

**MIXED-EFFECTS MODELING OF ASSOCIATION OF SYNDEMICS WITH HIV
VIRAL LOADS IN NAIROBI SEX WORKERS COHORT ON ANTIRETROVIRAL
THERAPY**

Akuku Isaiah Gumbe

W62/88215/2016

A thesis submitted in partial fulfillment for the award of the degree of Master of Science in Medical Statistics in the Institute of Tropical and Infectious Diseases, University of Nairobi

November 2018

DECLARATION

I, Akuku Isaiah Gumbe, confirm that the work presented in this thesis is my own.
Where information has been derived from other sources, I confirm that this has been
Indicated in the thesis.

Signature..... Date.....

This thesis has been submitted to the University with our approval as university supervisors

Dr. Nelson Owuor, PhD

School of Mathematics, University of Nairobi

Signature..... Date.....

Dr. Anne Wangómbe, PhD

School of Mathematics, University of Nairobi

Institute of Tropical and Infectious Diseases, University of Nairobi

Signature..... Date.....

ABSTRACT

Background: Sex workers are amongst the key populations substantially burdened by HIV/AIDS. However, there is definitive proof that antiretroviral therapy (ART), if consistently used, has individual and public health benefits due to viral suppression and onward HIV sexual transmission reduction. The standard HIV/AIDS clinical care approaches mostly emphasize on biomedical and health interventions and may ignore synergistic epidemics (syndemics) which, when present, may be associated with the likelihood of elevated HIV viral loads.

Objectives: To examine and model the association of syndemics with HIV viral load outcomes as well as antiretroviral therapy (ART) adherence among sex workers living with HIV in the Sex Worker Outreach Project (SWOP)–City cohort in Kenya using linear mixed-effects.

Method: Data collected between 2013—15th October 2018 from SWOP–City were modeled using linear mixed-effects for progression of viral load marker.

Results: None of HIV-syndemics considered was statistically significant on univariable random intercept model. On the multivariable model, only HIV-STI syndemic and condom use were statistically significant effect ($p < 0.01$). A unit increase in the months since baseline was associated with 0.055110 Box-Cox transformed viral load in the random intercept model, however, this effect wasn't statistically significant ($p > 0.005$). Poor ARV adherence variable was associated with 3.408471 increase in Box-Cox transformed viral load ($p = 0.0225$) and 1.237 positive change in transformed viral load ($p = 0.0119$) baseline effects. An interaction of poor last ARV adherence with month yielded 0.1528 negative change suggestive of programme intervention along the follow-up but not significant ($p = 0.0990$). The effect of longitudinal ART adherence in lessening syndemics was associated with a reduction in Box-Cox and \log_{10} transformed viral loads, nonetheless, these effects were not statistically significant ($p > 0.05$). An increase in linkage to psychosocial support was associated with 0.10077 reductions in \log_{10} viral load but not significant ($p = 0.454$). Linkage to psychosocial support played a key role in modifying the relationship between log HIV viral and poor adherence ($p < 0.01$).

Conclusions: The effect of syndemics was not directly associated with poor viral loads in the presence of adherence and psychosocial support. Linkage to psychosocial support modified the effect of syndemics on viral load evolutions. The study underscores the need of enhancement of linkage to psychosocial support.

DEDICATION

To the little, John Alex Akuku. And to the memory of my lovely grandparents, Elizabeth Mikwa Gumbe and Isaiah Gumbe Aluoch, whose immense love and support inspired my life.

ACKNOWLEDGMENTS

I'm very grateful to my thesis supervisors, Dr. Nelson Owuor and Dr. Anne Wangómbe, for their incessant guidance and unwavering support, and to the fraternity of Institute of Tropical and Infectious Disease, the University of Nairobi for their untold support.

I'm very thankful to the University of Manitoba/University of Nairobi – Partners for Health and Development in Africa (PHDA) for supplying the SWOP data. I wish to convey my sincere appreciation to Dr. Joshua Kimani, the Clinical Research Director–PHDA and Mr. Festus Muriuki, the Monitoring and Evaluation Manager–PHDA, Mr. Julius Munyao the SWOP City Coordinator, for their assistance. I'm also immensely thankful and greatly indebted to Mr. Byron Mbilu, the Data Manager for all his selfless support right from data preparation, taking his precious time to assist in this and responding to queries that molded this thesis.

I lack words to describe the professionalism and friendliness of the entire team at the SWOP City Clinic and for their invaluable assistance. I'm extremely grateful to all the research participants whose data were retrospectively used in this study.

I'm grateful to my classmates for their very useful suggestions and the constructive comments that shaped this thesis substantially. I'm so grateful to my family for their enduring support, love and continually being there for me. It would be callous not to mention my research mentor, Dr. Delia Grace (International Livestock Research Institute) and the initial career and funding mentors, the Consortium for National Health Research who have inspired my research trajectory over the years. Above all, I thank God for the life and for everything.

LIST OF TABLES

| | |
|--|----|
| Table 1: Sociodemographic and epidemiological profiles of sex workers | 24 |
| Table 2: Multivariate Analyses of sociodemographic and epidemiologic fixed effects | 37 |
| Table 3: Solution to key population type fixed effects | 38 |
| Table 4: Solution to Age at last visit fixed effects | 38 |
| Table 5: Solution to baseline condom use fixed effects | 39 |
| Table 6: Solution to baseline condom violence fixed effects | 39 |
| Table 7: Solution to baseline use of alcohol during sex fixed effects | 39 |
| Table 8: Solution to any STI fixed effects | 40 |
| Table 9: Solution to baseline GBV in the last 3 months fixed effects | 40 |
| Table 10: Solution to Nutritional assessment fixed effects | 40 |
| Table 11: Solution to last CTX adherence fixed effects | 41 |
| Table 12: Solution to last ARV adherence fixed effects | 41 |
| Table 13: Multivariable analyses of longitudinal ART adherence effect | 42 |
| Table 14: Solution to Multivariable Fixed Effects under \log_{10} and Box-Cox transformations | 44 |
| Table 15: Solution to modification effect on \log_{10} viral load of linkage to psychosocial support | 45 |
| Table 16: Random intercept Linear mixed model fit by maximum likelihood | 47 |
| Table 17: Random slope Linear mixed model fit by maximum likelihood | 48 |
| Table 18: Likelihood ratio test | 48 |
| Table 19: Variance of the random effects | 49 |
| Table 20: Within cluster resampling – Multiple outputation | 50 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1: HIV viral density by key population type | 26 |
| Figure 2: Log-transformed HIV viral load by key population..... | 27 |
| Figure 3: Log-transformed HIV viral load with a linear model of baseline age..... | 28 |
| Figure 4: Stratified log HIV by baseline age | 29 |
| Figure 5: Stratified plot of log HIV VL with linear model..... | 30 |
| Figure 6: Residuals vs Fitted plot | 31 |
| Figure 7: Normal Q-Q plot | 32 |
| Figure 8: histogram plot of the log-transformed HIV viral load | 33 |
| Figure 9: Box-Cox transformation of viral load +1 | 34 |
| Figure 10: Temporal means of log viral load..... | 35 |
| Figure 11: Temporal medians of log viral load | 35 |
| Figure 12: Individual profile plots by months since baseline | 36 |

LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|-----------|---|
| ART | Antiretroviral therapy |
| MSM | Men who have sex with men |
| MSW | Men who have sex with women |
| MSMW | Men who have sex with men and women |
| PWID/PWUD | People who inject drugs/People who use drugs |
| TasP | Treatment as prevention |
| WHO | World health organization |
| UoM | University of Manitoba |
| UoN | University of Nairobi |
| HIV | Human Immunodeficiency Syndrome |
| PLHIV | People living with HIV |
| PHDA | Partners in Health and Development in Africa |
| HIV/AIDS | Human Immunodeficiency Virus/Acquired Immunodeficiency Disease Syndrome |
| N | Normally distributed |
| GBV | Gender-Based Violence |
| ARV | Antiretroviral |
| CTX | Cotrimoxazole |

TABLE OF CONTENTS

| | |
|--|-----|
| DECLARATION | i |
| ABSTRACT | ii |
| DEDICATION | iii |
| ACKNOWLEDGEMENTS | iv |
| LIST OF TABLES | v |
| LIST OF FIGURES | vi |
| LIST OF ABBREVIATIONS AND ACCRONYMS | vii |
| CHAPTER ONE | 1 |
| 1. INTRODUCTION | 1 |
| 1.0 Background | 1 |
| 1.1 Application of syndemic theory to HIV in sex workers on ART in Kenya | 3 |
| 1.2 Problem Statement | 7 |
| 1.3 Study Justification and Significance | 8 |
| 1.4 Objectives..... | 10 |
| 1.4.1 General Objective | 10 |
| 1.4.2 Specific Objectives | 10 |
| 1.5 Assumptions of the Study | 10 |
| CHAPTER TWO | 12 |
| 2. LITERATURE REVIEW | 12 |
| 2.0 Introduction | 12 |
| 2.1 Methods for Longitudinal Clustered Data Analysis..... | 12 |
| 2.2 Mixed-Effects Models for Longitudinal Data..... | 13 |
| 2.3 Linear Mixed Models | 13 |
| 2.3.1 Random Slope and Random Intercept Model | 14 |

| | | |
|---------------|--|----|
| 2.3.2 | Conditional and Marginal distributions: Random intercept model..... | 16 |
| 2.3.3 | Conditional and Marginal distributions: Random slope model | 17 |
| CHAPTER THREE | | 18 |
| 3. | METHODOLOGY | 18 |
| 3.0 | Introduction | 18 |
| 3.1 | Study Design | 18 |
| 3.2 | Sample size..... | 18 |
| 3.3 | Ethical Consideration and Data Management | 19 |
| 3.3.1 | Ethical considerations | 19 |
| 3.3.2 | Data Management | 19 |
| 3.4 | Variables and Measures | 20 |
| 3.4.1 | Sociodemographic covariates | 20 |
| 3.4.2 | Other baseline variables | 20 |
| 3.4.3 | Outcome variable | 20 |
| 3.4.4 | Medication adherence | 20 |
| 3.4.5 | Syndemic variables/factors | 20 |
| 3.4.6 | Effect modification variable | 21 |
| 3.4.7 | Data Analysis | 21 |
| CHAPTER FOUR | | 24 |
| 4. | RESULTS | 24 |
| 4.0 | Introduction | 24 |
| 4.2 | Sociodemographic and epidemiologic profiles of the sex worker cohort | 24 |
| 4.3 | Viral load distributional properties | 25 |
| 4.3.1 | Key population and age-stratified viral load distributions on a logarithmic scale, log ₁₀ copies per mL..... | 25 |

| | | |
|--------------------|---|----|
| 4.2.1 | Box-Cox transformation of the HIV viral loads | 32 |
| 4.2.2 | Temporal trends in the mean and median values of the viral load copies | 34 |
| 4.3 | Individual profile plots | 36 |
| 4.4 | Analyses of sociodemographic and epidemiologic fixed effects | 36 |
| 4.5 | The influence of syndemics on HIV viral load outcomes and ART adherence | 38 |
| 4.5.1 | Univariable Models by log ₁₀ copies per mL | 38 |
| 4.5.2 | Estimates of baseline and longitudinal ART adherence effect on syndemics | 41 |
| 4.5.3 | Multivariable Model by log ₁₀ copies per mL and Box-Cox transformed viral loads 43 | |
| 4.6 | Syndemics and viral load outcomes and effect modification by linkage to psychosocial support..... | 45 |
| 4.7 | Random effects model for viral load by last adherence | 47 |
| 4.7.1 | Random intercept model of Box-Cox transformed HIV viral load and adherence with month interaction..... | 47 |
| 4.7.2 | Within cluster resampling – Multiple outputation | 49 |
| 4.7.3 | Both random intercept and random slope effects | 51 |
| CHAPTER FIVE | | 52 |
| 5. | DISCUSSIONS AND CONCLUSIONS..... | 52 |
| 5 | Introduction | 52 |
| 5.2 | Discussion | 52 |
| 5.3 | Limitations | 53 |
| 5.4 | Recommendations | 54 |
| 5.5 | Future research | 54 |
| 5.6 | Conclusions | 54 |
| REFERENCES | | 56 |
| APPENDICES | | 69 |

Appendix A: Ethical Approval Letter 69

Appendix B: R scripts – Exploring log viral load longitudinal data 71

Appendix C: R scripts for – the evolutions in \log_{10} viral load over time by adherence 72

Appendix D: R scripts – Test for the need of the random effects 74

Appendix E: R scripts – Box-Cox transformation of HIV viral loads 77

Appendix F: R scripts – Model fit by Maximum Likelihood 78

Appendix G: R scripts – Subsetting data by random selection of unique rows for multiple
outputation..... 80

CHAPTER ONE

1. INTRODUCTION

In the pre-analysis part of this thesis, the context and motivation of the study are presented. Statistical literature on longitudinal mixed-effects model for continuous data is reviewed. In the second part of this thesis, mixed-effects model was applied in the analysis of HIV-infection data to draw inferences and estimate the association/correlation parameters.

In Chapter 1, background to the study is presented and Sex Worker Outreach Project–Kenya that motivated this study introduced (Section 1.0), together with the relevant theme specifically application of syndemics to understanding human immunodeficiency virus (HIV) in sex workers initiating ART in Kenya (Section 1.1), problem statement (Section 1.2), justification and significance (Section 1.3) and objectives (Section 1.4). Chapter 2 gives a literature review of the mixed-effects model and a systematic framework of building it. Chapter 3 presents the statistical methodology of the study, data exploration and analysis software implementations in R.

Chapter 4 presents the results from the analysis of SWOP–Kenya data. Chapter 5 presents the discussion of the findings and close with conclusions, recommendations, and future research.

1.0 Background

In many longitudinal medical studies, laboratory and clinical measurements are usually measured at baseline, and patients are then followed over time consequent to this baseline time point to study associations of these outcomes with predictors. Similarly, in HIV prevention, care and treatment, the observational study interest is in monitoring the evolutions of patients' clinical and/or laboratory measures over time, the predictors that may affect the evolutions, and potentially how such changes may result in variations in morbidity, mortality, among other outcomes – if need be.

Since HIV is a chronic and life-threatening infection that requires regular and reliable medical care, insights into the predictors of disease progression are helpful in treatment initiation and therapeutic monitoring decisions. This is where HIV-RNA is useful for assessment of disease progression. And so, when patients' longitudinal data with repeated measures over time are accessible, analytical approaches that describe the evolutions or changes in these measures, and exploring the associations of these changes with predictors are employable.

Inspired by public health and epidemiological thinking, Baer, Singer, and Susser (2013) and Singer (1996) introduced the term “syndemic”, a portmanteau of “synergy” and “epidemic”, “pandemic” and “endemic” (Douglas-Vail, 2016), and defined it as the “aggregation of two or more diseases or other health conditions in a population in which there is some level of deleterious biological or behaviour interface that exacerbates the negative health effects of any or all of the diseases involved” (Singer, Bulled, Ostrach, & Mendenhall, 2017).

Moreover, syndemic or synergistic epidemic is beyond being just a portmanteau or a synonym for comorbidity (The Lancet, 2017) and can provide a framework applicable in the analysis of sex workers’ HIV-infection data. Epidemiologically, HIV is known to affect sex workers disproportionately (Djomand, Quaye, & Sullivan, 2014; Kerr et al., 2013; McKinnon et al., 2012; Okal et al., 2013), especially in Kenya.

This study is motivated by data from SWOP–Kenya, a programme run by Partners for Health and Development in Africa (PHDA) under the University of Manitoba and University of Nairobi research collaboration. PHDA builds on a legacy of successful collaborative implementation of HIV/AIDS research, prevention, Care and Treatment since 1980. The SWOP–Kenya, under PHDA has been in existence since 2008 and equally provides clinical & preventative services to sex workers in Nairobi and environs and operates nine clinics/sites. Seven SWOP clinics are dedicated to serving most-at-risk populations (male and female) while two other sites are under the University of Nairobi–Centre for Excellence and University of Maryland programme.

At the SWOP facilities, the sex workers—in a friendly, acceptable and confidential manner—access “the minimum HIV prevention, Care and Treatment package” covering information on safer sex practices, condom use, HIV testing and counseling, STI screening and treatment, risk reduction counselling, ARV and HIV basic care, family planning, TB screening and referral, Pre–, Post Exposure Prophylaxis, and linkage to psychosocial support.

And so, with such developments in HIV care, treatments, and prevention, mainly ART, the longstanding HIV/AIDS clinical management, that usually happen largely in outpatient setups – as with SWOP–Kenya, presents opportunities to model the sex workers’ VL evolutions given the syndemics. Given this, this study aims at modeling the association of longitudinal HIV-RNA (viral load) trajectories with syndemics in sex workers initiating ART, while appropriately taking

into account the statistical complexity of the data arising from the repeated measures over time using linear mixed-effects model.

1.1 Application of syndemic theory to HIV in sex workers on ART in Kenya

According to the UNAIDS report, Kenya has the joint fourth-largest HIV epidemic in the world (alongside Mozambique and Uganda) with 1.6 million PLHIV in 2016 (UNAIDS, 2017). While the number of deaths from AIDS-related illnesses was 36,000 in the year 2016, it had gradually reduced from 64,000 in 2010. Newly infected adult women in 2016 were 34 000 while 22 000 adult men. The UNAIDS data shows that women disproportionately affected by HIV accounted for about 57% of the 1.6 million PLHIV in Kenya.

The most HIV affected groups in Kenya are the sex workers. Nationally, the urban FSW population has been previously estimated as 138,420 although, six years ago, mapped FSW population in all the Kenyan towns was reported as 103,298 (Odek et al., 2014). Okal et al. (2013)'s estimates of the size of HIV high-risk MSM was 11,042 (10,000–22,222), FSW 29,494 (10,000–54,467) and about 6,107 (5,031–10,937) people who inject drugs/intravenous drug users (PWID/IDU) within Nairobi. In April 2009 in Nairobi central business district alone, 6,904 male and female sex workers working nightly have been previously enumerated (Kimani et al., 2013). These numbers may have plausibly increased.

Estimates from 2016 show that 29.3% of FSWs are living with HIV (MoH/NACC, 2016). Musyoki et al. (2015)'s respondent-driven sampling study in Nairobi also reported an overall prevalence in FSWs of 29.5 % (95 % CI 24.7–34.9) agreeing with Kenyan Ministry of Health/National AIDS Control Council estimates. However, systematic review and meta-analysis in Kenya revealed a pooled prevalence of 45.1% in FSW in comparison with 7.7% in female general population (Baral et al., 2012).

Previous prevalence among MSM in 2010 was 18.2% (IOM, 2011; MoH/NACC, 2016); in 2011, an estimated 18.3% among PWID (IBBS, 2012) majority of whom are concentrated in Nairobi and Mombasa (NACC, 2014). Data has also suggested that MSM, particularly those selling sex, contribute significantly to the Kenyan HIV epidemic (McKinnon et al., 2014; Muraguri et al., 2015).

Syndemic risk has been touted as an ecological construct and thus a function of determinants (Batchelder, Gonzalez, Palma, Schoenbaum, & Lounsbury, 2015). In Kenya, the behaviours predisposing key populations to increased risk of acquisition and transmission of HIV are unlawful (Githuka et al., 2014). Consequently, local ecologies of security of FSWs – in Nairobi for instance (Lorway et al., 2018), underpin HIV infection vulnerabilities (Jana, Basu, Rotheram-Borus, & Newman, 2004; Kerrigan et al., 2015; Moore et al., 2014; Reza-Paul et al., 2012) e.g sex workers and clients condom negotiations and the potential risk of HIV superinfection.

The aforementioned local ecologies may have repercussions in disease management among the HIV-infected sex workers. Apparently, the stigma, criminalization, and violence (Global Network of Sex Work Projects, 2015) enable these key populations to be hard to reach in HIV routine surveillance, thus obstructing their access to HIV treatment, care, and prevention services (Githuka et al., 2014). Even if they are reached, barriers may exist which affect HIV care retention along the care continuum (Wawrzyniak et al., 2015) as reported elsewhere in individuals not necessarily sex workers. Nonetheless, in Kenya, sex workers' access to services has since been prioritized in the National AIDS Strategic Plan III 2009–2013 (NACC, 2012).

In terms of antiretroviral treatment (ART), Kenya adopted the WHO's recommendations to treat immediately individuals diagnosed with HIV (UNAIDS, 2016) and although the country has had a long-standing HIV national prevention programme, ART coverage, is strikingly low in key populations, with values from 6% among MSM to 34% among FSW (PEPFAR [U.S. President's Emergency Plan for AIDS Relief], 2017). This coverage concern is alluded to in a study by Prakash et al. (2018) in Nairobi exploring FSW programme exposure – intervention depth, and behavioural outcomes – that equally found that 35% of the FSWs were not exposed to any HIV prevention programme.

In HIV treatment, care and prevention setups, standard treatment cascades are normally initiated with a diagnosis followed by referral to appropriate health care providers who then initiate a treatment plan with adherence by the HIV-infected persons (Blank & Eisenberg, 2014). So if access to health service providers and essential resources are existing, decreased infectiousness and better health outcomes follow when all components of the treatment cascade are fulfilled.

This is especially important in treatment as prevention (TasP). However, barriers may exist in every cascade step.

Obviously, adherence to HIV medication is important in the realization of long-term maintenance of VL suppression throughout the HIV treatment course (Günthard et al., 2014). WHO Adherence meeting participants in June, 2001, defined adherence as “the extent to which a person’s behaviour – taking medication, following a diet, and/or executing lifestyle changes, corresponds with agreed recommendations from a health care provider” (WHO, 2003) and emphasized the idea of the patient being an active collaborator in the treatment process instead of being “passive, acquiescent recipient of expert advice”. It has operational sub-units such as dosage, schedule and dietary adherence (Nilsson Schönnesson, Diamond, Ross, Williams, & Bratt, 2006).

Substance usage has been established to influence adherence to ART. Addressing ART adherence and substance use collectively is appropriate given their complex link (Parsons, Rosof, Punzalan, & Maria, 2005). High prevalence of violence syndemic, substance abuse, and HIV/AIDS among poor urban women is associated with poor outcomes of HIV (Hatcher, Smout, Turan, Christofides, & Stoeckl, 2015; Meyer, Springer, & Altice, 2011). Meyer et al. note that among HIV-infected, sexual risk-taking – including the use of substance during sex–and drug risk-taking (injection, drugs or alcohol or needle sharing) have profound effects on outcomes while Hatcher et al. reported significant violence association with lower ART use and VL suppression.

Symptoms of trauma, drinking, illegal drug use, including limited social support are connected with suboptimal utilization of HIV treatment (Meade, Hansen, Kochman, & Sikkema, 2009). The multifaceted poverty interaction may result in a syndemic of HIV-infection and mental illness. For instance, individuals reporting depression are likely to report nonadherence to HIV regimes and less likely to have low HIV VL levels (Tedaldi et al., 2012). Psychosocial factors such as distress and low self-efficacy are predictors of adherence (Naar-King et al., 2006). Among MSM at coastal Kenya, depressive symptoms and associated psychosocial factors that exacerbate the HIV burden and prognosis have been established (Secor et al., 2015).

Application of syndemic theory has been underscored in understanding the treatment side effect acting as a syndemic factor exposing PLHIV to a higher risk of developing negative health outcomes and creating circumstances where psychosocial factors are likely to arise (Gagnon, 2018).

Broadly, the Quality of healthcare outcomes is dependent upon patients' adherence to the treatment regimens recommended (Institute of Medicine (US) Committee on Health and Behavior: Research, Practice, 2001; Martin, Williams, Haskard, & Dimatteo, 2005). Given this, accurate assessment of adherence behaviour is essential for efficient and effective planning of treatment, and to ensure that health outcome changes are attributable to the regimen recommended, the scientific evaluation of treatment protocols, and even public health. Therefore to achieve the ideal clinical benefit, HIV-infected individuals receive medication continuously (Kim, Lee, Park, Bang, & Lee, 2018).

The syndemic interaction between the HIV and microbial and chronic diseases is also feasible. For instance, HIV tuberculosis (TB) syndemic increases the likelihood of death with HIV and also increases viral loads (HIV replication) and viral heterogeneity, more so at TB infection sites (Kwan & Ernst, 2011). The TB-HIV syndemicity is directly associated with VL and inversely with CD4 cell count (Sathekge, Maes, Van de Wiele, & Wiele, 2013). Chronic conditions may include diabetes (Byg, Bazzi, Funk, James, & Potter, 2016). At SWOP, longitudinal assessment of other sexually transmitted infections and chronic diseases is also done.

Within the Kenyan context, the concept of syndemic theory is very essential in sex workers as it allows for an assessment of (bio)social structures and risk environments that unite to yield disproportionate disease burden to specific members of society and challenges the reductionist discourse of disease (Douglas-Vail, 2016). The concept reaches beyond simple diseases and populations associations to grasp the health and society connections, exploring routes of transmission and interconnected health problems, that end in the excessive burden of disease (Singer, 2010).

This study looked at six syndemics: alcohol/substance use, chronic illness/TB-infected, microbial infections (STI/HCV/HBV/HPV), condomless sex and (intimate partner) violence/physical abuse and their interaction with psychosocial support.

1.2 Problem Statement

There are few peer reviewed studies on effects/associations of syndemics with HIV viral load trajectories in sex workers on ART that could influence policy. For instance, broadly using a specific search term on PubMed: (syndemic*[title/abstract] OR "synergistic pandemics"[title/abstract] OR "synergistic epidemics"[title/abstract]) AND "viral load"[title/abstract] AND (hiv[title/abstract] OR hiv/aids[title/abstract]) AND ("0001/01/01"[PDAT] : "2018/09/06"[PDAT]), gives only 12 studies. Few that exist have reported generalizability concerns.

However, there could be more studies on related risks and/or methodological factors or on segments of syndemics theory – for example, single PubMed database search: ("structural factors"[title/abstract] OR "social factors"[title/abstract] OR "psychosocial factors"[title/abstract] OR "psycho-social factors"[title/abstract] OR "behavioral factors"[title/abstract] OR "behavioural factors"[title/abstract] OR "biobehavioral factors"[title/abstract] OR "bio-behavioral factors"[title/abstract] OR "biobehavioural factors"[title/abstract] OR "bio-behavioural factors"[title/abstract] OR "individual factors" OR "biological factors"[title/abstract]) AND hiv[title/abstract] AND "viral load"[title/abstract] AND ("0001/01/01"[PDAT] : "2018/09/06"[PDAT]), yields only 95 PubMed studies globally.

Characteristically, it is easy to also notice that most analyses arising from the studies have inadequate social epidemiology insights on dynamic drivers of diseases (Noppert, Kubale, & Wilson, 2017) such as HIV. However, the contemporary setting of HIV epidemiology in Kenya may perhaps gain from approaches to understanding how syndemic problems sustain HIV spread or exacerbates HIV progression disproportionately in key populations.

Noticeably, two studies in MSM or minority men in the United States, have robustly reported on associations between selected syndemics and (non)adherence and (detectable) viral load (Friedman et al., 2015; Harkness et al., 2018). However, the Friedman et al.'s study has confessed reliability and validity concerns as well as over-representation by older participants. Few studies (e.g., Secor et al. 2015) in Kenya have considered syndemic theory to understand HIV in both seronegative and seropositive individuals but not within the context of viral load evolutions and ART adherence paradigm.

In the SWOP–Kenya, HIV-infected sex workers are observed over time. Modeling longitudinal data arising from such observations has many challenges in terms of both statistical and computational aspects. Statistical challenges occur due to complex dependence structures or the correlated nature (Cho, 2016). Regrettably, continuous collection, although occasionally nearly possible for certain outcomes is infeasible for virtually all variables. Realistically, the time-varying data are usually collected just when participants are observed (Hernán, McAdams, McGrath, Lanoy, & Costagliola, 2009).

The longitudinally collected progression markers on individuals may be missing attributable to some other reasons. Improper handling of missing data may decrease statistical power and result in biased parameter estimates (Matta, Flournoy, & Byrne, 2017), information loss, increased standard errors and leads to weakened generalizability (Dong & Peng, 2013) and therefore affect the validity of findings. This is where Linear Mixed-Effects Models (LMM) comes in handy.

1.3 Study Justification and Significance

There is a need of approaches to understanding HIV among Sex workers. Among sex workers, many syndemics such as alcohol/substance use, chronic illness/TB-infection, microbial infections (STI/HCV/HBV/HPV), condomless sex, (intimate partner) violence/physical abuse and nutritional insecurity – sex work among FSW is somewhat largely driven by poverty, and these have not been given adequate attention by research. Understanding these syndemic factors' relation with viral load evolutions and ART interruption can enlighten strategies/programmatic decisions for enhancing health outcomes or medication adherence and sex workers' retention in care in resource-poor settings – e.g in Kenya with evolving legal framework that broadly creates sex work criminalization conditions (KELIN, 2016).

There are public health implications for non-adherence. TasP is anchored on adherence because of viral suppression or an 'undetectable' viral load and reduced HIV transmissibility (WHO, 2012) and associated usefulness of early treatment (M. S. Cohen et al., 2011). The aforementioned reduced infectiousness and improved health outcomes take place when all basics of the treatment cascade are satisfied but syndemics may exist (Blank & Eisenberg, 2014). The magnitude of syndemics among sex workers' may vary considerably and all these underscores the rationale for this study.

This study, therefore, additionally fills any knowledge and/or research gaps in empirical HIV prevention literature on syndemics at the interface of biomedical, behavioural and systems research – especially in Kenya.

Observational studies may be suitable to show effectiveness/association (Yang et al., 2010) but are frequently critiqued for inferring uncorroborated causal inferences. In terms of statistical analysis, generating evidence to back up causal inference is generally the major aim. The SWOP–Kenya cohort has the strength of combining the classic observational, longitudinal cohort design with population-based cohort facility-level study of HIV/AIDS and HIV-related health problems.

Statistical methods for longitudinal studies are widely applicable to cohort studies. In observational health research where there is no interference with the care of patients, especially in longitudinal HIV/AIDS cohort observational and behavioral research, which is the focus of this study, data are often incomplete since some participants drop out (Wen, Terrera, & Seaman, 2018). The inherent features of the data collected are therefore crucial in deciding the finest statistical method to employ because of these methodological challenges.

A standard method used to model repeated measurements is the "repeated measures analysis of variance" (RM-ANOVA). RM-ANOVA is suitable for discrete predictors, a complete dataset – listwise/case deletion is used in case of missing data, and when measurements are taken at similar occasions (balanced) for all the participants (Fitzmaurice, Laird, & Ware, 2011) and an assumption of homogeneity of variances/equal covariance between all the observed outcomes (McCulloch, 2005) which is untenable in longitudinal studies.

In classical regression, more so multiple linear regression, the assumption of independence of measurements (outcome variable) and the independent identical distribution of error terms is maintained. However, in longitudinal measurements from the same patient tend to be alike (intra-patient (intra-cluster) correlation) but vary between patients (inter-patient (inter-cluster) variability), hence the need to be accounted for in the data analysis hence wrong to utilize standard methods.

Due to the longitudinal nature of the repeated measurements, special analysis procedures are needed to account for dependence. LMM are employable to systematically allow incorporation

of fixed-effects and random-effects when there is non-independence in the data. Since longitudinal follow-up studies are also affected by missing data (Karahalios et al., 2013) frequently causing complications in the analysis, LMM can be efficiently used to analyze it with the missing at random assumption. There are limited studies on modeling prognosis of HIV markers in sex workers in-care.

This study appreciates the utility of LMM for longitudinal VL data analysis. The outcomes provide useful evidence in the identification of potential hotspots for TasP within SWOP–Kenya and potential scale to Kenya HIV programmes.

1.4 Objectives

1.4.1 General Objective

To examine and model the association of syndemics with HIV viral load outcomes as well as ART adherence among sex workers living with HIV in the SWOP–City cohort using linear mixed-effects.

1.4.2 Specific Objectives

- i) To examine the influence of syndemics on HIV viral load and medication adherence
- ii) To determine whether the effect of syndemics on viral load is modified by linkage to psychosocial support.
- iii) To determine if a random intercept model is sufficient or the longitudinal model needs to have random slope effects.

1.5 Assumptions of the Study

It was assumed that data are “missing at random” (MAR). Data missing mechanisms are often of three types: MAR, “missing completely at random” (MCAR), and “missing not at random” (MNAR) (Little & Rubin, 2002). MAR implies that there could be systematic differences between the observed and the missing values, however, these can be entirely described by other observed variables (i.e., the data missingness could be entirely described by variables on which full information is available). In the SWOP–Kenya data, missing observations are expected to occur for these different reasons.

For example, in the SWOP –Kenya, if HIV viral load values (or adherence values) is missing conditional of age and gender only, having complete data of gender and age variables would

amount to a MAR mechanism. Similarly, a situation in which the missing values are greatly prevalent in a subgroup taking a particular regimen relative to another was considered MAR since the distributions of missing and observed values are similar within the age/gender strata (Bhaskaran & Smeeth, 2014).

Data missingness is a common issue in epidemiologic/medical studies and arises from a lack of response, because of nonresponse (refusal), loss to follow-up or due to clinical-care related decisions (Abraham et al., 2011). The missing observations could also be because of measurement thresholds outside the testable range.

CHAPTER TWO

2. LITERATURE REVIEW

2.0 Introduction

This chapter is a statistical literature review mostly based on Molenberghs and Verbeke (2000)'s longitudinal linear mixed models. Fitting of the LMMs is reviewed leading to the final model fitted in Chapter 3. Estimation and inferences for fixed effects are reviewed. The literature on the need for random intercept and random slope models for estimation and inference for variance components is also reviewed.

2.1 Methods for Longitudinal Clustered Data Analysis

Multiple indicators have been increasingly used to assess the outcomes of interest in human and biomedical sciences and many other scientific applications. Many of the kinds of data collected, including observational data, have a hierarchical or clustered structure (Costa, Colosimo, Vaz, Silva, & Amorim, 2017) due to repeated or sequential measures from a set of investigational units over time, a defining feature of longitudinal studies. The type of data that therefore arise is typically dependent, attributable to these repeated observations resulting in within-subject dependence from subject-specific characteristics (Galbraith, Daniel, & Vissel, 2010). Within a cluster, observations are “more alike” than those observations from other different clusters.

In longitudinal studies, analysis of the data may encompass modeling the marginal probability response (marginal analysis) i.e interest is only on the average response (Pavlou, Ambler, Seaman, & Omar, 2015). It may also involve, transition modeling that focuses on how the outcome (Y_{it}) is dependent on previous values of Y and other variables (i.e., a conditional model) or linear mixed models (random effects models) based on assumption of normality focusing on how coefficients of regression vary over individuals (Long, Loeber, & Farrington, 2009). Longitudinal data is therefore essential in the study of trends of change and the factors affecting it, between and within individuals.

In random-effects models, the assumption is that there is a correlation between observations in the same cluster (individual), whereas no correlation exists between observations from different ones (Costa et al., 2017). The nature of the data collected is influential in the best statistical approach to be taken (Galbraith et al., 2010).

2.2 Mixed-Effects Models for Longitudinal Data

Mixed-effects model use fixed and random effects together in the same data analysis. According to Molenberghs & Verbeke (2000), these models assume that observations from a individual subject share “a set of latent, unobserved, random effects” useful in generating “an association structure between the repeated measurements” and hence longitudinal analysis should recognize the serial correlation between observations from the similar unit (Laird & Ware, 1982). With repeated normally distributed data, a very general and flexible, class of parametric models is got from a random-effects approach, hence the linear mixed models. Other models for discrete data are the Generalized Linear Mixed Models and Liang and Zeger (1986)’s generalized estimating equations. The focus of this study is LMM.

2.3 Linear Mixed Models

Linear mixed models are also called Random effects models, Variance components models, Mixed-effects models. Linear mixed models, as an extension of simple linear models, are regression models that take into account both variation that is explained by the explanatory variables of interest i.e fixed effects, and variation that is not explained by these explanatory variables i.e random effects. The random effects, in essence, give structure to the error term ϵ . One clear advantage with this longitudinal method is the normality assumption (linear) that individual-level and group-level information is incorporated in the same model (Long et al., 2009).

They are predominantly useful in situations where there is dependence on the data, for instance, is arising from a hierarchical structure. A practical example is: within Kenyatta National Hospital, patients may well be sampled from within doctors, and equally, doctors sampled from practices hence variability treated as within group or between groups. The observations are then from clusters. Linear Mixed models are useful in the longitudinal analysis of data since they are both flexible and extensively applicable, and given that software implementation are available for fitting them (Van Montfort, Oud, & Satorra, 2010).

LMM is an extension of a linear regression model with random intercept and slope. The general LMM form is similar for longitudinal and clustered observations.

2.3.1 Random Slope and Random Intercept Model

2.3.1.1 Classical linear regression

Given a set of data $\{y_i, x_{i1}, \dots, x_{ip}\}_1^n$ with n units, a classical linear regression, tries to model the relationship between the dependent variable y and the p -vector of independent variables x by fitting a linear equation to the data observed.

$$(1) \quad y_i = \beta_0 \mathbf{1} + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i = \mathbf{X}'_i \boldsymbol{\beta} + \varepsilon_i, \text{ where } i = 1, \dots, n, \text{ and } \varepsilon_i \sim N(0, \sigma^2)$$

$\mathbf{X}'\boldsymbol{\beta}$ is the inner product between $\boldsymbol{\beta}$ and \mathbf{x}_i : denoted: $\mathbf{y} = \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ in matrix form,

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}'_1 \\ \mathbf{X}'_2 \\ \vdots \\ \mathbf{X}'_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

- \mathbf{Y} is a vector of observed values y_i values response variable.
- \mathbf{X} is the design matrix of row-vectors \mathbf{x}_i or of n -dimensional column-vectors \mathbf{x}_j , which are explanatory variables,
- $\boldsymbol{\beta}$ is a $(p+1)$ -dimensional parameter vector, where β_0 is the intercept term
- $\boldsymbol{\varepsilon}$ is a vector of values ε_i , error term.

In LMM, if $\mathbf{y}_i = (y_i \dots \dots y_{1n_i})^T$ is a vector of repeated measurements and T the transpose, then the general model is $\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\beta}$ is a population average regression coefficients vector of fixed effects and \mathbf{b}_i is a subject-specific regression coefficients vector that describes the deviation of evolution of the i th individual from the average population evolution. \mathbf{X}_i is an $n_i \times p$ and \mathbf{Z}_i $n_i \times q$ matrices of known covariates. The residual components $\boldsymbol{\varepsilon}_i$ are assumed to be independent $\mathbf{N}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_{n_i})$, where $\sigma_\varepsilon^2 \mathbf{I}_{n_i}$ is dependent on i only through its size n_i (Molenberghs & Verbeke, 2001). \mathbf{b}_i and $\boldsymbol{\varepsilon}_i$ are independent.

2.3.1.2 Random intercept model

Random intercept model is the simplest mixed model for longitudinal data.

Notation: Considering n patients in a cohort study, where follow up is over time t

$$(2) \quad Y_i = \beta_0 + \beta_1 t_{i1} + \varepsilon_{ij}, \text{ where } i = 1, \dots, \dots, n,$$

where $y_{ij} = 1, \dots, n_i$, Where Y_{ij} are measurements of the individual taken at time j ,

and Y_{ijt} = measurement for subject i taken at time t (t_{ij})

Model (2) assumes independence as in model (1) above. That is, they have the same intercept for all subjects and the same slope for all subjects. In random intercept model, patients/individuals have their own starting points – own intercepts b_{0i} : $i = 1, \dots, n$ but the slope β_1 remains the same.

$$(3) \quad Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + \varepsilon_{ij},$$

$$b_{0i} \sim N(0, \sigma_b^2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

ε_{ij}, b_{0i} , are independent

Each cluster/subject has its own intercept β_0 and b_{0i} and reflects the inter-subject variability at baseline. Apart from time, fixed effects, such as treatment group, demographic information, interactions among others, can be added into the model (3).

2.3.1.3 Random slope model

In Random slope model, each patient/cluster also has his/her own slope b_{1i}

$$(4) \quad Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} + \varepsilon_{ij},$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

$$\mathbf{D} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}, \varepsilon_{ij}, \text{ independent of } b_{0i}, b_{1i}.$$

In this model, each individual has his/her own $\beta_1 + b_{1i}$ in addition to individual intercept $\beta_0 + b_{0i}$. It allows the profiles to cross each other when plotted. Fixed effects too can be included in (4).

The random effects covariance: There are now two random effects: b_{0i}, b_{1i} and their covariance $\sigma_{01} = \sigma_{10}$. Positive covariance shows that the subject higher at baseline also higher

evolution while negative covariance shows that the subject higher at baseline, slower evolution.
The covariance unrestricted (unstructured).

2.3.2 Conditional and Marginal distributions: Random intercept model

Take $E(Y|b)$ and $\text{Var}(Y|b)$

$$(5) \quad Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} + \varepsilon_{ij},$$

$$(6) \quad E(Y_{ij}|b_{0i}) = \beta_0 + \beta_1 t_{ij} + b_{0i}$$

$$(7) \quad \text{Var}(Y_{ij}|b_{0i}) = \sigma_\varepsilon^2$$

$$Y_{ij}|b_{0i} \sim N(\beta_0 + \beta_1 t_{ij} + b_{0i}, \sigma_\varepsilon^2)$$

Marginal distribution (marginal over the random intercepts):

Take $E(Y)$ and $\text{Var}(Y)$

$$(8) \quad Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} + \varepsilon_{ij},$$

$$(9) \quad E(Y_{ij}) = \beta_0 + \beta_1 t_{ij}$$

$$(10) \quad \text{Var}(Y_{ij}) = \sigma_b^2 + \sigma_\varepsilon^2$$

$$Y_{ij} \sim N(\beta_0 + \beta_1 t_{ij}, \sigma_b^2 + \sigma_\varepsilon^2)$$

The implied intra-subject correlation: Measurements from the same subject share a random effect and this means that, there is a correlation structure. Considering two measurements from the same subject: Y_{ij} and $Y_{ik}, k \neq j$

$$(11) \quad \text{Cov}(Y_{ij}, Y_{ik}) = \sigma_b^2$$

Correlation between the two measurements from the same subject is given by:

$$(12) \quad \text{Corr}(Y_{ij}, Y_{ik}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2}$$

Exchangeable correlation/compound symmetry: also known as intra-cluster (intra-class) correlation. That is, any two measurements taken have the same correlation. This correlation

structure is called compound symmetry. It has a constant correlation between any two measurements regardless of the time interval, for instance, $\text{Corr}(Y_{11}, Y_{15}) = \text{Corr}(Y_{1j}, Y_{19})$. However, it may be too restrictive.

2.3.3 Conditional and Marginal distributions: Random slope model

The conditional distribution is given by

(13) $E(Y_{ij}|b_{0i}, b_{1i}) = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij}$ and $\text{Var}(Y_{ij}|b_{0i}, b_{1i}) = \sigma_\varepsilon^2$; while the marginal distribution is

(14) $E(Y_{ij}) = \beta_0 + \beta_1 t_{ij}$ and $\text{Var}(Y_{ij}) = \sigma_1^2 t_{ij}^2 + 2\sigma_{01} t_{ij} + \sigma_0^2 + \sigma_\varepsilon^2$ and is a function of time.

The intra-cluster correlation:

Y_{ij} and $Y_{ik}, k \neq j$,

(15) $\text{Cov}(Y_{ij}, Y_{ik}) = \sigma_1^2 t_{ij} t_{ik} + \sigma_{01} (t_{ij} + t_{ik}) + \sigma_0^2$ and the ICC a function of time

(16)
$$\text{Corr}(Y_{ij}, Y_{ik}) = \frac{\sigma_1^2 t_{ij} t_{ik} + \sigma_{01} (t_{ij} + t_{ik}) + \sigma_0^2}{\sqrt{\sigma_1^2 t_{ij}^2 + 2\sigma_{01} t_{ij} + \sigma_0^2 + \sigma_\varepsilon^2} \sqrt{\sigma_1^2 t_{ik}^2 + 2\sigma_{01} t_{ik} + \sigma_0^2 + \sigma_\varepsilon^2}}$$

In general, LMM conditional distributions is given by

(17) $E(\mathbf{Y}_i | \mathbf{b}_i) = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i$; $\text{Var}(\mathbf{Y}_i | \mathbf{b}_i) = \sigma_\varepsilon^2 \mathbf{I}_{n_i}$ and the marginal:

(18) $E(\mathbf{Y}_i) = \mathbf{X}_i \boldsymbol{\beta}$ and $\text{Var}(\mathbf{Y}_i) = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \sigma_\varepsilon^2 \mathbf{I}_{n_i}$.

CHAPTER THREE

3. METHODOLOGY

3.0 Introduction

In the SWOP–Kenya data, viral load measurements are from each sex worker (cluster) hence age effects (within cluster temporal changes) and cohort effects (differences between clusters at baseline) and predictors can be studied. In this chapter, the study design, sample size and variables/measures is described. Detailed graphical methods for the viral load data exploration were undertaken for its underlying structure. LMM study methods described in chapter 2 were applied to the SWOP data to describe the log viral load evolutions with and model the syndemics while attempting to model intra-cluster correlation.

3.1 Study Design

This was a retrospective study from a well-established cohort using data collected routinely from sex workers' clinical database for HIV prevention, Care and Treatment maintained by University of Manitoba and University of Nairobi. Only HIV-infected sex workers were eligible for this study, aged ≥ 18 and on ART between August 2013 and August 2018. Systematic Random Sampling method was used and simple random sampling used to determine the starting point for the systematic random sampling.

3.2 Sample size

The sample size was calculated by taking the median month on ART (24 months). Since statistical power is used in retrospective studies, sample size estimation was done following the work of Diggle and Diggle (2002) and Liu and Liang (1997) on sample size calculation for longitudinal studies. This was implemented using the R software longpower statistical package in R version 3.5.1. The percentage effect of treatment done with 80% power with the level of significance, $\alpha = 0.05$. The calculation carried out for a random intercept model. The treatment effect was obtained by calculating the mean change in treatment using initial viral and the follow up viral loads and found to be 74%.

```
> require(longpower)
```

```
> require(lme4)
```

```
> i.swop<-read.csv(file.choose())
```

```
> sample <- lmer(VL ~ months, random=~1|id, i.swop)
```

```
> lmpower(sample, pct.change = -.74, t = seq(0,24), power = 0.80) # n=291*2
```

N=582 and adjusted for 5% data management and missingness to 615.

This sample size calculation agreed with the simulations of Wang and Xue (2016) in the presence of the effect of medication adherence and modifying role of linkage to psychosocial support and the effect of the syndemic covariates.

3.3 Ethical Consideration and Data Management

3.3.1 Ethical considerations

Administrative permission to use the SWOP data was obtained from the University of Manitoba/University of Nairobi – PHDA. It was carried out in compliance with the principles encompassed in the Helsinki Declaration. As a requirement, it was presented to the Ethics Review Committee of the Kenyatta National Hospital/the University of Nairobi –KNH/UoN ERC for ethical review and approval and as part of clinical database use and as a subset of research studies carried out by the University of Manitoba and the University of Nairobi (Appendix A).

To ensure the patients remain anonymous and guarantee confidentiality, all unique personal identifiers in the analysis data were removed, and instead replaced with pseudonyms and were suitably coded to render it unidentifiable, during file conversion from the programme database to Excel spreadsheet.

3.3.2 Data Management

Meticulous data cleaning was carried. Often with data obtained from programmes, several files were found that were subsets of each other and with information that was required for this study. R software was used to merge and append as necessary, by making use of merge() or append() functions. The dplyr package's left, right, inner, anti, outer and semi join functions were used for inclusion of columns of data.

The datasets and their file subsets had unique alphanumeric identifiers and systematic numeric codes. Other subsets had purely long numeric code identifiers. Separation of the recurring codes was done in order to have the digits only comparable to the other data subsets' numeric part of the alphanumeric identifiers.

Effective data merge devoid of information loss was done after separation alphanumeric into digits and characters by use of R software's `colsplit` function 'within the `reshape2` package implemented as `s<-colsplit(df$id, "(?<=\\p{L})(?=[\\d+])", c("char", "digit"))`). The numeric part was then used for joining/merging the files of data. This was enhanced with Microsoft Excel's `RIGHT` and `LEN` split functions.

3.4 Variables and Measures

3.4.1 Sociodemographic covariates

These included age, categorical education (completed primary, did not complete primary, completed secondary, did not complete secondary, completed tertiary, did not complete tertiary, never attended school). Categorical marital status (i.e, single, married and separated/divorced, widowed), binary key population type (FSW and MSMW).

3.4.2 Other baseline variables

Baseline \log_{10} transformed viral load and recent Nutritional status.

3.4.3 Outcome variable

Continuous HIV-RNA (viral load) and Box-Cox transformed HIV viral load measured at the SWOP–City clinic. During modelling, the viral load was \log_{10} transformed to correct right skewness.

3.4.4 Medication adherence

Adherence to ART and CTX measured on an ordinal scale self-reported since the previous visit. An ordinal variable with options comprised of: Good; Fair; Poor and in some cases, informative “not applicable” due to clinical assessments. Self-report adherence measures have been found to have reliable concurrent and predictive validity with measures of disease progression including viral load (Lazo et al., 2007; Nieuwkerk & Oort, 2005; Wilson et al., 2002).

3.4.5 Syndemic variables/factors

Binary intimate partner violence/physical abuse, binary alcohol and substance use, binary chronic illness/TB-infected, binary microbial infections (STI/HCV/HBV/HPV) and binary condomless sex.

3.4.6 Effect modification variable

Binary linkage to psychosocial support – including information on safer sex practices, medically assisted therapy/naloxone, condom information.

3.4.7 Data Analysis

The data was analyzed using R Core Team (2018). The lme4 package (Bates, Mächler, Bolker, & Walker, 2014) and the nlme package (Pinheiro, Bates, DebRoy, Sarkar, & Team, 2018), later implementations of lme4 (lmerTest) and potentially lme4cens if some of the viral loads are left censored (Kuhn & Roeder, 2018) was used. However, lme4cens wasn't used since log to base 10 transformations were considered.

3.4.7.1 Graphical methods for exploring longitudinal data

Graphical methods enable exploration of mean structure and provision of an idea as to reasonable functional form for time for log viral load evolutions (i.e linear relationship with time or quadratic) (Molenberghs et al., 2000). These methods also give an idea of what kind of random effects structure/should be included in the model (random intercepts or random slopes) and explore the covariance structure (Appendix B).

3.4.7.2 Estimation of longitudinal medication adherence effect

In the SWOP–Kenya, HIV viral loads measurements, denoted here as Y_{ij} , were taken repeatedly in an individual i ($i=1 \dots n$). Overall, Y_{ij} for the retrospective period under this study were measured 5 times ($j=1 \dots 5$). The levels of self-reported medication adherence (M_i) were also assessed in each individual in every visit time-point. For individual i , let X_i connote a syndemic (or a vector of syndemics) and Z_i indicate individual i 's set of covariates e.g level of education, gender etc. while t_{ij} , the visit time-point, importantly, $t_{ij} = 0$, at baseline. $\mathbf{Y}_i = (Y_{i1} \dots Y_{ij})^T$ represents the transposed (T) matrix of the viral load outcomes.

Conversely, if HIV viral loads measures were taken one-off (cross-sectional), a classical linear regression with a single-level of medication adherence mediation (M) would be presented as:

$$(19) \quad Y = \beta_{01} + \beta_c X + \varepsilon_1 ; Y = \beta_{02} + \beta_{c'} X + \beta_b M + \varepsilon_2 ; \text{ and } M = \beta_{03} + \beta_a X + \varepsilon_3,$$

where: β_c = the relationship between the predictor (syndemics) variable and outcome variable (viral load). $\beta_{c'}$ = the relationship between predictor (syndemic(s)) variable and the outcome

variable after adjusting for the mediation effect, β_b = the relationship between the mediator variable and the outcome variable after adjusting for effects of the predictor variable; β_a = the relationship between the predictor variable and the mediator; β_{01} , β_{02} and β_{03} = the intercepts; ε_1 , ε_2 , and ε_3 = the corresponding error terms; and σ_1^2 , σ_2^2 and σ_3^2 = the corresponding variances.

And the medication adherence's mediating effect could be described by: $\beta_c - \beta_c'$ and $\beta_a\beta_b$ (Judd & Kenny, 1981a, 1981b; D. MacKinnon, 2008) which are all equivalent (D. P. MacKinnon, Warsi, & Dwyer, 1995).

But then this is generally different in the longitudinal mediation models since assumption of independence is violated (P. Cohen, West, Aiken, West, & Aiken, 1984; Steenbergen & Jones, 2002) and leads to biased standard errors. This makes longitudinal mediation models to be complex due to multilevel nature (Pan, Liu, Miao, & Yuan, 2018). Using Wang and Xue (2016)'s format, longitudinal effect of X on M maybe be modeled as:

$$(20) \quad \mathbf{M}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_a\mathbf{X}_i + \boldsymbol{\zeta}^T\mathbf{Z}_i + \boldsymbol{\varepsilon}_i, \text{ where } \boldsymbol{\zeta} \text{ is a coefficients' vector for the covariates' vector } \mathbf{Z}; \boldsymbol{\varepsilon}_i \text{ is normally distributed random error with mean}=0.$$

According to Laird and Ware (1982), the linear mixed-effects model for the evolutions of outcome \mathbf{Y} given X (or vector of X), M and \mathbf{Z} is:

$$(21) \quad Y_{ij} = \beta_0 + \beta_x^c X_i + \beta_m^c M_i + \beta_t t_{ij} + \beta_x^l X_i t_{ij} + b M_i t_{ij} + \boldsymbol{\phi}^T \mathbf{Z}_{it} + \varepsilon_{ij}, \text{ where } b \text{ quantifies } M\text{'s longitudinal effect on } \mathbf{Y} \text{ controlling for } X \text{ (or } \mathbf{X}) \text{ and } \mathbf{Z} \text{ and } \beta_x^l \text{ computes the longitudinal effect of } X \text{ on } \mathbf{Y} \text{ adjusted for } M \text{ and } \mathbf{Z} \text{ (the } Z \text{ here is different from that in equation (17) and (18) in chapter 2). } \beta_x^c \text{ and } \beta_m^c \text{ are baseline effects for } X \text{ and } M \text{ correspondingly. Details of this, including hypothesis testing for both effects, can be got in Wang and Xue (2016)'s papers where the expectation of } E(Y_{ij}|X_i, M_i, \mathbf{Z}_i) \text{ and } E(Y_{ij}|X_i, \mathbf{Z}_i) \text{ are given and the difference in the coefficient's of } X_i t_{ij} \text{ method for the longitudinal effects (a*b for fixed effects model) is presented.}$$

3.4.7.3 Estimation of effect of linkage to psychosocial support (effect modification)

The effect of linkage to psychosocial support was modelled as interaction effects/effect modifiers. This helped in understanding the effect of the interventions in the SWOP –Kenya

project. Psychosocial support may have a relationship as previously seen by Friedman et al. (2017) on HIV viral load. This was assessed/verified with SWOP data by inclusion of an interaction term followed by the test its significance.

3.4.7.4 Estimation of the need of, or the effect of random effects

This was conducted by computation of the evolutions in \log_{10} viral load over time by adherence and model fitting under maximum likelihood (ML) (Appendix C). Test for the need of the random slopes and the random intercepts was conducted and the model for population average fitted values and subject-specific predictions obtained (Appendix F and D). Within cluster resampling method/multiple outputation was applied to the data as described by Follmann and Fay (2010) and Follmann, Proschan, and Leifer (2003) due to large of proportion of non-clustered part of the SWOP data (Appendix G).

CHAPTER FOUR

4. RESULTS

4.0 Introduction

In Chapter 4, results from the application of the mixed-effects linear approach to the study of the association of syndemics with viral loads in HIV-infected sex workers are presented. The sociodemographic data of FSW and MSM are compared.

4.2 Sociodemographic and epidemiologic profiles of the sex worker cohort

Table 1 below shows the results from a Chi-square test of differences between FSW and MSMW. Age, marital status and highest education level with the exception of GBV in the past three months, were statistically significant all with p-value<0.0001.

Table 1: Sociodemographic and epidemiological profiles of sex workers

| N = 621 | FSW | MSMW | p-value |
|----------------------------|---------------|-------------|---------|
| | 502 [80.8%] | 119 [19.2%] | |
| | Frequency [%] | | |
| Age | | | |
| 18–25 | 26 [5.2] | 42 [35.3] | <0.0001 |
| 26–30 | 74 [14.7] | 34 [28.6] | |
| 31–35 | 109 [21.7] | 22 [18.5] | |
| 36–40 | 110 [21.9] | 11 [9.2] | |
| 41–45 | 83 [16.5] | 7 [5.9] | |
| 45+ | 100 [19.9] | 3 [2.5] | |
| Marital status | | | |
| Divorced/Separated/Widowed | 36 [45.6] | 6 [15.9] | <0.0001 |
| Married | 0 [0.0] | 2 [5.2] | |
| Single | 43 [54.4] | 30 [78.9] | |
| Highest education level | | | |

| N = 621 | FSW | MSMW | p-value |
|--------------------------|---------------|-------------|---------|
| | 502 [80.8%] | 119 [19.2%] | |
| | Frequency [%] | | |
| Secondary or below | 76 [96.2] | 25 [65.8] | <0.0001 |
| College or above | 3 [3.8] | 13 [34.2] | |
| GBV in the last 3 months | | | |
| Yes | 15 [19.2] | 2 [5.6] | 0.05673 |
| No | 63 [80.8] | 34 [94.4] | |

4.3 Viral load distributional properties

4.3.1 Key population and age-stratified viral load distributions on a logarithmic scale, \log_{10} copies per mL

Figure 1 shows a histogram of HIV viral loads by key population type. The viral loads (VL) are right-skewed hence violates the normality assumption. To normalize the VL data, a transformation of the form $\log(x+1)$ was done to the VLs, where x is follow up VL. One was added since $\log_{10}(0) = \text{infinity}$. When direct log transformation was first carried out, there were virally suppressed groups that were below the limit of detection (LOD) of the laboratory machine (usually <50 copies/ml). This meant that there were values of zeroes corresponding to the LOD.

A subsequent plot is illustrated in figure 2 below showing the transformations. After the transformation, the data appears normally distributed.

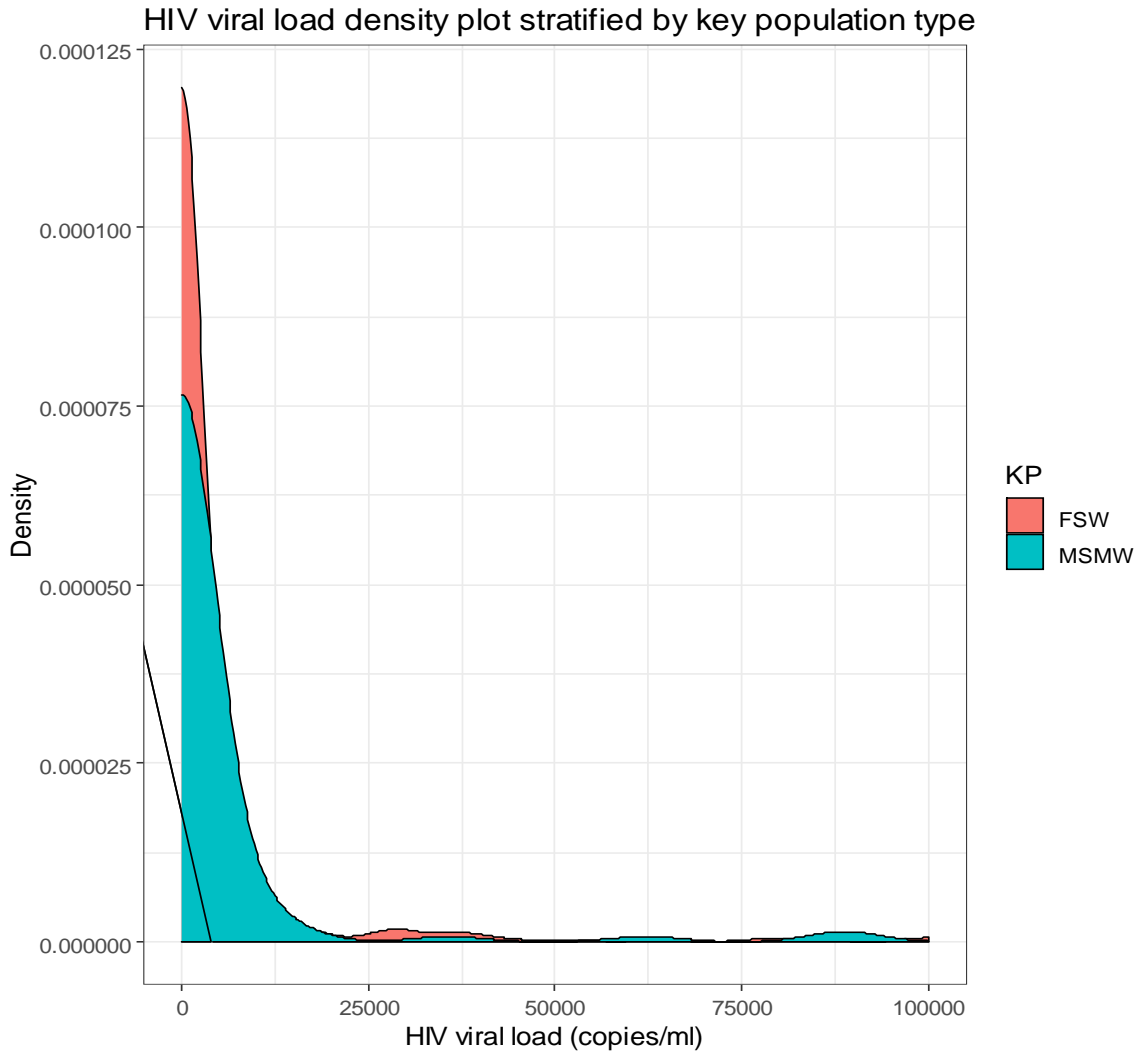


Figure 1: HIV viral density by key population type

Log transformred HIV viral load density plot by key population type

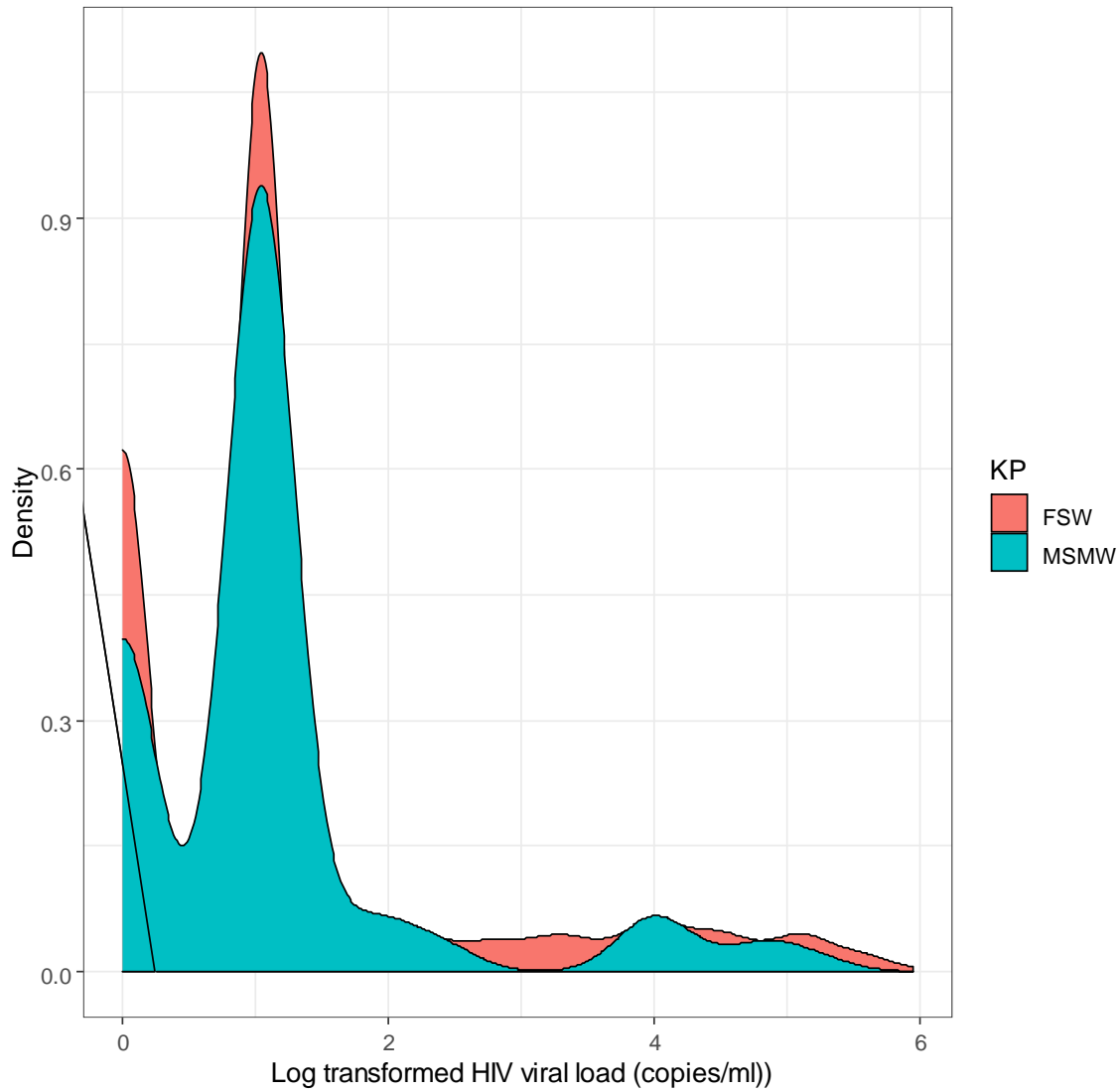


Figure 2: Log-transformed HIV viral load by key population

A linear model was fitted with ggplot2 disregarding the key population type and other variables. The model was fitted with log viral load as the response and age as the predictor. From the output below, it seems that older sex workers have somewhat higher viral loads. Figure 3 – 5 illustrates the viral load distributions.

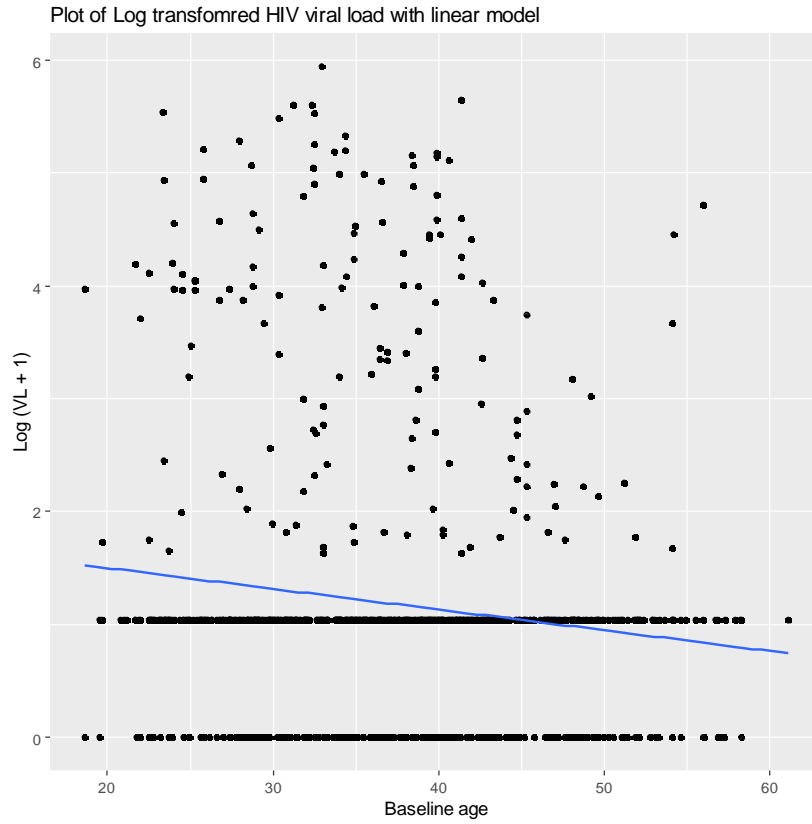


Figure 3: Log-transformed HIV viral load with a linear model of baseline age

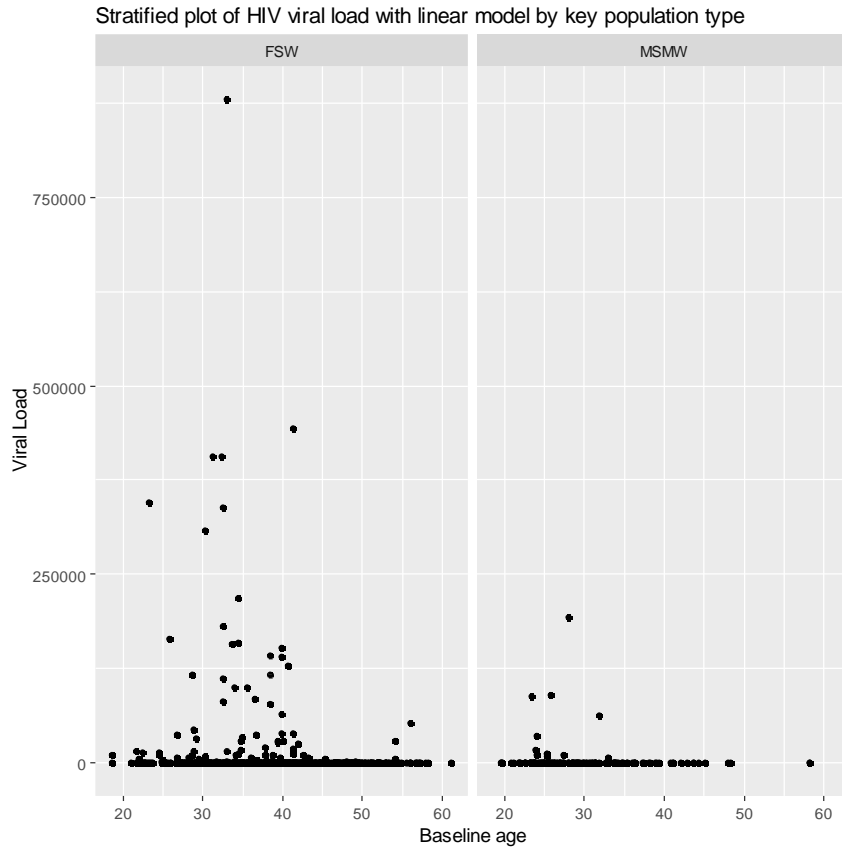


Figure 4: Stratified log HIV by baseline age

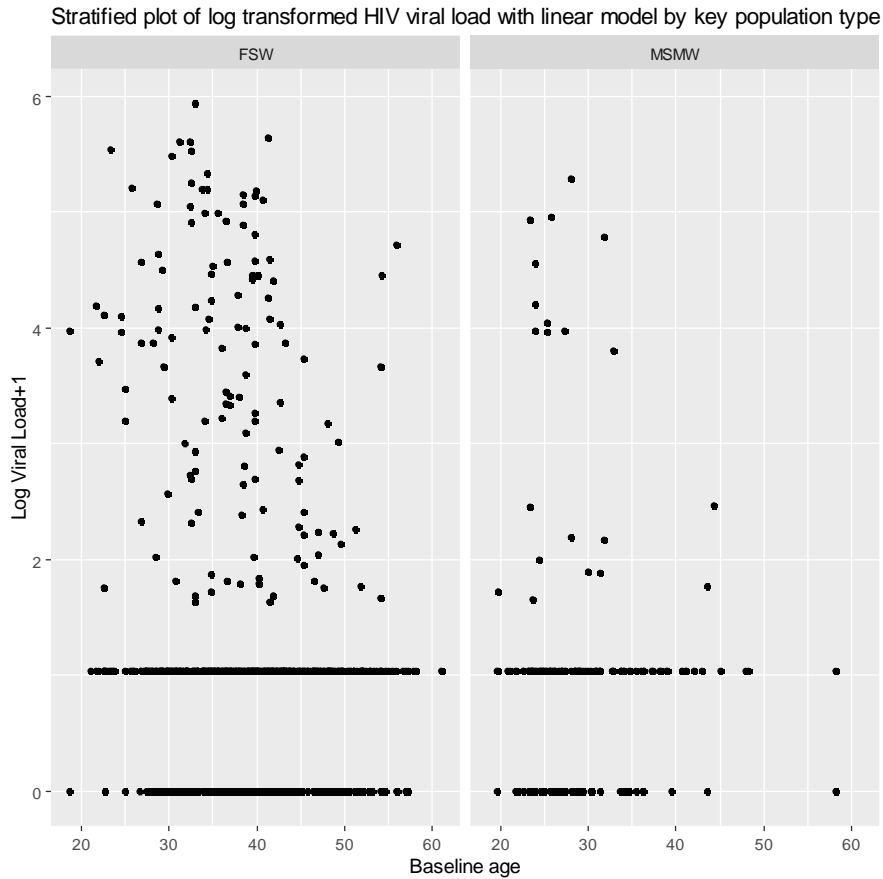


Figure 5: Stratified plot of log HIV VL with a linear model

A plot of the residuals is shown in figure 6 below. The red line seems to be nearly flat, similar to the faint grey dashed line around zero. The plot appears to suggest the presence of unusual dataset given the points above and below the red line.

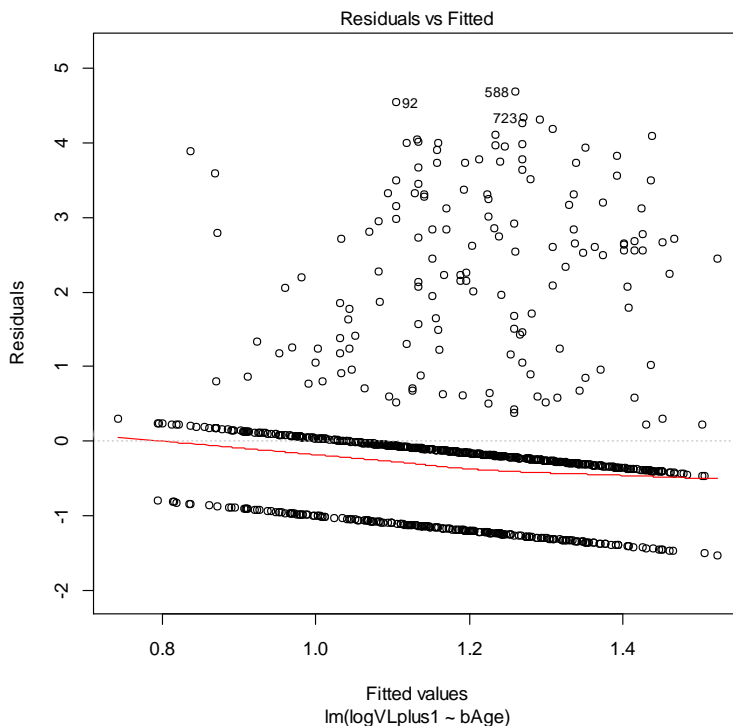


Figure 6: Residuals vs Fitted plot of log HIV VL +1

On subsequent fitting of qqplot shown in Figure 7, the points do not fall onto the diagonal dashed grey line. This plot exemplifies all the characteristics of badly-behaved residuals against fits plot. The expectation is that the residuals bounce randomly around the zero line and that no single residual stands out from the pattern. It also shows that there are two sets of observations – those contributing one and those with at least one observation. This shows that the assumption of linearity is not feasible.

The independence assumption is violated. This is because the sex worker's data were collected multiple times leading to clusters of observations per sex worker of different sizes. It's therefore reasonable to note that data from the same sex worker are more alike hence correlated.

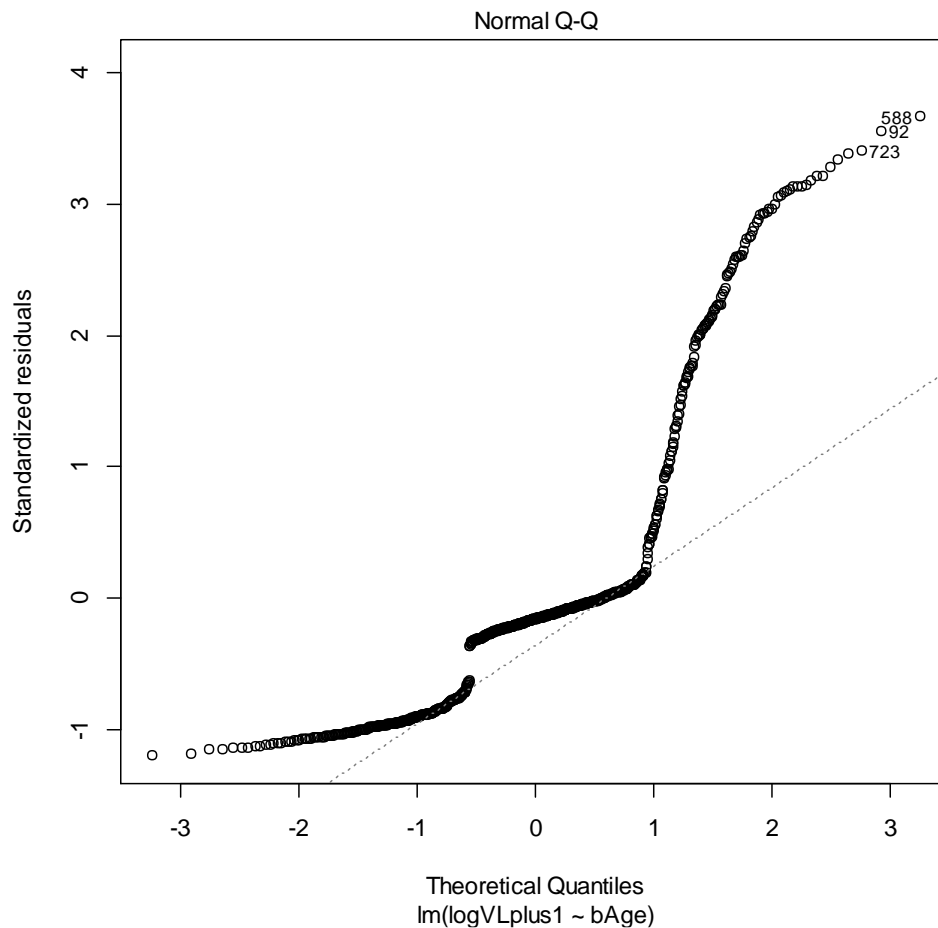


Figure 7: Normal Q-Q plot

4.2.1 Box-Cox transformation of the HIV viral loads

Linear mixed effects model assumption is that the model residuals are normally distributed. The HIV viral loads were highly positively skewed. The statistical literature typically recommends log normalization already implemented and shown in Figure 2 and a Box-Cox transformation (Box & Cox, 1964) for such positively skewed data. However, even after log transformation, the viral loads still appear skewed. A histogram plot of the log-transformed HIV viral load shows the multimodal appearance of the viral load data as shown in Figure 8 below. This still violates the normality assumptions.

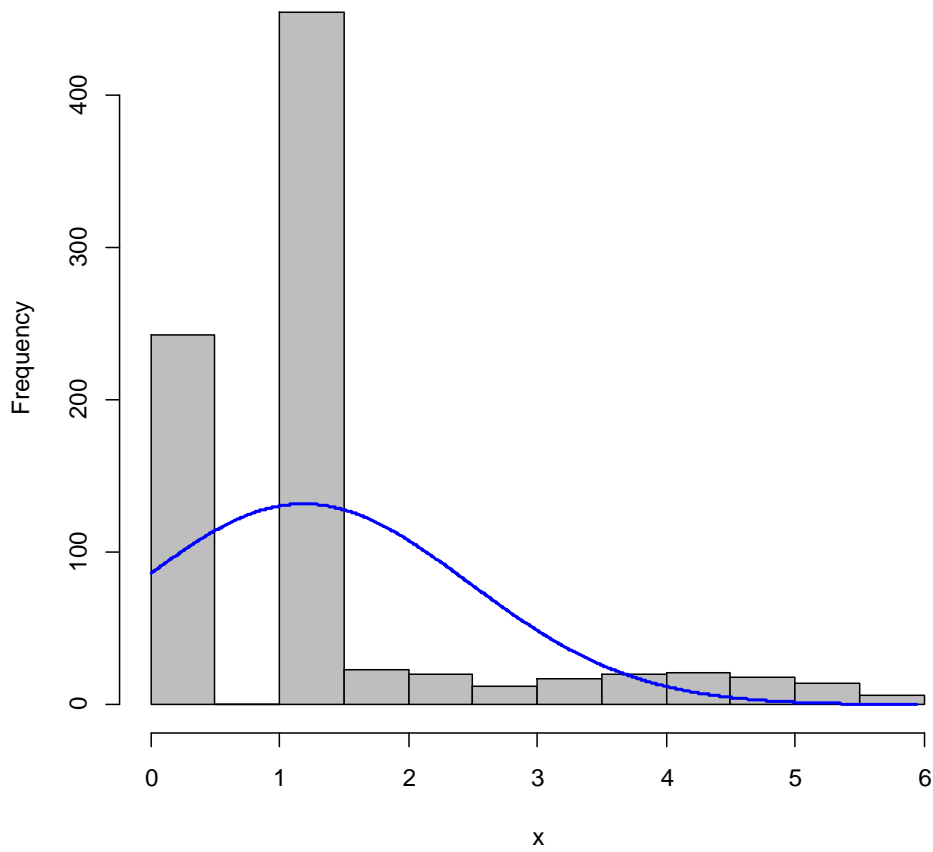


Figure 8: histogram plot of the log-transformed HIV viral load

So a Box-Cox transformation of viral load +1 was implemented to handle this and a resulting histogram plot showed normally distributed viral loads (Appendix E).

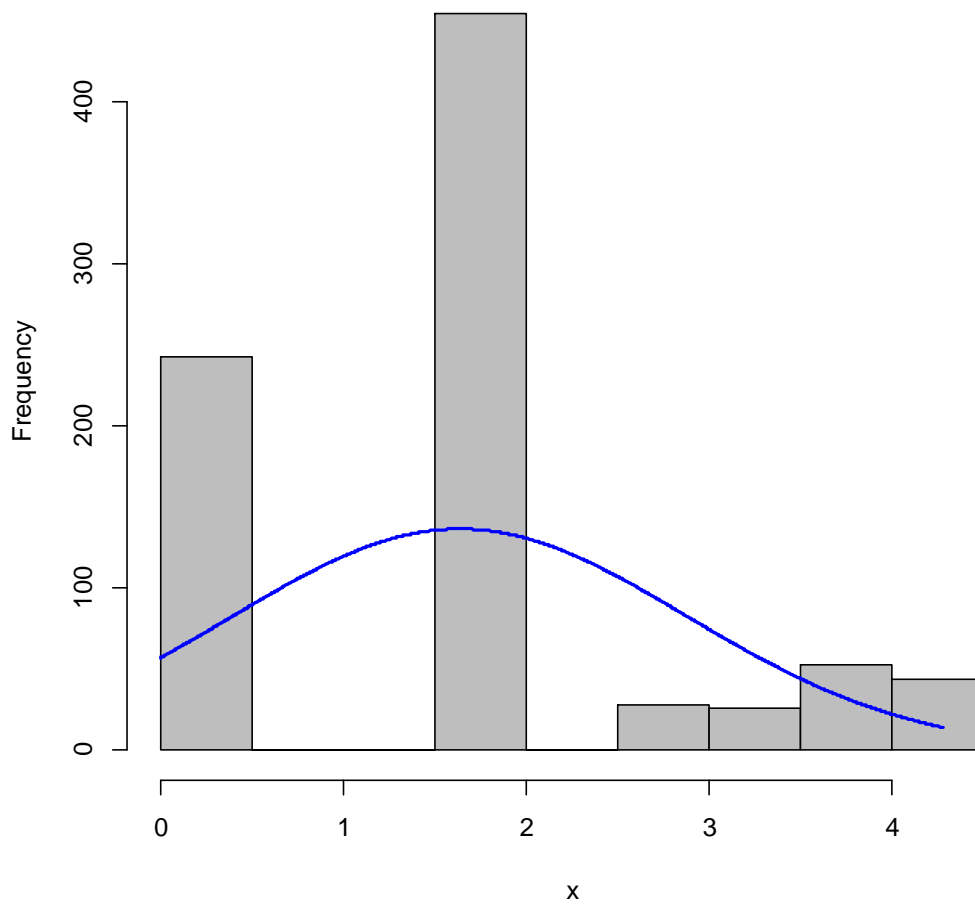


Figure 9: Box-Cox transformation of viral load +1

4.2.2 Temporal trends in the mean and median values of the viral load copies

Figure 8 shows a summary of the mean log viral loads since the initial viral measurement at SWOP–City. The temporal means are clustered around $\log(\text{viral load} + 1) = 1$, that is, a higher number is virally suppressed when back transformed. There is higher variability in MSMW compared to the FSW. However, it appears that the FSWs have had a longer follow up given the time range and as well have outliers.

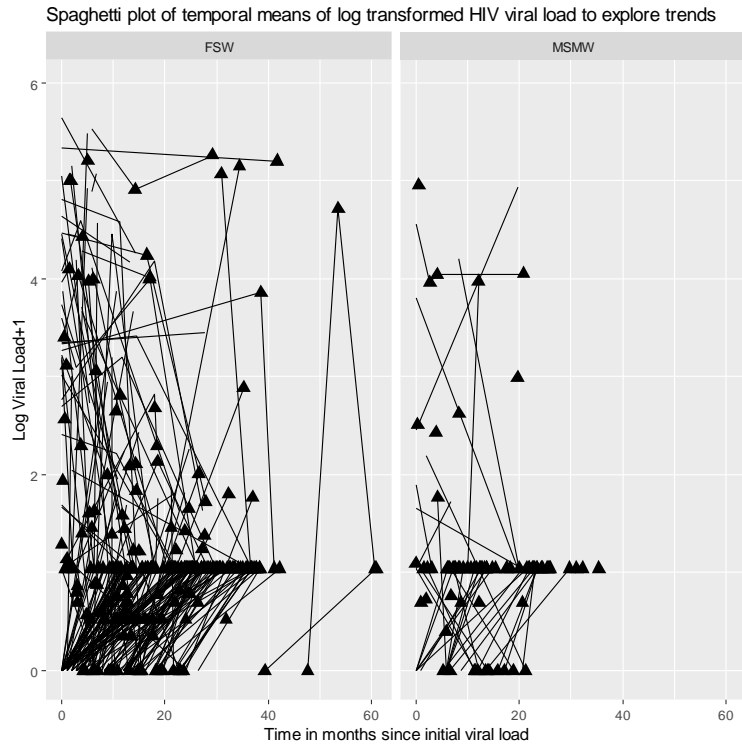


Figure 10: Temporal means of log viral load

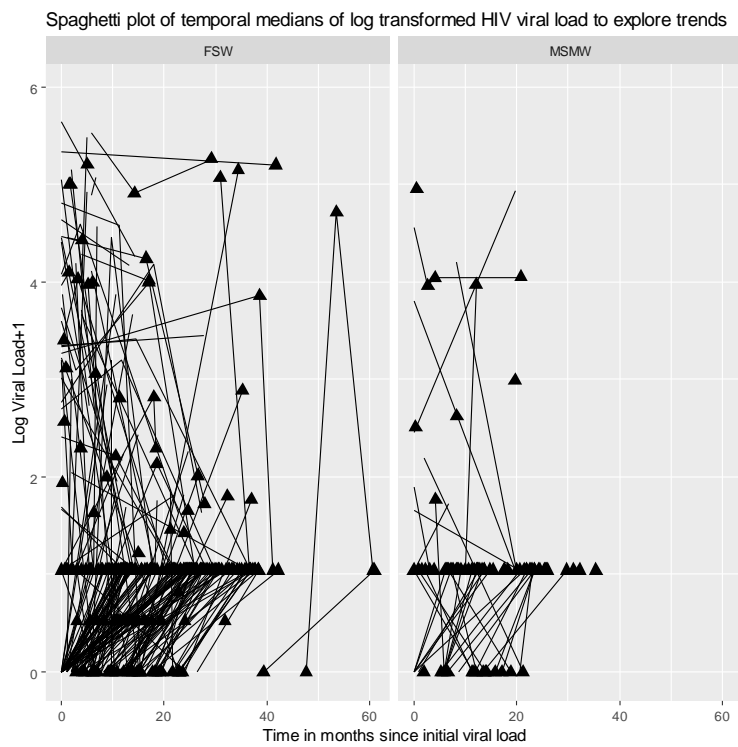


Figure 11: Temporal medians of log viral load

The change depicted by the plot of the medians is quite trivial when compared with the temporal means given the logarithmic transformations.

4.3 Individual profile plots

The observation plans were quite irregular. And since 67.5% of the data wasn't clustered, the profile plot of the whole data was taken, given that a random sample of the data taken consisted of only one observation. It is clear from the plot that each sex worker has his/her own intercept. Some sex workers initiated ART at higher levels while others began with relatively low viral loads.

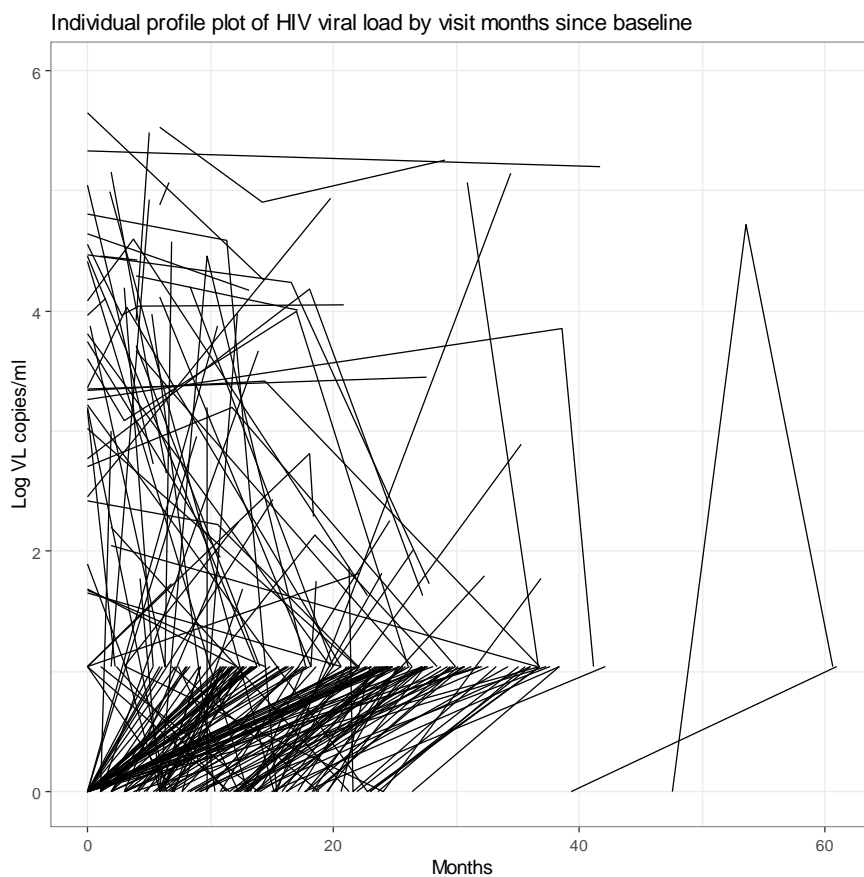


Figure 12: Individual profile plots by months since baseline

4.4 Analyses of sociodemographic and epidemiologic fixed effects

On multivariate analysis of the relationship of sociodemographic factors at baseline with log viral load, none of them was statistically significant. Table 2 below shows the results from the analysis.

Table 2: Multivariate Analyses of sociodemographic and epidemiologic fixed effects

| Fixed effects | Value | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|----------------------------|--------------|-----------------------|---------------------------|----------------|--------------------|
| Intercept | 0.3503355 | 1.2395178 | 40 | 0.282639 | 0.7789 |
| Baseline Age | -0.0007971 | 0.0184604 | 40 | -0.043180 | 0.9658 |
| Baseline viral load | 0.6411213 | 0.0793368 | 40 | 8.081006 | 0.0000 |
| MSMW | -0.4252316 | 0.4079186 | 40 | -1.042442 | 0.3035 |
| Marital Status | | | | | |
| Single | -0.0987403 | 0.6495295 | 40 | -0.152018 | 0.8799 |
| Divorced | -0.1854457 | 0.6900373 | 40 | -0.268747 | 0.7895 |
| Separated | -0.0722443 | 0.6338960 | 40 | -0.113969 | 0.9098 |
| Widowed | -0.1313249 | 0.9026399 | 40 | -0.145490 | 0.8851 |
| First CD4 | -0.0000111 | 0.0004260 | 40 | -0.025995 | 0.9794 |
| BMI | 0.0035646 | 0.0199652 | 40 | 0.178540 | 0.8592 |
| Baseline WHO 2 | 0.2013966 | 0.2292234 | 40 | 0.878604 | 0.3849 |
| Baseline WHO 3 | 0.6832387 | 0.5773912 | 40 | 1.183320 | 0.2437 |
| Baseline WHO 4 | 0.2582573 | 0.6795783 | 40 | 0.380026 | 0.7059 |
| Education | | | | | |
| Did not complete primary 2 | 0.2827462 | 0.2783826 | 40 | 1.015675 | 0.3159 |
| Completed secondary | -0.0470213 | 0.3141607 | 40 | -0.149673 | 0.8818 |
| Did not complete secondary | 0.0430963 | 0.3404536 | 40 | 0.126585 | 0.8999 |
| Completed tertiary | 0.5968396 | 0.5348929 | 40 | 1.115811 | 0.2712 |

| | | | | | |
|-----------------------|------------|-----------|----|-----------|--------|
| Never attended school | -0.0464107 | 0.6827156 | 40 | -0.067980 | 0.9461 |
|-----------------------|------------|-----------|----|-----------|--------|

4.5 The influence of syndemics on HIV viral load outcomes and ART adherence

4.5.1 Univariable Models by log₁₀ copies per mL

All the sex workers whose data were analysed did not show signs of Mycobacterium tuberculosis (TB) infections. Hence HIV-TB syndemic was not included in the models.

The intercept (FSW) of the estimate is statistically significant, however, the estimate of MSMW is not significant as shown in Table 3 below. This shows that log viral load is not predictable by key population type.

Table 3: Solution to key population type fixed effects

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|---------------|----------|----------------|--------------------|---------|----------|
| Intercept | 1.16401 | 0.05171 | 667.89660 | 22.510 | <0.00001 |
| MSMW | -0.08508 | 0.12009 | 703.63305 | -0.708 | 0.479 |

Table 4 below shows that a unit increase in age at last visit is associated with 0.016 decreases in log viral load. The effect is statistically significant.

Table 4: Solution to Age at last visit fixed effects

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|----------------|-----------|----------------|--------------------|---------|----------|
| Intercept | 1.761223 | 0.212368 | 687.951193 | 8.293 | <0.00001 |
| Age last visit | -0.015589 | 0.005273 | 683.677545 | -2.956 | 0.00322 |

Condom use and having experienced condom violence does not predict log viral load as indicated in Table 5 and 6. The p-values show a lack of significant effects. The estimate for condom use is so negligible too to amount to any effect.

Table 5: Solution to baseline condom use fixed effects

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|----------------------|-------------------------|-----------------------|---------------------------|----------------|--------------------|
| Intercept | 1.544x10 ⁻¹⁴ | 1.302 | 112.4 | 0.000 | 1.000 |
| Use condoms | 1.302 | 1.308 | 112.3 | 0.996 | 0.322 |

Table 6: Solution to baseline condom violence fixed effects

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|----------------------|-----------------|-----------------------|---------------------------|----------------|--------------------|
| Intercept | 1.286452 | 0.399882 | 67.726983 | 3.217 | 0.00199 |
| Condom violence | 0.001944 | 0.237458 | 73.751998 | 0.008 | 0.99349 |

There is a no syndemic association of alcohol use during sex with log viral load as shown in Table 7 below. A statistical significance with none use is however present as depicted by the p-value of 0.000159.

Table 7: Solution to baseline use of alcohol during sex fixed effects

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|------------------------|-----------------|-----------------------|---------------------------|----------------|--------------------|
| Intercept | 0.8562 | 0.2184 | 104.5845 | 3.920 | 0.000159 |
| Use alcohol during sex | 0.1775 | 0.1046 | 90.9301 | 1.697 | 0.093070 |

Having an STI or having experience GBV in the past 3 months at baseline were not associated with a unit change in log viral load outcome. As exemplified in Table 8 and 9 below.

Table 8: Solution to any STI fixed effects

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|----------------------|-----------------|-----------------------|---------------------------|----------------|--------------------|
| Intercept | 1.13885 | 0.04743 | 676.67235 | 24.012 | <0.00001 |
| Any STI | 0.28903 | 0.26316 | 669.21145 | 1.098 | 0.272 |

Table 9: Solution to baseline GBV in the last 3 months fixed effects

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|--------------------------|-----------------|-----------------------|---------------------------|----------------|--------------------|
| Intercept | 1.5472 | 0.4360 | 91.4078 | 3.549 | 0.000613 |
| GBV in the last 3 Months | -0.2009 | 0.2348 | 95.1682 | -0.856 | 0.394253 |

On nutritional assessments, it was established that normal weight, being obese, overweight and underweight did not yield statistical significance in the prediction of a unit increase in log viral load as shown in Table 10 below.

Table 10: Solution to Nutritional assessment fixed effects

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|----------------------|-----------------|-----------------------|---------------------------|----------------|--------------------|
| Intercept | 1.06673 | 0.11888 | 764.40180 | 8.973 | <0.00001 |
| Nutrition Assessment | | | | | |
| Normal | 0.16322 | 0.14175 | 731.98252 | 1.151 | 0.2499 |
| Obese | 0.01046 | 0.15084 | 722.62082 | 0.069 | 0.9447 |
| Overweight | 0.02301 | 0.15630 | 728.08250 | 0.147 | 0.8830 |
| Underweight | 0.58325 | 0.30746 | 699.36943 | 1.897 | 0.0582 |

A statistically significant association was found for poor adherence to CTX and ARV medications with p-values of 0.0185 and <0.001 respectively. Table 11 and 12 shows the results.

Table 11: Solution to last CTX adherence fixed effects

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|--------------------|----------|----------------|--------------------|---------|----------|
| Intercept | 1.09848 | 0.11535 | 760.54478 | 9.523 | <0.00001 |
| Last CTX adherence | | | | | |
| Fair | 0.70415 | 0.40170 | 600.72317 | 1.753 | 0.0801 |
| Good | 0.03144 | 0.12636 | 744.01651 | 0.249 | 0.8036 |
| Poor | 1.11367 | 0.47197 | 768.95627 | 2.360 | 0.0185 |

Table 12: Solution to last ARV adherence fixed effects

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|--------------------|----------|----------------|--------------------|---------|----------|
| Intercept | 1.09075 | 0.11326 | 756.58847 | 9.630 | <0.00001 |
| Last ARV adherence | | | | | |
| Fair | 0.83222 | 0.40756 | 638.74475 | 2.042 | 0.041569 |
| Good | 0.02869 | 0.12517 | 738.62302 | 0.229 | 0.818745 |
| N/A | 0.05660 | 0.22860 | 699.24731 | 0.248 | 0.804524 |
| Poor | 1.47060 | 0.44153 | 773.55219 | 3.331 | 0.000907 |

4.5.2 Estimates of baseline and longitudinal ART adherence effect on syndemics

A unit increase in the months since baseline was associated with 0.055110 Box-Cox transformed viral load, however, this effect wasn't statistically significant ($p > 0.005$). A unit increase in the Poor ARV adherence variable was associated with 3.408471 in Box-Cox transformed viral load ($p = 0.0225$). Longitudinal ART adherence was associated with a reduction in Box-Cox transformed viral loads, nonetheless, these effects were not statistically significant as shown in Table 13 below.

Table 13: Multivariable analyses of longitudinal ART adherence effect

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|-------------------------------|-----------------|-----------------------|---------------------------|----------------|--------------------|
| Intercept | 1.452759 | 1.555093 | 111.035577 | 0.934 | 0.3522 |
| Condom use | 0.814601 | 1.419900 | 106.356418 | 0.574 | 0.5674 |
| STI | 2.493874 | 1.178182 | 111.871006 | 2.117 | 0.0365 |
| Sex under alcohol | 0.021026 | 0.157045 | 111.333307 | 0.134 | 0.8937 |
| Use drugs | -0.193695 | 0.278021 | 111.033599 | -0.697 | 0.4875 |
| GBV last 3 months | 0.436554 | 0.337270 | 101.008574 | 1.294 | 0.1985 |
| Sex Alcohol: month | 0.002879 | 0.010557 | 108.838961 | 0.273 | 0.7856 |
| Use Drugs: month | 0.011056 | 0.020340 | 104.494743 | 0.544 | 0.5879 |
| GBV last 3 Months: month | -0.013711 | 0.020098 | 110.360585 | -0.682 | 0.4965 |
| Nutritional assessment | | | | | |
| Normal weight | -1.811510 | 1.076736 | 109.947301 | -1.682 | 0.0953 |
| Obese | -2.685814 | 1.132014 | 110.410357 | -2.373 | 0.0194 |
| Overweight | -1.918072 | 1.064696 | 111.173872 | -1.802 | 0.0743 |
| Underweight | -2.913506 | 1.511005 | 111.271489 | -1.928 | 0.0564 |
| Normal: month | 0.079027 | 0.113806 | 101.194580 | 0.694 | 0.4890 |
| Obese: month | 0.094694 | 0.117310 | 102.313567 | 0.807 | 0.4214 |
| Overweight: month | 0.066506 | 0.115501 | 104.774918 | 0.576 | 0.5660 |
| Month | -0.055110 | 0.106997 | 101.316395 | -0.515 | 0.6076 |
| Good ARV Adherence | 0.774124 | 0.941507 | 109.669341 | 0.822 | 0.4127 |
| N/A ARV Adherence | 2.182887 | 1.189886 | 110.921060 | 1.835 | 0.0693 |

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|---------------------------------|-----------------|-----------------------|---------------------------|----------------|--------------------|
| Poor Last ARV Adherence | 3.408471 | 1.472739 | 111.724423 | 2.314 | 0.0225 |
| Months: Good last ARV Adherence | -0.009743 | 0.062218 | 107.373371 | -0.157 | 0.8759 |
| Months: N/A last ARV Adherence | -0.021690 | 0.076786 | 96.122805 | -0.282 | 0.7782 |

4.5.3 Multivariable Model by log₁₀ copies per mL and Box-Cox transformed viral loads

From a multivariable model, only STI and condom use had a statistically significant effect. The results of multivariable analyses are shown in Table 13 below.

Table 14: Solution to Multivariable Fixed Effects under log₁₀ and Box-Cox transformations

| | Log ₁₀ transformation | | | Box-Cox transformation | | |
|--------------------------|----------------------------------|----------------|--------------------|------------------------|----------------|--------------------|
| Fixed effects | Estimate | t-value | Pr(> t) | Estimate | t-value | Pr(> t) |
| Intercept | -1.523546 | -0.787 | 0.4332 | -1.16365 | -0.611 | 0.5426 |
| MSMW | 0.522025 | 0.665 | 0.5082 | 0.57266 | 0.761 | 0.4490 |
| Age last visit | -0.009883 | -0.598 | 0.5519 | -0.01171 | -0.736 | 0.4640 |
| Use condoms | 3.684080 | 1.988 | 0.0498 | 3.87133 | 2.123 | 0.0364 |
| Condom violence | -0.037274 | -0.130 | 0.8973 | 0.02043 | 0.074 | 0.9413 |
| Use alcohol during sex | 0.099010 | 0.4552 | 0.4552 | 0.09571 | 0.764 | 0.4476 |
| GBV in the last 3 Months | 0.137668 | 0.424 | 0.6731 | 0.06532 | 0.212 | 0.8331 |
| STI | 2.887143 | 2.251 | 0.0268 | 2.00675 | 1.588 | 0.1156 |
| Nutrition Assessment | | | | | | |
| Normal | 1.961354 | 1.481 | 0.1422 | 1.68226 | 1.289 | 0.2005 |
| Obese | 1.399240 | 1.085 | 0.2807 | 1.12909 | 0.887 | 0.3772 |
| Overweight | 1.793974 | 1.347 | 0.1813 | 1.60480 | 1.224 | 0.2241 |
| Underweight | 0.596436 | 0.333 | 0.7396 | -0.22127 | -0.125 | 0.9006 |
| Last CTX Adherence | | | | | | |
| Fair | -2.706450 | -1.544 | 0.1260 | -2.26829 | -1.310 | 0.1936 |
| Good | -3.089226 | -1.904 | 0.0606 | -3.03946 | -1.929 | 0.0572 |
| Last ARV Adherence | | | | | | |
| Good | 0.585044 | 0.581 | 0.5639 | 0.64564 | 0.678 | 0.5007 |
| N/A | 1.749141 | 1.556 | 0.1256 | 1.99664 | 1.880 | 0.0653 |

4.6 Syndemics and viral load outcomes and effect modification by linkage to psychosocial support

STI was included in the model since it was statistically significant in the multivariable model fitted before. The interaction between STI and linkage to psychosocial support had statistically no significant association with change in log viral load. Statistically significant modification effect was reported between poor CTX adherence and linkage to psychosocial with a p-value of 0.00858. The sex workers linked to psychosocial support had 3.76283 increase for a unit increase in linkage to psychosocial support variable. Further, a unit increase in linkage to psychosocial support was associated with 0.10077 reductions in log₁₀ viral load, nonetheless, this was not statistically significant at alpha 0.05 (p=0.454). The underweight sex workers linked to psychosocial support had 0.6904 decreases in viral load.

Similarly, significance in effect modification in terms of interaction between poor adherence to ARV and linkage to psychosocial support with a p-value of 0.0196 was established.

Table 15: Solution to modification effect on log₁₀ viral load of linkage to psychosocial support

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|------------------------------------|-----------|----------------|--------------------|---------|------------|
| Linkage to psychosocial support | | | | | |
| Intercept | 1.16233 | 0.05041 | 668.33412 | 23.056 | <0.0000001 |
| Linked | -0.10077 | 0.13448 | 701.65477 | -0.749 | 0.454 |
| Age Last Visit | | | | | |
| Intercept | 1.759 | 0.2325 | 685.3 | 7.564 | <0.0000001 |
| AgeLastVisit:Linked | 0.0003448 | 0.01468 | 676.4 | 0.023 | 0.98127 |
| STI | | | | | |
| Intercept | 1.1582 | 0.0508 | 664.2407 | 22.798 | <0.0000001 |
| STI:Linked | 0.1513 | 0.5485 | 707.0791 | 0.276 | 0.783 |
| Nutrition Assessment – underweight | | | | | |
| Intercept | 0.6904 | 0.3326 | 0.066 | 2.076 | 0.0383 |
| Underweight:Linked | -0.6904 | 1.022 | 0.08027 | 0.676 | 0.4994 |

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|----------------------|-----------------|-----------------------|---------------------------|------------------|--------------------|
| Last CTX Adherence | | | | | |
| Intercept | 1.11890 | 0.11911 | 751.39542 | 9.394 | <0.0000001 |
| Intercept(Fair) | 0.89861 | 0.43230 | 609.10972 | 2.079 0.03806 | 0.03806 |
| Intercept(Good) | 0.02633 | 0.13168 | 732.11429 | 0.200 | 0.84160 |
| Intercept(Poor) | 0.63612 | 0.50535 | 751.39542 | 1.259 | 0.20851 |
| Fair:Linked | -1.17521 | 1.19390 | 537.87470 - | -0.984 | 0.32539 |
| Good:Linked | 0.22390 | 0.49339 | 754.59293 | 0.454 | 0.65011 |
| Poor:Linked | 3.76283 | 1.42813 | 801.07046 | 2.635 | 0.00858 |
| Last ARV Adherence | | | | | |
| Intercept | 1.11018 | 0.11726 | 744.92270 | 9.468 | <0.00000001 |
| Intercept (fair) | 0.90813 | 0.42990 | 606.99023 | 2.112 | 0.0351 |
| Intercept(Good) | 0.02366 | 0.13095 | 723.95962 | 0.181 | 0.8567 |
| Intercept(N/A) | 0.06615 | 0.24620 | 689.39726 | 0.269 | 0.7883 |
| Intercept(Poor) | 1.10081 | 0.47019 | 756.01469 | 2.341 | 0.0195 |
| Fair:Linked | -0.66395 | 1.39780 | 788.90935 - | 0.475 | 0.6349 |
| Good:Linked | 0.22161 | 0.49225 | 749.85811 | 0.450 | 0.6527 |
| N/A:Linked | 0.12615 | 0.72488 | 701.64283 | 0.174 | 0.861 |
| Poor:Linked | 3.29823 | 1.41071 | 797.45617 | 2.338 | 0.0196 |

4.7 Random effects model for viral load by last adherence

4.7.1 Random intercept model of Box-Cox transformed HIV viral load and adherence with month interaction

Tables 15 and 16 shows the results from linear mixed-effects fits by random intercept and only random slopes. The intercepts are highly statistically significant.

Good ARV adherence is associated with a decrease in viral load by 0.2522 but not statistically significant. A unit increase in the month since baseline was associated with -0.001811 change in transformed viral load. However, an interaction of Good last ARV adherence and month since baseline did not yield a statistically significant result.

A unit increase in the variable Poor Last ARV Adherence was associated with 1.237 positive change in transformed viral load baseline effects. An interaction with month yielded 0.1528 negative change suggestive of programme intervention along the follow-up.

Table 16: Random intercept Linear mixed model fit by maximum likelihood

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|--------------------------------|-------------|----------------|--------------------|---------|----------|
| Intercept | 1.710 | 0.1330 | 839 | 12.856 | 0.0000 |
| Month | -0.001811 | 0.01512 | 830.1 | -0.120 | 0.9047 |
| Fair ARV Adherence | 0.2499 | 0.5211 | 740.3 | 0.480 | 0.6317 |
| Good ARV Adherence | -0.2522 | 0.1488 | 837.1 | -1.695 | 0.0904 |
| N/A ARV Adherence | 0.02378 | 0.3026 | 815.7 | 0.079 | 0.9374 |
| Poor Last ARV Adherence | 1.237 | 0.4907 | 823.7 | 2.522 | 0.0119 |
| Month: Fair last ARV Adherence | 0.03171 | 0.03622 | 643.9 | 0.875 | 0.3816 |
| Month: Good last ARV Adherence | 0.01333 | 0.0157 | 830.9 | 0.849 | 0.3960 |
| Month: N/A last ARV Adherence | -0.00003732 | 0.02425 | 831.8 | -0.002 | 0.9988 |
| Month: Poor last ARV Adherence | -0.1528 | 0.09255 | 838.9 | -1.651 | 0.0990 |

Table 17: Random slope Linear mixed model fit by maximum likelihood

| Fixed effects | Estimate | Standard Error | Degrees of freedom | t-value | Pr(> t) |
|---------------------------------|-----------|----------------|--------------------|---------|----------|
| Intercept | 1.727139 | 0.133695 | 840.0 | 12.919 | 0.0000 |
| Month | -0.005885 | 0.015298 | 840.0 | -0.385 | 0.7006 |
| Fair ARV Adherence | 0.2499 | 0.510392 | 840.0 | 0.528 | 0.5975 |
| Good ARV Adherence | -0.222942 | 0.149175 | 840.0 | -1.494 | 0.1354 |
| N/A ARV Adherence | -0.002194 | 0.300550 | 840.0 | -0.007 | 0.9942 |
| Poor Last ARV Adherence | 1.250819 | 0.489553 | 840.0 | 2.555 | 0.0108 |
| Months: Fair last ARV Adherence | 0.036011 | 0.034965 | 840.0 | 1.030 | 0.3033 |
| Months: Good last ARV Adherence | 0.014355 | 0.015880 | 840.0 | 0.904 | 0.3663 |
| Months: N/A last ARV Adherence | 0.002876 | 0.024532 | 840.0 | 0.117 | 0.9067 |
| Months: Poor last ARV Adherence | -0.145579 | 0.093040 | 840.0 | -1.565 | 0.1180 |

Residual variance associated with random slopes was 1.501 with SE of 1.225.

Table 18: Likelihood ratio test

| | Model | DF | AIC | BIC | LogLik | Deviance | Chisq. | Chi | Pr(>Chisq) |
|-----------------|-------|----|--------|--------|---------|----------|--------|-----|------------|
| LMM (slope) | 1 | 4 | 2780.0 | 2799.0 | -1386.0 | 2772.0 | | | |
| LMM (intercept) | 2 | 4 | 2770.5 | 2789.5 | -1381.2 | 2762.5 | 9.5294 | 0 | <0.0000001 |

The fixed parts of the two models seem to be identically close to that of each other, however, the standard errors associated with intercept and levels of adherence are different. On doing a likelihood ratio test (LRT) by analysis of variance (anova(modelML_slp,modelML_int) to compare the models, the results showed that the two models were significantly different (p-value<0.0001) as shown in Table 17 below. Given the small standard errors obtained from linear mixed-effects model fit by maximum likelihood, it was an easy pick of LMM. The broad analysis interest was to understand the growth of the cohort effect.

All the models were fitted by maximum likelihood (ML) estimation. A chi-square test for comparing the log likelihoods of the two models was done. The likelihood ratio test/ χ^2 test is generally conservative for significance testing of a random effect since the null value ($\sigma^2=0$) lies at the feasible space boundary. However, looking at the scenario, strong evidence existed for rejecting the null hypothesis that the two models are the same. The model with random effects of sex worker from the models is $1386.0 - 1381.2=4.8$ log-likelihood units better. Chi-square test was then done, and since twice the log-likelihood of the value is χ^2 distributed, the direct p-value computation gave a value of -2.710908 from the R code: `pchisq(2*4.8, df=1, lower.tail=FALSE, log.p=TRUE)/log(10)`. This is a p-value of approximately $10^{2.710908}$ (or 0.001945772) which is overwhelmingly statistically significant.

The variance of the random intercept effect and slope was 0.2043 and 1.501 respectively and these capture individual sex worker variation not explained by fixed effects. The residual variance is 1.2877 which relates to within individual sex worker variation in which case where no fixed effects differ within individual sex worker. The residual variance is the conditional variance of the random effects model shown in Table 18 below.

Table 19: Variance of the random effects

| Random effects | Groups | Name | Variance | Std. Dev |
|------------------|----------|-----------|----------|----------|
| Random intercept | ID | Intercept | 0.2043 | 0.452 |
| | Residual | | 1.2877 | 1.135 |
| Random slopes | ID | Month | 0.000 | 0.000 |
| | Residual | | 1.501 | 1.225 |

4.7.2 Within cluster resampling – Multiple outputation

Sixty-seven point five percent of the data wasn't clustered, that is, they had single rows of data per sex worker. The data was subsetted and unique rows of data sampled with replacement in order to apply independent data statistical method (linear regression) in the analysis (Appendix G). The estimates appeared closely similar to those fit by linear mixed effects method. The intercepts fitted by both methods are both highly statistically significant at 5% level of significance. A unit increase in variable poor adherence was associated with 1.043836 positive change (increase) in Box-Cox transformed viral load at $p < 0.05$. However, interaction with a

number of months since baseline, it was associated with 0.139589 negative change (reduction) in Box-Cox transformed viral load.

Table 20: Within cluster resampling – Multiple outputation

| | Sample 1 | | Sample 2 | | Sample 3 | | Sample 4 | | Average |
|-------------------|-----------|----------|-----------|----------|-----------|----------|-----------|----------|----------|
| Fixed effects | Estimate | Pr(> t) | Estimate | Pr(> t) | Estimate | Pr(> t) | Estimate | Pr(> t) | Estimate |
| Intercept | 1.732383 | 0.0000 | 1.70143 | 0.0000 | 1.761369 | 0.0000 | 1.77266 | 0.0000 | 1.741961 |
| Residual variance | 1.317098 | | 1.330892 | | 1.30116 | | 1.344535 | | 1.323421 |
| Slope variance | 0.315324 | | 0.2939389 | | 0.2922447 | | 0.2953493 | | 0.299214 |
| Adherence | | | | | | | | | |
| Fair | -0.616729 | 0.2725 | -0.02879 | 0.9577 | -0.071289 | 0.8951 | -0.31152 | 0.5667 | -0.25708 |
| Good | -0.129688 | 0.3847 | -0.21241 | 0.1578 | -0.162801 | 0.2773 | -0.20210 | 0.1855 | -0.17675 |
| N/A | -0.054627 | 0.8670 | -0.05812 | 0.8534 | 0.288852 | 0.3509 | 0.36628 | 0.2436 | 0.135596 |
| Poor | 1.043836 | 0.0309 | 1.18093 | 0.0121 | 1.014850 | 0.0349 | 1.00355 | 0.0401 | 1.060792 |
| Month | | | -0.01204 | 0.5161 | -0.012758 | 0.4513 | -0.01350 | 0.4475 | -0.00957 |
| Adherence* month | | | | | | | | | |
| Fair | 0.106749 | 0.0220 | 0.10768 | 0.0214 | 0.109995 | 0.0166 | 0.03276 | 0.4174 | 0.089296 |
| Good | 0.003623 | 0.8443 | 0.01188 | 0.5379 | 0.010554 | 0.5514 | 0.01336 | 0.4711 | 0.009854 |
| N/A | 0.018586 | 0.4888 | 0.3583 | 0.3583 | -0.001035 | 0.9686 | -0.01009 | 0.7143 | 0.09144 |
| Poor | -0.139589 | 0.1126 | 0.0654 | 0.0654 | -0.133425 | 0.1265 | -0.13269 | 0.1354 | -0.08508 |

4.7.3 Both random intercept and random slope effects

The study's last objective was whether a random intercept would be sufficient or the longitudinal model needs to have a random slope as well. It appeared that the random intercept was adequate – or a random slope as seen in Section 4.8.1.

Further, it was found that there was probable un-identifiability of random effects parameters and residual variance if both random effects were fitted. This is because the number of random effects was the total number of observations, this is reminiscent of the irregular observation plans in the cohort. On inclusion of both random intercepts and random slopes, the number of random coefficients resulting was equal to a number of observations and as such the model could not be fitted. Implemented in R as: `fit1<-`

`(lmerTest::lmer(vlbox~month*lastARVAdherence+(month|id),data=i.swop))`. Graphical summaries of the data such as the individual profile plots and temporal trends detailed in Chapter 4, Section 4.3.2 and 4.4 also portrayed adequacy of just a random intercept model.

CHAPTER FIVE

5. DISCUSSIONS AND CONCLUSIONS

5 Introduction

This Chapter presents a comprehensive discussion of the results of the linear mixed-effects method and corresponding findings provided in Chapter 4.

5.2 Discussion

The objective of this study was to examine and model the association of syndemics with HIV viral load outcomes as well as ART adherence among sex workers living with HIV in the SWOP–City cohort using linear mixed-effects. In the background of global efforts of ART scale-up, this study's results draw substantial issues of sex workers' HIV treatment.

The log viral load temporal means clustered around $\log(\text{viral load} + 1) = 1$, this suggested high numbers of sex workers were virally suppressed. This is an important aspect of TasP. Graphical summaries showed the adequacy of random intercept model.

Investigating the longitudinal trajectories of HIV viral load following ART initiation is important in identifying HIV viral load change points. There were high numbers of left-censored observations (15%) at LOD, 71% having had 10 HIV viral load copies and overall, 92% with viral load copies below 350 copies/ml.

A number of syndemic factors were looked at in the study. Condom use and having experienced condom violence, alcohol use during sex, STI, experiencing GBV did not have a statistically significant effect in the prediction of log viral load. Poor CTX and poor ARV adherence were significantly associated with poor low log viral loads all with $p < 0.05$. On the fitting multivariable model, STI and condom use had a statistically significant effect ($p < 0.05$). Linkage to psychosocial support played a key role in modifying the relationship between log HIV viral and poor adherence (for CTX $p = 0.00858$ and ARV $p = 0.0196$). All other effect modifications by linkage to support was not statistically significant.

Characteristically, medium and high linkage to psychosocial support tend to be associated with higher levels of HIV viral load suppression to LOD and lower HIV viral load means (Friedman et al., 2017) and medians.

Consistent with earlier studies (Bain-Brickley, Butler, Kennedy, & Rutherford, 2011; Duff et al., 2016; Kranzer, Lawn, Johnson, Bekker, & Wood, 2013; Ortego, Huedo-Medina, Vejo, & Llorca, 2011; Ramadhani et al., 2007; Rupérez et al., 2015), sex workers with good adherence had higher inclination to viral suppression. Syndemics, mostly psychosocial ones, are additively associated with poor ART adherence (Blashill et al., 2015) but when these are tackled, this effect is lessened.

It is vital to assess hypotheses under a suitable random effect structure. For instance, often it is inadequate to just include random intercepts in the LMM. Random slopes may perhaps be essential. However, looking at the output presented in Chapter four, the results demonstrate that LMM with a random intercept was fairly adequate. This implies that every sex worker studied started at a specific own intercept of viral load at ART initiation and then viral load diminishes with ongoing care – especially for those virally suppressed.

This study considered fitting GLS since there were many sex many sex workers on ART who contributed single rows of data (one observation). There was a statistically significant difference between GLS model and LMM as demonstrated by likelihood ratio test and Chi-square test. Given the small standard errors obtained from LMM fit by maximum likelihood, the LMM characteristically important in understanding the evolutions of viral loads. The main interest in fitting the intercept was to understand the growth of the cohort effect the slopes.

In many observational studies, patients are usually followed over a period of time and are evaluated at regular intervals/times in a schedule common to all of them. Practically, this is infeasible in situations like that of sex workers following dynamic observation plans. This dynamism results in irregular plans of observations. Trying to fit a random intercept model in this study led to many random effects equivalent to the number of observations. However, the random intercept model was well enough for the analysis of the cohort effect and individual trajectories of log viral load.

5.3 Limitations

Since this is an observational cohort study, recall and social desirability biases may have arisen given the self-reported measurement of adherence. Limited time for the sex workers recently initiating ART may have been a challenge, however, the LMM method of analysis may take care of this. The initial interest of the study was to look ARV and CTX adherence as time-dependent

covariates, however, data for last adherence from the cohort was available and therefore used. Therefore, the longitudinal effect of adherence was not evaluated.

5.4 Recommendations

Given the statistical significance of the effect modification, the SWOP–City programme may deem it necessary to increase linkage to psychosocial support. This model has shown that it is a useful component in the reduction of viral load copies in the cohort. An interaction of linkage and poor adherence gave a statistically significant result. Adherence to ART is important has public health implications. Considering the structural and social contexts that tend to shape the sex workers ART experiences and affect outcomes of treatment.

5.5 Future research

Typically, HIV patients follow semiannual person-visits after ART initiation. However, irregular observation plans present challenges and opportunities for further research. Such irregular plans may have some potentially informative follow-ups or visiting schemes within the cohort, as such; future research may need to consider outcome dependent visits/follow-ups and predictors of visit times. Similarly, it appears that informative and non-informative cluster sizes could possibly be modeled to understand why some sex workers contribute few repeated observations may be useful too for the future studies to look at arbitrary patterns of data missingness.

Future studies may need to consider time-dependent/time-varying syndemics. Further research on HIV care among sex workers is necessary, alongside structural and community or outreach-led interventions to sustain the sex workers access to and in-care retention. Additional statistical methods, specifically mixed-effect modeling of time-to-retention in care may prove useful in understanding the HIV care continuum.

5.6 Conclusions

This study found low to moderate effects of syndemics at the sex workers level. Syndemics have an influence on longitudinal HIV viral load outcomes, however, in the presence of medication (ART and CTX) adherence and linkage to psychosocial support, the effect appears to diminish. The relationship between syndemics and HIV viral load outcomes become insignificant as linkage to psychosocial support is sustained. Random intercept model was sufficient in the analysis of viral load longitudinal subject-specific trajectories and may prove useful in the clinical assessment of an individual sex worker with elevated viral load and at the programme

level to identify potential hotspot of viral transmission. The results of this study suggest that for this sample of sex workers with a history of syndemic problems either medication adherence or linkage to psychosocial support was associated with comparatively reduced levels of log (viral load +1).

REFERENCES

- Abraham, A. G., Lau, B., Deeks, S., Moore, R. D., Zhang, J., Eron, J., ... Gange, S. J. (2011). Missing Data on the Estimation of the Prevalence of Accumulated Human Immunodeficiency Virus Drug Resistance in Patients Treated With Antiretroviral Drugs in North America. *American Journal of Epidemiology*, *174*(6), 727–735. <https://doi.org/10.1093/aje/kwr141>
- Baer, H. A., Singer, M., & Susser, I. (2013). *Medical anthropology and the world system : critical perspectives* (3rd ed.). Santa Barbara, CA: Praeger.
- Bain-Brickley, D., Butler, L. M., Kennedy, G. E., & Rutherford, G. W. (2011). Interventions to improve adherence to antiretroviral therapy in children with HIV infection. In *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/14651858.CD009513>
- Baral, S., Beyrer, C., Muessig, K., Poteat, T., Wirtz, A. L., Decker, M. R., ... Kerrigan, D. (2012). Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, *12*(7), 538–549. [https://doi.org/10.1016/S1473-3099\(12\)70066-X](https://doi.org/10.1016/S1473-3099(12)70066-X)
- Batchelder, A. W. A., Gonzalez, J. S. J., Palma, A., Schoenbaum, E., & Lounsbury, D. W. (2015). A social ecological model of syndemic risk affecting women with and at-risk for HIV in impoverished urban communities. *American Journal of Community Psychology*, *56*(3–4), 229–240. <https://doi.org/10.1007/s10464-015-9750-y>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4, *67*(1). <https://doi.org/10.18637/jss.v067.i01>
- Bhaskaran, K., & Smeeth, L. (2014). What is the difference between missing completely at random and missing at random? *International Journal of Epidemiology*, *43*(4), 1336–1339. <https://doi.org/10.1093/ije/dyu080>
- Blank, M. B., & Eisenberg, M. M. (2014). The Syndemic of the Triply Diagnosed: HIV Positives with Mental Illness and Substance Abuse or Dependence. *Clinical Research in HIV/ AIDS*, *1*(1), 1006. Retrieved from <https://www.jscimedcentral.com/HIV->

AIDS/download.php?file=HIV-AIDS-1-1006.pdf

- Blashill, A. J., Bedoya, C. A., Mayer, K. H., O’Cleirigh, C., Pinkston, M. M., Remmert, J. E., ... Safren, S. A. (2015). Psychosocial Syndemics are Additively Associated with Worse ART Adherence in HIV-Infected Individuals. *AIDS and Behavior*, *19*(6), 981–986.
<https://doi.org/10.1007/s10461-014-0925-6>
- Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *211*–252. Retrieved from
<https://www.semanticscholar.org/paper/An-Analysis-of-Transformations-Box-Cox/6e820cf11712b9041bb625634612a535476f0960>
- Byg, B., Bazzi, A. R., Funk, D., James, B., & Potter, J. (2016). The Utility of a Syndemic Framework in Understanding Chronic Disease Management Among HIV-Infected and Type 2 Diabetic Men Who Have Sex with Men. *Journal of Community Health*, *41*(6), 1204–1211. <https://doi.org/10.1007/s10900-016-0202-x>
- Cho, H. (2016). The analysis of multivariate longitudinal data using multivariate marginal models. *Journal of Multivariate Analysis*, *143*, 481–491.
<https://doi.org/10.1016/j.jmva.2015.10.012>
- Cohen, M. S., Chen, Y. Q., McCauley, M., Gamble, T., Hosseinipour, M. C., Kumarasamy, N., ... Fleming, T. R. (2011). Prevention of HIV-1 Infection with Early Antiretroviral Therapy. *New England Journal of Medicine*, *365*(6), 493–505.
<https://doi.org/10.1056/NEJMoa1105243>
- Cohen, P., West, S. G., Aiken, L. S., West, S. G., & Aiken, L. S. (1984). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge.
<https://doi.org/10.4324/9781410606266>
- Costa, A. G. F. C., Colosimo, E. A., Vaz, A. B. M., Silva, J. L. P., & Amorim, L. D. (2017). Marginal models for the association structure of hierarchical binary responses. *Journal of Applied Statistics*, *44*(10), 1827–1838. <https://doi.org/10.1080/02664763.2016.1238042>
- Diggle, P., & Diggle, P. (2002). *Analysis of longitudinal data*. Oxford University Press. Retrieved from <https://global.oup.com/academic/product/analysis-of-longitudinal-data->

9780198524847?cc=us&lang=en&

- Djomand, G., Quaye, S., & Sullivan, P. S. (2014). HIV epidemic among key populations in west Africa. *Current Opinion in HIV and AIDS*, 9(5), 506–513.
<https://doi.org/10.1097/COH.0000000000000090>
- Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, 2(1), 222. <https://doi.org/10.1186/2193-1801-2-222>
- Douglas-Vail, M. (2016). Syndemics theory and its applications to HIV/AIDS public health interventions. *International Journal of Medical Sociology and Anthropology ISSN*, 4(1), 81–90.
- Duff, P., Goldenberg, S., Deering, K., Montaner, J., Nguyen, P., Dobrer, S., & Shannon, K. (2016). Barriers to Viral Suppression Among Female Sex Workers: Role of Structural and Intimate Partner Dynamics. *Journal of Acquired Immune Deficiency Syndromes (1999)*, 73(1), 83–90. <https://doi.org/10.1097/QAI.0000000000001022>
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis*. Wiley. Retrieved from <https://www.wiley.com/en-us/Applied+Longitudinal+Analysis%2C+2nd+Edition-p-9780470380277>
- Follmann, D., & Fay, M. (2010). Exact Inference for Complex Clustered Data Using Within-Cluster Resampling. *Journal of Biopharmaceutical Statistics*, 20(4), 850–869.
<https://doi.org/10.1080/10543401003618884>
- Follmann, D., Proschan, M., & Leifer, E. (2003). Multiple Outputation: Inference for Complex Clustered Data by Averaging Analyses from Independent Data. *Biometrics*, 59(2), 420–429.
<https://doi.org/10.1111/1541-0420.00049>
- Friedman, M. R., Coulter, R. W. S., Silvestre, A. J., Stall, R., Teplin, L., Shoptaw, S., ... Plankey, M. W. (2017). Someone to count on: social support as an effect modifier of viral load suppression in a prospective cohort study. *AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV*, 29(4), 469–480.
<https://doi.org/10.1080/09540121.2016.1211614>
- Friedman, M. R., Stall, R., Silvestre, A. J., Wei, C., Shoptaw, S., Herrick, A., ... Silvestre, A. J.

- (2015). Effects of syndemics on HIV viral load and medication adherence in the multicentre AIDS cohort study. *Aids*, 29(9), 1087–1096.
<https://doi.org/10.1097/QAD.0000000000000657>
- Gagnon, M. (2018). “There is a chain of connections”: using syndemics theory to understand HIV treatment side effects. *AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV*, 30(7), 910–913. <https://doi.org/10.1080/09540121.2018.1450478>
- Galbraith, S., Daniel, J. A., & Vissel, B. (2010). A study of clustered data and approaches to its analysis. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 30(32), 10601–10608. <https://doi.org/10.1523/JNEUROSCI.0362-10.2010>
- Githuka, G., Hladik, W., Mwalili, S., Cherutich, P., Muthui, M., Gitonga, J., ... KAIS Study Group. (2014). Populations at increased risk for HIV infection in Kenya: results from a national population-based household survey, 2012. *Journal of Acquired Immune Deficiency Syndromes (1999)*, 66 Suppl 1(Suppl 1), S46-56.
<https://doi.org/10.1097/QAI.0000000000000137>
- Global Network of Sex Work Projects. (2015). Stigma and Discrimination Experienced by Sex Workers Living with HIV, 14. Retrieved from www.hivgaps.org for more information.
- Günthard, H. F., Aberg, J. A., Eron, J. J., Hoy, J. F., Telenti, A., Benson, C. A., ... International Antiviral Society-USA Panel. (2014). Antiretroviral Treatment of Adult HIV Infection. *JAMA*, 312(4), 410. <https://doi.org/10.1001/jama.2014.8722>
- Harkness, A., Bainter, S. A., O’Cleirigh, C., Mendez, N. A., Mayer, K. H., & Safren, S. A. (2018). Longitudinal Effects of Syndemics on ART Non-adherence Among Sexual Minority Men. *AIDS and Behavior*, 22(8), 2564–2574. <https://doi.org/10.1007/s10461-018-2180-8>
- Hatcher, A. M., Smout, E. M., Turan, J. M., Christofides, N., & Stoeckl, H. (2015). Intimate partner violence and engagement in HIV care and treatment among women Intimate partner violence and engagement in HIV care and treatment among women : a systematic review and meta-analysis. *Aids*, 2015, (SEPTEMBER).
<https://doi.org/10.1097/QAD.0000000000000842>

- Hernán, M. A., McAdams, M., McGrath, N., Lanoy, E., & Costagliola, D. (2009). Observation plans in longitudinal studies with time-varying treatments. *Statistical Methods in Medical Research*, 18(1), 27–52. <https://doi.org/10.1177/0962280208092345>
- IBBS. (2012). *Implementing Integrated Bio- Behavioral Surveys among Key Populations at Higher Risk of HIV Exposure with an Emphasis in Respondent Driven Sampling*. Retrieved from <https://www.iom.int/jahia/webdav/shared/shared/mainsite/activities/health/hiv-population/IBBS-Study-Final.pdf>
- Institute of Medicine (US) Committee on Health and Behavior: Research, Practice, and P. (2001). *Health and Behavior: The Interplay of Biological, Behavioral, and Societal Influences*. National Academies Press (US). Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK43749/>
- IOM. (2011). *Integrated Biological and Behavioural Surveillance Survey among Migrant Female Sex Workers in Nairobi, Kenya*. International Organization for Migration. Retrieved from <https://www.iom.int/jahia/webdav/shared/shared/mainsite/activities/health/hiv-population/IBBS-Study-Final.pdf>
- Jana, S., Basu, I., Rotheram-Borus, M. J., & Newman, P. A. (2004). The Sonagachi Project: A Sustainable Community Intervention Program. *AIDS Education and Prevention*, 16(5), 405–414. <https://doi.org/10.1521/aeap.16.5.405.48734>
- Judd, C. M., & Kenny, D. A. (1981a). *Process analysis: Estimating Mediation in Treatment Evaluations*. *Evaluation Review* (Vol. 5). Cambridge: Cambridge University Press. <https://doi.org/10.1177/0193841X8100500502>
- Judd, C. M., & Kenny, D. A. (1981b). Process Analysis. *Evaluation Review*, 5(5), 602–619. <https://doi.org/10.1177/0193841X8100500502>
- Karahalios, A., Baglietto, L., Lee, K. J., English, D. R., Carlin, J. B., & Simpson, J. A. (2013). The impact of missing data on analyses of a time-dependent exposure in a longitudinal cohort: A simulation study. *Emerging Themes in Epidemiology*, 10(1), 6. <https://doi.org/10.1186/1742-7622-10-6>

- KELIN. (2016). *Punitive Laws Affecting Sex*. Nairobi, Kenya.
- Kerr, L. R. F. S., Mota, R. S., Kendall, C., Pinho, A. D. A., Mello, M. B., Guimarães, M. D. C., ... Rutherford, G. (2013). HIV among MSM in a large middle-income country. *Aids*, 27(3), 427–435. <https://doi.org/10.1097/QAD.0b013e32835ad504>
- Kerrigan, D., Kennedy, C. E., Morgan-Thomas, R., Reza-Paul, S., Mwangi, P., Win, K. T., ... Butler, J. (2015). A community empowerment approach to the HIV response among sex workers: effectiveness, challenges, and considerations for implementation and scale-up. *The Lancet*, 385(9963), 172–185. [https://doi.org/10.1016/S0140-6736\(14\)60973-9](https://doi.org/10.1016/S0140-6736(14)60973-9)
- Kim, J., Lee, E., Park, B.-J., Bang, J. H., & Lee, J. Y. (2018). Adherence to antiretroviral therapy and factors affecting low medication adherence among incident HIV-infected individuals during 2009–2016: A nationwide study. *Scientific Reports*, 8(1), 3133. <https://doi.org/10.1038/s41598-018-21081-x>
- Kimani, J., McKinnon, L. R., Wachihi, C., Kusimba, J., Gakii, G., Birir, S., ... Gelmon, L. (2013). Enumeration of Sex Workers in the Central Business District of Nairobi, Kenya. *PLoS ONE*, 8(1), e54354. <https://doi.org/10.1371/journal.pone.0054354>
- Kranzer, K., Lawn, S. D., Johnson, L. F., Bekker, L.-G., & Wood, R. (2013). Community viral load and CD4 count distribution among people living with HIV in a South African Township: implications for treatment as prevention. *Journal of Acquired Immune Deficiency Syndromes (1999)*, 63(4), 498–505. <https://doi.org/10.1097/QAI.0b013e318293ae48>
- Kuhn, M., & Roeder, I. (2018). Linear Mixed Models with Censored Observations.
- Kwan, C., & Ernst, J. D. (2011). HIV and tuberculosis: A deadly human syndemic. *Clinical Microbiology Reviews*, 24(2), 351–376. <https://doi.org/10.1128/CMR.00042-10>
- Laird, N. M., & Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4), 963–974. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7168798>
- Lazo, M., Gange, S. J., Wilson, T. E., Anastos, K., Ostrow, D. G., Witt, M. D., & Jacobson, L. P. (2007). Patterns and Predictors of Changes in Adherence to Highly Active Antiretroviral Therapy: Longitudinal Study of Men and Women. *Clinical Infectious Diseases*, 45(10),

1377–1385. <https://doi.org/10.1086/522762>

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22. <https://doi.org/10.1093/biomet/73.1.13>

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). Retrieved from <https://www.wiley.com/en-us/Statistical+Analysis+with+Missing+Data%2C+2nd+Edition-p-9780471183860>

Liu, G., & Liang, K. Y. (1997). Sample size calculations for studies with correlated observations. *Biometrics*, *53*(3), 937–947. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9290224>

Long, J. D., Loeber, R., & Farrington, D. P. (2009). Marginal and Random Intercepts Models for Longitudinal Binary Data With Examples From Criminology. *Multivariate Behavioral Research*, *44*(1), 28–58. <https://doi.org/10.1080/00273170802620071>

Lorway, R., Lazarus, L., Chevrier, C., Khan, S., Musyoki, H. K., Mathenge, J., ... Blanchard, J. (2018). Ecologies of security: On the everyday security tactics of female sex workers in Nairobi, Kenya. *Global Public Health*, 1–14. <https://doi.org/10.1080/17441692.2018.1442487>

MacKinnon, D. (2008). *Introduction to Statistical Mediation Analysis*. Routledge. <https://doi.org/10.4324/9780203809556>

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A Simulation Study of Mediated Effect Measures. *Multivariate Behavioral Research*, *30*(1), 41–62. https://doi.org/10.1207/s15327906mbr3001_3

Martin, L. R., Williams, S. L., Haskard, K. B., & Dimatteo, M. R. (2005). The challenge of patient adherence. *Therapeutics and Clinical Risk Management*, *1*(3), 189–199. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18360559>

Matta, T. H., Flournoy, J. C., & Byrne, M. L. (2017, October 28). Making an unknown unknown a known unknown: Missing data in longitudinal neuroimaging studies. *Developmental Cognitive Neuroscience*. <https://doi.org/10.1016/j.dcn.2017.10.001>

McCulloch, C. E. (2005). Repeated Measures ANOVA, R.I.P.? *CHANCE*, *18*(3), 29–33.

<https://doi.org/10.1080/09332480.2005.10722732>

- McKinnon, L. R., Gakii, G., Juno, J. A., Izulla, P., Munyao, J., Ileri, N., ... Kimani, J. (2014). High HIV risk in a cohort of male sex workers from Nairobi, Kenya. *Sexually Transmitted Infections*, *90*(3), 237–242. <https://doi.org/10.1136/sextrans-2013-051310>
- McKinnon, L. R., Kaul, R., Kimani, J., Nagelkerke, N. J., Wachihi, C., Fowke, K. R., ... Plummer, F. A. (2012). HIV-specific CD8 + T-cell proliferation is prospectively associated with delayed disease progression. *Immunology and Cell Biology*, *90*(3), 346–351. <https://doi.org/10.1038/icb.2011.44>
- Meade, C. S., Hansen, N. B., Kochman, A., & Sikkema, K. J. (2009). Utilization of Medical Treatments and Adherence to Antiretroviral Therapy among HIV-Positive Adults with Histories of Childhood Sexual Abuse. *AIDS Patient Care and STDs*, *23*(4), 259–266. <https://doi.org/10.1089/apc.2008.0210>
- Meyer, J. P., Springer, S. A., & Altice, F. L. (2011). Substance Abuse, Violence, and HIV in Women: A Literature Review of the Syndemic. *Journal of Women's Health*, *20*(7), 991–1006. <https://doi.org/10.1089/jwh.2010.2328>
- MoH/NACC. (2016). *Kenya AIDS Response progress Report, 2016* (Vol. 23). <https://doi.org/10.1258/ijisa.2012.011361.HIV>
- Molenberghs, G., & Verbeke, G. (2001). A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modeling*, *1*(4), 235–269. <https://doi.org/10.1177/1471082X0100100402>
- Molenberghs, G., Verbeke, G., & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4419-0300-6>
- Moore, L., Chersich, M. F., Steen, R., Reza-Paul, S., Dhana, A., Vuylsteke, B., ... Scorgie, F. (2014). Community empowerment and involvement of female sex workers in targeted sexual and reproductive health interventions in Africa: a systematic review. *Globalization and Health*, *10*(1), 47. <https://doi.org/10.1186/1744-8603-10-47>
- Muraguri, N., Tun, W., Okal, J., Broz, D., Raymond, H. F., Kellogg, T., ... Geibel, S. (2015).

- HIV and STI Prevalence and Risk Factors Among Male Sex Workers and Other Men Who Have Sex With Men in Nairobi, Kenya. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 68(1), 91–96. <https://doi.org/10.1097/QAI.0000000000000368>
- Musyoki, H., Kellogg, T. A., Geibel, S., Muraguri, N., Okal, J., Tun, W., ... Kim, A. A. (2015). Prevalence of HIV, sexually transmitted infections, and risk behaviours among female sex workers in Nairobi, Kenya: results of a respondent driven sampling study. *AIDS and Behavior*, 19 Suppl 1(Suppl 1), S46-58. <https://doi.org/10.1007/s10461-014-0919-4>
- Naar-King, S., Templin, T., Wright, K., Frey, M., Parsons, J. T. J., & Lam, P. (2006). Psychosocial factors and medication adherence in HIV-positive youth. *AIDS Patient Care and STDs*, 20(1), 44–47. <https://doi.org/10.1089/apc.2006.20.44>
- NACC. (2012). Kenya National AIDS Strategic Plan 2009/10-2012/13. Retrieved from http://guidelines.health.go.ke:8000/media/Kenya_National_Aids_Strategic_Plan_2009-2010.2012-2013-.pdf
- NACC. (2014). Kenya AIDS Response Progress Report Progress towards Zero. *NASCOP, Ministry of Health, Government of Kenya.*, (March), 37. https://doi.org/http://www.unaids.org/sites/default/files/country/documents/KEN_narrative_rep
- Nieuwkerk, P. T., & Oort, F. J. (2005). Self-reported adherence to antiretroviral therapy for HIV-1 infection and virologic treatment response: a meta-analysis. *Journal of Acquired Immune Deficiency Syndromes (1999)*, 38(4), 445–448. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15764962>
- Nilsson Schönnesson, L., Diamond, P. M., Ross, M. W., Williams, M., & Bratt, G. (2006). Baseline predictors of three types of antiretroviral therapy (ART) adherence: A 2-year follow-up. *AIDS Care*, 18(3), 246–253. <https://doi.org/10.1080/09540120500456631>
- Noppert, G. A., Kubale, J. T., & Wilson, M. L. (2017). Analyses of infectious disease patterns and drivers largely lack insights from social epidemiology: contemporary patterns and future opportunities. *J Epidemiol Community Health*, 4(11), 350–355. [https://doi.org/10.1016/S2214-109X\(16\)30265-0](https://doi.org/10.1016/S2214-109X(16)30265-0). Cost-effectiveness

- Odek, W. O., Githuka, G. N., Avery, L., Njoroge, P. K., Kasonde, L., Gorgens, M., ... Moses, S. (2014). Estimating the size of the female sex worker population in Kenya to inform HIV prevention programming. *PloS One*, *9*(3), e89180.
<https://doi.org/10.1371/journal.pone.0089180>
- Okal, J., Geibel, S., Muraguri, N., Musyoki, H., Tun, W., Broz, D., ... Raymond, H. F. (2013). Estimates of the size of key populations at risk for HIV infection: men who have sex with men, female sex workers and injecting drug users in Nairobi, Kenya. *Sexually Transmitted Infections*, *89*(5), 366–371. <https://doi.org/10.1136/sextrans-2013-051071>
- Ortego, C., Huedo-Medina, T. B., Vejo, J., & Llorca, F. J. (2011). Adherence to highly active antiretroviral therapy in Spain. A meta-analysis. *Gaceta Sanitaria*, *25*(4), 282–289.
<https://doi.org/10.1016/J.GACETA.2010.10.016>
- Pan, H., Liu, S., Miao, D., & Yuan, Y. (2018). Sample size determination for mediation analysis of longitudinal data. *BMC Medical Research Methodology*, *18*(1), 32.
<https://doi.org/10.1186/s12874-018-0473-2>
- Parsons, J. T., Rosof, E., Punzalan, J. C., & Maria, L. Di. (2005). Integration of Motivational Interviewing and Cognitive Behavioral Therapy to Improve HIV Medication Adherence and Reduce Substance Use Among HIV-Positive Men and Women: Results of a Pilot Project. *AIDS Patient Care and STDs*, *19*(1), 31–39. <https://doi.org/10.1089/apc.2005.19.31>
- Pavlou, M., Ambler, G., Seaman, S., & Omar, R. Z. (2015). A note on obtaining correct marginal predictions from a random intercepts model for binary outcomes. *BMC Medical Research Methodology*, *15*, 59. <https://doi.org/10.1186/s12874-015-0046-6>
- PEPFAR [U.S. President's Emergency Plan for AIDS Relief]. (2017). *Kenya Country Operational Plan Strategic Direction Summary*. Retrieved from <http://www.pepfar.gov/documents/organization/250290.pdf>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R.-C. (2018). nlme: Linear and Nonlinear Mixed Effects Models. Retrieved from <https://cran.r-project.org/package=nlme>
- Prakash, R., Bhattacharjee, P., Blanchard, A., Musyoki, H., Anthony, J., Kimani, J., ... Moses, S. (2018). Effects of exposure to an intensive HIV-prevention programme on behavioural

- changes among female sex workers in Nairobi, Kenya. *African Journal of AIDS Research*, 17(2), 99–108. <https://doi.org/10.2989/16085906.2017.1377268>
- R Core Team. (2018). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Retrieved from <http://www.R-project.org/>
- Ramadhani, H. O., Thielman, N. M., Landman, K. Z., Ndosi, E. M., Gao, F., Kirchherr, J. L., ... Crump, J. A. (2007). Predictors of Incomplete Adherence, Virologic Failure, and Antiviral Drug Resistance among HIV-Infected Adults Receiving Antiretroviral Therapy in Tanzania. *Clinical Infectious Diseases*, 45(11), 1492–1498. <https://doi.org/10.1086/522991>
- Reza-Paul, S., Lorway, R., O'Brien, N., Lazarus, L., Jain, J., Bhagya, M., ... Steen, R. (2012). Sex worker-led structural interventions in India: a case study on addressing violence in HIV prevention through the Ashodaya Samithi collective in Mysore. *The Indian Journal of Medical Research*, 135, 98–106. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22382190>
- Rupérez, M., Pou, C., Maculve, S., Cedeño, S., Luis, L., Rodríguez, J., ... Paredes, R. (2015). Determinants of virological failure and antiretroviral drug resistance in Mozambique. *Journal of Antimicrobial Chemotherapy*, 70(9), 2639–2647. <https://doi.org/10.1093/jac/dkv143>
- Sathekge, M., Maes, A., Van de Wiele, C., & Wiele, C. Van De. (2013). FDG-PET imaging in HIV infection and tuberculosis. *Seminars in Nuclear Medicine*, 43(5), 349–366. <https://doi.org/10.1053/j.semnuclmed.2013.04.008>
- Secor, A. M., Wahome, E., Micheni, M., Rao, D., Simoni, J. M., Sanders, E. J., & Graham, S. M. (2015). Depression, substance abuse and stigma among men who have sex with men in coastal Kenya. *AIDS (London, England)*, 29 Suppl 3(0 3), S251-9. <https://doi.org/10.1097/QAD.0000000000000846>
- Singer, M. (1996). A dose of drugs, a touch of violence, a case of aids: conceptualizing the SAVA syndemic. *Free Inquiry in Creative Sociology*, 24(2), 99–110. Retrieved from <https://ojs.library.okstate.edu/osu/index.php/FICS/article/view/1346>

- Singer, M. (2010). Pathogen-pathogen interaction: A syndemic model of complex biosocial processes in disease. *Virulence, 1*(1), 10–18. <https://doi.org/10.4161/viru.1.1.9933>
- Singer, M., Bulled, N., Ostrach, B., & Mendenhall, E. (2017). Syndemics and the biosocial conception of health. *The Lancet, 389*(10072), 941–950. [https://doi.org/10.1016/S0140-6736\(17\)30003-X](https://doi.org/10.1016/S0140-6736(17)30003-X)
- Steenbergen, M. R., & Jones, B. S. (2002). Modeling Multilevel Data Structures. *American Journal of Political Science, 46*(1), 218. <https://doi.org/10.2307/3088424>
- Tedaldi, E. E. M., van den Berg-Wolf, M., Richardson, J., Patel, P., Durham, M., Hammer, J., ... SUN Study Investigators. (2012). Sadness in the SUN: Using computerized screening to analyze correlates of depression and adherence in HIV-infected adults in the United States. *AIDS Patient Care and STDs, 26*(12), 718–729. <https://doi.org/10.1089/apc.2012.0132>
- The Lancet. (2017, March 4). Syndemics: health in context. *The Lancet, 389*(10072), 881. [https://doi.org/10.1016/S0140-6736\(17\)30640-2](https://doi.org/10.1016/S0140-6736(17)30640-2)
- UNAIDS. (2016). Prevention Gap Report 2016. *Unaids, 83*(12), 286. <https://doi.org/10.1371/journal.pone.0154893>.
- UNAIDS. (2017). *UNAIDS Data 2017*. <https://doi.org/978-92-9173-945-5>
- Van Montfort, K., Oud, J. H. L., & Satorra, A. (2010). Longitudinal research with latent variables. In *Longitudinal Research with Latent Variables* (pp. 1–301). <https://doi.org/10.1007/978-3-642-11760-2>
- Wang, C., & Xue, X. (2016). Power and sample size calculations for evaluating mediation effects in longitudinal studies. *Statistical Methods in Medical Research, 25*(2), 686–705. <https://doi.org/10.1177/0962280212465163>
- Wawrzyniak, A. A. J., Rodríguez, A. E., Falcon, A. A. E., Chakrabarti, A., Parra, A., Park, J., ... Metsch, L. R. (2015). Association of individual and systemic barriers to optimal medical care in people living with HIV/AIDS in Miami-Dade County. *Journal of Acquired Immune Deficiency Syndromes, 69 Suppl 1*(Suppl 1), S63–S72. <https://doi.org/10.1097/QAI.0000000000000572>

Wen, L., Terrera, G. M., & Seaman, S. R. (2018). Erratum: Methods for handling longitudinal outcome processes truncated by dropout and death. *Biostatistics*.
<https://doi.org/10.1093/biostatistics/kxy001>

WHO. (2003). *Adherence to long-term therapies: Evidence for action*. World Health Organization (Vol. 2). [https://doi.org/10.1016/S1474-5151\(03\)00091-4](https://doi.org/10.1016/S1474-5151(03)00091-4)

WHO. (2012). *Antiretroviral treatment as prevention (TASP) of HIV and TB*. World Health Organization. <https://doi.org/WHO/HIV/2012.12>

Wilson, T. E., Barrón, Y., Cohen, M., Richardson, J., Greenblatt, R., Sacks, H. S., ... Women's Interagency HIV Study. (2002). Adherence to Antiretroviral Therapy and Its Association with Sexual Behavior in a National Sample of Women with Human Immunodeficiency Virus. *Clinical Infectious Diseases*, 34(4), 529–534. <https://doi.org/10.1086/338397>

Yang, W., Zilov, A., Soewondo, P., Bech, O. M., Sekkal, F., & Home, P. D. (2010). Observational studies: Going beyond the boundaries of randomized controlled trials. *Diabetes Research and Clinical Practice*, 88(SUPPL. 1), 3–9.
[https://doi.org/10.1016/S0168-8227\(10\)70002-4](https://doi.org/10.1016/S0168-8227(10)70002-4)

APPENDICES

Appendix A: Ethical Approval Letter



UNIVERSITY OF NAIROBI
COLLEGE OF HEALTH SCIENCES
P O BOX 19676 Code 00202
Telegrams: varsity
Tel:(254-020) 2726300 Ext 44355

KNH-UON ERC
Email: uonknh_erc@uonbi.ac.ke
Website: <http://www.erc.uonbi.ac.ke>
Facebook: <https://www.facebook.com/uonknh.erc>
Twitter: [@UONKNH_ERC](https://twitter.com/UONKNH_ERC) https://twitter.com/UONKNH_ERC



KENYATTA NATIONAL HOSPITAL
P O BOX 20723 Code 00202
Tel: 726300-9
Fax: 725272
Telegrams: MEDSUP, Nairobi

Ref. No.KNH/ERC/R/39

19th February 2018

Dr.Joshua Kimani
Co-Investigator
UNTID
College of Health Sciences
University of Nairobi

Dear Dr. Kimani

Re: Approval of Annual Renewal – Use of clinical care database by the University of Nairobi/University of Manitoba Research team to evaluate HIV prevention, care and treatment in Kenya (P258/09/2008)

Refer to your communication dated February 8, 2018.

This is to acknowledge receipt of the study progress report and hereby grant annual extension of approval for ethical research protocol P258/09/2008.

The approval dates are 18th February 2018 – 17th February 2019.

This approval is subject to compliance with the following requirements:

- a) Only approved documents (informed consents, study instruments, advertising materials etc) will be used.
- b) All changes (amendments, deviations, violations etc.) are submitted for review and approval by KNH- UoN ERC before implementation.
- c) Death and life threatening problems and severe adverse events (SAEs) or unexpected adverse events whether related or unrelated to the study must be reported to the KNH- UoN ERC within 72 hours of notification.
- d) Any changes, anticipated or otherwise that may increase the risks or affect safety or welfare of study participants and others or affect the integrity of the research must be reported to KNH- UoN ERC within 72 hours.
- e) Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. (*Attach a comprehensive progress report to support the renewal*).
- f) Clearance for export of biological specimens must be obtained from KNH- UoN-Ethics & Research Committee for each batch of shipment.

Protect to discover

- g) Submission of an *executive summary* report within 90 days upon completion of the study
This information will form part of the data base that will be consulted in future when processing related research studies so as to minimize chances of study duplication and/or plagiarism.

Ensure that the ethical renewal is renewed timely as per KNH-UoN ERC requirements.

For more details consult the KNH- UoN ERC website <http://www.erc.uonbi.ac.ke>

Yours sincerely,



PROF. M.L. CHINDIA
SECRETARY, KNH-UON ERC

c.c. The Principal, College of Health Sciences, UoN
The Deputy Director CS, KNH
The Chairperson, KNH-UoN ERC

Protect to discover

Appendix B: R scripts – Exploring log viral load longitudinal data

##Individual profile plots

```
# plotting for all the sex workers, first
i.swop<-read.table(file="C:/MyPath/i.swop.csv", header=TRUE, sep="") #reading full swop data
View(i.swop)
#dropping the variables not needed currently
dropvars<-names(i.swop) %in% c("N", "var")
i.swop <- i.swop[!dropvars]
View(i.swop)
#plot
library(ggplot2) #needs ggplot2 R package
p_swop<- ggplot(data = i.swop, aes(x = visit, y = lgviral, group = sw_id))
p_swop + geom_line()
#all sex workers, separately by adherence category
p_swop + geom_line() + facet_grid(. ~ lastARVadherence)
#plotting individual profiles for a randomly selected 15 patients
#first reading in data file containing swop data for 10 randomly selected sex workers

i.swop<-read.table(file="C:/MyPath/i.swop.csv", header=TRUE, sep="")
View(i.swop)
library(ggplot2) #needs ggplot2 R package too

is.factor(i.swop$id) #id variable needs to be in factor form

as.factor(i.swop$id) #convert to factor
p_swop <- ggplot(data = i.swop, aes(x = months, y = logVLplus1, group = id)) #defining graph
#plotting
p_swop + geom_line()
#Average profile plots
swop<-read.table(file="C:/MyPath/i.swop.csv", header=TRUE, sep="") #reading full swop data
View(i.swop)
#####
#getting average profile plot
```

```
#####
#first getting mean at each visit time point
swopmean_v<-aggregate(i.swop$ logVLplus1, by=list(i.swop$months), FUN=mean)

View(swopmean_v)
names(swopmean_v)
names(swopmean_v)<-c("months","mean_ logVLplus1") #renaming variables accordingly
View(swopmean_v)
#simple line plot as average profile plot
plot(swopmean_v$months, swopmean_v$mean_ logVLplus1, type="l", ylab="Average Log HIV
Viral Load", xlab="Month")
```

```
#####
```

#Obtaining average profile plot by adherence category

```
#####
```

```
#first obtaining mean at each visit time point by adherence
swopmean_v_byadhr<-aggregate(i.swop$ logVLplus1,
by=list(i.swop$months,i.swop$lastARVadherence), FUN=mean)
View(swopmean_v_byadhr)
names(swopmean_v_byadhr)
names(swopmean_v_byadhr)<-c("Months","Adherence","Mean") #renaming variables
accordingly
View(swopmean_v_byadhr)

#average profile plots by adherence
swopmean_byadhr<- ggplot(data = swopmean_v_byadhr, aes(x = Visit, y = Mean, group =
lastARVAdherence))
swopmean_byadhr + geom_line(mapping=aes(colour=lastARVAdherence))
```

Appendix C: R scripts for – the evolutions in log₁₀ viral load over time by adherence

#Tests for fixed effects

```
library(nlme) #call the nlme library
#fitting random intercept model
```

```

randint_logVLplus1 <- lme(logVLplus1 ~ month*lastARVadherence, data = i.swop, random = ~
1|id, na.action=na.omit)
summary(randint_logVLplus1)

#fitting random slope model via restricted maximum likelihood (REML) – default
randslp_lgviral <- lme(logVLplus1 ~ month*lastARVdherence, data = i.swop, random = ~
1+visit|id, na.action=na.omit)
summary(randslp_logVLplus1)

#####
#fitting same model under ML
#####
#first call the nlme library
library(nlme)
#fit random slope model
randslp_lgviralML <- lme(lgviral ~ visit*adhere, random = ~ 1+visit|id, method="ML")
summary(randslp_lgviralML)

#testing for interaction using likelihood ratio test
#fitting reduced model - without interaction
randslp_lgviralML_red <- lme(logVLplus1 ~ monthst+lastARVadherence, random = ~
1+visit|id, method="ML")
summary(randslp_lgviralML_red)
#log likelihood for this model is xxxx.xxx
#log likelihood for model with interaction (above) will be yyyy.yyy
#computing likelihood ratio test statistic
lrt<--2*(xxxx.xxx+yyyy.yyy)
lrt

#likelihood ratio test for significance - using Chi-Square with 2 degrees of freedom
pchisq(lrt, df=2, lower.tail=FALSE)
##repeating interaction test using anova function
anova(randslp_lgviralML, randslp_lgviralML_red)
#repeating test for interaction - F test

```

```
anova(randslp_lgviralML, type="marginal")
```

Appendix D: R scripts – Test for the need of the random effects

```
#fitting and testing for random slope
```

```
randslp_lgviral<- lme(logVLplus1 ~ months*lastARVadherence, random = ~ 1+months|id)
```

```
summary(randslp_lgviral)
```

```
#maximum likelihood value = xxxx.xxx
```

```
#fitting random intercept model
```

```
randint_lgviral<- lme(logVLplus1 ~ months*lastARVadherence, random = ~ 1|id)
```

```
summary(randint_lgviral)
```

```
#maximum lik. value = xxxx.xxx
```

```
#likelihood ratio test statistic
```

```
lik_slp<--2*(xxxx.xxx+yyyy.yyy)
```

```
#mixture of chi-square test with 1 and 2 df
```

```
p_value<-0.5*pchisq(lik_slp, df=1, lower.tail=FALSE)+0.5*pchisq(lik_slp, df=2,  
lower.tail=FALSE)
```

```
p_value
```

```
#####
```

```
#testing for random intercept
```

```
#####
```

```
#random intercept model
```

```
randint_lgviralML<- lme(logVLplus1 ~ months*lastARVadherence, random = ~ 1|id,  
method="ML")
```

```
summary(randint_lgviralML)
```

```
#fit model without any random effect - just ordinary multiple linear regression
```

```
linreg_lgviral<-lm(logVLplus1 ~ months*lastARVadherence)
```

```
summary(linreg_lgviral)
```

```
#extract max lik value
```

```
logLik(linreg_lgviral) #xxxx.xxx
```

```

#likelihood ratio stat
lik_int<--2*(xxxx.xxx+yyyy.yyy)
#mixture of chi-square with 0 and 1 df
p_value_int<-0.5*pchisq(lik_int, df=1, lower.tail=FALSE)
p_value_int

#####
#estimates of the random effects
#####
random.effects(randslp_lgviral)
#####
#fitted values
#####
fitted(randslp_lgviral, level=0) #the cohort/community viral load – population average
fitted(randslp_lgviral, level=1) #a sex worker viral load – subject specific

#####

#####LINEAR MIXED MODELS#####

#####RANDOM EFFECTS#####

i.swop<-read.csv(file.choose())

#first call the nlme library

library(nlme)

library(lme4)

library(lmerTest)

###Random intercepts

modell1 <- gls(logVLplus1 ~ months*lastARVAdherence, data =
i.swop,na.action=na.omit,method="ML")

summary(modell1)

```

```

model2 <- lmerTest::lmer(logVLplus1 ~ months*lastARVAdherence +(1 | id), data =
i.swop,na.action=na.omit,REML=FALSE)

summary(model2)

model3 <- lme(logVLplus1 ~ months*lastARVAdherence, random = ~1|id, data =
i.swop,na.action=na.omit,method="ML")

summary(model3)

lrt1<-anova(model1, model2)

lrt1

lrt3<-anova(model1, model2)

lrt3

lrt5<-anova(model1, model3)lrt5

##Random slopes

model4 <- lmerTest::lmer(logVLplus1 ~ months*lastARVAdherence +(1 | id), data =
i.swop,na.action=na.omit,method="ML")

summary(model4)

lrt2<-anova(model1, model4)

lrt2

model5 <- lmer(logVLplus1 ~ months*lastARVAdherence, random = ~months|id, data =
i.swop,na.action=na.omit,REML=FALSE)

summary(model5)

model6 <- lme(logVLplus1 ~ months*lastARVAdherence, random = ~1+months|id, data =
i.swop,na.action=na.omit,REML=FALSE)

summary(model6)

m2 <- lmer(logVLplus1 ~ months*lastARVAdherence + (1+month|id), i.swop)

```



```

summary(m2)

randslp_lgviral<- lme(logVLplus1 ~ months*lastARVAdherence, random = ~ 1+months|id,data
= i.swop,na.action=na.omit,method="ML")

summary(randslp_lgviral)

randslp_lgviral<- lme(logVLplus1 ~ month*lastARVAdherence, random = ~ 1+month|id,data =
i.swop,na.action=na.omit)

summary(randslp_lgviral)

fit1<-(lmerTest::lme(logVLplus1~months*lastARVAdherence+(months|id),data=i.swop))

fit1

pchisq(2*25.045,df=1,lower.tail=FALSE,log.p=TRUE)/log(10)

## -11.83312

p<-10 ^-11.83312

p

```

Appendix E: R scripts – Box-Cox transformation of HIV viral loads

```

#####
#####BOX-COX TRANSFORMATION OF VIRAL LOADS#####
#####
plotNormalHistogram(i.swop$logVLplus1) ##Appears multimodal even after log transformation

if(!require(MASS)){install.packages("MASS")}

if(!require(rcompanion)){install.packages("rcompanion")}

library(MASS)

Box = boxcox(i.swop$VLplus1 ~ 1, # Transform VL as a single vector

lambda = seq(-2,2,0.1) # using default values -2 to 2 by 0.1

)

```

```

Cox = data.frame(Box$x, Box$y) # Creating a data frame having the results

Cox

Cox2 = Cox[with(Cox, order(-Cox$Box.y)),] # Ordering the created data frame by decreasing y

Cox2

Cox2[1,] # Displaying the lambda value with the greatest log likelihood

lambda = Cox2[1, "Box.x"] # Extracting the lambda produced

lambda

T_box = (i.swop$VLplus1 ^ lambda - 1)/lambda # Transform the original data

T_box

library(rcompanion)

plotNormalHistogram(T_box)

##Back-transformation to HIV viral loads

VL_1 = exp(log(-0.2222222 * T_box + 1) / -0.2222222)

VL_1

VL = (exp(log(-0.2222222 * T_box + 1) / -0.2222222))-1 ##1 was added to each viral load

VL

```

Appendix F: R scripts – Model fit by Maximum Likelihood

```
#####model fit by Maximum Likelihood#####
```

```

library(nlme)

library(lme4)

library(lmerTest)

i.swop<-read.csv(file.choose())

model_nlme<- lme(vlbox ~ lastARVAdherence*month, random = ~month|id, data =
i.swop,na.action=na.omit,method="ML") ##Random slope

```

```

summary(model_nlme) ##nlme

#By lmerTest/lme4

modelML_int <- lmerTest::lmer(vlbox ~ month + (1|id),
data=i.swop,na.action=na.omit,REML=FALSE)

summary(modelML_int)

coef(modelML_int)

modelML_int1 <- lmer(vlbox ~ lastARVAdherence*month + (1|id),
data=i.swop,na.action=na.omit,REML=FALSE)

summary(modelML_int1)

#Random slope only by lmerTest/lme4

modelML_slp <- lmerTest::lmer(vlbox ~ month + (month+ 0| id),
data=i.swop,na.action=na.omit,REML=FALSE)

summary(modelML_slp) ##by month only

coef(modelML_slp)

modelML_slp1 = lmerTest::lmer(vlbox ~ lastARVAdherence*month + (month+ 0| id),
data=i.swop,na.action=na.omit,REML=FALSE)

summary(modelML_slp1)

coef(modelML_slp1)

anova(modelML_int,modelML_slp)

anova(modelML_slp,modelML_int)

##Unidentifiable random effects when fitting both intercept and slopes

modelML_RE <- lmer(vlbox ~ lastARVAdherence*month + (1+month|id),
data=i.swop,na.action=na.omit,REML=FALSE)

summary(modelML_RE)

```

Appendix G: R scripts – Subsetting data by random selection of unique rows for multiple outputation

###Multiple outputation–within cluster resampling #4 samples

```
#####
```

```
i.swop<-read.csv(file.choose())
```

```
library(plyr)
```

```
###Sample 1
```

```
i.swop1 <- ddply(i.swop,.(id),
```

```
  function(x) {
```

```
    x[sample(nrow(x),size=1),]
```

```
  })
```

```
write.csv(i.swop1, file = "i.swop1.csv")
```

```
###Sample 2
```

```
i.swop2 <- ddply(i.swop,.(id),
```

```
  function(x) {
```

```
    x[sample(nrow(x),size=1),]
```

```
  })
```

```
write.csv(i.swop2, file = "i.swop2.csv")
```

```
###Sample 3
```

```
i.swop3 <- ddply(i.swop,.(id),
```

```
  function(x) {
```

```
    x[sample(nrow(x),size=1),]
```

```
  })
```

```
write.csv(i.swop3, file = "i.swop3.csv")
```

```
###Sample 4
```

```
i.swop4 <- ddply(i.swop,.(id),
```

```
  function(x) {
```

```
    x[sample(nrow(x),size=1),]
```

```
  })
```

```

write.csv(i.swop4, file = "i.swop4.csv")
#####Within cluster resampling – multiple outputation for syndemics#####
i.swop1<-read.csv(file.choose())
i.swop2<-read.csv(file.choose())
i.swop3<-read.csv(file.choose())
i.swop4<-read.csv(file.choose())

lm_model1<-
lm(vlbox~CDs+STI+Sex_Alcohol+Use_Drugs+GBV_3Mth+NutritionAssessment,data=i.swop1
)
summary(lm_model1)

lm_model2<-
lm(vlbox~CDs+STI+Sex_Alcohol+Use_Drugs+GBV_3Mth+NutritionAssessment,data=i.swop2
)
summary(lm_model2)

lm_model3<-
lm(vlbox~CDs+STI+Sex_Alcohol+Use_Drugs+GBV_3Mth+NutritionAssessment,data=i.swop3
)
summary(lm_model3)

lm_model4<-
lm(vlbox~CDs+STI+Sex_Alcohol+Use_Drugs+GBV_3Mth+NutritionAssessment,data=i.swop4
)
summary(lm_model4)

#####Within cluster resampling – multiple outputation for viral load and#####
#####ARV adherence*month interaction#####

model_vl<-
lm(vlbox~lastARVAdherence+lastARVAdherence*month,data=i.swop1,na.action=na.omit,RE
ML=FALSE)
summary(model_vl)

vcov(model_vl)[2,2] ##Variance of the slope
(summary(model_vl)$sigma)**2 ##Residual variance

model_vl2<-
lm(vlbox~lastARVAdherence+lastARVAdherence*month,data=i.swop2,na.action=na.omit,RE
ML=FALSE) ##REML disregarded in the model

```

```

summary(model_vl2)
(summary(model_vl2)$sigma)**2 ##Residual variance
vcov(model_vl2)[2,2] ##Variance of the slope

model_vl3<-
lm(vlbox~lastARVAdherence+lastARVAdherence*month,data=i.swop3,na.action=na.omit,RE
ML=FALSE)
summary(model_vl3)
vcov(model_vl3)[2,2] ##Variance of the slope
(summary(model_vl3)$sigma)**2 ##Residual variance

model_vl4<-
lm(vlbox~lastARVAdherence+lastARVAdherence*month,data=i.swop4,na.action=na.omit,RE
ML=FALSE)
summary(model_vl4)
vcov(model_vl4)[2,2] ##Variance of the slope
(summary(model_vl4)$sigma)**2 ##Residual variance

```