM with main effects covariates, with and without cross-product term were evaluated via LRT as explained in the method's section. The LRT p-value comparing log-likelihood of nested GLMMs yielded was <0.05, thus implying presence of meaningful variation in prevalent HIV infection by age (p-value <0.001).

Additionally, to align HIV prevalence estimates reporting to age groups used by WHO/UNAIDS age groups (i.e., 15 to 19, 20 to 24, 25 to 29, 30 to 34, 35 to 39 and 40 to

44), age-group specific analysis of trends in HIV prevalence were conducted as displayed in Figure 6.2a for urban areas and Figure 6.2b for rural areas. HIV burden was higher in urban sites than rural sites, and the sharp decline in HIV prevalence trends between 1994 and 2011 in the 15 to 19, 20 to 24 year-olds was most marked in urban areas. HIV prevalence in the 15 to 24 year-olds declined in rural areas, but less than in urban areas.

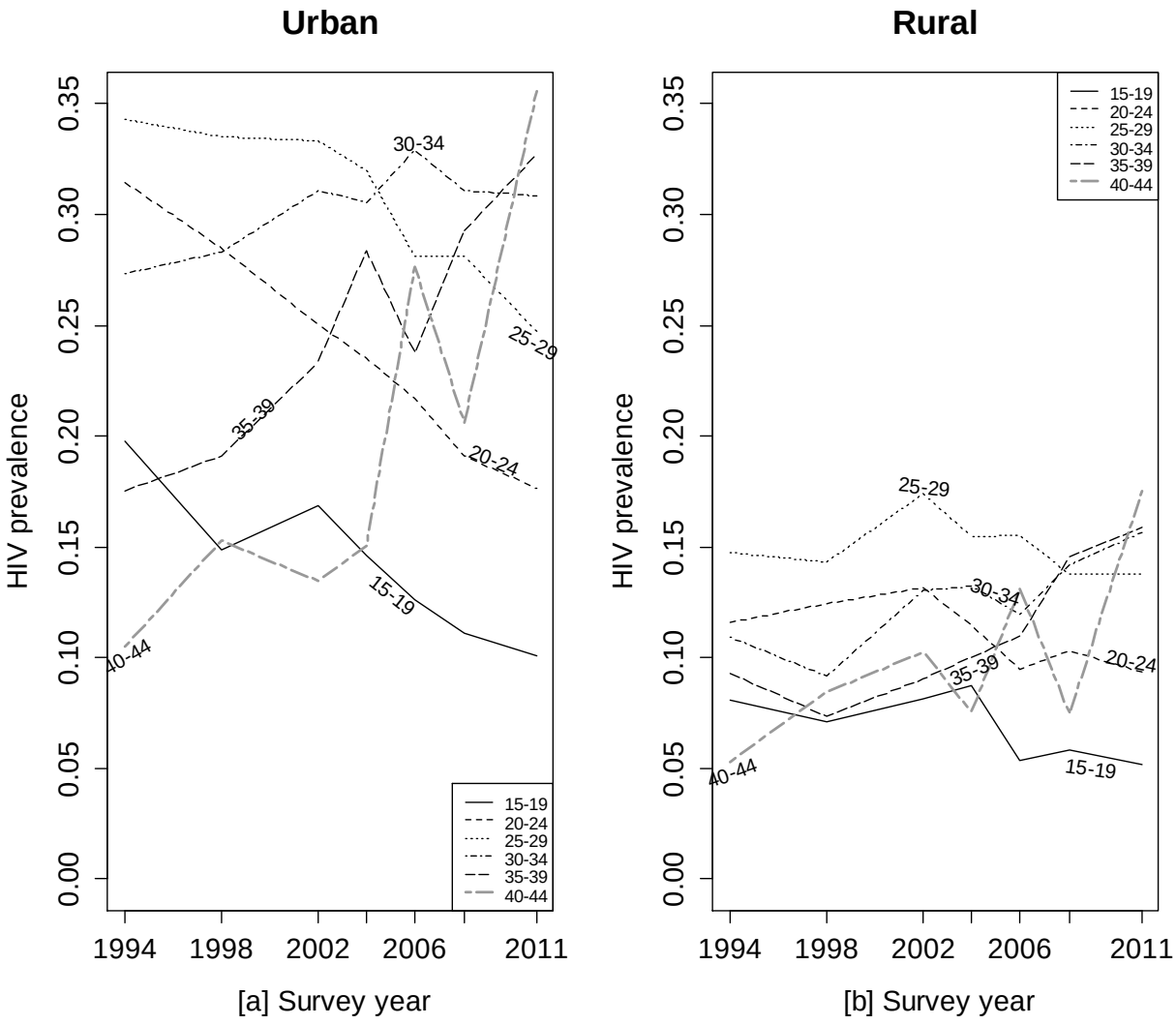

Figure 6.2.Age group-specific HIV prevalence trends stratified by urban and rural residence among pregnant women aged 15 to 44 who participated in the ANC-HIV-SS between 1994 and 2011 in Zambia. The labels on the lines (i.e., 15-19; 20-24; 35-29; 30-34; 35-39 and 40-44) indicate the age group. Declining HIV prevalence in the 15-19 and 20-24 age groups is noted, and a more pronounced burden is seen in urban than rural areas.

6.15.1. The trends in HIV prevalence were different by urban and rural location of sentinel site (i.e., proxy for residence)

To investigate whether HIV prevalence varies by urban or rural residence, a cross-product interaction term between residence and survey year (i.e., centered on 1994) was created. Using the LRT, two nested GLMMs, with and without cross-product term were compared as explained in the method section. The p-value for the LRT that assessed whether there was meaningful variation in prevalent HIV infection by urban or rural residence yielded a statistically significant (p-value <0.001). Because LRT p-value was less than 0.05, separate analyses for investigating trends in HIV prevalence were conducted for urban and rural areas. Figure 6.3a and Figure 6.3b shows HIV prevalence trends within the 15 to 24 year-olds and 25 to 44 year-olds in urban and rural areas, respectively.

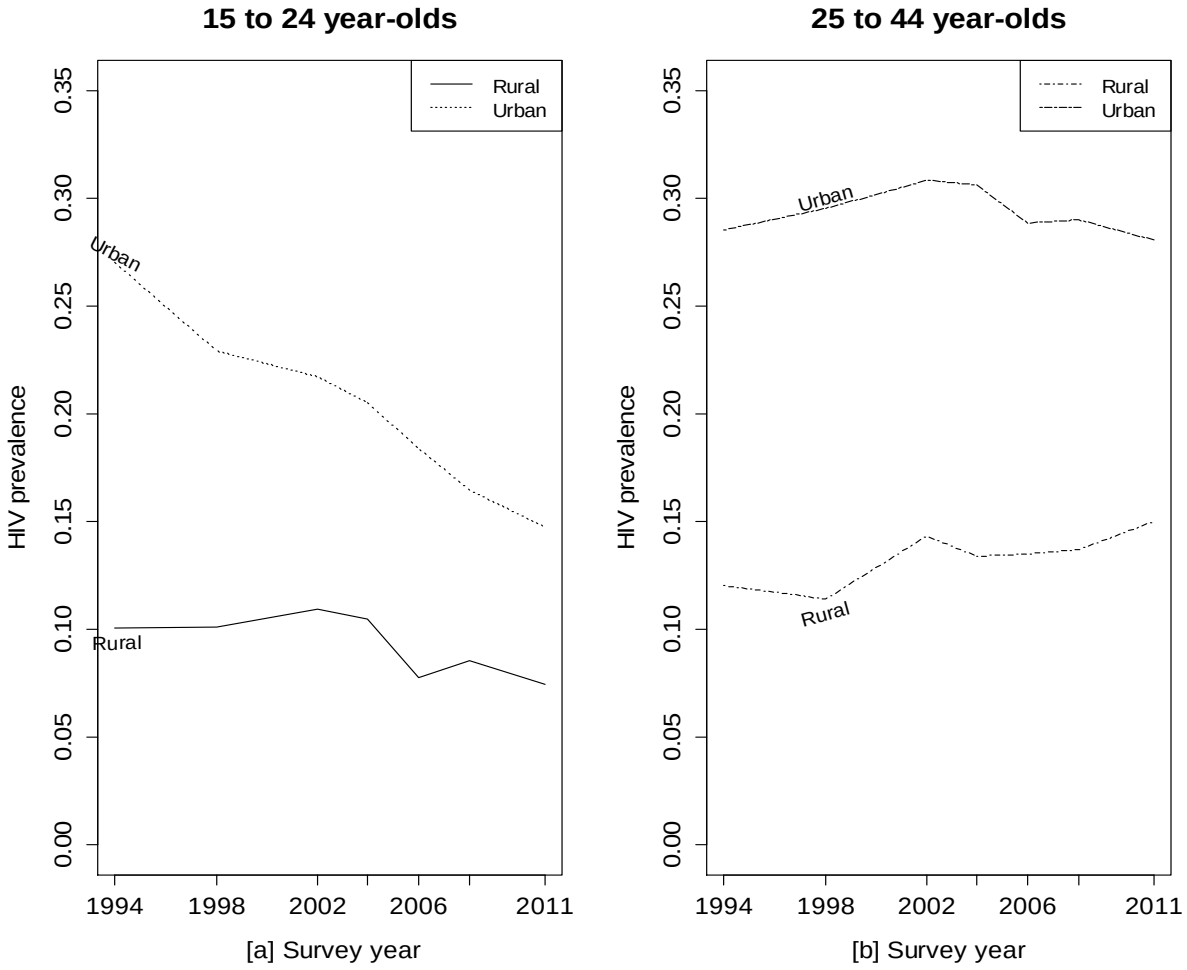

Figure 6.3. HIV prevalence among 15 to 24 year-olds pregnant women (Fig. 6.3a) and 25-44 year-olds pregnant women (Fig. 6.3b) stratified according to urban and rural residence for pregnant women attending antenatal care and surveyed in the ANC-HIV-SS from between 1994 and 2011 in Zambia.

6.15.2. HIV prevalence trends, although in the same direction, differed by sentinel site

To address prior literature that the burden of prevalent HIV infection is heterogeneously distributed in different geographic areas of Zambia, a cross-product interaction term was created between centered survey year and sentinel site. The LRT based on two nested GLMMs, with and without a cross-product interaction term, were compared (see Methods). The p-value for the LRT that assessed whether HIV prevalence

trends varies by sentinel site yielded a statistically significant (p-value <0.001). Consequently, separate investigations of HIV prevalence trends by sentinel sites were conducted. Figure 6.3 shows different but largely declining HIV prevalence trends by urban and rural residence among 15 to 24 year-olds in most urban between 1994 and 2011. Figure 6.4 and Figure 6.4 shows site-specific HIV prevalence trends among 15 to 24 year-olds pregnant women in urban and rural sites indicating largely declining but non-uniform HIV prevalence trends across sentinel sites between 1994 and 2011

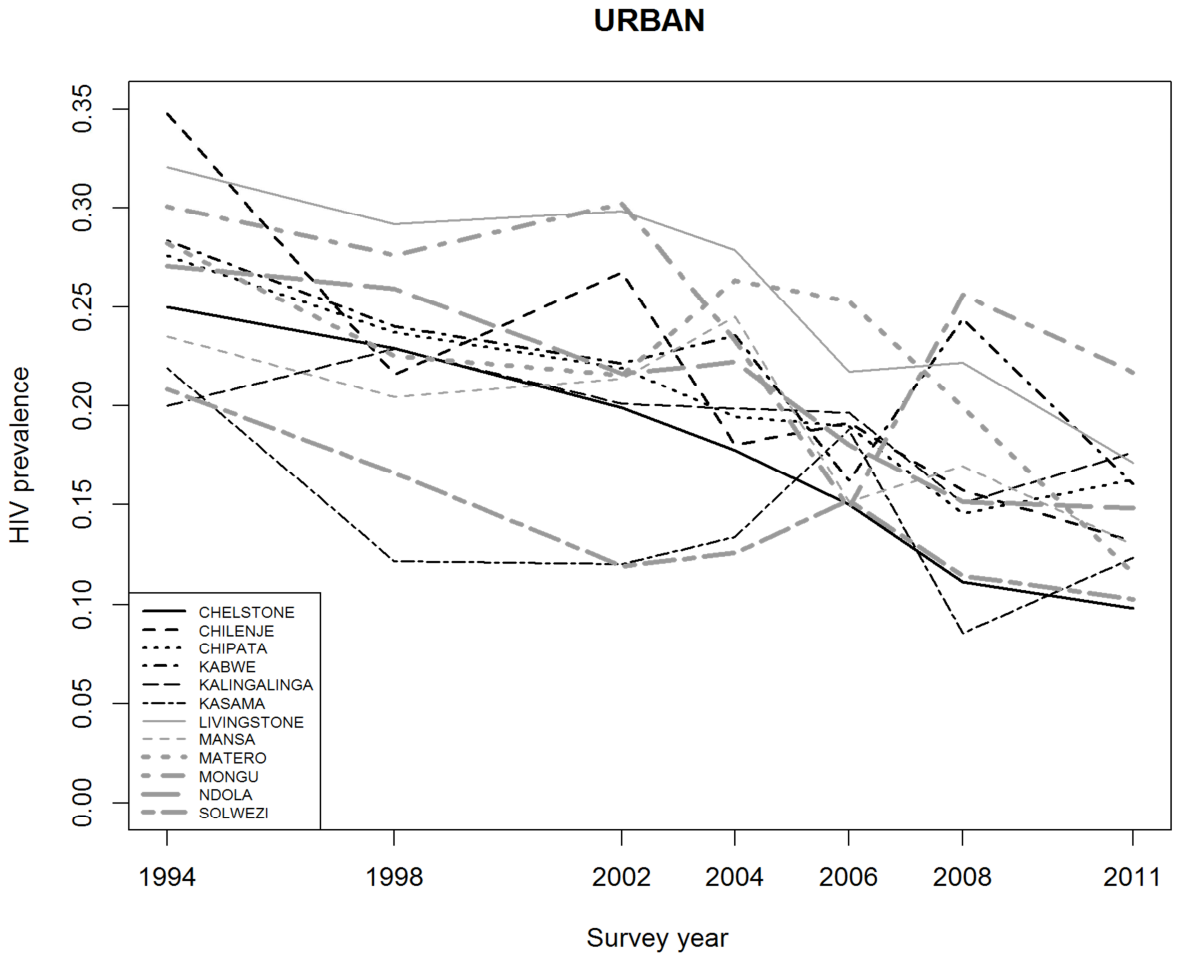

Figure 6.4. Site-specific HIV prevalence trends among 15 to 24 year-olds pregnant women recruited from sentinel sites located in urban areas in Zambia based on ANC-HIV-SS data collected between 1994 and 2011. An overall pattern of declining HIV prevalence trends noted most sites (Chelstone, Chilenje Livingstone, Matero, Ndola, and Solwezi). HIV prevalence increased between 2008 and 2011 in Chipata,

Kalingalinga and Mongu. The HIV prevalence trends in the remaining three sites (i.e., Kasama) were less clear (i.e., fluctuating and without a consistent direction).

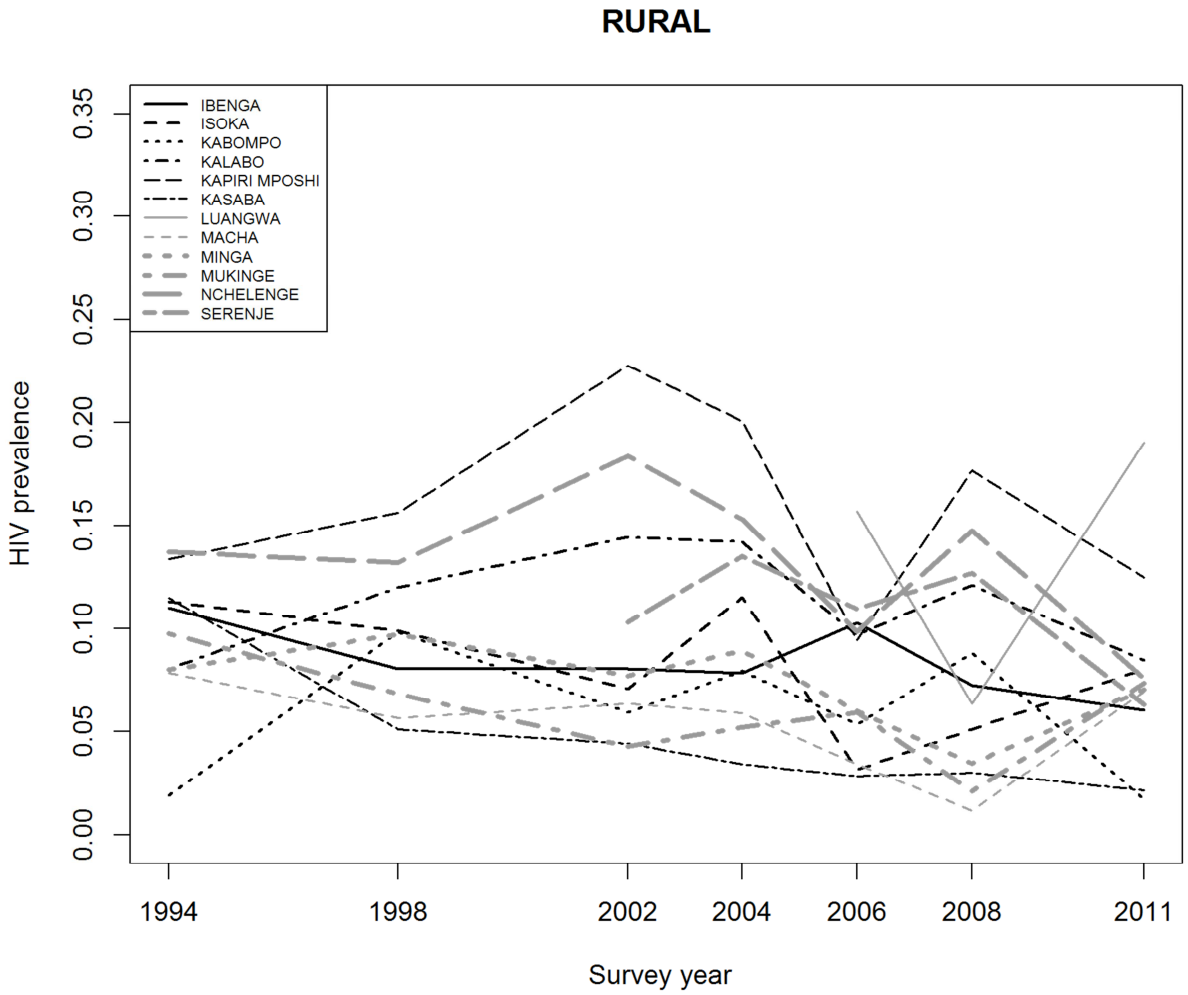

Figure 6.5.Site-specific HIV prevalence trends among pregnant women recruited at sites located in rural areas in Zambia based on ANC-HIV-SS data collected between 1994 and 2011. Upward swings in HIV prevalence noted in five sites between 2008 and 2011; HIV prevalence rose in Macha (1.2% to 7.0%), Minga (3.5% to 7.1%), and Mukinge (2.1% to 7.6%). Declining HIV prevalence was observed in Kasaba from 3.0 % in 2008 to 2.2 % in 2011 and a sharp drop was noted in Kabompo from 8.8% in 2008 to 1.7% in 2011. HIV prevalence trends among pregnant women aged 15 to 24 years in Isoka rose from 3.1 % in 2006 to 8.0% in 2011. The patterns in HIV prevalence trends in the remaining six sites (e.g., Kalabo) were less clear and displayed fluctuating pattern.

6.16. Overall HIV prevalence trends in sentinel sites located in urban areas and in rural areas

HIV prevalence among 15 to 24 year-olds in urban areas fell from 27% in 1994 to 16.2% in 2011, and HIV prevalence for rural sites fell from 10% in 1994 to 7.4% in

2011 (Table 5). Figure 6.3a and Figure 6.3b shows HIV prevalence trends among 15 to 24 year-olds and 25-44 year-olds pregnant women between 1994 and 2011 according to rural and urban location of sentinel sites.

Using the LRT for assessing whether HIV prevalence trends were meaningfully different between sites (p-value <0.001). Because LRT p-value was < 0.05, separate analyses of trends in HIV prevalence were performed for each sentinel sites. Table 6.3 and Table 6.4 presents HIV prevalence estimates for specific survey years per site for the seven survey rounds in the 15 to 24 year-olds.

6.16.1. Site-specific assessment of trends in HIV prevalence diverse but largely declining patterns in HIV prevalence

HIV prevalence trends in all the 24 sites were investigated by modeling the relationship between survey calendar year and prevalent HIV infection non-linearly. Centered survey year was fit using RCS function with knots at 10^{th}, 50^{th} and 90^{th} percentile in a GLM. Slight nonlinear trends were detected in some sites (e.g., Kabompo, Kalabo, Kapiri Mposhi and Mansa). HIV prevalence trends (Figure 6.4 and Figure 6.5) seemed to decline linearly in most sites, but modest nonlinear HIV prevalence trends were detected in some (e.g., Kabompo, Kapiri Mposhi, Kalabo and Nchelenge). Further, HIV prevalence in most rural sites were lower (e.g., Kabompo, Kalabo and Mukinge) than HIV prevalence in urban sites (e.g., Chelstone, Livingstone, and Ndola). **Error! Reference source not found.** and **Error! Reference source not found.** shows smoothed trends in HIV prevalence between 1994 and 2011 in selected sites.

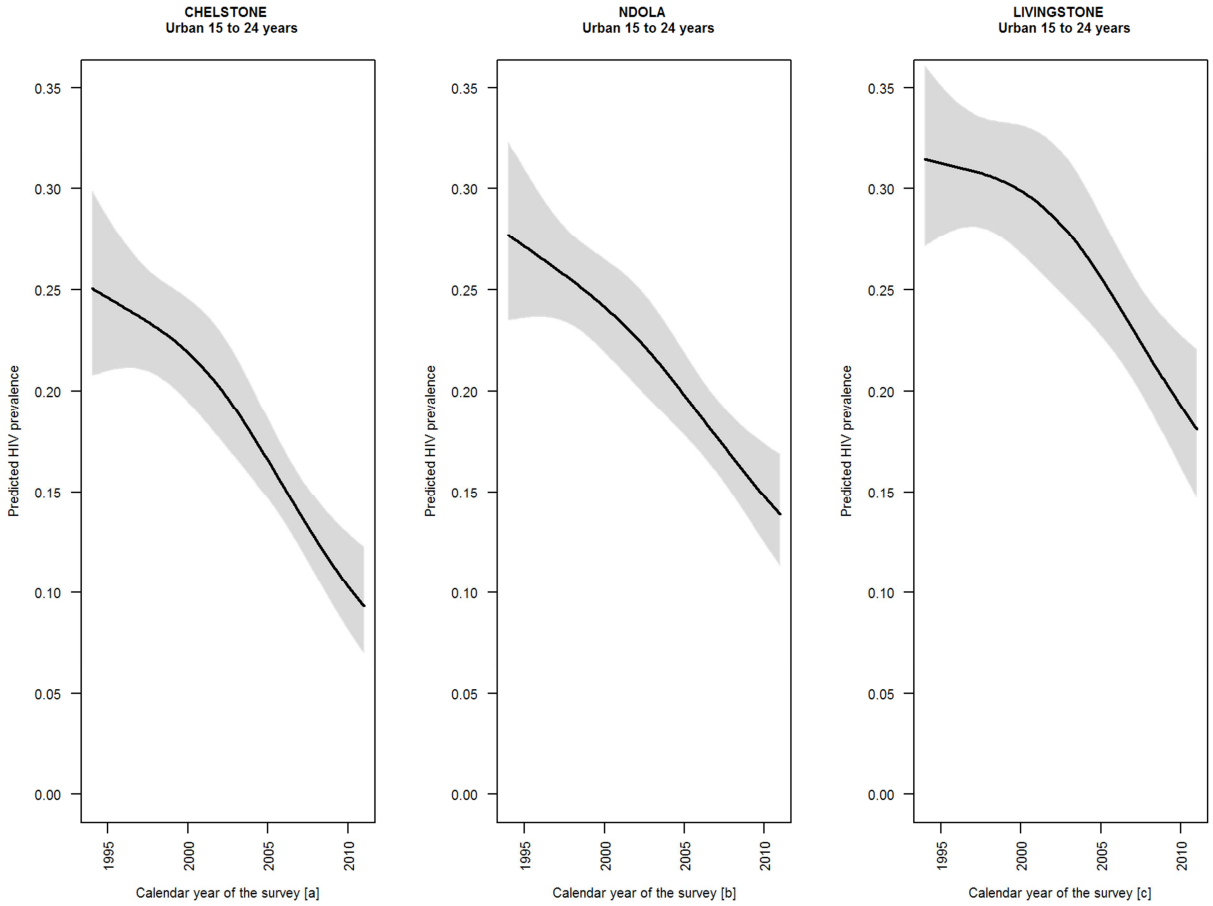

Figure 6.6.Smoothed HIV prevalence trends among pregnant women aged 15 to 24 years in selected sentinel sites (i.e., Chelstone, Ndola and Livingstone) in urban areas showing the observed nearly linear pattern of decline in the burden of prevalent HIV infections. Data were collected between 1994 and 2011 during ANC-HIV-SS in Zambia. Non-linear relationship between survey year and log-odds of prevalent HIV infections were explored by fitting survey as a continuous variable using restricted cubic splines function with knots placed at 10^{th}, 50^{th} and 90^{th} percentile via a generalized linear model [GLM] with a logit link function. The grey shades represent the 95% CI.

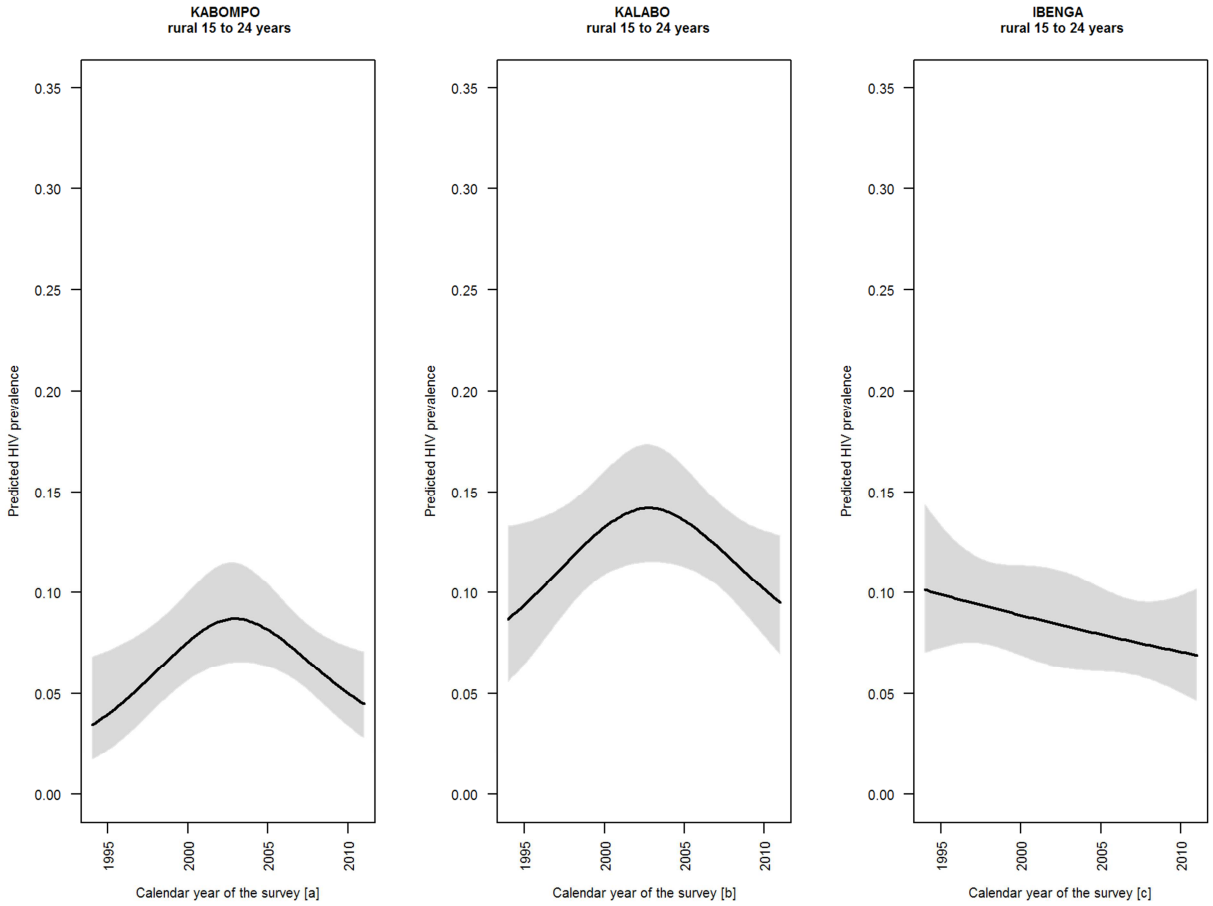

Figure 6.7. Smoothed HIV prevalence as a function calendar survey year among pregnant women aged 15 to 24 years in selected sentinel sites (i.e., Kabompo, Kalabo and Ibenga) urban areas showing the observed nearly linear pattern of decline in the burden of prevalent HIV infections. Data were collected between 1994 and 2011 during ANC-HIV-SS in Zambia. Non-linear relationship between survey year and log-odds of prevalent HIV infections were explored by fitting survey as a continuous variable using restricted cubic splines function with knots placed at 10^{th}, 50^{th} and 90^{th} percentile via a generalized linear model [GLM] with a logit link function. The grey zone indicates the 95% CI.

6.16.2. Site-specific HIV prevalence trends in sentinel sites: urban

Data from 12 sites located in urban areas were used to assess trends in HIV prevalence trends among 15 to 24 year-olds between 1994 and 2011 (Table 6.3 and Table 6.4). Along with site-specific HIV prevalence estimates by survey year, respective 95% Wilson confidence interval are presented in Table 6.3 and Table 6.4. HIV prevalence

trends were declines that were largely linear in six urban sites (Chelstone, Chilenje, Livingstone, Matero, Ndola and Solwezi) between 1994 and 2011, but a recent upward swing was noted in 2011 in two of the urban sites (Kalingalinga in Lusaka Province and Chipata in Eastern Province) that had exhibited a prior near-linear decline in HIV prevalence between 1994 and 2008. Near-linear declining trends in HIV prevalence were observed in most sites (Figure 6.4 and Figure 6.5). HIV prevalence estimates for the 25 to 44 year-olds are provided in Table 6.5 and Table 6.6.

6.16.3. Site-specific HIV prevalence trends in sentinel sites: rural

Site-specific estimates of HIV prevalence and corresponding 95% Wilson C.I.s are presented for 15 to 24 year-olds based on data from 12 sites located in rural areas in Table 6.3and Table 6.4. Although less profound declines in HIV prevalence in rural sites are noted compared to declines observed in urban sites, declining HIV prevalence trends were observed in between 1994 and 2008 in Isoka in Northern Province, Kasaba in Luapula Province, Macha in Southern Province, and Mukinge in Northern Province. Between 2006 and 2011, HIV prevalence in Ibenga declined from 10.3% to 6.1% in 2011.

HIV prevalence trends among pregnant women aged 15 to 24 years in Isoka rose from 3.1 % in 2006 to 8.0% in 2011. HIV prevalence in rural sites was generally lower than in most urban sites except Luangwa, where HIV prevalence was 19.5% in 2002, and dropped to 6.3% in 2008, but swung upwards in 2011 at 19.0%. Further assessment of HIV prevalence trends among 25 to 44 year-olds are provided in Table 6.5 and Table 6.6.

HIV prevalence swung upward in 2011 in rural Isoka, Macha, and Mukinge sites where consistent declines in HIV prevalence had been noted between 1994 and 2008 as shown in **Error! Reference source not found.**. In contrast, HIV prevalence in Kasaba declined steadily from 11.5% in 1994 to 2.2% in 2011. Trends in HIV prevalence in Kabompo in Northwestern Province were less clear, but a sharp drop in HIV prevalence was observed from 8.8% in 2008 to 1.7% in 2011. A similar sharp drop in HIV prevalence were observed in Mukinge in Northwestern Province, when HIV prevalence dropped from 5.9% in 2006 to 2.1% in 2008 as shown in Table 6.4.

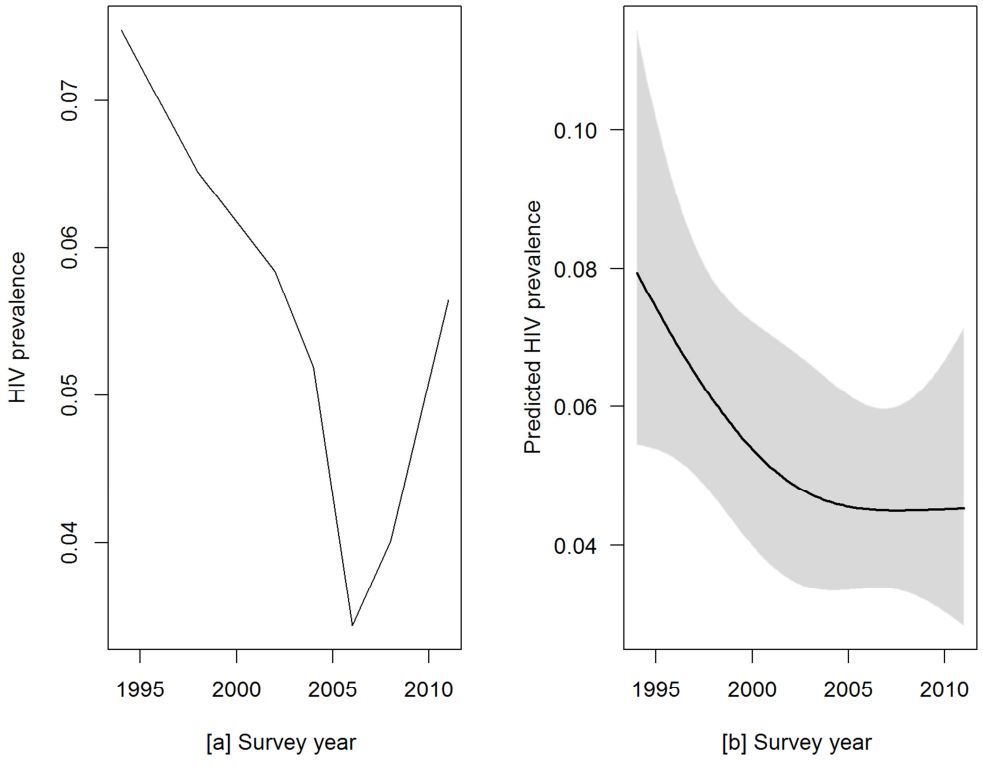

Figure 6.8. HIV prevalence trends among pregnant women aged 15 to 24 year-olds recruited between 1994 and 2011 in the Macha, one of the rural sentinel sites for the ANC-HIV-SS located in southern Zambia. Data were collected in 1994, 1998, 2002, 2004, 2006, 2008, and 2011. Fig.6.8a shows non-smoothed HIV prevalence trends in Macha between 1994 and 2011. Centered survey calendar year (i.e., survey year minus 1994) was fitted using restricted cubic spline function with knots located at the 10^{th}, 50^{th} and 90^{th} percentiles in Fig.6.8b.

6.16.4. Heat map for representing interaction between calendar survey year and age

To facilitate the interpretation of statistical multiplicative interaction effects between continuous survey calendar year and pregnant woman's age, I created a heatmap to capture visually the statistical multiplicative interaction effects between pregnant woman's age and survey calendar year based on a GLMM that included a main effects (age and survey year) and cross-product interaction term (age*survey year). Consistent with findings reported in prior studies, the odds of prevalent HIV infection were lower for younger pregnant women in recent surveys compared to older women as depicted in Figure 6.9.

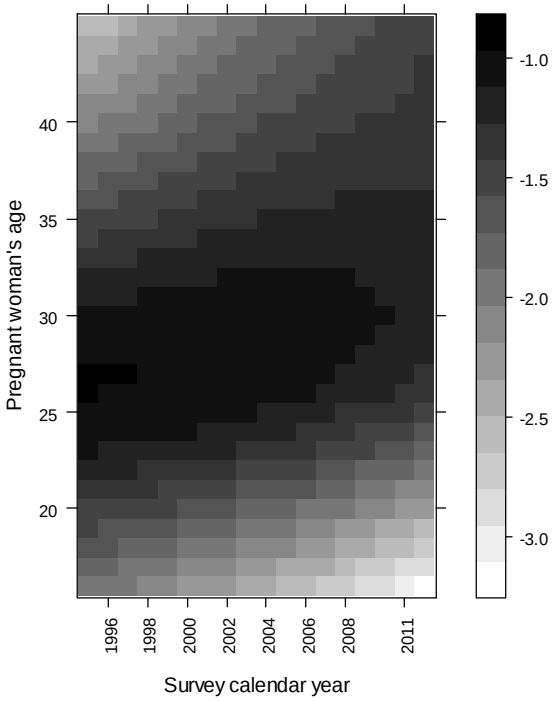

Figure 6.9. Heat map indicating statistical multiplicative interaction effects of the cross-product term between survey year and age based on ANC-HIV-SS data collected between 1994 and 2011 in Zambia. The legend key indicates color intensity values with heavy color intensity corresponding to greater odds of prevalent HIV infections and light color intensity representing lesser odds.

6.16.5. HIV prevalence trends by educational attainment: 1994 to 2011

HIV prevalence was highest among women with highest educational attainment (Figure 6.10). HIV prevalence declined profoundly among pregnant women who self-reported more education. Between 1994 and 2011 in rural areas, HIV prevalence among pregnant women in the lowest category (i.e., 0 to 4 schooling years) of educational attainment changed very little, with slight increase between 1994 and 2004 but a slight decline in 2011,(Figure 6.10). However, HIV prevalence in the lowest category declined only slightly in urban areas. Between 1994 and 2002, HIV prevalence was highest among pregnant women who reported incomplete senior secondary school (i.e., 10-11) in urban areas (Figure 6.10).

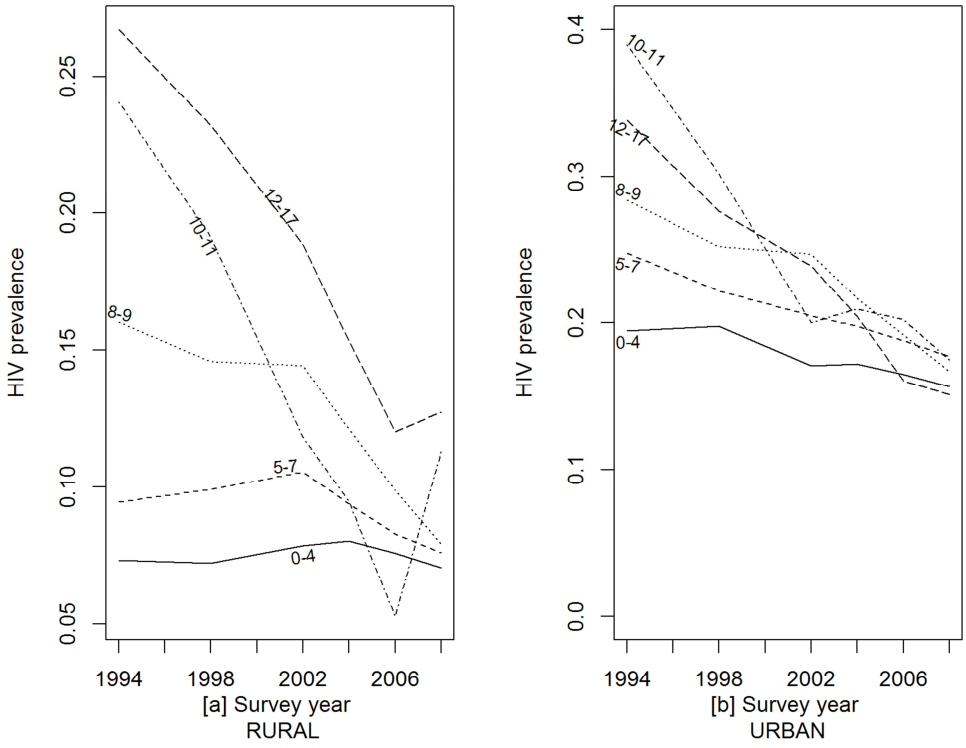

Figure 6.10. HIV prevalence trends by educational attainment among pregnant women aged 15 to 24 years based on the ANC-HIV-SS data collected between 1994 and 2008 inclusive. The labels on the curves (i.e.,

0 to 4; 5 to 7; 8 to 9; 10 to 11; and 12 to 17) represent self-reported number of schooling years completed by pregnant women.

6.16.6. Trends in HIV prevalence assessed using age-only adjusted GLMM

The LRT were applied to two nested GLMM models, one with and without a cross-product term, where the cross-product was used to evaluate statistical multiplicative interaction between survey calendar year and rural/urban residence, the p-value of for the LRT was <0.05. Because the computed LRT p-value <0.001, separate age-only adjusted GLMM (i.e., sentinel sites as random components) were fitted for urban and rural areas sites. Figure 6.11 and Figure 6.12 shows the predicted probabilities of prevalent HIV infection for pregnant women at specific ages based on the age-only adjusted GLMM, where age was fit using RCS function with knots at the 10^{th}, 50^{th} and 90^{th} percentile. To enhance communication, graphed estimates are presented and numerical values of estimates are presented in Table 6.7 and Table 6.8 for urban and rural areas respectively.

6.16.7. Declining predicted probability of HIV prevalence between 1994 through 2011 across all ages in urban and rural areas

Figure 6.11 show shows decline in predicted HIV prevalence in urban areas for age 15, 19, 21 and 24 years, with a higher predicted probability with increasing age, specifically in earlier surveys

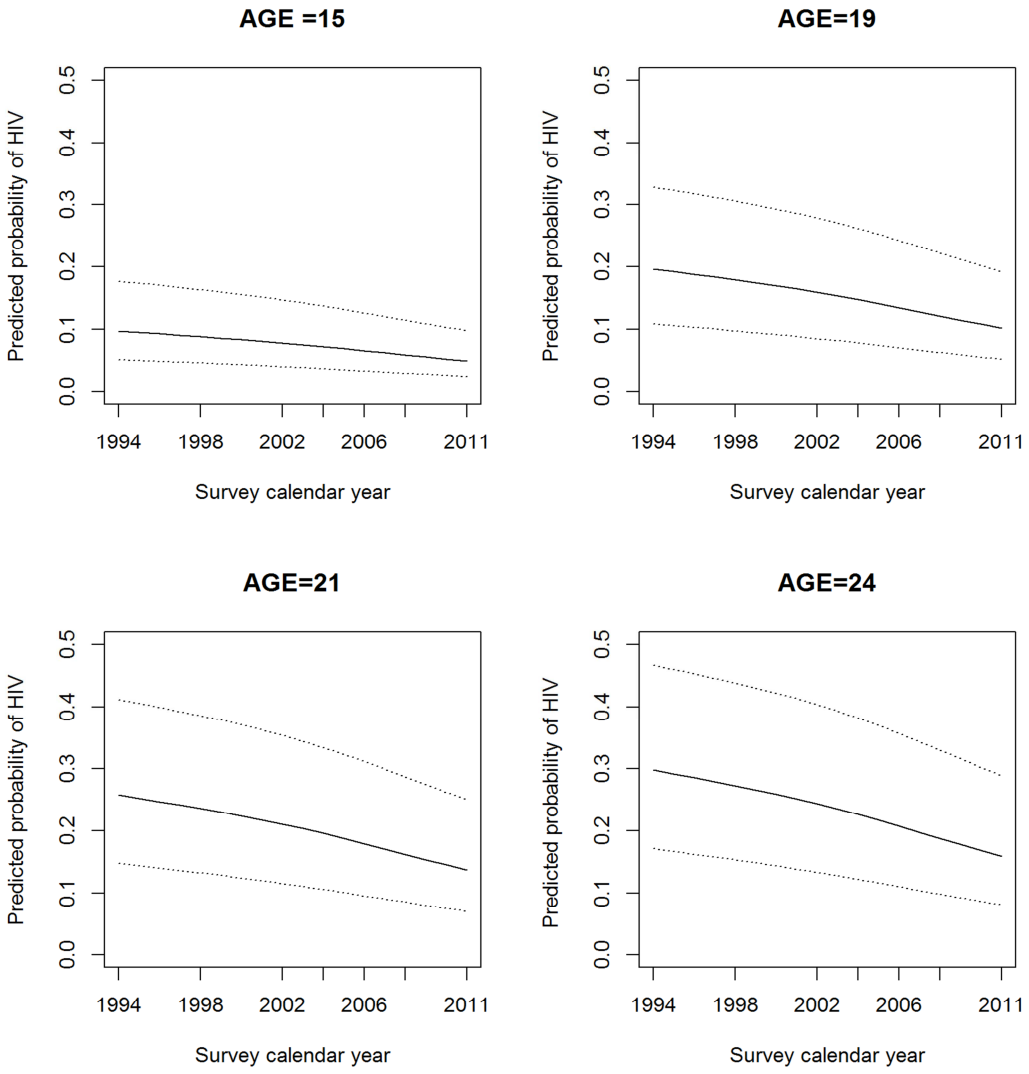

Figure 6.11. The predicted probabilities of prevalent HIV infection for specific age value (i.e., 15, 19, 21 and 24) across different survey calendar years 1994-2011 among pregnant women aged 15 to 24 years who participated in the ANC-HIV-SS in sentinel sites in urban areas in Zambia. Predicted probabilities were computed from a GLMM that examined the relationship between survey year and prevalent HIV infections. The age-only adjusted GLMM was using logit link function and random component for sentinel site. Parameter estimates were by Laplacian approximation. Age and survey calendar year were nonlinearly fit as continuous variables using RCS function with three pre-specified knots the 10th, 50th and 90th percentiles.

The trend in HIV prevalence seen in urban areas was similar to the trend is seen in rural areas (Figure 6.12), but a lower predicted probability of prevalent HIV infection for pregnant women was noted compared to pregnant women in urban areas (Figure 6.11).

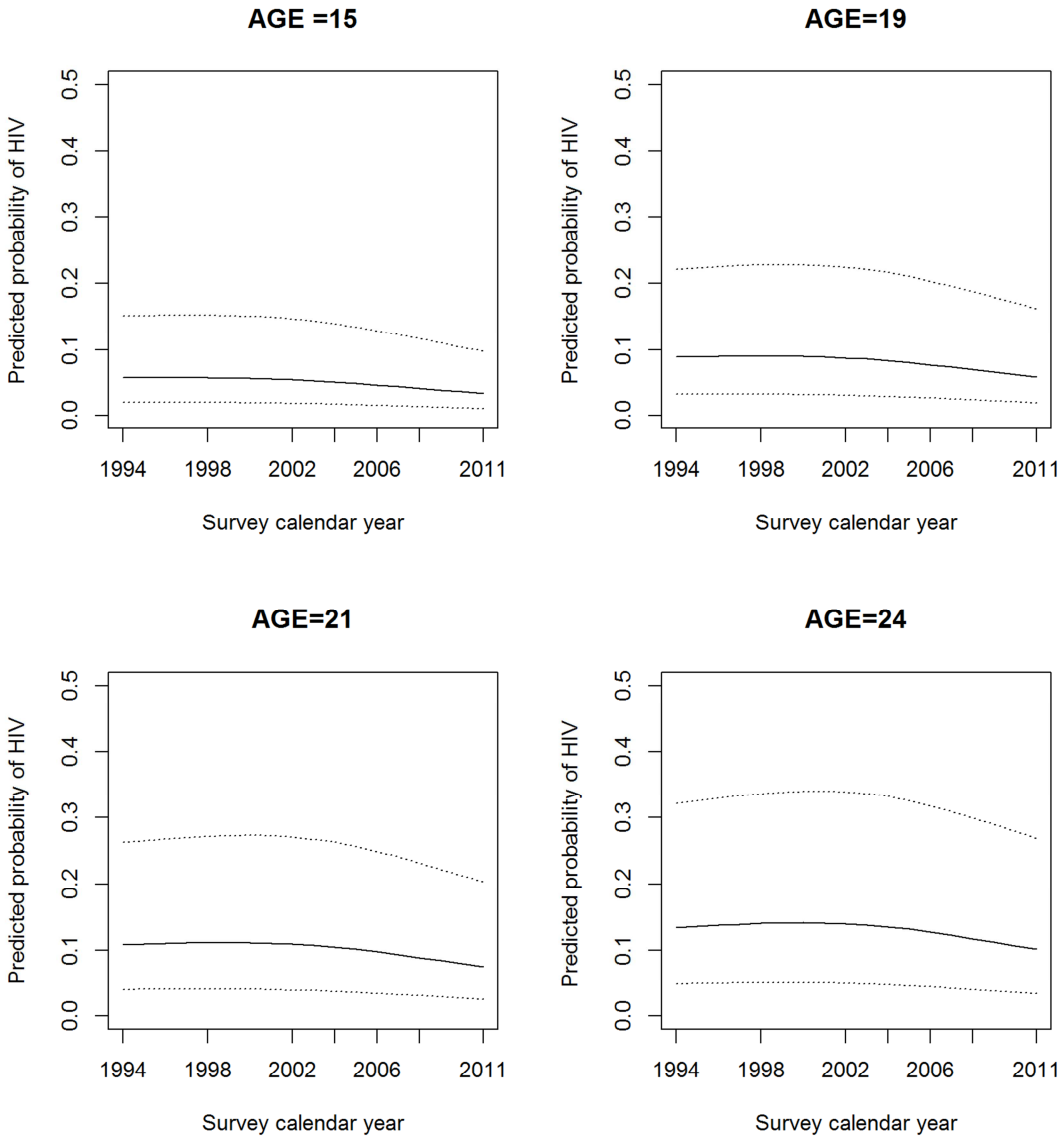

Figure 6.12. The predicted probabilities of prevalent HIV infection in across seven survey rounds during the 1994-2011 among pregnant women aged 15 to 24 years who participated in ANC-HIV-SS in sentinel sites in rural areas in Zambia. The predicted probabilities were estimated from the regression model of the relationship between survey calendar year and prevalent HIV infections using age only-adjusted GLMM with a logit link function and a random component for sentinel sites. Age and survey calendar year were fit nonlinearly as continuous variables using RCS function with three pre-specified knots at the 10th, 50th and 90th percentiles.

6.17. Elevated odds of prevalent HIV infection with increasing age among urban pregnant women

Compared to a pregnant woman age 15 years in 1994, the odds of prevalent HIV infection for a 24 year-old pregnant woman were higher in both urban (OR=3.96, 95% CI: 3.83, 4.09) and rural areas (OR=2.53, 95% CI: 2.40, 2.66). As shown in Table 6.7 and Table 6.8 based on age 15 years as the referent age, gradual increase in odds of prevalent HIV infection were noted with increasing age among pregnant women aged 15 to 24 years within each of the seven survey rounds.

Figure 6.13 shows that odds of prevalent HIV infections among pregnant women aged 15 to 24 years in urban areas were slightly more than 3.5 times higher for a 24 year-old compared to 15 year-olds in 2011 (OR=3.65, 95% CI: 3.54, 3.77).

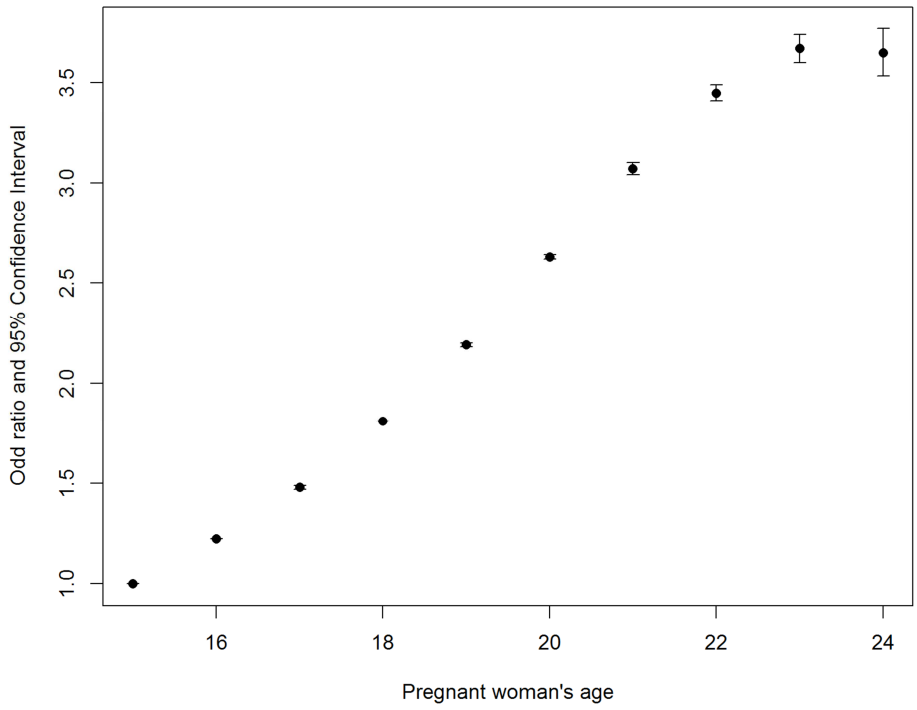

Figure 6.13. Odds ratios for specific ages for pregnant women aged 15 to 24 years based on data collected from the ANC-HIV-SS in urban areas in Zambia in 2011. The odds ratios were computed from the age-only adjusted GLMM where age was fit using restricted cubic spline function with knots placed at 10th, 50th, and 90th percentile. The graph shows that the odds of prevalent HIV infections rose steadily with increasing age of pregnant woman.

6.17.1. Elevated odds of prevalent HIV infections with increasing age among rural pregnant women

Among pregnant women in rural areas, and using 15-year-olds as referent group, the odds of prevalent HIV did not change much over time, but appear to have increased for rural pregnant women. For example, the odds of prevalent HIV infections of a 24 year-old compared to a 15 year-olds in 2011 were OR=3.22, 95% CI: 3.07, 3.38. On the other hand, the prevalent HIV infection for 24 years compared to 15 year old in 1994 was OR=2.53, 95% CI: 2.40, 2.66 in 1994 in rural areas.

6.18. Discussion

ANC-based HIV surveillance data were used to examine trends in HIV prevalence among pregnant women, and findings revealed a profound decline in HIV prevalence among pregnant women aged 15 to 24 years over the 18 year period studied in Zambia. An overall decline in urban areas was a near halving of the prevalence rate, from 27% in 1994 to 14.7% in 2011. In rural areas, a lower HIV prevalence was noted, and this also declined substantially from 10% in 1994 to 7.4% in 2011, a less dramatic fall than seen in urban areas. Prevalence remains high, reflecting the intensity of HIV transmission in southern Africa.[10]

Comparison of trends in HIV prevalence by educational attainment category indicated downward trends in HIV prevalence within all groups of educational attainment, but more profound decline in HIV prevalence were noted among pregnant women in urban areas who were in categories representing more education than among pregnant women in categories with less education.

The odds of prevalent HIV infections for pregnant women within age range 15 to 24 years indicated that the odds of prevalent HIV infection increased gradually with increasing age of pregnant woman. Even though the predicted probability of prevalent HIV infections declined at all ages during the period 1994 through 2011, the predicted probability was higher for older than younger women, reflecting the increasing aggregate risk of HIV infection over decades of sexual activity.

Even though my analysis revealed an overall downward trend in HIV prevalence, the burden of HIV infections and trends in HIV prevalence across the 24 sites examined

were heterogeneous. For example, there were more sites with downwards trends in urban areas than in rural areas. Further, HIV prevalence swung upward in 2011 in three urban sites (Kasama, Kalingalinga and Chipata) and three five sites (Isoka, Luangwa, Macha, and Minga), suggesting a need for special prevention interventions in this communities.

The findings from my study are comparable to the findings by Kayeyi et al (2012) who examined trends in HIV prevalence in Zambia among 15 to 24 year-olds using DHS-based HIV prevalence data and ANC-based HIV prevalence data. Kayeyi et al (2012) noted downward trends in HIV prevalence from 1994 to 2008 based on ANC-HIV surveillance data as well as a heterogeneous burden of prevalent HIV infections across sites. Further, estimates by Kayeyi et al (2012) based on DHS-based HIV prevalence estimates revealed a decline in HIV prevalence from 2001 to 2007consistent with other earlier reports.[40, 146]

Further to providing prevalence estimates, unlike Kayeyi et al 2012, I have provided 95% Wilson's CI for estimated HIV prevalence across sites and within subgroup to enable the reader to judge the precision of HIV prevalence estimates. A remarkable strength of the study by Kayeyi et al (2012) was the use of population-based data from the 2001 and 2007 DHS in Zambia. Extending my analysis with DHS-based HIV prevalence data was not possible as there been no new DHS-based HIV prevalence data since 2007 in Zambia, preparation of the third round of DHS are underway. A further novel approach in the current analysis was use of restricted cubic splines function to model continuous variables in the age adjusted analysis to explore non-linear trends in HIV prevalence within sites and overall. Further, the current study considered within-

site clustering by modeling sentinel site as a random effect similar to the approach by Stringer et al (2008).[168, 267]

The 2011 ANC-HIV-SS data revealed spikes in HIV prevalence in sites such as Isoka, Luangwa, Macha, Minga, and Mukinge. The upward swings HIV prevalence noted in some sites are worrisome, but can be interpreted in the context of the downward trends in HIV prevalence in the 15 to 24 year-olds. The spike in HIV prevalence in the mentioned sites might reflect a true increase in HIV prevalence or might be due to random variation. The extent to which recruitment of fewer pregnant women per site in 2011 (i.e., 360 per site compared to 500 per site prior years) has impacted HIV prevalence estimates is uncertain. Further data are needed to confirm the observed spikes in HIV prevalence in 2011 as true increases in HIV prevalence in affected sites (i.e., Isoka, Luangwa, Macha, Minga, and Mukinge). Predicted prevalence estimates regression models fitted using RCS function provides more conservative results that are consistent with national trends.

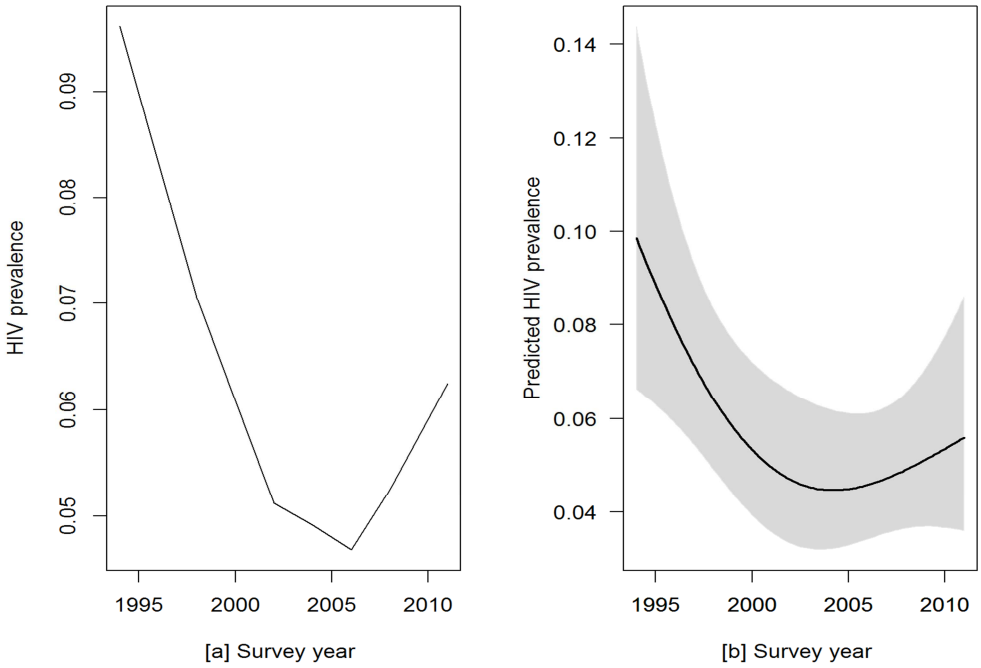

Figure 6.14. Unsmoothed trends in HIV prevalence trends for Mukinge (Fig. 13a), with prevalence trends after survey years were fitted using RCS for data collected in 1994, 1998, 2002, 2004, 2006, 2008 and 2011 (Fig. 6.14). ANC-HIV-SS data used and GLM fitted to generate Fig.6.14b.

The upward spikes in HIV prevalence in a number of communities raises concerns: reminiscent of the HIV epidemic dynamics in Uganda, where after a remarkable decline in HIV prevalence, the burden of prevalent HIV infection swung upwards.[268-270] However, more data are required to confirm that the observed spikes in selected communities are an indication of the true increases in HIV prevalence, given the continued overall declining trends. Examination of trends in HIV prevalence by flexibly fitting survey years as RCS suggested a decline in HIV prevalence, with high uncertainty, as displayed by the confidence interval (Figure 6.14). Estimation of HIV prevalence

trends in sites where spikes have been observed requires further investigations to uncover dynamics of the HIV infections.

My findings are both encouraging and raise some concerns. Observed overall decline in HIV prevalence in 15 to 24 year-olds is encouraging because the fall in HIV prevalence invariably suggests a likely decline in HIV incidence and is consistent with trends reported in other SSA countries.[271-274] AIDS-related mortality rate is lower at least in the 15 to 24 year-old and therefore may not be a contributing factor to the observed decline, assuming HIV prevalence is not affected by migration rates. Although the observed HIV prevalence estimates among pregnant women suggest a decline between 1994 and 2011, the HIV burden remains high in most sites. Therefore, intensified research are required to understand what may be driving the high prevalence in those sites where HIV prevalence spiked upwards, and devising preventive interventions for curbing HIV transmission.[275-278]

Given the availability of cART, intensive combination prevention strategies to thwart the occurrence of new HIV infections remain the best option for managing the HIV epidemic, including expanded cART coverage to reduce infectiousness of HIV-infected persons.[29] Data from the ANC-HIV-SS is useful in assessing changes in HIV prevalence over time because ANC-based HIV surveillance is more frequently (every 2-4 years) conducted compared to population based DHS-based HIV surveillance which less frequently conducted, most recently in 2001 and 2007 in Zambia. ANC will highlight women of childbearing age, but it is hard to imagine an effective prevention intervention that would not be reflected in data from that population.

There are several interventions aimed at modifying sexual risky behavior that have been implemented in Zambia, with greater intensity in urban areas where HIV prevalence has been high. It is difficult to singly credit an intervention as having influenced the trends in HIV prevalence among the 15 to 24 year-olds. However, one may also attribute the decline in HIV prevalence among 15 to 24 year-olds to the cumulative effects from numerous interventions implemented since the advent of the HIV epidemic.[30, 279-292]

HIV prevention interventions have focused on modifying risky sexual behavior modifying interventions. For example, these interventions include community-level interventions such as incorporating HIV and AIDS educational in the school curriculum, condom promotion, and influencing social norms to promulgate reduction in the number sexual partners.[30, 282, 283, 290-294] An increased intensity of interventions in urban areas may have contributed to this marked decline in HIV prevalence, though we cannot be sure. The reduction in infectiousness of cART-treated HIV-infected persons ("treatment as prevention") may have also contributed in recent years, but would not have been a factor in earlier downward trends before HIV therapy was available widely. Because of the historically high HIV prevalence in urban areas, compared to rural areas, its rates of decline appear more marked, but caution must be exercised in making inferences about the comparative effectiveness of interventions in urban than rural areas. Also Wilson (2012) suggest that economic boom in copper mining towns corresponded to substantial reduction in transactional sex and multiple partnership.[295]

The HIV assays used for HIV serostatus ascertainment were not consistent across the seven survey rounds.[24, 182] The HIV assays used for HIV screening and confirmatory testing had were of very high sensitivities and specificities, and stringent

HIV testing algorithms guidelines were adhered to in all survey years. As with most studies based on ANC-based HIV surveillance data, quality control efforts for assuring the validity of HIV serostatus were focused on specimens that are reactive (seropositive) to the screening HIV antibody testing.[185, 228, 232] All specimens designated as HIV seropositive in the initial site screening were further confirmatory-tested using an HIV ELISA assay but only 5% to 10% of the specimens designated as HIV seronegative in site screening were further tested as part of reference laboratory quality control testing. Misclassification of HIV positive specimens as HIV seronegative were assumed to be few given the stringent survey HIV testing algorithm.[182, 185]

6.18.1. Study limitations

Present study findings must be interpreted in the context of the following limitation of the ANC-HIV-SS data. Neither the sentinel sites nor the pregnant women were selected via random sampling. The sentinel sites were selected on the basis the need to achieve nationwide geographic coverage of the survey, and the potential of the site to recruit the set sample size per site. Therefore, although geographic distribution of sites provides impressive geographic coverage, the sentinel sites used for ANC-HIV-SS may not be representative of all health centers used as antenatal care clinics in the nine provinces of Zambia, estimated at 1400 health centers countrywide in 2005.[296, 297]

Nonetheless in my study, I assumed that urban sites are a likely reasonable representation of other health centers with an urban catchment population, while rural sites likely represented health centers in a rural catchment population. Because the pregnant women recruited during ANC-based HIV surveillance are a convenient sample,

generalization of study findings is limited to that population. Young women are an important sentinel population to assess HIV transmission dynamics.[24, 274, 298, 299].

No records were maintained in ANC-HIV-SS to track the number of pregnant women who refused to participate in the survey. Therefore, it is not possible to ascertain the survey response rate. Anecdotal evidence suggests that nearly all eligible pregnant women were recruited for the ANC-based HIV prevalence surveys at both urban and rural sites, but this cannot be confirmed.

Some pregnant women might have been recruited in more than one round of ANC-based HIV surveillance, therefore might be a source of within-pregnant woman clustering effects. Because the ANC-based HIV surveillance follows an unlinked and anonymous strategy, pregnant women who participated in more than one survey during the 1994 to 2011 period cannot be identified, and consequently this source of within-pregnant woman clustering effects may be unaccounted for and may lead to invalid inferences (i.e., p-values and confidence intervals).

The findings of this study may be influenced by selection bias. Selection bias may threaten validity of inferences proportion of pregnant women do and not seek antenatal care are different. However, according the Ministry of Health in Zambia, more than 95% of pregnant women will seek antenatal care at least once during pregnancy.[36] However, in some parts of rural Zambia, home-based deliveries are still common and some women may miss antenatal care, so the generalizability of my findings is limited to pregnant women who sought antenatal care. [296, 297, 300-302]

HIV serostatus was objectively ascertained using serological means but the other covariates were captured via self-report. Here as in other studies, self-reported data are subject to recall, social desirability, and intentional mis-report bias. Based on the assumption that self-report biases are constant across survey years, self-report bias is unlikely to have influenced the findings to the extent of changing any of my substantive conclusions. The ANC-based HIV surveillance data do not contain information on the sexual behavior of pregnant women or of their spouse. Sexual behavior information can shed light on the observed patterns of HIV prevalence, but ANC-HIV-SS lack sexual behavior data by design in an effort to increase participation rates.

Drastic changes in the population structure of the catchment area of the sentinel site may influence the trends in HIV prevalence. For example if there is immigration of HIV seropositive women in catchment, HIV prevalence may swing upwards if they are captured during ANC-based HIV surveillance. HIV prevalence also declines in the face of high death rates, a factor to consider in some communities, especially in the pre-treatment era.[303]

As the number of HIV-infected persons accessing cART increase, appropriateness of using the number of prevalent infection in the 15 to 24 year-old to approximate the number of new HIV infections may become somewhat less valid.[24] Younger women who were infected in infancy through mother to child transmission (MTCT) may survive into the 15 to 24 year-old age group, and may become pregnant. However the survival of HIV infected babies has only improved in recent times when cART has become relatively more available than in the past, and therefore may not materially impact the observed trends in this study.

6.18.2. Strengths of the study

The study benefited from the large number of pregnant women (i.e., 82,086) from a wide geographic coverage (i.e., sites from different regions of Zambia), and with diverse social, economic and cultural backgrounds. Even if there were some selection bias, the large sample size for the study improves its potential generalizability to pregnant women. HIV serostatus of survey specimens were serologically-confirmed. The three stage testing protocol ensured that all reported HIV seropositive were confirmed, but despite the stringency of HIV testing process, some HIV seropositive specimens might have been missed specimens classified as HIV seronegative specimens or some specimens were collected from HIV infected but antibody-negative pregnant women. The number of false negative specimens is unlikely to be high, and may not change the substantive conclusion of the study because the assays used were of high sensitivity and specificities.

The sampling and covariate measurement methods implemented in six rounds of the surveys (i.e., 1994 to 2008) were similar. Although slightly different methods were used in 2011, the study population and sampling approached remained unchanged. Unlike population-based survey for HIV prevalence measurement, where refusal bias is one of the main threats to validity, ANC-based HIV surveillance is less threatened by refusal bias because nearly all eligible pregnant women who present at the antenatal care clinic were likely recruited. Social desirability bias may threaten the validity of study findings. However, the nurse-administered questionnaires possibly augmented data collection from most women, even from poorly literate or disabled (e.g., blind) women, and enabled clarification of questions. This study provides a unique opportunity to examine trends in HIV prevalence over an extended period (i.e., 18 years).[304]

HIV prevalence trends within sentinel sites were investigated by fitting survey year as a RCS function of each site. Although linear assumptions might be relaxed by categorization, and subsequent use of indicator variables in regression model, categorization as a means of relaxing linearity is less efficient statistically, and may lead biased parameter estimates. Further, within-category nonlinear effects may be missed, and wide categories may be a source of residual confounding. On the other hand, RCS function facilitates flexible modeling of relationship between continuous variable and outcome variable, and diminishes amount of residual confounding for potential confounders.

6.18.3. Conclusion

The overall prevalence of HIV among pregnant women aged 15 to 24 years declined profoundly between 1994 and 2011. The decline in HIV prevalence was noted among pregnant women aged 15 to 24 years in urban areas and among pregnant women aged 15 to 24 years in rural areas. On the other hand, site-specific HIV prevalence estimates highlighted heterogeneous trend and burden of HIV prevalence, revealing possible upwards swings in HIV prevalence some sites in 2011. Although downward trends in HIV prevalence were observed in both urban and rural areas, HIV infection burden was lower in rural than urban sites. Given that the economic and health consequences are eminent the young and middle age groups, the historically low overall prevalence estimates observed in 2011 are encouraging, although observed spikes in HIV prevalence must be investigated.

Table 6.1. Characteristics by HIV serostatus of pregnant women aged 15 to 44 years who were recruited in the ANC-HIV-SS in Zambia, 1994-2011

	N	Combined (N=82086)		Seropositive (n=15505)		Seronegative (n=66581)		P-value∞
		n	%	n	%	n	%	
Survey calendar year	82561							
1994		9724	11.8	1981	12.8	7743	11.6	<0.001
1998		11718	14.3	2296	14.8	9422	14.2	
2002		12838	15.6	2559	16.5	10279	15.4	
2004		12404	15.1	2407	15.5	9997	15.0	
2006		13223	16.1	2348	15.1	10875	16.3	
2008		13298	16.2	2403	15.5	10895	16.4	
2011		8881	10.8	1511	9.7	7370	11.1	
Age [years]	82561							
Median		24		25		23		<0.001
IQR‡		20 to 29		22 to 29		20 to 29		
Missing		—		—		—		
Age groups [years]	82561							
15-19		17562	21.4	1954	12.6	15608	23.4	<0.001
20-24		27121	33.0	5117	33.0	22004	33.0	
25-29		18975	23.1	4626	29.8	14349	21.6	
30-34		11289	13.8	2569	16.6	8720	13.1	
35-39		5648	6.9	1032	6.7	4616	6.9	
40-44		1491	1.8	207	1.3	128481	1.9	
Residence	82561							
Rural		34686	42.3	3905	25.2	30781	46.2	<0.001
Urban		47400	57.7	11600	74.8	35800	53.8	
Educational attainment	72299							
Median		7		8		7		<0.001
IQR		5 to 9		7 to 9		5 to 9		
Missing††		10260	12.5	1700	11.0	8560	12.9	
Number of children	80918							
0 [Primagravida]		23208	28.9	3582	23.4	19626	30.1	<0.001
1		17779	22.1	3909	25.5	13870	21.3	
≥2		39456	49.0	7839	51.1	31617	48.6	
Missing		1643	2.0	175	1.1	1468	2.2	

∞ χ^2 test for differences in HIV ; ∞Wilcoxon sum rank test for differences in continuous values
‡IQR —Interquartile range; ∞Because of large sample size, p-value reported may or may not bear public health-relevant meanings; ††Missing educational attainment data include 8881 from 2011 not collected

Table 6.2 Distribution of pregnant women age 15 to 44 years surveyed during the ANC-based HIV sentinel surveillance in Zambia from 1994 through 2011 by selected characteristics

Characteristic	1994 (N=9760)		1998 (11907)		2002 (13051)		2004 (12404)		2006 (13260)		2008 (13298)		2011 (8881)	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Age [years]														
Median	23		23		23		24		24		25		25	
IQR	20 to 29		20 to 28		20 to 28		20 to 29		20 to 29		20 to 29		20 to 30	
Missing	—		—		—		—		—		—		—	
Age groups [years]														
15 — 19	2228	23	2975	25	3169	24	2397	19	2758	21	2394	18	1768	20
20 — 24	3314	34	4126	35	4376	34	4468	36	4312	33	4237	32	2450	28
25 — 29	1968	20	2539	21	2892	22	2834	23	3187	24	3412	26	2252	25
30 — 34	1363	14	1395	12	1604	12	1717	14	1822	14	2005	15	1440	16
35 — 39	697	7	689	6	792	6	749	6	945	7	1014	8	785	9
40 — 44	190	2	183	2	218	2	239	2	236	2	236	2	196	2
Missing	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Residence														
Rural	4238	43	4797	40	5539	42	5237	42	5684	43	5362	40	4043	46
urban	5522	57	7110	60	7212	58	7167	58	7576	57	7936	60	4838	54
Missing	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Marital status														
Divorced*	930	10	307	3	287	2	162	1	111	1	108	1	NC	NC
Married	8685	89	10351	88	10413	86	10529	85	11130	84	11185	84	NC	NC
Single	33	0	997	9	1280	11	1570	13	1943	15	1916	14	NC	NC
Widowed	53	1	64	1	88	1	62	1	6262	0	52	0	NC	NC
Missing	59		188		983		81		14		37		NC	NC
Education attainment [number of schooling years completed)														
Median	7		7		7		7		7		8		NC	NC
IQR	5 to 9		5 to 9		5 to 9		5 to 9		6 to 9		6 to 9		NC	NC
Missing	1016		103		95		51		236		78		8881	100
Educational attainment classification														
< 12	8166	93	10983	93	11750	91	10954	89	11005	84	10922	83	NC	NC
≥ 12	578	7	821	7	1206	9	1399	11	2019	16	2298	17	NC	NC
Missing	1016		103		95		51		236		78		8881	100

Characteristic	1994 (N=9760)		1998 (11907)		2002 (13051)		2004 (12404)		2006 (13260)		2008 (13298)		2011 (8881)	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Number of children														
0	2435	26	3527	38	3829	37	3637	37	4024	38	3576	34	2322	37
1	2016	21	2597	28	2937	28	2725	28	2929	27	2980	28	1709	27
≥2	5046	53	3190	34	3561	34	3465	35	3733	35	4003	38	2189	35
Missing														
HIV serostatus														
HIV negative	7743	80	9422	80	10279	80	9997	81	10875	82	10895	82	7370	83
HIV positive	1981	20	2296	20	2559	20	2407	19	2348	18	2403	18	1511	17
Missing	36	0	189		213		0		0		0		0	

*Divorced and separated; All the variables except “Partner age difference (p-value =0.754)” were statistically significantly different across survey years, p-value <0.0001. Because our large sample sizes across the years, caution must be exercised in interpretation of p-value. [Statistical significance NOT equivalent to public health significance]; NC—Data missing because not collected; Less than 0.5 rounded to zero and some cells may not add to 100 due rounding off; —No missing values

Table 6.3 Trends in HIV prevalence among pregnant women ages 15 to 24 years by ***urban*** sentinel site surveyed during the Zambia antennal attendees HIV sentinel surveillance in Zambia, 1994 through 2011

SITE	1994	1998	2002	2004	2006	2008	2011
	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)
Chelstone	268 25.0 (20.2-30.5)	471 22.9 (19.4-26.9)	447 19.9 (16.5-23.9)	411 17.8 (14.4-21.7)	460 15.0 (12.0-18.6)	377, 11.1 (8.3-14.7)	163 9.8 (6.1-15.3)
Chilenje	273 34.8 (29.4-40.6)	287 21.6 (17.2-26.7)	426 26.8 (22.8-31.2)	427 18.0 (14.7-22.0)	345 19.1 (15.3-23.6)	400 15.8 (12.5-19.6)	175 13.1 (8.9-18.9)
Chipata	261 27.6 (22.5-33.3)	287 23.7 (19.1-28.9)	283 21.9 (17.5-27.1)	288 19.4 (15.3-24.4)	290 19.0 (14.9-23.9)	268 14.6 (10.8-19.3)	203 16.3 (11.8-21.9)
Kabwe	275 28.4 (23.4-34.0)	304 24.0 (19.6-29.1)	280 22.1 (17.7-27.4)	263 23.6 (18.8-29.1)	277 16.2 (12.4-21.0)	254 24.4 (19.5-30.0)	187 16.0 (11.5-22.0)
Kalingalinga	280 20.0 (15.7-25.1)	284 22.9 (18.4-28.1)	338 20.1 (16.2-24.7)	403 19.9 (16.2-24.0)	427 19.7 (16.2-23.7)	404 15.1 (11.9-18.9)	306 17.6 (13.8-22.3)
Kasama	251 21.9 (17.2-27.4)	304 12.2 (9.0-16.3)	299 12.0 (8.8-16.2)	306 13.4 (10.0-17.7)	282 18.8 (14.7-23.8)	280 8.6 (5.8-12.4)	178 12.4 (8.3-18.0)
Livingstone	337 32.0 (27.3-37.2)	432 29.2 (25.1-33.6)	315 29.8 (25.1-35.1)	165 27.9 (21.6-35.2)	290 21.7 (17.4-26.8)	275 22.2 (17.7-27.5)	187 17.1 (12.4-23.2)
Mansa	268 23.5 (18.8-28.9)	342 20.5 (16.5-25.1)	299 21.4 (17.1-26.4)	249 24.5 (19.6-30.2)	284 15.1 (11.4-19.8)	236 16.9 (12.7-22.3)	169 13.0 (8.8-18.9)
Matero	248 28.2 (23.0-34.1)	293 22.5 (18.1-27.6)	487 21.6 (18.1-25.4)	482 26.3 (22.6-30.5)	423 25.3 (21.4-29.6)	376 19.9 (16.2-24.3)	181 11.6 (7.7-17.1)
Mongu	276 30.1 (25.0-35.7)	286 27.6 (22.8-33.1)	311 30.2 (25.4-35.5)	279 23.3 (18.7-28.6)	250 14.8 (10.9-19.7)	289 25.6 (20.9-30.9)	175 21.7 (16.2-28.4)
Ndola	288 27.1 (22.3-32.5)	613 25.9 (22.6-29.5)	578 21.6 (18.5-25.2)	450 22.2 (18.6-26.3)	394 18.0 (14.5-22.1)	475 15.2 (12.2-18.7)	236 14.8 (10.9-19.9)
Solwezi	120 20.8 (14.5-28.9)	295 16.6 (12.8-21.3)	310 11.9 (8.8-16.0)	302 12.6 (9.3-16.8)	289 15.2 (11.5-19.8)	297 11.4 (8.3-15.6)	175 10.3 (6.6-15.7)
Total	3145	4198	4373	4025	4011	3931	2335

95% Wilson confidence interval for HIV prevalence and *Number of pregnant women at each site [<25 years]

Table 6.4. Trends in HIV prevalence among pregnant women ages 15 to 24 years by ***rural*** sentinel site surveyed during the Zambia antennal attendees HIV sentinel surveillance in Zambia, 1994 through 2011

SITE	1994	1998	2002	2004	2006	2008	2011
	N* HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)
Ibenga	218 11.0 (7.5-15.9)	261 8.0 (5.3-12.0)	223 8.1 (5.2-12.4)	217 7.8 (4.9-12.2)	175 10.3 (6.6-15.7)	179 7.3 (4.3-12.0)	165 6.1 (3.3-10.8)
Isoka	274 11.3 (8.1-15.6)	313 9.9 (7.1-13.7)	296 7.1 (4.7-10.6)	269 11.5 (8.2-15.9)	254 3.1 (1.6-6.1)	273 5.1 (3.1-8.4)	174 8.0 (4.9-13.1)
Kabompo	159 1.9 (0.6-5.4)	163 9.8 (6.1-15.3)	219 5.9 (3.5-9.9)	302 7.9 (5.4-11.6)	336 5.4 (3.4-8.3)	295 8.8 (6.1-12.6)	176 1.7 (0.6-4.9)
Kalabo	149 8.1 (4.7-13.5)	225 12.0 (8.4-16.9)	249 14.5 (10.6-19.4)	274 14.2 (10.6-18.9)	300 9.7 (6.8-13.5)	297 12.1 (8.9-16.3)	177 8.5 (5.2-13.5)
Kapiri Mposhi	284 13.4 (9.9-17.8)	492 15.7 (12.7-19.1)	312 22.8 (18.5-27.7)	239 20.1 (15.5-25.6)	244 9.4 (6.4-13.7)	226 17.7 (13.3-23.2)	152 12.5 (8.2-18.7)
Kasaba	261 11.5 (8.2-15.9)	274 5.1 (3.1-8.4)	158 4.4 (2.2-8.9)	205 3.4 (1.7-6.9)	142 2.8 (1.1-7.0)	167 3.0 (1.3-6.8)	93 2.2 (0.6-7.5)
Luangwa	NC	NC	172 19.2 (14.0-25.7)	NC*	248 15.7 (11.7-20.8)	252 6.3 (3.9-10.1)	84 19.0 (12.1-28.7)
Macha	280 7.9 (5.2-11.6)	282 5.7 (3.5-9.0)	268 6.3 (4.0-9.9)	271 5.9 (3.7-9.4)	264 3.4 (1.8-6.4)	173 1.2 (0.3-4.1)	187 7.0 (4.1-11.5)
Minga	287 8.0 (5.4-11.7)	288 9.7 (6.8-13.7)	298 7.7 (5.2-11.3)	280 8.9 (6.1-12.8)	283 6.0 (3.8-9.4)	202 3.5 (1.7-7.0)	170 7.1 (4.1-11.9)
Mukinge	205 9.8 (6.4-14.6)	205 6.8 (4.1-11.1)	281 4.3 (2.5-7.3)	250 5.2 (3.1-8.7)	269 5.9 (3.7-9.4)	143 2.1 (0.7-6.0)	163 7.4 (4.3-12.4)
Nchelenge	262 13.7 (10.1-18.4)	295 13.2 (9.8-17.6)	283 18.4 (14.3-23.3)	274 15.3 (11.5-20.1)	274 9.9 (6.9-14.0)	257 14.8 (11.0-19.6)	158 7.6 (4.4-12.8)
Serenje	NC	NC	281 10.3 (7.3-14.4)	259 13.5 (9.9-18.2)	246 11.0 (7.7-15.5)	236 12.7 (9.1-17.6)	174 6.3 (3.6-11.0)
Total	2379	2798	3040	2840	3035	2700	1873

95% Wilson confidence interval for HIV prevalence
*Number of pregnant women at each site [<25 years]
NC— Data not collected [Luangwa and Serenje were introduced as sites in 2002] and NC* excluded from the ANC-HIV-SS data set [unreliable]

Table 6.5. Trends in HIV prevalence among pregnant women ages 25 to 44 years by urban sentinel site surveyed during the Zambia antennal attendees HIV sentinel surveillance in Zambia, 1994 through 2011

SITE	1994	1998	2002	2004	2006	2008	2011
	N* HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)
Chelstone	182 23.1 (17.6-29.7)	308 30.8 (25.9-36.2)	339 30.1 (25.5-35.2)	330 27.0 (22.5-32.0)	532 28.2 (24.5-32.2)	420 29.5 (25.4-34.1)	180 20.0 (14.8-26.4)
Chilenje	181 35.9 (29.3-43.1)	212 34.4 (28.4-41.1)	362 35.1 (30.3-40.1)	365 30.1 (25.7-35.0)	353 26.1 (21.8-30.9)	494 31.2 (27.2-35.4)	193 19.7 (14.7-25.9)
Chipata	227 33.9 (28.1-40.3)	198 32.3 (26.2-39.1)	166 36.1 (29.2-43.7)	210 34.8 (28.6-41.4)	225 31.1 (25.4-37.4)	252 28.6 (23.3-34.4)	167 30.5 (24.1-37.9)
Kabwe	214 31.3 (25.5-37.8)	197 31.0 (24.9-37.7)	217 37.3 (31.2-43.9)	235 31.5 (25.9-37.7)	221 35.3 (29.3-41.8)	279 32.3 (27.0-38.0)	167 29.9 (23.5-37.3)
Kalingalinga	224 24.1 (19.0-30.1)	203 32.0 (26.0-38.7)	248 35.1 (29.4-41.2)	287 37.3 (31.9-43.0)	369 36.9 (32.1-41.9)	393 38.9 (34.2-43.8)	396 37.4 (32.8-42.2)
Kasama	217 26.3 (20.9-32.5)	230 17.4 (13.0-22.8)	218 27.1 (21.6-33.3)	191 18.8 (13.9-25.0)	213 20.7 (15.8-26.6)	257 13.2 (9.6-17.9)	210 18.6 (13.9-24.4)
Livingstone	254 32.3 (26.8-38.3)	248 33.5 (27.9-39.6)	204 34.3 (28.1-41.1)	132 37.9 (30.1-46.4)	226 33.2 (27.4-39.6)	254 36.2 (30.6-42.3)	173 41.6 (34.5-49.1)
Mansa	191 23.6 (18.1-30.1)	243 22.2 (17.5-27.9)	196 23.5 (18.1-29.9)	251 31.1 (25.7-37.1)	214 21.5 (16.5-27.5)	263 19.8 (15.4-25.0)	190 25.8 (20.1-32.4)
Matero	136 28.7 (21.7-36.8)	196 37.8 (31.3-44.7)	292 36.6 (31.3-42.3)	317 34.7 (29.7-40.1)	367 38.1 (33.3-43.2)	422 28.4 (24.3-32.9)	178 25.3 (19.5-32.1)
Mongu	199 26.1 (20.5-32.6)	203 27.6 (21.9-34.1)	183 34.4 (27.9-41.6)	200 35.0 (28.7-41.8)	187 18.7 (13.8-24.9)	221 36.2 (30.1-42.7)	176 24.4 (18.7-31.3)
Ndola	211 28.4 (22.8-34.9)	394 30.5 (26.1-35.2)	416 24.3 (20.4-28.6)	411 29.9 (25.7-34.5)	403 26.8 (22.7-31.3)	514 27.6 (23.9-31.6)	260 30.0 (24.8-35.8)
Solwezi	109 26.6 (19.2-35.6)	193 25.9 (20.2-32.5)	185 16.2 (11.6-22.2)	213 20.2 (15.3-26.1)	226 20.4 (15.6-26.1)	236 20.3 (15.7-25.9)	213 25.4 (20.0-31.6)

95% Wilson confidence interval for HIV prevalence
*Number of pregnant women at each site [25 to 44 years]

Table 6.6. Trends in HIV prevalence among pregnant women ages 25 to 44 years by urban sentinel site surveyed during the Zambia antennal attendees HIV sentinel surveillance in Zambia, 1994 through 2011

SITE	1994	1998	2002	2004	2006	2008	2011
	N* HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)	N HIV% (95% CI)
Ibenga	134 11.9 (7.5-18.5)	130 13.8 (8.9-20.8)	169 11.8 (7.8-17.6)	156 16.0 (11.1-22.6)	155 12.3 (8.0-18.4)	202 17.3 (12.7-23.1)	196 19.4 (14.5-25.5)
Isoka	202 10.9 (7.3-15.9)	272 13.6 (10.0-18.2)	223 6.3 (3.8-10.3)	222 9.9 (6.6-14.5)	270 10.4 (7.3-14.6)	222 7.2 (4.5-11.4)	171 8.2 (4.9-13.3)
Kabompo	157 8.3 (4.9-13.7)	94 8.5 (4.4-15.9)	159 10.7 (6.8-16.5)	195 13.3 (9.3-18.8)	243 4.1 (2.3-7.4)	243 6.6 (4.1-10.4)	210 4.8 (2.6-8.5)
Kalabo	129 11.6 (7.2-18.3)	121 9.9 (5.8-16.5)	174 14.4 (9.9-20.3)	221 13.1 (9.3-18.2)	197 19.3 (14.4-25.4)	221 18.6 (14.0-24.2)	188 25.0 (19.4-31.6)
Kapiri Mposhi	210 12.9 (9.0-18.1)	285 15.8 (12.0-20.5)	213 20.7 (15.8-26.6)	256 20.3 (15.8-25.7)	247 23.5 (18.6-29.1)	279 26.2 (21.4-31.6)	209 27.8 (22.1-34.2)
Kasaba	217 13.8 (9.9-19.0)	212 5.7 (3.3-9.6)	147 8.2 (4.7-13.7)	177 9.0 (5.6-14.2)	121 3.3 (1.3-8.2)	154 8.4 (5.0-13.9)	163 3.1 (1.3-7.0)
Luangwa	NC	NC	170 27.1 (20.9-34.2)	NC*	242 14.5 (10.6-19.4)	233 13.3 (9.5-18.3)	76 26.3 (17.7-37.2)
Macha	214 10.7 (7.3-15.6)	208 9.6 (6.3-14.4)	250 9.2 (6.2-13.4)	246 9.8 (6.6-14.1)	232 10.8 (7.4-15.4)	178 1.7 (0.6-4.8)	181 10.5 (6.8-15.8)
Minga	196 12.2 (8.4-17.6)	188 9.0 (5.7-14.0)	225 13.3 (9.5-18.4)	218 14.7 (10.6-20.0)	229 12.7 (9.0-17.6)	166 9.6 (6.0-15.1)	189 14.3 (10.0-20.0)
Mukinge	164 9.1 (5.6-14.5)	193 10.9 (7.2-16.1)	215 13.0 (9.2-18.2)	247 12.1 (8.6-16.8)	232 10.8 (7.4-15.4)	183 8.2 (5.0-13.1)	191 10.5 (6.9-15.6)
Nchelenge	232 16.4 (12.2-21.7)	194 13.4 (9.3-18.9)	214 19.6 (14.9-25.5)	219 14.6 (10.5-19.9)	223 17.9 (13.5-23.5)	243 21.8 (17.1-27.4)	208 17.3 (12.8-23.0)
Serenje	NC	NC	240 17.9 (13.6-23.3)	240 13.8 (10.0-18.7)	250 18.0 (13.7-23.2)	338 15.7 (12.2-19.9)	188 16.5 (11.9-22.5)

95% Wilson confidence interval for HIV prevalence
*Number of pregnant women at each site [25 to 44]
NC— Data not collected [Luangwa and Serenje were introduced as sites in 2002]
NC*— Data excluded from the main ANC report because of problematic ANC-SS data file [unreliable]

Table 6.7 Odds ratio and 95% confidence interval for the relationship between pregnant woman's age and prevalent HIV infection based on the GLMM in which survey years was fitted using restricted cubic splines using data from the antenatal clinic attendees [age 15 to 24 years] collected in urban areas in Zambia between 1994 and 2011

Age [years]	1994 OR (95% CI)	1998 OR (95% CI)	2002 OR (95% CI)	2004 OR (95% CI)	2006 OR (95% CI)	2008 OR (95% CI)	2011 OR (95% CI)
15	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)
16	1.23 (1.23-1.23)	1.23 (1.22-1.23)	1.22 (1.22-1.22)	1.22 (1.22-1.22)	1.22 (1.22-1.22)	1.22 (1.22-1.22)	1.22 (1.22-1.22)
17	1.51 (1.51-1.51)	1.50 (1.50-1.51)	1.50 (1.49-1.50)	1.49 (1.49-1.50)	1.49 (1.49-1.49)	1.49 (1.48-1.49)	1.48 (1.48-1.49)
18	1.85 (1.85-1.86)	1.84 (1.84-1.85)	1.83 (1.83-1.84)	1.83 (1.82-1.83)	1.81 (1.80-1.81)	1.81 (1.81-1.82)	1.81 (1.80-1.81)
19	2.27 (2.26-2.28)	2.25 (2.25-2.26)	2.23 (2.23-2.24)	2.23 (2.22-2.23)	2.19 (2.19-2.20)	2.21 (2.20-2.21)	2.19 (2.19-2.20)
20	2.75 (2.74-2.76)	2.72 (2.71-2.73)	2.69 (2.68-2.71)	2.68 (2.66-2.69)	2.63 (2.62-2.64)	2.65 (2.64-2.66)	2.63 (2.62-2.64)
21	3.24 (3.21-3.26)	3.20 (3.17-3.22)	3.16 (3.13-3.18)	3.14 (3.11-3.16)	3.07 (3.05-3.10)	3.10 (3.07-3.12)	3.07 (3.05-3.10)
22	3.67 (3.62-3.71)	3.61 (3.57-3.66)	3.56 (3.52-3.61)	3.54 (3.49-3.58)	3.45 (3.40-3.49)	3.48 (3.44-3.53)	3.45 (3.40-3.49)
23	3.94 (3.86-4.02)	3.87 (3.79-3.95)	3.81 (3.73-3.89)	3.78 (3.70-3.85)	3.67 (3.59-3.74)	3.71 (3.64-3.79)	3.67 (3.59-3.74)
24	3.96 (3.83-4.09)	3.88 (3.76-4.01)	3.81 (3.69-3.93)	3.78 (3.66-3.90)	3.65 (3.54-3.77)	3.71 (3.59-3.82)	3.65 (3.54-3.77)

OR-Odds ratio
CI-95% confidence interval
Age value of 15 years was used as reference age value for computation of odds ratio and 95% confidence limits
Age was fit as a restricted cubic spline function with knots located at percentiles 10%, 50%, and 90% in GLMM with logit link function and random intercept for sentinel site to account for possible intra-site clustering

.

Table 6.8 Odds ratio (OR) and 95% confidence interval for the relationship between pregnant woman's age and prevalent HIV infection based on the GLMM in which survey years was fitted using restricted cubic splines using data from the antenatal clinic attendees [age 15 to 24 years] collected in rural areas in Zambia between 1994 and 2011

Age [years]	1994 OR (95% CI)	1998 OR (95% CI)	2002 OR (95% CI)	2004 OR (95% CI)	2006 OR (95% CI)	2008 OR (95% CI)	2011 OR (95% CI)
15	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)
16	1.12 (1.12-1.12)	1.13 (1.13-1.13)	1.14 (1.14-1.14)	1.14 (1.14-1.14)	1.14 (1.14-1.15)	1.15 (1.15-1.15)	1.15 (1.15-1.16)
17	1.26 (1.26-1.26)	1.28 (1.27-1.28)	1.29 (1.29-1.30)	1.30 (1.30-1.31)	1.31 (1.31-1.31)	1.32 (1.31-1.32)	1.33 (1.33-1.33)
18	1.42 (1.41-1.42)	1.44 (1.44-1.45)	1.47 (1.47-1.48)	1.49 (1.48-1.49)	1.54 (1.53-1.54)	1.51 (1.51-1.52)	1.54 (1.53-1.54)
19	1.59 (1.58-1.60)	1.63 (1.62-1.64)	1.67 (1.66-1.68)	1.69 (1.68-1.70)	1.77 (1.76-1.78)	1.74 (1.73-1.75)	1.77 (1.76-1.78)
20	1.78 (1.77-1.79)	1.84 (1.82-1.85)	1.90 (1.88-1.91)	1.93 (1.91-1.94)	2.04 (2.02-2.05)	1.99 (1.97-2.00)	2.04 (2.02-2.05)
21	1.98 (1.95-2.00)	2.05 (2.03-2.08)	2.13 (2.11-2.16)	2.17 (2.15-2.20)	2.32 (2.29-2.35)	2.26 (2.23-2.29)	2.32 (2.29-2.35)
22	2.18 (2.13-2.22)	2.28 (2.23-2.32)	2.38 (2.33-2.43)	2.43 (2.38-2.48)	2.63 (2.57-2.68)	2.54 (2.49-2.60)	2.63 (2.57-2.68)
23	2.37 (2.29-2.45)	2.49 (2.41-2.57)	2.62 (2.53-2.71)	2.68 (2.60-2.77)	2.93 (2.84-3.03)	2.82 (2.73-2.92)	2.93 (2.84-3.03)
24	2.53 (2.40-2.66)	2.68 (2.54-2.82)	2.83 (2.69-2.98)	2.92 (2.77-3.07)	3.22 (3.07-3.38)	3.09 (2.94-3.24)	3.22 (3.07-3.38)

OR-Odds ratio
CI-95% confidence interval
Age value of 15 years was used as reference age value for computation of odds ratio and 95% confidence limits
Age was fit as a restricted cubic spline function at percentiles 10%, 50%, and 90% in GLMM with logit link function and random intercept for sentinel site to account for possible intra-site clustering

Table 6.9 shows HIV prevalence by selected characteristics among pregnant women who sought antenatal care in urban areas during the survey period for the ANC-HIV-SS between from 1994 through 2011

Characteristics	1994	1998	2002	2004	2006	2008	2011
	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)
Age group							
15-24	3145 27.0 (25.5-28.6)	4198 22.9 (21.7-24.2)	4373 21.7 (20.5-23.0)	4025 20.5 (19.3-21.8)	4011 18.4 (17.2-19.6)	3931 16.5 (15.3-17.7)	2335 14.7 (13.4-16.2)
Educational attainment category							
0 to 4	313 18.8 (14.9-23.6)	512 20.7 (17.4-24.4)	515 17.1 (14.1-20.6)	402 18.4 (14.9-22.5)	307 16.3 (12.6-20.8)	249 15.3 (11.3-20.3)	NC
5 to 7	1391 24.9 (22.7-27.2)	1801 21.9 (20.1-23.9)	1690 20.5 (18.6-22.5)	1400 20.8 (18.7-23.0)	1174 17.6 (15.6-19.9)	1034 18.2 (16.0-20.6)	NC
8 to 9	860 29.8 (26.8-32.9)	1224 23.0 (20.8-25.5)	1300 24.7 (22.4-27.1)	1165 20.6 (18.4-23.0)	1182 20.2 (18.0-22.6)	1248 16.3 (14.3-18.4)	NC
10 to 11	185 38.9 (32.2-46.1)	252 30.6 (25.2-36.5)	320 20.0 (16.0-24.7)	355 20.8 (16.9-25.4)	390 22.1 (18.2-26.4)	429 16.8 (13.5-20.6)	NC
12 to 17	226 36.3 (30.3-42.7)	409 24.9 (21.0-29.4)	548 23.9 (20.5-27.6)	682 21.1 (18.2-24.3)	879 16.0 (13.8-18.6)	954 15.1 (13.0-17.5)	NC
Parity [Number of children birthed by pregnant woman]							
0	1286 24.8 (22.5-27.2)	2053 19.0 (17.4-20.8)	2148 20.6 (19.0-22.4)	2065 17.5 (15.9-19.2)	2199 15.4 (14.0-17.0)	1942 12.9 (11.5-14.5)	1354 13.6 (11.9-15.5)
1	965 31.1 (28.2-34.1)	1274 27.6 (25.2-30.1)	1393 23.3 (21.2-25.6)	1236 23.9 (21.6-26.3)	1189 22.1 (19.9-24.6)	1301 18.9 (16.9-21.1)	716 17.0 (14.5-20.0)
≥2	758 24.9 (22.0-28.1)	871 25.3 (22.5-28.2)	832 21.9 (19.2-24.8)	718 23.5 (20.6-26.8)	623 21.7 (18.6-25.1)	688 21.8 (18.9-25.0)	265 14.3 (10.6-19.1)

CI- Confidence interval

95% Confidence interval estimated by Wilson's method

[†]Number of school years completed: Categories reflect the school system in Zambia. The 10 to 11 years was included to reflect women who drop out due to pregnancy

Trends in HIV prevalence by survey year within educational attainment grouping within groups corresponding to the educational attainment groups based on the education system in Zambia for rural areas. Wide 95% confidence intervals for the estimated HIV prevalence for the ≥ 12 schooling years group is due to relatively smaller number of observations. Even though HIV prevalence is generally high, there was an overall distinctive decline in HIV prevalence from 32.9% in 1994 to 16.9% in 2008 for ≥ 12 schooling year's group. Little change in HIV prevalence in the 0-4 schooling years.

Table 6.10 shows HIV prevalence by selected characteristics among pregnant women who sought antenatal care in rural areas during the survey period for the ANC-HIV-SS between from 1994 through 2011

Characteristics	1994	1998	2002	2004	2006	2008	2011
	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)
Age group							
15-24	2379 10.0 (8.9-11.3)	2798 10.1 (9.1-11.3)	3040 10.9 (9.9-12.1)	2840 10.5 (9.4-11.6)	3035 7.7 (6.8-8.7)	2700 8.5 (7.5-9.6)	1873 7.4 (6.3-8.7)
Educational attainment category							
0 to 4	657 7.5 (5.7-9.7)	820 7.0 (5.4-8.9)	865 7.9 (6.2-9.8)	758 9.5 (7.6-11.8)	646 6.3 (4.7-8.5)	463 7.6 (5.5-10.3)	NC
5 to 7	949 9.5 (7.8-11.5)	1389 9.9 (8.4-11.5)	1411 10.6 (9.1-12.3)	1220 10.0 (8.4-11.8)	1225 7.3 (6.0-8.9)	1144 8.0 (6.6-9.8)	NC
8 to 9	310 16.8 (13.0-21.3)	474 13.5 (10.7-16.9)	589, 14.4 (11.8-17.5)	647, 12.4 (10.0-15.1)	752, 9.4 (7.6-11.7)	690 8.1 (6.3-10.4)	NC
10 to 11	32 21.9 (11.0-38.8)	50 22.0 (12.8-35.2)	93 11.8 (6.7-19.9)	131 10.7 (6.5-17.1)	208 5.3 (3.0-9.2)	177 11.3 (7.4-16.8)	NC
12 to 17	36 27.8 (15.8-44.0)	65 21.5 (13.3-33.0)	82 23.2 (15.4-33.4)	80 11.2 (6.0-20.0)	158 12.0 (7.8-18.0)	204 12.7 (8.8-18.0)	NC
Parity [Number of children birthed by pregnant woman]							
0	947 9.7 (8.0-11.8)	1215 8.7 (7.3-10.4)	1335 9.5 (8.1-11.2)	1266 9.5 (8.0-11.2)	1445 7.3 (6.0-8.7)	1254 8.3 (6.9-10.0)	635 8.3 (6.4-10.8)
1	715 11.7 (9.6-14.3)	844 12.7 (10.6-15.1)	930 11.7 (9.8-13.9)	839 11.6 (9.6-13.9)	864 7.9 (6.3-9.9)	791 8.5 (6.7-10.6)	369 9.5 (6.9-12.9)
≥2	633 8.8 (6.9-11.3)	739 9.5 (7.6-11.8)	775 12.4 (10.3-14.9)	733 10.8 (8.7-13.2)	723 8.6 (6.7-10.8)	655 9.0 (7.0-11.4)	261 9.6 (6.6-13.8)

CI- Confidence interval

95% Confidence interval estimated by Wilson's method

[†]Number of school years completed: Categories reflect the school system in Zambia. The 10 to 11 years was included to reflect women

Trends in HIV prevalence by survey year within educational attainment grouping within groups corresponding to the educational attainment groups based on the education system in Zambia for rural areas. Wide 95% confidence intervals for the estimated HIV prevalence for the ≥ 12 schooling years group is due to relatively smaller number of observations. Even though HIV prevalence is generally high, there was an overall distinctive decline in HIV prevalence from 32.9% in 1994 to 16.9% in 2008 for ≥ 12 schooling year's group. Little change in HIV prevalence in the 0-4 schooling years.

Table 6.11 displays the HIV prevalence estimates and 95% confidence interval trends by selected characteristic based on ANC-based HIV prevalence data conducted in urban and rural areas of Zambia between 1994 and 2011

Variable name	1994	1998	2002	2004	2006	2008	2011
	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)
15-24	3145 27.0 (25.5-28.6)	4198 22.9 (21.7-24.2)	4373 21.7 (20.5-23.0)	4025 20.5 (19.3-21.8)	4011 18.4 (17.2-19.6)	3931 16.5 (15.3-17.7)	2335 14.7 (13.4-16.2)
25-34	1887 31.5 (29.4-33.6)	2373 31.7 (29.8-33.6)	2551 32.6 (30.8-34.4)	2679 31.5 (29.8-33.3)	2951 29.7 (28.1-31.4)	3391 29.2 (27.7-30.7)	2080 27.1 (25.2-29.0)
35-44	458 16.4 (13.3-20.0)	452 18.4 (15.1-22.2)	475 21.5 (18.0-25.4)	463 25.7 (21.9-29.9)	585 24.4 (21.1-28.1)	614 27.9 (24.5-31.5)	423 33.1 (28.8-37.7)
Educational attainment [number of schooling years completed categorized according Zambia school system]							
0-4	2360 24.5 (22.8-26.3)	3024 24.0 (22.5-25.6)	2877 23.6 (22.1-25.2)	2658 23.9 (22.3-25.6)	2428 23.4 (21.8-25.2)	2379 23.1 (21.5-24.9)	Not collected (NC)
5-7	602 17.9 (15.1-21.2)	1008 21.1 (18.7-23.8)	924 19.5 (17.1-22.2)	789 20.3 (17.6-23.2)	680 19.9 (17.0-23.0)	631 20.3 (17.3-23.6)	NC
8-9	440 40.0 (35.5-44.6)	430 34.9 (30.5-39.5)	492 27.4 (23.7-31.5)	527 28.3 (24.6-32.3)	594 26.9 (23.5-30.6)	682 25.5 (22.4-28.9)	NC
10-11	1258 30.8 (28.4-33.4)	1780 27.0 (25.0-29.1)	2012 28.8 (26.8-30.8)	1927 26.8 (24.9-28.9)	2042 24.8 (23.0-26.7)	2323 23.1 (21.4-24.9)	NC
12-17	504 40.1 (35.9-44.4)	781 29.1 (26.0-32.3)	1094 28.4 (25.8-31.2)	1227 25.9 (23.5-28.4)	1652 21.3 (19.4-23.3)	1883 21.7 (19.9-23.6)	NC
Parity [Number of children birthed by pregnant woman							
0	1403 26.4 (24.2-28.8)	2206 20.1 (18.5-21.9)	2357 21.9 (20.3-23.6)	2310 19.0 (17.5-20.7)	2485 16.9 (15.4-18.4)	2248 15.2 (13.7-16.7)	1609 15.0 (13.3-16.8)
1	1187 33.2 (30.6-35.9)	1594 31.2 (29.0-33.5)	1795 26.9 (24.9-29.0)	1737 28.4 (26.3-30.5)	1806 26.8 (24.8-28.9)	1985 23.9 (22.1-25.9)	1204 22.0 (19.8-24.4)
≥2	2731 25.6 (24.0-27.3)	3223 26.6 (25.1-28.1)	3247 27.3 (25.8-28.8)	3112 27.4 (25.9-29.0)	3255 26.2 (24.8-27.8)	3703 26.8 (25.4-28.2)	2025 26.7 (24.8-28.7)
Marital status							

Variable name	1994	1998	2002	2004	2006	2008	2011
Married	4794 27.0 (25.8-28.3)	6027 25.6 (24.5-26.7)	6004 25.0 (23.9-26.1)	6020 25.6 (24.5-26.7)	6370 23.9 (22.8-24.9)	6703 23.0 (22.0-24.0)	NC
Single	638 30.7 (27.3-34.4)	220 33.6 (27.7-40.1)	207 33.8 (27.7-40.5)	91 39.6 (30.1-49.8)	73 53.4 (42.1-64.4)	69 53.6 (42.0-64.9)	NC
Divorced	20 70.0 (48.1-85.5)	690 23.8 (20.7-27.1)	799 26.0 (23.1-29.2)	996 20.0 (17.6-22.6)	1101 17.9 (15.7-20.3)	1141 19.7 (17.5-22.1)	NC

Table 6.12 displays the HIV prevalence estimates and 95% confidence interval trends by selected characteristic based on ANC-based HIV prevalence data conducted among pregnant women in urban and rural Zambia between 1994 and 2011

Variable name	1994	1998	2002	2004	2006	2008	2011
	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)
15-24	5524 19.7 (18.7-20.8)	6996 17.8 (16.9-18.7)	7413 17.3 (16.5-18.2)	6865 16.4 (15.5-17.3)	7046 13.8 (13.0-14.6)	6631 13.2 (12.4-14.1)	4208 11.5 (10.5-12.5)
25-34	3317 23.6 (22.2-25.1)	3862 24.3 (22.9-25.6)	4426 25.4 (24.2-26.7)	4551 24.5 (23.3-25.8)	4999 23.3 (22.2-24.5)	5417 23.5 (22.4-24.6)	3692 21.6 (20.3-22.9)
35-44	883 12.5 (10.4-14.8)	860 13.3 (11.2-15.7)	999 15.1 (13.0-17.5)	988 17.0 (14.8-19.5)	1178 17.9 (15.8-20.2)	1250 20.3 (18.2-22.6)	981 23.5 (21.0-26.3)
Educational attainment [number of schooling years completed]							
0-4	4052 18.8 (17.7-20.1)	5387 17.9 (16.9-18.9)	5455 17.9 (16.9-18.9)	5029 18.1 (17.1-19.2)	4855 16.6 (15.6-17.7)	4752 16.8 (15.8-17.9)	Not collected (NC)
5-7	544 36.4 (32.5-40.5)	549 31.9 (28.1-35.9)	648 25.6 (22.4-29.1)	708 24.7 (21.7-28.0)	872 21.2 (18.6-24.1)	930 21.7 (19.2-24.5)	NC
8-9	1706 27.4 (25.3-29.5)	2458 23.9 (22.2-25.6)	2938 25.4 (23.9-27.0)	2937 22.7 (21.3-24.3)	3210 20.4 (19.0-21.8)	3454 19.4 (18.1-20.7)	NC
10-11	1832 10.9 (9.6-12.4)	2412 13.2 (11.9-14.6)	2517 12.4 (11.1-13.7)	2280 13.2 (11.9-14.7)	2035 11.8 (10.5-13.3)	1786 13.5 (12.0-15.2)	NC
12-17	573 39.3 (35.4-43.3)	912 27.6 (24.8-30.6)	1280 28.3 (25.9-30.8)	1395 24.4 (22.3-26.8)	2009 20.9 (19.2-22.7)	2296 20.9 (19.2-22.6)	NC
Parity [Number of children birthed by pregnant woman]							

Variable name	1994	1998	2002	2004	2006	2008	2011
0	2430 19.8 (18.3-21.4)	3471 16.2 (15.0-17.5)	3763 17.6 (16.4-18.8)	3637 5.7 (14.6-16.9)	4009 13.4 (12.4-14.5)	3576 12.8 (11.8-13.9)	2322 13.4 (12.0-14.8)
1	5021 18.7 (17.7-19.8)	5691 19.4 (18.4-20.4)	6194 20.4 (19.4-21.4)	6032 20.0 (19.1-21.1)	6292 19.1 (18.2-20.1)	6742 20.2 (19.2-21.1)	3484 21.8 (20.4-23.2)
≥2	2010 24.7 (22.9-26.7)	2556 24.7 (23.1-26.4)	2881 22.0 (20.5-23.5)	2725 22.9 (21.3-24.5)	2918 20.8 (19.4-22.3)	2980 19.6 (18.2-21.1)	1709 19.5 (17.7-21.4)
Marital status							
Married	8653 19.8 (19.0-20.6)	10204 19.2 (18.5-20.0)	10251 19.8 (19.0-20.6)	10529 19.5 (18.8-20.3)	11106 18.2 (17.4-18.9)	11185 18.1 (17.4-18.8)	NC
Single	33 51.5 (35.2-67.5)	970 19.9 (17.5-22.5)	1248 21.2 (19.0-23.5)	1570 16.9 (15.1-18.8)	1930 14.0 (12.6-15.7)	1916 16.8 (15.1-18.5)	NC
Divorced	979 23.9 (21.3-26.7)	361, 29.9 (25.4-34.8)	365 28.5 (24.1-33.3)	224 29.9 (24.3-36.2)	173 34.1 (27.5-41.4)	160 34.4 (27.5-42.0)	NC

ble 6.13 displays the HIV prevalence estimates and 95% confidence interval trends by selected characteristic based on ANC-based HIV prevalence data conducted in rural areas in Zambia between 1994 and 2011

Variable name	1994	1998	2002	2004	2006	2008	2011
	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)	n HIV % (95% CI)
15-24	2379 10.0 (8.9-11.3)	2798 10.1 (9.1-11.3)	3040 10.9 (9.9-12.1)	2840 10.5 (9.4-11.6)	3035 7.7 (6.8-8.7)	2700 8.5 (7.5-9.6)	1873 7.4 (6.3-8.7)
25-34	1430 13.1 (11.5-15.0)	1489 12.4 (10.8-14.2)	1875 15.7 (14.2-17.5)	1872 14.5 (13.0-16.2)	2048 14.1 (12.6-15.6)	2026 13.9 (12.5-15.5)	1612 14.5 (12.9-16.3)
35-44	425 8.2 (6.0-11.2)	408 7.6 (5.4-10.6)	524, 9.4 (7.1-12.1)	525 9.3 (7.1-12.1)	593 11.5 (9.1-14.3)	636 13.1 (10.7-15.9)	558 16.3 (13.5-19.6)
Educational attainment [number of schooling years completed categorized according Zambia school system]							
0-4	1692 10.9 (9.5-12.5)	2363 10.0 (8.9-11.3)	2578 11.5 (10.3-12.8)	2371 11.6 (10.4-12.9)	2427 9.8 (8.7-11.1)	2373 10.5 (9.4-11.8)	Not collected (NC)
5-7	104 21.2 (14.4-30.0)	119 21.0 (14.7-29.2)	156 19.9 (14.4-26.8)	181 14.4 (10.0-20.2)	278 9.0 (6.2-12.9)	248 11.3 (7.9-15.8)	NC
8-9	448	678	926	1010	1168	1131	NC

Variable name	1994	1998	2002	2004	2006	2008	2011
	17.6 (14.4-21.4)	15.6 (13.1-18.6)	18.0 (15.7-20.6)	15.0 (12.9-17.3)	12.6 (10.8-14.6)	11.7 (9.9-13.7)	
10-11	1230 7.5 (6.1-9.1)	1404 7.5 (6.3-9.1)	1593 8.2 (7.0-9.7)	1491 9.5 (8.1-11.1)	1355 7.7 (6.4-9.3)	1155 9.9 (8.3-11.7)	NC
12-17	69 33.3 (23.4-45.1)	131 19.1 (13.3-26.7)	186 27.4 (21.5-34.2)	168 13.7 (9.3-19.7)	357 19.0 (15.3-23.4)	413 16.9 (13.6-20.9)	NC
Parity (i.e., number of children birthed by pregnant woman)							
0	1027 10.7 (9.0-12.8)	1265 9.4 (7.9-11.1)	1406 10.3 (8.8-12.0)	1327 9.9 (8.5-11.7)	1524 7.7 (6.5-9.2)	1328 8.8 (7.4-10.5)	713 9.7 (7.7-12.1)
1	2290 10.5 (9.3-11.8)	2468 10.0 (8.8-11.2)	2947 12.9 (11.7-14.2)	2920 12.2 (11.0-13.4)	3037 11.5 (10.4-12.7)	3039 12.1 (11.0-13.3)	1459 14.9 (13.1-16.8)
≥2	823 12.5 (10.4-15.0)	962 13.9 (11.9-16.3)	1086 13.9 (12.0-16.1)	988 13.2 (11.2-15.4)	1112 11.1 (9.4-13.0)	995 11.1 (9.3-13.2)	505 13.5 (10.8-16.7)
Marital status							
Married	3859 10.8 (9.9-11.8)	4177 10.0 (9.2-11.0)	4247 12.5 (11.5-13.5)	4509 11.4 (10.5-12.4)	4736 10.5 (9.6-11.4)	4482 10.7 (9.8-11.6)	NC
Single	13 23.1 (8.2-50.3)	280 10.4 (7.3-14.5)	449 12.5 (9.7-15.9)	574 11.5 (9.1-14.4)	829 8.9 (7.2-11.1)	775 12.4 (10.3-14.9)	NC
Divorced	341 11.1 (8.2-14.9)	141 24.1 (17.8-31.8)	158 21.5 (15.8-28.6)	133 23.3 (16.9-31.2)	100 20.0 (13.3-28.9)	91 19.8 (12.9-29.1)	NC

CHAPTER 7

AGE, PERIOD, AND COHORT EFFECTS ON HIV PREVALENCE AMONG PREGNANT WOMEN IN ZAMBIA, 1994 THROUGH 2011

7.1. Background

Sub-Saharan Africa (SSA) comprises 48 of 54 countries in Africa, and is home to approximately 900 million people, 13% of the global human population but in 2011 accounted for 69% of the 34 million people living with the human immunodeficiency virus (HIV), the virus that causes acquired immunodeficiency virus (AIDS).[10, 11] As in prior years, compared to other regions globally, the highest number of people newly infected with HIV in 2011 were in SSA, accounting for 1.8 million out of the estimated 2.5 million people who had new HIV infection worldwide.[10]

Despite the seriousness and extent of HIV epidemic in SSA, considerable success in prevention of spread of HIV infections has been achieved, especially in recent years. For example, 13 countries of the 25 countries worldwide that recorded more than a 50% reduction in HIV prevalence are located in SSA, with historically high HIV burden. [10] Compared to 2001, the estimated numbers of people newly infected with HIV in 2011 were fewer by 700,000, suggesting a net decline in global HIV incidence.[10]

The number of people living with HIV increased by 3 million in 2011 reaching 34 million, compared to 31 million people living with HIV in 2005.[23] The marked geographic variation in the burden of HIV infections suggests the synergistic influence of social, biologic, environmental, community, and cultural factors on the HIV burden, and further emphasizes how HIV epidemic is driven by region-unique factors.[10]

7.2. Increasing population of HIV survivors due to cART and high rate of new HIV infections in SSA contributed to high HIV burden in 2011

The differential burden in HIV infections between and within countries globally may reflect the differential distribution of risk factors for HIV infection in different communities.[24] HIV-related mortality depletes community HIV prevalence, but access to effective cART improves survival of HIV-infected persons. HIV burden is a function of new HIV infections, immigration and outmigration of HIV-infected persons into a specific geographic area, and improved survival of HIV-infected persons due to cART. Low rate of new HIV infection dwarfs the HIV prevalence as does HIV-related mortality. Further, "treatment as prevention concept", enunciates that effective chemotherapy (cART) among adults can lower HIV transmission rates in a similar manner that chemotherapeutic interventions limit HIV transmission from mother to child, thus may retard the expansion of the HIV epidemic (if behavioral risk factors remain constant).[29]

HIV preventive and treatment interventions implemented over the years as well as maturation of the HIV epidemic are likely to have influenced the observed decline in HIV incidence in recent years.[29, 305, 306] Specifically, as HIV epidemic mature and become more widespread, more people are awakened to the risk factors for HIV infection, and may adopt less risky sexual behavior. Overtime, corresponding reduction in the number of new HIV infections might unfold as more and more people shun risk sexual behavior such as unprotected sexual intercourse (e.g., increased correct and consistent condom).[307-309] Also, there can be saturation phenomena, such as the high prevalence seen in highest risk persons, with early high death rates; prevalence can decline merely as a function of these saturation dynamics.[31, 32, 306]

Innumerous HIV preventive and treatment interventions designed to curb the spread of HIV infections rests on information from prior research undertakings that have identified factors associated with the heightened spread of HIV infections. Risk factors for HIV infection are complexly interrelated across individual-level factor and community-level factors. [25] Highlighting the interrelatedness of risk factors for HIV infections, Vermund and Hayes (2013) emphasizes that dampening the HIV epidemic require concerted and multipronged interventions.[29] However, to design interventions that are appropriate to local environment require adequate understanding of the HIV epidemic dynamics in affected communities, and hence the importance of monitoring trends in HIV incidence and prevalence.[20, 29, 33] The availability, and implementation of preventive and treatment interventions have not been even across period and regions. Therefore, differential capture of preventive interventions and knowledge may obtain across birth cohort, sociodemographic and geographic areas.[310] There is limited data on age, period and birth cohort influence on HIV prevalence.

7.3. Zambia has a generalized HIV epidemic

Zambia has a generalized HIV epidemic (W.H.O's definition of HIV prevalence > 1% in the general population). Zambia's first official report of an AIDS case was in 1984, and HIV/AIDS has since emerged as a prominent public health problem. By 1994, the estimated HIV prevalence among pregnant women in Livingstone, an urban area setting, spiked to 32%. Based on the 2012 UNAIDS report of the global HIV epidemic, the estimated number of people living with HIV in Zambia in 2011 could range from 900,000 and 1,100,000 out the estimated 13.2 million people, an HIV burden nearly as

great as the estimated 1.2 million people living with HIV infection in the United States of America, a country with 24 times the number of persons (314 million in 2012).[10, 34, 35] Direct measurement of HIV incidence rate is a hampered by logistical, technical, and financial challenges, and thus HIV incidence, although a preferable measure of progression of HIV epidemic, is less widely reported.[10, 24, 173, 176-179]

7.4. Population-based surveys: gold standard for estimating HIV prevalence

The chief data sources for HIV prevalence estimates in Zambia are antenatal care based HIV sentinel surveillance (ANC-HIV-SS) and population-based demographic and health surveys (PBS-DHS). PBS generated HIV prevalence estimates are regarded "gold standard" provided the surveys are not threatened by low participation rate or affected by methodological constraints (e.g., incomplete sampling frame).[36, 171] Further, PBS-based HIV prevalence estimates are considered more trustworthy because they are based on data obtained from surveys that use statistically robust methods of sampling participants.[36, 155, 171, 172] However, selection bias may threaten the external validity of the DHS-based HIV prevalence estimates if those who agreed and those who did not agree to participate in the survey are different, or if key respondents are unavailable at the time of the survey (men migrating for work, for example).[36, 171, 172]

7.5. Two PBS-DHS HIV prevalence data points may be inadequate for monitoring HIV prevalence trends

Trends in HIV prevalence are sufficiently described when multiple time points of HIV prevalence data are exist. As of June 2013, there were only two time points with

HIV prevalence estimates based on the DHS in 2001 and 2007. Undoubtedly, DHS-HIV prevalence estimates in 2001 and 2007 DHS have provided useful information on the HIV burden, but HIV prevalence trends assessment using two time points, while informative, may not inadequate. Most countries in SSA countries have corroborated DHS-based HIV prevalence trend analysis with evidence from ANC-HIV-SS based HIV prevalence trend analysis.[24, 36, 40] For example, the Zambian ANC-HIV-SS has seven time points of HIV prevalence estimates in at least 22 sites between 1994 and 1998 and 24 sites between 2002 and 2011 (i.e., 1994, 1998, 2002, 2004, 2006, 2008, and 2011.[182]

7.6. Both ANC-HIV-SS and PBS-DHS based HIV prevalence estimates provide critical information for monitoring HIV prevalence trends

The ANC-HIV-SS in Zambia was first launched in 1990 for monitoring HIV prevalence trends among pregnant women, and since 2001 has complemented DHS-based HIV prevalence monitoring efforts. HIV prevalence estimates based on the ANC-HIV-SS, notwithstanding biases inherent in ANC-based HIV surveillance methods (e.g., potential for selection bias and exclusion of non-pregnant women and men) has been until 2001, the leading provider of key HIV prevalence estimates for understanding and assessing magnitude of the HIV epidemic in Zambia.[24, 41, 157]

7.7. Age, period and birth cohort effects may influence HIV prevalence

Thoughtful assessment of age, period, and birth cohort on HIV prevalence may help identify exposures that are driving the HIV epidemic in the population, and may guide prevention and treatment programs. Few studies have conducted focused

investigation of the simultaneous influence of age, period and birth cohort effects on HIV prevalence. Prior research efforts have focused on assessing trends in HIV prevalence by age group and by survey period, and independent effect of age, period and cohort effects on the reported decline and/or stabilization of HIV prevalence in the younger generation has not been investigated. To enhance a better understanding of the growth and evolution of the HIV epidemic that may inform future prevention and treatment interventions, assessment of age, period, and birth-cohort effects is critical.[66-68, 74, 311] .

7.8. Key variables for investigation of classical age, period and birth cohort analysis

7.8.1. Age

Age, as a measure of time since birth, is related to innumerous health and economic outcomes, and consequently used as study factor in many epidemiologic and demographic studies.[62] Briefly, person's age may reflect physiologic changes overtime; accumulation of social and cultural experience; and change in social status.[63, 74] With regard to HIV risk, age-related social, cultural and peer pressure norms may influence sexual behavior of men and women.[312]

7.8.2. Period

Period effect represented by calendar year of the survey, refers to effects specific to respective calendar years (e.g., events in the community or interventions introduced specific period), and may influence all age groups simultaneously.[312] Period effects might be attributed to cultural, economic, environmental and social influences (e.g., civil

war disturbances; disease or epidemic outbreaks; and public health preventive interventions.[74, 312]

With respect to HIV risk, many interventions might qualify as sources of period effects (e.g., with increasing availability of information about HIV/AIDS in Zambia, persons across all age values may become familiar about HIV as a causal agent for AIDS including preventive strategies against HIV spread).[310] Factors that may qualify as a period effects included preventive program such as expanded condom use for prevention of HIV transmission, implementation of new scientific innovations such as cART leading to improved survival of HIV-infected persons and lower transmissibility of HIV from adherent persons, and policy changes by a government such as increased educational opportunities.[310]

7.8.3. Birth cohort

Birth cohort effects arise when certain period related factors exert their influence differentially across persons with different age values (i.e., persons born in the same years are similarly influenced). Persons born in the same year or period (i.e., same birth cohort) are likely to have experienced similar cultural, economic and social factors, therefore conceivable that persons in the same birth cohort might have similar behaviors, and because of having been in similar contextual settings, might have been exposed to similar risk factors for HIV infection.

Within the framework of risk for HIV infection, the obtaining contextual setting (i.e., period and birth cohort), may shape the sexual behaviors and attitudes (e.g. attitude toward condom use or frequency of multiple sexual partners), consequently influencing

the risk of acquiring HIV infection. On one hand, younger persons who have grown up in era saturated with innumerous HIV preventive information may have greater self-efficacy in adopting safer sexual behavior practices as their sexual life buds.[310] On the other hand, older people, might be more restrained to adapt newer sets of beliefs and behaviors especially if the beliefs and behaviors are relates to sexuality. The above exemplify cohort effects because the individual's beliefs and behaviors are invariant, but different birth cohorts' breeds lines of sexual and cultural behavior that less risk. Expressed differently, cohort effects are present when period-related factors exert their effects differentially across different age values, and may be regarded as age-period statistical multiplicative interaction effects [i.e., cross-product term between age and survey calendar year].[313-316]

7.8.4. Several alternatives methods for investigating age, period and birth-cohort effects

Several statistical methods have been proposed for investigating age, period, and cohort effects for various outcomes, each with specific assumptions that enable estimation of unique regression coefficients.[317] [68, 74, 318, 319] Using any of the age-period-cohort regression methods in an investigation of age, period and birth-cohort effects on a specific condition require closer examination of the specific assumptions. However, certain assumptions may not defensible some settings. [60, 61, 320]

7.8.5. Identification problem

Nearly all regression-based methods for investigating age, period and birth cohort effects introduce constraints to overcome. In particular, what is popularly known as the

identification problem is the failure of a regression model to yield unique parameter values because the variables entered in the regression are linearly dependent.[315] Identification problems hamper estimation of unique parameters for age, period and birth-cohort effects from a fitted regression model in which the three intrinsically time-related variables are modeled simultaneously. [68, 69, 314, 321-324] As would be expected, estimated parameters may vary depending on constraints that were used.[317, 325-328]

Parameter estimates from a regression model may not be unique (i.e., identification problem) if the modeled variables are linearly dependent, and additivity of effects is assumed. The variables age, period and birth cohort are linearly dependent because birth-cohort of a person can be approximated with accuracy if the investigator knows the person's age and survey year. For example, a pregnant woman aged 21 years recruited in the 2011 survey round would principally belong to the 1990-1996 birth-cohorts.

7.9. Classified random intercept effect model (CCREM) was used to estimate age, period and birth cohort effects on HIV prevalence

To avoid identification problem in the current study analyses, I used the cross-classified random effect model for investigating age, period and birth cohort (CCREM-based APC) proposed by Yang and Land (2006) was used to estimate age, period and birth-cohort effects on HIV prevalence among pregnant women in Zambia. Because the outcome was dichotomous (i.e., HIV seropositive and HIV seronegative), generalized linear mixed model (GLMM) form of the CCREM-based APC method was applied to assess age, period and cohort effects on HIV prevalence using repeated cross-sectional data.[66-68, 74] To use the CCREM-based APC method, period and birth-cohort are considered as proxies for contextual settings factors (e.g., social, social, historical

environments or unmeasured influences specific to period and birth cohorts) in which persons are embedded. [66, 68, 74] Possible intra-group clustering by period and birth-cohort were accounted for by modeling period and birth-cohort as random effects in GLMM.[68]

7.9.1. Overview of the CCREM-based APC method

Armed with the theory that the identification problem arises when effects of linearly dependent variables (i.e., age, period, and birth cohort) are modeled under the assumption of additive effects in standard logistic regression, Yang and Land (2006) contrived the CCREM-based APC method that facilitated non-additive modeling of age, period and birth cohort effects.[60, 329] With the understanding that additivity of effects assumption in modeling age, period and birth-cohort leads to the identification problem, Yang and Land (2006) suggested that age could be fit in a nonlinear (i.e., a quadratic function) model, and period and birth cohort variables are fit as random effects.[68, 74]

7.9.2. CCREM-based APC analysis accounts for the cross-classification of respondents within period and birth cohorts

CCREM-based APC method has some attractive properties. First, a CCREM-based APC approach accounts for the inherent cross-classification of individual-level observations by period and birth cohorts. Second, the identification problem is overcome when age, period, and birth-cohort variables are fit in a non-additive manner (i.e., age is fit nonlinearly using restricted cubic splines, and period groups and birth cohort groups as random group effects).

Third, CCREM-based APC analysis accommodates additional explanatory variables (e.g. educational attainment) beyond the three intrinsically related variables (i.e., age period and birth cohort).[68, 74] Despite the attractive features of CCREM-based APC modeling, Yang and Land (2006) have stressed that the approach is not a definitive panacea but merely avoid the identification problem in age, period and birth cohort effects estimation. [66-68, 74]

7.9.3. Within-period and within-cohort clustering is captured by modeling period and birth cohort as random effects

The elegance of the CCREM-based APC method lies in its ability to account for within-cohort and within-period clustering via modeling of period and birth cohort groups as random components, cross-classification, and also in its ability to quantify the variation in the regression intercept across the different levels of the grouping variable (i.e., period and birth cohort) using a probability distribution. Fixed effect model estimates a fixed effect coefficient for each grouping category (i.e., one for the period group and another for birth cohort group).[66-68, 265].

7.9.4. Person-level covariates are regarded as fixed (i.e., level-1) variables and period and Birth cohort variables as random components (i.e., level-2)

Person-level information (e.g., age and educational attainment) was considered fixed effect level-1 variables, whereas period groups and birth cohort groups regarded as random components level-2 variables yielding a multilevel data structure that was cross-classified. Expressed differently, in repeated cross-sectional data, participant observations are cross-classified by period and birth cohort that are regarded as higher

level contexts. [68] Thus, the data structure described above may be viewed as a multilevel data structure. Period and birth-cohort group levels represented contextual settings factors.[66, 68, 74, 330] Although period and birth cohort represented as level-2 variables, the resulting cross-classified data structure does not constitute a cleanly nested hierarchy, but requires use of statistical methods that help account for cross-classification.[74] Disregarding the multilevel and cross-classified data structure would detract from the substantive statistical utility of an APC analyses based on repeated cross-sectional data.[66, 68, 74]

Review of literature shows that the majority of classical age, period and birth-cohort analyses are based on longitudinal study designs: people in the same birth cohort are tracked over a long period of time.[328] However Yang and Land (2006) suggested that age and survey year from repeated cross-sectional surveys may be used to create "synthetic birth cohorts" groups, i.e., proxies for real birth cohorts.[66, 68, 74]

The "synthetic birth cohorts" (i.e., created from person's age and survey years in repeated cross-sectional data) thus created can be applied to conduct CCREM-based APC analyses while acknowledging the following. Tracking of individual persons in a synthetic birth cohort is impossible as would be in real birth cohort, i.e., a longitudinal study because synthetic birth cohort are created from cross-sectional data.[74] However, synthetic birth cohorts are regarded as proxy for real birth cohorts. Therefore synthetic birth cohort from repeated cross-sectional data can be traced and used as proxies for cohort effects.[312, 317, 328]

7.9.5. Knowledge gap

Data on the age, period, and birth cohort effects are limited because very few studies have examined simultaneously the unique influence of age, period, and cohort effects on HIV prevalence.[66, 68, 72-75] Notably, Houweling et al (1999) examined age, period and cohort effects on HIV incidence trends among drug users in France, and highlighted that age-period-cohort (APC) analyses may disentangle age, period and cohort effects, and provide enhanced picture of the growth and direction of the HIV epidemic.[72, 76, 77] Further, applications of age-period-cohort analysis in HIV research were reported by Rosinska et al (2011) using surveillance data in Poland.[75] As more data become available via repeated cross-sectional surveys, opportunities for investigating age, period, and birth cohort effects using CCREM-based APC method have arisen providing addition epidemiologic armamentarium for assessing trends in HIV prevalence.

7.10. Specific aim

The current study was designed to examine the age, period and birth-cohort effects on HIV prevalence using data collected from pregnant women from ANC-based HIV surveillance in Zambia between 1994 and 2011.

7.11. Methods

7.11.1. Study design and study population

Secondary analysis of repeated cross-sectional survey data collected from 82,086 pregnant women aged 15 to 24 years who participated in the ANC-HIV-SS [i.e., 1994,

1998, 2002, 2004, 2006, 2008, and 2011]. Details of the design and data collection methods of the ANC-HIV-SS program have been described previously, and in Chapter 4.[24, 41, 157, 182] Briefly, ANC-HIV-SS is a series of surveys done every 2 to 4 years, focused on estimating and monitoring HIV prevalence trends among pregnant women seeking antenatal care in Zambia. ANC-HIV-SS were conducted in 1994, 1998, 2002, 2004, 2006, 2008, and 2011. The number of sites varied only slightly over time: 22 sites from 1994 to 2002 and 24 sentinel sites from 2004 to 2011.

7.11.2. Inclusion criteria

Pregnant women who sought antenatal care at health centers designated as sentinel sites for ANC-HIV-SS during the four-month survey period in specified survey years were recruited for the purpose of estimating HIV prevalence.

7.11.3. Target sample size

Each site targeted 500 pregnant women, except for Ndola and Lusaka sites where around 800 pregnant women were recruited per site. However, in 2011, due to change in survey protocol the target number of pregnant women to be recruited per site was set at 360.

7.11.4. Sociodemographic variable collected via questionnaire [i.e. 1994 to 2008]

Study nurses trained on the survey protocol identified and recruited eligible pregnant women in a chronologically consecutive manner. The study nurse interviewed each eligible pregnant woman, on her first antenatal clinic visit for the current pregnancy,

to collect sociodemographic data (e.g., age, education, residence) using a standard questionnaire (i.e., 1994 through 2008). A revised protocol used in 2011 mandated abstraction of information from pregnant woman's routine antenatal medical record card, consequently limiting data collection only to those variables that were abstracted from pregnant woman's routine medical record card (i.e., age, parity (i.e., number of children birth by pregnant woman).

7.11.5. HIV serostatus were serologically confirmed using commercial HIV test kits

HIV serostatus of serum/plasma specimen prepared from blood provided by each pregnant woman was determined using pre-specified survey-specific HIV screening algorithms (i.e., screening and confirmatory testing) as explained in prior studies and in Chapter 4. [24, 157, 183] Briefly, part of the blood collected routinely on the first antenatal care visit from each pregnant woman for routine syphilis screening was portioned in a de-identified container, and assigned a distinctive identity number (ID) for survey reporting. Site-screening using rapid HIV tests, reference-laboratory confirmatory testing using HIV ELISA, and tie-breaking testing, where HIV test result from site-screening and reference laboratory confirmatory testing were different (e.g., Western Blot confirmation or Bionor HIV-1/2).[24, 183]

HIV test kits used for HIV testing across the seven survey years were dissimilar HIV test kits, but the HIV testing algorithm used was nearly consistent across survey years. Even though the impact of using HIV test kits manufactured by different companies may impact the HIV prevalence trends, the fact that HIV test kits with high sensitivities and specificities (>99%) were used provide adequate confidence that the

different HIV test kits used across the years may not materially alter the estimated HIV prevalence trends. Details of the HIV assays used have been reported in prior publications and in Chapter 4.

7.12. Data management

Survey round specific data sets were cleaned and checked for consistency, and merged as explained in Chapter 4 using R version 3.0, a freely available statistical and computing software.[188] [331] The merged data set consisted 82,086 records of pregnant women aged 15 to 44 from the ANC-HIV-SS rounds conducted between 1994 and 2011 (i.e., 1994, 1998, 2002, 2004, 2006, 2008 and 2011). Variables relevant to the research question, and collected in all the seven survey rounds (e.g., age, educational attainment, survey year, residence site location [i.e., urban or rural], and parity) and consistently coded were the focus of merging of seven data sets. Also focused on was educational attainment variable, although data on educational attainment were not collected in 2011, was included in the merged data. The number of pregnant women recruited per survey between 1994 and 2011 ranged from a low of 8881 in 2011 to 13298 in 2008.

7.13. Fixed variables (level-1) used in the CCREM-based APC analyses

The variables used in the CCREM-based APC analyses were pre-specified guided by subject matter literature and availability of variables in data sets from ANC-HIV-SS rounds.

7.13.1. Age

Pregnant woman's age measured as complete years lived at the time the time of interview.

7.13.2. Educational attainment

Pregnant women's educational attainment measured as number of years of schooling completed. Pregnant women were asked the following question to capture information on educational attainment [i.e. number of school-years completed]: "How many years did you go to school?"[78, 157] The highest educational attainment value recorded in a continuous format for pregnant women in 1998 and 2002 was 12 schooling years. Beyond 12 schooling years, educational attainment was recorded as "greater than 12 school-years" [i.e., categorical]. For the 1994, and 2004 through 2011 data, educational attainment was captured as a continuous variable.

To accommodate the different coding systems during analyses, the following approaches were applied. First, because there were relatively fewer pregnant women [i.e., 1015 out of 82290] with greater than 12 schooling years, pregnant women with educational attainment beyond 12 school-years were combined with secondary school graduates into a "12 school-years or greater" category in all survey years to facilitate comparability across all survey years.

Second, for the 1994, 2004 to 2008 survey years in which educational attainment was recorded in continuous form, educational attainment was truncated at 17 school-years. Pregnant women who reported 17 schooling year or greater were designated as having completed 17 schooling years [i.e., 15 out of 82290 pregnant women had

educational attainment greater than 17 school-years]. Completion of 17 schooling years was regarded as a reasonably high number of schooling years for the study population based on the school system in Zambia [i.e., 7 year in primary school; 5 years in secondary; and assumed 5 years in university or college].

7.13.3. Pregnant woman's area of residence

Sentinel sites location was regarded as a proxy for residence of pregnant woman according to the urban-rural classification of geographic areas of the Government of the Republic of Zambia (GRZ). Misclassification of residence for some pregnant women is likely because some pregnant women from urban areas might have sought antenatal care from rural areas or vice versa. Indicator variable coding system was used to code categorical variable (i.e., residence coded as "1" if the site is in an urban area and as "0" if the site is rural area).

Given that there were more pregnant women in urban residence category than rural residence category, use of the urban category as the reference category was the methodologically favored approach because of benefits that arise from variance stabilization are maximized when the largest category is used as the referent.[204] However, the difference in the number of pregnant women in urban and rural areas was not considerable, and rural residence category was used as reference category to facilitate a communication framework that is consistent with prior reports.[204]

7.13.4. Parity (i.e., number of women birthed by pregnant women)

Parity was coded in a categorical format for the 1994 data, but as a continuous variable in some years, and truncated at seven in other survey years. To capture data from all the survey, parity was coded according to the 1994 coding (i.e., zero, one and two or more children), and as indicator variables. Pregnant women with no children were used as reference category.

7.14. Random component variables (level-2) used in the CCREM-based APC analyses

7.14.1. Birth cohort were derived from pregnant women's birth year

Pregnant woman's birth year was not recorded in all survey years, but was computed by subtracting pregnant woman's self-reported age from calendar year of the survey. Figure 7.1 illustrate birth-cohort computations. Pregnant women were grouped into nine 5-year intervals as birth cohorts and a one 2-year interval birth-cohort [i.e. 1994-1996] based on birth year beginning with 1950-1954 birth cohorts through 1995-1996 birth cohorts. Because observations in the 1995-1996 birth cohorts were fewer, pregnant women in the 1990-1994 birth-cohort and those in the 1995-1996 birth-cohort were coalesced into a single birth-cohort, the 1990-1996.

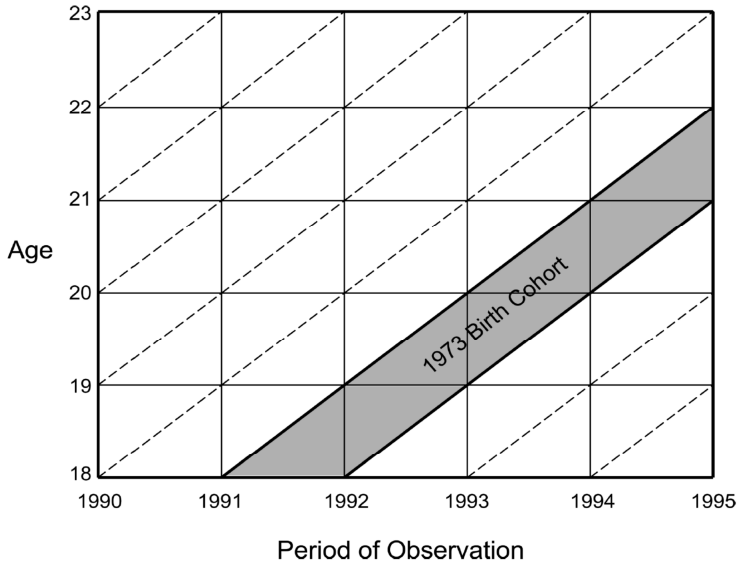

Figure 7.1. The diagram to illustrate a 1973 synthetic birth cohort, published by Reither et al (2009). Subtracting a respondent's age from the period of observation (i.e., survey calendar year) enables creation of a synthetic birth cohort (i.e., 1991-18=1973)

7.14.2. Survey calendar years were regarded as period

The seven survey calendar years in which data were collected for the ANC-HIV-SS were designated as period group effects (i.e., 1994, 1998, 2002, 2004, 2006, 2008, and 2011).

7.15. Statistical data analysis

Statistical analyses were conducted using the R program, a freely available statistical and computing.[188] [331] The library lme4 in R version 3.0 was applied for analysis.[188, 262]

7.15.1. Descriptive statistics for the analytic sample

Median, which is less sensitive to outliers than mean, and interquartile range were computed to describe the characteristics of the continuous variables for the overall sample, and by HIV serostatus.[211] Counts and proportions were used to describe categorical variables. To compare distribution of pregnant women by selected characteristics (e.g., age, educational attainment, residence, period, and birth cohort) between pregnant women who were and were not HIV seropositive, I used Chi-square test for categorical variables and the Wilcoxon sum rank test for continuous variables.[211] Counts and proportion were computed for missing values on covariates of interest.

7.15.2. Age was categorized to enable estimation of age-group specific HIV prevalence

To describe age distribution of pregnant women in the sample, and to facilitate estimation of HIV specific HIV prevalence, the continuous variable age (i.e., full-years interval between birth year and time of interview) was categorized according to the guidelines by the World Health Organization (W.H.O) for HIV prevalence reporting: 15-19, 20-24, 25-29, 30-34, 35-39, and 40-44 years. Within each age-group, the proportion of HIV seropositive pregnant women was computed, along with 95% Wilson confidence interval. To closely match the age category used by WHO/UNAIDS for reporting HIV prevalence, pregnant women were grouped according to the following categories: 15-24, 25-34, and 35-44, and pregnant women $\geq$45 and girls <15 excluded from the ANC-HIV-SS.

To assess the proportion of pregnant women who had completed at least 12 schooling years (i.e., completion of 12 schooling years is pre-requisite for college, university and some employment opportunities), a binary categorical variable was created: less than 12 schooling coded as "0" and greater or equal to 12 schooling years coded as "1". Noteworthy, continuous variables (i.e., age and educational attainment) were not categorized in all CCREM-based APC analyses.

7.15.3. Birth-cohort and period cross-classified data structure

To explore the extent of cross-classification table, cross-tabulations of pregnant women of birth-cohort groups and period groups were created. HIV prevalence estimates and the corresponding 95% Wilson CI were computed within birth-cohorts for the period 1994 through 2011.

7.16. Age, period and cohort effects were examined using CCREM-based APC modeling

Random intercept CCREM-based APC methods were used to assess age, period and birth cohort influence on HIV prevalence. The random component of the CCREM captured the variability of the overall mean HIV prevalence from cohort to cohort, and from period to period. The logit link function was applied because the outcome variable HIV serostatus, was dichotomously defined, and therefore assumed to follow a binomial distribution. Phrased differently, period group effects and cohort group effects modeled as random effects were used to explain period specific and birth-cohort specific variation in HIV prevalence.

7.16.1. Laplace approximation used for estimating likelihood function for parameter estimation in CCREM-based APC modeling

Because the likelihood function for binary response outcomes does not have a closed form solution, parameter estimation for the CCREM was achieved via maximum likelihood estimation, approximated by Laplacian approximation which uses high dimensional integration to estimate the maximum likelihood function.[267, 328, 331, 332] Laplacian approximation facilitates approximation of the closed likelihood function.

7.16.2. Assumptions of the GLMM-based CCREM

First, the random effects model was that the random effects were independent of the fixed effects (i.e., explanatory covariates). Second, the outcome variable, HIV serostatus was binomially distributed.

7.17. Ten CCREM (i.e., random intercept) were fit to the ANC-HIV-SS data to investigate various aspects of the research questions

To address various aspects of the research question, ten CCREM-based APC models, Model #1 through Model # 10 were fit to investigate age, period, and birth cohort effects on HIV prevalence. The model intercepts were defined as mean log-odds of prevalent HIV infection among pregnant women without period group and without birth cohort group effect. [333, 334]

7.17.1. Within-cohort and within-period ICC calculated using unconditional model

The variance estimates derived the unconditional Model #1 were applied to estimate the intra-class correlation (ICC) for period groups and cohort groups which represented the proportion of the total variance that was due to the period group influence and due to cohort group influence respectively. The unconditional model contains random components (i.e., period and birth-cohort variables). Fixed effects variables are not a part of the unconditional model (e.g., age or other individual-level variables). [335] Guided by statistical literature, the residual variance for a logistic regression model was assumed equivalent to $\pi^2/3$ as shown below.[267]

$$ICC_{period} = \left(\frac{\sigma^2_{period}}{\sigma^2_{Birth\ cohort} + \sigma^2_{period} + \left(\frac{\pi^2}{3} \right)} \right) \text{and } ICC_{Birth\ cohort} = \left(\frac{\sigma^2_{Birth\ cohort}}{\sigma^2_{Birth\ cohort} + \sigma^2_{period} + \left(\frac{\pi^2}{3} \right)} \right)$$

7.17.2. Age-only adjusted CCREM-APC analysis

Model #2 was equivalent to the classical age-period-cohort analysis. Age was the only fixed effect covariate, and birth cohort groups and period groups as random effects covariates. To relax the assumption of linearity, age was modeled nonlinearly, using restricted cubic spline (RCS) function with four pre-specified knots.

The odds of prevalent HIV infection and the corresponding 95% CI were computed from the specific covariate (e.g., age) regression parameter estimate and their standard errors. Briefly, log-odds of prevalent HIV infection for age 15, 19, 26, 29 34, and 39 years were compared to log-odds of prevalent HIV infection for age 24 years (i.e., median age).

7.17.3. Computation of period and birth cohort specific effects from CCREM-based APC

To compute period group specific effects and birth cohort group specific effects, the difference between respective group specific estimates (e.g., 1970-1974 birth cohort effects) and overall mean effect (i.e., intercept) were computed. Under this framework, the null hypothesis was absence of periods or absence of birth cohort effects. Lack of deviation from the value of the intercept implied a lack of group specific effect (i.e., deviation equivalent zero within a period and birth cohort group).

7.17.4. Difference between the group (e.g., 1950-1954) effect and intercept was equivalent to group specific cohort effect

Given that the deviation from the intercept (i.e., expected mean effect for the population) represents cohort group specific effect, its exponentiation (i.e., group specific effect minus model intercept) would yield the cohort group-specific odds ratio.[334, 336] Consequently, birth cohort effect of each of the nine birth-cohorts groups were computed as described above. Model intercept was regarded as the expected value or mean effects of population studied, and served as the referent value for estimation of group specific cohort effects.[74, 334] Period were represented by survey calendar year, and period group specific effects for the seven survey period were computed in a similar manner as described for computation of cohort group specific effects, by subtracting model intercept [reference category] from each of the seven period group effects.[74, 334]

For all subsequent models, group specific odds ratio for prevalent HIV infections for a specific birth cohort or period were computed by subtracting the CCREM-based APC overall model intercept from the intercept of a specific cohort or period (e.g.,

intercept for 1970-1974 birth cohort minus model intercept, which was regarded as a measure of deviation from the population mean), and subsequent exponentiation of computed difference.[74]

7.17.5. Age and residence adjusted CCREM-APC analysis

To assess the influence of urban and rural residence, Model #3 was adjusted for age and residence (i.e., urban and rural site location represented were regarded as proxy for pregnant woman's urban and rural residence) modeled as fixed effect covariates, and birth cohort and period factors were modeled as random effects. As in Model #2, linearity assumption in the relationship between age and log-odds of prevalent HIV infection was relaxed by fitting age as RCS function with four pre-specified knots.

Odds ratios and 95% CI for age, birth cohort, and residence were computed from regression parameter estimates and corresponding standard errors. Similar to analyses in Model #2, log-odds of prevalent HIV infection for ages 15, 19, 26, 29, 34, and 39 years were compared to log-odds of prevalent HIV infection for age 24 years (i.e., median age). Further, indicator variable coding was applied to residence, with rural residence designated as the reference group. The specific cohort group effects and specific period group effects were computed as explained for Model #2.

7.17.6. Age and educational attainment adjusted CCREM for APC analysis

To examine educational attainment effects, Model #4 was restricted to data collected between 1994 and 2008 because educational attainment was not captured in the 2011 ANC-HIV-SS. To explore non-linear relationship and avoid identification problem,

age and educational attainment were modeled as continuous variables using RCS function with four pre-specified knots. For fitting Model #4, educational attainment was truncated at 12 schooling years for all the survey years as explained earlier. Model #4 was adjusted for age, educational attainment and parity (i.e., number of children birthed by pregnant woman).[182]

7.17.7. Age, parity, residence and educational attainment adjusted CCREM for APC analysis using multiply imputed data

Complete case analysis can lead to information loss, selection bias, and incorrect inference if the missing completely at random (MCAR) assumption is not tenable. To accommodate observations with incomplete data on some variables of interest (e.g., educational attainment), and to avoid complete case analysis, Model #5 (i.e., adjusted for age, residence, educational attainment, birth cohort and period) was fit to 10 multiply imputed data sets.

7.18. Ten imputed data sets were created to estimates fixed effects of the CCREM

Missing data on variables were filled-in using multiple imputations performed by the Amelia package in R statistical and computing program, which uses combinations of bootstrap and maximum expectation to impute missing values.[188, 337] First, 10 copies of the original data set were created where each of the 10 data sets had missing values filled by random values generated from the specified predictive model for multiple imputations.

The predictive imputation model within Amelia program seeks to capture the quintessential features of the distribution of missing data through the relationship of

subject with others who do not have missing data, based on similarity of covariates and HIV serostatus.[338] Data were assumed to be missing at random (MAR) a less stringent assumption compared to MCAR assumption.[188] The pre-specified analytical models for each research question were fit to each of the 10 data sets with multiply imputed data generated by the Amelia program.[337] The parameter estimates and their corresponding standard errors were combined subsequently using rules proposed by Rubin (1976). [190, 191, 337]

Fitting analytical models to multiply imputed data is preferable because multiple imputations technique supersedes other methods for handling missing data methods (e.g. stratification on missing data, conditional mean imputation, or complete case analysis).[339, 340] [341-345] The elegance of the multiple imputation technique lies in its ability to integrate within-imputed-data set variability and between-imputed-data set variability in the computation of the estimated parameters and their standard errors during the post-imputation stage.[338, 346]

7.18.1. Age, parity and educational attainment adjusted CCREM for APC analysis using multiply imputed data stratified by residence

Based on prior reports of differential geographical distribution of prevalent HIV infections, with heavier burdens in the urban than rural areas, I conducted separate analyses for urban and rural pregnant women. To determine whether my findings were robust, and sensitive to highest value at which educational attainment was truncated, Model #6 was fit using educational attainment truncated at 17 schooling years as explained earlier.

To assess whether urban or rural residence modified age effects on HIV prevalence, I assessed statistical significance of multiplicative cross-product terms between age and residence. Two nested models were fit to the same data. The first GLMM contained age as the primary exposure variable and HIV as the outcome variable, and was adjusted for urban residence. The second model was similar to the first but also contained a cross-product term (i.e., age*residence). A likelihood ratio test (LRT) was applied to compare the two nested models, and the LRT for this analysis of statistical multiplicative interaction was not pre-specified. Because the p-value associated with LRT was significant, stratified CCREM-based APC analyses by urban and rural residence were performed.

Model # 7 was restricted to rural data while model # 8 was restricted to urban data. Both Model #7 and Model # 8 were adjusted for age, educational attainment, and parity. Because educational attainment (not collected in 2011) was considered, only the 1994 through 2008 data were used for these analyses. For Model #7 and Model #8, educational attainment was truncated at 17 schooling years such that all values $\geq$17 years were coded as 17 years of education. Further, Model # 9 and Model# 10 were age-only adjusted models for rural and urban areas respectively for the period 1994 through 2011.

7.18.2. Age fitted using restricted cubic spline function (RCS)

Age was fit as a continuous variable RCS function with four pre-specified knots to relax the linearity assumption. Because the relationship between age and log-odds prevalent HIV infection may not be captured adequately if the functional form of age is improperly expressed.[149, 264, 347-349] RCS functions transform a continuous

variable in such a way that the curve is linear before the first knot, represents a piecewise cubic polynomial between adjacent knots, and is linear after the last knot.[149, 350] RCS function of age and educational attainment in regression models may effectively capture non-linear relationships, and minimize residual confounding. Consequently, modeling continuous variables in their continuous format obviates subjective categorizations and enhances control of potential confounding variables.

7.18.3. LRT test used to compare nested model models for assessing linearity assumption

To evaluate the linearity assumption for continuous variables, two models were fit. First, CCREM-based APC was fit where a linear relationship between continuous variable (e.g., age and educational attainment) and log-odds of prevalent HIV infections was assumed. Second, CCREM-based APC was fitted assuming a non-linear relationship (i.e.. using RCS) between continuous variable (e.g., age and educational attainment) and log-odds of prevalent HIV infections.[149, 263]

The difference in the log-likelihood of the two nested models was evaluated using the LRT to assess the tenability of the linearity assumption, based on the null hypothesis of no difference in log-likelihood.[149, 264, 265] Significant p-values (<0.05) for the LRT implied that the linearity assumption may not be tenable, and models where continuous variables were fit using RCS functions may be better to adequately capture the relationship between age and log-odds of prevalent HIV infection.[149, 263] Assessment of the linearity assumption was conducted for age and educational attainment. For example, detection of a nonlinear relationship between age and log-odds

of prevalent HIV infections implies that the relationship is non-constant over the range of age observed age values.[266]

7.18.4. P-values associated with LRT for comparing models with and without random components are conservative

Inclusion of random effects covariates were not based on statistical tests but were pre-specified. Therefore, I did not assess whether random group effects were significantly different. Bates and Pinheiro (2000) cautions using LRT statistic for assessment of the significance random component in GLMM.[265, 331, 351]

7.18.5. Sensitivity analyses to check the robustness of estimates from fitted models

As a type of sensitivity check to assess the robustness of parameter estimates, two CCREM-based APC models were fit. First, Model #5 adjusted for age, parity, residence and educational attainment truncated at 12 schooling years. Second, Model #6 covariates had the same covariates as in Model #5 except educational attainment was truncated at 17 schooling years [i.e., 1994, and 2004 through 2008]. To assess the robustness of the estimated parameters, the parameter estimates from Model #5 and Model # 6 were compared. Further sensitivity checks using models fitted to data with and without imputed values were performed.

7.18.6. Ethics

The Ethics and Research sub-Committees in Zambia cleared the ANC-HIV-SS and Vanderbilt University Institution Review Board (IRB) reviewed and approved this secondary analysis of ANC-HIV-SS data.

7.19. Results

7.19.1. Description of study sample

The study sample comprised 82561 pregnant women aged 15 to 44 years with HIV serostatus data from the seven rounds of ANC-HIV-SS. These included 9760 women in 1994, 11907 in 1998, 13051 in 2002, 12404 in 2004, 13260 in 2006, 13298 in 2008, and 8610 in 2011. An estimated 99.4% of pregnant women (i.e., 82086) in the present study had serologically confirmed HIV serostatus, either seropositive or seronegative. The proportion of pregnant women recruited from rural sites were lower (42.3%) compared with proportion of pregnant women recruited urban sites (57.7%).

The median age for the study sample of pregnant women in the 1994 through 2011 surveys was 24 years and the IQR was 20 to 29 years. The median age for HIV seronegative women was 23 years (IQR: 20 to 29 years), compared to 25 years (IQR: 22 to 29 years) in HIV seropositive women (p-value=0.001). The sample median educational attainment was 7 years [IQR: 5 to 9 school-years]. The number of schooling years completed varied from 0 to 17+ years (truncated at 17 year to ensure plausible values).

The reported statistically significant findings should be interpreted with caution because of the large sample size in this study; statistically significant differences in age distribution by HIV serostatus may or may not be substantively important as shown in Figure 7.2. Among pregnant women born aged 15 to 44 years in the study sample, 67,105 (81.3%) were born between 1970 and 1989. Figure 7.2 shows survey year-specific age distribution for the studied pregnant women by age group.

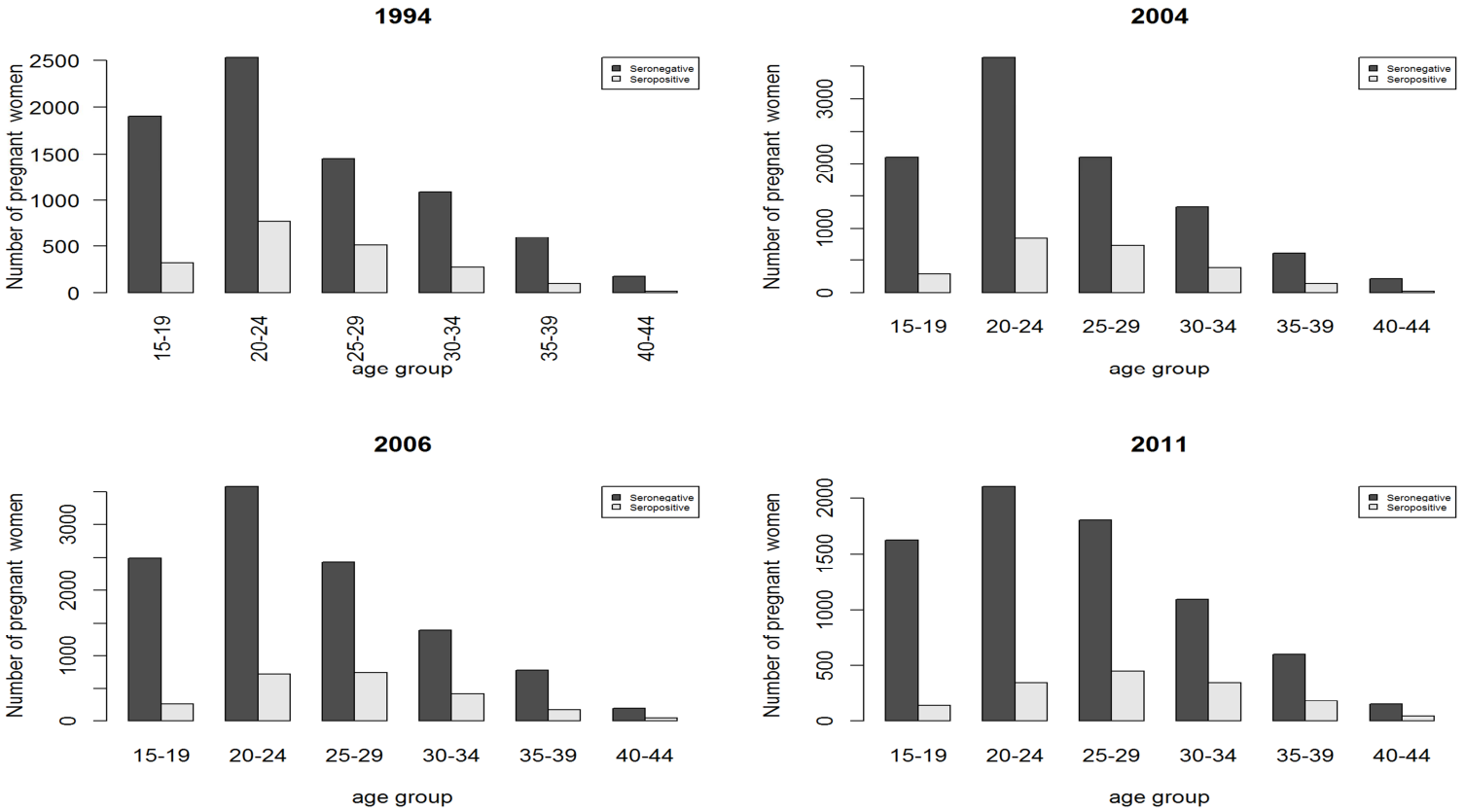

Figure 7.2. Distribution of pregnant women by HIV serostatus by age group, beginning age group 15 to 19 years and ending with age group 40 to 44 years based on ANC-HIV-SS data collected in 1994, 2004, 2004 and 2011 [1998, 2002 and 2008 not presented because year-specific distributions were identical]. Fewer pregnant women were in the 40 to 44 year age group, while 91% of the studied pregnant women ANC-HIV-SS in Zambia between 2006 and 2011 were < 35 years old.

7.19.2. Low proportion of missing data on covariates used in the CCREM-APC analyses

Four fixed covariates (i.e., age, educational attainment, parity, and residence) included in fitted multivariable regression analyses (Table 7.2). With few exceptions (i.e., education), the proportions of missing data on the covariates used for regression analyses were generally low across the seven survey rounds. For example, among 9724 pregnant women who had an HIV serostatus result in 1994, educational attainment data were incomplete for 1014 (10.4%) pregnant women. Among the 13,223 pregnant women recruited in 2006, only 236 (1.8%) had missing educational attainment data. Data on educational attainment were missing in <1% of pregnant women recruited in 1998, 2004, and 2008 as shown Table 7.2.

Table 7.2 shows considerable missing data were noted for parity (i.e., number of children birthed by pregnant women) in the 2011 data where 1366 pregnant women (15.5%) out of 8881 had missing data on parity, but the proportions of pregnant women with incomplete data on parity in the remaining survey years were low. For example, 263 out 9724 pregnant women (2.7%) in 1994 did not have data on parity.

7.19.3. Highest proportion of seropositive pregnant women in the 1974-1979 birth cohort

Within-birth cohort proportions of HIV seropositive pregnant women among the nine birth cohorts ranged from 0.1% in the 1950-1954 birth-cohorts to 27.5% (1975-1979 birth cohorts. The proportion of HIV seropositive women were relatively high in the 1970-1974 birth-cohort (20%) and the 1980-1984 birth-cohort (25%).

7.19.4. Distinctive decline of HIV prevalence in the 15 to 24 year-olds

Table 7.2 and Figure 7.3 indicate a nearly linear fall in HIV prevalence among 15 to 24 year-olds from 19.7% in 1994 to 11.5% in 2011. Although a slight fall was observed between 2002 and 2011, Figure 7.3 indicates a fairly stable burden of HIV infection in the 25 to 34 year-olds remained stable, lowest in 2011 (21.6%). HIV prevalence in the 35-44 year-olds increased notably from 12.5% in 1994 to 23.5% in 2011. The smaller numbers of pregnant women in the 35 to 44 year-old age group compared to the 15 to 24 year-olds warrants caution when interpreting HIV prevalence in the 35 to 44 year-olds. HIV prevalence estimates among pregnant women aged 15 to 24, 25 to 34 and 35-44 are presented in Table 7.2. Wilson's score method was applied to compute the 95% CI.

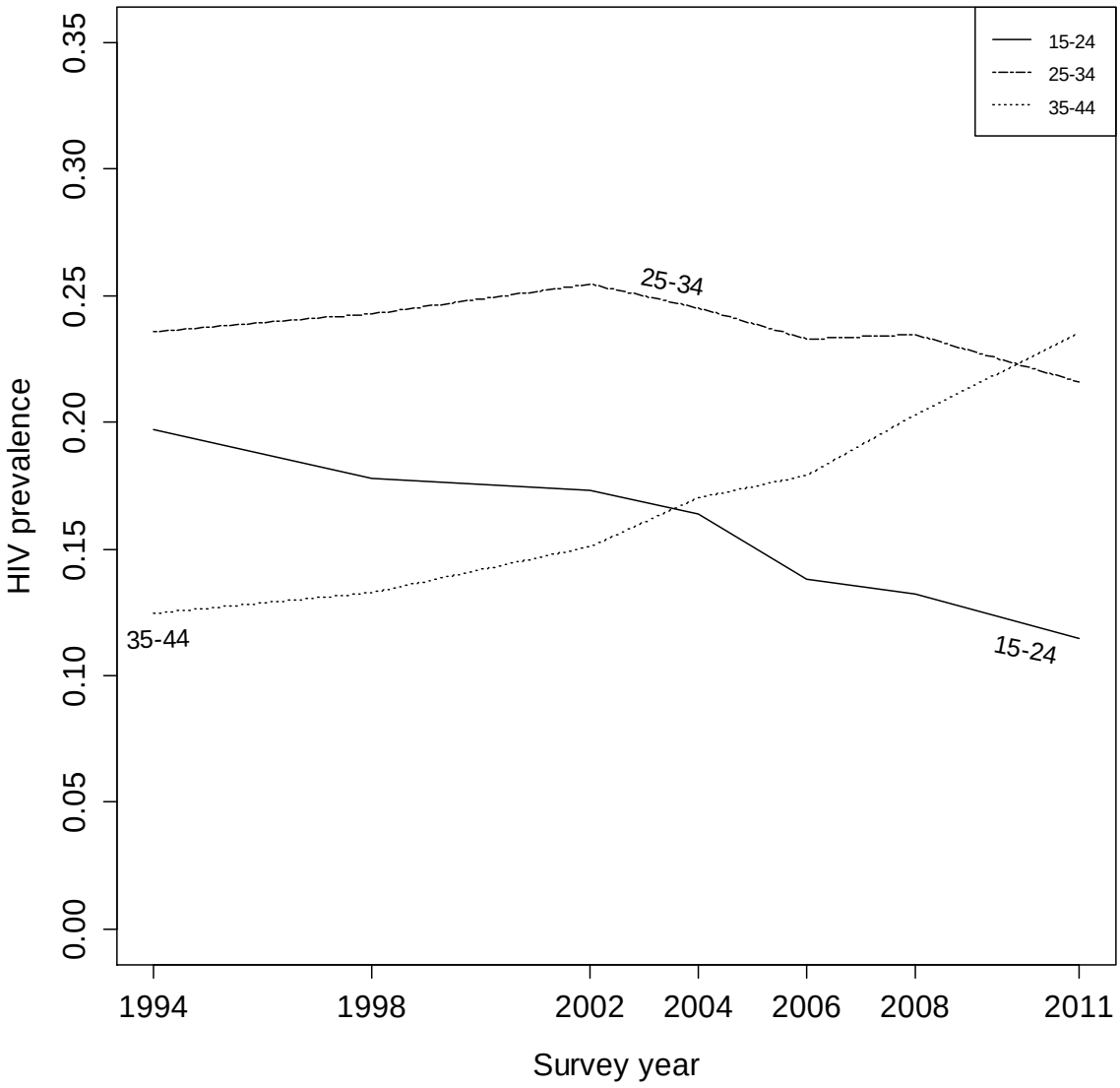

Figure 7.3. Age-group specific HIV prevalence trends among pregnant in the for age groups 15 to 24, 25 to 34 and 35 to 44 based on ANC-HIV-SS data collected in Zambia between 1994 and 2011. UNAIDS recommends using prevalent HIV infections in the 15-24 year-olds as proxy for number of new HIV infections. Graph indicates a distinctive decline in HIV prevalence in the 15 to 24 year-olds, from 19.7% in 1994 to 11.5% in 2011. HIV prevalence among 25 to 34 year-olds between 1994 and 2011 fluctuated around 24%. HIV prevalence in the 35-44 year-olds rose from 12.5% in 1994 to 23.5% in 2011.

7.19.5. Overall HIV prevalence appear to be increasing as birth-cohort ages

Table 7.3 and Figure 7.4. presents within-cohort HIV prevalence trends for the period 1994 to 2011, suggesting a tendency of increasing HIV prevalence in recent birth-

cohorts compared to older birth cohorts. HIV prevalence among pregnant women within the 1990-1996 birth-cohort increased from 8.5% (95% CI: 6.5%, 11.0%) in 2006 to 10% (95% CI: 8.9%, 11.2%) in 2011, and HIV prevalence in the 1960-1964 birth cohort decreased from 20.1% (95% CI: 18.1%, 22.3%) in 1994 to 11.5% (95% CI: 4.0, 29%) in 2011.

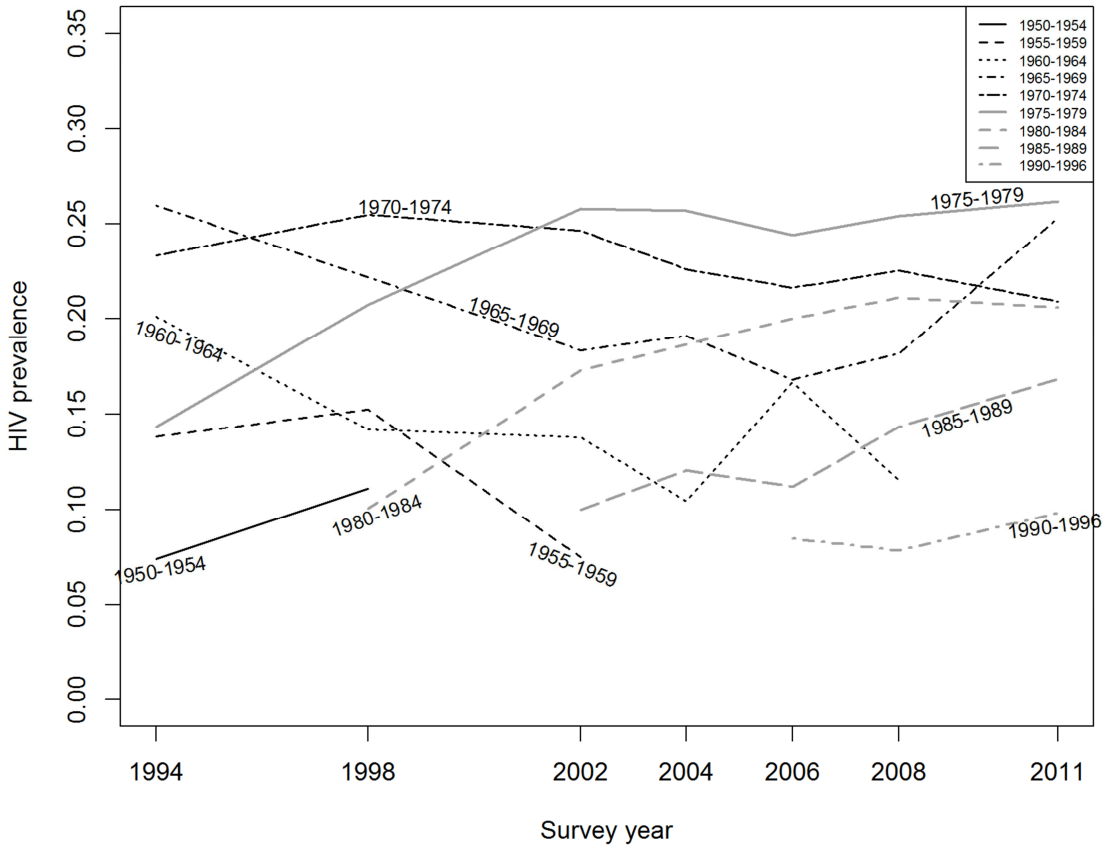

Figure 7.4. Trends in HIV prevalence within birth-cohorts from 1950-1954 through 1990-1996 over the survey periods 1994 through 2011 based on ANC-HIV-SS data among pregnant women in Zambia. For pre-1974 birth-cohorts, HIV prevalence started off at a high prevalence level, and then declined as the birth-cohort aged, presumably due to deaths and lower risk behaviors in older women. On the other hand, post-1974 birth cohorts (e.g., 1985-1989 and 1990-1996) started off with low HIV prevalence, and then swung upwards as the epidemic expanded.

7.19.6. Nonlinearity assumption relaxed using RCS function for the relationship age and log-odds of HIV prevalence

To assess whether association between age and log-odds of prevalent HIV infections was linear, LRT was applied to evaluate the difference between the log-likelihoods of two nested models, one model fitted assuming a linear relationship and second model fitted assuming non-linear relationship.[149, 264] The p-value <0.001 associated with LRT was less than 0.001. Therefore, age was modeled as a continuous variable using RCS function since the significant LRT suggested that the relation between age and log-odds of prevalent HIV infection may not be linear.[149]

7.19.7. Nonlinear function adequately captured relationship between educational attainment and log-odds of HIV prevalence

The relationship between educational attainment (i.e., number of schooling years) and log-odds of prevalent HIV infection may not be linear. To decide which model captured the relationship between educational attainment and the log-odds of prevalent HIV infection adequately between linearly and nonlinearly fitted models (i.e., model of educational attainment fit using a linear function and model in which educational attainment fit using a RCS function), LRT test for nested models was applied. The LRT statistic yielded a statistically significant p-value <0.001, implying that the relationship between educational attainment and log-odds of prevalent HIV infection may be not be linear, and consequently, educational attainment was fit using RCS function in the CCREM-based APC model.[149, 204]

7.20. Findings based on the CCREM-based APC regression analyses

Table 7.4 through Table 7.6 presents findings from random intercept CCREM-based APC analyses (i.e., GLMM) where two higher level (e.g., level-2) grouping covariates are cross-classified, and modeled as random components. The outcome variable in the current analyses was HIV serostatus (i.e., dichotomous).

7.20.1. Intercept-only CCREM-based APC model

Table 7.4 presents findings based on Model # 1, the intercept-only logistic regression model (i.e., unconditional model without fixed effects covariates). Using Model #1 variance estimates, within-period and within-cohort ICC were 5.8% and 0.34% respectively. The estimated predicted probability of prevalent HIV infection for a typical pregnant woman in this population was 16.0%. Assuming normal distribution, and considering period and cohort group effects, predicted probability of prevalent HIV infection at 95% CI would vary from 14.0% and 21.6%.

7.20.2. Prominent birth-cohort influence in 1970-1974 and 1975-1979 birth cohorts associated with elevated odds of prevalent HIV infection

Model #2 was the age-only adjusted CCREM-based APC model, and birth cohort group effects and period group effects were defined as digressions from the intercept (i.e., mean log-odds of expected prevalent HIV infection).[74] Birth cohort effects were most prominent in the 1975-1979 birth-cohort (OR=1.36, 95% CI: 1.31, 1.42), slightly but not substantively higher than 1970-1974 birth cohort (OR=1.34, 95% CI: 1.28, 1.40). Further results are presented in Table 6.

Compared to the estimated overall mean of the log-odds of prevalent HIV infection of a typical pregnant woman in this population, the odds of prevalent HIV infection were 36% higher among for a pregnant woman in the 1975-1979 birth-cohort. Based on the 95 % CI for the 1975-1979 birth-cohort group effect, the odds of prevalent HIV infection of a pregnant woman in that birth-cohort could be from 31% to 42% higher than for a typical pregnant woman in the studied population.

7.20.3. Reduced odds of prevalent HIV infection in the 1985-1989 and 1990-1996 birth-cohorts

The odds ratio and 95% CI for the odds of prevalent HIV infection for 1985-1989 birth-cohort (OR = 0.88, 95% CI: 0.84, 0.93) and 1990-1996 birth-cohort (OR=0.76, 95% CI: 0.69, 0.84) suggest protective birth cohort effects. The 95% CI does not include OR=1.0, and can therefore be considered meaningfully different from the null value.

7.20.4. Higher odds of prevalent HIV infections in urban compared to rural women

Table 7.4 also presents findings based on Model #3, which was a CCREM-based APC adjusted for age and residence. The odds of prevalent HIV infection were 2.5 times higher among pregnant women in urban areas compared to pregnant women in rural areas (OR=2.53, 95% CI: 2.44, 2.64).

7.20.5. Separate age-only adjusted model were performed for urban and rural areas

Because there was evidence of multiplicative statistical interaction between age and residence (p-value <0.001), separate CCREM-based APC analyses were performed for urban and rural areas for age-only adjusted models, and results are presented in Table 7.5.

Model #7 was based on the data collected from rural areas in which age was the only fixed effect covariate. The elevated odds of prevalent HIV infections were observed for birth-cohorts of 1970-1974 (OR=1.16, 95% CI: 1.07, 1.26) and 1975-1979 (1.28, 95% CI: 1.19, 1.38). Similarly, based on Model #8 the age-only adjusted model fitted to fitted to data captured from pregnant women in urban areas, the odds of prevalent HIV infections were more prominent for birth-cohorts 1970-1974 (OR=1.41, 95% CI: 1.34, 1.49) and 1975-1979 (OR=1.35, 95% CI: 1.29, 1.41). Among pregnant women in urban areas, birth cohort effects were prominent and significant for the 1965-1969 birth-cohort effects (OR=1.33, 95% CI: 1.23, 1.44), but significant for rural pregnant women (OR=1.00, 95% CI: 0.89, 1.12). Detailed reports of birth cohort effects are in Table 7.4

7.20.6. Pronounced odds of prevalent HIV infections for the 1970-1974 and 1975-1979 birth cohort in urban than rural areas

Separate CCREM-based APC were fitted for rural and urban areas, Model #9 and Model #10 respectively, and were adjusted for three covariates: age, parity, and educational attainment. Estimates of odds ratio and 95% derived from Model #9 and Model #10 are in Table 7.6. Table 7.6 presents findings from Model # 9 and Model # 10, the fully-adjusted models for rural and urban areas, respectively. Protective birth cohort influence were observed among pregnant women in urban areas for the 1985-1989 birth cohort (OR = 0.79, 95% CI: 0.74, 0.84) and the 1990-1996 birth-cohort (OR = 0.68, 95% CI: 0.60, 0.76). Similarly, among pregnant women in rural areas, birth cohort effects were protective for the 1985-1989 birth cohort (OR=0.88, 95% CI: 0.79, 0.97), but not for the 1990-1996 birth cohort (OR=0.94, 0.79, 1.12).

7.20.7. Prominent birth cohort effects (i.e., 1985-1989 and 1990-1996) after adjusting for parity and educational attainment for urban areas

Observations from 2011 round of ANC-HIV-SS were excluded from analyses that assessed educational attainment effects because educational attainment was not collected in 2011. Model #9 and Model #10 were fitted to data collected between 1994 through 2008. Estimated parameters did not change materially after adjustment for age, educational attainment and parity in Model # 9 for rural areas. Model #10 in Table 8 indicate that birth cohort effects remained protective for the 1985-1989 birth cohort in urban areas (OR=0.82, 95% CI: 0.77, 0.87) and the urban 1990-1996 birth cohort (OR=0.71, 95% CI: 0.63, 0.80). The significant protective cohort effects observed in the age-only adjusted model for rural areas in Model # 7 disappeared following adjustment for educational attainment and parity in Model # 9 for the 1985-1989 birth cohort (OR=0.97, 95% CI: 0.88, 1.06) and the rural 1990-1996 birth cohort was still not significantly protective (OR=0.89, 95% CI: 0.76, 1.03) as indicated in Table 7.6.

7.20.8. Period effects were no significantly different across survey rounds

Period group effects were not statistically significant for both age-only, and the age, educational attainment and parity adjusted CCREM-based APC models. For example based on Model # 11, period group effects for the 1994 period (OR=1.02, 95% CI: 0.98, 1.06) and the 2011 period (OR=0.99, 95% CI: 0.96, 1.03) were not significant. Table 7.4 presents further details of period specific odds ratios and 95% CI.

7.20.9. Non-linear age effects detected for pregnant women between 15 to 44 years old

Pregnant women who were 26 year old had slightly increased odds of prevalent HIV infection compared to 24 year olds in both urban (OR=1.13, 95% CI: 1.02, 1.25) and rural areas (OR=1.09, 95% CI: 1.09,1.16). Overall, as age value increased, age influence measured as odds ratio for prevalent HIV infection compared to age 24 years were not materially different the null value (OR=1.0).

7.21. Sensitivity analyses for different truncation of educational attainment

There was no substantive difference in estimated log-odds of prevalent HIV infections from sensitivity analysis of CCREM-based APC regression model that included educational attainment truncated at 12 schooling years, and at 17 schooling years. Table 7.5 presents estimates of odd ratios and 95% CI based on Model # 4 in which educational attainment was truncated at 12 schooling years (i.e., schooling years greater than 12 years were designated as 12), and the Model # 5 in which educational attainment was truncated 17 years (i.e., schooling years greater than 17 were designated as 17). The odds ratio and 95% CI were nearly equivalent values. For example, based on Model #4 and Model #5, the odds ratio and 95% CI for the 1975-1979 birth-cohorts were OR=1.27, 95% CI: 1.20, 1.34 and OR=1.27, 95% CI: 1.21, 1.34), respectively.

7.22. Sensitivity analyses for complete and imputed data

Multiple imputation of missing data was conducted to replace missing values, and Model #6 fitted to multiply imputed data sets. Like Model #5, educational attainment variable used in Model #6 was truncated 17 schooling years. The estimated birth cohort

effects for 1970-1974 and 1975-1979 birth were OR=1.34, 95% CI: 1.27, 1.41 and OR=1.28, 95% CI: 1.21, 1.34), respectively.

7.23. Reduced odds prevalent HIV infections among pregnant women with ≥ 2 children compared pregnant women with no children

There were three categories for parity (i.e., number of children birthed by pregnant woman): no children; one child and ≥2 children. Compared to pregnant women with no children, the odds of prevalent HIV infection for pregnant with one child were slightly higher for pregnant women in urban areas (OR=1.11, 95% CI: 1.00, 1.20), but not for pregnant women in rural areas (OR=1.00, 95% CI: 0.83, 1.20) as indicated in Table 7.5. Based on Model #9 and Model #10, the odds of prevalent HIV infections for pregnant women who had ≥2 children were less than odds of prevalent HIV infection pregnant women with no children among pregnant women in rural areas (OR=0.71, 95% CI: 0.58, 0.90) and urban areas (OR=0.81, 95% CI: 0.72, 0.90) as shown in Table 7.4.

7.24. Graphs for birth cohort group, period and age effects adjusted to median age

Figure 7.5 generated using CCREM-based APC model adjusted to median age (24 years). Using this approach, predicted probability of prevalent HIV infection was greatest (24.6%) for pregnant women in the 1970-1974 birth-cohorts as shown Figure 7.5b and estimates for 1950-1954 and 1990-1996 birth-cohorts given the wide 95% CI. For the age-only adjusted CCREM-based APC model, and conditioning on median age of 24 years, the predicted probability of a typical pregnant woman increased from 15% (1994) to a peak of 25.1% (2008), and dropped to 22.1% (2011). Figure 7.6 and Figure 7.7 presents period effects and age effects respectively.

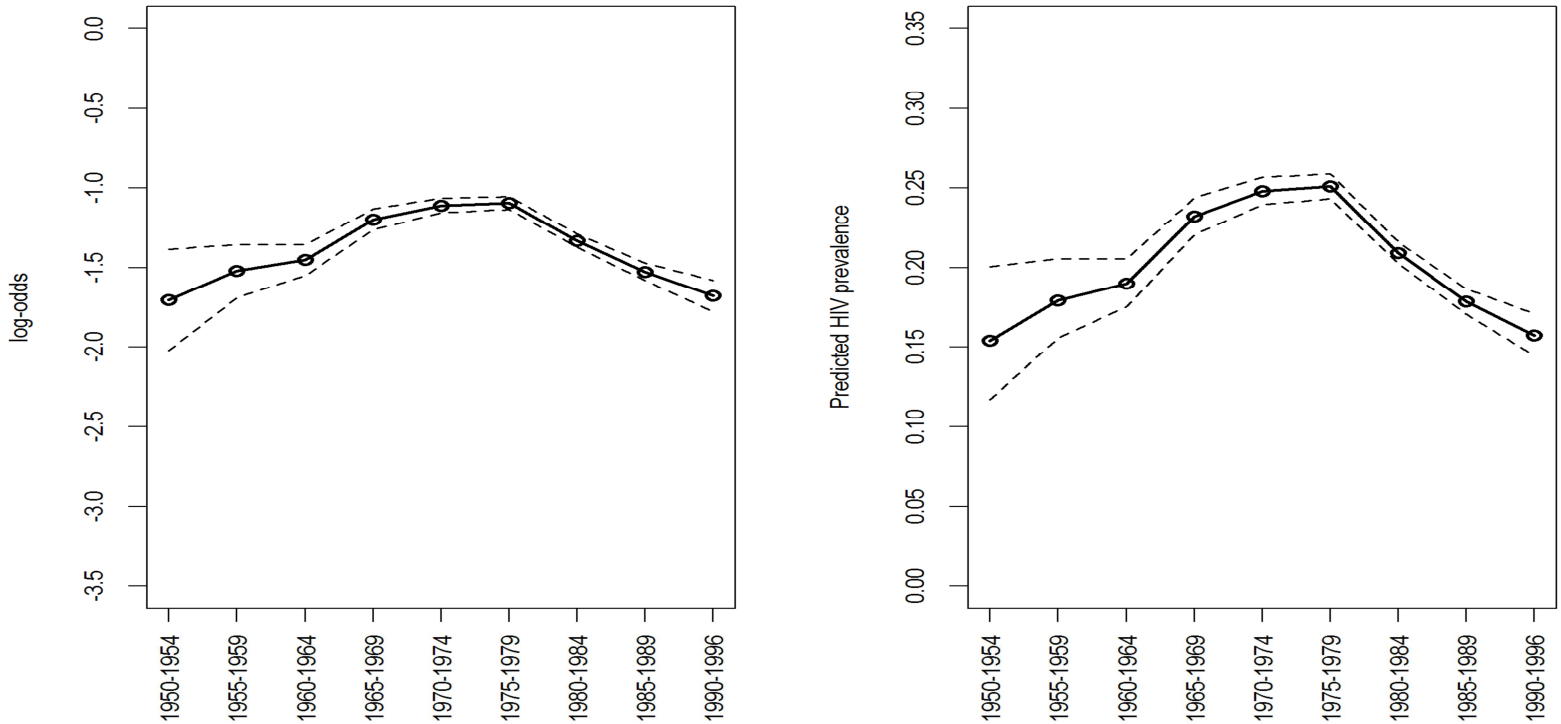

Figure 7.5. Log-odds of prevalent HIV infections in pregnant women attending antenatal care in a specific survey years, adjusted to median age of 24 years based on the age-only adjusted CCREM-based APC regression using ANC-HIV-SS data collected in seven surveys in Zambia between 1994 and 2011. The second graph shows cohort group effects represented as predicted probabilities of prevalent HIV infection at median age (i.e., 24 years) averaged over all the seven periods estimated from the age-only adjusted CCREM-based APC. An indication of prominent birth-cohort affects 1975-1979 birth-cohorts, followed by the 1970-1974 birth-cohorts. The 1990-1995 and 1995-1996 birth-cohorts were coalesced into 1990-1996 birth-cohorts to provide more stable estimates. Age-only adjusted CCREM-based APC parameters were estimated using Laplacian approximation of maximum likelihood. Fewer observations in the 1990-1994 and 1995-1996 birth cohorts; therefore these birth-cohorts were coalesced to avoid imprecise estimates of predicted probabilities.

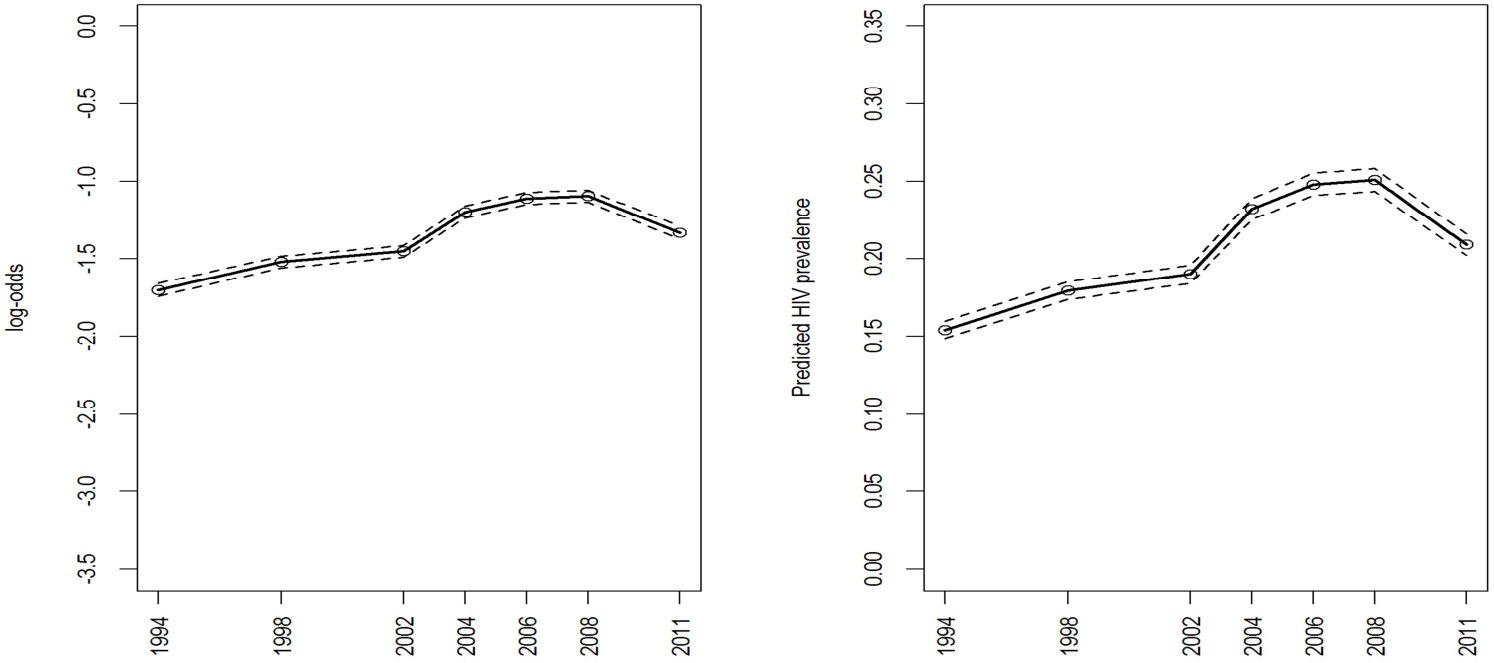

Figure 7.6. Estimated period group effects for the log-odds of prevalent HIV infection for pregnant woman in a specific period groups adjusted to median age (i.e., 24 years) and estimated from the CCREM-based APC with random period and birth cohort components and age as the only fixed covariate. The second graph shows estimated random group period effects from the age-only adjusted model, represented as predicted probabilities of being HIV seropositive for each survey calendar year at the median age (i.e., 24 years) averaged over all nine birth-cohort. The period effects for a 24 year-old, despite being less prominent, show a rise from a predicted probability of 15.4% in 1994 to a peak in 2008 of 25.1%, dropping in 2011 to 20.9%. The estimates are based on age-only adjusted CCREM-based APC analyses using ANC-HIV-SS data collected in Zambia between 1994 and 2011.

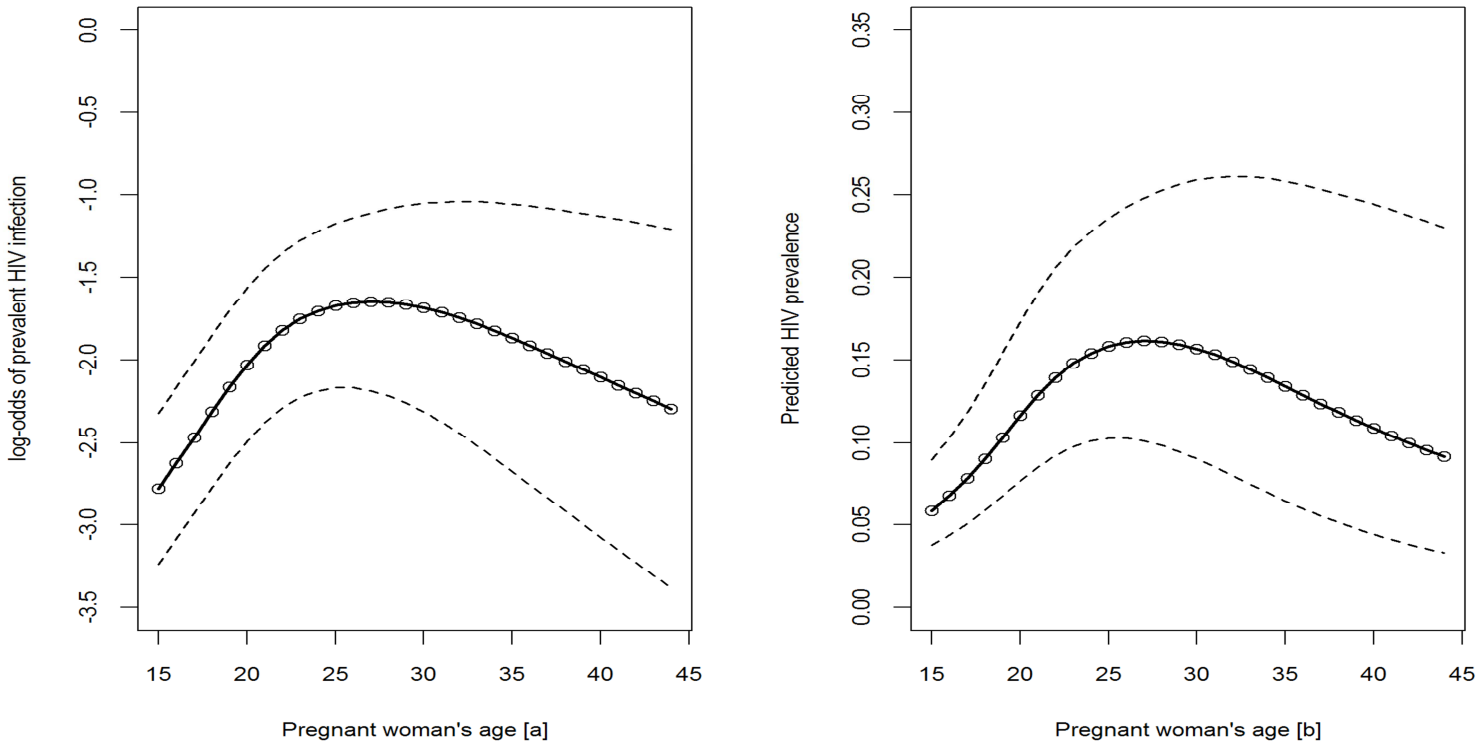

Figure 7.7. Age-only adjusted CCREM-based APC analysis [i.e., age, period and birth cohort] based data collected from pregnant women during ANC-HIV-SS conducted between 1994 and 2011 in Zambia. Fig 37a shows a ccurvilinear relationship between age and log-odds of odds of prevalent HIV infection. Fig 37b shows a nonlinear relationship age and predicted HIV prevalence. The log-odds of prevalent HIV infection increases, and peaks at age 27 years, and subsequently declines as age increases, with corresponding widening of 95% CI.

7.25. Discussion

Based on the method proposed by Yang and Land (2006) for estimating age, period and cohort effects in repeated cross-sectional surveys and using ANC-HIV-SS data for seven survey rounds between 1994 and 2011, the odds of prevalent HIV infection were relatively elevated for pregnant women in the 1970-1974 and 1975-1979 birth cohorts. On the other hand, protective birth cohort group influences were observed for pregnant women in urban areas who belonged to the 1985-1989 and 1990-1996 birth cohorts.

Although protective birth cohort influences were observed for pregnant women who belonged to the 1985-1989 birth cohort in rural areas, the protective influence did not persist following adjustment further adjustment of the regression model with educational attainment and parity. These findings of considerable birth cohort effects are contrary to my prior null hypothesis of no age, period and birth cohort effects.[66, 68, 74] Birth-cohort-group effects for the 1990-1974 and 1975-1779 birth cohorts were more pronounced in urban than rural areas. Examination of within-cohort HIV prevalence trends revealed increasing HIV prevalence in the 1985-1989 and the 1990-1996 birth cohorts.

The relationship between age and the odds of prevalent HIV infection among pregnant women were curvilinear. Profound protective age influences were observed for pregnant women who were ≤24 years of age. The odds of prevalent HIV infection for 24 year-old pregnant women were compared with the odds of prevalent HIV infections for selected age values. As an example, the analysis revealed that comparison of odds of prevalent HIV infection of pregnant women aged less than 24 years with odds of

prevalent HIV infection for pregnant woman aged 24 year-olds revealed that age effects for pregnant women younger than 24 years were protective. The protective effects of younger age did not persist as the age of the pregnant woman increased.

Even though the predicted probability of prevalent HIV infections (i.e., adjusted to median age of 24 years) increased gradually , peaked in 2008 but dropped in 2011 over the considered time period as shown in Figure 10, the estimated period group effects were not considerably different across the seven period for the time spanning 1994 through 2011. Further analysis revealed that period group effects were not substantively different from the mean odds of prevalent HIV infection among pregnant (i.e., no period and cohort influences), represented by the intercept of the model.

Most pregnant women in the 1970-1974 birth cohort were probably entering the age group in which most young women become sexually active around 1985, the year when the first AIDS case was reported in Zambia.[39, 86] Similarly, elevated odds of prevalent HIV infections were observed among pregnant women in the 1965-1969 and 1975-1979 birth cohorts. It would seem that a combination of factors (e.g., inaccurate information on HIV transmission mechanism and the obtaining sociocultural orientation possibly) drove up the odds of prevalent HIV infection exert their influence differentially by age. Pressed for marriage, young women may have greater higher likelihood of encountering an infected man, including and particularly the husband.[352] On the other hand, other studies have reported protective effects of marriage, while transmission may also occur within marriages.[353, 354]

The protective birth cohort group effects present among pregnant women in the 1985-1989 and 1990-1996 birth cohorts are consistent with reports of a decline in the number of persons in the 15-24 year-olds, newly infected with HIV in recent years.[10, 23, 24] The drop in the odds of new HIV infection among pregnant women might be a consequence of multiple preventive and treatment interventions intervention aimed at curbing the spread of HIV infection, and women in recent birth cohort became sexually active in an era of intensified prevention efforts. [281, 282, 355] Further, heightened estimated odds of prevalent HIV infection at age 27 years coincides with the when most women have potentially increased sexual activity (e.g., new marriage).

The current study is the first study in Zambia to examine age, period and birth cohort effects simultaneously. Additionally, no study has examined within birth cohort trends in HIV prevalence. However, the observed reduced odds of prevalent HIV in recent cohort is consistent with reports of reduced HIV prevalence in 15 to 24 year-olds, the age-group used for approximating new HIV infections [i.e., persons in 15 to 24 years age group are likely to have initiated sexual intercourse recently, and HIV-related mortality in the younger age group].[24] Consistent with prior findings of declining HIV prevalence, HIV prevalence declined overall but within birth cohort assessment of HIV prevalence trends revealed slight upward swing in HIV prevalence trends in the 1985-1989 and 1990-1996 in birth cohorts (i.e., 15 to 24 year olds) in 2011.[24, 36]

Further, as reported first in the study by Fylkesnes et al (2001), the prevalent HIV infection burden in Zambia is heavier in urban than rural areas, and my analyses also indicated more pronounced birth cohort group effects for 1970-1974 and 1975-1979 birth cohorts, particularly in urban areas.[36, 40].

Consistent with the use of HIV prevalence in the 15 to 24 years age group to approximate new HIV infections, the observed slight upward swings in HIV prevalence within the 1985-1989 and 1990-1996 birth-cohorts could be interpreted as possible upward shift in number of young women newly infected with HIV. However, there are fewer data points beyond 2008 to reliably confirm this observation. It is possible, though not probable, that the observed increase in estimated HIV prevalence might be a consequence of random variation. Further, increasing burden of HIV infection may be a function of new HIV infections and improved survival of HIV infected persons, especially with increased access to cART.

HIV prevalence in the recent birth cohorts (i.e., 1985-1989 and 1990-1996) is of great public health interest as women in the 1985-1989 and 1990-1996 birth cohorts are largely in the 15-24 age groups, the age group used for approximating number of women with new HIV infections. The observed decline in HIV prevalence among pregnant women in most pre-1975 birth cohorts were mainly due to HIV-related mortality, and reduced fertility among HIV positive women.[356] Further, women beyond 40 years are less likely to become pregnant, and consequently the number of older women within any cohort would naturally diminish as the cohort ages.

Compared to earlier birth cohorts such as pre-1970-1974, recent birth cohorts (i.e., post-1970-1979) started off at a lower HIV prevalence at first entry into the survey sample. This may signify reduced burden in the younger age group as has been reported in earlier studies.[24, 146, 152] One can speculate that awareness of the HIV epidemic

has increased and some risk factors for HIV infections has decreased among young women, possibly due the widespread HIV prevention interventions implemented over the years (i.e., HIV prevention program gained momentum in late 1990s). Consequently, one may speculate that young men and women have become more cautious regarding risky sexual behavior, contributing to the observed drop in prevalent HIV infection in the 15 to 24 year-olds.

7.25.1. Limitations

Pregnant women recruited for the ANC-HIV-SS were not a random sample; neither were the sentinel sites randomly selected. However, sites used for data collection were geographically well spread across the country to represent rural and urban areas. The pregnant women were not randomly selected, and our sample was not population-based, and comes with biases inherent in convenient sampling strategies, and therefore the study findings may be threatened by selection bias. Additionally, most women prefer concentrating child bearing in early stage of their reproductive period, therefore selection bias of younger women into the study is more likely.

Dzekedzeke & Fylkesnes (2006) compared ANC-HIV-SS-based and population-based (DHS) HIV prevalence estimates for the period 2001-2002 in Zambia, and found congruence in HIV prevalence estimates in urban and rural areas based on the data from 2001-2002 ANC-HIV-SS and from the 2001-2002 DHS.[357] For example rural areas ANC-HIV-SS-based and population-based HIV prevalence estimates were 11.5% and 10.8%, while HIV prevalence estimates for urban areas were 25.4% and 23.2% respectively. Further, aggregated weighted national HIV prevalence were 16.9% (ANC-

HIV-SS) and 15.6% (DHS). The findings by Dzekedzeke & Fylkesnes (2006) findings bolster confidence in the current study findings, but it is possible that HIV prevalence estimates from the two methods might not be similar over the fully considered period of my study (i.e., 1994 through 2011).[40]

We cannot guarantee accuracy in the covariate measurement, although strategies were place to minimize occurrence of measurement error. Therefore, it is possible that error in this study as in most research studies might be attributed mis-measurement of study variables.[358] All variables used in the present study, except HIV serostatus, were captured via self-report, and as with self-report data, the findings of this study are valid to the extent that self-reported variables (i.e., age, parity, and educational attainment) are unbiased.[157] Because the key focus of the current analyses was on trends in the odds of prevalent HIV infections over time, and based on the assumption that self-report bias is constant in the 7 survey rounds over the 17-year period, the substantive conclusion will remain unaltered. .Further, the accuracy of data on the antenatal record card from where data were abstracted in 2011 survey cannot be guaranteed.

Because primary data collection was achieved via cross-sectional observational study design, causal inferences cannot be made with confidence. Although large sample enables high statistical power for detecting differences in age distribution between HIV seropositive and HIV seronegative pregnant women, statistically significant p-values for Wilcoxon rank sum test assessing difference in age distribution of HIV seronegative and HIV seropositive pregnant may not be of public-health relevance because the large

sample size enables detection of even very slight differences. The assumption used in the present study that pregnant women in various birth cohorts were not differentially affected by emigration and immigration cannot be confirmed given non-availability of data on migration patterns.

The current analyses did not account include period-level and cohort-level covariates due to non-availability of data to enable assessment of the influence of period-level and cohort-level factors on prevalent HIV infections. Therefore, cross-level statistical interactions assessment could not be conducted. For example, HIV-related deaths were probably higher in earlier birth-cohort, before the introduction of cART compared to recent birth cohort. Along the same line, data on some variables that may influence prevalent HIV infection are not routinely collected in ANC-HIV-SS data [e.g., educational attainment at first pregnancy; number of life time sexual partners]. Therefore, residual confounding due to uncontrolled factors is a possibility because of non-adjustment for some factors (e.g., socioeconomic status).

The number of period groups and birth cohort groups were fewer because there were only seven survey rounds, and age of pregnant women ranged from 15 to 44 years (i.e., nine five-year birth cohorts). Because some studies with similar number of period groups and birth cohorts groups have yielded informative results, and given our large sample size, I proceeded with the analysis. Further, data for the current analysis spanned a relatively short time period of 18years, than would be adequate in most conventional age-period-cohort analyses. However, our study demonstrated the application of CCREM-based APC analysis to HIV surveillance data, and revealed insightful

information such as a suggestion of falling odds of HIV prevalence in recent birth cohort compared to older birth cohorts.

Birth cohorts were created via categorization of continuous variables (i.e., birth year) into nine birth cohort groups. Inaccurate age values could imply that pregnant women were grouped in the wrong birth cohort. Information bias from non-differential misclassification of pregnant women in birth cohort is possible, although the direction of bias would be difficult to predict because there are more than two categories of birth cohorts.[204] Whereas non-differential misclassification biases the measure of association is diluted towards the null when an association exist, non-differential bias when there are more than two categories may be away from the null or towards the null value (e.g. OR=1.0).[204]

Improvement in the HIV testing methods is unlikely to have influenced substantively influenced the estimated trends in odds of prevalent HIV infection over the considered period (i.e., 1994 through 2011). The assays used had specificities and sensitivities greater than 98%. Reassuringly, stringent laboratory testing algorithms were applied in all HIV testing across the seven survey years. Further, given the high background HIV prevalence in this population, HIV seropositive specimens were less likely to be missed compared to environments where the background HIV prevalence is low, given high sensitivities and specificities of HIV assay used,

7.26. Strengths of the study

Because nearly all pregnant women who attended antenatal care clinics during the survey period at sentinel sites were included in the survey sample, the validity of the

ANC-HIV-SS data is minimally threatened by non-response bias, which is of great concern when the proportion of pregnant women who refused to participate in the survey in high, and a major threat to validity for population-based surveys.[36, 204, 359] For example, estimated parameters might be biased if some women refused to participate, resulting in difference in the distribution of characteristics between pregnant women who participated and those who did not participate in the study.[204]

HIV serostatus was objectively determined using commercial diagnostic HIV test kits, and a stringent HIV testing algorithm was implemented that screened-out false HIV seropositive, and false HIV seronegative specimens, thereby limiting HIV serostatus misclassification.[157, 360] The stringent HIV testing criteria provide confidence in the HIV testing strategy. Therefore, because there are two categories [i.e., HIV seropositive and HIV seronegative], non-differential misclassification of HIV serostatus will diminish the strength of the exposure-disease association.[361] Note that non-differential misclassification will not always bias parameter estimates towards the null when the covariate has more than three categories [e.g., single, married, divorced and separated].[62, 204, 361]

The study used a novel methodological approach introduced by Yang and Land (2006) to examine age, period and birth cohort effects on HIV prevalence, which does require using non-defensible constraints for estimating unique parameter estimates. The methodology used in this study may serve as an additional armamentarium of potential tools for understanding the huge HIV epidemic in SSA. Further, considerable proportion of data were missing for educational attainment in 1994 but it was reassuring that

proportions of missing data on other covariates were less than 5%, and sensitivity analysis using multiply imputed data yielded congruent parameter estimates.

Based on CCREM-based APC analysis, there is a suggestion of falling odds of prevalent HIV infection in recent birth cohorts compared to expected odds of prevalent HIV infection in the population. Further examination of HIV prevalence by birth cohort revealed overall declining HIV prevalence for all birth cohorts but increasing HIV prevalence in the 1975-1979, 1985-189, 1990-1996 birth cohorts, suggesting increasing prevalence due to continued occurrence of new HIV as women age as indicated in Figure 6.

Prior work suggest targeted HIV preventive and treatment interventions may curb HIV spread, although well-spread preventive interventions are well-suited and more efficient in generalized epidemic. Although the overall prevalence estimates indicate falling HIV prevalence in the 15 to 24 year-olds, the current analysis provides an early warning of potentially upward swing HIV prevalence within 1985-1989 and 1990-1996 birth cohorts. Therefore, intensifying HIV preventive interventions to curb the suggested upward swing is an imperative public strategy. Analysis of HIV prevalence by period with age groups without adjusting for age and birth cohort group affects revealed a distinct decline in HIV prevalence in the 15 to 24 age group.

Because the study sample was limited to pregnant women aged 15 to 44 years who attended antenatal care during the ANC-HIV-SS during the survey period in 1994, 1998, 2002, 2004, 2006, 2008 and 2011, this sample of pregnant women may not be representative of all women in Zambia in the reproductive age group. Generalizability of

study findings to all pregnant women might be enhanced based on the following reasons. First, Ministry of Health in Zambia has reported that at least 95% of the pregnant women attend antenatal clinic at least once during any pregnancy. Second, sample size of the current study was large (i.e., 82,086). However, even though prior studies have demonstrated comparable HIV prevalence trends, generalization of the findings from this study should be made with caution. Third, pregnant women were drawn from wide geographic area, and widespread social, cultural and economic background, improving generalizability.

The estimates of age, period and cohort effects may be sensitive to the model selected for parameter estimation. Although other methods have not been used to assess the age-period-cohort effects using ANC-HIV-SS data, Yang and Land (2006) CCREM-based APC method seem well-suited to assessing age, period, and cohort effects on HIV prevalence using ANC-HIV-SS. Further evaluation of age, period and birth cohort effects using population-based data from DHS will help improve understanding of age, period and birth cohort effects. The emergent data sources from DHS which are repeated cross-sectional surveys provides excellent opportunities for using CCREM-based APC analysis for assessment of age, period and cohort effects, but longitudinal data remain superior in providing conclusive causal evidence regarding age, period and cohort effects; sadly fewer longitudinal data exist.

7.27. Conclusion

In conclusion, pregnant women in recent birth cohorts (i.e., younger women) appear to have reduced odds of prevalent HIV infection compared to pregnant women in

older birth cohorts (i.e., older pregnant women). For nearly all birth cohorts examined HIV prevalence within all birth cohorts starts off low and increase gradually. Therefore, intensive efforts should be aimed at increasing safer sexual behaviors are required prior to sexual activity stage. The increased HIV prevalence in most recent birth cohort persists may suggest increasing number of new HIV infections as well as improved survival. Further, my study shows how novel methods can be applied to existing databases to provide key public health information that can be used for guiding intervention, and monitoring the direction of the HIV epidemic. Further investigations into the factors related to age, period and birth cohorts that drive the HIV epidemic could yield critical results furthering understanding of the HIV epidemic.

Table 7.1. Characteristics of the pregnant women [15 to 44 year-olds] with and without evidence of HIV infection (i.e., HIV seropositive or HIV seronegative) recruited in the ANC-HIV-SS in conducted Zambia, from 1994 through 2011

	Combined (N=82086)		Seropositive (n=15505)		Seronegative (n=66581)		P-value
	Median	IQR	Median	IQR†	Median	IQR	
Age	24	20 to 29	25	22 to 29	23	20 to 29	0.001
Missing							
15—19	17562	21.4	1954	12.6	15608	23.4	0.001
20—24	27121	33.0	5117	33.0	22004	33.1	
25—29	18975	23.1	4626	23.1	14349	21.6	
30—34	11289	13.8	2569	13.8	8720	13.1	
35—39	5648	6.9	1032	6.7	4616	6.9	
40—44	1491	1.8	207	1.3	1284	1.9	
Educational attainment*							
School-years†	7	5 to 9	8	7 to 9	7	5 to 9	0.001
Missing∞							
	n	%	n	%	n	%	
Residence							
Rural	34686	42.3	3905	25.2	30781	46.2	0.001
Urban	47400	57.7	11600	74.8	35800	53.8	
Period							
1994	9724	11.9	1981	12.8	7743	11.6	0.001
1998	11718	14.3	2296	14.8	9422	14.2	
2002	12838	15.6	2559	16.5	10279	15.4	
2004	12404	15.1	2407	15.5	9997	15.0	
2006	13223	16.1	2348	15.1	10875	16.3	
2008	13298	16.2	2403	15.5	10895	16.4	
2011	8881	10.8	1511	9.8	7370	11.1	
Birth Cohort							
1950— 1954	198	0.2	15	0.1	183	0.3	0.001
1955— 1959	983	1.2	137	0.9	846	1.3	
1960— 1964	2948	3.6	490	3.2	2458	3.7	
1965— 1969	6157	7.5	1337	8.6	4820	7.2	
1970— 1974	13162	16.0	3102	20.0	10060	15.1	
1975— 1979	18476	22.5	4259	27.5	14217	21.4	
1980— 1984	21046	25.6	3868	25.0	17178	25.8	
1985— 1989	14022	17.1	1839	11.9	12183	18.3	
1990— 1996‡	5094	6.2	458	3.0	4636	7.0	

†IQR-interquartile range; ‡-1990-1994 and 1995-1996 birth cohorts coalesced because of fewer observations
∞Missing number include 2011 missing data for educational attainment which was not collected: see Table 3 for year specific break down of missing data; *As explained earlier, educational attainment data for 1998 and 2002 surveys was a recorded in a mixed manner [i.e., zero to 12 schooling years [i.e., continuous] and beyond 12 school-years was recorded as "greater than 12 school-years" [i.e., categorical]. To make the analyses comparable across all survey years, educational attainment greater than 12 years was recorded as 12 years for this table, suffice to say that this is less attractive, and imperfect approach. However, there were comparatively fewer pregnant women with educational attainment beyond 12 years. The median and IQR values for school truncated at 12 years and non-truncated remained unaltered; Although not to encourage use of OR as statistical test, an odds ratio confidence interval that does not include OR=1.0 is equivalent to a statistically significant association.

Table 7.2. HIV-1 prevalence by selected characteristics among pregnant women age 15 to 44 years surveyed during the ANC-HIV-SS in Zambia 1994 through 2011

Variable	1994	1998	2002	2004	2006	2008	2011
	n HIV% [95% CI]	n HIV% [95% CI]	n HIV% [95% CI]	n HIV% [95% CI]	n HIV% [95% CI]	n HIV% [95% CI]	n HIV% [95% CI]
Age-group [years]*							
15-24	5524 19.7 (18.7-20.8)	6996 17.8 (16.9-18.7)	7413 17.3 (16.5-18.2)	6865 16.4 (15.5-17.3)	7046 13.8 (13.0-14.6)	6631 13.2 (12.4-14.1)	4208 11.5 (10.5-12.5)
25-34	3317 23.6 (22.2-25.1)	3862 24.3 (22.9-25.6)	4426 25.4 (24.2-26.7)	4551 24.5 (23.3-25.8)	4999 23.3 (22.2-24.5)	5417 23.5 (22.4-24.6)	3692 21.6 (20.3-22.9)
35-44	883 12.5 (10.4-14.8)	860 13.3 (11.2-15.7)	999 15.1 (13.0-17.5)	988 17.0 (14.8-19.5)	1178 17.9 (15.8-20.2)	1250 20.3 (18.2-22.6)	981 23.5 (21.0-26.3)
Missing	—	—	—	—	—	—	—
Residence							
Rural	4234 10.9 (10.0-11.9)	4695 10.6 (9.8-11.5)	5439 12.4 (11.6-13.3)	5237 11.8 (11.0-12.7)	5676 10.4 (9.6-11.2)	5362 11.1 (10.3-12.0)	4043 11.5 (10.5-12.5)
Urban	5490 27.7 (26.5-28.9)	7023 25.6 (24.6-26.6)	7399 25.4 (24.5-26.5)	7167 25.0 (24.0-26.0)	7547 23.3 (22.3-24.2)	7936 22.8 (21.9-23.7)	4838 21.6 (20.5-22.8)
Educational attainment							
0-4	4052 18.8 (17.7-20.1)	5387 17.9 (16.9-18.9)	5455 17.9 (16.9-18.9)	5029 18.1 (17.1-19.2)	4855 16.6 (15.6-17.7)	4752 16.8 (15.8-17.9)	NC
5-7	1832 10.9 (9.6-12.4)	2412 13.2 (11.9-14.6)	2517 12.4 (11.1-13.7)	2280 13.2 (11.9-14.7)	2035 11.8 (10.5-13.3)	1786 13.5 (12.0-15.2)	NC
8-9	544, 36.4 (32.5-40.5)	549 31.9 (28.1-35.9)	648 25.6 (22.4-29.1)	708 24.7 (21.7-28.0)	872 21.2 (18.6-24.1)	930 21.7 (19.2-24.5)	NC
10-11	1706 27.4 (25.3-29.5)	2458 23.9 (22.2-25.6)	2938 25.4 (23.9-27.0)	2937 22.7 (21.3-24.3)	3210 20.4 (19.0-21.8)	3454 19.4 (18.1-20.7)	NC
12-17	573 39.3 (35.4-43.3)	912 27.6 (24.8-30.6)	1280 28.3 (25.9-30.8)	1395 24.4 (22.3-26.8)	2009 20.9 (19.2-22.7)	2296, 20.9 (19.2-22.6)	NC
Missing	1014	—	—	51	236	78	8881**

Variable	1994	1998	2002	2004	2006	2008	2011
Parity (i.e., number of children birthed pregnant women)							
0	2430 19.8 (18.3-21.4)	3471 16.2 (15.0-17.5)	3763 17.6 (16.4-18.8)	3637 15.7 (14.6-16.9)	4009 13.4 (12.4-14.5)	3576 12.8 (11.8-13.9)	2322 13.4 (12.0-14.8)
1	2010 24.7 (22.9-26.7)	2556 24.7 (23.1-26.4)	2881 22.0 (20.5-23.5)	2725 22.9 (21.3-24.5)	2918 20.8 (19.4-22.3)	2980 19.6 (18.2-21.1)	1709 19.5 (17.7-21.4)
≥2	5021 18.7 (17.7-19.8)	5691 19.4 (18.4-20.4)	6194 20.4 (19.4-21.4)	6032 20.0 (19.1-21.1)	6292 19.1 (18.2-20.1)	6742 20.2 (19.2-21.1)	3484 21.8 (20.4-23.2)
Missing	263	0	0	10	4	0	1366

*Pregnant women with missing age were excluded based the inclusion criteria [i.e. 15 to 44 years]
NC — Data not collected in specific survey year, and **Educational attainment data not collected in 2011

Table 7.3. HIV prevalence among pregnant women 15-44 years olds within birth cohorts [i.e., 1950-1954 through 1990-1996 birth cohort] using ANC-HIV-SS-based HIV prevalence data collected between 1994 through 2011

Birth Cohort	Period [Calendar year of the survey]						
	1994	1998	2002	2004	2006	2008	2011
	n HIV% (95% CI)‡	n HIV% (95% CI)	n HIV% (95% CI)	n HIV% (95% CI)	n HIV% (95% CI)	n HIV% (95% CI)	n HIV% (95% CI)
1950-1954	189 7.4 (4.5-12.0)	9 11.1 (0.6-43.5)	—	—	—	—	—
1955-1959	694 13.8 (11.5-16.6)	249 15.3 (11.3-20.3)	40 7.5 (2.6-19.9)	—	—	—	—
1960-1964	1357 20.1 (18.1-22.3)	836 14.2 (12.0-16.8)	406 13.8 (10.8-17.5)	239 10.5 (7.2-15.0)	84 16.7 (10.2-26.1)	26 11.5 (4.0-29.0)	—
1965-1969	1960 26.0 (24.1-28.0)	1470 22.2 (20.1-24.4)	955 18.3 (16.0-20.9)	749 19.1 (16.4-22.1)	589 16.8 (14.0-20.0)	347 18.2 (14.5-22.6)	87 25.3 (17.3-35.3)
1970-1974	3301 23.3 (21.9-24.8)	2895 25.5 (23.9-27.1)	2118 24.6 (22.9-26.5)	1717 22.6 (20.7-24.6)	1434 21.6 (19.6-23.8)	1190 22.5 (20.2-25.0)	507 20.9 (17.6-24.7)
1975-1979	2223 14.3 (13.0-15.9)	4156 20.7 (19.5-22.0)	3334 25.8 (24.3-27.3)	2834 25.7 (24.1-27.3)	2568 24.4 (22.8-26.1)	2219 25.4 (23.6-27.3)	1142 26.2 (23.7-28.8)
1980-1984	—	2103 10.1 (8.9-11.4)	4713 17.3 (16.3-18.4)	4468 18.7 (17.6-19.8)	4041 20.0 (18.8-21.3)	3696 21.1 (19.8-22.4)	2025 20.6 (18.9-22.4)
1985-1989	—	—	1272 10.0 (8.5-11.8)	2397 12.1 (10.8-13.4)	3904 11.2 (10.3-12.2)	4113 14.4 (13.3-15.5)	2336 16.9 (15.4-18.4)
1990-1996†	—	—	—	—	603 8.5 (6.5-11.0)	1707 7.9 (6.7-9.2)	2513 10.0 (8.9-11.2)

†Include 1990-1996 and 1994-1996 birth-cohorts; merged due to fewer observations
—No pregnant women in the category, ‡ Within-birth cohort estimated HIV prevalence with corresponding 95% Wilson confidence interval. Pregnant women are cross-classified and nested within period [survey calendar year] and birth-cohort thereby creating a multilevel data structure: individual-level covariates are level-1 and period and birth-cohort variables are level-2. Note the wide confidence that characterize estimates with cells with fewer observations, reflecting uncertainty in estimated values, and highlighting the importance of confidence interval reporting. Graphical impression of this Table 6 is in Figure 4

Table 7.4. Estimated log-odds and standard errors, as well as the corresponding odds ratio and 95% confidence interval for CCREM-APC adjusted for age and residence using data from ANC-HIV-SS in Zambia, 1994 through 2011

	Model #1	Model #2	Model #3

	β	se(β)	β ([se(β])	OR,95% CI	β [se(β)]	OR,95% CI
Fixed effects	-1.6578	0.1584				
Age						
15	—	—	—	0.34 (0.33-0.35)	—	0.36 (0.35-0.37)
19	—	—	—	0.63 (0.62-0.64)	—	0.64 (0.63-0.66)
26	—	—	—	1.05 (1.02-1.08)	—	1.05 (1.02-1.09)
29	—	—	—	1.04 (0.93-1.17)	—	1.06 (0.95-1.19)
34	—	—	—	0.89 (0.66-1.19)	—	0.96 (0.71-1.28)
39	—	—	—	0.70 (0.44-1.11)	—	0.81 (0.51-1.28)
Residence						
Rural [Reference]	—	—	—	—		1.0
urban	—	—	—	—	0.930 (0.02)	2.53 (2.44-2.64)
Random effects			β ([se(β*])	OR, 95% CI	β [se(β*)]	OR, 95% CI
Period						
1994	-0.0906	0.0470	0.026 (0.021)	1.03 (0.99-1.07)	0.047 (0.026)	1.05 (1.00-1.10)
1998	-0.1427	0.0456	-0.016 (0.020)	0.98 (0.95-1.02)	-0.031 (0.024)	0.97 (0.92-1.02)
2002	-0.0232	0.0450	0.021 (0.019)	1.02 (0.98-1.06)	0.030 (0.024)	1.03 (0.98-1.08)
2004	-0.0002	0.0452	0.002 (0.019)	1.00 (0.96-1.04)	0.003 (0.024)	1.00 (0.96-1.05)
2006	-0.0121	0.0453	-0.032 (0.019)	0.97 (0.93-1.01)	-0.041 (0.024)	0.96 (0.92-1.01)
2008	0.0949	0.0454	-0.003 (0.020)	1.00 (0.96-1.04)	-0.022 (0.024)	0.98 (0.93-1.03)
2011	0.1753	0.0482	0.002 (0.021)	1.00 (0.96-1.05)	0.014 (0.027)	1.01 (0.96-1.07)
Birth cohort						
1950-1954	-0.5652	0.2205	-0.300 (0.163)	0.74 (0.54-1.02)	-0.323 (0.168)	0.72 (0.52-1.01)
1955-1959	-0.0595	0.0995	-0.120 (0.086)	0.89 (0.75-1.05)	-0.180 (0.088)	0.84 (0.70-0.99)
1960-1964	0.1276	0.0641	-0.049 (0.050)	0.95 (0.86-1.05)	-0.058 (0.052)	0.94 (0.85-1.04)
1965-1969	0.4319	0.0508	0.206 (0.033)	1.23 (1.15-1.31)	0.214 (0.035)	1.24 (1.16-1.33)
1970-1974	0.5222	0.0451	0.292 (0.023)	1.34 (1.28-1.40)	0.314 (0.026)	1.37 (1.30-1.44)
1975-1979	0.4762	0.0436	0.308 (0.020)	1.36 (1.31-1.42)	0.320 (0.023)	1.38 (1.32-1.44)
1980-1984	0.1528	0.0440	0.073 (0.021)	1.08 (1.03-1.12)	0.087 (0.024)	1.09 (1.04-1.14)
1985-1989	-0.2857	0.0477	-0.124 (0.028)	0.88 (0.84-0.93)	-0.112 (0.030)	0.89 (0.84-0.95)
1990-1996	-0.7756	0.0645	-0.276 (0.050)	0.76 (0.69-0.84)	-0.249 (0.052)	0.78 (0.70-0.86)

† Merged due to fewer observations in 1990-1996 and 1994-1996 birth-cohorts. Model #2 was age-adjusted and Model #3 was adjusted with age and location of site [proxy for pregnant woman's residence] Age was fitted using restricted cubic splines [RCS] function with 4 pre-specified knots

Table 6 shows odds ratio estimates and corresponding 95% confidence interval [CI]. Null hypothesis for this analysis was that there is no age, period or cohort effect. Distinct odds of prevalent HIV infection in 1970-1974 and 1975-1979 birth-cohorts.

Table 7.5. Odds ratio and corresponding 95% confidence intervals, and log-odds and standard errors for the logit cross-classified random effect age-period and cohort model of prevalent HIV infection

	Model #4		Model #5†		Model #6‡	
	β ([se(β])	OR,95% CI	β ([se(β])	OR,95% CI	β [se(β)]	OR,95% CI
Fixed effects					—	
Age						

	Model #4		Model #5†		Model #6‡	
	β ([se(β])	OR,95% CI	β ([se(β])	OR,95% CI	β [se(β)]	OR,95% CI
15	—	0.31 (0.31-0.32)	—	0.31 (0.3-0.33)	—	0.32 (0.31-0.33)
19	—	0.58 (0.57-0.59)	—	0.58 (0.56-0.60)	—	0.58 (0.56-0.60)
26	—	1.11 (1.07-1.15)	—	1.11 (1.05-1.16)	—	1.11 (1.05-1.17)
29	—	1.16 (1.02-1.32)	—	1.16 (0.96-1.40)	—	1.16 (0.97-1.41)
34	—	1.04 (0.75-1.45)	—	1.04 (0.64-1.68)	—	1.04 (0.64-1.69)
39	—	0.85 (0.51-1.43)	—	0.84 (0.39-1.79)	—	0.84 (0.39-1.80)
Educational attainment [number of school-years completed						
0	—	0.72 (0.72-0.73)	—	0.72 (0.72-0.73)	—	0.72 (0.72-0.73)
4	—	0.76 (0.76-0.76)	—	0.76 (0.76-0.77)	—	0.77 (0.76-0.77)
9	—	1.3 (1.28-1.31)	—	1.29 (1.27-1.32)	—	1.29 (1.27-1.31)
11	—	1.17 (1.11-1.24)	—	1.17 (1.08-1.28)	—	1.19 (1.11-1.28)
12	—	1.02 (0.94-1.11)	—	1.02 (0.90-1.16)	—	1.05 (0.94-1.18)
Residence						
Rural [Reference]	Reference	1.0	Reference	1.0	Reference	1.0
Urban	0.823 (0.023)	2.28 (2.18-2.30)	0.830 (0.033)	2.40 (2.15-2.40)	0.829 (0.033)	2.40 (2.14-2.40)
Parity [Number of children]						
0 [Reference]	Reference	1.0	Reference	1.0	Reference	1.0
1	0.083 (0.031)	1.09 (1.02-1.15)	0.080 (0.046)	1.20 (0.99-1.2)	0.084 (0.046)	1.20 (0.99-1.20)
≥2	-0.235 (0.035)	0.79 (0.74 -0.85)	-0.241 (0.052)	0.90 (0.71-0.9)	-0.240 (0.052)	0.90 (0.71-0.90)
Random effects			β ([se(β*])	OR, 95% CI	β [se(β*)]	OR, 95% CI
Period						
1994	0.064 (0.031)	1.07 (1.00-1.13)	0.046 (0.028)	1.05 (0.99-1.11)	0.045 (0.028)	1.05 (0.99-1.11)
1998	-0.019 (0.029)	0.98 (0.93-1.04)	-0.018 (0.027)	0.98 (0.93-1.03)	-0.019 (0.026)	0.98 (0.93-1.03)
2002	0.044 (0.028)	1.04 (0.99-1.10)	0.044 (0.026)	1.04 (0.99-1.10)	0.042 (0.026)	1.04 (0.99-1.10)
2004	0.007 (0.028)	1.01 (0.95-1.06)	0.010 (0.026)	1.01 (0.96-1.06)	0.010 (0.026)	1.01 (0.96-1.06)

	Model #4		Model #5†		Model #6‡	
	β ([se(β])	OR,95% CI	β ([se(β])	OR,95% CI	β [se(β)]	OR,95% CI
2006	-0.054 (0.029)	0.95 (0.90-1.00)	-0.048 (0.026)	0.95 (0.91-1.00)	-0.048 (0.026)	0.95 (0.91-1.00)
2008	-0.040 (0.029)	0.96 (0.91-1.02)	-0.033 (0.027)	0.97 (0.92-1.02)	-0.029 (0.026)	0.97 (0.92-1.02)
2011	NC	NC	NC	NC	NC	NC
Birth cohort						
1950-1954	-0.158 (0.157)	0.85 (0.63-1.16)	-0.189 (0.157)	0.83 (0.61-1.13)	-0.190 (0.158)	0.83 (0.61-1.13)
1955-1959	-0.111 (0.089)	0.90 (0.75-1.07)	-0.097 (0.087)	0.91 (0.77-1.08)	-0.097 (0.087)	0.91 (0.77-1.08)
1960-1964	-0.033 (0.054)	0.97 (0.87-1.08)	-0.018 (0.053)	0.98 (0.89-1.09)	-0.017 (0.053)	0.98 (0.89-1.09)
1965-1969	0.196 (0.038)	1.22 (1.13-1.31)	0.207 (0.037)	1.23 (1.14-1.32)	0.208 (0.036)	1.23 (1.15-1.32)
1970-1974	0.285 (0.030)	1.33 (1.25-1.41)	0.289 (0.028)	1.34 (1.26-1.41)	0.291 (0.028)	1.34 (1.27-1.41)
1975-1979	0.239 (0.027)	1.27 (1.20-1.34)	0.241 (0.026)	1.27 (1.21-1.34)	0.244 (0.025)	1.28 (1.21-1.34)
1980-1984	0.019 (0.028)	1.02 (0.96-1.08)	0.018 (0.027)	1.02 (0.97-1.07)	0.020 (0.026)	1.02 (0.97-1.07)
1985-1989	-0.175 (0.036)	0.84 (0.78-0.90)	-0.180 (0.034)	0.84 (0.78-0.89)	-0.181 (0.034)	0.83 (0.78-0.89)
1990-1996*	-0.253 (0.075)	0.78 (0.67-0.90)	-0.261 (0.074)	0.77 (0.67-0.89)	-0.266 (0.074)	0.77 (0.66-0.89)

*Merged due to fewer observations in 1990-1996 and 1994-1996 birth-cohorts. Model #2 was age-adjusted and Model #3 was adjusted with age and location of site [proxy for pregnant woman's residence] Age was fitted using restricted cubic splines [RCS] function with 4 pre-specified knots
†Model #4 but educational attainment was truncated at 12 years, with pregnant women having greater than 12 years of education regarded as 12 years.
‡Model #5 was fitted to multiply imputed data, adjusted for age, parity, education and residence. Educational attainment was truncated at 17 years for model #5, where greater than 17 were regarded as 17 years. Truncation though imperfect facilitated comparison across year. Because fewer observations beyond 12 years, there was no meaning differences between 12-year and 17-year truncation. Ten multiply imputed data sets were created.
NC-data not collected Table 9 shows odds ratio estimates and corresponding 95% confidence interval [CI]. The null hypothesis for this analysis is that there is no age, period or cohort effect: indicated by uniform coefficients within period and cohort groups. The odds of prevalent HIV infection were most distinct in the 1970-1974 and 1975-1979 birth-cohorts. Although not exactly the same values, the estimates from the seven models are not materially different.

Table 7.6. Urban-rural stratified odds ratio and corresponding 95% CI, and log-odds and standard errors for the logit cross-classified random effect age-period and cohort model of prevalent HIV infection

	Age-only adjusted model†		Fully-adjusted-models†	
	Rural (Model #7)	Urban (Model #8)	Rural (Model # 9)	Urban (Model#10)
Fixed effects				

Age	β ([se(β])	OR,95% CI	β [se(β)]	OR,95% CI	β ([se(β])	OR,95% CI	β [se(β)]	OR,95% CI
15	—	0.31 (0.29-0.33)		0.32 (0.31-0.33)	—	0.35 (0.33-0.36)	—	0.38 (0.36-0.41)
19	—	0.57 (0.53-0.61)		0.59 (0.57-0.61)	—	0.64 (0.62-0.67)	—	0.65 (0.61-0.69)
26	—	1.13 (1.02-1.25)		1.09 (1.03-1.16)	—	1.04 (0.99-1.11)	—	1.07 (0.97-1.17)
39	—	1.21 (0.83-1.75)		1.13 (0.9-1.41)	—	1.04 (0.84-1.28)	—	1.09 (0.79-1.52)
34	—	1.09 (0.43-2.79)		0.98 (0.55-1.74)	—	0.91 (0.54-1.55)	—	1.00 (0.43-2.30)
39		0.89 (0.2-3.87)		0.77 (0.31-1.92)	—	0.75 (0.33-1.73)	—	0.85 (0.23-3.16)
0	—	—	—	—		0.70 (0.70-0.71)	—	0.76 (0.75-0.77)
4	—					0.73 (0.73-0.74)		0.79 (0.79-0.8)
9	—	—	—	—		1.40 (1.36-1.45)	—	1.24 (1.22-1.27)
11	—	—	—	—		1.52 (1.31-1.76)	—	1.13 (1.03-1.23)
12	—	—	—	—		1.47 (1.18-1.84)	—	0.99 (0.86-1.14)
Parity (Number of children)								
0	—	—	—	—			1.0	
1	—	—	—	—		0.00 (0.093)	1.00 (0.83-1.2)	0.104 (0.053)
≥2	—	—	—	—		-0.34 (0.103)	0.71 (0.58-0.9)	-0.212 (0.060)
Period								
	β* ([se(β*])	OR, 95% CI	β* [se(β*)]	OR, 95% CI	β* ([se(β*])	OR, 95% CI	β* [se(β*)]	OR, 95% CI
1994	-0.004 (0.039)	1.00 (0.92-1.08)	0.031 (0.022)	1.03 (0.99-1.08)	0.008 (0.045)	1.01 (0.92-1.10)	0.022 (0.019)	1.02 (0.98-1.06)
1998	-0.046 (0.038)	0.96 (0.89-1.03)	-0.019 (0.021)	0.98 (0.94-1.02)	-0.039 (0.044)	0.96 (0.88-1.05)	-0.011 (0.018)	0.99 (0.95-1.03)
2002	0.061 (0.036)	1.06 (0.99-1.14)	0.009 (0.020)	1.01 (0.97-1.05)	0.092 (0.041)	1.10 (1.01-1.19)	0.010 (0.018)	1.01 (0.97-1.05)
2004	0.016 (0.037)	1.02 (0.95-1.09)	0.001 (0.020)	1.00 (0.96-1.04)	0.037 (0.042)	1.04 (0.96-1.13)	0.002 (0.018)	1.00 (0.97-1.04)
2006	-0.052 (0.037)	0.95 (0.88-1.02)	-0.016 (0.020)	0.98 (0.95-1.02)	-0.075 (0.042)	0.93 (0.85-1.01)	-0.015 (0.018)	0.99 (0.95-1.02)
2008	-0.004 (0.037)	1.00 (0.93-1.07)	-0.012 (0.020)	0.99 (0.95-1.03)	-0.021 (0.043)	0.98 (0.90-1.06)	-0.010 (0.018)	0.99 (0.96-1.03)
2011	0.031 (0.039)	1.03 (0.96-1.11)	0.005 (0.022)	1.01 (0.96-1.05)	NC	NC	NC	NC
Birth cohort								
1950-1954	-0.089 (0.131)	0.91 (0.71-1.18)	-0.248 (0.198)	0.78 (0.53-1.15)	-0.042 (0.115)	0.96 (0.77-1.20)	-0.213 (0.203)	0.81 (0.54-1.20)
1955-1959	-0.081 (0.107)	0.92 (0.75-1.14)	-0.167 (0.104)	0.85 (0.69-1.04)	-0.023 (0.098)	0.98 (0.81-1.18)	-0.130 (0.104)	0.88 (0.72-1.08)
1960-1964	-0.125 (0.077)	0.88 (0.76-1.03)	-0.012 (0.061)	0.99 (0.88-1.11)	-0.071 (0.074)	0.93 (0.81-1.08)	0.023 (0.061)	1.02 (0.91-1.15)
1965-1969	0.001 (0.057)	1.00 (0.89-1.12)	0.286 (0.039)	1.33 (1.23-1.44)	0.015 (0.058)	1.02 (0.91-1.14)	0.306 (0.039)	1.36 (1.26-1.46)
1970-1974	0.152 (0.042)	1.16 (1.07-1.26)	0.346 (0.027)	1.41 (1.34-1.49)	0.138 (0.045)	1.15 (1.05-1.25)	0.356 (0.027)	1.43 (1.36-1.50)

1975-1979	0.246 (0.038)	1.28 (1.19-1.38)	0.301 (0.023)	1.35 (1.29-1.41)	0.190 (0.041)	1.21 (1.12-1.31)	0.283 (0.023)	1.33 (1.27-1.39)
1980-1984	0.060 (0.039)	1.06 (0.98-1.15)	0.045 (0.024)	1.05 (1.00-1.10)	-0.008 (0.043)	0.99 (0.91-1.08)	0.018 (0.023)	1.02 (0.97-1.07)
1985-1989	-0.033 (0.049)	0.97 (0.88-1.06)	-0.197 (0.032)	0.82 (0.77-0.87)	-0.132 (0.054)	0.88 (0.79-0.97)	-0.240 (0.032)	0.79 (0.74-0.84)
1990-1996*	-0.120 (0.076)	0.89 (0.76-1.03)	-0.343 (0.060)	0.71 (0.63-0.80)	-0.061 (0.089)	0.94 (0.79-1.12)	-0.391 (0.060)	0.68 (0.60-0.76)

†Estimates for model #9 and Model #10 were based on multiply 10 imputed data sets.
Model # 7 and Model #8 were adjusted for age, parity and educational attainment (i.e., measured as number of schooling years)
Model #9 and Model # 10 were age-only adjusted models

CHAPTER 8

ASSOCIATION BETWEEN EDUCATIONAL ATTAINMENT AND PREVALENT HIV INFECTION AMONG PREGNANT WOMEN

8.1. Background

Many factors have been explored to better understand what drives the serious HIV epidemic in sub-Saharan Africa (SSA), including number sexual partners, background HIV prevalence, condom use, partner mixing, viral load and sub-type factors, co-infections, and HIV stigma.[56] Most of the HIV epidemic in SSA is in the context of serve economic challenges at both the personal and governmental level.[10, 11, 39, 41, 48, 50, 52, 362, 363] One factor that has not been clarified is the role of educational attainment in HIV risk, a topic that I chose to study using a unique antenatal surveillance database in Zambia.[52, 81, 157]

Identifying factors associated with increased odds of HIV incidence and prevalence provides key information for development and implementation of preventive HIV interventions.[20, 29, 107] Admittedly, population-wide HIV preventive interventions are cost effective in countries where the HIV epidemic is generalized, and where resources are resources limited. Therefore, identifying sub-groups with high HIV incidence and HIV prevalence is a key step in tracking the changes in the dynamics of the HIV epidemic, and may provide key information for targeting HIV prevention and treatment program.[164, 363]

Given its importance as key components of one of the social determinants of health, educational attainment has been included as a key component for human

development index (i.e., education, and income, health) and as a benchmark for the Millennium Development Goal 2 that seeks universal primary school education.[48, 78, 79] Contrary to the intuitive expectation of greater risk of infectious diseases among the poor, illiterate, and less educated, many studies conducted in SSA have reported higher odds of prevalent HIV infections among more educated persons, particularly in earlier years of the HIV epidemic (i.e., 1980s and early 1990s).[42, 80] Hargreaves and Glynn (2002) highlighted in their systematic review that 20th century studies that had examined the association between educational attainment and prevalent HIV infections reported disparate findings.[51, 81]

The association between education attainment and HIV is complex, and although several studies have examined the association, few have explicitly indicated what educational attainment represents as a study factor (i.e., educational attainment as a proxy for literacy or socioeconomic dimension).[50, 51] Focused on studies conducted in SSA between 1987 and 2003, Hargreaves et al. (2008) conducted a systematic review to understand the relationship between educational attainment and prevalent HIV infection, and noted a shift towards elevated HIV prevalence among the least educated.[81] Although Hargreaves et al. (2008) hypothesized that a positive association between educational attainment and prevalent HIV infection might wane as the HIV epidemic progressed, inconsistent reports on the association between educational attainment and prevalent HIV infection have continued to appear in literature.[81, 196]

Recent data indicate a reduction in the proportion of persons with higher educational attainment living with HIV infection among 15 to 24 year-olds in most countries in SSA.[4, 24, 87] Various explanations have been advanced to explain the

noted fall in the odds of prevalent HIV infections among educated people in recent times compared to earlier years of the HIV epidemic, including the view that HIV preventive interventions have differentially benefited educated people than less educated people.[40, 58, 140]

Educated people are likely to have a superior ability to obtain; process and comprehend HIV-related prevention and treatment information; they theoretically might be more likely to adopt safer sexual behavior patterns than less educated people. Further, education can awaken people to judge their environments and may enable them make salutary changes in their life styles.[364] Higher educational attainment may facilitate greater understanding of risk factors for HIV spread that consequently enables behavior change.[127, 140] All of this, however, does not address why a number of studies early in the SSA HIV epidemic reported higher education to be a risk factor, not a protective factor.[50, 51, 58, 81, 140]

The validity of estimates from most research studies that examined association between educational attainment and prevalent HIV infections were threatened by methodological limitations inherent in cross-sectional studies.[52, 55, 56, 141] Most studies were cross-sectionally designed, therefore causal inference may only be suggestive.[204] Like my current study, majority of prior studies relied on data collected for other purposes. Consequently, without a guiding conceptual framework or without a directed acyclic graph (i.e., diagrammatic strategy for presenting expert-knowledge assumptions about a relationship when choosing or identifying potential confounders and identifying potential sources of bias in an epidemiologic analysis) during study design stage, collection of data on salient potential could be missed.[52, 365-371] Further, few

studies have been focused specifically on investigating the association between educational attainment and HIV infection. [52, 80]

Educational attainment does not have a direct biologic effect on risk of HIV infection, but its effects are mediated by other risk factors that are biologically connected to HIV infection. Because educated people are more receptive to health interventions, there are likely to use condoms during sexual intercourse than less educated people. Therefore, I investigated the association between educational attainment and prevalent HIV infection among pregnant women attending antenatal care clinics (ANC) in Zambia varied between 1994 and 2008, accounting for within-site clustering (i.e., data were collected from multiple sites).

8.2. Methods

8.2.1. Study design and population

Data used in the current study were collected in six cross-sectional surveys conducted to monitor HIV prevalence among pregnant women between 1994 and 2008 (i.e.1994, 1998, 2002, 2004, 2006 and 2008). A similar survey in 2011 did not include educational attainment data and was not used. Secondary data analysis of the ANC-HIV-SS data were restricted to 15 to 44 years.[157]

8.2.2. Rationale for using prevalent HIV infections

To limit the influence of survival bias that arise when relationship are examined using prevalent cases, the current analyses were restricted to pregnant women who were

15 to 24 year-olds based on the premise that prevalent HIV infection in 15 to 24 year-old represented incident HIV infections as recommended by UNAIDS.[24] Pregnant women who were 15 to 24 year-old were assumed to have recently commenced sexual intercourse, and only a tiny fraction of whom would be expected to die by their 24th birthday.[185]

Longitudinal studies for HIV incidence conducted for HIV incidence estimation are fewer because of the inherent logistical and technical challenges stereotypic of longitudinal data collection.[372] . Furthermore, the Hawthorne Effect, that persons being studied may change their behaviors, can render a true cohort less valid in estimating community seroincidence.[176, 178, 179] Though progress is being made, here are considerable logistic and cost challenges in the direct measurement of incident HIV infections.[174, 175, 243]

8.2.3. Survey response

Survey response rate was assumed nearly 100% because all eligible pregnant women who attended the antenatal clinic during the four-month survey period were included in the sample used for estimating HIV prevalence. Some eligible pregnant women could have been missed, but the number of pregnant women left out of the survey sample is thought to be very tiny since almost all pregnant women provide blood for routine syphilis testing during their first antenatal care visit. The remnant blood was used for the unlinked anonymous HIV testing survey.[157]

8.3. Methods

8.3.1. Sociodemographic and reproductive history information

To ensure similar data collection procedures and comparable implementation of survey protocol in all the sentinel sites, survey staff were trained on survey methods and procedure during a one-day pre-survey training workshops. Pregnant woman's characteristics (e.g., pregnant woman's educational attainment, age, marital status, and number of children birthed by woman) were captured by self-report via a standard questionnaire administered by a survey nurse. Educational attainment, the primary exposure variable for the current analyses were captured via the following question "How many years did you go to school?". The primary exposure variable was educational attainment measured as pregnant woman's self-reported number of years of schooling completed (i.e., continuous variable). HIV serostatus was the outcome variable, defined as HIV seropositive if HIV specific antibodies were confirmed, otherwise reported as HIV seronegative. Other covariates included in the analysis included age, parity, and marital status. Potential effect modifiers were residence and survey calendar year.

8.3.2. HIV screening, confirmatory and quality control testing

Blood collected for survey HIV reporting were screened for HIV specific antibodies survey specific HIV testing algorithm. The final HIV serostatus were determined according to the three-stage HIV testing algorithm. First, sentinel site based serological testing for HIV specific antibodies (i.e., screening) using rapid HIV test assays and reference laboratory HIV confirmatory testing using ELISA were performed according test assay manufacturer specifications, and with adherence to WHO guidelines

for anonymous and unlinked HIV screening. Specimens with positive HIV results on the rapid HIV screen were subjected to confirmatory testing using ELISA at the reference laboratories: the Tropical Diseases Research Centre (TDRC) in Ndola and the University Teaching Hospital (UTH) Virology Laboratory in Lusaka.

Quality control HIV testing were performed on 5% of the plasma/serum samples in 1994 and 1998 and 10% of the plasma/serum samples in 2002, 2004, 2006, and 2008 that tested HIV seronegative during HIV screening at the site, using an ELISA. A third tie-breaker test (a second ELISA or in 2011, a Western blot (Table 1) was conducted on specimens where screening and confirmatory HIV results were discrepant, and the tie-breaker result reported as final HIV serostatus result of the specimen. Details of the HIV testing, quality assurance procedures and specific HIV test assays are explained in Chapter 4, and the ANC-HIV-SS protocol and prior publications.[24, 157] Further, HIV test assays were not consistent across survey rounds.

8.4. Statistical Methods for the secondary data analysis

To examine the relationship between educational attainment and prevalent HIV infection using ANC-HIV-SS data collected between 1994 and 2008. Because pregnant women were recruited from multiple sentinel sites, intra-site clustering was regarded a possible threat to validity of standard error estimation and subsequent inference. Pregnant women were regarded to be nested in respective sentinel sites yielding a hierarchical data structure. Data analyses were performed using Stata 12.1 and R-software version 3.0.[193]

8.5. Descriptive statistics

The distribution of continuous variables was summarized by the median and interquartile ranges (IQR) and the distribution of categorical variables were described by counts and percentages for each of the six survey rounds.[211] Wilcoxon Rank sum tests was used to compare distribution of continuous variables between HIV seropositive and HIV seronegative pregnant women, and Chi-square tests were used to compare proportions of categorical variables between HIV seropositive and HIV seronegative groups.[211]

8.5.1. Multivariable random intercept generalized linear mixed model (GLMM) were fitted to account for possible within-site clustering

ANC-HIV-SS data were captured consistently in 22 sites from 1994 through 1998 and from 24 sites between 2004 and 2008. To account for possible within-site clustering of pregnant that may threaten the validity of standard errors for estimated parameters (i.e., less trustworthy inferences), sentinel sites were modeled as random components in GLMM. Pregnant women who sought antenatal care from the same health center (i.e., sentinel site) might have similar sociodemographic characteristics. Therefore, GLMM were fitted to estimate the odds ratios (OR) for the association between educational attainment and prevalent HIV infection. The outcome variable HIV serostatus was dichotomous (i.e., HIV seropositive or HIV seronegative), therefore was assumed to follow a binomial distribution invariably required a logit link function for the GLMM.

All the odds ratios and 95% CI were computed from the fitted GLMM were adjusted for age, parity and marital status. Educational attainment was fitted as restricted cubic splines function with four pre-specified knots and age as a linear function. The

analyses were conducted on the premise that GLMM with a logit link function and sentinel site as random component accounted for possible within-site clustering, and was regarded efficient compared to standard logistic regression, which does not account for clustering.

8.5.2. Laplacian approximation of maximum likelihood used to compute model parameters

GLMM parameters for the relationship between educational attainment and prevalent HIV infections were estimated via maximum likelihood estimation using Laplacian approximation.[267, 328] Because the outcome variable was dichotomous, GLMM did not have closed form solutions to the maximum likelihood function via integration.[267] Consequently, Laplacian approximation, one of the several methods for approximating likelihood function solutions using iterations that avoid integration during maximum likelihood estimation, yet facilitates approximation of marginal likelihood were used.[188, 193, 328]

8.6. Assessment of statistical multiplicative interaction between educational attainment and specific covariates

8.6.1. Educational attainment and residence

To assess whether the association between education attainment and prevalent HIV infection varied according to the residence of the pregnant women (i.e., rural-urban location of sentinel site), two nested generalized linear mixed models (GLMM) with logit link functions one with main effects only and another with main effects along with a cross-product term (i.e., education*residence) were fit to the same data.[62] The Log-

likelihood values of the nested models were compared using the LRT. Apriori specified p-value for deciding a presence of meaningful statistical multiplicative interaction was 0.2.[38] Because p-value for LRT for the education-residence cross-product term was <0.20, separate analyses according to residence (i.e., level of effect modifying variable) were conducted.

8.6.2. Educational attainment and calendar time

Whether the association between educational attainment and survey calendar year varied between 1994 and 2008 was assessed using LRT. Survey years were centered by subtracting 1994 from the each of the survey years, and cross-product term created using centered survey calendar year and educational attainment. The centered values of survey calendar year (i.e., 0, 4, 8, 10, 12, and 14 years) eased the convergence of full GLMM when some values of a variable are high (1998, 2002 and 2004). Two nested models with and without cross-product term (i.e., education*survey year) were fit to the same data, and their log-likelihood compared. P-value <0.2 was interpreted as detection of substantively meaningful multiplicative interaction, and therefore warranted separate analyses by survey year to assess education-prevalent HIV infection association.

8.7. Incomplete data filled in by multiple imputations

Overall, other than in 1994 where educational attainment data were missing in an estimated 10%, fewer observations had missing values on covariates relevant to the current analysis. Therefore, analyses for the 1994 data were conducted using ten multiply imputed data sets and also based on complete case analysis that eliminated

observations with missing values. Handling missing data by a multiple imputation technique was prioritized because multiple imputation technique incorporates key sources of variability in the imputation process (i.e., variability inherent in sampling, within imputed data set variance, and between-imputed data set variance). However, when I found that parameter estimates and standard error based on complete case analysis and from analyses based on multiply imputed data were not materially different, estimates from analyses based on complete case analyses results were reported. Fewer observations had missing data on variables relevant to the analyses in 1998, 2002, 2004, 2006 and 2008. Therefore, analyses for 1998, 2002, 2004, 2006 and 2008 followed complete case analyses strategy.

8.8. Continuous variables were modeled using restricted cubic spline functions

Educational attainment was modeled as continuous variable using restricted cubic spline function (RCS) with four pre-specified knots to relax the linearity assumption and explore nonlinear relationship between education attainment and prevalent HIV infection.[373-375] Pregnant women's age was modeled linearly as a continuous variable.

8.9. Intra-class correlation coefficient

Intra-class correlation coefficient (ICC) was computed to quantify the proportion of variation attributed to between-site variation in HIV prevalence using variance estimates derived from the random intercept GLMM with sentinel site fitted as random component, but without fixed effects covariates (e.g., age, educational attainment) as adjustment

variables (i.e., an unconditional model). The variance of fixed effects covariates in logistic regression was regarded equivalent to $\pi^2/3$. [376] The ICC was computed using the formulae below.

$$\text{Intra-class correlation (ICC)} = \left(\frac{\sigma^2_{\text{SITE}}}{\sigma^2_{\text{SITE}} + \left(\pi^2 \diagup 3 \right)} \right)$$

8.10. Comparison of mixed effects and fixed effects model

LRT could have been used to decide whether to include random effects component (i.e., sentinel site) in the regression model, by comparing the nested model with same fixed effects covariates, with and without a variance component (i.e., sentinel site), but Bates and Pinheiro (2000) cautions the use of the LRT for assessing significance of the random components. The asymptotic reference distribution assumption of χ^2 on which the LRT rest may not hold when the value of the variance component is on or near the boundary of the (σ^2 =0) of feasible space, and consequently p-value for the LRT statistic based on the null hypothesis would be conservative. [267, 331]. Therefore, odds ratios and 95% CI for the relationship between educational attainment and prevalent HIV infection were pre-specified GLMM.

8.11. Meta-analysis were conducted to compute pooled estimates and assess between-survey year heterogeneity

Because the p-value for the LRT suggested presence statistical multiplicative interaction between educational attainment and survey calendar year (i.e., centered) <0.2, year-specific analyses of the association between educational attainment and prevalent

HIV infection were conducted. Similar to the approach by Zheng et al (2010), year-specific analyses were regarded as different studies and a DerSimonian-Laird random effect model meta-analysis of survey year estimates of odds ratio and 95% CI performed for specified levels of educational attainment.[212, 213, 377]

Specifically, the adjusted odds ratios and 95% CI for association of educational attainment and prevalent HIV infections in the six survey years (i.e., 1994, 1998, 2002, 2004, 2006 and 2008) for four levels of educational attainment were computed: zero number years of educational attainment; four years of educational attainment; nine years of educational attainment; and 12 years of educational compared to seven years of educational attainment were estimated, based on the GLMM adjusted for age, parity and marital status as described earlier.[214, 378]

First, odds ratios and 95% confidence intervals for *zero* educational attainment versus *seven* years educational attainment for each of the six survey years were meta-analyzed. Second, odds ratios and 95% confidence intervals for *four* years educational attainment versus *seven* years educational attainment for each of the six survey years were meta-analyzed. Third, odds ratios and 95% confidence intervals for *nine* years educational attainment versus *seven* years educational attainment for each of the six survey years were meta-analyzed. Fourth, odds ratios and 95% confidence intervals for *12 years* educational attainment versus seven years educational attainment for each of the six survey years meta-analyzed. Educational attainment of seven schooling years were used as referent based on the assumption that completion of primary school education is adequate to enhance health literacy. To measure heterogeneity of survey-specific odds ratios, DerSimonian and Laird-estimated Q statistics and I-squared were computed and

odds ratio estimates presented as forest plots.[213, 214] Pooled estimates were computed using “metafor" package in R statistical and computing program.[188, 210]

8.12. Ethical review

The ANC-HIV-SS was approved by ethics committee in Zambia as indicated in survey methods section in Chapter 4. The Vanderbilt University institution review board (IRB) granted permission for this secondary analysis of data.

8.13. Results

8.13.1. Descriptive summary of study sample

Preliminary analyses were focused on 82,086 pregnant women aged 15 to 44 years recruited for ANC-HIV-SS between 1994 and 2011. Mean age of pregnant women of the study sample increased only slightly between 1994 and 2011 (Figure 8.1). Overall, 54.8% (44,962/82,086) pregnant women aged 15 to 24 years but the investigation of the association were focused on 40,754 pregnant women 15 to 24 years in 1994 through 2008 because 4208 pregnant women from 2011 survey were excluded from the analysis because educational attainment data were not collected. The analytic sample investigating the association between educational attainment and prevalent HIV infection comprised 5542, 7101, 7545, 6865, 7070 and 6631 pregnant women from 1994, 1998, 2002, 2004, 2006, and 2008 respectively (Figure 8.1).

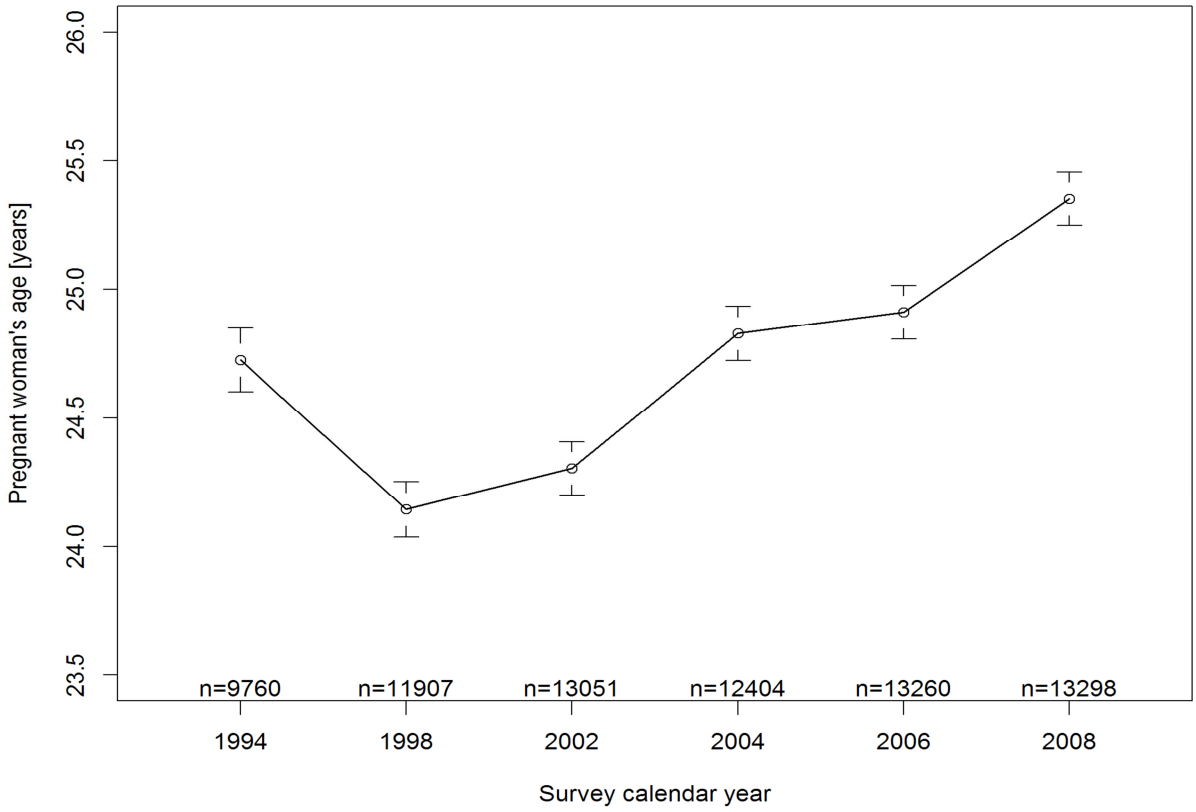

Figure 8.1. Mean age of pregnant women and 95% confidence intervals based on ANC-HIV-SS data in Zambia collected between 1994 and 2008. Little change in pregnant women's mean age, from 24.1 years in 1998 to 25.3 years in 2008. The "n" above the x-axis denotes the number of 15 to 44 year-olds pregnant women included in the preliminary analyses.

8.13.2. Proportion of pregnant who self-reported completion of at least 12 schooling years increased between 1994 and 2008

Overall, proportion of pregnant women who self-reported that they had completed at least 12 schooling years rose slightly between 1994 and 2008 (Figure 8.2). Table 2 and Table 3 present HIV prevalence trends by educational attainment between 1994 and 2008 for urban and rural areas respectively, and indicate falling HIV prevalence among 15 to 24 year-olds.

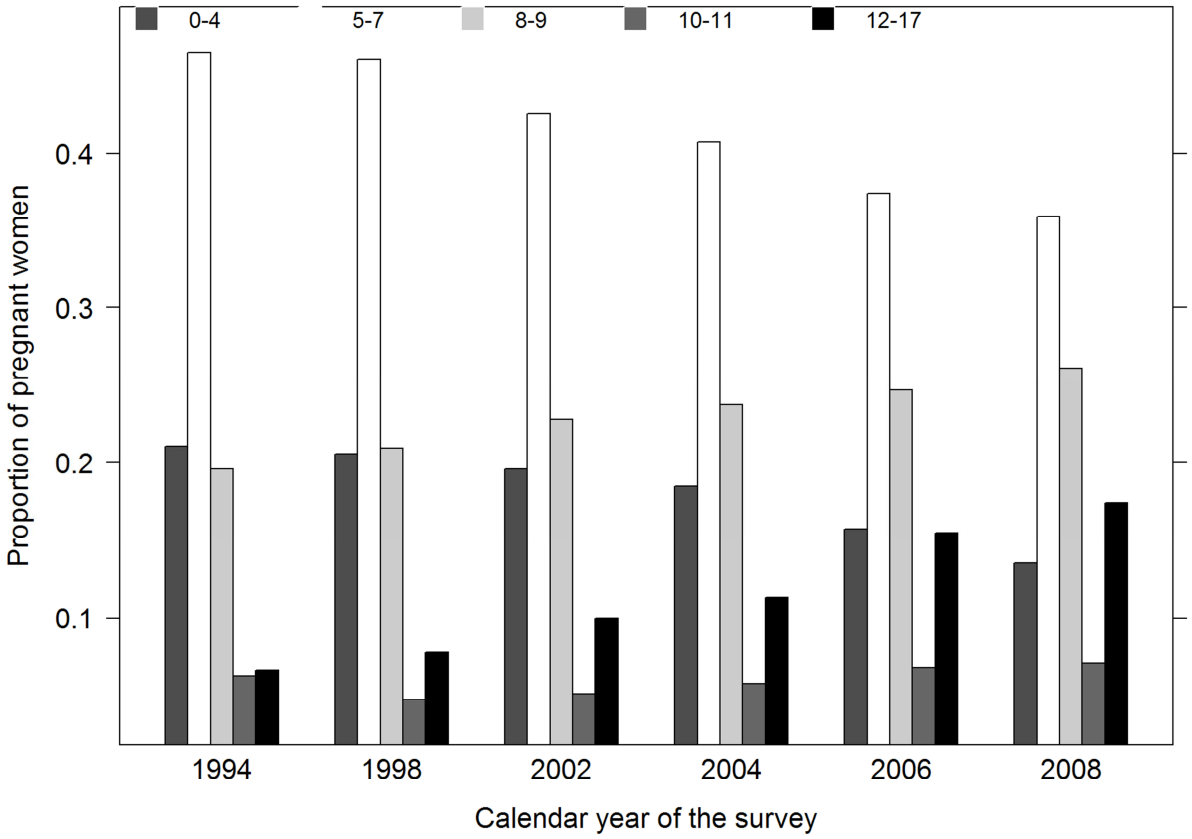

Figure 8.2. Distribution of educational attainment among pregnant women based on ANC-HIV-SS data collected in Zambia between 1994 and 2011. Educational attainment categories based on the education system in Zambia (lower primary (0-4); upper primary (5-7); junior secondary (8-9); incomplete senior secondary (10-11) and 12 to 19 years represent complete senior secondary and college or university educational attainment. The bars are sequentially arranged from beginning with 0-4 year's category and ending with 12-17 years category

8.13.3. Assessment of multiplicative interaction: educational attainment and residence

Apriori, it was decided that p-value <0.2 for the LRT would suggest presence of statistical multiplicative statistical interaction between educational attainment and residence. Because LRT p-value was <0.001 (i.e., implied that there was a meaningful

variation in the relationship between educational attainment and prevalent HIV infection by residence) separate analyses were conducted for urban and rural areas.

8.13.4. Assessment of multiplicative interaction: educational attainment and survey year

The p-value <0.001 for the LRT for assessment of nested models with and without cross-product term between educational attainment and calendar year of survey suggested the presence of multiplicative statistical interaction between educational attainment and survey year. Therefore, year-specific analyses, stratified by residence, were conducted. Figure 8.3 represent education-year interaction.

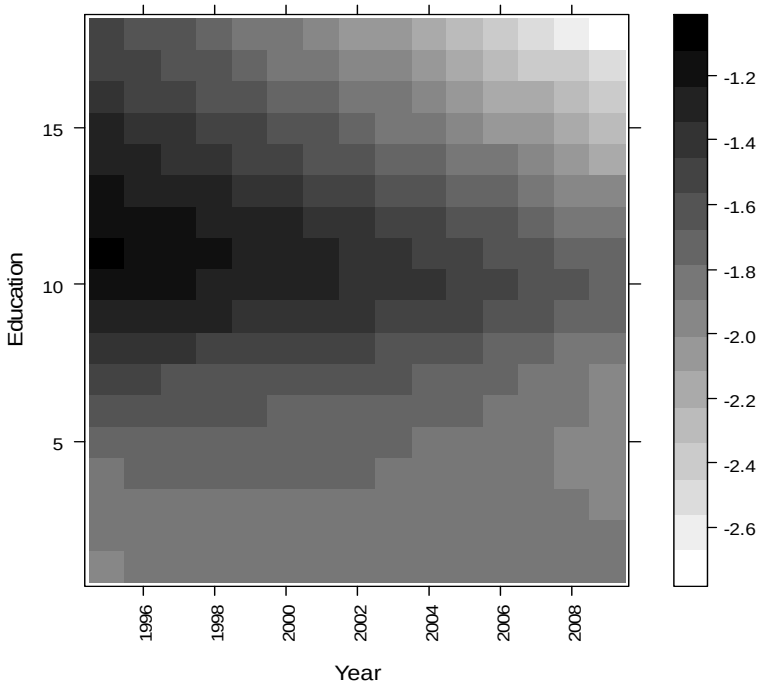

Figure 8.3. Heat map representation of statistical multiplicative interaction between educational attainment measured as number of schooling years completed and survey calendar year based on ANC-HIV-SS data collected between 1994 and 2008 in Zambia. Darker shades correspond to higher odds of prevalent HIV infection. The key legend displays color intensity. The maximum plausible number school years that can be completed by a 24 year was set at 17 schooling years based on the following assumptions (i.e., Primary school (7 years); secondary school (5 years); and post-secondary school (5 years).

8.14. HIV prevalence trends by category of educational attainment

Figure 8.4 presents profound fall HIV prevalence by educational attainment category between 1994 through 2008. Details are in Table 8.1 and Table 8.2.

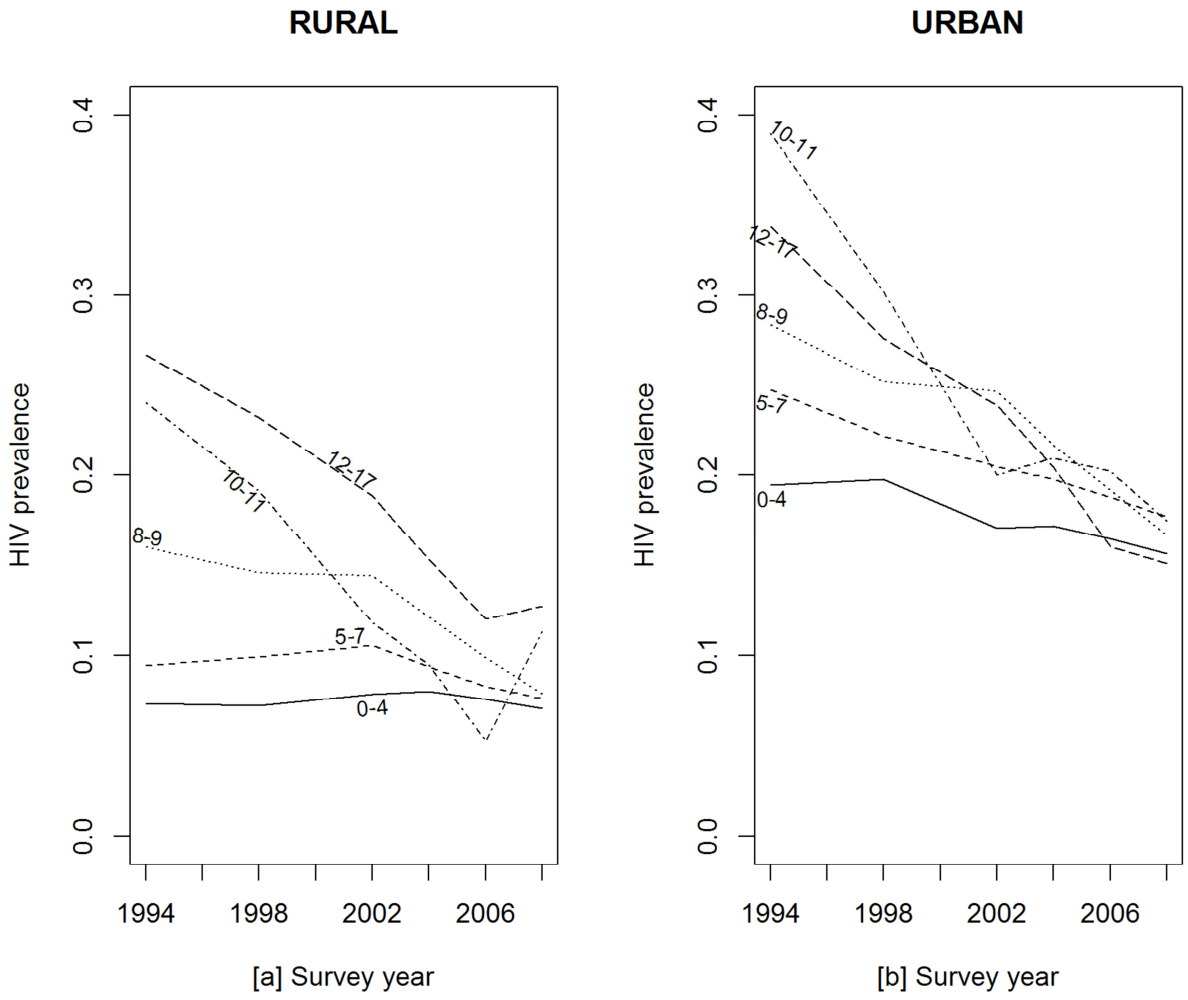

Figure 8.4. Education-category specific trends in HIV prevalence among 15 to 24 year-olds pregnant women based on ANC-HIV-SS data collected in Zambia between 1994 and 2008. Educational attainment categories were created according to educational attainment categories based on school system in Zambia. Fig. 4a and Fig. 4b represent rural and urban sites data, respectively. The line labels (i.e., 0-4; 5-7; 8-9; 10-11; and 12-17 indicate the number of school years completed by pregnant women.

8.14.1. Non-statistically significant association observed and in different direction during the 1994 to 2008 period in 15 to 24 year-olds urban and rural areas

Adjusted odds ratios and 95% CIs computed based on parameter estimates from the GLMM adjusted age, marital status and parity are presented in Table 8.3 and Table 8.4. The adjusted odds ratio for the association between educational attainment and prevalent HIV infection were largely not statistically significant as presented in Figure 8.5 and Figure 8.6. However, pregnant women in urban areas in 2008 who reported completion of 12 schooling years had lower odds compared to those who completed 7 schooling years adjusted odds ratio (AOR=0.55, 95% CI: 0.32, 0.97).

Additionally, the association between educational attainment and prevalent HIV infection among women in urban areas in 2006 was protective with decreasing educational attainment. For example, odds of prevalent HIV infections were higher lower for pregnant women who had self-reported completion of four schooling years compared to pregnant women who had completed seven schooling years, AOR=0.85, 95% CI: 0.74, 0.97 as shown in Figure 8.7.

The association between educational attainment and prevalent HIV infection among urban and rural areas pregnant women tended to be in different direction between 1994 and 2008. For example, pregnant women who self-reported having completed 12 schooling year tended to have higher odds of prevalent HIV infection compared to pregnant women who self-reported having completed seven schooling years 1994 and 1998, AOR=1.23, 95% CI: 0.68, 2.21 and AOR=1.24, 95% CI: 0.73, 2.08 respectively.

Conversely, odds of prevalent HIV infections tended towards a protective association for pregnant women in urban areas with no formal education compared to

pregnant women with 7 years of education in 1994 and 1998, AOR=0.85, 95% CI: 0.71, 1.02 and AOR=0.78, 95% CI: 0.57, 1.08, respectively. Notably, chance occurrence of the observed association cannot be ruled. Figure 8.5 provides presents graphically the association between educational attainment and prevalent HIV infection in 1994 and 1998.

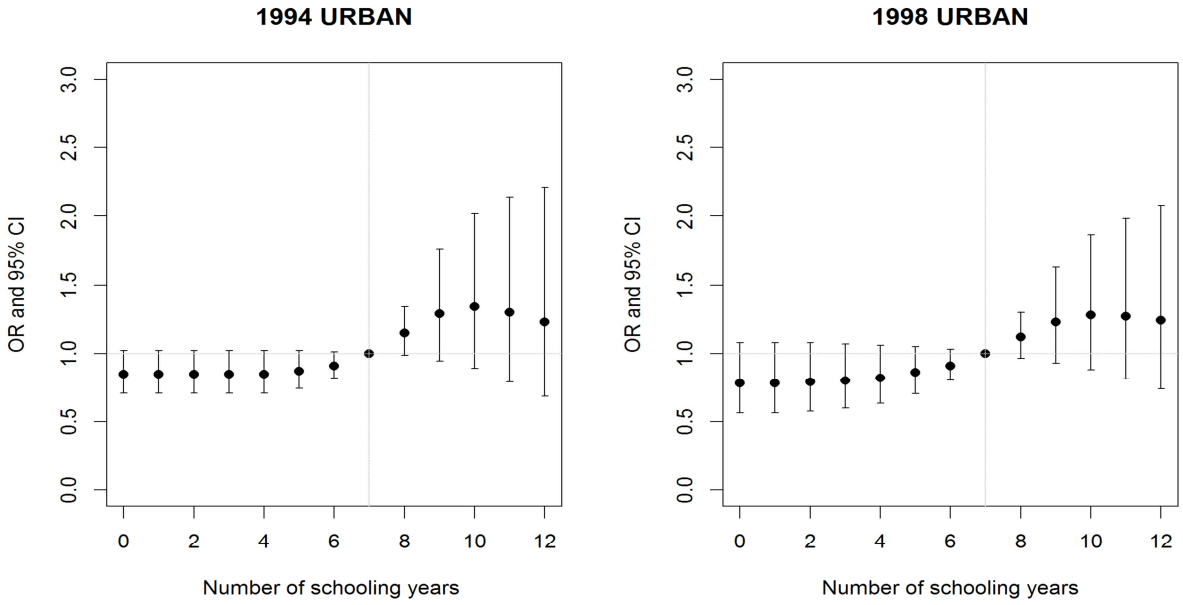

Figure 8.5. Odds ratio and 95% CIs for association between educational attainment and prevalent HIV infection among pregnant women aged 15 to 24 years in sites located in urban areas in Zambia based on ANC-HIV-SS data collected between 1994 and 1998. The referent educational attainment was completion of 7 schooling years.

8.14.2. Higher education attainment tended to protective in urban areas beginning 2002

Among pregnant women in urban areas in 2002 and 2004, the odds of prevalent HIV infections for pregnant women who had self-reported having completed 12 schooling years compared to pregnant women who self-reported completion of seven

years of schooling were AOR=0.61, 95% CI: 0.35, 1.07 and AOR=0.66, 95% CI: 0.52, 1.64, respectively.

Figure 8.6 and Figure 8.7 presents the odds ratios and 95% CI estimates for the association between educational attainment and prevalent HIV infections in urban areas for the period 2002 through 2008. There was no significant association between educational attainment and prevalent HIV infections in urban areas, although the observed association tended to be protective with increasing educational attainment in survey years 2002, 2004, and 2008.

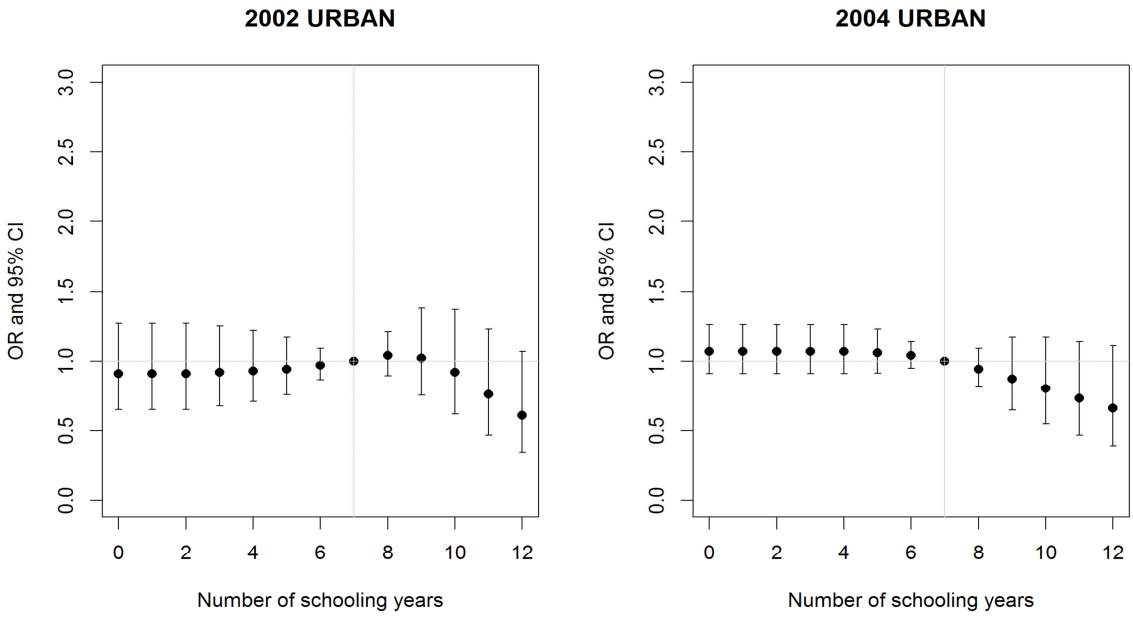

Figure 8.6. Odds ratio and 95% CI for the association between educational attainment and prevalent HIV infection among pregnant women aged 15 to 24 years in sites located in urban areas based on ANC-HIV-SS data collected between 2002 and 2004. Pregnant women who had completed 7 schooling years were used as referent group.

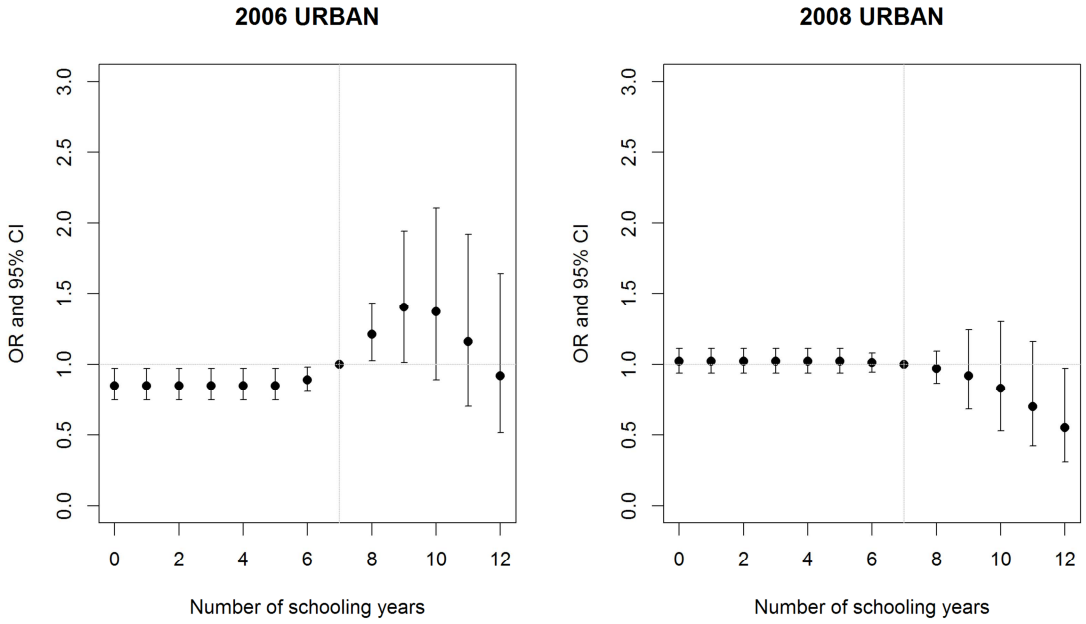

Figure 8.7. Odds ratio and 95% CI for the association between educational attainment and prevalent HIV infection among pregnant women aged 15 to 24 years in sites located in urban areas in Zambia based on the ANC-HIV-SS between 2006 and 2008. Pregnant women who had completed 7 schooling years were used as referent group.

8.14.3. Non-significant positive association between educational attainment and prevalent HIV infection in later years in rural areas

The odds of prevalent HIV infections in pregnant women in rural areas who completed at 12 schooling years versus pregnant women who completed seven schooling years among surveyed pregnant were AOR=3.09, 95% CI: 0.95, 10.04) in 1994; AOR=1.24, 95% CI: 0.43, 3.57 in 1998; AOR=2.38, 95% CI: 0.91, 6.23 in 2002, 1.08, 95% CI: 0.40, 2.88 in 2004; AOR=1.33, 95% CI: 0.40, 2.88 in 2006 and AOR=1.70, 95% CI: 0.61, 4.73 in 2008. Wide confidence intervals indicate imprecise estimates.

In rural areas, the odds of prevalent HIV infections tended to be protective, with decreasing educational attainment, although chance occurrence cannot be ruled as shown in Figure 8.8 and Figure 8.9. For example, odds ratio of prevalent HIV infection among

pregnant women who had not received any formal education (i.e., zero number of schooling years) compared to the odds of prevalent HIV infection for pregnant women who had received 7 years of education were AOR=0.58, 95% CI: 0.17, 2.02 in 1994; AOR=0.82, 95% CI: 0.27, 2.45 in 1998; AOR=0.71, 95% CI: 0.27, 1.87 in 2002, OR=0.74, 95% CI: 0.29, 1.90 in 2004 and AOR=0.82, 0.36, 1.83 in 2006 and AOR=0.76, 95% CI: 0.37, 1.54 in 2008.

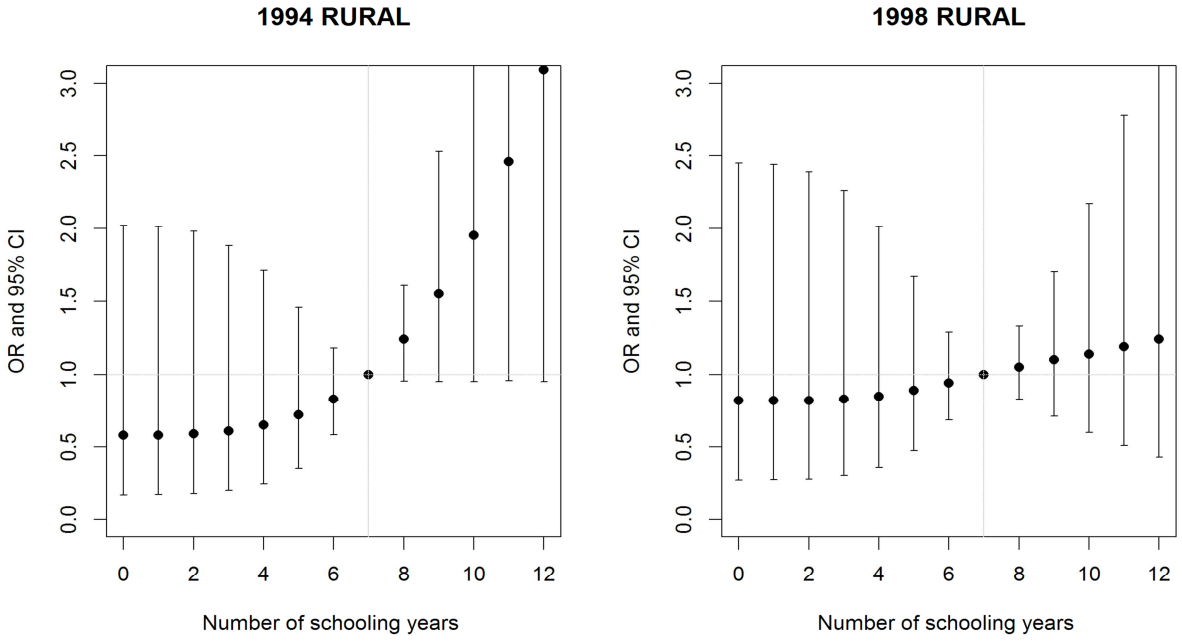

Figure 8.8. Odds ratio and 95% CI for the association between educational attainment and prevalent HIV infection among pregnant women aged 15 to 24 years in rural who participated in the ANC-HIV-SS conducted between 1994 and 2008 in Zambia. Pregnant women who had completed 7 schooling years were used as referent group.

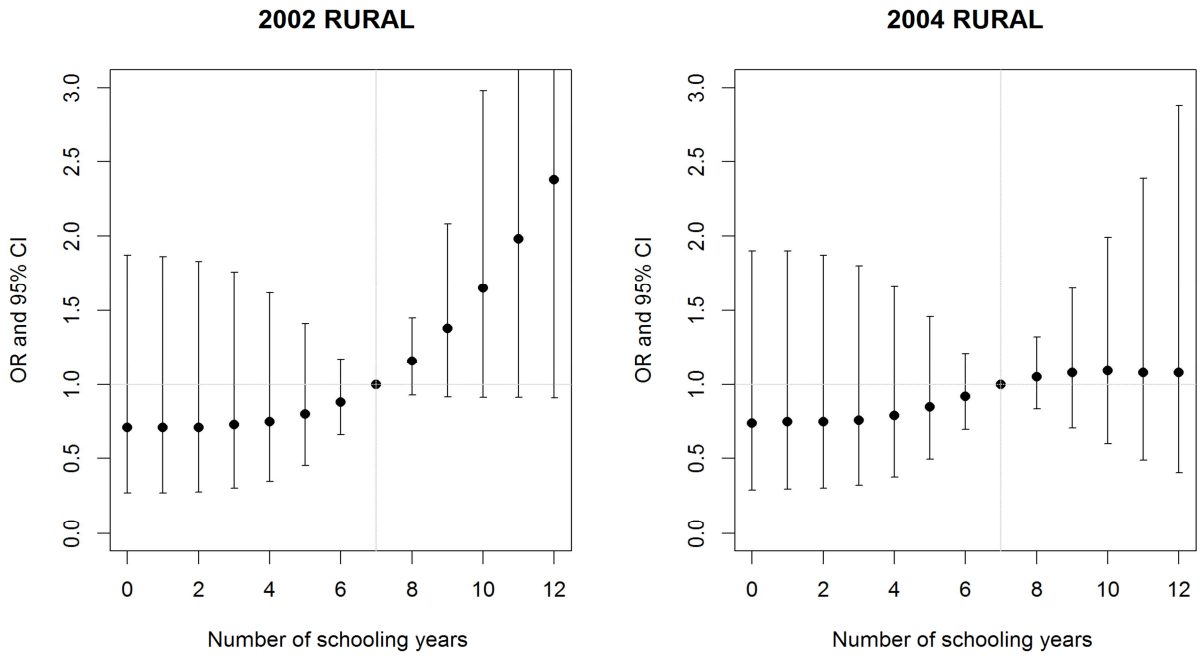

Figure 8.9. Odds ratio and 95% confidence intervals for the association between educational attainment and prevalent HIV infection among pregnant women aged 15 to 24 years in sites located in rural areas in Zambia for survey round 2002 and 2004. The referent level for educational attainment was completion of 7 schooling years and GLMM adjusted for age, parity and marital status. Educational attainment modeled using RCS.

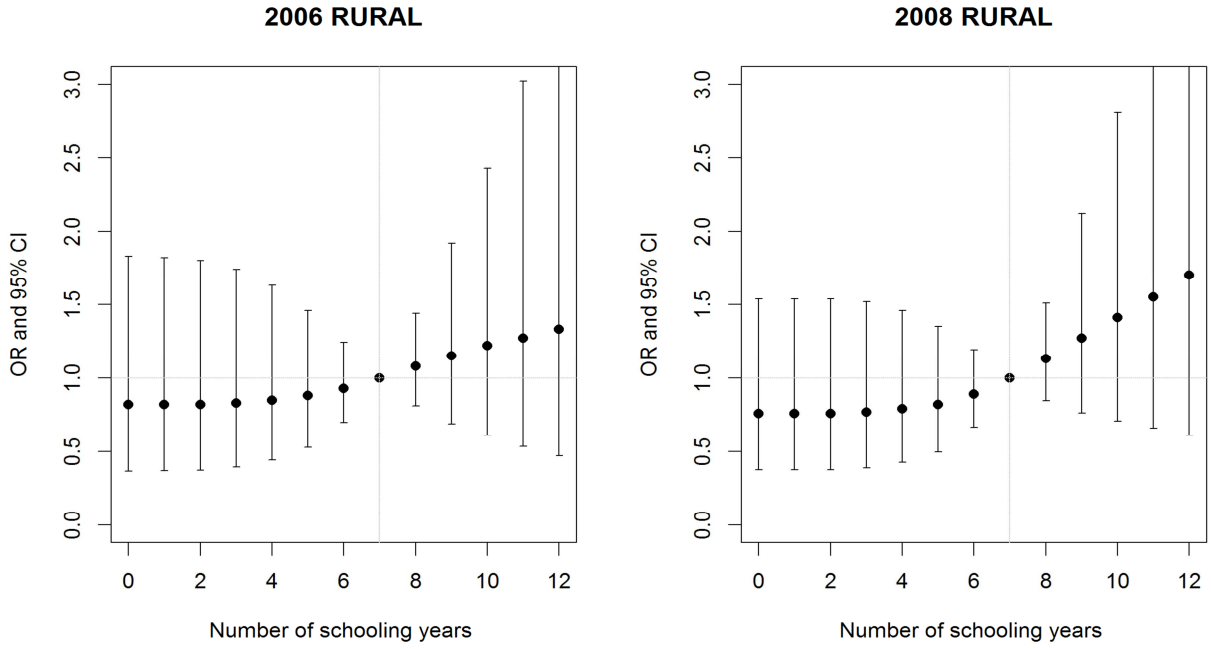

Figure 8.10. Estimated odds ratio and 95% confidence intervals for the association between educational attainment and prevalent HIV infection among pregnant women aged 15 to 24 years in sites located in rural areas in Zambia for survey round 2002 and 2004. The referent level for educational attainment was completion of 7 schooling years and GLMM adjusted for age, parity and marital status.

8.14.4. Greater odds of prevalent HIV infection among pregnant women in 2008 who reported single, divorced marital status, and had at one child

Among pregnant women recruited from urban areas, single pregnant women and divorced pregnant women had significantly greater odds of prevalent HIV infections compared to married pregnant women, AOR=1.35, 95% CI: 1.06, 1.72 and AOR=3.37, 95% CI: 1.20, 9.49, respectively. Table 4 indicate that pregnant women who reported being divorced in 2006 were more likely to be HIV-infected compared to pregnant women who were married (AOR=8.89, 95% CI: 3.36, 23.5); wide 95% CI indicate imprecision of the estimate. The observed association between marital status and prevalent HIV infection in 1994, 1998, 2002 and 2004 may be explained by chance.

Table 8.4shows that based on rural sites data in 2008, pregnant women who reported being divorced had greater odds of prevalent HIV infections compared to pregnant women who indicated marital status as married in 1998, 2004, and 2008: AOR=2.65, 95% CI: 1.46, 4.83 in 1998; AOR=2.99, 95% CI: 1.44, 6.18 in 2004; AOR=5.39, 95 CI: 2.40, 12.0 in 2008.

The odds of prevalent HIV infections were significant, and greater among pregnant women in urban areas in 2008 who self-reported having one child (AOR=3.37, 95% CI: 1.20, 9.49) and two children (AOR=1.35, 95% CI: 1.35, 1.06, 1.72) compared to pregnant women who self-reported no children. Similarly, among pregnant women in 2006 pregnant women with one child had greater odds of prevalent HIV infections compared to pregnant women with no children (AOR=1.07, 95% CI: 3.36, 23.5).

Among pregnant women in urban areas, parity of at least two children tended to be protective in 1994, 2002, and 2004, although random error could not be ruled out in 1998

(Table 8.3). Similar patterns were observed in rural areas where parity of ≥ 2 children tended to be protective in all survey years but not beyond chance occurrence in all except in 1998 as shown in Table 8.4.

8.14.5. Meta-analysis of year-specific odds ratios and 95% CIs

Using the method applied by Zheng et al (2010) of meta-analyzing odds ratios and 95% CIs of different study cohorts, odds ratio and 95% CI of survey year-specific odds and 95% CI for selected levels of education (i.e., zero education, four years education, nine years education and 12 years education, compared to seven years of schooling) estimated from GLMM (i.e., adjusted for three covariates: age, parity and marital status) for 1994, 1998, 2002, 2004, 2006 and 2008 were meta-analyzed and results are shown in .Figure 8.11 and Figure 8.12 [377]

8.14.6. Higher educational attainment tended to be protective but may be explained chance in urban areas

The pooled odds ratio estimate from meta-analysis of year-specific odds ratio estimates indicate that pregnant women in urban sites who self-reported having completed 12 years of schooling tended to have lower odds of prevalent HIV infections compared to pregnant women who self-reported completion of 7 schooling years OR=0.82, 95% CI: 0.62, 1.10) but the observed association may be explained by chance because the 95% CI includes OR=1.0 as shown in Figure 8.11.

8.14.7. Higher odds of prevalent HIV infection suggested among pregnant women in rural areas who had self-reported higher education

Chance occurrence may explain the noted elevated odds of prevalent HIV infection observed among pregnant women in rural areas between 1994 and 2008. Based on the meta-analysis of year-specific odds ratio estimates, the pooled odds ratio and 95% CI for the association between educational attainment and prevalent HIV infection tended to be protective but not beyond chance with decreasing educational attainment (OR=0.58, 95% CI: 0.58, 1.07 for 4 versus 7 years of schooling years), and positive but beyond chance with increasing educational attainment (OR=1.64, 95% CI: 1.08, 2.51 for 12 versus 7 years of schooling) as shown in Figure 8.12.[377]

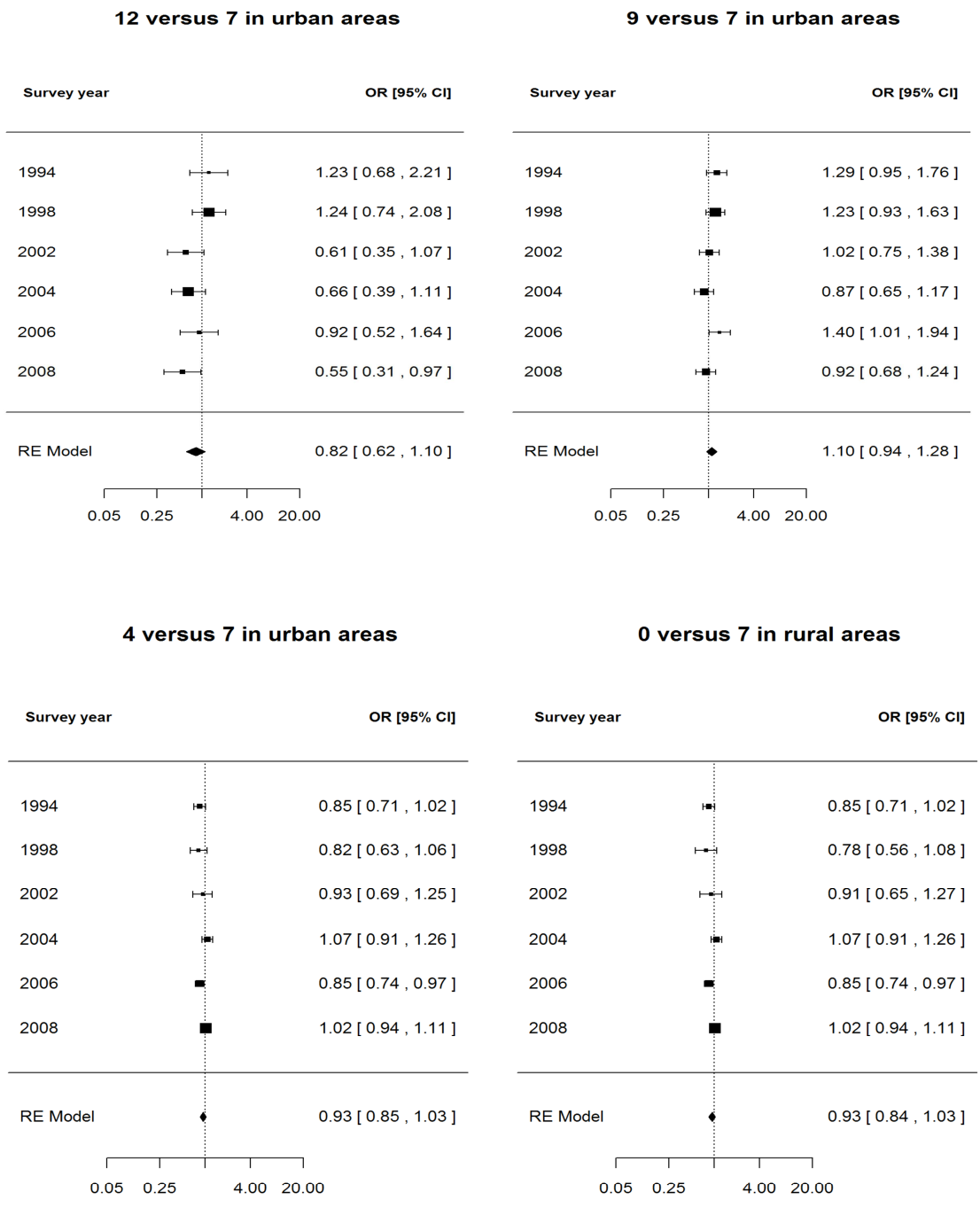

Figure 8.11. Forest plots for the meta-analysis of the association between educational attainment and prevalent HIV infection generated using random effect model using DerSimonian and Laird method used to assess between-survey heterogeneity and to calculate the overall odds ratio for sites in *urban* areas. Odds ratios and 95% confidence intervals were computed using GLMM and adjusted age, parity, and marital status. The square boxes represent the odds ratio and its size is inversely proportion to the variance of log odds ratio. The horizontal bars represent the 95% confidence interval and the diamond shaped object at the bottom corresponds to the overall estimate of odds ratio.

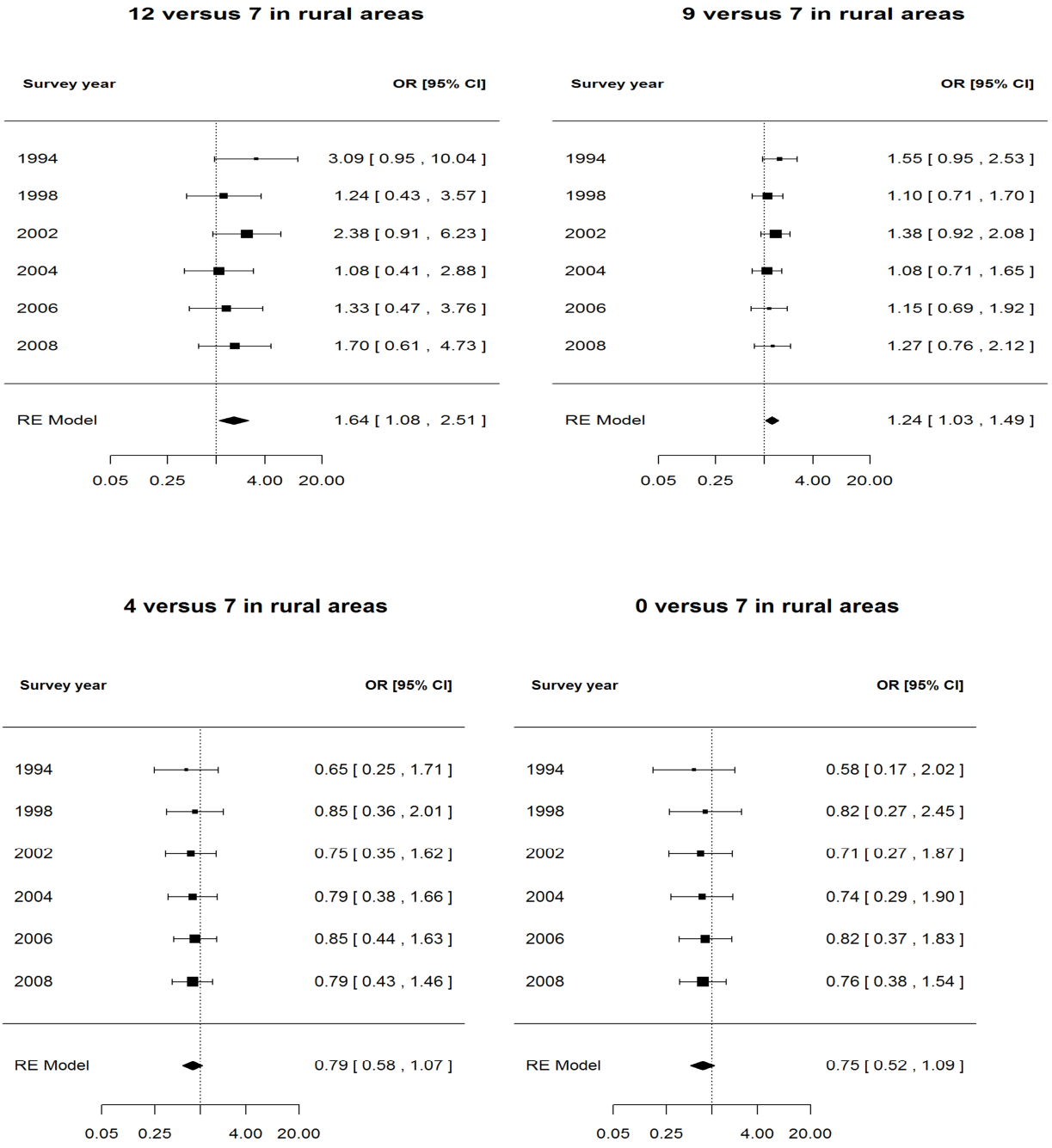

Figure 8.12. Forest plots for the meta-analysis of the association between educational attainment and prevalent HIV infection generated using random effect model using DerSimonian and Laird method used to assess between-survey heterogeneity and to calculate the overall odds ratio for sites in *rural* areas. Odds ratios and 95% confidence intervals were computed using GLMM and adjusted age, parity, and marital status. The square boxes represent the odds ratio and its size is inversely proportion to the variance of log odds ratio. The horizontal bars represent the 95% confidence interval and the diamond shaped object at the bottom corresponds to the overall estimate of odds ratio.

8.15. Discussion

Although the observed association is not beyond chance occurrence, based on the ANC-HIV-SS data, pregnant women with higher educational attainment in urban areas tended to have reduced odds of prevalent HIV infections between 2002 and 2008, yet higher educational attainment tended to be associated with elevated odds of prevalent HIV infections in earlier years (i.e., 1994 and 1998). Conversely, increasing educational attainment tended towards a positive association with prevalent HIV infection in rural areas between 1994 and 2008, although the not statistically significant.

Based on the method applied by Zheng et al (2010), rounds of ANC-HIV-SS in 1994, 1998, 2002, 2004, 2006 and 2008 were regarded as six separate studies, and adjusted odds ratios for selected levels of educational attainment compared referent level (i.e., seven schooling years) meta-analyzed yielding a pattern towards a protective association but non-statistically significant with increasing educational attainment among pregnant women in urban Zambia as shown in Figure 8.11.

However, among pregnant women in rural areas, the association between educational attainment and prevalent HIV infection was positive as educational attainment increased but protective yet mostly not significant with decreasing educational attainment as shown in Figure 8.12. To compute an estimate of the pooled odds ratio and 95% CI, the random effect model meta-analysis based on DerSimonian and Laird method was used to account for between-survey variability. Although not significant statistically, the protective association pattern observed are nearly consistent with prior reports in South Africa, Uganda and Zambia.[46, 140, 141, 197]

Examination of HIV prevalence by educational attainment created according to the school system in Zambia (i.e., lower primary, upper primary, junior secondary, incomplete senior secondary and ≥ senior secondary), HIV prevalence declined overall, but profoundly in the category of pregnant women who reported higher educational attainment, consistent with prior reports.[24]

Although not significant, the elevated odds of prevalent HIV infection among pregnant women in rural areas with increasing educational attainment is worrying, and seems consistent with patterns that were observed in most urban areas in earlier years of the HIV epidemic, the 1980s and 1990s. Studies conducted in earlier years of the HIV epidemic found increased odds of HIV prevalence among persons with higher levels of educational attainment.[52] To explain the elevated odds of prevalent HIV infections among educated categories in earlier years of the HIV epidemic (i.e., 1980s and 1990s), a number of investigators reasoned that economic and resource empowerment associated with higher educational attainment also enables behaviors that increase the risk of HIV acquisition.[80]

With respect to the rural environment where background HIV prevalence is lower compared to urban areas, higher educational attainment might be a risk factor, as was case in earlier HIV epidemics stage in urban areas. The seemingly contrasting direction in the association between educational attainment and prevalent HIV infections in urban and rural areas, although largely not significant, underscores the critical roles of self-efficacy in adoption of safer sexual behavior and also the importance of contextual factors.[25, 140]

Educational attainment does not have direct biological influence on risk of HIV infection, may exert influence via other factors that may biologic influence. Further, limitations are often encountered in measurement of sexual behavioral characteristics and other factors related to educational attainment with respect to risk of HIV infection.[379] Further, sexual behavior may be influenced by several factors including age, marital status, culture and social norms, health promotions, and educational attainment.[379] It seems like higher educational attainment can be both a risk factor and protective factor depending on the stage of the epidemic and contextual settings.

8.15.1. Strength of the study

The current study pooled data from 7 survey rounds conducted over a period of 15 years inclusive between 1994 and 2008, covering the critical phase of the evolution of the HIV epidemic in Zambia.[157] Although different sample of pregnant women were used, ANC-HIV-SS data facilitated assessment of the association between educational attainment and prevalent HIV infection using a nearly consistent ANC population of pregnant women.[157] While comparing a series of cross-sectional surveys from different rounds of ANC-HIV-SS is a strength, temporal changes in the composition of the study population between 1994 and 2008 is possible, but not likely to be drastic. The application of random effect meta-analytic to adjusted odds ratio estimates from year-specific estimates improved the robustness of the analysis by integrating between study variability in the analyses.[201, 212]

The large study size and consistent set of covariates used for adjustment in the multivariable regression model for assessing the relationship between educational

attainment and prevalent HIV infections enabled comparability of the analyses across survey years. Therefore, it may be reasonable to assume that consistency in operational definition of covariates across survey years and consistency in set of adjustment covariates (i.e., age, parity, and marital status) for GLMM obviated variations in odds ratio estimates that would have arisen from using different sets of covariates and definition of covariates for modeling the relationship between educational attainment and prevalent HIV infections. Further, the fact that pregnant women in the study were drawn from diverse geographic areas and socioeconomic settings are strengths of the study, as well as the long period (i.e., 1994 to 2011) examined.

Educational attainment was modeled as a continuous variable using restricted cubic spline functions, and possible within-site clustering were accounted for by modeling sentinel site as a random components, consequently enhancing the validity of current's study inference. In contrast, most prior studies that examined the association between education attainment and prevalent HIV infection categorized the educational attainment measure, even where the data was collected in continuous form as number of schooling years completed. [46, 56, 141] Modeling educational attainment as a continuous variable in GLMM obviated creation of educational attainment categories using subjective cut-points. Further, fitting educational attainment using restricted cubic splines function relaxed the linearity assumption, and also facilitated flexible modeling of non-linear association with log-odds of prevalent HIV infection.

For all the analyses conducted, variables were assumed to have been measured correctly; therefore using continuous variables rather than categorized variable (e.g., age) massively minimized possible residual confounding. Further, use of continuous form of

educational attainment and the large study size enhanced the power of the study to detect association between educational attainment and prevalent HIV infection. As revealed by systematic reviews on the relationship educational attainment and prevalent HIV infections, most past and current literatures are dominated by studies that have used categorical measure of educational attainment, and have limited ability to examine non-linear associations within categories.

The use of standardized survey procedures in all the sentinel sites may limit, but may not eliminate heterogeneity in implementation of survey procedures across sites over in all the six survey rounds. Further efforts focused on improving comparability of data across survey sites included pre-survey personnel training at a central location.

8.15.2. Limitations

The study findings must be interpreted in the milieu of the following limitations. First, the study was conducted using routine HIV surveillance data, therefore restricted to covariates that are routinely collected for ANC-HIV-SS. For example, educational attainment was not captured in 2011; consequently 2011 data were excluded from the analyses of the association between educational attainments. Admittedly, ANC-HIV-SS, although a cornerstone source for HIV prevalence data in most sub-Saharan Africa countries, including Zambia, is subject to selection biases, and findings may have limited generalizability.

The sociodemographic structure and characteristics of pregnant women being tested may influence the magnitude of the association between educational attainment and prevalent HIV infection. Prior researchers have noted that the pregnant women

characteristics of pregnant women who attended and those who did not attend antenatal care clinic may be different, thus threatening the validity of the estimated odds ratio. Further, selection bias of young women into the study, who have higher propensity for reproduction, might be profound. The estimated odds ratios and 95% CIs were derived from a sample that excluded non-pregnant but sexually active women. Consequently, the study findings are may not be generalized to women outside the study population.

Pregnant women's sociodemographic and pregnancy history information were self-reported via nurse administered questionnaire. Marital status reported by some women may have been influenced by report bias arising from social desirability bias because pregnancies outside formal marital arrangement are normally frowned upon. Therefore, some women, especially young women may be more inclined to report being married as their marital status to avoid social ridicule.

The assumption that prevalent HIV infections in pregnant women aged 25 to 44 years represented HIV infections that were longstanding, and that prevalent infections in pregnant women aged 15 to 24 years represented recent HIV infection disregarded the possibility that new HIV infection may occur 25 to 44 year-olds, and also the fact that some of the 19 to 24 year-olds might be long-term infections (e.g., pregnant women aged 24 years but HIV infected at age 15). The rigid age bounds created may not be realistic. Use of biologic assay for identifying new HIV infections would limit misclassification, and enable use of data from all ages to assess the association. These are not yet available in Zambia and, in fact, have only recently been validated.[174, 175, 380]

Some of the pregnant women were not old enough to have completed secondary school education [12th grade] while others were not old enough (e.g., <19 years) to have completed secondary school education. One way to improve robustness of the estimates would be to include an interaction term between age and education or to restrict the analyses to 20 to 24 years olds who were old enough to complete at least 12 schooling years. These analyses were not conducted because they were not pre-specified.

The extent to which different diagnostic HIV assays used over the years affected the estimated odds ratio and 95% CIs is not estimable. However, HIV testing using different assay may not materially influence the estimated odds ratios and 95% CI because a nearly consistent HIV testing algorithms was used that minimized the chances of false positive and false negative was implemented, and HIV diagnostic assays used were of high sensitivity and specificity. Further, misclassification of serostatus by HIV assays would be largely non-differential over the years.

Data on contextual characteristics of the catchment areas for the sentinel sites were not captured, and therefore lack of consideration of contextual factors may have limited my investigation of the association between educational attainment and prevalent HIV infection.[30, 55] Sexual behavior information not collected, therefore the current analysis lack consideration of sexual behavior information that has been reported to provide insight in the dynamics of the HIV epidemics. However, sexual behavior information is important but often plagued by measurement challenges regarding sexual behavior construct. ANC-HIV based surveillance data does not include information on how long a pregnant woman has stayed in the catchment areas, therefore lack ability to

differential long-term resident from short-term residents might also resulted in misclassification of women infected in urban areas as rural or vice versa

The quality of educational attainment across survey years were not assessed (i.e., 1994, 1998, 2002, 2004, 2006 and 2008). Therefore, the implicit assumption that quality of educational attainment was even across the survey years might be not be realistic because the quality of education received by pregnant women might have changed over the years, and may be different between rural and urban areas. Residual confounding might be present due to variables that might have been left out or due to misclassification during measurement of variables.

Residual confounding cannot be rule as an explanation of the result. However, sources of residual confounding were minimized by avoiding categorization of continuous variables, may be present because of some important variables were not included as adjustment covariates because they were not measured or because covariates were imprecisely measured or recorded. Further, pregnant women who reported greater than 12 years of educational attainment were regarded as having completed 12 years, limiting ability to examining relationship beyond greater than 12 years of educational attainment. However, two out of the six survey rounds data (i.e., 1998 and 2002) indicated educational attainment beyond 12 years as a category.

There is need to intensify education-oriented interventions in rural areas where there higher educational attainment tended to be associated with increased odds of prevalent HIV infection. Directing future research towards contextual settings factors

related to increased risk of HIV infection may unravel important modifiable factors for targeting HIV prevention efforts, including empowerment of women.[141, 381, 382]

8.15.3. Conclusion

Based on data from a period of more than a decade, 1994 through 2008, and consistent with prior research findings, the current study findings suggest that the association between educational attainments has waned in urban areas but has not waned in rural areas, although estimated odds ratios may be subject to the influence of random error. It was beyond capacity of the current study to identify the factors that influence the differential association between educational attainment and prevalent HIV infection in rural and urban sites.

Future studies should conduct more nuanced investigations focusing on background information on sexual behavior characteristics and contextual characteristics to help investigate the observed differences. Whether educational attainment construct reflect a measure of socioeconomic status or literacy is cannot be inferred from the current study. Therefore, it is an imperative for future studies to examine the relationship between literacy and HIV in rural settings.

Inconsistent results may stem from methodologically different approaches applied in different studies (i.e., sampling of subjects, study population, and covariates controlled for as potential confounders) as well as limited number of primary studies with rigorous data collection and analysis approaches. Educational attainment is variously measured across studies.[142] Additionally, the multivariable regression modeling approaches and adjustment variables differ across studies that have examined the relationship between

educational attainment and HIV infection.[41, 80, 81, 141] Additionally, exposure and covariate may differ across studies, but most studies have used serologically-confirmed HIV serostatus as the outcome.[39, 56, 80, 141]

Clarifying the association between educational attainment and prevalent HIV infection has been critical feature in the design and implementation of HIV prevention and treatment interventions: provide insight into how the interventions can be crafted to suit the local contextual settings.[39, 42, 52, 81, 146, 196, 362] Undoubtedly, education is an important social determinant of health, but not an accurate indicator of socioeconomic status because it does not capture several factors at individual-level and society-level (e.g., income, material possession, and education) that define socioeconomic status.[49, 126, 383]

Table 8.1. HIV prevalence by educational attainment among pregnant women aged 15 to 24 years surveyed in the Zambia Antenatal Attendees Sentinel Surveillance in *urban* sites, 1994 to 2008

Educational attainment[†]	1994	1998	2002	2004	2006	2008
0 to 4	313 18.8 (14.9-23.6)	512 20.7 (17.4-24.4)	515 17.1 (14.1-20.6)	402 18.4 (14.9-22.5)	307 16.3 (12.6-20.8)	249 15.3 (11.3-20.3)
5 to 7	1391 24.9 (22.7-27.2)	1801 21.9 (20.1-23.9)	1690 20.5 (18.6-22.5)	1400 20.8 (18.7-23.0)	1174 17.6 (15.6-19.9)	1034 18.2 (16.0-20.6)
8 to 9	860 29.8 (26.8-32.9)	1224 23.0 (20.8-25.5)	1300 24.7 (22.4-27.1)	1165 20.6 (18.4-23.0)	1182 20.2 (18.0-22.6)	1248 16.3 (14.3-18.4)
10 to 11	185 38.9 (32.2-46.1)	252 30.6 (25.2-36.5)	320 20.0 (16.0-24.7)	355 20.8 (16.9-25.4)	390 22.1 (18.2-26.4)	429 16.8 (13.5-20.6)
≥ 12	228 36.0 (30.0-42.4)	378 24.6 (20.5-29.2)	526 24.1 (20.7-28.0)	683 21.1 (18.2-24.3)	879 16.0 (13.8-18.6)	955 15.1 (12.9-17.5)

[†]Number of school years completed: Categories reflect the school system in Zambia. The 10 to 11 years was included to reflect women who drop out due to pregnancy

Table 8.2. HIV prevalence by educational attainment among pregnant women aged 15 to 24 years surveyed in the Zambia Antenatal Attendees Sentinel Surveillance in *rural* sites, 1994 to 2008

Educational attainment[†]	1994	1998	2002	2004	2006	2008
0 to 4	657 7.5 (5.7-9.7)	820 7.0 (5.4-8.9)	865 7.9 (6.2-9.8)	758 9.5 (7.6-11.8)	646 6.3 (4.7-8.5)	463 7.6 (5.5-10.3)
5 to 7	949 9.5 (7.8-11.5)	1389 9.9 (8.4-11.5)	1411, 10.6 (9.1-12.3)	1220 10.0 (8.4-11.8)	1225 7.3 (6.0-8.9)	1144 8.0 (6.6-9.8)
8 to 9	310 16.8 (13.0-21.3)	474 13.5 (10.7-16.9)	589 14.4 (11.8-17.5)	647 12.4 (10.0-15.1)	752 9.4 (7.6-11.7)	690 8.1 (6.3-10.4)
10 to 11	32 21.9 (11.0-38.8)	50 22.0 (12.8-35.2)	93 11.8 (6.7-19.9)	131 10.7 (6.5-17.1)	208 5.3 (3.0-9.2)	177 11.3 (7.4-16.8)
≥12	37 27.0 (15.4-43.0)	57 21.1 (12.5-33.3)	76 23.7 (15.5-34.4)	81 11.1 (6.0-19.8)	160 12.5 (8.2-18.5)	204 12.7 (8.8-18.0)

[†]Number of school years completed: Categories reflect the school system in Zambia. The 10 to 11 years was included to reflect women who drop out due to pregnancy

Table 8.3. Odds ratio and 95% confidence interval (CI) for fixed parameters from a GLMM that assessed the relationship between educational attainment and prevalent HIV infection among pregnant women in urban sentinel sites beginning 1994 through 2013

Educational attainment‡	1994 OR (95% CI)	1998 OR (95% CI)	2002 OR (95% CI)	2004 OR (95% CI)	2006 OR (95% CI)	2008 OR (95% CI)
0	0.85 (0.71-1.02)	0.78 (0.57-1.08)	0.91 (0.65-1.27)	1.07 (0.90-1.26)	0.85 (0.74-0.97)	1.02 (0.94-1.11)
1	0.85 (0.71-1.02)	0.78 (0.57-1.08)	0.91 (0.65-1.27)	1.07 (0.90-1.26)	0.85 (0.74-0.97)	1.02 (0.94-1.11)
2	0.85 (0.71-1.02)	0.79 (0.58-1.08)	0.91 (0.65-1.27)	1.07 (0.90-1.26)	0.85 (0.74-0.97)	1.02 (0.94-1.11)
3	0.85 (0.71-1.02)	0.80 (0.60-1.07)	0.92 (0.67-1.25)	1.07 (0.90-1.26)	0.85 (0.74-0.97)	1.02 (0.94-1.11)
4	0.85 (0.72-1.02)	0.82 (0.63-1.06)	0.93 (0.70-1.22)	1.07 (0.90-1.26)	0.85 (0.74-0.97)	1.02 (0.94-1.11)
5	0.87 (0.74-1.02)	0.86 (0.70-1.05)	0.94 (0.76-1.17)	1.06 (0.91-1.23)	0.85 (0.75-0.97)	1.02 (0.94-1.11)
6	0.91 (0.82-1.01)	0.91 (0.81-1.03)	0.97 (0.85-1.09)	1.04 (0.94-1.14)	0.89 (0.81-0.98)	1.01 (0.95-1.08)
7 [Ref]	1.00	1.00	1.00	1.00	1.00	1.00
8	1.15 (0.98-1.34)	1.12 (0.96-1.30)	1.04 (0.89-1.21)	0.94 (0.81-1.09)	1.21 (1.03-1.43)	0.97 (0.86-1.09)
9	1.29 (0.95-1.76)	1.23 (0.93-1.63)	1.02 (0.76-1.38)	0.87 (0.65-1.17)	1.40 (1.01-1.94)	0.92 (0.69-1.24)
10	1.34 (0.89-2.02)	1.28 (0.88-1.86)	0.92 (0.61-1.37)	0.80 (0.54-1.17)	1.37 (0.89-2.11)	0.83 (0.54-1.30)
11	1.30 (0.79-2.14)	1.27 (0.81-1.98)	0.76 (0.47-1.23)	0.73 (0.46-1.14)	1.16 (0.70-1.92)	0.70 (0.42-1.16)
12	1.23 (0.68-2.21)	1.24 (0.73-2.08)	0.61 (0.35-1.07)	0.66 (0.39-1.11)	0.92 (0.52-1.64)	0.55 (0.32-0.97)
Marital status						
Married [Ref]	1.00	1.00	1.00	1.00	1.00	1.00
Divorced	1.14 (0.89-1.46)	1.02 (0.67-1.58)	1.00 (0.62-1.62)	1.16 (0.78-1.73)	8.89 (3.36-23.5)	3.37 (1.20-9.49)
Single	—	1.08 (0.85-1.37)	1.17 (0.93-1.48)	1.13 (0.92-1.38)	1.03 (0.82-1.29)	1.35 (1.06-1.72)
Parity						
0 [Ref]	1.00	1.00	1.00	1.00	1.00	1.00
1	1.09 (0.88-1.36)	1.16 (0.96-1.41)	0.86 (0.71-1.05)	0.89 (0.76-1.06)	1.07 (3.36-23.5)	3.37 (1.20-9.49)
≥2	0.70 (0.53-0.92)	0.81 (0.64-1.04)	0.67 (0.51-0.86)	0.73 (0.58-0.90)	1.03 (0.82-1.29)	1.35 (1.06-1.72)

GLMM adjusted urban areas adjusted for age [continuous variable], marital status and parity fit using via Laplacian approximation of maximum likelihood
‡Fitted using restricted cubic splines (RCS) with four knots at the 5th, 35th, 65th and 95th percentile and *Marital status coded as unmarried and married in 1994. NB. ∞Educational attainment measured as number of school years completed. OR=odds ratio; CI=Confidence interval

Table 8.4. Odds ratio and 95% confidence interval (CI) for fixed parameters from a GLMM that assessed the relationship between educational attainment and prevalent HIV infection among pregnant women in *rural* sentinel sites beginning 1994 through 2013

Educational attainment∞	1994 OR (95% CI)	1998 OR (95% CI)	2002 OR (95% CI)	2004 OR (95% CI)	2006 OR (95% CI)	2008 OR (95% CI)
0	0.58 (0.17- 2.02)	0.82 (0.27-2.45)	0.71 (0.27-1.87)	0.74 (0.29-1.90)	0.82 (0.36-1.83)	0.76 (0.37-1.54)
1	0.58 (0.17- 2.01)	0.82 (0.27-2.44)	0.71 (0.27-1.86)	0.75 (0.29-1.90)	0.82 (0.37-1.82)	0.76 (0.37-1.54)
2	0.59 (0.18- 1.98)	0.82 (0.28-2.39)	0.71 (0.28-1.83)	0.75 (0.30-1.87)	0.82 (0.37-1.80)	0.76 (0.37-1.54)
3	0.61 (0.20- 1.88)	0.83 (0.31-2.26)	0.73 (0.30-1.76)	0.76 (0.33-1.80)	0.83 (0.40-1.74)	0.77 (0.39-1.52)
4	0.65 (0.25- 1.71)	0.85 (0.36-2.01)	0.75 (0.35-1.62)	0.79 (0.38-1.66)	0.85 (0.44-1.63)	0.79 (0.42-1.46)
5	0.72 (0.35- 1.46)	0.89 (0.47-1.67)	0.80 (0.46-1.41)	0.85 (0.49-1.46)	0.88 (0.53-1.46)	0.82 (0.50-1.35)
6	0.83 (0.59- 1.18)	0.94 (0.69-1.29)	0.88 (0.67-1.17)	0.92 (0.70-1.21)	0.93 (0.70-1.24)	0.89 (0.67-1.19)
7 [Ref]	1.00	1.00	1.00	1.00	1.00	1.00
8	1.24 (0.95- 1.61)	1.05 (0.83-1.33)	1.16 (0.93-1.45)	1.05 (0.84-1.32)	1.08 (0.81-1.44)	1.13 (0.85-1.51)
9	1.55 (0.95- 2.53)	1.10 (0.71-1.70)	1.38 (0.92-2.08)	1.08 (0.71-1.65)	1.15 (0.69-1.92)	1.27 (0.76-2.12)
10	1.95 (0.95- 4.00)	1.14 (0.60-2.17)	1.65 (0.92-2.98)	1.09 (0.59-1.99)	1.22 (0.61-2.43)	1.41 (0.71-2.81)
11	2.46 (0.95- 6.33)	1.19 (0.51-2.78)	1.98 (0.91-4.30)	1.08 (0.49-2.39)	1.27 (0.54-3.02)	1.55 (0.66-3.64)
12	3.09 (0.95-10.04)	1.24 (0.43-3.57)	2.38 (0.91-6.23)	1.08 (0.40-2.88)	1.33 (0.47-3.76)	1.70 (0.61-4.73)
Marital status						
Married [Ref]	1.00	1.00	1.00	1.00	1.00	1.00
Divorced*	1.49 (0.78-1.49)	2.65 (1.46-4.83)	1.76 (0.88-3.54)	2.99 (1.44-6.18)	1.39 (0.47-4.16)	5.39 (2.40-12.0)
Single	—	1.21 (0.74-1.98)	1.11 (0.72-1.66)	1.10 (0.75-1.65)	0.93 (0.63 -1.38)	1.57 (1.06 -_2.33)
Parity						
0 [Ref]	1.00	1.00	1.00	1.00	1.00	1.00
1	1.02 (0.69-1.49)	1.08 (0.77-1.52)	0.97 (0.70-1.36)	0.93 (0.67-1.29)	0.80 (0.55-1.25)	0.77 (0.52-1.13)
≥2	0.70 (0.43-1.43)	0.57 (0.36-0.88)	0.89 (0.60-1.34)	0.78 (0.52-1.17)	0.79 (0.49-4.16)	0.74 (0.46-1.20)

GLMM adjusted rural areas adjusted for age [continuous variable], marital status and parity fit using via Laplacian approximation of maximum likelihood
‡Fitted using restricted cubic splines (RCS) with four knots at the 5th, 35th, 65th and 95th percentile and *Marital status coded as unmarried and married in 1994.
∞Educational attainment measured as number of school years completed. NB. OR=odds ratio; CI=Confidence interval

CHAPTER 9

COMMENTARY AND NEXT STEPS

Monitoring HIV incidence and prevalence is a key step in observing the progression and direction of the HIV epidemic, and invariably an important public health activity. Based on the ANC-HIV-SS data collected from pregnant women between 1994 and 2011, HIV prevalence among 15 to 24 year-olds dropped from 27 % in 1994 to 14.7% in 2011 in urban sites. The fall in HIV prevalence was less profound in rural sites where HIV prevalence declined from 10% in 1994 to 7.4% in 2011. The current report is based on analyses that used the UNAIDS-recommended age group (i.e., 15 to 24 year-olds) for approximating HIV incidence (i.e., number of new HIV infections). The observed decline in HIV prevalence in 15 to 24 year-olds are encouraging, assuming prevalent HIV infections in 15 to 24 year-olds provide a valid estimate of HIV incidence, and consequently indicative of a drop in the number of new HIV infections. Even though the HIV prevalence in the 15 to 24 year-olds has declined, the HIV infection burden among pregnant women in Zambia is still higher compared to western countries.[10, 14] The ANC-HIV-SS based HIV prevalence estimates, although informative, may be subject to biases, but the noted high HIV prevalence estimates highlights a lurking source of HIV infections for the general population, given the main route of HIV infections in Zambia is unprotected sexual intercourse, and pregnant women may represent sexually active population.

Noteworthy in this report and as in prior reports is that overall HIV prevalence trend analysis estimates enshrouded the heterogeneous HIV burden and prevalence trends

revealed by site-specific HIV prevalence trends analysis in both urban and rural areas.[24] For example, some sites exhibited potentially worrying upwards swings in HIV prevalence in 2011, and should trigger closer examination of site-specific HIV prevalence trends using robust statistical methods (e.g., restricted cubic splines for flexible modeling trends), including the data collection procedures, population structures of catchment and impact of change of HIV diagnostic criteria in 2011 on HIV identification. Examination of PBS-based HIV prevalence in the catchment areas of sentinel sites that displayed unstable HIV prevalence estimates may provide more enlightening explanation.[153]

The dwindling numbers of prevalent HIV infections observed between 1994 and 2011 underscores the need and potential benefit that may ensue from intensifying prevention messages that promotes avoidance of risky sexual behavior in young people, and should galvanize future interventions.[26] Evidence, although largely derived from cross-sectionally collected data, indicate that HIV-related risky sexual behavior initiated in adolescence and climaxes in young adulthood.[294, 384] Therefore, it might be more effective prevention-wise, to dwarf risky sexual behaviors that typify youthful sexual exuberance via “ABC” creeds of delayed onset of sexual intercourse among adolescents (Abstinence), reduced number of sexual companions (be faithful), and unswerving use of barrier to HIV infections (Condom) that worked remarkably in Uganda and Thailand.[385-387] Further, sussing factors that may be contributing to the observed upward swings in HIV prevalence in selected sites require focused research, including conducting population-based surveys in the catchment areas of sites with unstable HIV prevalence estimates to help understand the noted trends. Here as in prior literature,

decline in HIV prevalence may also stem from several factors including higher rates of AIDS-related mortality; out-migration of HIV infected people, and reduced fertility among HIV positive women.[153]

It is also possible that drastic changes in the population structure of the sentinel site catchment area might lead to changes in HIV prevalence, and possibly to unstable HIV prevalence estimates. Whether the reduction in the sample size per site in 2011 (360 versus 500 in earlier years) affected the HIV prevalence estimation process or use of different HIV test assays influenced HIV infection identification cannot be decided using the available data. Assuming strong influence stemming from site sample size reduction and HIV test assays change, one would have expected similar variances in the estimated HIV prevalence across all sites in 2011. However, smoothing the estimates using restricted cubic splines function for survey year yielded more conservative HIV prevalence trends estimates (Figure 6.8).

The analyses were reliant on secondary data, and therefore I acknowledge the dependence on variables (i.e., definitions and measurement) available in the ANC-HIV-SS data sets. Further, I cannot rule out errors and variances in the data that may arise from self-report nature of sociodemographic data contained in the ANC-HIV-SS data but assumed that data were accurately captured and recorded. Against this backdrop, it is reasonable to assert that the findings are valid to the extent that my assumptions of random distribution of errors in the ANC-HIV-SS data are defensible. Additionally, short of random sampling and longitudinal design, the large sample size, repeated cross-sectional design, and diverse geographic coverage of the study were strengths of the current study.

To avoid subjective categorization, the relationship between educational attainment as a continuous variable (i.e. number of schooling year completed modeled flexibly using restricted cubic spline function) and HIV prevalence was examined. Notably, all the studies included in my meta-analysis for specific aim 1 used categorical definition of educational attainment. Therefore, findings from analyses in which educational attainment is defined as a categorical variable may be different from findings from analyses educational attainment is expressed as continuous variable. My analysis also accounted for possible intra-site clustering that may arise from using ANC-HIV-SS data that was collected from multiple sites, and possibly yielded standard errors that incorporated possible intra-site clustering of pregnant women.

Significant protective association was noted for pregnant women who self-reported to have completed 12 schooling year compared to pregnant women who self-reported 7 schooling years in urban areas in 2008. No significant association between educational attainment and prevalent HIV infection in both urban and rural prior to 2008, although increasing educational attainment tended to be protective in pregnant women in urban beginning 2002. However, increasing educational attainment tended to be associated with increased odds of prevalent HIV infection among pregnant women in rural areas between 1994 to 2008.

One can speculate that in rural settings where HIV prevalence is lower than in urban areas, higher educational attainment may be linked to high risk behaviors (e.g., travel to urban areas, ability to set up short and long-term sexual relationships with strangers or new comers).[54] Similar results were reported by Yahya-Malima et al. (2007) in rural Tanzania based on population-based study, despite using categorical

definition of education attainment.[388] Although survival bias cannot be completely ruled out due to use of prevalent HIV infections, assessing the estimated odds ratio for education-HIV association in 15 to 24 year-olds in my analyses is likely to have minimized influence of survival bias: UNAIDS recommends approximation of the number of incident HIV infections by number of prevalent HIV infections in the 15 to 24 year-olds. Using ANC-HIV-SS data in 1994, 1998 and 2002, Sandoy et al. (2006) reported significant association between educational attainment and prevalent HIV infection. Their analysis included pregnant women aged 15 to 49 year, therefore more likely affected by survival bias than mine (i.e., based on 15 to 24 year-olds). Further, their analysis did not account for possible intra-site clustering, and they categorized continuous age and educational variable, whereas I used continuous variable.[45]

Meta-analysis based investigation of the association between educational attainment and prevalent HIV infection yielded pooled odds ratio whose 95% CI included null value of 1.0 (i.e., OR=1.18, 95% CI: 0.93, 1.50), but the relationship tended to be slightly protective when study year was used as an explanatory variable in a meta-regression (OR=0.91, 95% CI: 0.97, 1.07). The findings are consistent with Hargreaves et al.(2008) hypothesis of a waning association between educational attainment and prevalent HIV infection.

The odds of prevalent HIV infection were highest and significant among pregnant women in urban sites who self-reported to have been born between 1965 and 1979. Similar patterns among pregnant women in rural sites were observed but the odds of prevalent HIV infections were significant and more pronounced in women in the 1975-1979 birth cohort. The findings might support the view that pregnant women who had

attained sexual maturity around the time the HIV epidemic was emerging in SSA may have a different set of values regarding sexual behavior, hence the elevated odds of prevalent HIV infections. Encouragingly, pregnant women in recent birth cohorts in urban areas had lower odds of prevalent HIV infection. Whether this is a reflection of the greater intensity of preventive interventions or an augmentation of the view that younger people are more receptive to HIV preventive intervention is but a reasonable speculation.[310] Kayeyi et al (2013) have reported declining trends in pre-marital sex and multiple sexual relationship in Zambia among 15 to 24 year-olds based on the Sexual Risk Behavior Survey conducted between 2000 and 2009.[30] The findings of lower odds of prevalent HIV infections are encouraging and may trigger research into specific factors that could explain heightened (e.g., 1975-1979 birth cohort) and lowered (e.g., 1990-1996 birth cohort in urban areas) odds of prevalent HIV infection.

9.1. Next steps

Performing separate meta-analyses on the association between educational attainment and prevalent HIV infection infections using studies conducted in men and in non-pregnant women may provide complement findings from the meta-analysis conbducted in pregnant women. Age, period and cohort analyses using models (CCREM) that incorporate sexual behavioral data based on the Zambia Sexual Behavior Survey data for 1998, 2000, 2003, 2005, and 2009 may elucidate the age, period, and cohort effects observed in the ANC-HIV-SS data. Further, directly measured HIV incidence estimates would provide more informative means of monitoring HIV progression than HIV prevalence estimates, especially in generalized epidemic setting,

where increasing population of HIV-infected persons due to expanding access to cART, might complicate interpretation of HIV prevalence trends.[298] Therefore, HIV prevalence data, although useful in guiding health service delivery, may become less informative in monitoring progression of HIV with persons infected in childhood growing into the 15 to 24 year-old population. Against the backdrop, future research should focus on using methods (i.e., under validation) that are able to distinguish established and new HIV infections based on biomarkers in cross-sectionally collected blood specimen (e.g., BED-CEIA and avidity assay).

REFERENCES

1. UNAIDS, *Together we will end AIDS*, 2012.

2. Fowler, M.G., Gable AR, Lampe MA, Etima M, Owor M, *Perinatal HIV and its prevention: progress toward an HIV-free generation.* Clinical Perinatology, 2010. **37**(4): p. 699-719.

3. Simon, V., D.D. Ho, and Q. Abdool Karim, *HIV/AIDS epidemiology, pathogenesis, prevention, and treatment.* Lancet, 2006. **368**(9534): p. 489-504.

4. UNAIDS, *UNAIDS report on the global AIDS epidemic*, 2010.

5. Mofenson, L.M. and M.G. Fowler, *Interruption of materno-fetal transmission.* AIDS, 1999. **13 Suppl A**: p. S205-14.

6. Fowler, M.G., et al., *Reducing the risk of mother-to-child human immunodeficiency virus transmission: past successes, current progress and challenges, and future directions.* American journal of obstetrics and gynecology, 2007. **197**(3 Suppl): p. S3-9.

7. Hira, S.K., et al., *Perinatal transmission of HIV-I in Zambia.* BMJ, 1989. **299**(6710): p. 1250-2.

8. UNAIDS, *UNAIDS WORLD AIDS DAY REPORT*, 2011, Joint United Nations Programme on HIV/AIDS.

9. Cavarelli, M. and G. Scarlatti, *Human immunodeficiency virus type 1 mother-to-child transmission and prevention: successes and controversies.* Journal of internal medicine, 2011. **270**(6): p. 561-79.

10. UNAIDS, *UNAIDS report on the global AIDS epidemic*, 2012.

11. World Bank, *Fact Sheet: Infrastructure in Sub-Saharan Africa*, 2013.

12. Sarngadharan, M.G., et al., *Antibodies reactive with human T-lymphotropic retroviruses (HTLV-III) in the serum of patients with AIDS.* Science, 1984. **224**(4648): p. 506-8.

13. UNAIDS, *AIDS epidemic update*, 2009: Geneva, Switzerland.

14. UNAIDS, *World AIDS Day Report*, 2012.

15. Poteat, T., et al., *HIV risk among MSM in Senegal: a qualitative rapid assessment of the impact of enforcing laws that criminalize same sex practices.* PloS one, 2011. **6**(12): p. e28760.

16. Baral, S., et al., *Assessing Priorities for Combination HIV Prevention Research for Men Who have Sex with Men (MSM) in Africa.* AIDS and behavior, 2012.

17. Baral, S., et al., *Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis.* The Lancet infectious diseases, 2012. **12**(7): p. 538-49.

18. Sabin, M.L., et al., *HIV in men who have sex with men in sub-Saharan Africa.* The Lancet infectious diseases, 2012. **12**(7): p. 505-6.

19. National HIV/AIDS/STI/TB Council of Zambia, *HIV Prevention Response and Modes of Transmission Analysis*, 2009.

20. Vermund, S.H., K.L. Allen, and Q.A. Karim, *HIV-prevention science at a crossroads: advances in reducing sexual risk.* Current opinion in HIV and AIDS, 2009. **4**(4): p. 266-73.

21. Joint United Nations Programme on AIDS and World Health Organization, *AIDS Epidemic update*, 2006: Geneva, Switzerland.

22. Foundation, K.F., *U.S. Global Health Policy*, 2011.

23. UNAIDS, *GLOBAL HIV/AIDS RESPONSE:Epidemic update and health sector progress towards Universal Access*, 2011.

24. Kayeyi, N., et al., *Decline in HIV prevalence among young women in Zambia: national-level estimates of trends mask geographical and socio-demographic differences.* PloS one, 2012. **7**(4): p. e33652.

25. Poundstone, P.E., S. A. Strathdee, and D. D. Celentano, *The Social Epidemiology of Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome.* Epidemiologic reviews, 2004. **28**.

26. Vermund, S.H. and R.J. Hayes, *Combination prevention: new hope for stopping the epidemic.* Current HIV/AIDS reports, 2013. **10**(2): p. 169-86.

27. Perez, J.L. and R.D. Moore, *Greater effect of highly active antiretroviral therapy on survival in people aged > or =50 years compared with younger people in an urban observational cohort.* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2003. **36**(2): p. 212-8.

28. Hentrich, M., et al., *Highly active antiretroviral therapy (HAART) improves survival in HIV-associated Hodgkin's disease: results of a multicenter study.* Annals of oncology : official journal of the European Society for Medical Oncology / ESMO, 2006. **17**(6): p. 914-9.

29. Vermund, S.H. and R.J. Hayes, *Combination Prevention: New Hope for Stopping the Epidemic.* Current HIV/AIDS reports, 2013.

30. Kayeyi, N., et al., *Decline in Sexual Risk Behaviours among Young People in Zambia (2000-2009): Do Neighbourhood Contextual Effects Play a Role?* PloS one, 2013. **8**(5): p. e64881.

31. Levin, B.R., J.J. Bull, and F.M. Stewart, *Epidemiology, evolution, and future of the HIV/AIDS pandemic.* Emerging infectious diseases, 2001. **7**(3 Suppl): p. 505-11.

32. Levin, B.R., J.J. Bull, and F.M. Stewart, *The intrinsic rate of increase of HIV/AIDS: epidemiological and evolutionary implications.* Mathematical biosciences, 1996. **132**(1): p. 69-96.

33. Kurth, A.E., et al., *Combination HIV prevention: significance, challenges, and opportunities.* Current HIV/AIDS reports, 2011. **8**(1): p. 62-72.

34. Zambia National HIV/AIDS/STI/TB Council, *HIV Prevention Response and Modes of Transmission Analysis*, 2009.

35. United States Census bureau. *States and Country Quick facts*. [Report] 2013 14th March 2013 [cited 2013 5th May 2013].

36. Central Statistical Office (CSO), et al. *Zambia Demographic and Health Survey 2007*. 2009 [cited 2012 June 6]; Available from: http://www.measuredhs.com/pubs/pdf/FR211/FR211%5Brevised-05-12-2009%5D.pdf.

37. Central Statistics Office Zambian Ministry of Health & ORC Macro, *Zambia Demographic and Health Survey 2007*, 2008: Calverton, Maryland,USA,.

38. Rothman, K.J. and C. Poole, *A strengthening programme for weak associations.* International journal of epidemiology, 1988. **17**(4): p. 955-9.

39. Fylkesnes, K., et al., *Studying dynamics of the HIV epidemic: population-based data compared with sentinel surveillance in Zambia.* AIDS, 1998. **12**(10): p. 1227-34.

40. Fylkesnes, K., et al., *Declining HIV prevalence and risk behaviours in Zambia: evidence from surveillance and population-based surveys.* AIDS, 2001. **15**(7): p. 907-16.

41. Fylkesnes, K., Musonda, R. M., Kasumba, K.,Ndhlovu, Z.,Mluanda, F.,Kaetano, L.,Chipaila, C. C.,, *The HIV epidemic in Zambia: socio-demographic prevalence patterns and indications of trends among childbearing women.* AIDS, 1997. **11**(3): p. 339-45.

42. Mmbaga, E.J., et al., *Education attainment and the risk of HIV-1 infections in rural Kilimanjaro Region of Tanzania, 1991-2005: a reversed association.* Sexually transmitted diseases, 2007. **34**(12): p. 947-53.

43. Mmbaga, E.J., et al., *Trends in HIV-1 prevalence and risk behaviours over 15 years in a rural population in Kilimanjaro region of Tanzania.* AIDS research and therapy, 2007. **4**: p. 23.

44. Kron, J., *In Uganda, an AIDS Success Story Comes Undone,* in *New York Times*2012, New York Times: USA.

45. Sandoy, I.F., et al., *Antenatal clinic-based HIV prevalence in Zambia: declining trends but sharp local contrasts in young women.* Trop Med Int Health, 2006. **11**(6): p. 917-28.

46. Michelo, C., I.F. Sandoy, and K. Fylkesnes, *Marked HIV prevalence declines in higher educated young people: evidence from population-based surveys (1995-2003) in Zambia.* AIDS, 2006. **20**(7): p. 1031-8.

47. Sandoy, I.F., et al., *Associations between sexual behaviour change in young people and decline in HIV prevalence in Zambia.* BMC Public Health, 2007. **7**: p. 60.

48. Winkleby, M.A., et al., *Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease.* American journal of public health, 1992. **82**(6): p. 816-20.

49. Williams, D.R., *Race, socioeconomic status, and health. The added effects of racism and discrimination.* Annals of the New York Academy of Sciences, 1999. **896**: p. 173-88.

50. Hargreaves, J.R., et al., *The association between school attendance, HIV infection and sexual behaviour among young people in rural South Africa.* Journal of epidemiology and community health, 2008. **62**(2): p. 113-9.

51. Hargreaves, J.R., *Socioeconomic status and risk of HIV infection in an urban population in Kenya.* Tropical medicine & international health : TM & IH, 2002. **7**(9): p. 793-802.

52. Hargreaves, J.R. and J.R. Glynn, *Educational attainment and HIV-1 infection in developing countries: a systematic review.* Tropical medicine & international health : TM & IH, 2002. **7**(6): p. 489-98.

53. Ross, C.E. and J. Mirowsky, *Refining the association between education and health: the effects of quantity, credential, and selectivity.* Demography, 1999. **36**(4): p. 445-60.

54. Smith, J., et al., *Education attainment as a predictor of HIV risk in rural Uganda: results from a population-based study.* Int J STD AIDS, 1999. **10**(7): p. 452-9.

55. Kayeyi, N., I.F. Sandoy, and K. Fylkesnes, *Effects of neighbourhood-level educational attainment on HIV prevalence among young women in Zambia.* BMC public health, 2009. **9**: p. 310.

56. Fylkesnes, K., et al., *The HIV epidemic in Zambia: socio-demographic prevalence patterns and indications of trends among childbearing women.* AIDS, 1997. **11**(3): p. 339-45.

57. Stroeken, K., et al., *HIV among out-of-school youth in Eastern and Southern Africa: a review.* AIDS care, 2012. **24**(2): p. 186-94.

58. Zuilkowski, S.S. and M.C. Jukes, *The impact of education on sexual behavior in sub-Saharan Africa: a review of the evidence.* AIDS care, 2012. **24**(5): p. 562-76.

59. Jhun, H.J., H. Kim, and S.I. Cho, *Time trend and age-period-cohort effects on acute myocardial infarction mortality in Korean adults from 1988 to 2007.* J Korean Med Sci, 2011. **26**(5): p. 637-41.

60. Holford, T.R., *Understanding the effects of age, period and cohort on incidence and mortality rates.* Annual Review of Public Health, 1991. **12**: p. 425-457.

61. Holford, T.R., *The estimation of age, period and cohort effects for vital rates.* Biometrics, 1983. **39**(2): p. 311-24.

62. Szklo, M. and F.J. Nieto, *EPIDEMIOLOGY: Beyond the Basics*. 2nd ed2007, Boston: Jones and Bartlett Publishers.

63. Land, Y., *Age-Period-Cohort Analysis of Cross-Sectional Surveys Fixed or Random Effects?* Sociological Methods and Research, 2008. **36**(3): p. 297-326.

64. Clayton, D. and E. Schifflers, *Models for temporal variation in cancer rates. I: Age-period and age-cohort models.* Statistics in medicine, 1987. **6**(4): p. 449-67.

65. Chen, X., et al., *Secular trends in adolescent never smoking from 1990 to 1999 in California: an age-period-cohort analysis.* American journal of public health, 2003. **93**(12): p. 2099-104.

66. Yang, Y.K.C.L., *Age-period-cohort analysis of repeated cross-section surveys:fixed or random effects.* Sociological Methods and Research, 2008. **36**(3): p. 297-326.

67. Yang, Y., W.J. Fu, and K.C. Land, *A methodological comparison of age-period-cohort models:the intrinsic estimator and conventional generalized linear models.* Sociological Methodology, 2004. **34**(1): p. 75-110.

68. Yang, Y. and K.C. Land, *A mixed model approach to age-period-cohort analysis of repeated cross-section surveys, with an application on trends in verbal test scores.* Sociological Methodology, 2006. **36**(1): p. 75-97.

69. Reither, E.N., R.M. Hauser, and Y. Yang, *Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States.* Social science & medicine, 2009. **69**(10): p. 1439-48.

70. Morgenthau, J.E. and N.J. Sokoloff, *The sexual revolution: myth or fact?* Pediatric clinics of North America, 1972. **19**(3): p. 779-89.

71. Green, E.C., et al., *Uganda's HIV prevention success: the role of sexual behavior change and the national response.* AIDS and behavior, 2006. **10**(4): p. 335-46; discussion 347-50.

72. Houweling, H., et al., *An age-period-cohort analysis of 50,875 AIDS cases among injecting drug users in Europe.* International journal of epidemiology, 1999. **28**(6): p. 1141-8.

73. Castilla, J., M. Pollan, and G. Lopez-Abente, *The AIDS epidemic among Spanish drug users: a birth cohort-associated phenomenon.* American journal of public health, 1997. **87**(5): p. 770-4.

74. Yang, Y. and K.C. Land, *Age-Period-Cohort Analysis: New Models, Methods and Empirical Applications*. First ed, ed. N. Keiding, B.J.T. Morgan, and P.H. Wikle2013, New York: CRC Press.

75. Rosinska, M. and A. Zielinski, *Recent increase in HIV rate by age, cohort, period analysis of surveillance data suggests changes in HIV epidemiology in Poland.* Central European journal of public health, 2011. **19**(3): p. 123-7.

76. Houweling, H., et al., *A birth cohort analysis of AIDS in Europe: high incidence among young persons at risk.* AIDS, 1998. **12**(1): p. 85-93.

77. Wand, H., et al., *Characterizing trends in HIV infection among men who have sex with men in Australia by birth cohorts: results from a modified back-projection method.* Journal of the International AIDS Society, 2009. **12**: p. 19.

78. Gakidou, E., et al., *Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis.* Lancet, 2010. **376**(9745): p. 959-74.

79. Nations, U., *Millenium Development Goals*, 2010: New York.

80. Smith, J., et al., *Education attainment as a predictor of HIV risk in rural Uganda: results from a population-based study.* International journal of STD & AIDS, 1999. **10**(7): p. 452-9.

81. Hargreaves, J.R., et al., *Systematic review exploring time trends in the association between educational attainment and risk of HIV infection in sub-Saharan Africa.* AIDS, 2008. **22**(3): p. 403-14.

82. Monge, S., et al., *Educational level and HIV disease progression before and after the introduction of HAART: a cohort study in 989 HIV seroconverters in Spain.* Sexually transmitted infections, 2011. **87**(7): p. 571-6.

83. Hegazi, A., et al., *Literacy, education and adherence to antiretroviral therapy in The Gambia.* AIDS care, 2010. **22**(11): p. 1340-5.

84. Buve, A., et al., *Multicentre study on factors determining differences in rate of spread of HIV in sub-Saharan Africa: methods and prevalence of HIV infection.* AIDS, 2001. **15 Suppl 4**: p. S5-14.

85. Buve, A., et al., *The multicentre study on factors determining the differential spread of HIV in four African cities: summary and conclusions.* AIDS, 2001. **15 Suppl 4**: p. S127-31.

86. Glynn, J.R., et al., *Why do young women have a much higher prevalence of HIV than young men? A study in Kisumu, Kenya and Ndola, Zambia.* AIDS, 2001. **15 Suppl 4**: p. S51-60.

87. Gouws, E., et al., *The epidemiology of HIV infection among young people aged 15-24 years in southern Africa.* AIDS, 2008. **22 Suppl 4**: p. S5-16.

88. Gouws, E., V. Mishra, and T.B. Fowler, *Comparison of adult HIV prevalence from national population-based surveys and antenatal clinic surveillance in countries with generalised epidemics: implications for calibrating surveillance data.* Sexually transmitted infections, 2008. **84 Suppl 1**: p. i17-i23.

89. Melbye, M., et al., *Evidence for heterosexual transmission and clinical manifestations of human immunodeficiency virus infection and related conditions in Lusaka, Zambia.* Lancet, 1986. **2**(8516): p. 1113-5.

90. Vermund, S.H. and A.J. Leigh-Brown, *The HIV Epidemic: High-Income Countries.* Cold Spring Harbor Perspectives in Medicine, 2012. **2**(5): p. a007195.

91. Sandala, L., et al., *'Dry sex' and HIV infection among women attending a sexually transmitted diseases clinic in Lusaka, Zambia.* AIDS, 1995. **9 Suppl 1**: p. S61-8.

92. Coombs, R.W., P.S. Reichelderfer, and A.L. Landay, *Recent observations on HIV type-1 infection in the genital tract of men and women.* AIDS, 2003. **17**(4): p. 455-80.

93. Moss GB, C.D., D'Costa L, Plummer FA, Ndinya-Achola JO, Reilly M, Holmes KK, Piot P, Maitha GM, Hillier SL, et al, *Association of cervical ectopy with heterosexual transmission of human immunodeficiency virus: results of a study of couples in Nairobi, Kenya.* J Infect Dis., 1991 **164**((3)): p. 588-91.

94. Birdthistle, I.J., et al., *Child sexual abuse and links to HIV and orphanhood in urban Zimbabwe.* Journal of epidemiology and community health, 2011. **65**(12): p. 1075-82.

95. Boerma, J.T.W., S.Wr, *Integrating Demographic and Epidemiological Approaches to Research on HIV/AIDS:The Proximate-Determinants Framework.* Journal of Infectious Diseases, 2005. **191**(Suppl 1): p. S61-71.

96. Afonso, J.M., M.G. Morgado, and G. Bello, *Evidence of multiple introductions of HIV-1 subtype C in Angola.* Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 2012. **12**(7): p. 1458-1465.

97. Spira, S., et al., *Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance.* The Journal of antimicrobial chemotherapy, 2003. **51**(2): p. 229-40.

98. Lingappa, J.R., et al., *Genomewide association study for determinants of HIV-1 acquisition and viral set point in HIV-1 serodiscordant couples with quantified virus exposure.* PloS one, 2011. **6**(12): p. e28632.

99. Kaul, R., et al., *Biological factors that may contribute to regional and racial disparities in HIV prevalence.* American journal of reproductive immunology, 2011. **65**(3): p. 317-24.

100. Susser, I. and Z. Stein, *Culture, sexuality, and women's agency in the prevention of HIV/AIDS in southern Africa.* American journal of public health, 2000. **90**(7): p. 1042-8.

101. Meekers, D. and A.E. Calves, *'Main' girlfriends, girlfriends, marriage, and money: the social context of HIV risk behaviour in sub-Saharan Africa.* Health Transit Rev, 1997. **7 Suppl**: p. 361-75.

102. Guest, G., et al., *Fear, hope and social desirability bias among women at high risk for HIV in West Africa.* J Fam Plann Reprod Health Care, 2005. **31**(4): p. 285-7.

103. Bryans, A., F. Cornish, and J. McIntosh, *The potential of ecological theory for building an integrated framework to develop the public health contribution of health visiting.* Health & social care in the community, 2009. **17**(6): p. 564-72.

104. Stokols, D., *Translating social ecological theory into guidelines for community health promotion.* American journal of health promotion : AJHP, 1996. **10**(4): p. 282-98.

105. Barnighausen, T. and F. Tanser, *Rethinking the role of the local community in HIV epidemic spread in sub-Saharan Africa: a proximate-determinants approach.* HIV therapy, 2009. **3**(5): p. 435-445.

106. Cote, A.M., et al., *Transactional sex is the driving force in the dynamics of HIV in Accra, Ghana.* AIDS, 2004. **18**(6): p. 917-25.

107. Chen, L., et al., *Sexual risk factors for HIV infection in early and advanced HIV epidemics in sub-Saharan Africa: systematic overview of 68 epidemiological studies.* PloS one, 2007. **2**(10): p. e1001.

108. Cohen, M.S., et al., *Reduction of concentration of HIV-1 in semen after treatment of urethritis: implications for prevention of sexual transmission of HIV-1. AIDSCAP Malawi Research Group.* Lancet, 1997. **349**(9069): p. 1868-73.

109. Freeman, E.E., et al., *Herpes simplex virus 2 infection increases HIV acquisition in men and women: systematic review and meta-analysis of longitudinal studies.* AIDS, 2006. **20**(1): p. 73-83.

110. Weiss, H.A., M.A. Quigley, and R.J. Hayes, *Male circumcision and risk of HIV infection in sub-Saharan Africa: a systematic review and meta-analysis.* AIDS, 2000. **14**(15): p. 2361-70.

111. Siegfried, N., et al., *HIV and male circumcision--a systematic review with assessment of the quality of studies.* The Lancet infectious diseases, 2005. **5**(3): p. 165-73.

112. Vermund, S.H., *Limitations in characterization of heterosexual HIV transmission risk: commentary on the models of Downs and De Vincenzi.* Journal of acquired immune deficiency syndromes and human retrovirology : official publication of the International Retrovirology Association, 1996. **11**(4): p. 385-7.

113. Grosskurth, H., et al., *Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomised controlled trial.* Lancet, 1995. **346**(8974): p. 530-6.

114. Korenromp, E.L., et al., *Determinants of the impact of sexually transmitted infection treatment on prevention of HIV infection: a synthesis of evidence from the Mwanza, Rakai, and Masaka intervention trials.* The Journal of infectious diseases, 2005. **191 Suppl 1**: p. S168-78.

115. Hayes, R., et al., *Treatment of sexually transmitted infections for HIV prevention: end of the road or new beginning?* AIDS, 2010. **24 Suppl 4**: p. S15-26.

116. Ng, B.E., et al., *Population-based biomedical sexually transmitted infection control interventions for reducing HIV infection.* Cochrane database of systematic reviews, 2011(3): p. CD001220.

117. Kirunga, C.T. and J.P. Ntozi, *Socio-economic determinants of HIV serostatus: a study of Rakai District, Uganda.* Health Transit Rev, 1997. **7 Suppl**: p. 175-88.

118. Quigley, M., et al., *Sexual behaviour patterns and other risk factors for HIV infection in rural Tanzania: a case-control study.* AIDS, 1997. **11**(2): p. 237-48.

119. Gomez, C.A., *Preventing HIV in U.S. women and girls: a call for social action.* Womens Health Issues, 2011. **21**(6 Suppl): p. S287-94.

120. Kim, J., et al., *Exploring the role of economic empowerment in HIV prevention.* AIDS, 2008. **22 Suppl 4**: p. S57-71.

121. Piot, R.R., S, *Squaring the circle: AIDS, poverty, and human development.* PLoS medicine, 2007. **4**(e314).

122. Kongnyuy, E.J., et al., *Wealth and sexual behaviour among men in Cameroon.* BMC international health and human rights, 2006. **6**: p. 11.

123. Kirunga, C.T. and J.P. Ntozi, *Socio-economic determinants of HIV serostatus: a study of Rakai District, Uganda.* Health transition review : the cultural, social, and behavioural determinants of health, 1997. **7 Suppl**: p. 175-88.

124. Berhan, Y. and A. Berhan, *Meta-analysis on risky sexual behaviour of men: Consistent findings from different parts of the world.* AIDS care, 2012.

125. Ross, C.E. and J. Mirowsky, *Does employment affect health?* J Health Soc Behav, 1995. **36**(3): p. 230-43.

126. Braveman, P.A., et al., *Socioeconomic status in health research: one size does not fit all.* JAMA : the journal of the American Medical Association, 2005. **294**(22): p. 2879-88.

127. Peterson C, S.A., *Personal control and health promotion.* Social Science and Medicine, 1989. **28**(8): p. 819-28.

128. Msamanga, G., et al., *Socio-economic and demographic factors associated with prevalence of HIV infection among pregnant women in Dar es Salaam, Tanzania.* East African medical journal, 2006. **83**(6): p. 311-21.

129. Niyonsenga, T., et al., *Measuring Socioeconomic Inequality in the Incidence of AIDS: Rural-Urban Considerations.* AIDS Behav, 2012.

130. Msisha, W.M., et al., *Socioeconomic status and HIV seroprevalence in Tanzania: a counterintuitive relationship.* International journal of epidemiology, 2008. **37**(6): p. 1297-303.

131. Williams, D.R., et al., *Race, socioeconomic status, and health: complexities, ongoing challenges, and research opportunities.* Annals of the New York Academy of Sciences, 2010. **1186**: p. 69-101.

132. Cancer epidemiology, *No association between education level and pancreatic cancer incidence in the European prospective investatigation into cancer and nutrition.* International Journal of Cancer Epidemiology, Detection, and Prevention, 2010. **34**.

133. Fuchs, V.R., *Setting priorities in health education and promotion.* Health Aff (Millwood), 1983. **2**(4): p. 56-69.

134. Kemptner, D., H. Jurges, and S. Reinhold, *Changes in compulsory schooling and the causal effect of education on health: evidence from Germany.* Journal of health economics, 2011. **30**(2): p. 340-54.

135. Farrell, P. and V.R. Fuchs, *Schooling and health: the cigarette connection.* Journal of health economics, 1982. **1**(3): p. 217-30.

136. Kilian, A.H., et al., *Reductions in risk behaviour provide the most consistent explanation for declining HIV-1 prevalence in Uganda.* AIDS, 1999. **13**(3): p. 391-8.

137. Senkoro, K.P., et al., *HIV incidence and HIV-associated mortality in a cohort of factory workers and their spouses in Tanzania, 1991 through 1996.* Journal of acquired immune deficiency syndromes, 2000. **23**(2): p. 194-202.

138. Gregson, S., et al., *A rural HIV epidemic in Zimbabwe? Findings from a population-based survey.* Int J STD AIDS, 2001. **12**(3): p. 189-96.

139. Glynn, J.R., et al., *Does increased general schooling protect against HIV infection? A study in four African cities.* Trop Med Int Health, 2004. **9**(1): p. 4-14.

140. Jukes, M.S., S. Bundy,D., *Education and vulnerability: the role of schools in protecting young women and girls from HIV in southern Africa.* AIDS, 2008. **22**: p. S41-S56.

141. Johnson, L.F., et al., *The effect of educational attainment and other factors on HIV risk in South African women: results from antenatal surveillance, 2000-2005.* AIDS, 2009. **23**(12): p. 1583-8.

142. Cohen, A.K., et al., *Education and obesity at age 40 among American adults.* Social science & medicine, 2013. **78**: p. 34-41.

143. de Walque, D., et al., *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. **10**(10): p. 993-1001.

144. Birdthistle, I., et al., *Is education the link between orphanhood and HIV/HSV-2 risk among female adolescents in urban Zimbabwe?* Social science & medicine, 2009. **68**(10): p. 1810-8.

145. Glynn, J.R., et al., *Age at menarche, schooling, and sexual debut in northern Malawi.* PloS one, 2010. **5**(12): p. e15334.

146. Sandoy, I.F., et al., *Antenatal clinic-based HIV prevalence in Zambia: declining trends but sharp local contrasts in young women.* Tropical medicine & international health : TM & IH, 2006. **11**(6): p. 917-28.

147. Fylkesnes, K., et al., *HIV infection among antenatal women in Zambia, 1990-1993.* AIDS, 1996. **10**(5): p. 555-6.

148. Pedhazur, E.J., *Multiple regression in behavioral research: Explanation and prediction*1982, New York, N.Y. 10017: CBC College Publishing.

149. Harrell, F.E., Jr., *Regression Modeling Strategies*. Spirnger series in statistics2001, New York: Springer.

150. Bakhshi, E., et al., *Let continuous outcome variables remain continuous.* Computational and mathematical methods in medicine, 2012. **2012**: p. 639124.

151. Turner, E.L., J.E. Dobson, and S.J. Pocock, *Categorisation of continuous risk factors in epidemiological publications: a survey of current practice.* Epidemiologic perspectives & innovations : EP+I, 2010. **7**: p. 9.

152. Michelo, C., et al., *Steep HIV prevalence declines among young people in selected Zambian communities: population-based observations (1995-2003).* BMC Public Health, 2006. **6**: p. 279.

153. Michelo, C., I. Sandoy, and K. Fylkesnes, *Antenatal clinic HIV data found to underestimate actual prevalence declines: evidence from Zambia.* Tropical medicine & international health : TM & IH, 2008. **13**(2): p. 171-9.

154. Central Statistics Office, *2010 Census of Population and Housing Preliminary Report.* 2011.

155. Central Statistics Office, *Summary Report for the 2000 Census of Population and Housing*, 2003, Central Statistics Office, Zambia: Lusaka.

156. Buve, A., *AIDS and hospital bed occupancy: an overview.* Tropical medicine & international health : TM & IH, 1997. **2**(2): p. 136-9.

157. Ministry of Health, *The Zambia Antental Clinic Sentinel surveillance Report 1994-2006*, 2006, Ministry of Health Lusaka.

158. Holtgrave, D.R., H.I. Hall, and J. Prejean, *HIV Transmission Rates in the United States, 2006-2008.* The open AIDS journal, 2012. **6**: p. 26-8.

159. Morison, L., et al., *Commercial sex and the spread of HIV in four cities in sub-Saharan Africa.* AIDS, 2001. **15 Suppl 4**: p. S61-9.

160. Kamanga, J.C. *Is This Enough Evidence to Support Targeted Intervention to Reduce Burden of STI's and*

HIV among Female Sex Workers. in *14th International Conference on AIDS and STIs in Africa, Abuja, Nigeria, 12/4-9*. 2005.

161. Simooya, O.O., et al., *'Behind walls': a study of HIV risk behaviours and seroprevalence in prisons in Zambia.* AIDS, 2001. **15**(13): p. 1741-4.

162. Hira, S.K., et al., *Condom and nonoxynol-9 use and the incidence of HIV infection in serodiscordant couples in Zambia.* International journal of STD & AIDS, 1997. **8**(4): p. 243-50.

163. Stephenson, R., et al., *The influence of motivational messages on future planning behaviors among HIV concordant positive and discordant couples in Lusaka, Zambia.* AIDS care, 2008. **20**(2): p. 150-60.

164. Kapina, M., et al., *HIV incidence rates and risk factors for urban women in Zambia: preparing for a microbicide clinical trial.* Sexually transmitted diseases, 2009. **36**(3): p. 129-33.

165. Celum, C., et al., *Effect of aciclovir on HIV-1 acquisition in herpes simplex virus 2 seropositive women and men who have sex with men: a randomised, double-blind, placebo-controlled trial.* Lancet, 2008. **371**(9630): p. 2109-19.

166. Heffron, R., et al., *High prevalent and incident HIV-1 and herpes simplex virus 2 infection among male migrant and non-migrant sugar farm workers in Zambia.* Sexually transmitted infections, 2011. **87**(4): p. 283-8.

167. Bello, G.A., J. Chipeta, and J. Aberle-Grasse, *Assessment of trends in biological and behavioural surveillance data: is there any evidence of declining HIV prevalence or incidence in Malawi?* Sexually transmitted infections, 2006. **82 Suppl 1**: p. i9-13.

168. Stringer, E.M., et al., *Declining HIV prevalence among young pregnant women in Lusaka, Zambia.* Bulletin of the World Health Organization, 2008. **86**(9): p. 697-702.

169. Mpairwe, H., et al., *HIV risk perception and prevalence in a program for prevention of mother-to-child HIV transmission: comparison of women who accept voluntary counseling and testing and those tested anonymously.* Journal of acquired immune deficiency syndromes, 2005. **39**(3): p. 354-8.

170. Smith, P.W.F., A. Berrington, and P. Sturgis, *A comparison of graphical and structural equation for analysis of longitudinal data*. Methodology of longitudinal surveys2009, UK: John Wiley and Sons.

171. USAID, *Demographic and Health Survey: sampling and household listing manual*, 2012, MEASURE DHS/ICF International

Calverton, Maryland USA.

172. Razak, F., D.J. Corsi, and S.V. Subramanian, *Change in the body mass index distribution for women: analysis of surveys from 37 low- and middle-income countries.* PLoS medicine, 2013. **10**(1): p. e1001367.

173. Mastro, T.D., et al., *Estimating HIV Incidence in Populations Using Tests for Recent Infection: Issues, Challenges and the Way Forward.* Journal of HIV AIDS surveillance & epidemiology, 2010. **2**(1): p. 1-14.

174. Brookmeyer, R., et al., *Estimation of HIV incidence using multiple biomarkers.* American journal of epidemiology, 2013. **177**(3): p. 264-72.

175. Laeyendecker, O., et al., *HIV incidence determination in the United States: a multiassay approach.* The Journal of infectious diseases, 2013. **207**(2): p. 232-9.

176. Mangione-Smith, R., et al., *An observational study of antibiotic prescribing behavior and the Hawthorne effect.* Health services research, 2002. **37**(6): p. 1603-23.

177. Roethlisberger, F.H. and W.J. Dickson., *Management and the Worker*, 1939: Cambridge,

MA.

178. McCarney, R., et al., *The Hawthorne Effect: a randomised, controlled trial.* BMC medical research methodology, 2007. **7**: p. 30.

179. Murray, M., et al., *The Hawthorne effect in the measurement of adolescent smoking.* Journal of epidemiology and community health, 1988. **42**(3): p. 304-6.

180. World Health Organization, *United nations Joint Programme on AIDS second generation surveillance for HIV: the next decade*, WHO/UNAIDS, Editor 2006: Geneva.

181. Mertens, T., et al., *Prevention indicators for evaluating the progress of national AIDS programmes.* AIDS, 1994. **8**(10): p. 1359-69.

182. Ministry of Health, *National HIV/syphilis sentinel surveillance protocol*, 2010, Ministry of Health of Zambia

183. Fylkesnes, K., et al., *The HIV epidemic in Zambia: socio-demographic prevalence patterns and indications of trends among childbearing women*

HIV infection among antenatal women in Zambia, 1990-1993

The first Zambian population-based HIV survey: saliva-based testing is accurate and acceptable

Studying dynamics of the HIV epidemic: population-based data compared with sentinel surveillance in Zambia

Declining HIV prevalence and risk behaviours in Zambia: evidence from surveillance and population-based surveys. AIDS, 1997. **11**(3): p. 339-45.

184. Colebunders, R.L. and W.L. Heyward, *Surveillance of AIDS and HIV infection:Opportunities and challenges.* Health Policy, 1990. **15**: p. 1-11.

185. Joint United nations Program on AIDS/World Health Organization (UNAIDS/WHO), *Guidelines for conducting HIV sentinel serosurveys among pregnant women and other groups*, 2003.

186. Joint United Nations Programme on AIDS and World Health Organization working group on Global HIV/AIDS on and STI Surveillance, *Guidelines for conducting HIV sentinel serosurvey among pregnant women and other groups*, 2003.

187. Centers for Disease Control and Prevention, *Epi Info*, 2008: Atlanta, U.S.A.

188. R Core Team (2012), *R: A language and environment for statistical computing* in *R Foundation for Statistical Computing*2012: Vienna, Austria.

189. King, G., et al., *Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation.* American Political Science Review, 2001. **95**(1): p. 49-69.

190. Sterne, J.A., et al., *Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls.* BMJ, 2009. **338**: p. b2393.

191. Rubin, B.D., *Inference and missing data.* Biometrika, 1976. **63**(581-592).

192. Gelman, A., King, G., Liu, C., , *Not Asked and Not Answered: Multiple Imputation for Multiple Surveys.* Journal of the American Statistical Association, 1999. **93**: p. 846-857.

193. StataCorp, *Stata Statistical Software: Release 12*, 2011, College Station: TX.

194. Lagarde, E., et al., *Condom use and its association with HIV/sexually transmitted diseases in four urban communities of sub-Saharan Africa.* AIDS, 2001. **15 Suppl 4**: p. S71-8.

195. Lagarde, E., et al., *Educational level is associated with condom use within non-spousal partnerships in four cities of sub-Saharan Africa.* AIDS, 2001. **15**(11): p. 1399-408.

196. Hargreaves, J.R. and L.D. Howe, *Changes in HIV prevalence among differently educated groups in Tanzania between 2003 and 2007.* AIDS, 2010. **24**(5): p. 755-61.

197. de Walque, D., et al., *Changing association between schooling levels and HIV-1 infection over 11 years in a rural population cohort in south-west Uganda.* Tropical medicine & international health : TM & IH, 2005. **10**(10): p. 993-1001.

198. Sobrino-Vegas, P., et al., *Educational gradient in HIV diagnosis delay, mortality, antiretroviral treatment initiation and response in a country with universal health care.* Antiviral therapy, 2012. **17**(1): p. 1-8.

199. Hartmann, K.E., D.B. Matchar, and S.M. Chang, *Assessing Applicability of Medical Test Studies in Systematic Reviews*, in *Methods Guide for Medical Test Reviews*, S.M. Chang, et al., Editors. 2012: Rockville (MD).

200. Higgins, J.P.T. and S. Green, *Cochrane handbook for systematic reviews of interventions*. Cochrane book series2011, San Francisco, CA: John Wiley.

201. Borenstein, M., et al., *Introduction to Meta-Analysis*2009, West Sussex, UK: Wiley.

202. Stillwell, S.B., et al., *Evidence-based practice, step by step: asking the clinical question: a key step in evidence-based practice.* The American journal of nursing, 2010. **110**(3): p. 58-61.

203. Stillwell, S.B., et al., *Evidence-based practice, step by step: searching for the evidence.* The American journal of nursing, 2010. **110**(5): p. 41-7.

204. Rothman, J.K., Greenland, S, Lash, L.L, *Modern Epidemiology*. 3rd ed2008, New York: Lippincolt William & Wilkin.

205. Hammerstrom, K., Wade, A., Jorgensen, A.M.K.,, *Searching for studies:A guide to information retrieval for Campbell Systematic Reviews.* Campbell Systematic Reviews, 2010.

206. Lefebvre, C., Higgin,J.P.T,, *Cochrane Handbook for Systematic Reviews of Interventions*, ed. C.B. series2008, San Francisco: John Wiley.

207. Hopewell, S., et al., *A comparison of handsearching versus MEDLINE searching to identify reports of randomized controlled trials.* Statistics in medicine, 2002. **21**(11): p. 1625-34.

208. Mulrow, C. and D. Cook, *Systematic reviews: Synthesis of Best Evidence for Health Care Decisions*1998, Philadephia: ACP.

209. Edwards, P., et al., *Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records.* Statistics in medicine, 2002. **21**(11): p. 1635-40.

210. Viechtbauer, W., *Conducting meta-analyses in R with the metafor package.* Journal of Statistical Software, 2010. **36**(3): p. 1-48.

211. Rosner, B., *Fundamentals of Biostatistics* 6ed2006, Belmont, CA: Thompson Higher Education.

212. DerSimonian, R. and R. Kacker, *Random-effects model for meta-analysis of clinical trials: an update.* Contemporary clinical trials, 2007. **28**(2): p. 105-14.

213. DerSimonian, R., *Meta-analysis in the design and monitoring of clinical trials.* Statistics in medicine, 1996. **15**(12): p. 1237-48; discussion 1249-52.

214. DerSimonian, R. and N. Laird, *Meta-analysis in clinical trials.* Controlled clinical trials, 1986. **7**(3): p. 177-88.

215. Chu, H., S. Chen, and T.A. Louis, *Random Effects Models in a Meta-Analysis of the Accuracy of Two Diagnostic Tests Without a Gold Standard.* Journal of the American Statistical Association, 2009. **104**(486): p. 512-523.

216. Sterne, J.A.C., *Meta-analysis in stata: updated colection from stata journal*2009: StataCorp.

217. Hedges, L.V. and I. Olkin, *Statistical methods for meta-analysis*1994, Orlando, FL: Academic Press.

218. Kwiek, J.J., et al., *Socio-demographic characteristics associated with HIV and syphilis seroreactivity among pregnant women in Blantyre, Malawi, 2000-2004.* Malawi medical journal : the journal of Medical Association of Malawi, 2008. **20**(3): p. 80-5.

219. Yahya-Malima, K.I., et al., *The silent HIV epidemic among pregnant women within rural Northern Tanzania.* BMC public health, 2006. **6**: p. 109.

220. Lipsey, M.W. and D.B. Wilson, *Practical Meta-analysis*. Applied social research series, ed. L. Bickman and D.J. Rog2000, Sage Publications.

221. Duval , S. and R. Tweedie, *Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication*

Bias in Meta-Analysisr. Biometrics, 2000. **56**(2): p. 455-463.

222. Leimu, R. and J. Koricheva, *Cumulative meta-analysis: a new tool for detection of temporal trends and publication bias in ecology.* The Royal Society, 2004. **271**(1551).

223. Lawoyin, T.O. and D.A. Adewole, *Predictors of maternal HIV infection at the primary care level in inner city Ibadan.* International journal of STD & AIDS, 2004. **15**(3): p. 165-8.

224. Egger, M., et al., *Bias in meta-analysis detected by a simple, graphical test.* BMJ, 1997. **315**(7109): p. 629-34.

225. Ringquist, E., *Meta-Analysis for Public Management and Policy*, ed. M.R. Anderson2013, San Francisco, CA: Josse Bass.

226. Monge, S.J.I., Pérez-Hoyos S, Ferreros I, García-Olalla P, Muga R, Del Romero J, Belda J, Castilla J, Bolúmar F, Del Amo J; GEMES, *Educational level and HIV disease progression before and after the introduction of HAART: a cohort study in 989 HIV seroconverters in Spain.* Sexually Transmitted Infections, 2012. **87**(7): p. 571-576.

227. Zaba, B., et al., *Effect of HIV infection on pregnancy-related mortality in sub-Saharan Africa: secondary analyses of pooled community-based data from the network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA).* Lancet, 2013. **381**(9879): p. 1763-71.

228. Fabiani, M., et al., *A high prevalence of HIV-1 infection among pregnant women living in a rural district of north Uganda severely affected by civil strife.* Transactions of the Royal Society of Tropical Medicine and Hygiene, 2006. **100**(6): p. 586-93.

229. Utulu, S.N. and T.O. Lawoyin, *Epidemiological features of HIV infection among pregnant women in Makurdi, Benue State, Nigeria.* Journal of biosocial science, 2007. **39**(3): p. 397-408.

230. Crampin, A.C., et al., *Use of antenatal clinic surveillance to assess the effect of sexual behavior on HIV prevalence in young women in Karonga district, Malawi.* Journal of acquired immune deficiency syndromes, 2008. **48**(2): p. 196-202.

231. Mwandagalirwa, K., et al., *Local differences in human immunodeficiency virus prevalence: a comparison of social venue patrons, antenatal patients, and sexually transmitted infection patients in eastern kinshasa.* Sexually transmitted diseases, 2009. **36**(7): p. 406-12.

232. Kiptoo, M., et al., *Survey on prevalence and risk factors on HIV-1 among pregnant women in North-Rift, Kenya: a hospital based cross-sectional study conducted between 2005 and 2006.* BMC international health and human rights, 2009. **9**: p. 10.

233. Etukumana, E.A., T.D. Thacher, and A.S. Sagay, *HIV risk factors among pregnant women in a rural Nigerian hospital.* The West Indian medical journal, 2010. **59**(4): p. 424-8.

234. Kayibanda, J.F., et al., *HIV Prevalence Comparison Between Antenatal Sentinel Surveillance and Demographic and Health Survey in Rwanda.* The open AIDS journal, 2011. **5**: p. 29-36.

235. Urassa, M., et al., *HIV prevalence and sexual behaviour changes measured in an antenatal clinic setting in northern Tanzania.* Sexually transmitted infections, 2006. **82**(4): p. 301-6.

236. Ntaganira, J., et al., *Intimate partner violence among pregnant women in Rwanda.* BMC women's health, 2008. **8**: p. 17.

237. Nakubulwa, S., et al., *Association between HSV-2 and HIV serostatus in pregnant women of known HIV serostatus attending Mulago hospital antenatal clinic, Kampala, Uganda.* Journal of infection in developing countries, 2009. **3**(10): p. 803-6.

238. Kuate, S., et al., *Time trends and regional differences in the prevalence of HIV infection among women attending antenatal clinics in 2 provinces in Cameroon.* Journal of Acquired Immunodeficiency Syndrome, 2009. **52**(2): p. 258-264.

239. WHO, *Global Health Observatory*, 2011.

240. Leclerc, P.M., A.P. Matthews, and M.L. Garenne, *Fitting the HIV epidemic in Zambia: a two-sex micro-simulation model.* PloS one, 2009. **4**(5): p. e5439.

241. Leclerc, P.M. and M. Garenne, *Commercial sex and HIV transmission in mature epidemics: a study of five African countries.* International journal of STD & AIDS, 2008. **19**(10): p. 660-4.

242. Mertens, R., B. Jans, and X. Kurz, *A computerized nationwide network for nosocomial infection surveillance in Belgium.* Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America, 1994. **15**(3): p. 171-9.

243. Laeyendecker, O., et al., *Specificity of four laboratory approaches for cross-sectional HIV incidence determination: analysis of samples from adults with known nonrecent HIV infection from five African countries.* AIDS research and human retroviruses, 2012. **28**(10): p. 1177-83.

244. Braunstein, S.L., et al., *Dual testing algorithm of BED-CEIA and AxSYM Avidity Index assays performs best in identifying recent HIV infection in a sample of Rwandan sex workers.* PloS one, 2011. **6**(4): p. e18402.

245. Batzing-Feigenbaum, J., et al., *Country-wide HIV incidence study complementing HIV surveillance in Germany.* Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, 2008. **13**(36).

246. Batzing-Feigenbaum, J., et al., *Implications of and perspectives on HIV surveillance using a serological method to measure recent HIV infections in newly diagnosed individuals: results from a pilot study in Berlin, Germany, in 2005-2007.* HIV medicine, 2009. **10**(4): p. 209-18.

247. Hargrove, J.W., et al., *Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay.* AIDS, 2008. **22**(4): p. 511-8.

248. Kim, A.A., et al., *Evaluating the BED capture enzyme immunoassay to estimate HIV incidence among adults in three countries in sub-Saharan Africa.* AIDS research and human retroviruses, 2010. **26**(10): p. 1051-61.

249. Dobbs, T., et al., *Performance characteristics of the immunoglobulin G-capture BED-enzyme immunoassay, an assay to detect recent human immunodeficiency virus type 1 seroconversion.* Journal of clinical microbiology, 2004. **42**(6): p. 2623-8.

250. Brookmeyer, R., *On the statistical accuracy of biomarker assays for HIV incidence.* Journal of acquired immune deficiency syndromes, 2010. **54**(4): p. 406-14.

251. Brookmeyer, R., *Should biomarker estimates of HIV incidence be adjusted?* AIDS, 2009. **23**(4): p. 485-91.

252. Laeyendecker, O., et al., *Factors associated with incorrect identification of recent HIV infection using the BED capture immunoassay.* AIDS research and human retroviruses, 2012. **28**(8): p. 816-22.

253. Laeyendecker, O., et al., *Pregnancy does not affect HIV incidence test results obtained using the BED capture enzyme immunoassay or an antibody avidity assay.* PloS one, 2010. **5**(10): p. e13259.

254. Han, M., et al., *[Application of BED-CEIA to estimate the human immunodeficiency virus prevalence among injecting drug users in Chongqing municipality]*. Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi, 2007. **28**(11): p. 1077-80.

255. Kana, K., et al., *HIV-1 incidence estimates among young Thai men using IgG-capture BED-enzyme immunoassay (BED-CEIA) during 2005-2006.* Journal of the Medical Association of Thailand = Chotmaihet thangphaet, 2009. **92 Suppl 1**: p. S112-6.

256. Ma, W.J. and N. Wang, *[Validity of using BED-CEIA to estimate HIV-1 incidence and factors influencing its validity-systematic review]*. Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi, 2010. **31**(9): p. 1056-61.

257. Shen, S., et al., *[Evaluation of BED-CEIA assay with dried blood spot specimens in China]*. Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi, 2009. **30**(3): p. 273-6.

258. Mullis, C.E., et al., *Differential Specificity of HIV Incidence Assays in HIV Subtypes A and D-Infected Individuals from Rakai, Uganda.* AIDS research and human retroviruses, 2013.

259. Quinn, T.C., et al., *Feasibility of pooling sera for HIV-1 viral RNA to diagnose acute primary HIV-1 infection and estimate HIV incidence.* AIDS, 2000. **14**(17): p. 2751-7.

260. Kandala, N.B., et al., *Millennium development goal 6 and HIV infection in Zambia: what can we learn from successive household surveys?* AIDS, 2011. **25**(1): p. 95-106.

261. Newcombe, R.G., *Interval estimation for the difference between independent proportions: Comparison of eleven methods.* Statistics in medicine, 1998. **17**: p. 119-126.

262. Bates, D., *Linear mixed-effects models using S4 classes*, 2013.

263. Dupont, W., *Statistical Modelling for Biomedical Researchers*. 2 ed2009: Cambridge University Press.

264. Hosmer, W.H. and S. Lemeshow, *Applied Logistic Regression*. 2nd ed. Wiley Series in Probability and Statistics2000, New York: John Wiley and Sons, inc.

265. Thiele, J. and B. Markussen, *Potential of GLMM in modelling invasive spread.* CAB Reviews,, 2012. **7**(016).

266. Dalal, D.K. and J.Z. Michael *Centering or Not Centering in Multilevel Models? The Role of the Group Mean and the Assessment of Group Effects.* Organizational Research Methods. **15**(3): p. 339-362.

267. Sijders, T.A.B. and R.J. Bosker, *Introduction to basic and advanced multilevel modelling* 2011, Los Angelos: SAGE.

268. Shafer, L.A., et al., *HIV prevalence and incidence are no longer falling in southwest Uganda: evidence from a rural population cohort 1989-2005.* AIDS, 2008. **22**(13): p. 1641-9.

269. Shafer, L.A., et al., *The role of the natural epidemic dynamics and migration in explaining the course of the HIV epidemic in rural Uganda: a modelling study.* International journal of epidemiology, 2011. **40**(2): p. 397-404.

270. Kirungi, W.L., et al., *Trends in antenatal HIV prevalence in urban Uganda associated with uptake of preventive sexual behaviour.* Sexually transmitted infections, 2006. **82 Suppl 1**: p. i36-41.

271. Beyrer, C., et al., *The global HIV epidemics in men who have sex with men (MSM): time to act.* AIDS, 2013.

272. Gouws, E. and P. Cuchi, *Focusing the HIV response through estimating the major modes of HIV transmission: a multi-country analysis.* Sexually transmitted infections, 2012. **88 Suppl 2**: p. i76-85.

273. Abdool Karim, S.S., et al., *Global epidemiology of HIV-AIDS.* Infectious disease clinics of North America, 2007. **21**(1): p. 1-17, vii.

274. Mahy, M., J.M. Garcia-Calleja, and K.A. Marsh, *Trends in HIV prevalence among young people in generalised epidemics: implications for monitoring the HIV epidemic.* Sexually transmitted infections, 2012. **88 Suppl 2**: p. i65-75.

275. Long, E.F. and R.R. Stavert, *Portfolios of Biomedical HIV Interventions in South Africa: A Cost-Effectiveness Analysis.* Journal of general internal medicine, 2013.

276. Vermund, S.H., et al., *Translation of Biomedical Prevention Strategies for HIV: Prospects and Pitfalls.* Journal of acquired immune deficiency syndromes, 2013. **63 Suppl 1**: p. S12-25.

277. Marrazzo, J.M. and W. Cates, *Interventions to prevent sexually transmitted infections, including HIV infection.* Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 2011. **53 Suppl 3**: p. S64-78.

278. Flash, C., D. Krakower, and K.H. Mayer, *The Promise of Antiretrovirals for HIV Prevention.* Current infectious disease reports, 2012. **14**(2): p. 185-93.

279. Sandoy, I.F., et al., *Targeting condom distribution at high risk places increases condom utilization-evidence from an intervention study in Livingstone, Zambia.* BMC public health, 2012. **12**: p. 10.

280. Sandoy, I.F., et al., *Condom availability in high risk places and condom use: a study at district level in Kenya, Tanzania and Zambia.* BMC public health, 2012. **12**: p. 1030.

281. Jones, D.J., et al., *Sexual risk reduction among Zambian couples.* SAHARA J : journal of Social Aspects of HIV/AIDS Research Alliance / SAHARA , Human Sciences Research Council, 2009. **6**(2): p. 69-75.

282. Bunnell, R., J. Mermin, and K.M. De Cock, *HIV prevention for a threatened continent: implementing positive prevention in Africa.* JAMA : the journal of the American Medical Association, 2006. **296**(7): p. 855-8.

283. Bunnell, R., et al., *Changes in sexual behavior and risk of HIV transmission after antiretroviral therapy and prevention interventions in rural Uganda.* AIDS, 2006. **20**(1): p. 85-92.

284. Agha, S., *Declines in casual sex in Lusaka, Zambia: 1996-1999.* AIDS, 2002. **16**(2): p. 291-3.

285. Agha, S., *Intention to use the female condom following a mass-marketing campaign in Lusaka, Zambia.* American journal of public health, 2001. **91**(2): p. 307-10.

286. Agha, S. and T. Kusanthan, *Equity in access to condoms in urban Zambia.* Health policy and planning, 2003. **18**(3): p. 299-305.

287. Agha, S., P. Hutchinson, and T. Kusanthan, *The effects of religious affiliation on sexual initiation and condom use in Zambia.* The Journal of adolescent health : official publication of the Society for Adolescent Medicine, 2006. **38**(5): p. 550-5.

288. Agha, S., *An evaluation of the effectiveness of a peer sexual health intervention among secondary-school students in Zambia.* AIDS education and prevention : official publication of the International Society for AIDS Education, 2002. **14**(4): p. 269-81.

289. Yoder, P.S., R. Hornik, and B.C. Chirwa, *Evaluating the program effects of a radio drama about AIDS in Zambia.* Studies in family planning, 1996. **27**(4): p. 188-203.

290. Underwood, C., et al., *Reducing the risk of HIV transmission among adolescents in Zambia: psychosocial and behavioral correlates of viewing a risk-reduction media campaign.* The Journal of adolescent health : official publication of the Society for Adolescent Medicine, 2006. **38**(1): p. 55.

291. Bloom, S.S., et al., *Looking for change in response to the AIDS epidemic: trends in AIDS knowledge and sexual behavior in Zambia, 1990 through 1998.* Journal of acquired immune deficiency syndromes, 2000. **25**(1): p. 77-85.

292. Tate, J., et al., *Measurement of HIV prevention indicators: a comparison of the PLACE method and a household survey in Zambia.* AIDS and behavior, 2010. **14**(1): p. 209-17.

293. UNICEF, *Zambia fact sheet*, 2013.

294. Singh, K., et al., *Age, poverty and alcohol use as HIV risk factors for women in Mongu, Zambia.* African health sciences, 2011. **11**(2): p. 204-10.

295. Wilson, N., *Economic booms and risky sexual behavior: evidence from Zambian copper mining cities.* Journal of health economics, 2012. **31**(6): p. 797-812.

296. Kyei, N.N., C. Chansa, and S. Gabrysch, *Quality of antenatal care in Zambia: a national assessment.* BMC pregnancy and childbirth, 2012. **12**: p. 151.

297. Kyei, N.N., O.M. Campbell, and S. Gabrysch, *The influence of distance and level of service provision on antenatal care use in rural Zambia.* PloS one, 2012. **7**(10): p. e46475.

298. Karim, Q.A., et al., *Stabilizing HIV prevalence masks high HIV incidence rates amongst rural and urban women in KwaZulu-Natal, South Africa.* International journal of epidemiology, 2011. **40**(4): p. 922-30.

299. Monasch, R. and M. Mahy, *Young people: the centre of the HIV epidemic.* World Health Organization technical report series, 2006. **938**: p. 15-41; discussion 317-41.

300. Gabrysch, S., V. Simushi, and O.M. Campbell, *Availability and distribution of, and geographic access to emergency obstetric care in Zambia.* International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics, 2011. **114**(2): p. 174-9.

301. Gabrysch, S., et al., *Tracking progress towards safe motherhood: meeting the benchmark yet missing the goal? An appeal for better use of health-system output indicators with evidence from Zambia and Sri Lanka.* Tropical medicine & international health : TM & IH, 2011. **16**(5): p. 627-39.

302. Gabrysch, S., et al., *The influence of distance and level of care on delivery place in rural Zambia: a study of linked national data in a geographic information system.* PLoS medicine, 2011. **8**(1): p. e1000394.

303. Anderson, K.G., A.M. Beutel, and B. Maughan-Brown, *HIV risk perceptions and first sexual intercourse among youth in Cape Town South Africa.* International family planning perspectives, 2007. **33**(3): p. 98-105.

304. Sala, C., et al., *Modelling the trend of bovine spongiform encephalopathy prevalence in France: Use of restricted cubic spline regression in age-period-cohort models to estimate the efficiency of control measures.* Preventive veterinary medicine, 2009. **90**(1-2): p. 90-101.

305. Case, K.K., et al., *Understanding the modes of transmission model of new HIV infection and its use in prevention planning.* Bulletin of the World Health Organization, 2012. **90**(11): p. 831-838A.

306. Pisani, E., et al., *Back to basics in HIV prevention: focus on exposure.* BMJ, 2003. **326**(7403): p. 1384-7.

307. Abdool, Q. and Karim, *Barriers to preventing human immunodeficiency virus in women: experiences from KwaZulu-Natal, South Africa.* Journal of the American Medical Women's Association, 2001. **56**(4): p. 193-6.

308. Maharaj, P. and J. Cleland, *Risk perception and condom use among married or cohabiting couples in KwaZulu-Natal, South Africa.* International family planning perspectives, 2005. **31**(1): p. 24-9.

309. McGrath, N., et al., *Sexual behaviour in a rural high HIV prevalence South African community: time trends in the antiretroviral treatment era.* AIDS, 2013.

310. Fedor, M.T., *HIV/AIDS Prevention Knowledge: age-period-cohort analysis of trends in Malawi from 1992 to 2010*, in *Population Association of America*2013: New Orleans, LA.

311. Yang, T.T., et al., *Population inversion between I((2)P(1/2)) and I((2)P(3/2)) of atomic iodine generated by the energy transfer from NCl(a(1)Delta) to I((2)P(3/2)).* Optics letters, 1992. **17**(24): p. 1803-5.

312. Yang, Y., *Social inequalities in hapiness in the United States, 1972 to 2004: An Age-Period-Cohort analysis.* American Sociological Review, 2008. **73**: p. 204-226.

313. Jacobs, D.R., *Interpreting age, period and cohort effects in plasma lipids and serum insulin using repeated measures regression analysis: the Cardia study.* Statistics in medicine, 1999. **18**(1955-1979).

314. Robinson, W.R., et al., *Birth cohort effects among US-born adults born in the 1980s: foreshadowing future trends in US obesity prevalence.* International journal of obesity, 2013. **37**(3): p. 448-54.

315. Keyes, K.M., et al., *What is a cohort effect? Comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971-2006.* Social science & medicine, 2010. **70**(7): p. 1100-8.

316. Keyes, K.M. and G. Li, *A multiphase method for estimating cohort effects in age-period contingency table data.* Annals of epidemiology, 2010. **20**(10): p. 779-85.

317. Harding, D.J., *Recent advances in age-period-cohort analysis: a commentary on Dregan and armstrong and Reither, Hauser and Yang.* Social Science and Medicine, 2009. **69**: p. 1449_1451.

318. Zheng, H., Y. Yang, and K.C. Land, *Variance Function Regression in Hierarchical Age-Period-Cohort Models: Applications to the Study of Self-Reported Health.* American sociological review, 2011. **76**(6): p. 955-983.

319. Zheng, T., et al., *Time trend and age-period-cohort effect on incidence of bladder cancer in Connecticut, 1935-1992.* International journal of cancer. Journal international du cancer, 1996. **68**(2): p. 172-6.

320. Holford, T.R., *Approaches to fitting age-period-cohort models with unequal intervals.* Statistics in medicine, 2006. **2525:977–993**: p. 977–993.

321. Robinson, W.R., et al., *Birth cohort effects on abdominal obesity in the United States: the Silent Generation, Baby Boomers and Generation X.* International journal of obesity, 2012.

322. Carstensen, B., *Age-period-cohort models for the Lexis diagram.* Statistics in medicine, 2007. **26**(15): p. 3018-3048.

323. Carstensen, B., *Age-Period-Cohort models:Statistical inference in the Lexis diagram*, 2005, University of Copenhagen, Institute of Public health: Copehegan.

324. Clayton, D. and E. Schifflers, *Models for temporal variation in cancer rates. II: Age-period-cohort models.* Statistics in medicine, 1987. **6**(4): p. 469-81.

325. Rodgers, W.L., *Estimable functions of age, period and cohort effects.* American Sociological Review 1982. **47**: p. 774-787.

326. Glenn, N.D., *Cohort analyst futile quest:statistical attempts to separate age, period and cohort effects.* American Sociological Review, 1976. **41**(900-904).

327. Mason, W.B. and S.E. Fienberg, *Cohort analysis in social research:beyond identification problem*1985: Springer-Verag, New York.

328. Diggle, P.J., et al., *Analysis of longitudinal data*. 2 ed, ed. A.C. Atkinson, et al.2002: Oxfor Press.

329. Ananth, C.V., et al., *Small-for-gestational-age births in the United States: an age-period-cohort analysis.* Epidemiology, 2004. **15**(1): p. 28-35.

330. Ryder, R.W., et al., *Fertility rates in 238 HIV-1-seropositive women in Zaire followed for 3 years post-partum.* AIDS, 1991. **5**(12): p. 1521-7.

331. Pinheiro, J.C. and M.B. Bates, *Mixed-Effects Models in S and S-PLUS*, ed. J. Chamber, et al.2000: Springer-Velarg.

332. Singer, J.D. and J.B. Willet, *Applied Longitudinal Data Analysis*2003: Oxford Press.

333. Dickinson, L.M. and A. Basu, *Multilevel modeling and practice-based research.* Annals of family medicine, 2005. **3**(1).

334. Miech, R. and S. Koester, *Trends in U.S., past-year marijuana use from 1985 to 2009: an age-period-cohort analysis.* Drug and alcohol dependence, 2012. **124**(3): p. 259-67.

335. Raundenbush, S.W.a.B., A.S, *Hierarchical linear models: Applications and data analysis methods*, ed. Newbury2000, Lincolnwood, IL: Scientific Software Company.

336. Keyes, K.M. and R. Miech, *Age, period, and cohort effects in heavy episodic drinking in the US from 1985 to 2009.* Drug and alcohol dependence, 2013.

337. Honaker, J., G. King, and M. Blackwell, *Amelia II: A Program for Missing Data.* Journal of Statistical Software, 2011. **45**(7).

338. Sander Greenland, G. and W.D. Finkle, *A Critical Look at Methods for Handling Missing Covariates in Epidemiologic Regression Analyses.* American journal of epidemiology, 1995. **142**(12): p. 1255-1264.

339. Floyd, S., et al., *Underestimation of HIV prevalence in surveys when some people already know their status, and ways to reduce the bias.* AIDS, 2013. **27**(2): p. 233-42.

340. Mbougua, J.B., et al., *Nonlinear multiple imputation for continuous covariate within semiparametric Cox model: application to HIV data in Senegal.* Statistics in medicine, 2013.

341. Graham, J.W., *Missing data analysis: making it work in the real world.* Annual review of psychology, 2009. **60**: p. 549-76.

342. Baraldi, A.N. and C.K. Enders, *An introduction to modern missing data analyses.* Journal of school psychology, 2010. **48**(1): p. 5-37.

343. He, Y., *Missing data analysis using multiple imputation: getting to the heart of the matter.* Circulation. Cardiovascular quality and outcomes, 2010. **3**(1): p. 98-105.

344. DeSouza, C.M., A.T. Legedza, and A.J. Sankoh, *An overview of practical approaches for handling missing data in clinical trials.* Journal of biopharmaceutical statistics, 2009. **19**(6): p. 1055-73.

345. Mackinnon, A., *The use and reporting of multiple imputation in medical research - a review.* Journal of internal medicine, 2010. **268**(6): p. 586-93.

346. Rubin, D.B., *Multiple imputations for nonresponse in survey*1987, New York: John Wiley and Sons.

347. Crainiceanu. C, R.D., Claesken.G, Wand.M.P, *Exact Likelihood Ratio Tests for Penalized Splines.* Biometrika, 2005. **92**: p. 91-103.

348. WS, G., *Inference in smoothing spline analysis of variance.* Journal of Royal Statiscal Society, 2002. **64**: p. 887-898.

349. Crainiceanu. C, R.D., *Likelihood ratio test in linear mixed models with one variance component.* Journal of Royal Statiscal Society, 2004. **66**: p. 909-926.

350. Harrel, E.F., *Regression Modelling Strategies*. Springer series in statistics, ed. Springer2001.

351. Twisk, W.R.J., *Applied Multilevel Analysis*. Practical guides to biostatistics and Epidemilogy, ed. S. Ellenberg, et al.2006: Cambridge Press.

352. de Walque, D. and R. Kline, *The association between remarriage and HIV infection in 13 sub-Saharan African countries.* Studies in family planning, 2012. **43**(1): p. 1-10.

353. Santelli, J.S., et al., *Behavioral, Biological, and Demographic Risk and Protective Factors for New HIV Infections among Youth, Rakai, Uganda.* Journal of acquired immune deficiency syndromes, 2013.

354. Dunkle, K.L., et al., *New heterosexually transmitted HIV infections in married or cohabiting couples in urban Zambia and Rwanda: an analysis of survey and clinical data.* Lancet, 2008. **371**(9631): p. 2183-91.

355. Bunnell, R., et al., *HIV transmission risk behavior among HIV-infected adults in Uganda: results of a nationally representative survey.* AIDS, 2008. **22**(5): p. 617-24.

356. Dzekedzeke, K., S. Siziya, and K. Fylkesnes, *The impact of HIV infection on adult mortality in some communities in Zambia: a cohort study.* Tropical medicine & international health : TM & IH, 2008. **13**(2): p. 152-61.

357. Dzekedzeke, K. and K. Fylkesnes, *Reducing uncertainties in global HIV prevalence estimates: the case of Zambia.* BMC Public Health, 2006. **6**: p. 83.

358. Jurek, A.M., et al., *Exposure-measurement error is frequently ignored when interpreting epidemiologic study results.* European journal of epidemiology, 2006. **21**(12): p. 871-6.

359. Robert M. Groves, R.M., et al., *Survey Methodology*. 2 ed. Wiley series in survey methodology2009: Wiley.

360. Jurek, A.M., G. Maldonado, and S. Greenland, *Adjusting for outcome misclassification: the importance of accounting for case-control sampling and other forms of outcome-related selection.* Annals of epidemiology, 2013. **23**(3): p. 129-35.

361. Jurek, A.M., et al., *Proper interpretation of non-differential misclassification effects: expectations vs observations.* International journal of epidemiology, 2005. **34**(3): p. 680-7.

362. Hargreaves, J.R., et al., *Explaining continued high HIV prevalence in South Africa: socioeconomic factors, HIV incidence and sexual behaviour change among a rural cohort, 2001-2004.* AIDS, 2007. **21 Suppl 7**: p. S39-48.

363. Shi, C.F., F.G. Kouyoumdjian, and J. Dushoff, *Intimate partner violence is associated with HIV infection in women in Kenya: A cross-sectional analysis.* BMC public health, 2013. **13**(1): p. 512.

364. Cohen, A.K. and S.L. Syme, *Education: a missed opportunity for public health intervention.* American journal of public health, 2013. **103**(6): p. 997-1001.

365. Evans, D., et al., *Combining directed acyclic graphs and the change-in-estimate procedure as a novel approach to adjustment-variable selection in epidemiology.* BMC medical research methodology, 2012. **12**: p. 156.

366. Howards, P.P., et al., *"Toward a clearer definition of confounding" revisited with directed acyclic graphs.* American journal of epidemiology, 2012. **176**(6): p. 506-11.

367. Schipf, S., et al., *[Directed acyclic graphs (DAGs) - the application of causal diagrams in epidemiology]*. Gesundheitswesen, 2011. **73**(12): p. 888-92.

368. Weng, H.Y., et al., *Methods of covariate selection: directed acyclic graphs and the change-in-estimate procedure.* American journal of epidemiology, 2009. **169**(10): p. 1182-90.

369. VanderWeele, T.J., M.A. Hernan, and J.M. Robins, *Causal directed acyclic graphs and the direction of unmeasured confounding bias.* Epidemiology, 2008. **19**(5): p. 720-8.

370. VanderWeele, T.J. and J.M. Robins, *Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect.* American journal of epidemiology, 2007. **166**(9): p. 1096-104.

371. Hernan, M.A., et al., *Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology.* American journal of epidemiology, 2002. **155**(2): p. 176-84.

372. Lopman, B., et al., *HIV incidence in 3 years of follow-up of a Zimbabwe cohort--1998-2000 to 2001-03: contributions of proximate and underlying determinants to transmission.* International journal of epidemiology, 2008. **37**(1): p. 88-105.

373. Rutherford, M.J., J.R. Thompson, and P.C. Lambert, *Projecting cancer incidence using age-period-cohort models incorporating restricted cubic splines.* The international journal of biostatistics, 2012. **8**(1): p. 33.

374. Marrie, R.A., N.V. Dawson, and A. Garland, *Quantile regression and restricted cubic splines are useful for exploring relationships between continuous variables.* Journal of clinical epidemiology, 2009. **62**(5): p. 511-7 e1.

375. Lambert, P.C., et al., *Analysis of ambulatory blood pressure monitor data using a hierarchical model incorporating restricted cubic splines and heterogeneous within-subject variances.* Statistics in medicine, 2001. **20**(24): p. 3789-805.

376. Vermunt, J.K., *Mixed-Effects Logistic Regression Models for Indirectly Observed Discrete Outcome Variables.* Multivariate behavioral research, 2005. **40**(3): p. 281-301.

377. Zheng, W., et al., *Circulating 25-hydroxyvitamin D and risk of epithelial ovarian cancer: Cohort Consortium Vitamin D Pooling Project of Rarer Cancers.* American journal of epidemiology, 2010. **172**(1): p. 70-80.

378. Zeleniuch-Jacquotte, A., et al., *Circulating 25-hydroxyvitamin D and risk of endometrial cancer: Cohort Consortium Vitamin D Pooling Project of Rarer Cancers.* American journal of epidemiology, 2010. **172**(1): p. 36-46.

379. Hargreaves, J.R., et al., *Changes over time in sexual behaviour among young people with different levels of educational attainment in Tanzania.* Journal of the International AIDS Society, 2012. **15 Suppl 1**: p. 1-7.

380. Eshleman, S.H., et al., *Use of a multifaceted approach to analyze HIV incidence in a cohort study of women in the United States: HIV Prevention Trials Network 064 Study.* The Journal of infectious diseases, 2013. **207**(2): p. 223-31.

381. Pronyk, P.M., et al., *Effect of a structural intervention for the prevention of intimate-partner violence and HIV in rural South Africa: a cluster randomised trial.* Lancet, 2006. **368**(9551): p. 1973-83.

382. Pronyk, P.M., et al., *A combined microfinance and training intervention can reduce HIV risk behaviour in young female participants.* AIDS, 2008. **22**(13): p. 1659-65.

383. Niyonsenga, T., et al., *Measuring Socioeconomic Inequality in the Incidence of AIDS: Rural-Urban Considerations.* AIDS and behavior, 2012.

384. Dariotis, J.K., et al., *Changes in sexual risk behavior as young men transition to adulthood.* Perspectives on sexual and reproductive health, 2008. **40**(4): p. 218-25.

385. Vermund, S.H., *Modeling interventions to assess HIV epidemic impact in Africa.* Journal of acquired immune deficiency syndromes, 2011. **58**(2): p. 121-4.

386. Stoneburner, R.L. and D. Low-Beer, *Sexual partner reductions explain human immunodeficiency virus declines in Uganda: comparative analyses of HIV and behavioural data in Uganda, Kenya, Malawi, and Zambia.* International journal of epidemiology, 2004. **33**(3): p. 624.

387. Stoneburner, R.L. and D. Low-Beer, *Population-level HIV declines and behavioral risk avoidance in Uganda.* Science, 2004. **304**(5671): p. 714-8.

388. Yahya-Malima, K.I., et al., *High potential of escalating HIV transmission in a low prevalence setting in rural Tanzania.* BMC public health, 2007. **7**: p. 103.