THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF DEPARTMENT

ESTIMATION OF THE TRANSITION PROBABILITIES BETWEEN HIV DISEASE PROGRESSION CATEGORIES

SHANGLUN LI
SPRING 2018

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Mathematics and Statistics
with interdisciplinary honors in Mathematics and Statistics

Reviewed and approved* by the following:

Le Bao
Professor of Statistics
Thesis Supervisor

Sergei Tabachnikov
Professor of Mathematics
Honors Adviser

Murali Haran
Professor of Statistics
Honors Adviser

*Signatures are on file in the Schreyer Honors College.

Abstract

In this project, we investigate the effect of treatment, age, and gender to the progression of AIDS by using CD4 data. We fit a Markov chain model, and use the Maximum Likelihood approach to estimate the transition probability matrices. Some numerical analysis are applied on the resulting transition probabilities, such as comparisons of confidence intervals of transition probabilities across different subgroups.

Table of Contents

List of Figures

List of Tables

Acknowledgements

The research was supervised by Dr.Bao, who is an associate professor in Statistics at Penn State. Dr.Bao has made several contributions in understanding HIV/AIDS epidemics. He is currently a key technical advisor of the UNAIDS Reference Group on HIV Estimates, Modeling and Projections who advises on the methods for calculating international AIDS statistics, and a core project team leader of the Diagnostics Modeling Consortium who aims to utilize modeling to guide the effective use of diagnostic technologies in resource-poor settings.

This research was sponsored by UNAIDS and the National Institutes of Health, National Institute of Allergy and Infectious Diseases, R01AI136664-01.

I also appreciate the opportunity, which is given by my parents, to study at Penn State University.

Chapter 1

Introduction

Figure 1.1: HIV and AIDS

The human immunodeficiency virus (HIV) is a lentivirus, a subgroup of retrovirus, that causes HIV infection and over time acquired immunodeficiency syndrome (AIDS).[1][2] AIDS is a condition in humans in which progressive failure of the immune system allows life-threatening opportunistic infections and cancers to thrive. Without treatment, average survival time after infection with HIV is estimated to be 9 to 11 years, depending on the HIV subtype.[3] In most cases, HIV is a sexually transmitted infection and occurs by contact with or transfer of blood, pre-ejaculate, semen, and vaginal fluids. Non-sexual transmission can occur from an infected mother to her infant through breast milk.[4][5][6] An HIV-positive mother can transmit HIV to her baby both during pregnancy and childbirth due to exposure to her blood or vaginal fluid.[7] Within these bodily fluids, HIV is present as both free virus particles and virus within infected immune cells.

Antiretroviral Therapy (ART) are medications that treat HIV. The drugs do not kill or cure the virus. However, when taken in combination they can prevent the growth of the virus. When the virus is slowed down, so is HIV disease. Antiretroviral drugs are referred to as ARV. Combination ARV therapy (cART) is referred to as highly active ART (HAART). [8]

HIV progression (often measured by CD4 counts) and death rates are key HIV indicators. The current software used by UNAIDS and 163 countries uses estimates derived from the ALPHA network, using data from Eastern and Southern Africa. It likely provides biased estimates for Asian population. I will work on modeling the progression and survival of HIV/AIDS with a cohort of more than 30,000 Chinese participants. It is the largest HIV cohort study in the world, but the data has never been fully explored and analyzed. We will apply the Markov Chain Model to those data, estimate the transmission rates between CD4 categories and the death rate within each category, and account of the possible CD4 measurement errors.

Chapter 2

Method

2.1 Data Description and Method Abstraction

HIV progression (often measured by CD4 counts) and death rates are key HIV indicators. The current software used by UNAIDS and 163 countries uses estimates derived from the ALPHA network, using data from Eastern and Southern Africa. It likely provides biased estimates for Asian population. We model the progression and survival of HIV/AIDS for hundred-thousands of Chinese participants. It is the largest HIV cohort study in the world, but the data has never been fully explored and analyzed by using statistical models. We estimate the transmission rates between CD4 categories and the death rate within each category in the Markov chain model, and account of the possible CD4 measurement errors. In this project, we use a subset of 30,000 individuals provided by National Center for AIDS/STD Control and Prevention(NCAIDS), Chinese Center for Disease Control and Prevention(China CDC). The results are only for illustration purpose and should not be viewed as the offical results for China.

2.2 Variable Description

There are 13 variables in the data - ID, gender, DateOfBirth, CD4count, CD4date, DiagnosisDateOfHIV, DiagnosisDateOfAIDS, ARTstart, art, DateOfDeath, PrimaryCauseOfDeath, CategoryOfPrimaryCOD, and ARTAge, as shown in the variable attribute table below.

Variable Attributes

Variable	**Description**	**Levels**
ID	The patients' identification number	16-digits number
gender	The gender of the patient	"Femal" and "Male"
DateOfBirth	The date of birth of the patient	Any dates
CD4count	The CD4 count of the patient	Positive number
CD4date	The CD4 measurement date of the patient	Any date
DiagnosisDateOfHIV	The date of diagnosis of HIV for patient	Any date
DiagnosisDateOfAIDS	The date of diagnosis of AIDS for patient	Any date
ARTstart	The date that patient receive ART treatment	Any date
art	Indicate whether the patient receive ART treatment	1 - ART; 9 - no ART
DateOfDeath	The date that the patient died	Any date
PrimaryCauseOfDeath	The cause of death	Text
CategoryOfPrimaryCOD	The category of the cause of death	Text

Table 2.1: Variable attributes table includes the name, description, levels for each variable.

2.3 Data Modification

From the original data frame, three sub-data frames were created:

- The state data frame: There were 8 possible states in the data frame. State 1 is the state that CD4 count is higher than 500; State 2 is the state that CD4 count is between 350 and 500; State 3 is the state that CD4 count is between 250 and 350; State 4 is the state that CD4 count is between 200 and 250; State 5 is the state that CD4 count is between 100 and 200; State 6 is the state that CD4 count is between 50 and 100; State 7 is the state that CD4 count is less than 50; and state 8 is the death state. Each row in the data frame represents one individual with each cell represents each CD4 measurement.

- The period data frame: each row represents one individual with the number of months between each CD4 measurement.

- The age data frame: the age has been categorized into three different age ranges - 15 to 35, 35 to 45, and 45 to 55 (otherwise, do not take into account, refer to section 2.4 for more details). Each row represents one individual with the age category at each CD4 measurement time.

2.4 Assumptions and Subset of Data

The data followed the below assumptions and subsetted:

- Each individual can only move the adjacent state (higher or lower) or stay in the same state for each month, The individuals which has more than one state change in one month were removed from the data set.

- Only the individuals with more than one measurement were taken into account.

- Only the individuals between 15 and 55 years old were taken into account.

2.5 Description of Method

Markov Chain model was used to fit the data in this study. It is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.[9] In this study, we can see that the state of CD4 count level for each individual depends only on the previous state of CD4 count level, which fits the Markov Chain model well.

This project aimed at estimating Markov transition probability matrix for different groups of individuals. Groups were separated at different age categories, gender, and if the individual received ART treatment. Thus, there are 12 different groups. Maximum likelihood estimation (MLE) were used in this case. Maximum likelihood is by far the most popular general method of estimation.

Its widespread acceptance is seen on the one hand in the very large body of research dealing with its theoretical properties, and on the other in the almost unlimited list of applications.[10] MLE is a method of estimating the parameters of a statistical model, given observations. MLE attempts to find the parameter values that maximize the likelihood function, given the observations. The resulting estimate is called a maximum likelihood estimate, which is also abbreviated as MLE.

R was used to implement the optimization of the likelihood function. We used optim() function to maximize the likelihood with method "L-BFGS-B". Method "L-BFGS-B" which allows box constraints, that is each variable can be given a lower and/or upper bound; that is, constraints of the form $l_i \leq x_i \leq u_i$ where l_i and u_i are per-variable constant lower and upper bounds, respectively (for each x_i, either or both bounds may be omitted). The initial value must satisfy the constraints. The method works by identifying fixed and free variables at every step (using a simple gradient method), and then using the L-BFGS method on the free variables only to get higher accuracy, and then repeating the process. [11] [12]

2.5.1 Transformation of Transition Probability

Since we want to maximize the likelihood function using method "L-BFGS-B", narrow intervals of constraints is not optimal for the function to operate. Thus, we did a transformation on the transition probabilities so that the lower bound and upper bound can be positive and negative infinity. We created a vector called $\alpha = (\alpha_1, \alpha_2, ..., \alpha_{19})$, and transformed the transition probability matrix as the following:

$$\begin{bmatrix} \theta_{11} & \theta_{12} & 0 & 0 & 0 & 0 & 0 & \theta_{18} \\ \theta_{21} & \theta_{22} & \theta_{23} & 0 & 0 & 0 & 0 & \theta_{28} \\ 0 & \theta_{32} & \theta_{33} & \theta_{34} & 0 & 0 & 0 & \theta_{38} \\ 0 & 0 & \theta_{43} & \theta_{44} & \theta_{45} & 0 & 0 & \theta_{48} \\ 0 & 0 & 0 & \theta_{54} & \theta_{55} & \theta_{56} & 0 & \theta_{58} \\ 0 & 0 & 0 & 0 & \theta_{65} & \theta_{66} & \theta_{67} & \theta_{68} \\ 0 & 0 & 0 & 0 & 0 & \theta_{76} & \theta_{77} & \theta_{78} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

where,

$$\theta_{11} = \frac{1}{1 + e^{\alpha_1} + e^{\alpha_2}}, \theta_{12} = \frac{e^{\alpha_1}}{1 + e^{\alpha_1} + e^{\alpha_2}}, \theta_{18} = \frac{e^{\alpha_2}}{1 + e^{\alpha_1} + e^{\alpha_2}}$$

$$\theta_{21} = \frac{e^{\alpha_3}}{1 + e^{\alpha_3} + e^{\alpha_4} + e^{\alpha_5}}, \theta_{22} = \frac{1}{1 + e^{\alpha_3} + e^{\alpha_4} + e^{\alpha_5}}, \theta_{23} = \frac{e^{\alpha_4}}{1 + e^{\alpha_3} + e^{\alpha_4} + e^{\alpha_5}}$$

$$\theta_{28} = \frac{e^{\alpha_5}}{1 + e^{\alpha_3} + e^{\alpha_4} + e^{\alpha_5}}, \theta_{32} = \frac{e^{\alpha_6}}{1 + e^{\alpha_6} + e^{\alpha_7} + e^{\alpha_8}}, \theta_{33} = \frac{1}{1 + e^{\alpha_6} + e^{\alpha_7} + e^{\alpha_8}}$$

$$\theta_{34} = \frac{e^{\alpha_7}}{1 + e^{\alpha_6} + e^{\alpha_7} + e^{\alpha_8}}, \theta_{38} = \frac{e^{\alpha_8}}{1 + e^{\alpha_6} + e^{\alpha_7} + e^{\alpha_8}}, \theta_{43} = \frac{e^{\alpha_9}}{1 + e^{\alpha_9} + e^{\alpha_{10}} + e^{\alpha_{11}}}$$

$$\theta_{44} = \frac{1}{1 + e^{\alpha_9} + e^{\alpha_{10}} + e^{\alpha_{11}}}, \theta_{45} = \frac{e^{\alpha_{10}}}{1 + e^{\alpha_9} + e^{\alpha_{10}} + e^{\alpha_{11}}}, \theta_{48} = \frac{e^{\alpha_{11}}}{1 + e^{\alpha_9} + e^{\alpha_{10}} + e^{\alpha_{11}}}$$

$$\theta_{54} = \frac{e^{\alpha_{12}}}{1 + e^{\alpha_{12}} + e^{\alpha_{13}} + e^{\alpha_{14}}}, \theta_{55} = \frac{1}{1 + e^{\alpha_{12}} + e^{\alpha_{13}} + e^{\alpha_{14}}}, \theta_{56} = \frac{e^{\alpha_{13}}}{1 + e^{\alpha_{12}} + e^{\alpha_{13}} + e^{\alpha_{14}}}$$
$$\theta_{58} = \frac{e^{\alpha_{14}}}{1 + e^{\alpha_{12}} + e^{\alpha_{13}} + e^{\alpha_{14}}}, \theta_{65} = \frac{e^{\alpha_{15}}}{1 + e^{\alpha_{15}} + e^{\alpha_{16}} + e^{\alpha_{17}}}, \theta_{66} = \frac{1}{1 + e^{\alpha_{15}} + e^{\alpha_{16}} + e^{\alpha_{17}}}$$
$$\theta_{67} = \frac{e^{\alpha_{16}}}{1 + e^{\alpha_{15}} + e^{\alpha_{16}} + e^{\alpha_{17}}}, \theta_{68} = \frac{e^{\alpha_{17}}}{1 + e^{\alpha_{15}} + e^{\alpha_{16}} + e^{\alpha_{17}}}, \theta_{76} = \frac{e^{\alpha_{18}}}{1 + e^{\alpha_{18}} + e^{\alpha_{19}}}$$
$$\theta_{77} = \frac{1}{1 + e^{\alpha_{18}} + e^{\alpha_{19}}}, \theta_{78} = \frac{e^{\alpha_{19}}}{1 + e^{\alpha_{18}} + e^{\alpha_{19}}}$$

2.5.2 Individual Fit Method

We create 12 separate functions that take the vector of α to calculate the log-likelihood of 12 different combination of treatment groups. Then, we use optimization function to get the 12 vectors of α that maximize the log-likelihood. Then, we transformed each α vector back to the transition probability matrix so that we got the final 12 transition probability matrices for all 12 combinations of treatment groups.

Then, we find the hessian matrix from the optim() function in R, and solve the matrix to find the covariance matrix for α. Bootstrap was used next to find the 95% confidence interval. The basic algorithm for bootstrap in this case is to generate 10,000 α vectors according to its distribution, then transform them back to 10,000 θ vectors. From the generated 10,000 θ vectors, we can find the 2.5 percentile and 97.5 percentile in the sample, which are the lower bound and upper bound for the 95% confidence interval for θ.

2.5.3 Combined Fit Method

Different from individual fit method, for combined fit method, we create a function for α to represent all possible combinations of treatment groups, as shown below:

$$\alpha = \beta_0 + \beta_1 \cdot \text{I.gender} + \beta_2 \cdot \text{I.age2} + \beta_3 \cdot \text{I.age3} + \beta_4 \cdot \text{I.ART}$$

where we have the table below for the different encoding for different combination of treatment groups:

	I.gender	I.age2	I.age3	I.ART
male, ART, age1	0	0	0	1
male, ART, age2	0	1	0	1
male, ART, age3	0	0	1	1
male, no ART, age1	0	0	0	0
male, no ART, age2	0	1	0	0
male, no ART, age3	0	0	1	0
female, ART, age1	1	0	0	1
female, ART, age2	1	1	0	1
female, ART, age3	1	0	1	1
female, no ART, age1	1	0	0	0
female, no ART, age2	1	1	0	0
female, no ART, age3	1	0	1	0

Table 2.2: Table of encoding

We used this encoding to find the likelihood of all combinations of treatment groups using one single likelihood function. Then, optim() function was used to find the parameters that maximize the likelihood function. Finally, we can transform back to 12 different transition probability matrices for 12 different combinations of treatment groups.

Chapter 3

Result

3.1 Individual Fit Results

After the Maximum Likelihood estimation for the transition probability matrix using individual fit method, the results were the following (The colors from light to heavy represent the probabilities in the intervals 0 to 0.00001, 0.00001 to 0.1, 0.1 to 0.25, 0.25 to 0.5, and 0.5 to 1, respectively.):

0.90583	0.09374	0.00000	0.00000	0.00000	0.00000	0.00000	0.00043
0.09734	0.78245	0.11968	0.00000	0.00000	0.00000	0.00000	0.00052
0.00000	0.18600	0.67871	0.13439	0.00000	0.00000	0.00000	0.00091
0.00000	0.00000	0.35949	0.44880	0.18986	0.00000	0.00000	0.00185
0.00000	0.00000	0.00000	0.17847	0.77560	0.04216	0.00000	0.00377
0.00000	0.00000	0.00000	0.00000	0.24071	0.62375	0.12769	0.00785
0.00000	0.00000	0.00000	0.00000	0.00000	0.31145	0.64052	0.04803
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 3.1: Transition probability matrix of male patients with ART treatment in the age range between 15 and 35.

As we can see from Table 3.1, death rate increases as CD4 count decreases (more unhealthy). In addition, for a fixed CD4 level, the probability of staying in that CD4 level for the next stage is the highest among all possible outcomes. We can also see that the transition probability of getting to a higher CD4 level (healthier) is higher than that of getting to a lower CD4 level (more unhealthy). For other transition probability matrices, see Appendix A.

Then, we use figure to compare each individual's transition probability across sub-population defined by definition of gender, ART group, and age group, as shown below. The blue arrow is the 95% confidence interval for the transition probability. In this section, only 4 out of 26 graphs are shown below. Refer to Appendix B for the rest of the graphs.

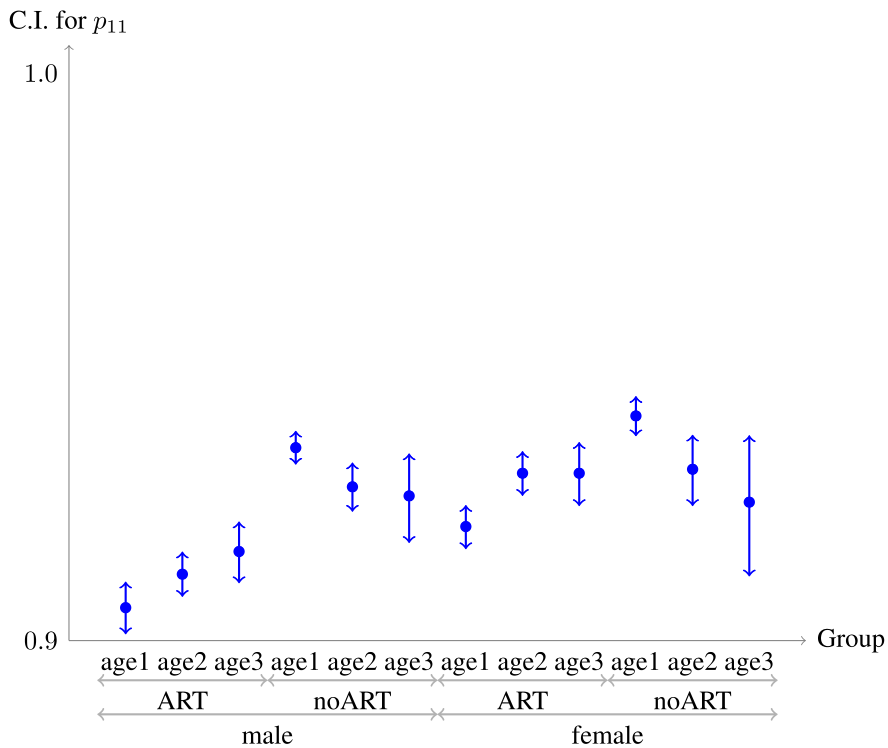

Figure 3.1: p_{11} comparison among different group combinations.

As we can see from Figure 3.1, for the individuals who receive ART treatment, the probability of staying in the same CD4 count level (state 1) is higher for older age group; while for the individuals who do not receive ART treatment, the probability of staying in the same CD4 count level (state 1) is lower for older age group. In addition, female seems to have a higher probability of staying in the same CD4 count level (state 1) than male. And, no ART group seems to have a higher probability of staying in the same CD4 count level (state 1) than ART group.

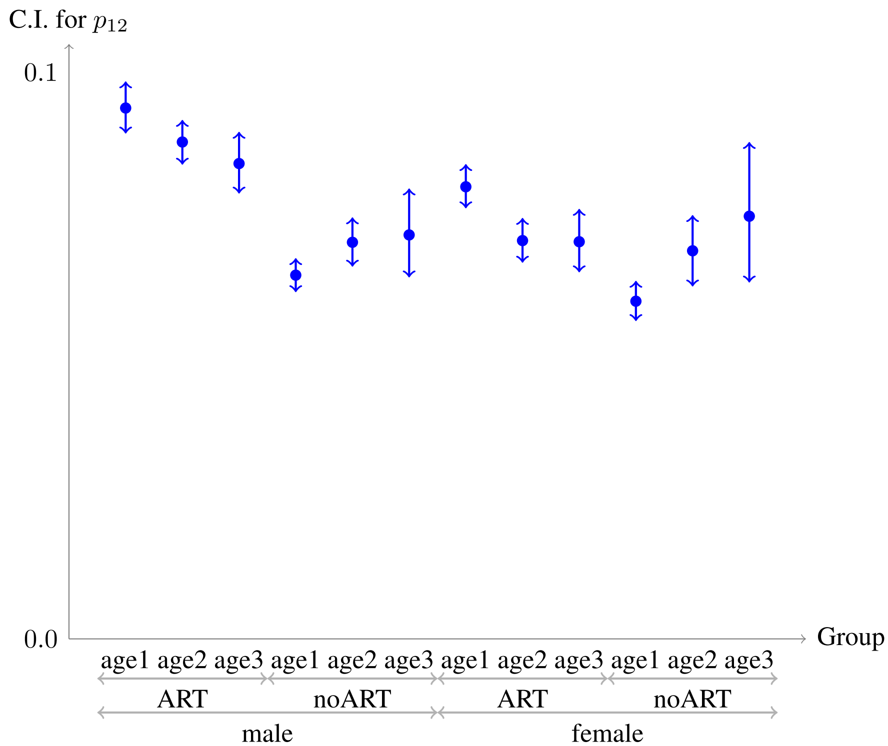

Figure 3.2: p_{12} comparison among different group combinations.

As we can see from Figure 3.2, for the individuals who receive ART treatment, the probability of moving to the lower CD4 count level (more unhealthy) is lower for older age group; while for the individuals who do not receive ART treatment, the probability of moving to the lower CD4 count level (more unhealthy) is higher for older age group. In addition, for ART treatment group, female seems to have a lower probability of moving to the lower CD4 count level (more unhealthy) than male. And, no ART group seems to have a lower probability of moving to the lower CD4 count level (more unhealthy) than ART group.

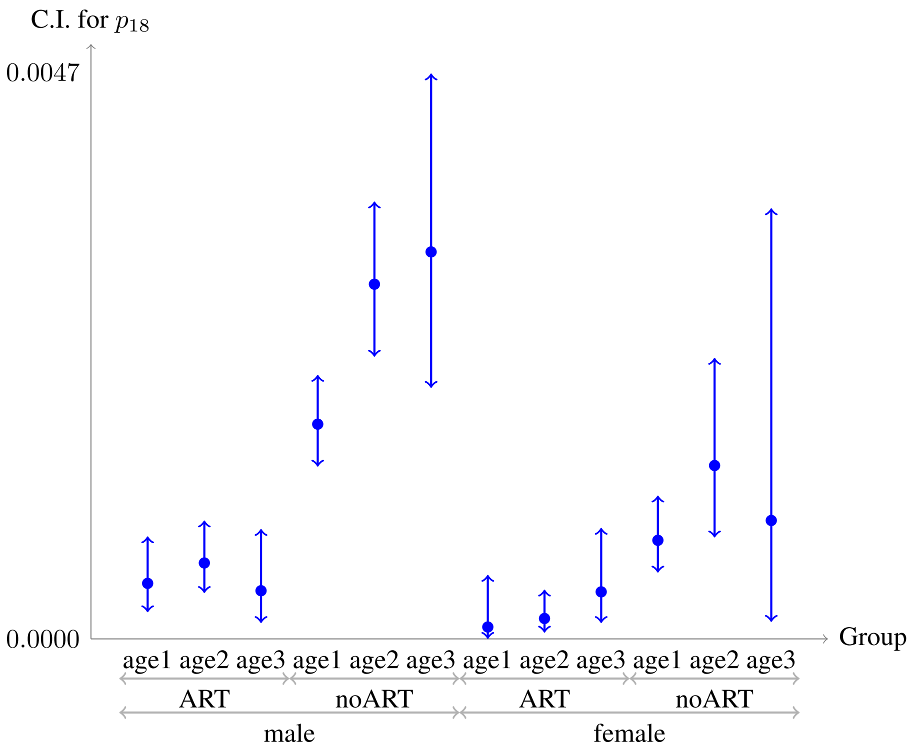

Figure 3.3: p_{18} comparison among different group combinations.

As we can see from Figure 3.3, for the individuals who receive ART treatment, the probabilities of death are similar among different age groups; while for the individuals who do not receive ART treatment, the probability of death is higher for older age group. In addition, female seems to have a lower probability of death than male. And, no ART group seems to have a higher probability of death than ART group.

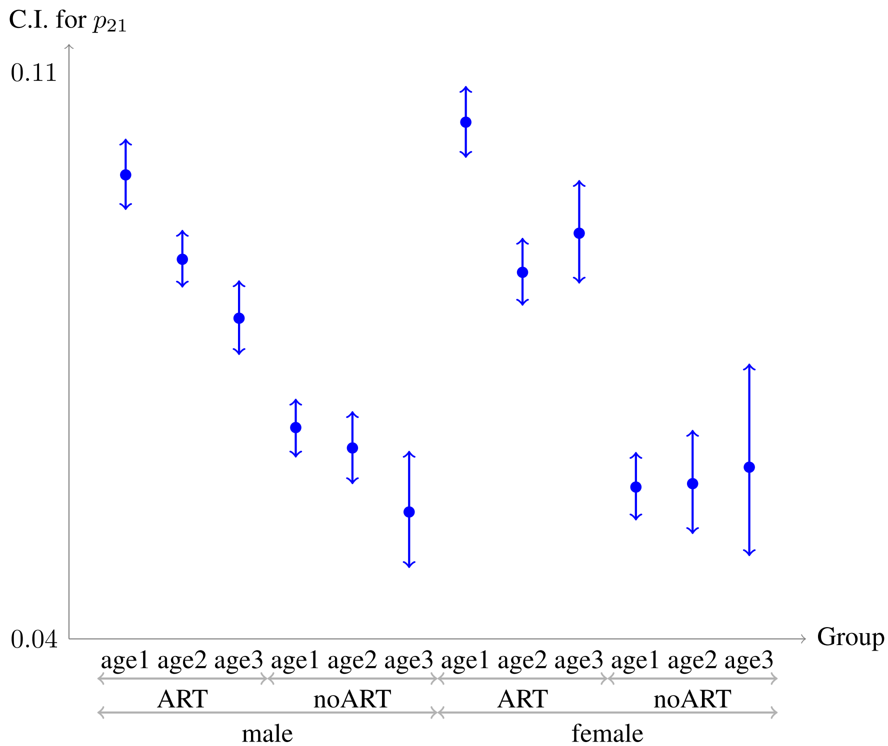

Figure 3.4: p_{21} comparison among different group combinations.

As we can see from Figure 3.4, for the individuals who receive ART treatment, the probability of moving to the higher CD4 count level (healthier) is higher for younger age group; while for the individuals who do not receive ART treatment, the probabilities of moving to the higher CD4 count level (healthier) are not different too much among different age groups. In addition, female and male seems to have similar probabilities of moving to the higher CD4 count level (healthier) than male. And, ART group seems to have a higher probability of moving to the higher CD4 count level (healthier) than no ART group.

3.2 Combined Fit Results and Comparison between Two Methods

For the transition matrices for different subgroups, refer to Appendix C. We also created a table with respect to the value $\log \frac{p_{ij}}{p_{ii}}$ for different subgroups to indicate the effect of the transition probability of getting better, worse, and death as shown in Appendix D.

In order to compare the result of individual fit and combined fit, we draw the scatterplot of the logit of the individual fitted estimated transition probabilities versus the logit of the combined fitted estimated transition probabilities with the $y = x$ line. As we can from Figure 3.5, the probabilities of moving to more unhealthy state and moving to death state are almost same by the two methods. Furthermore, the probabilities of getting healthier are always overestimated by the combined fit method compared to the individual fit method; while the probabilities of staying in the same CD4 level are mostly underestimated by the combined fit method compared to the individual fit method. Notice that for the probabilities of staying in the same CD4 level, some of them are on the $y = x$ line. After carefully analyzed, we found that these points stand for the p_{11} in the transition matrix as shown in Figure 3.6 (staying in state 1).

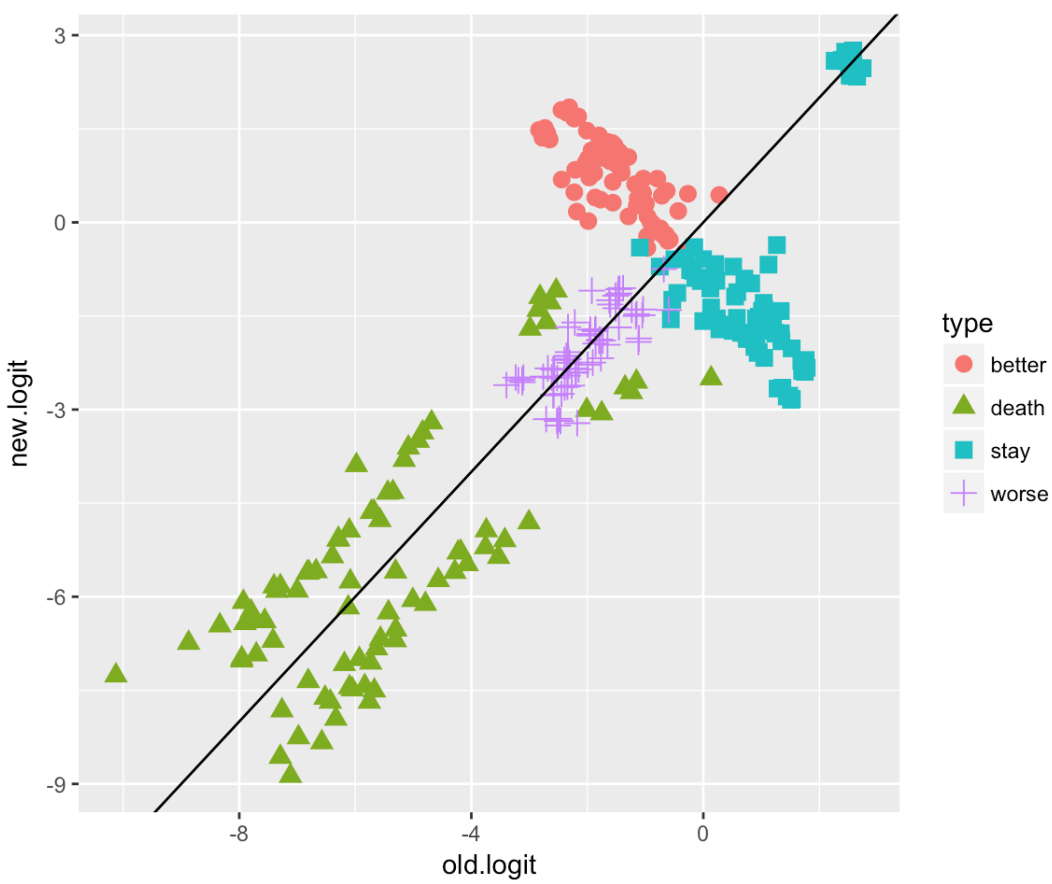

Figure 3.5: Scatterplot for different groups

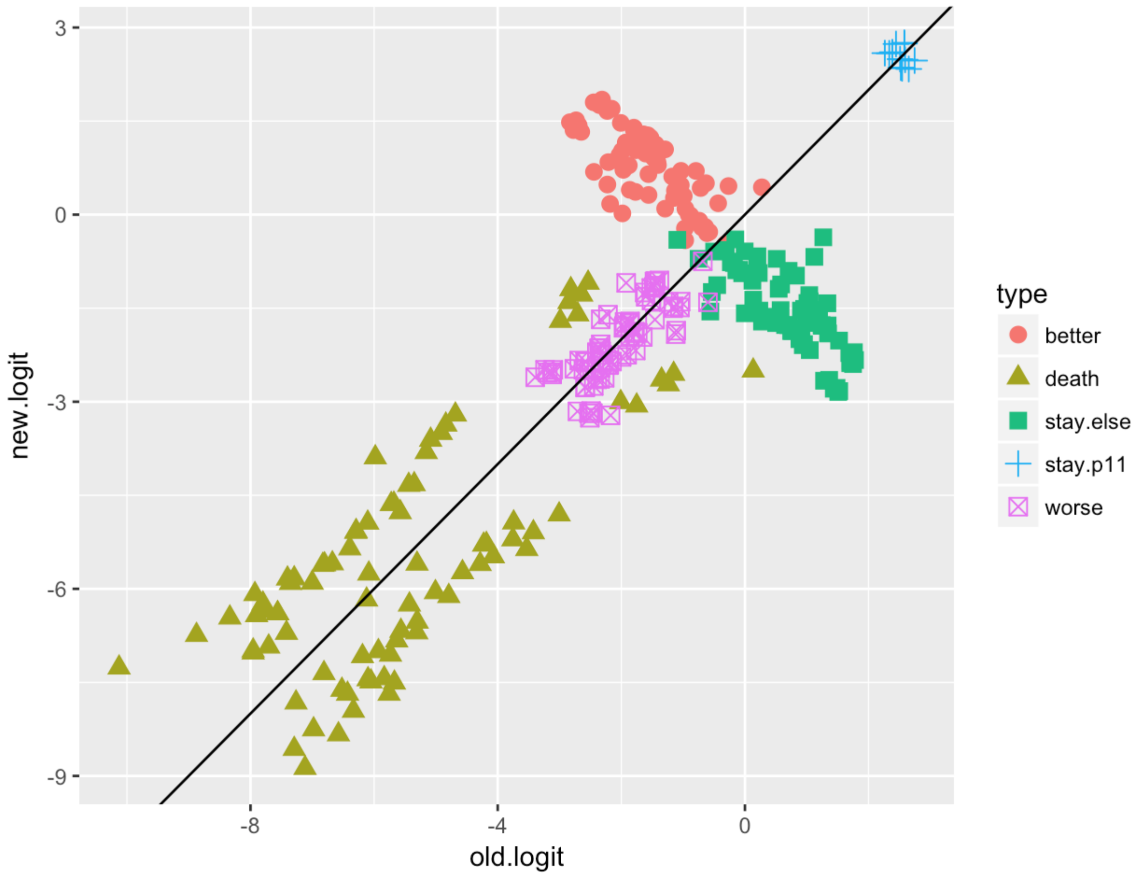

Figure 3.6: Modified scatterplot for different groups

Chapter 4

Discussion

In this project, we model the transition probabilities among CD4 categories and death for a large HIV cohort study in China. We have taken some key covariates into account such as gender, age, treatment status by using two different approaches. In the first approach, we fit the Markov chain model to each combination of those factors separately. Since the sample size is large in each combination, most of the transition probabilities have small standard errors. In the second approach, we jointly model the transition probabilities for all combinations, and assume the effects of gender, age, treatment are additive.

We have assumed that the CD4 counts are accurately measured. However, it is known that the CD4 measurements often have large uncertainties. As we can see from the raw dataset, there were some dramatic increase of CD4 counts in a few months which are unlikely, and should be due to the measurement errors. We should take the uncertainty of CD4 counts into account in the future analysis.

In this project, we use a subset of 30,000 individuals provided by National Center for AIDS/STD Control and Prevention(NCAIDS), Chinese Center for Disease Control and Prevention(China CDC). The results are only for illustration purpose and should never be viewed as the official results for China. We plan to extend this analysis to the full national representative Chinese cohort data.

Chapter 5

Appendix

5.1 Appendix A: Individual Fit Transition Probability Matrix Result

0.91177	0.08763	0.00000	0.00000	0.00000	0.00000	0.00000	0.00061
0.08689	0.80809	0.10434	0.00000	0.00000	0.00000	0.00000	0.00068
0.00000	0.14644	0.71841	0.13389	0.00000	0.00000	0.00000	0.00126
0.00000	0.00000	0.32306	0.48890	0.18637	0.00000	0.00000	0.00167
0.00000	0.00000	0.00000	0.16231	0.79338	0.03959	0.00000	0.00472
0.00000	0.00000	0.00000	0.00000	0.23561	0.65694	0.09830	0.00915
0.00000	0.00000	0.00000	0.00000	0.00000	0.26232	0.68293	0.05476
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.1: Transition probability matrix of male patients with ART treatment in the age range between 35 and 45.

0.91574	0.08391	0.00000	0.00000	0.00000	0.00000	0.00000	0.00036
0.07969	0.81904	0.10017	0.00000	0.00000	0.00000	0.00000	0.00111
0.00000	0.13527	0.74067	0.12299	0.00000	0.00000	0.00000	0.00108
0.00000	0.00000	0.28724	0.53027	0.18026	0.00000	0.00000	0.00223
0.00000	0.00000	0.00000	0.14221	0.82096	0.03250	0.00000	0.00432
0.00000	0.00000	0.00000	0.00000	0.19668	0.70755	0.08847	0.00729
0.00000	0.00000	0.00000	0.00000	0.00000	0.32803	0.61555	0.05642
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.2: Transition probability matrix of male patients with ART treatment in the age range between 45 and 55.

0.93408	0.06415	0.00000	0.00000	0.00000	0.00000	0.00000	0.00177
0.06606	0.84404	0.08843	0.00000	0.00000	0.00000	0.00000	0.00147
0.00000	0.13249	0.72271	0.14216	0.00000	0.00000	0.00000	0.00265
0.00000	0.00000	0.33234	0.42286	0.24043	0.00000	0.00000	0.00437
0.00000	0.00000	0.00000	0.14391	0.76727	0.07858	0.00000	0.01025
0.00000	0.00000	0.00000	0.00000	0.14544	0.64225	0.18915	0.02316
0.00000	0.00000	0.00000	0.00000	0.00000	0.09750	0.75464	0.14785
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.3: Transition probability matrix of male patients without ART treatment in the age range between 15 and 35.

0.92712	0.06996	0.00000	0.00000	0.00000	0.00000	0.00000	0.00291
0.06358	0.84676	0.08643	0.00000	0.00000	0.00000	0.00000	0.00322
0.00000	0.11595	0.73023	0.15001	0.00000	0.00000	0.00000	0.00381
0.00000	0.00000	0.29277	0.44336	0.25892	0.00000	0.00000	0.00495
0.00000	0.00000	0.00000	0.14222	0.76857	0.07419	0.00000	0.01502
0.00000	0.00000	0.00000	0.00000	0.17343	0.63338	0.14603	0.04716
0.00000	0.00000	0.00000	0.00000	0.00000	0.07996	0.69544	0.22460
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.4: Transition probability matrix of male patients without ART treatment in the age range between 35 and 45.

0.92570	0.07114	0.00000	0.00000	0.00000	0.00000	0.00000	0.00315
0.05560	0.85186	0.08898	0.00000	0.00000	0.00000	0.00000	0.00356
0.00000	0.12043	0.71382	0.16083	0.00000	0.00000	0.00000	0.00492
0.00000	0.00000	0.27627	0.47733	0.23817	0.00000	0.00000	0.00823
0.00000	0.00000	0.00000	0.11834	0.79204	0.07524	0.00000	0.01438
0.00000	0.00000	0.00000	0.00000	0.09882	0.74020	0.12938	0.03159
0.00000	0.00000	0.00000	0.00000	0.00000	0.13417	0.62589	0.23994
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.5: Transition probability matrix of male patients without ART treatment in the age range between 45 and 55.

0.92019	0.07977	0.00000	0.00000	0.00000	0.00000	0.00000	0.00004
0.10382	0.79630	0.9953	0.00000	0.00000	0.00000	0.00000	0.00035
0.00000	0.19516	0.66935	0.13489	0.00000	0.00000	0.00000	0.00060
0.00000	0.00000	0.42109	0.37780	0.20047	0.00000	0.00000	0.00064
0.00000	0.00000	0.00000	0.21549	0.74477	0.03789	0.00000	0.00184
0.00000	0.00000	0.00000	0.00000	0.29474	0.36354	0.33599	0.00573
0.00000	0.00000	0.00000	0.00000	0.00000	0.56810	0.36921	0.06269
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.6: Transition probability matrix of female patients with ART treatment in the age range between 15 and 35.

0.92962	0.07024	0.00000	0.00000	0.00000	0.00000	0.00000	0.00014
0.08523	0.81495	0.9959	0.00000	0.00000	0.00000	0.00000	0.00024
0.00000	0.16719	0.70744	0.12485	0.00000	0.00000	0.00000	0.00052
0.00000	0.00000	0.35126	0.45553	0.19103	0.00000	0.00000	0.00219
0.00000	0.00000	0.00000	0.19082	0.76271	0.04306	0.00000	0.00340
0.00000	0.00000	0.00000	0.00000	0.27094	0.55801	0.16490	0.00615
0.00000	0.00000	0.00000	0.00000	0.00000	0.43391	0.49954	0.06655
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.7: Transition probability matrix of female patients with ART treatment in the age range between 35 and 45.

0.92958	0.07007	0.00000	0.00000	0.00000	0.00000	0.00000	0.00035
0.09012	0.82143	0.08805	0.00000	0.00000	0.00000	0.00000	0.00041
0.00000	0.15376	0.72231	0.12357	0.00000	0.00000	0.00000	0.00037
0.00000	0.00000	0.34449	0.46531	0.18793	0.00000	0.00000	0.00227
0.00000	0.00000	0.00000	0.17779	0.77663	0.04233	0.00000	0.00326
0.00000	0.00000	0.00000	0.00000	0.26157	0.56868	0.16723	0.00252
0.00000	0.00000	0.00000	0.00000	0.00000	0.39346	0.53325	0.07329
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.8: Transition probability matrix of female patients with ART treatment in the age range between 45 and 55.

0.93967	0.05952	0.00000	0.00000	0.00000	0.00000	0.00000	0.00081
0.05875	0.85368	0.08664	0.00000	0.00000	0.00000	0.00000	0.00093
0.00000	0.12272	0.73319	0.14338	0.00000	0.00000	0.00000	0.00070
0.00000	0.00000	0.27563	0.46120	0.26112	0.00000	0.00000	0.00205
0.00000	0.00000	0.00000	0.19191	0.73904	0.06241	0.00000	0.00664
0.00000	0.00000	0.00000	0.00000	0.12102	0.66972	0.18078	0.02848
0.00000	0.00000	0.00000	0.00000	0.00000	0.10142	0.78041	0.11816
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.9: Transition probability matrix of female patients without ART treatment in the age range between 15 and 35.

0.93035	0.06826	0.00000	0.00000	0.00000	0.00000	0.00000	0.00139
0.05901	0.85589	0.08349	0.00000	0.00000	0.00000	0.00000	0.00161
0.00000	0.11265	0.74066	0.14434	0.00000	0.00000	0.00000	0.00235
0.00000	0.00000	0.27461	0.49893	0.22536	0.00000	0.00000	0.00110
0.00000	0.00000	0.00000	0.12693	0.79154	0.07660	0.00000	0.00494
0.00000	0.00000	0.00000	0.00000	0.17371	0.55572	0.24769	0.02288
0.00000	0.00000	0.00000	0.00000	0.00000	0.24369	0.55031	0.20600
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.10: Transition probability matrix of female patients without ART treatment in the age range between 35 and 45.

0.92486	0.07446	0.00000	0.00000	0.00000	0.00000	0.00000	0.00068
0.06113	0.84599	0.09064	0.00000	0.00000	0.00000	0.00000	0.00224
0.00000	0.11624	0.74531	0.13502	0.00000	0.00000	0.00000	0.00343
0.00000	0.00000	0.32398	0.32082	0.35475	0.00000	0.00000	0.00045
0.00000	0.00000	0.00000	0.17194	0.71238	0.10208	0.00000	0.01360
0.00000	0.00000	0.00000	0.00000	0.34732	0.38972	0.24618	0.01679
0.00000	0.00000	0.00000	0.00000	0.00000	0.21563	0.25093	0.53344
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.11: Transition probability matrix of female patients without ART treatment in the age range between 45 and 55.

5.2 Appendix B: Transition Probability Confidence Interval Comparison

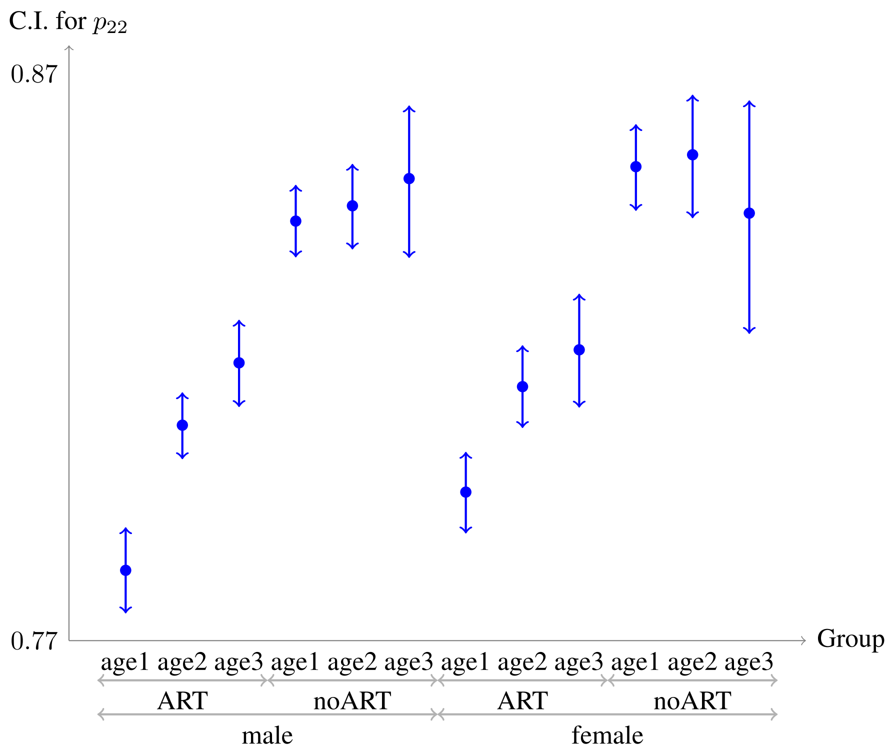

Figure 5.1: p_{22} comparison among different group combinations.

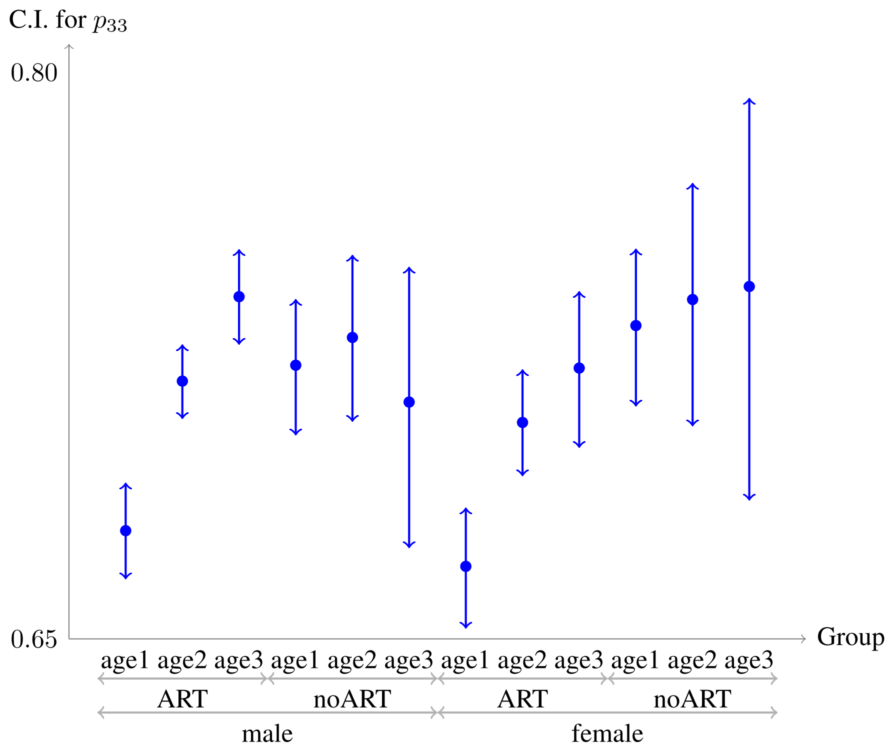

Figure 5.2: p_{33} comparison among different group combinations.

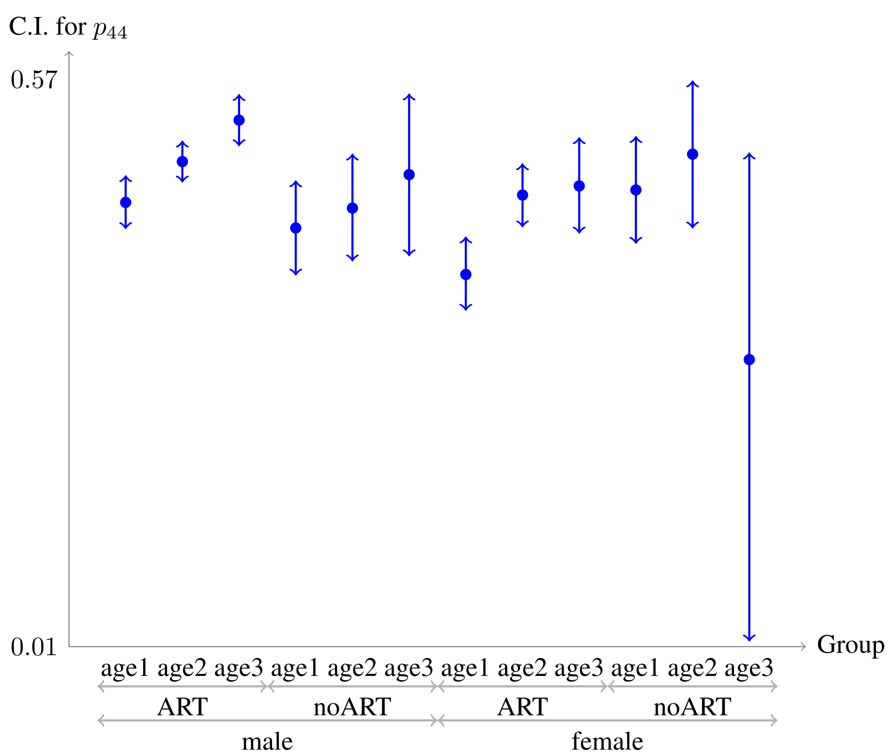

Figure 5.3: p_{44} comparison among different group combinations.

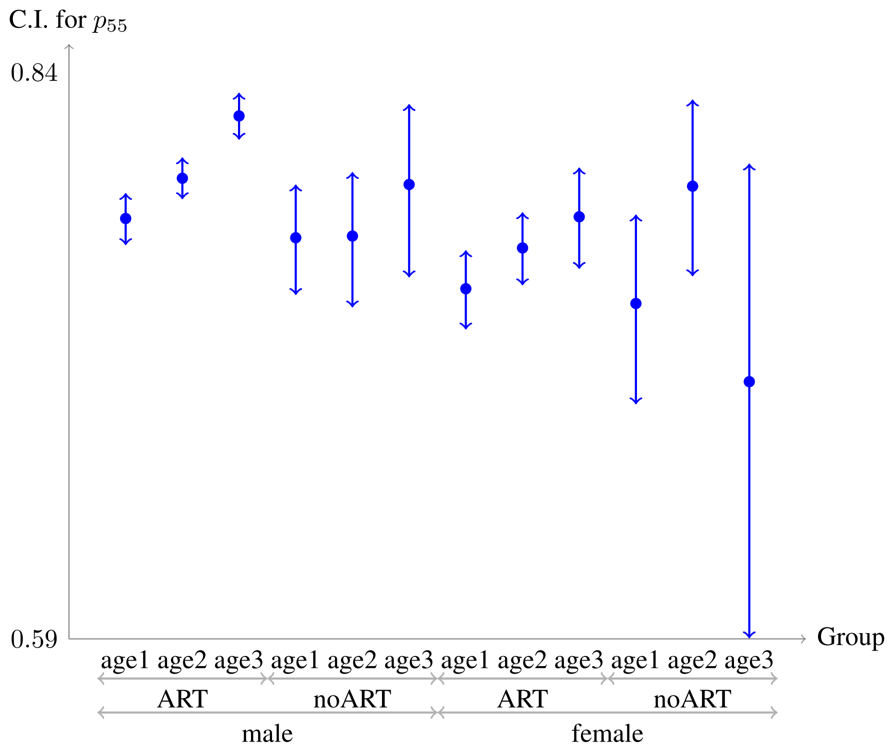

Figure 5.4: p_{55} comparison among different group combinations.

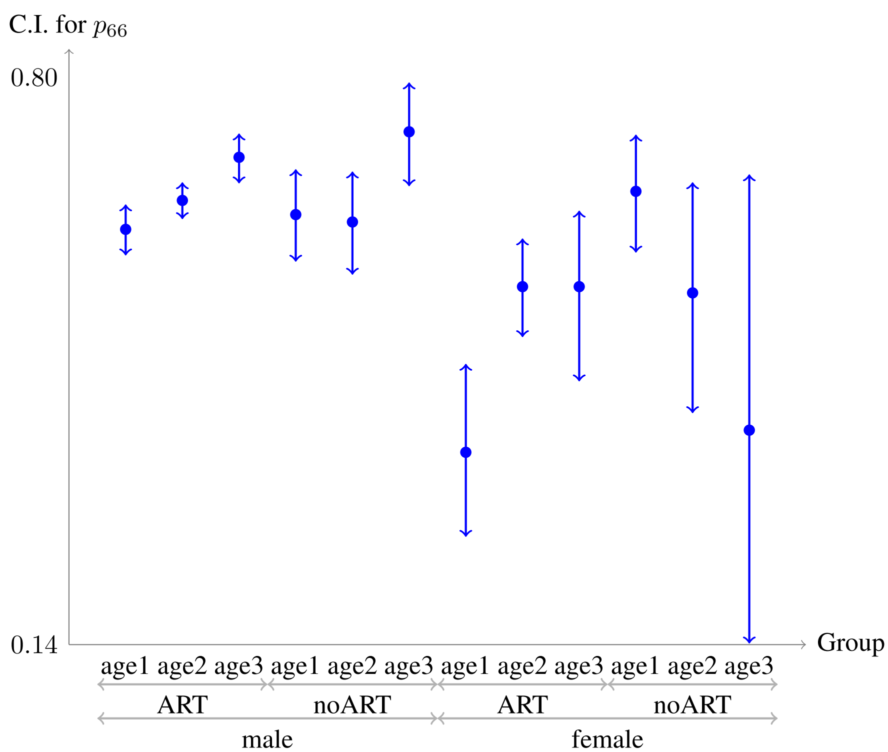

Figure 5.5: p_{66} comparison among different group combinations.

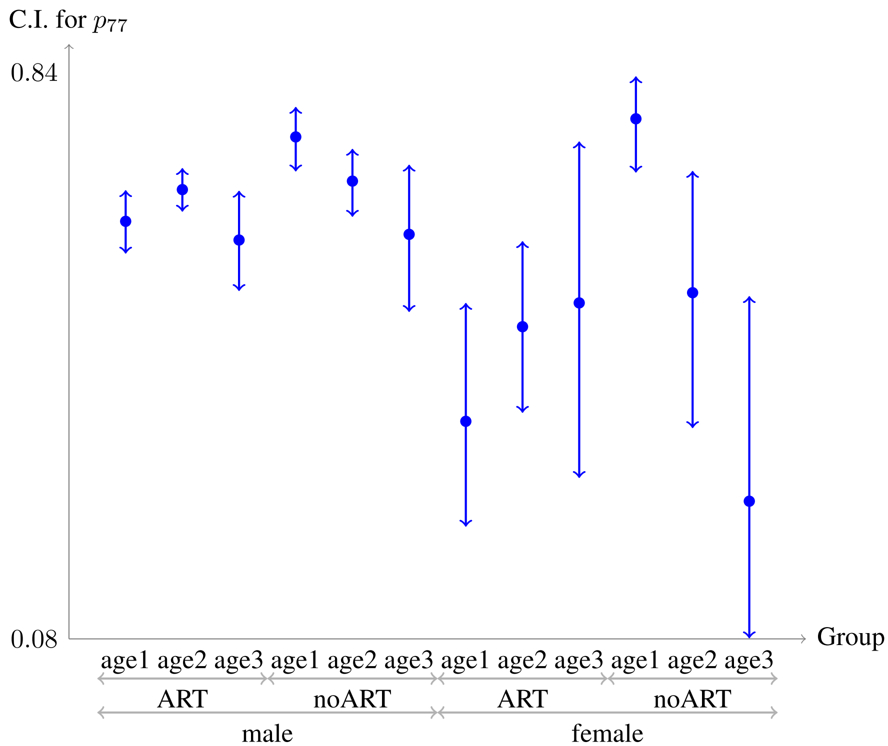

Figure 5.6: p_{77} comparison among different group combinations.

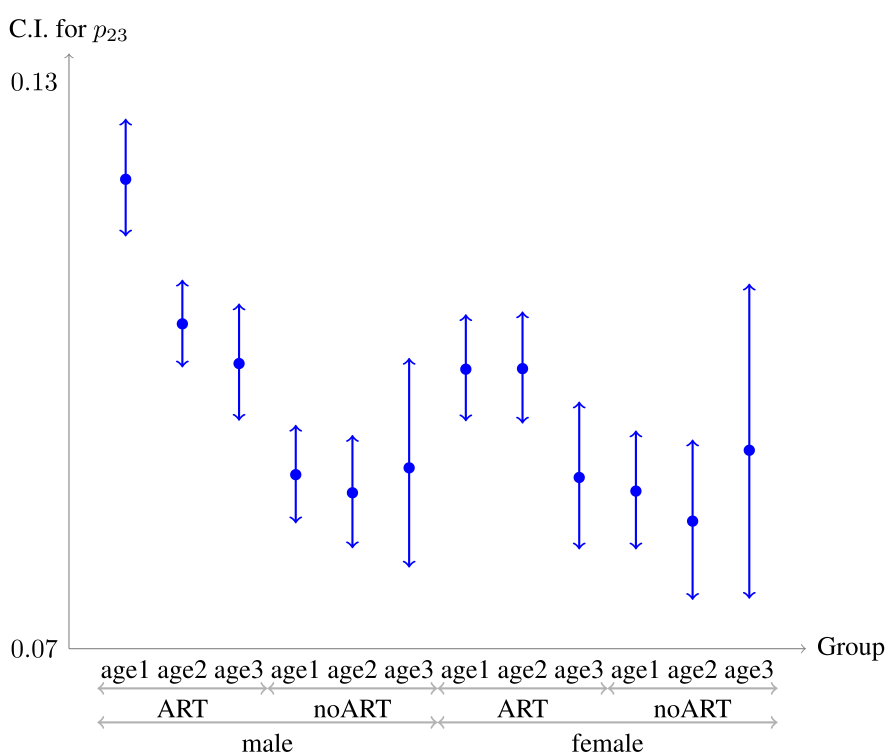

Figure 5.7: p_{23} comparison among different group combinations.

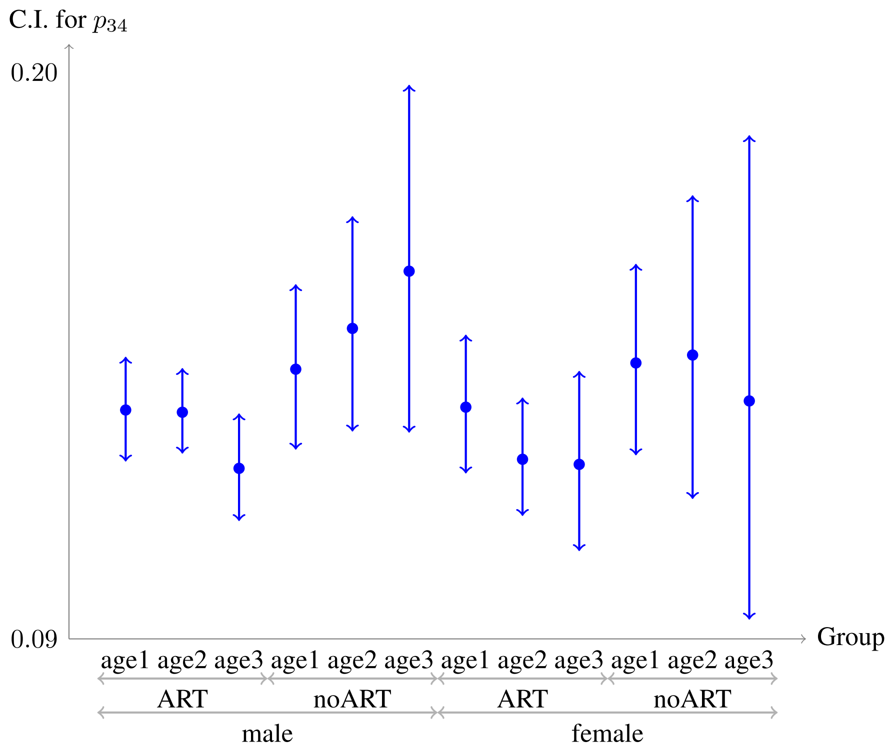

Figure 5.8: p_{34} comparison among different group combinations.

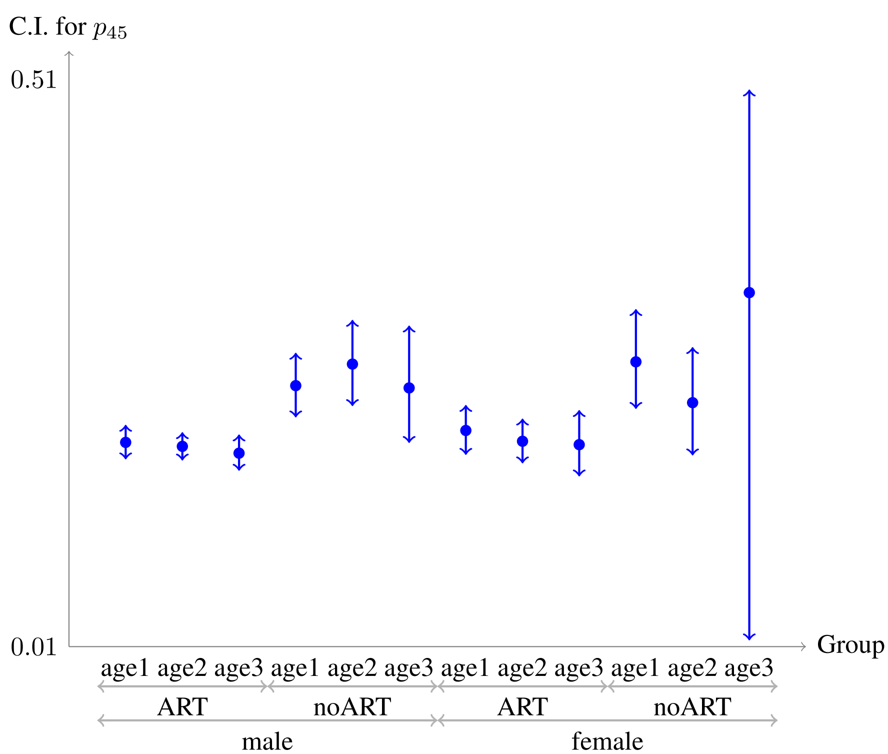

Figure 5.9: p_{45} comparison among different group combinations.

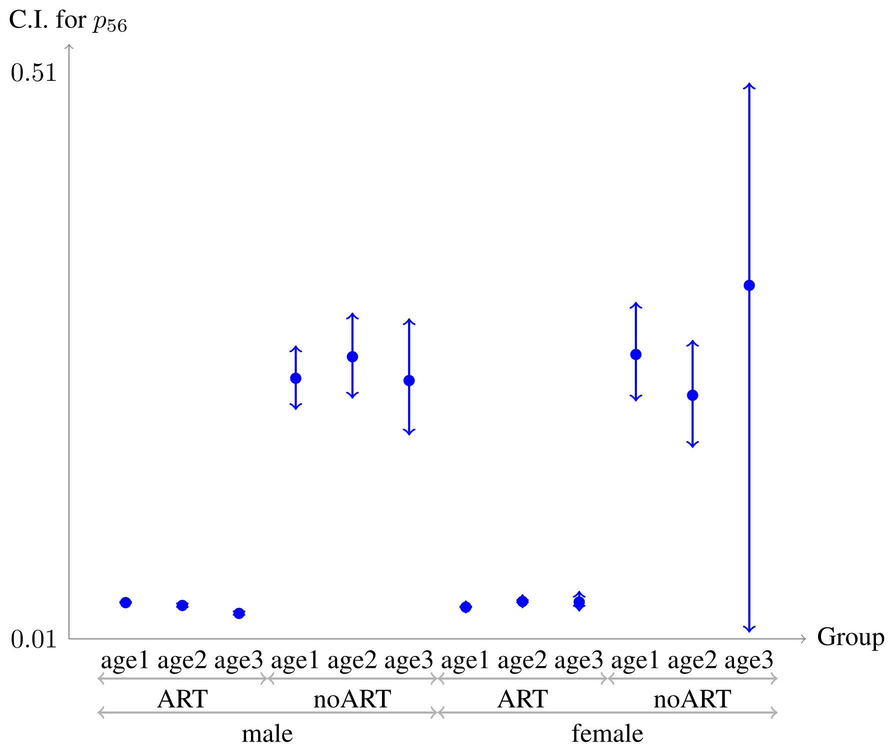

Figure 5.10: p_{56} comparison among different group combinations.

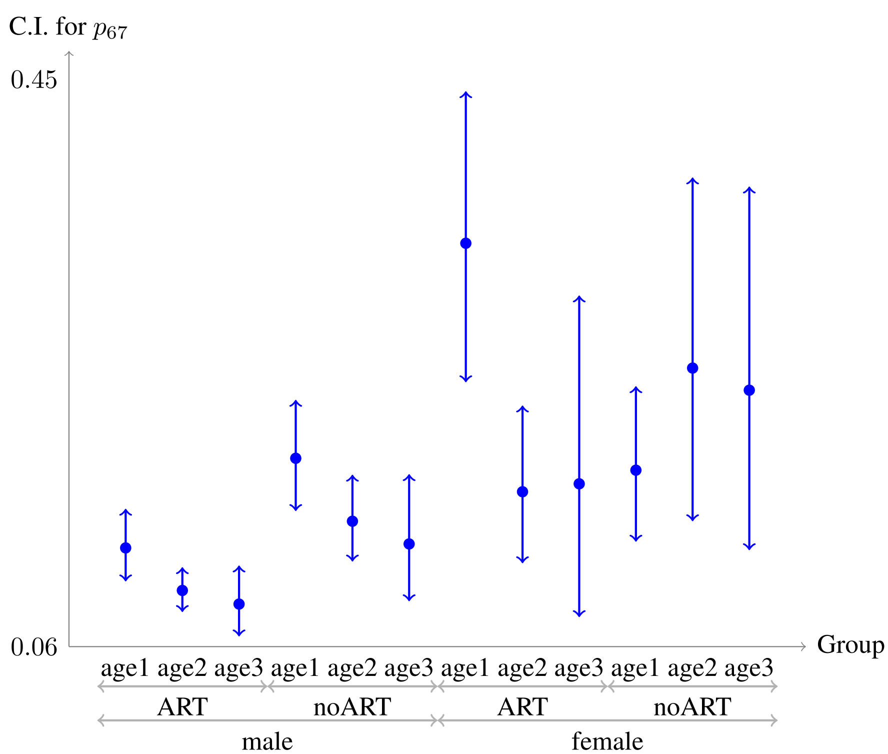

Figure 5.11: p_{67} comparison among different group combinations.

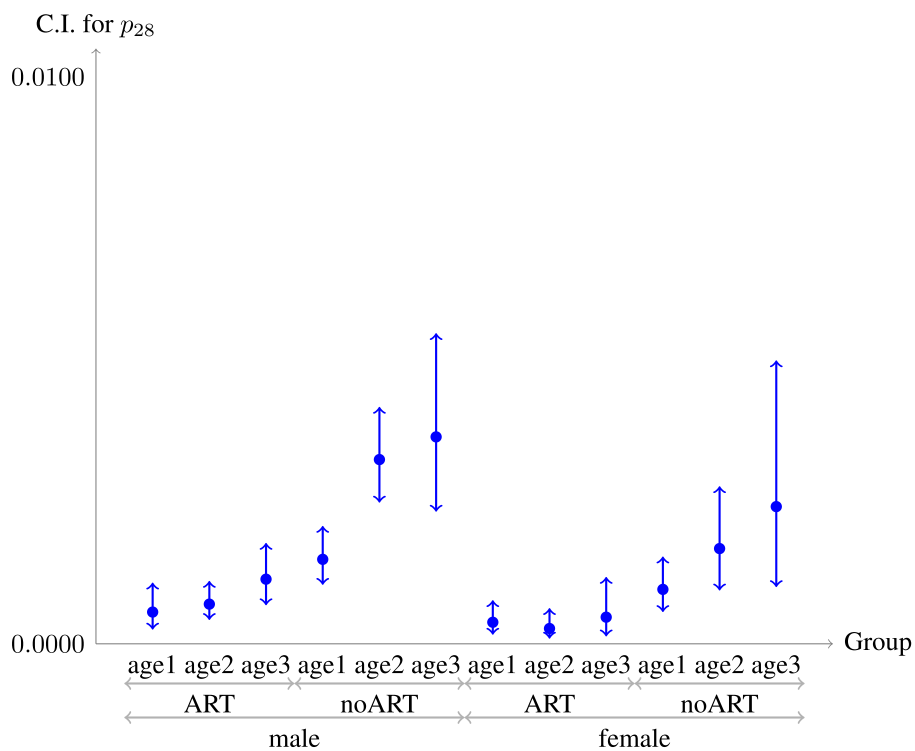

Figure 5.12: p_{28} comparison among different group combinations.

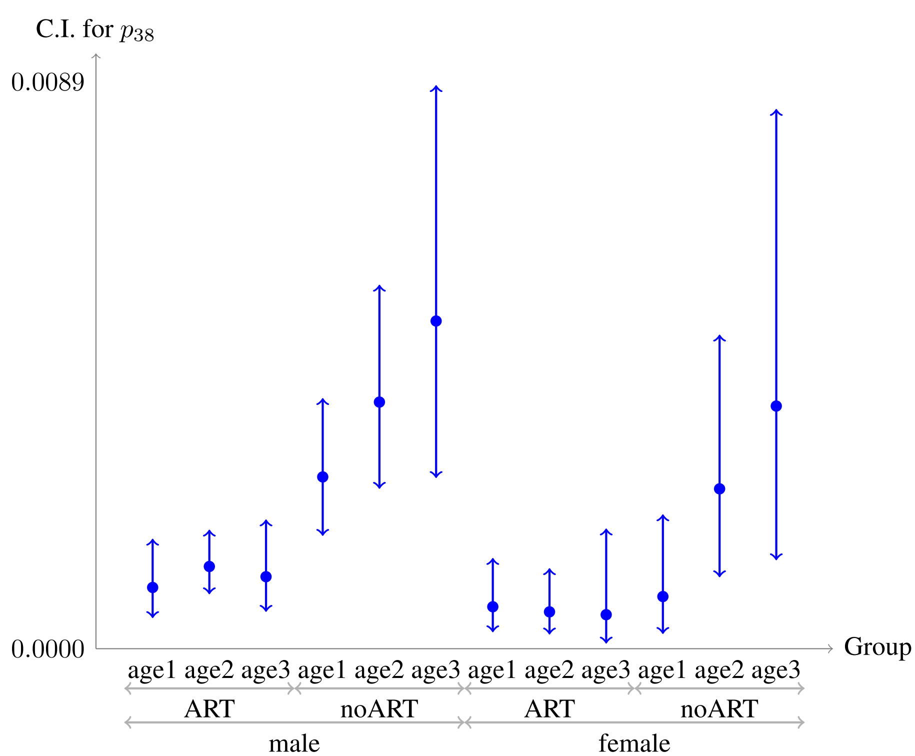

Figure 5.13: p_{38} comparison among different group combinations.

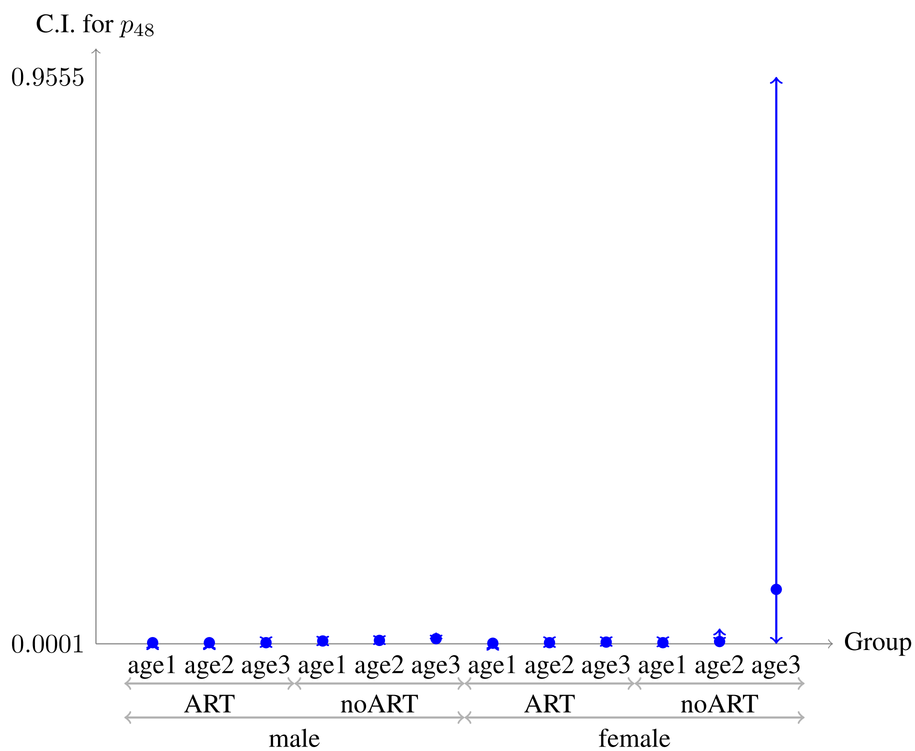

Figure 5.14: p_{48} comparison among different group combinations.

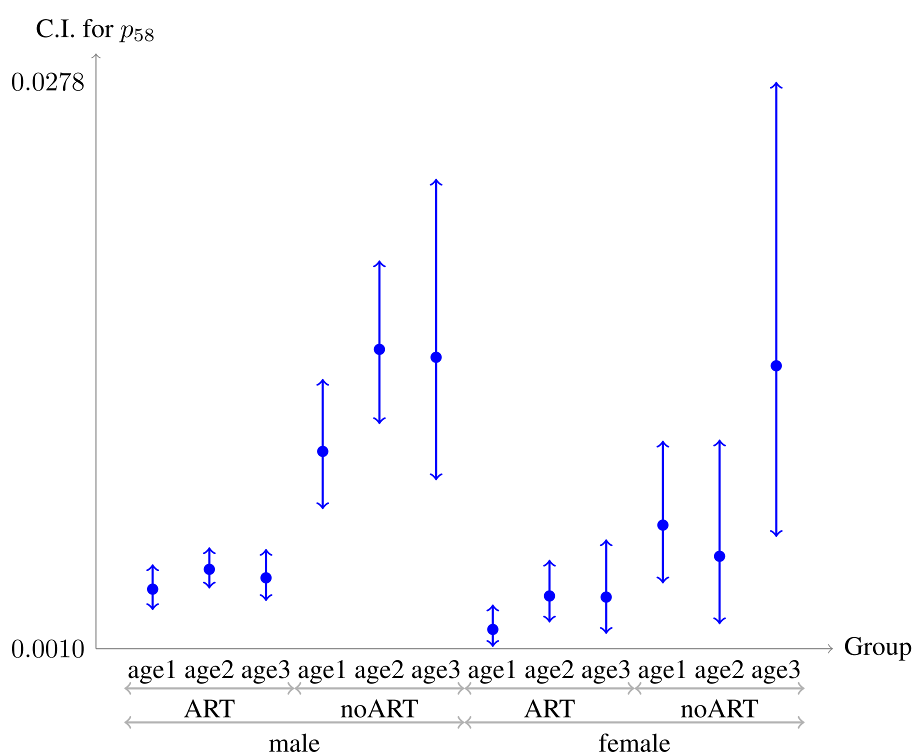

Figure 5.15: p_{58} comparison among different group combinations.

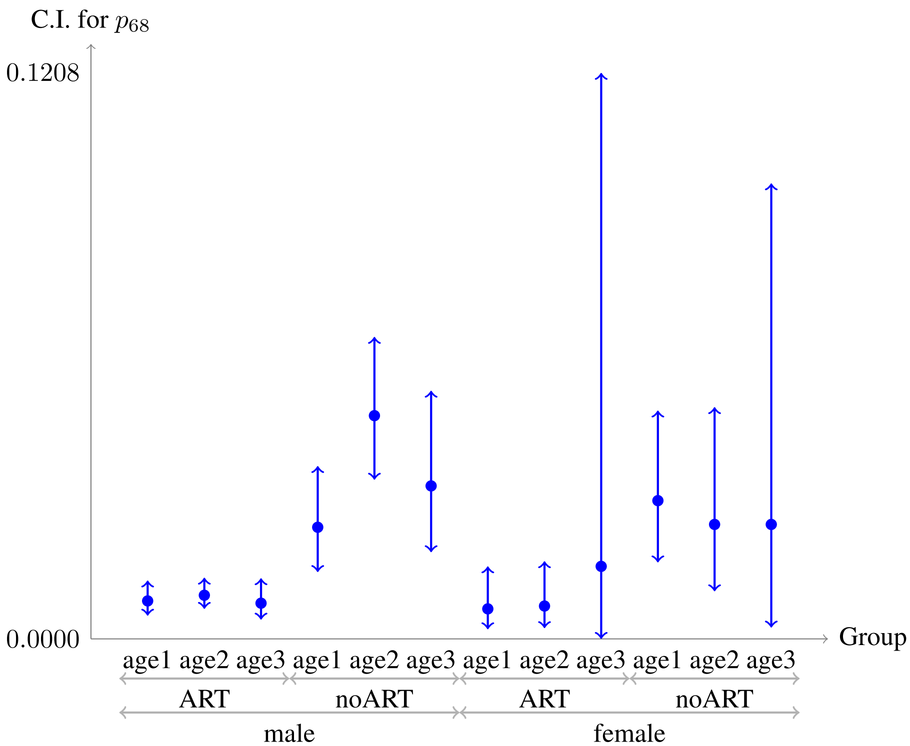

Figure 5.16: p_{68} comparison among different group combinations.

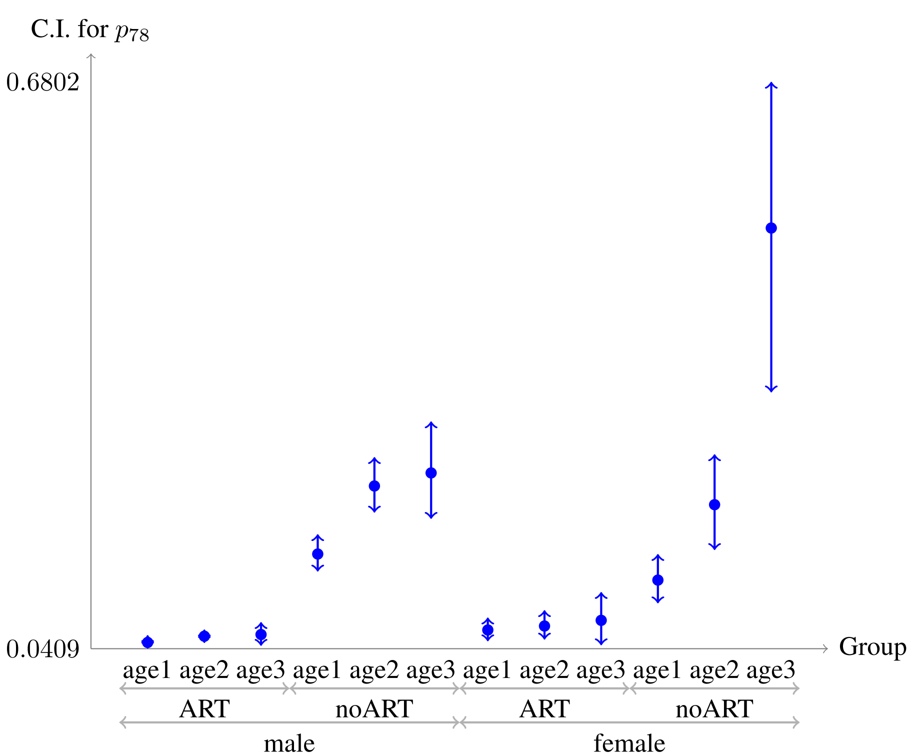

Figure 5.17: p_{78} comparison among different group combinations.

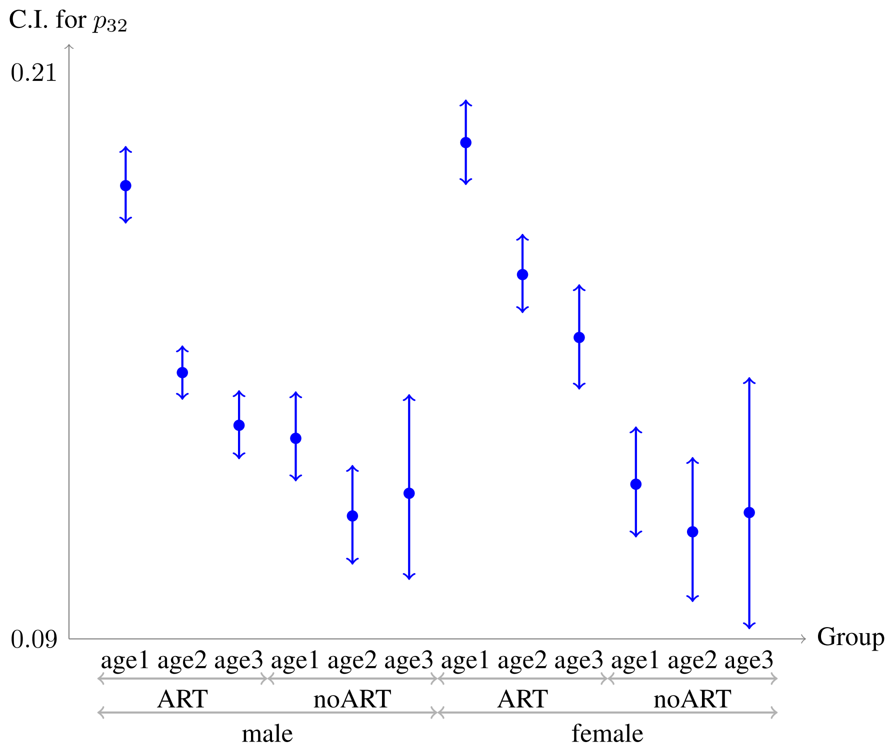

Figure 5.18: p_{32} comparison among different group combinations.

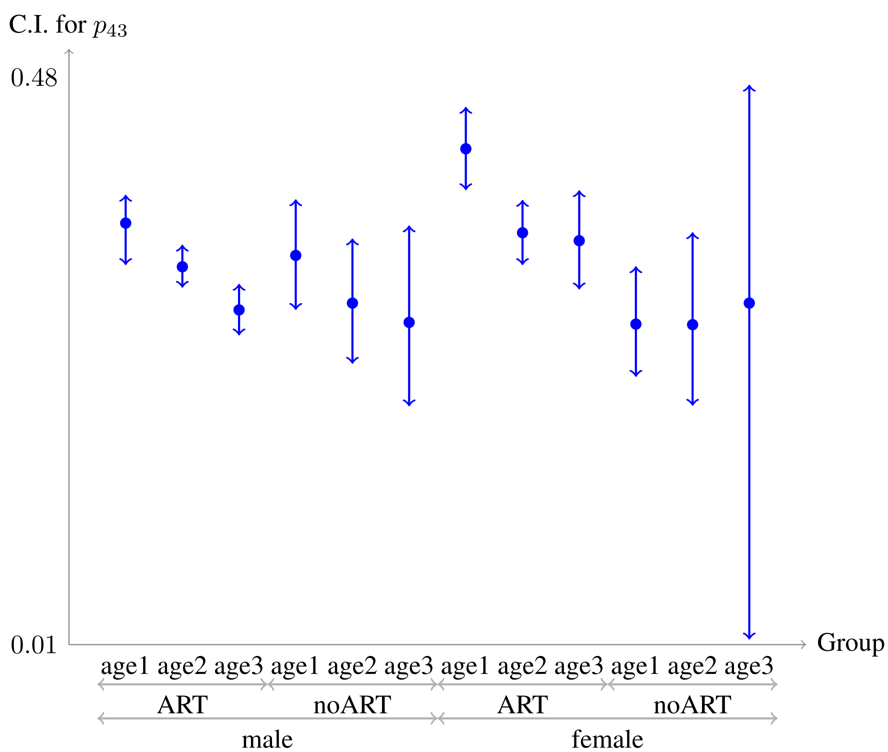

Figure 5.19: p_{43} comparison among different group combinations.

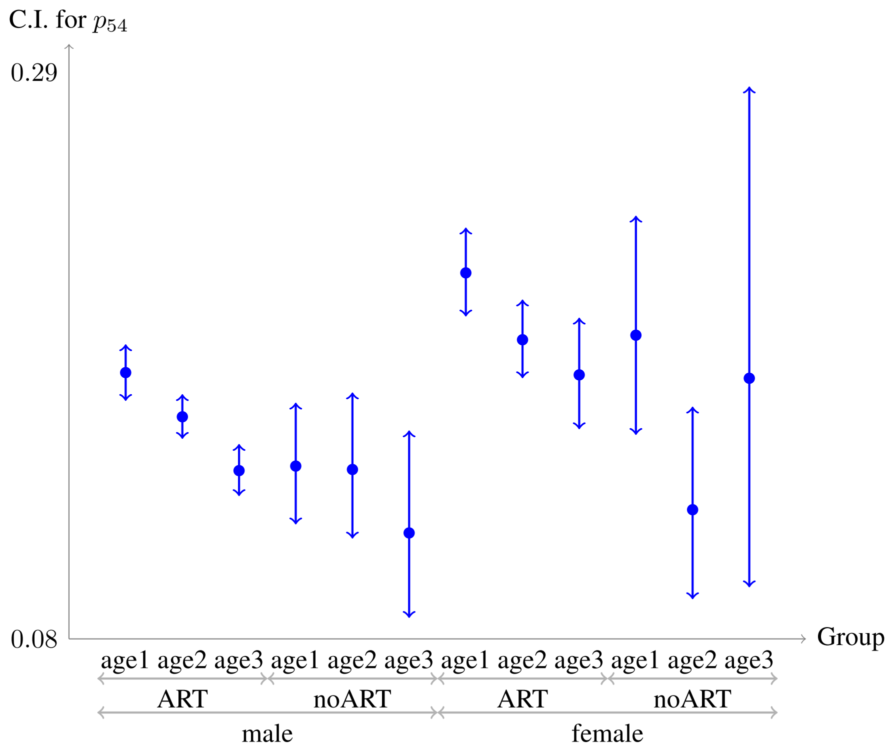

Figure 5.20: p_{54} comparison among different group combinations.

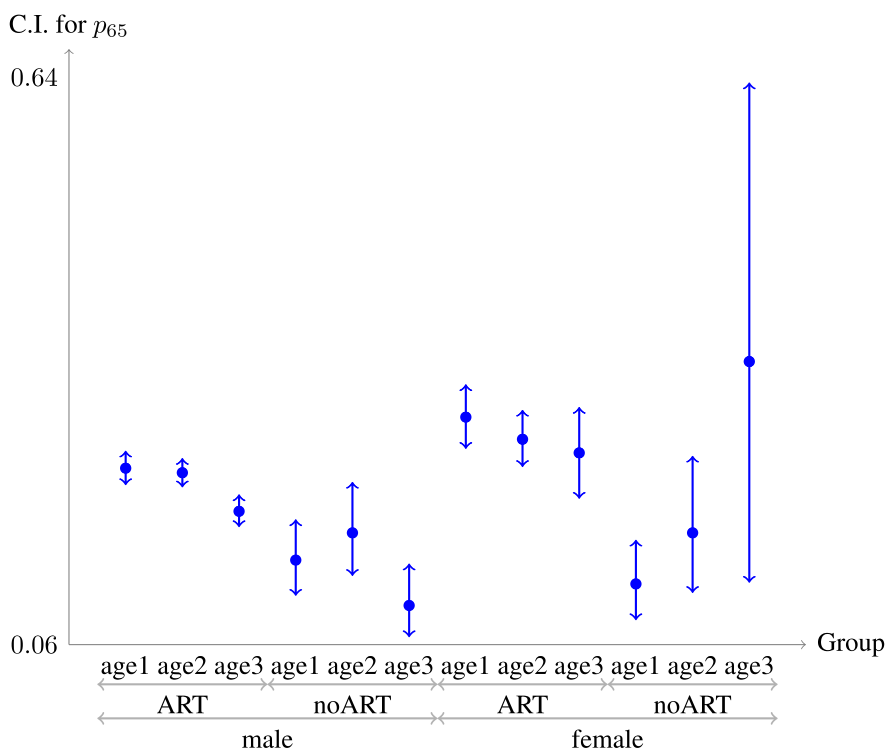

Figure 5.21: p_{65} comparison among different group combinations.

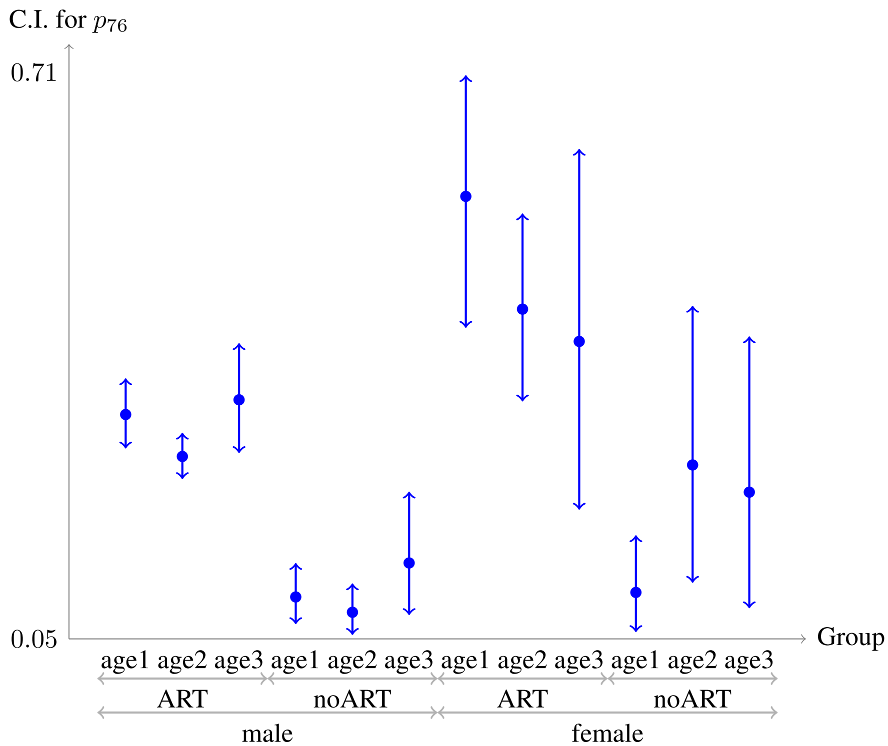

Figure 5.22: p_{76} comparison among different group combinations.

5.3 Appendix C: Combined Fit Transition Matrix Result

0.93020	0.06807	0.00000	0.00000	0.00000	0.00000	0.00000	0.00173
0.84050	0.06518	0.09266	0.00000	0.00000	0.00000	0.00000	0.00167
0.00000	0.71328	0.13387	0.15013	0.00000	0.00000	0.00000	0.00272
0.00000	0.00000	0.43101	0.32016	0.24263	0.00000	0.00000	0.00621
0.00000	0.00000	0.00000	0.77215	0.14535	0.07412	0.00000	0.00838
0.00000	0.00000	0.00000	0.00000	0.56651	0.14945	0.07412	0.03330
0.00000	0.00000	0.00000	0.00000	0.00000	0.66841	0.17786	0.15374
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.12: Transition probability matrix of male patients with ART treatment in the age range between 15 and 35.

0.93000	0.06708	0.00000	0.00000	0.00000	0.00000	0.00000	0.00291
0.85267	0.05779	0.08663	0.00000	0.00000	0.00000	0.00000	0.00292
0.00000	0.73767	0.10979	0.14882	0.00000	0.00000	0.00000	0.00371
0.00000	0.00000	0.47561	0.28083	0.23885	0.00000	0.00000	0.00471
0.00000	0.00000	0.00000	0.78375	0.12947	0.07370	0.00000	0.01307
0.00000	0.00000	0.00000	0.00000	0.64815	0.14558	0.16738	0.03888
0.00000	0.00000	0.00000	0.00000	0.00000	0.66862	0.13419	0.19719
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.13: Transition probability matrix of male patients with ART treatment in the age range between 35 and 45.

0.93116	0.06657	0.00000	0.00000	0.00000	0.00000	0.00000	0.00227
0.85827	0.05503	0.08306	0.00000	0.00000	0.00000	0.00000	0.00365
0.00000	0.75394	0.10229	0.14015	0.00000	0.00000	0.00000	0.00362
0.00000	0.00000	0.49980	0.25726	0.23584	0.00000	0.00000	0.00710
0.00000	0.00000	0.00000	0.80130	0.11716	0.06861	0.00000	0.01293
0.00000	0.00000	0.00000	0.00000	0.68912	0.12442	0.15721	0.02924
0.00000	0.00000	0.00000	0.00000	0.00000	0.60507	0.16292	0.23201
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.14: Transition probability matrix of male patients with ART treatment in the age range between 45 and 55.

0.91177	0.08788	0.00000	0.00000	0.00000	0.00000	0.00000	0.00035
0.79039	0.09801	0.11111	0.00000	0.00000	0.00000	0.00000	0.00049
0.00000	0.68794	0.18072	0.13043	0.00000	0.00000	0.00000	0.00091
0.00000	0.00000	0.44938	0.36193	0.18678	0.00000	0.00000	0.00192
0.00000	0.00000	0.00000	0.77854	0.17841	0.03983	0.00000	0.00322
0.00000	0.00000	0.00000	0.00000	0.59010	0.24617	0.15661	0.00712
0.00000	0.00000	0.00000	0.00000	0.00000	0.61844	0.33686	0.04470
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.15: Transition probability matrix of male patients without ART treatment in the age range between 15 and 35.

0.91270	0.08671	0.00000	0.00000	0.00000	0.00000	0.00000	0.00059
0.80710	0.08747	0.10456	0.00000	0.00000	0.00000	0.00000	0.00086
0.00000	0.71850	0.14967	0.13057	0.00000	0.00000	0.00000	0.00126
0.00000	0.00000	0.49654	0.31789	0.18412	0.00000	0.00000	0.00146
0.00000	0.00000	0.00000	0.79518	0.15992	0.03985	0.00000	0.00506
0.00000	0.00000	0.00000	0.00000	0.65688	0.23331	0.10171	0.00809
0.00000	0.00000	0.00000	0.00000	0.00000	0.66511	0.27324	0.06165
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.16: Transition probability matrix of male patients without ART treatment in the age range between 35 and 45.

0.91352	0.08602	0.00000	0.00000	0.00000	0.00000	0.00000	0.00046
0.81482	0.08354	0.10056	0.00000	0.00000	0.00000	0.00000	0.00108
0.00000	0.73583	0.13973	0.12321	0.00000	0.00000	0.00000	0.00123
0.00000	0.00000	0.52336	0.29209	0.18235	0.00000	0.00000	0.00220
0.00000	0.00000	0.00000	0.81316	0.14473	0.03711	0.00000	0.00500
0.00000	0.00000	0.00000	0.00000	0.69881	0.19951	0.09559	0.00609
0.00000	0.00000	0.00000	0.00000	0.00000	0.59820	0.32971	0.07209
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.17: Transition probability matrix of male patients without ART treatment in the age range between 45 and 55.

0.93911	0.06019	0.00000	0.00000	0.00000	0.00000	0.00000	0.00070
0.84524	0.06600	0.08787	0.00000	0.00000	0.00000	0.00000	0.00089
0.00000	0.70061	0.14534	0.15284	0.00000	0.00000	0.00000	0.00122
0.00000	0.00000	0.38287	0.35614	0.25828	0.00000	0.00000	0.00271
0.00000	0.00000	0.00000	0.74008	0.17708	0.07672	0.00000	0.00613
0.00000	0.00000	0.00000	0.00000	0.48253	0.17454	0.32125	0.02168
0.00000	0.00000	0.00000	0.00000	0.00000	0.60791	0.22431	0.16777
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.18: Transition probability matrix of female patients with ART treatment in the age range between 15 and 35.

0.93947	0.05935	0.00000	0.00000	0.00000	0.00000	0.00000	0.00118
0.85773	0.05853	0.08218	0.00000	0.00000	0.00000	0.00000	0.00156
0.00000	0.72680	0.11956	0.15197	0.00000	0.00000	0.00000	0.00167
0.00000	0.00000	0.42624	0.31516	0.25652	0.00000	0.00000	0.00208
0.00000	0.00000	0.00000	0.75515	0.15857	0.07668	0.00000	0.00960
0.00000	0.00000	0.00000	0.00000	0.57397	0.17677	0.22295	0.02631
0.00000	0.00000	0.00000	0.00000	0.00000	0.61268	0.17051	0.21681
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.19: Transition probability matrix of female patients with ART treatment in the age range between 35 and 45.

0.94021	0.05887	0.00000	0.00000	0.00000	0.00000	0.00000	0.00092
0.86350	0.05574	0.07881	0.00000	0.00000	0.00000	0.00000	0.00195
0.00000	0.74360	0.11151	0.14327	0.00000	0.00000	0.00000	0.00162
0.00000	0.00000	0.45105	0.29074	0.25506	0.00000	0.00000	0.00315
0.00000	0.00000	0.00000	0.77482	0.14400	0.07164	0.00000	0.00954
0.00000	0.00000	0.00000	0.00000	0.61609	0.15252	0.21141	0.01998
0.00000	0.00000	0.00000	0.00000	0.00000	0.54541	0.20365	0.25094
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.20: Transition probability matrix of female patients with ART treatment in the age range between 45 and 55.

0.92203	0.07783	0.00000	0.00000	0.00000	0.00000	0.00000	0.00014
0.79506	0.09927	0.10540	0.00000	0.00000	0.00000	0.00000	0.00026
0.00000	0.67229	0.19520	0.13210	0.00000	0.00000	0.00000	0.00040
0.00000	0.00000	0.39861	0.40202	0.19854	0.00000	0.00000	0.00084
0.00000	0.00000	0.00000	0.74092	0.21581	0.04093	0.00000	0.00234
0.00000	0.00000	0.00000	0.00000	0.50495	0.28883	0.20157	0.00466
0.00000	0.00000	0.00000	0.00000	0.00000	0.54287	0.41005	0.04708
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.21: Transition probability matrix of female patients without ART treatment in the age range between 15 and 35.

0.92297	0.07679	0.00000	0.00000	0.00000	0.00000	0.00000	0.00024
0.81177	0.08858	0.09918	0.00000	0.00000	0.00000	0.00000	0.00046
0.00000	0.70453	0.16221	0.13269	0.00000	0.00000	0.00000	0.00056
0.00000	0.00000	0.44494	0.35671	0.19771	0.00000	0.00000	0.00064
0.00000	0.00000	0.00000	0.76069	0.19446	0.04116	0.00000	0.00369
0.00000	0.00000	0.00000	0.00000	0.57826	0.28162	0.13468	0.00544
0.00000	0.00000	0.00000	0.00000	0.00000	0.59492	0.33892	0.06616
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.22: Transition probability matrix of female patients without ART treatment in the age range between 35 and 45.

0.92364	0.07617	0.00000	0.00000	0.00000	0.00000	0.00000	0.00019
0.81945	0.08459	0.09537	0.00000	0.00000	0.00000	0.00000	0.00058
0.00000	0.72245	0.15163	0.12538	0.00000	0.00000	0.00000	0.00055
0.00000	0.00000	0.47204	0.32990	0.19709	0.00000	0.00000	0.00098
0.00000	0.00000	0.00000	0.78112	0.17672	0.03849	0.00000	0.00367
0.00000	0.00000	0.00000	0.00000	0.62349	0.24408	0.12828	0.00415
0.00000	0.00000	0.00000	0.00000	0.00000	0.52386	0.40040	0.07575
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000

Table 5.23: Transition probability matrix of female patients without ART treatment in the age range between 45 and 55.

5.4 Appendix D: ART Effect Table

		male, age1	male, age2	male, age3	female, age1	female, age2	female, age3
Getting Worse							
$\log \frac{p_{12}}{p_{11}}$	ART	-2.61	-2.63	-2.64	-2.75	-2.76	-2.77
	noART	-2.34	-2.35	-2.36	-2.47	-2.49	-2.50
$\log \frac{p_{23}}{p_{22}}$	ART	0.35	0.40	0.41	0.29	0.34	0.35
	noART	0.13	0.18	0.19	0.06	0.11	0.12
$\log \frac{p_{34}}{p_{33}}$	ART	0.11	0.30	0.31	0.05	0.24	0.25
	noART	-0.33	-0.14	-0.13	-0.39	-0.20	-0.19
$\log \frac{p_{45}}{p_{44}}$	ART	-0.28	-0.16	-0.09	-0.32	-0.21	-0.13
	noART	-0.66	-0.55	-0.47	-0.71	-0.59	-0.52
$\log \frac{p_{56}}{p_{55}}$	ART	-0.67	-0.56	-0.54	-0.84	-0.73	-0.70
	noART	-1.50	-1.39	-1.36	-1.66	-1.55	-1.52
$\log \frac{p_{67}}{p_{66}}$	ART	0.52	0.14	0.23	0.61	0.23	0.33
	noART	-0.45	-0.83	-0.74	-0.36	-0.74	-0.64
Getting Better							
$\log \frac{p_{21}}{p_{22}}$	ART	2.56	2.69	2.75	2.55	2.68	2.74
	noART	2.09	2.22	2.28	2.08	2.22	2.27
$\log \frac{p_{32}}{p_{33}}$	ART	1.67	1.90	2.00	1.57	1.80	1.90
	noART	1.34	1.57	1.66	1.24	1.47	1.56
$\log \frac{p_{43}}{p_{44}}$	ART	0.30	0.53	0.66	0.07	0.30	0.44
	noART	0.22	0.45	0.58	-0.01	0.22	0.36
$\log \frac{p_{54}}{p_{55}}$	ART	1.67	1.80	1.92	1.43	1.56	1.68
	noART	1.47	1.60	1.73	1.23	1.36	1.49
$\log \frac{p_{65}}{p_{66}}$	ART	1.33	1.49	1.71	1.02	1.18	1.40
	noART	0.87	1.04	1.25	0.56	0.72	0.94
$\log \frac{p_{76}}{p_{77}}$	ART	1.32	1.61	1.31	1.00	1.28	0.99
	noART	0.61	0.89	0.60	0.28	0.56	0.27
Death							
$\log \frac{p_{18}}{p_{11}}$	ART	-6.29	-5.77	-6.02	-7.20	-6.68	-6.93
	noART	-7.87	-7.34	-7.59	-8.79	-8.25	-8.49
$\log \frac{p_{28}}{p_{22}}$	ART	-3.66	-2.99	-2.71	-4.31	-3.62	-3.35
	noART	-5.30	-4.62	-4.35	-5.94	-5.26	-4.98
$\log \frac{p_{38}}{p_{33}}$	ART	-3.90	-3.39	-3.34	-4.78	-4.27	-4.23
	noART	-5.29	-4.78	-4.73	-6.19	-5.67	-5.62
$\log \frac{p_{48}}{p_{44}}$	ART	-3.94	-4.09	-3.59	-4.88	-5.02	-4.53
	noART	-5.24	-5.38	-4.89	-6.17	-6.32	-5.82
$\log \frac{p_{58}}{p_{55}}$	ART	-2.85	-2.29	-2.20	-3.36	-2.80	-2.71
	noART	-4.01	-3.45	-3.37	-4.52	-3.96	-3.87
$\log \frac{p_{68}}{p_{66}}$	ART	-1.50	-1.32	-1.45	-2.09	-1.90	-2.03
	noART	-3.54	-3.36	-3.49	-4.13	-3.95	-4.07
$\log \frac{p_{78}}{p_{77}}$	ART	-0.15	0.38	0.35	-0.29	0.24	0.21
	noART	-2.02	-1.49	-1.52	-2.16	-1.63	-1.67

Table 5.24: Table of ART effect

Bibliography

[1] RA Weiss. How does hiv cause aids? *Science*, 260(5112):1273–1279, 1993.

[2] Daniel C. Douek, Mario Roederer, and Richard A. Koup. Emerging concepts in the immunopathogenesis of aids. *Annual Review of Medicine*, 60(1):471–484, 2009. PMID: 18947296.

[3] WHO UNAIDS. 2007 aids epidemic update. Dec 2007.

[4] Jennifer Mabuka, Ruth Nduati, Katherine Odem-Davis, Dylan Peterson, and Julie Overbaugh. Hiv-specific antibodies capable of adcc are common in breastmilk and are associated with reduced risk of transmission in women with high viral loads. *PLOS Pathogens*, 8(6):1–13, 06 2012.

[5] Marcia Claire Hahn, Robert A.; Inhorn. Anthropology and public health : bridging differences in culture and society (2nd ed.). page 449, 2009.

[6] Mead MN. Contaminants in human milk: weighing the risks against the benefits of breastfeeding. *Environmental Health Perspectives*, 116, Oct 2008.

[7] HIV.gov. Preventing mother-to-child transmission of hiv.

[8] What is antiretroviral therapy (art)? July 2014.

[9] Oxford Dictionaries—English. Markov chain—definition of markov chain in us english by oxford dictionaries.

[10] F. W. Scholz. *Maximum Likelihood Estimation*. American Cancer Society, 2006.

[11] Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[12] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, 23(4):550–560, December 1997.

Mr. LI SHANGLUN

+1 814 321 5411

klonterli@gmail.com

The Pennsylvania State University-Park, State College, PA, USA, 16802

Personal Highlights

- Efficient, independent, team-player, proactive, detail-oriented, outgoing, dependable, leadership
- Applied Statistics major and Computational Science minor for current graduate studies
- Sufficient research experiences, especially in mathematics and statistics applications relate to HIV progression and death rates, as well as Neuroscience
- Mastery of Microsoft Office Suite, Matlab, SAS, minitab, R Language, Python, and C/C++
- Native Chinese speaker, fluent in English, ASA (American Statistical Association) Member

Education Experience

The Pennsylvania State University - Park State College, PA, USA

Undergraduate Major(2): Mathematics, Statistics (Schreyer Honors College)

Graduate Major(1) & Minor(1): Applied Statistics, Computational Science *Aug. 2014 – present*

(Six semesters of Dean's List for undergraduate studies)

TA Experience

Penn State Math Department, Undergraduate Studies, Penn State University

- Grader for Math 435 Abstract Algebra, Aug. 2017 - Dec. 2017
- Conduct Review Session for Math 230 Multivariable, Jan. 2017 - Aug. 2017

Extracurricular Activities

- Penn State Chinese Undergrad Association, EP Dpt. Member, Host, Jan. 2017 - May. 2017
- Penn State Chinese Theatre & Movie Society Actor, PR Dpt. Member, Jan. 2015 - Dec. 2016
- Penn State Students Engaging Students Club (SES), Member, Aug. 2015 - Dec. 2015

Awards and Honors

- 2017 Spring, Penn State DataFest Most Creative Award
- 2017 Spring, William B. Forest Scholarship receiver
- 2017 Phi Kappa Phi Honor Society Member
- 2017 Phi Beta Kappa Honor Society Member
- 2016 Fall, Penn State MASS (Math Advanced Study Semester) Program participant

Research Experience

Mr. LI SHANGLUN

+1 814 321 5411

klonterli@gmail.com

The Pennsylvania State University-Park, State College, PA, USA, 16802

HIV Progression Studies, Undergrad Research Project (Honor Thesis), Penn State University

Primary Investigator *Apr. 2017 – Present*

- Research Topic & Contents: Using R language to analyze a HIV cohort contains three million observations by a Markov Model. Using R to locate the efficient channel to compute the estimation of transition probability matrix of the CD4 count according to different group (base on age, sex, or whether have ART treatment or not)
- Advising Professor: Prof. Le Bao

Neuroscience Studies, Undergrad Research Project, Penn State University

Research Assistant *May. 2017 – Present*

- Research Topic & Contents: Involving in the study relates to threshold-linear network, considering it is a valuable tool for understanding the relationship between network connectivity and emergent dynamics; majorly focusing on offer mathematical support to the group, enables the team to engineer networks with multiple dynamic patterns.
- Advising Professor: Prof. Carina Curto

Social Data Analytics Workshop Series, Dpt. Political Science, Penn State University

Participant *Aug. 2017 – Present*

- Workshop function: Working as a supplement for extending my knowledge in R language application and data science; the workshop happens on weekly basis and it will continue to the end of next spring; topics covered so far including Data Structures and Sub-setting in R, Reshaping Data and Text Data with R, Data Input/Output and R packages etc.
- Managing Professor: Prof. Burt L. Monroe