The estimation of transmitted drug resistance mutation strains probability in the treatment of HIV using the Beta-Binomial model

Urban Haankuku
Peter Njuho (njuhopm@unisa.ac.za)
University of South Africa https://orcid.org/0000-0001-8796-220X

Methodology

Keywords: Transmitted drug resistance mutation strains, Prior distribution, likelihood distribution, posterior distribution, Markov Chain Monte Carlo (MCMC), Transition probability matrix

Posted Date: February 12th, 2020

DOI: https://doi.org/10.21203/rs.2.23288/v1

Version of Record: A version of this preprint was published on November 16th, 2020. See the published version at https://doi.org/10.1089/AID.2020.0166.

The estimation of transmitted drug resistance mutation strains probability in the treatment of HIV using the Beta-Binomial model

Urban Haankuku[1] and Peter Njuho[2,*]

[1]Department of Statistics, University of Zambia, Zambia (uhaankuku@gmail.com)

[2]Department of Statistics, Science Campus, University of South Africa, South Africa (njuhopm@unisa.ac.za)

[*]Corresponding Author

Abstract

Background: The human immunodeficiency virus (HIV) is a viral infection that destroys the human immune system resulting in acquired immunodeficiency syndrome (AIDS). The Zambia HIV prevalence rate is among the highest in the sub-Saharan Africa. With no HIV vaccine and cure for HIV/AIDS, the antiretroviral (ARV) therapy that slows the spread of the virus remains the only option. The administration of ARV has challenges of Transmitted Drug Resistance Mutation strains (TDRMs) in the treatment of HIV naïve patients. To address these challenges, we propose use of transition probabilities when prescribing a switch from first-line to second-line or to third-line regimen on ARV drugs combination.

Methods: We formulate a statistical technique for determining an optimal ARV drugs combination. Using data from the Zambia National Health Research Authority, we compute transition probability matrix chart on ARV therapy combination of the first and second baseline regimens using a Beta-Binomial Hierarchical model. We check parameter convergence using history and density plots, and check model adequacy using the deviation information criteria.

Results: The estimates converge to the fitted model as demonstrated by the history and density plots. The transition probability matrices corresponding to, TDF+ETC+NVP, TDF+FTC+EFV, AZT+3TC+NVP, AZT+3TC+EFV, D4T+3TC+NVP and D4T+3TC+EFV provide an upper triangular matrix of the probabilities. We observe a higher probability of remaining in the same regimen state than that of moving to another state. The computed transition probability matrix chart aid in deciding on the most effective combination to prescribe to a patient, in the presence of TDRM test results. Based on transition probabilities TDF+ETC+NVP, TDF+FTC+EFV, AZT+3TC+EFV and D4T+3TC+EFV cannot be prescribed to the patients who test K101 E and 115Y F strains. The available option to the patients' remains either AZT + 3TC + NVP or D4T + 3TC + NVP. Combination AZT + 3TC + NVP with success probability of 0.97027 should be prescribed to the patient.

Conclusion: The transmission probabilities play a major role in aiding the physicians make an informed decision when prescribing an optimal drug combination. All newly diagnosed HIV individuals should have a TDRM test before any prescription on ARV therapy combination is made.

Keywords: Transmitted drug resistance mutation strains, Prior distribution, likelihood distribution, posterior distribution, Markov Chain Monte Carlo (MCMC), Transition probability matrix.

Background

The introduction of highly active antiretroviral therapy (ART) as treatment for HIV infection has greatly improved mortality and morbidity for adults and children living with HIV around the world [1]. According to [1], TDF is superior to AZT in terms of immunologic response and adherence and less frequent emergence of resistance. But how much the other drugs in the

regimens contributed to these findings remain unclear. Treatment failure remains a significant challenge, particularly for highly treatment-experienced patients, despite the success of ART combination in improving clinical outcomes. In addressing the problem, [2] provides guidance in selecting active tolerable drug combinations, that promote a reasonable quality of life, full adherence and a durable treatment response.

According to [3], the number of HIV-infected individuals with prior multiple treatment failures have been increasing with time. The success of ART in the patients with these characteristics is often compromised by the selection of drug-resistant viruses. Maintaining treatment HIV-infected individuals failing virologically and harbouring drug-resistant viruses, might ameliorate immunological deterioration, until new drugs became available [3]. The prescription of efavirenz to NNRTI-naive patients among heavily pre-treated patients is associated with a good virological response, while a high baseline viral load, a large number of protease inhibitor (PI) resistance mutations and nelfinavir prescription at baseline are associated with a poor virological response [4]. Lamivudine (3TC) and emtricitabine (FTC) are guideline choices for combination highly active antiretroviral therapy (HAART), where the former has a shorter intracellular half-life than the latter, which may be more likely to lead to the development of drug resistant HIV variants. No evidence of an increased risk of development of M184V and K65R in patients exposed to 3TC established [5]. According to [6], Emtricitabine and lamivudine showed differing resistance profiles when administered in combination with tenofovir disproxil fumarate and either efavirenz or a ritonavir-boosted PI. However, the prevalence of the M184V/I resistance mutation was significantly lower in patients who received emtricitabine and tenofovir disoproxil fumarate than in those who received lamivudine and tenofovir disoproxil fumarate.

Although HAART has been associated with improved clinical response to treatment, issues of adherence and viral resistance are major challenges limiting its success [7]. Further

studies evaluating the effects and safety of TDF + FTC + EFV as first-line treatment for patients with HIV are needed. The need to understand the occurrence of antiretroviral (ARV)-related adverse events (AEs) among patients receiving second-line antiretroviral therapy (ART) is important in preventing switches to more limited and expensive third-line regimens [8]. In ART-eligible pregnant women with HIV infection, ART is a safe and effective means of providing maternal virological suppression, decreasing infant mortality, and reducing mother to child transmission (MTCT) [9].

According to [10], continuous surveillance of resistance-associated mutations in ARV-naive HIV-1-infected individuals is necessary in order to promptly recognize any significant variation that may affect their clinical management, as well as to plan and optimize the first line regimen and to estimate the evolution of the genetic heterogeneity of HIV-1-resistant strains. Despite considerable uncertainty in the removal probability estimates for resistant strains, patients infected with sensitive strains may be less likely to transmit after diagnosis than patients infected with resistant strains [11]. According to [12], viral load, symptoms, CD4 counts, transmission route, and the duration of ART are associated with HIV-1 DR. Resistance to antiretroviral drugs can complicate the management of HIV-1infection and impair control of its spread. Screening for TDF is recommended to limit its local spread and to optimize HIV-1 therapy [13]. Test-and-treat programs are central to the global control of HIV, but transmitted drug resistance threatens the effectiveness of these programs [14].

A second-line regimen needs to be prepared in the national program to replace a fixed-dose combination of stavudine, lamivudine, and nevirapine which is extensively used as an antiretroviral regimen in developing countries. Despite its affordability, it has high chance of virological failure. Early detection of virological failure may provide more options and better treatment outcomes [15]. Transmitted drug resistance-associated mutations (M) can compromise treatment effectiveness in patients initiating ART and the prevalence can vary in

different clinical settings. Regular monitoring of M should be encouraged, especially with the scale-up of ART at higher CD4 levels [16].

Although the cost of HIVDR testing is high, the cost of second-line ART in developing countries is 4-5 times higher than first-line ART and therapy is life-long. Therefore, it is important to investigate what level of TDF will make routine HIVDR testing prior to starting ART cost-effective [17].

The beta-binomial model has high asymptotic relative efficiency for most of parameter space and offers an attractive and viable alternative to computing the maximum likelihood estimator [18]. Bayesian estimation for probability treatment failure provides complete distributions of means of groups and their differences, standard deviations and their differences, credible intervals for combinations means and their differences, and the data normality. A Bayesian posterior probability distribution allows for the extraction of information of interest such as mean, standard deviation, medians, credible interval, and highest density intervals. We use the distribution in obtaining summary statistics for each drug chain combination.

The challenges of TDRMs in the treatment of HIV naïve patients

The zidovudine and tenofovir cross-resistance testing is rarely available in resource-limited settings. According to [19], it is critical to identify the cross-resistance patterns associated with first-line stavudine failure. Whether patients are switched off of stavudine (d4T) as a result of virological failure or to avoid long-term toxicities, TDF will be more advantageous than AZT for the majority of patients in regions where genotypic resistance testing is not available. Such unavailability restricts the scope of switching to the World Health Organisation (WHO)-recommended standard second-line combinations (SLC) without HIV drug resistance (HIVDR)-testing in routine clinical practice [20]. In addition, first-line ART-failure exhibits high-level NRTI-resistance, with potential lower-efficacy of AZT compared to TDF. The world

health organisation [21] describes transmitted drug resistance mutation strains (TDRMs) as a significant ARV administration challenge, prevalent in the sub-Saharan African region.

The current literature review identifies TDRMs to be a considerable challenge on ARVs administration [22]. Investigation of the genotypic of transmitted drug resistance (TDR) in ART-naïve individuals in Surabaya, Indonesia using sequencing analysis revealed no primary mutations associated with drug resistance to integrase inhibitors were detected [23]. The introduction of two new non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the past 5 years and the identification of novel NNRTI-associated mutations have made it necessary to reassess the extent of phenotypic NNRTI cross-resistance [24].

Methods

Components of the Bayesian methods

We utilised data from HIV naïve patients who had been put on first and second baseline regimens in Zambia. Characteristics of individual patients such as age, gender, CD4 count, prescribed regimen, and the outcome, commencement date on ARV, time of failure, weight, number of patients on each combination and number of failures were collected within a period of 48 weeks. The probabilities of treatment failure and survival time estimates for each combination of ARV therapy were computed using a Beta-Binomial hierarchical model.

The perspective of Bayesian methods combine the likelihood function with the prior distribution through Bayes Theorem to produce the Posterior distribution. We provide inference based on quantitative information obtain using the posterior distribution. A schematic diagram of Bayesian methods demonstrates the connection and flow from data through Baye's rule to posterior distribution (figure 1).

<**Insert Fig 1**>

Construction of a three stage Beta-Binomial Hierarchical model

Let data y_i (i =1,2,…n) be independent and identically distributed, drawn from a Binomial distribution. Consider the likelihood function $p(y|\vartheta)$ and the prior distribution $p(\vartheta)$ which produces posterior distribution $p(\vartheta|y)$. Suppose ϑ_i, a parameter governing the data generating process is exchangeable from a standard population with distribution governed by a hyper-parameter ϕ where ϑ_i and ϕ are random variable parameters. We consider estimating the probability of treatment failure of ARV therapy combinations of first and second baseline regimens using a Beta-Binomial Hierarchical model. Hierarchical models are those with hierarchical structure to the parameters and potentially to the covariates if the model is a regression model.

Three stage hierarchy model follows:

Stage I: Consider the likelihood function, $p(y_i|\vartheta_i, \phi)$, with prior distribution, $p(\vartheta_i, \phi)$ where the likelihood depends on ϕ only through ϑ_i. Using the Bayes' theorem, the stage I, prior distribution has the form $p(y_i|\vartheta_i, \phi) = p(y_i|\vartheta_i)p(\vartheta_i|\phi)p(\phi)$ with ϕ as a hyper parameter with hyper prior distribution $p(\vartheta)$. Thus, the posterior distribution is proportional to the product of prior and the likelihood function.

Stage II: We consider the joint posterior distribution,

$$p(\vartheta, \phi \mid y) = \frac{p(y \mid \vartheta, \phi)p(\vartheta, \phi)}{p(y)} = \frac{p(y \mid \vartheta)p(\vartheta \mid \phi)p(\phi)}{p(y)}\ , \qquad (1)$$

where, $p(\vartheta, \phi|y) \propto p(y|\vartheta)p(\vartheta|\phi)p(\phi)$.

Using Bayes rule, we write

$$p(y) = \frac{p(y \mid \vartheta)}{p(\vartheta)}\ , \qquad (2)$$

which is a conditional probability, to give

$$p(y) = \int_{\vartheta \in \Theta} p(y \mid \vartheta) p(\vartheta) d\vartheta . \quad (3)$$

The use of the hyper prior provides more information leading to more accurate opinions on the behaviour of a parameter.

Stage III: The final stage of the Beta-Binomial Hierarchical model provides the posterior distribution as

$$p(\vartheta, \phi, x \mid y) = \frac{p(y \mid \vartheta) p(\vartheta \mid \phi) p(\phi \mid x) p(x)}{p(y)} . \quad (4)$$

Thus, $p(\vartheta, \phi, x|y) \propto p(y|\vartheta)p(\vartheta|\phi)p(\phi|x)p(x)$.

Consider the probability that a patient switch from the first baseline treatment to be ϑ_i, probability of treatment failure for combination *i*, which is the quantity of interest of the analysis. The complement of this probability is the kernel probability of a patient remaining on the first base regimen. We obtain a full probability model by combining the prior *Beta (a, b)*, the likelihood distribution $p(y_i|\vartheta_i, \phi)$ and the hyper prior. Thus,

$$p(y, \vartheta|a, b) = \prod_{i=1}^{N} \binom{n_i}{y_i} \vartheta_i^{y_i} (1 - \vartheta_i)^{n_i - y_i} \prod_{i=1}^{N} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \vartheta^{\alpha-1} (1 - \vartheta_i)^{b-1} \quad (5)$$

Hence, for each $p(\vartheta_i|y_i, a, b) \propto \vartheta_i^{a+y_i-1}(1-\vartheta_i)^{b+n_i-y_i}$, and $\vartheta_i|y_i \sim Beta(a + y_i, b + n_i - y_i)$. The parameter $\vartheta_i's$,'s has a *Beta* (*a*, *b*) distribution assuming the ϑ_i,'s to be independently and identically distributed. We analyse $Beta(a + y_i, b + n_i - y_i)$ by fixing *a*, *b* and *n*. The shape of the population of HIV patient's distribution requires estimates of the parameters of *a* and *b*. Using the properties of the beta distribution for expected value and variance, the parameters *a* and *b* are obtained where,

$$E(\vartheta) = \frac{a}{a+b} \quad \text{and} \quad Var(\vartheta) = \frac{ab}{(a+b)^2 (a+b-1)} \quad (6)$$

The full model becomes,

$$p(y,\vartheta,a,b)=\prod_{i=1}^{N}\binom{n_i}{y_i}\vartheta_i^{y_i}(1-\vartheta_i)^{n_i-y_i}\prod_{i=1}^{N}\frac{\Gamma(a+b)}{\Gamma a\Gamma b}\vartheta^{a-1}(1-\vartheta)^{b-1}p(a,b) \quad (7)$$

According to [25], 4% of patients initiated with ART develop TDRMs with a standard deviation of 0.1. It assumes a binomial model for the number of patients who experience TDRMs given θ. Hence, a prior distribution for θ from the conjugate family, θ ~ *Beta (a, b)* is selected. We use the Beta-Binomial Hierarchical model in the estimation. The TDRMs probabilities, θ vary because of differences in patients and the socio-economic status among the communities. Using (6) the values for *a* and *b* that correspond to the given values for the mean and standard deviation provide the benefits as a = 0.11 and b = 2.73. In estimating the probability of TDRMs, a binomial distribution model was fitted with a prior distribution of the Beta distribution and a Gamma distribution of hyper prior. WinBugs software was used to calculate the posterior distribution through the Markov Chains Monte Carlo (MCMC) simulations of the Gibbs sampling algorithm. There are two main issues to consider, convergence and efficiency, when using the MCMC algorithm.

- *Convergence:* It is essential to know how quickly the distribution of $\vartheta^{(t)}$ approaches $p(\vartheta|y)$.
- *Efficiency:* Necessary to know how well the distributions of $p(\vartheta|y)$ are estimated from $\vartheta^{(t)}$.

Using the data on treatment failure of the first baseline regimen, an MCMC algorithm was run. The WinBug syntax code (Appendix A1) used to implements the Bayesian model provided full MCMC chains for each parameter. These chains form the basis for estimating their posterior distributions and associated statistics, (i.e., means, medians, standard deviations, and credible intervals). We verified convergence using, history, density, Brooks, Gelman and Rubin plots, cross-correlation matrix and auto-correlation plots. The results indicate that the MC Chains were well mixed and no evidence of drift from the fitted model.

Results

The first baseline regimen results

We checked the convergence of the fitted model using plots and the deviance information criteria (DIC). The plots and DIC indicate a well-mixed MC chains and no evidence of drift from the fitted model. Figure 2 presents Brooks, Gelman and Rubin [26] plots of the parameter theta(bgr) obtained using the WinBUGS software. Brooks-Gelman-Rubin scale reduction factor is one of the convergence. For each of the six theta chains, the start-iteration ranged from 10101 to 17500.

<Insert Fig 2>

The Brooks, Gelman and Rubin (bgr) plots generate multiple chains which start from different locations and assess convergence by comparing within- and between-chain variability. For convergence, the plot concentrate at around one as denoted in red, and green for between chains variability (pooled) and plotting blue (average) for within chains variability. Convergence is finalised with an increased sample size after the burn-in iterations and a sufficient number of stationary samples. Results in figure 2 provide evidence that all parameters converged to a distribution.

Figure 3 presents history plots of beta parameters in table 1. For each parameter the iteration ranged from 10001 to 30000.

<Insert Fig. 3>

Based on results in figure 3, we find evidence of chains converging to the distributions. Figure 4 presents density plots of theta for the first baseline regimen, which are smooth kernel density estimates for each theta. A sample of 40 000 applied for each density plot size.

<Insert Fig. 4>

The density plots of theta for the first baseline regimen is consistent with the summary statistics. All parameters converged according to the auto-correlation statistic or plots and cross-correlation plots. The estimated values from the posterior distributions were used to produce the probability transition matrix for each ARV therapy combination.

Model checking

A Bayesian measure of model complexity and fit is the Deviance Information Criterion (DIC) which we use in model checking. Consider the deviance, defined as - 2 * log (likelihood) where likelihood is $p(y|\vartheta)$ including all the normalising constants comprising of all stochastic nodes given y, and theta as the immediate stochastic parents of y. The output for the DIC tool gives the posterior mean of the deviance denoted as Dbar, a point estimate of the deviance denoted as Dhat obtained by substituting in the posterior means and theta.bar, which is the average of theta parameters. Thus, Dhat = - 2 * log (p(y | theta.bar)). PD is the effective number of parameters, and the pD = Dbar - Dhat. In normal hierarchical models, pD = tr(H), where H is the 'hat' matrix that maps the observed data to their proper values. The model with the smallest DIC best predict a replicated dataset.

Using DIC for the Beta-binomial model, we make comparison of the first baseline regimen. The computation produced the posterior mean, Dbar=25.683, Dhat=19.579, pD=6.105 and DIC=31.788. The DIC for residual analysis produced, Dbar=35.179, Dhat=34.197, pD=0.983 and DIC=36.162. The DIC for model adequacy is 31.788 whereas for residual analysis it is 36.162.

Results of the first baseline regimen on ARV therapy combinations

Table 1 presents summary statistics on data for the first baseline regimen on ARV therapy combinations namely mean, standard deviation (SD), Monte Carlo (MC) error, point estimate

2.5 % percentile, median and point estimate 97.5% percentile, with the start sample at 10001 to 40000, for each of the six nodes.

<Insert Table 1>

The calculated values in table 1 show that MC error < 1 − 5% of posterior SD as the rule of the thumb in Bayesian data analysis have been satisfied. In general, a mean or median of the posterior samples for each parameter of interest, as a point estimate 2.5% and 97.5% percentiles of the posterior samples for each parameter give a 95% credible posterior interval. The interval within which the parameter lies with probability 0.95. The results indicate that the posterior distribution of P, the rate of treatment failure due to TDRMs, is approximately normal (judging from the plots in figures 3 and 4) with $\mu = 0.04124$ and $\sigma = 0.01779$ for theta (1). These numbers are computationally accurate to about ±8.874E-5 (MC error). Consequently, we report $\mu = 0.041$ and $\sigma = 0.018$, with median of 0.03873 and a credible interval of [0.014, 0.083]. Similar interpretation should follow for the rest of the theta nodes.

The second baseline regimen

[27] compared outcomes of second-line ART containing and not containing TDF in cohort studies from Zambia and the Republic of South Africa (RSA). Patients on TDF-containing second-line ART were less likely to develop treatment failure than patients on other regimens, for TDF to be an effective component of second-line ART in southern Africa. Despite Zambia being the first African country to introduce TDF as a component of first-line antiretroviral therapy (ART) on a wide scale, no available literature on the treatment failure of second baseline regimen [28]. Patterns of drug substitutions and regimen switches from stavudine (d4T) and zidovudine (AZT) regimens have been well described but data on TDF are more limited [29]. According to [29], regimen switches and virological suppression were similar for patients exposed to TDF, d4T and AZT, suggesting all regimens were equally effective.

We consider a non-informative prior in determining the prior distribution of the beta-binomial hierarchical model. The prior density combines with the likelihood distribution to define a proper joint probability model, $p(y,\vartheta)$. Proper non-normalised posterior density function as define by Bayesian inference has the form, $p(\vartheta|y) \propto p(y|\vartheta)p(\vartheta)$. Results based on a posterior distribution from a non-informative prior require checking for finite integral and ensuring that it is of practical form. Jeffrey's non-informative prior distribution is based on one to one transformation of its parameters, $\phi = h(\vartheta)$. An equivalent prior density on ϕ to the prior density $p(\vartheta)$, obtained through variable change is of the form

$$p(\phi) = p(\vartheta)\left|\frac{d\vartheta}{d\phi}\right| = p(\vartheta)\left|h'(\vartheta)\right|^{-1} \tag{8}$$

According to Gelman (2009), Jeffrey's principles are that any rule for determining the prior density $p(\vartheta)$, should yield comparable results if applied to the transformed parameters. The prior distribution becomes a critical part of the model specifications when the sample size is small. Consider Jeffrey's non-informative prior density $p(\vartheta) \propto [J(\vartheta)]^{1/2}$, where $J(\vartheta)$ is the Fisher information for ϑ expressed as

$$J(\vartheta) = E\left(\left(\frac{d\log p(y|\vartheta}{d\vartheta}\right)^2|\vartheta\right) = -E\left(\frac{d^2\log p(y|\vartheta)}{d\vartheta^2}|\vartheta\right). \tag{9}$$

Jeffrey's' prior model is invariant to parameterisation as illustrated at $J(\vartheta)$, $\vartheta = h^{-1}(\phi)$. Thus,

$$\begin{aligned} J(\phi) &= -E\left(\frac{d^2\log p(y|\phi)}{d\phi^2}\right) \\ &= -E\left(\frac{d^2\log p(y|\vartheta = h^{-1}(\phi)}{d\vartheta^2}\left|\frac{d\vartheta}{d\phi}\right|^2\right) \\ &= J(\vartheta)\left|\frac{d\vartheta}{d\phi}\right|^2 \end{aligned}$$

Jeffrey's prior density $p(\vartheta) \propto \vartheta^{-1/2}(1-\vartheta)^{-1/2}$ is a Beta (½, ½) density. Express the Bayes-Laplace uniform prior density as $\vartheta \sim \text{Beta}(1,1)$. The prior density, uniform in the exponential family and represented by the distribution, p (logit (ϑ)) ∝ constant, corresponds to the improper Beta (0, 0) density on ϑ. Consider using Jeffrey's non-informative prior distribution of Beta (½, ½). Results for the second baseline regimen obtained using WinBugs at *a = 0.5* and *b=0.5* follow.

Results of the second baseline regimen on ARV combinations

Results from the Brooks, Gelman and Rubin plots, and history and density plots of the parameters theta for the second baseline regimen confirm existence of evidence of convergence to the stipulated distribution. The density plots which are smoothed kernel density estimates for the parameters theta show consistency with the computed summary statistics in table 2. Similarly, the auto-correlation plots and cross-correlation plots confirm the convergence of parameters.

Table 2 presents summary statistics on data for the second baseline regimen on ARV therapy combinations namely mean, standard deviation (SD), MC error, point estimate 2.5 % percentile, median and point estimate 97.5% percentile, with the start sample at 100001 to 200000, for each of the five nodes.

<**Insert Table 2**>

Table 2 results show that the nodes ranged from 1 to 5 with corresponding start sample from 100001 to 200000. The mean value was maximum at node 2 and decreased to 0.1364 at node 5. Results from the plots obtained using the WinBUGS indicate the posterior distribution of P, the rate of treatment failure due to TDRMs, to be approximately normal. We report μ = 0.2275 and σ = 0.1207, with median of 0.2112 and a credible interval of [0.0441, 0.5020]. Similar interpretation should follow for the rest of the theta nodes.

Model checking for the second baseline

We compare the Bayesian model of the second baseline regimen on ARV combination, using DIC for the Beta-binomial model. The deviance information criteria (DIC) value=18.289 corresponding to Dbar=17.335, Dhat=16.382, and pD=0.954 confirms the adequacy of the model. We denote, Dbar = post. Mean of -2logL and Dhat = -2LogL at post. Mean of stochastic nodes.

Treatment of HIV as a unique stochastic process with Markov Chain properties

Treatment of HIV pandemic using ARV qualify as a stochastic process. As a time sequence representing the evolution of some system constituted by a variable whose change is subject to a random variation, the process satisfies the Markov properties. Hence, treatment of HIV using ARV combinations is a Markov process. This process is homogeneous in space because the transition probability depends on the difference between those state values. This chain is not Ergodic because it cannot go to every state it is not irreducible and not periodic. The chain has an absorbing state.

Let P be a k x k matrix with elements $\{p_{i,j} : i, j = 1,2,...,k\}$ of a random process $(y_0, y_1, \dots)$ with finite state space, $S = \{s_1, s_2, \dots, s_k\}$. The process is a Markov chain with transition matrix P if for all n, all $i, j \in \{1,...,k\}$ and all $i_0, \dots, i_{n-1} \epsilon \{1, \dots, k\}$ there is $P(Y_{n+1} = s_j | Y_0 = i_0, Y_1 = i_1, \dots, Y_{n-1} = i_{n-1}, Y_n = i) = P(Y_{n+1} = s_j | Y_n = i) = P_{i,j}$.

It implies that the future depends on the past through the present. The final goal in the use of Markov chains is the property of having a stationary distribution as Y_n approaches the stable distribution as n increases.

The transition probability matrix of patients switching regimen

We produce two matrices, the first being the transition probability matrix Q, the probability of a patient switching regimen from first to second baseline regimen. The second being a kernel probability matrix P, the probability of a patient remaining on first baseline regimen after initiation of ART. Consider the kernel matrix, P with elements $p_{i,j}$, *i=1,2,…,6* and *j=1,2,3* where the i_s are the six possible combinations used in Zambia, and j_s are the three line treatments. Let define a P transition matrix for a system with three states as,

$$\text{P=States} \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} \text{States} & & \\ 1 & 2 & 3 \end{array} \\ \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{2,3} & p_{3,3} \end{pmatrix} \end{array}$$

The Markov chain remains in state 1 with probability $p_{1,1}$ when in state 1. It moves to state 3 with probability $p_{1,3}$ and so on. The $p_{i,1}$ column presents the MAP obtained from the posterior distribution of each combination. The MAP is the highest posterior density for the first baseline treatment, and offers the maximum tolerable combination. We have $p_{i,2}$, as the MAP of second-line therapy and develop the summary statistics for first and second baseline regimen, a probability transition matrix. We consider 6 drug combinations namely TDF+FTC+NVP, TDF+EFT+EFV, AZT+3TC+NVP, AZT+3TC-EFV, DAT+3TC+NVP and D4T+3TC+EFV.

The elements of the P matrix were computed to provide the transition probability matrices for the first baseline regimen given in figure 5.

<Insert Fig. 5>

Each of the transition probability matrices corresponding to the 6 combinations, provides an upper triangular matrix, with the probability of remaining in the same state being higher than that of moving to another state. The probability deceases as the chain state moves from state 1 to 2, and 2 to 3. In all the states, there is no recorded direct move from state 1 to state 3. There is no recorded move from either state 2 or 3, with combination 6. The only

recorded moves were from states 1 to 2 and 2 to 3, with the rest remaining in the same state with high probabilities. The results indicate high probability of remaining in regimen 1. The cost and low advice from doctors regarding change of regimen are some of the reasons the patients remained in the same chain state, especially in state 1. An example is the case of Zambia. Taking TDF + FTC + EFV, which comes in tablet form and is easily provided by the Ministry of Ministry of Health in Zambia.

Consider formulation of TDF+FTC+EFV combination. Each tablet contains:

- 300 mg tenofovir disoproxil fumarate (Brand name, Viread). (TDF) -NRTI
- 200 mg emtricitabine (brand name: Emtriva). (FTC)-NRTI
- 600 mg efavirenz (brand name: Sustiva). (EFV)-NNRTI.

Table 3 presents the TDRMs for NNRTI and NRTI groups of classes of ARV used in Zambia for treatment of HIV.

<**Insert Table 3**>

A patient cannot be given a combination containing TDF and EFV if the transmitted drug resistance mutation (TDRM) strains test shows the presence of K101 E and 115Y F strains in the patient. The first baseline regimen combinations used in Zambia are,

1. TDF + FTC + NVP
2. TDF + FTC + EFV
3. AZT + 3TC + NVP
4. AZT + 3TC + EFV
5. D4T + 3TC + NVP
6. D4T + 3TC + EFV

Based on results in table 3, combinations 1, 2, 4 and 6 cannot be prescribed to the patients who got the indicated strains. An alternative option available to these patients' remains either

AZT + 3TC + NVP or D4T + 3TC + NVP. To choice between the two combinations requires use of the transition probabilities, where the combination with the highest probability of success is prescribed to the patient. Consider for instance results in figure 5 where AZT+3TC+NVP could be preferred because it has success probability of 0.97027, compared to D4T+3TC+NVP with success probability of 0.9625. The transition probabilities play a major role in guiding the physicians make an informed decision in respect of prescribing an optimal drug combination.

Conclusion

Transmitted Drug Resistance Mutation continue to pose a severe challenge in the treatment of HIV in the sub-Saharan Africa. The transition probability matrix of ARV combination aids the physicians make an informed decide on the most effective ARV combination to prescribe in the presence of TDRM test results. Our findings suggest that all newly diagnosed HIV individuals should have a TDRM test before any prescription on ARV combination is made. The physicians are assured of success upon prescribing an optimal ARV combination to HIV patients that is based on transition probabilities.

Abbreviations

AIDS: Acquired Immune Deficiency Syndrome; ARV: Anti-retroviral drugs; CD4: Cluster of Differentiation; HAART: High active antiretroviral therapy; HIV: Human immunodeficiency virus; NNRTI: Non-Nucleoside Reverse Transcriptase Inhibitor; NRTI: Nucleoside Reverse Transcriptase Inhibitor; RNA: Ribonucleic acid; TDRMs: Transmitted Drug Resistance Mutation strains; MCMC: Markov Chain Monte Carlo; AE: Adverse event; MTCT: Mother to child transmission; HIVDR: HIV drug resistance; DIC: Deviance information criteria.

Acknowledgements

The authors wish to thank the University of Zambia and the University of South Africa for sponsoring the work. We are grateful to the Zambia National Health Research Authority for providing access to the data and permission to publish the results. We greatly appreciate the constructive comments from the reviewers.

Authors' contribution

This work was done in collaboration between authors. UNH conceptualized the study and conducted the data analysis. PMN provided guidance in the analysis, conducted literature review and compiled entire manuscript. All authors read and approved the final manuscript.

Funding

University of Zambia contributed in employing UNH while University of South Africa contributed in employing PMN. No direct funds obtained for this work.

Availability of data and material

The data can be made available on request from the corresponding author.

Ethics approval and concept to participate

The permission to use data was obtained from the Zambia National Health Research Authority

Consent for publication

Not applicable

Competing interests

The authors declare that they have no conflicting interests.

Author detail

[1]Department of Statistics, University of Zambia, Zambia (uhaankuku@gmail.com)

[2]Department of Statistics, Science Campus, University of South Africa, South Africa (njuhopm@unisa.ac.za)

Appendix A1: Winbugs Syntax model codes

```
MODEL ARV Combination {#Hyper prior for the ARV combination of failure Rates
a~dgamma(.01,.01)
b~dgamma(.01,.01)
#Prior Distribution of the True failure Rates for (i in 1:k) {
#Prior distribution of ARV Combination i's True Rate
theta[i] ~ dbeta(.11,2.78)
#Likelihood of ARV combination i's Data
y[i] ~ dbin(theta[i],n[i]) }}
DATA list (k=6,
n = c(121, 781, 102, 32, 186, 56),
y = c( 5, 8, 3, 3, 7, 2))
INITIAL VALUES1 list(a=0.11,b=2.73)
```

Reference

[1] Spaulding A, Rutherford GW, Siegfried N. Tenofovir or zidovudine in three-drug combination therapy with one nucleoside reverse transcriptase inhibitor and one non-nucleoside reverse transcriptase inhibitor for initial treatment of HIV infection in

antiretroviral-naïve individuals. Cochrane Database Syst Rev 2010:CD008740. https://doi.org/10.1002/14651858.CD008740.

[2] Boyd MA, Hill AM. Clinical Management of Treatment-Experienced, HIV/AIDS Patients in the Combination Antiretroviral Therapy Era. PharmacoEconomics 2010;28:17–34. https://doi.org/10.2165/11587420-000000000-00000.

[3] Rodes B, García F, Gutierrez C, Martinez-Picado J, Aguilera A, Saumoy M, et al. Impact of drug resistance genotypes on CD4+ counts and plasma viremia in heavily antiretroviral-experienced HIV-infected patients. Journal of Medical Virology 2005;77:23–8. https://doi.org/10.1002/jmv.20395.

[4] Vray M, Meynard J-L, Dalban C, Morand-Joubert L, Clavel F, Brun-Vézinet F, et al. Predictors of the virological response to a change in the antiretroviral treatment regimen in HIV-1-infected patients enrolled in a randomized trial comparing genotyping, phenotyping and standard of care (Narval trial, ANRS 088). Antivir Ther (Lond) 2003;8:427–34.

[5] Bulteel N, Bansi-Matharu L, Churchill D, Dunn D, Bibby D, Hill T, et al. The emergence of drug resistant HIV variants at virological failure of HAART combinations containing efavirenz, tenofovir and lamivudine or emtricitabine within the UK Collaborative HIV Cohort. Journal of Infection 2014;68:77–84. https://doi.org/10.1016/j.jinf.2013.09.005.

[6] Marcelin AG, Charpentier C, Wirden M, Landman R, Valantin MA, Simon A, et al. Resistance profiles of emtricitabine and lamivudine in tenofovir-containing regimens. Journal of Antimicrobial Chemotherapy 2012;67:1475–8. https://doi.org/10.1093/jac/dks047.

[7] Omeje I, Okwundu CI. Effectiveness and safety of first-line fixed-dose TNF + EMC + EFZ for patients with HIV. In: The Cochrane Collaboration, editor. Cochrane Database

of Systematic Reviews, Chichester, UK: John Wiley & Sons, Ltd; 2008. https://doi.org/10.1002/14651858.CD007276.

[8] Onoya D, Hirasen K, van den Berg L, Miot J, Long LC, Fox MP. Adverse Drug Reactions Among Patients Initiating Second-Line Antiretroviral Therapy in South Africa. Drug Safety 2018;41:1343–53. https://doi.org/10.1007/s40264-018-0698-3.

[9] Sturt AS, Dokubo EK, Sint TT. Antiretroviral therapy (ART) for treating HIV infection in ART-eligible pregnant women. Cochrane Database of Systematic Reviews 2010. https://doi.org/10.1002/14651858.CD008440.

[10] Bonura F, Tramuto F, Vitale F, Perna AM, Viviano E, Romano N. Transmission of Drug-Resistant HIV Type 1 Strains in HAART-Naive Patients: A 5-Year Retrospective Study in Sicily, Italy. AIDS Research and Human Retroviruses 2010;26:961–5. https://doi.org/10.1089/aid.2009.0250.

[11] Kühnert D, Kouyos R, Shirreff G, Pečerska J, Scherrer AU, Böni J, et al. Quantifying the fitness cost of HIV-1 drug resistance mutations through phylodynamics. PLOS Pathogens 2018;14:e1006895. https://doi.org/10.1371/journal.ppat.1006895.

[12] Lu X, Zhao H, Zhang Y, Wang W, Zhao C, Li Y, et al. HIV-1 drug-resistant mutations and related risk factors among HIV-1-positive individuals experiencing treatment failure in Hebei Province, China. AIDS Research and Therapy 2017;14:4. https://doi.org/10.1186/s12981-017-0133-3.

[13] Sallam M, Şahin GÖ, Indriðason H, Esbjörnsson J, Löve A, Widell A, et al. Decreasing prevalence of transmitted drug resistance among ART-naive HIV-1-infected patients in Iceland, 1996–2012. Infection Ecology & Epidemiology 2017;7:1328964. https://doi.org/10.1080/20008686.2017.1328964.

[14] Wertheim JO, Oster AM, Johnson JA, Switzer WM, Saduvala N, Hernandez AL, et al. Transmission fitness of drug-resistant HIV revealed in a surveillance system transmission network. Virus Evolution 2017;3. https://doi.org/10.1093/ve/vex008.

[15] Sungkanuparph S, Manosuthi W, Kiertiburanakul S, Piyavong B, Chumpathat N, Chantratita W. Options for a second-line antiretroviral regimen for HIV type 1-infected patients whose initial regimen of a fixed-dose combination of stavudine, lamivudine, and nevirapine fails. Clin Infect Dis 2007;44:447–52. https://doi.org/10.1086/510745.

[16] Jiamsakul A, Sirivichayakul S, Ditangco R, Wong K-H, Li PC, Praparattanapan J, et al. Transmitted drug resistance in recently infected HIV-positive Individuals from four urban locations across Asia (2007–2010) – TASER-S. AIDS Research and Therapy 2015;12:3. https://doi.org/10.1186/s12981-015-0043-1.

[17] Phanuphak P, Sirivichayakul S, Jiamsakul A, Sungkanuparph S, Kumarasamy N, Lee MP, et al. Transmitted Drug Resistance and Antiretroviral Treatment Outcomes in Non-Subtype B HIV-1–Infected Patients in South East Asia: JAIDS Journal of Acquired Immune Deficiency Syndromes 2014;66:74–9. https://doi.org/10.1097/QAI.0000000000000108.

[18] Tripathi RC, Gupta RC, Gurland J. Estimation of parameters in the beta binomial model. Annals of the Institute of Statistical Mathematics 1994;46:317–31. https://doi.org/10.1007/BF01720588.

[19] Tang MW, Rhee S-Y, Bertagnolio S, Ford N, Holmes S, Sigaloff KC, et al. Nucleoside Reverse Transcriptase Inhibitor Resistance Mutations Associated with First-Line Stavudine-Containing Antiretroviral Therapy: Programmatic Implications for Countries Phasing Out Stavudine. J Infect Dis 2013;207:S70–7. https://doi.org/10.1093/infdis/jit114.

[20] Takou D, Fokam J, Teto G, Santoro M-M, Ceccherini-Silberstein F, Nanfack AJ, et al. HIV-1 drug resistance testing is essential for heavily-treated patients switching from first- to second-line regimens in resource-limited settings: evidence from routine clinical practice in Cameroon. BMC Infect Dis 2019;19:246. https://doi.org/10.1186/s12879-019-3871-0.

[21] WHO | World Health Statistics 2009. WHO n.d. https://www.who.int/whosis/whostat/2009/en/ (accessed January 30, 2020).

[22] Hardon A, Davey S, Gerrits T, World Health Organization, Universiteit van Amsterdam, Koninklijk Instituut voor de Tropen, editors. From access to adherence: the challenges of antiretroviral treatment: studies from Botswana, Tanzania and Uganda 2006. Geneva: World Health Organization; 2006.

[23] Kotaki T, Khairunisa SQ, Witaningrum AM, M MQY, Sukartiningrum SD, Diansyah MN, et al. HIV-1 transmitted drug resistance mutations among antiretroviral therapy-Naïve individuals in Surabaya, Indonesia. AIDS Res Ther 2015;12. https://doi.org/10.1186/s12981-015-0046-y.

[24] Melikian GL, Rhee S-Y, Varghese V, Porter D, White K, Taylor J, et al. Non-nucleoside reverse transcriptase inhibitor (NNRTI) cross-resistance: implications for preclinical evaluation of novel NNRTIs and clinical genotypic resistance testing. J Antimicrob Chemother 2014;69:12–20. https://doi.org/10.1093/jac/dkt316.

[25] Seu L, Mulenga LB, Siwingwa M, Sikazwe I, Lambwe N, Guffey MB, et al. Characterization of HIV drug resistance mutations among patients failing first-line antiretroviral therapy from a tertiary referral center in Lusaka, Zambia: Sequence Analysis of HIV DRMs in Lusaka, Zambia. Journal of Medical Virology 2015;87:1149–57. https://doi.org/10.1002/jmv.24162.

575 [26] Brooks SP, Gelman A. General Methods for Monitoring Convergence of Iterative
576 Simulations. Journal of Computational and Graphical Statistics 1998;7:434–55.
577 https://doi.org/10.1080/10618600.1998.10474787.
578 [27] Wandeler G, Keiser O, Mulenga L, Hoffmann CJ, Wood R, Chaweza T, et al. Tenofovir
579 in second-line ART in Zambia and South Africa: Collaborative analysis of cohort
580 studies. J Acquir Immune Defic Syndr 2012;61:41–8.
581 https://doi.org/10.1097/QAI.0b013e3182632540.
582 [28] Chi BH, Mwango A, Giganti M, Mulenga LB, Tambatamba-Chapula B, Reid SE, et al.
583 Early clinical and programmatic outcomes with tenofovir-based antiretroviral therapy in
584 Zambia. J Acquir Immune Defic Syndr 2010;54:63–70.
585 https://doi.org/10.1097/QAI.0b013e3181c6c65c.
586 [29] Njuguna C, Orrell C, Kaplan R, Bekker L-G, Wood R, Lawn SD. Rates of switching
587 antiretroviral drugs in a primary care service in South Africa before and after
588 introduction of tenofovir. PLoS ONE 2013;8:e63596.
589 https://doi.org/10.1371/journal.pone.0063596.
590 [30] Gelman A. Bayes, Jeffreys, Prior Distributions and the Philosophy of Statistics.
591 Statistical Science 2009;24:176–8. https://doi.org/10.1214/09-STS284D.
592
593
594

Figures

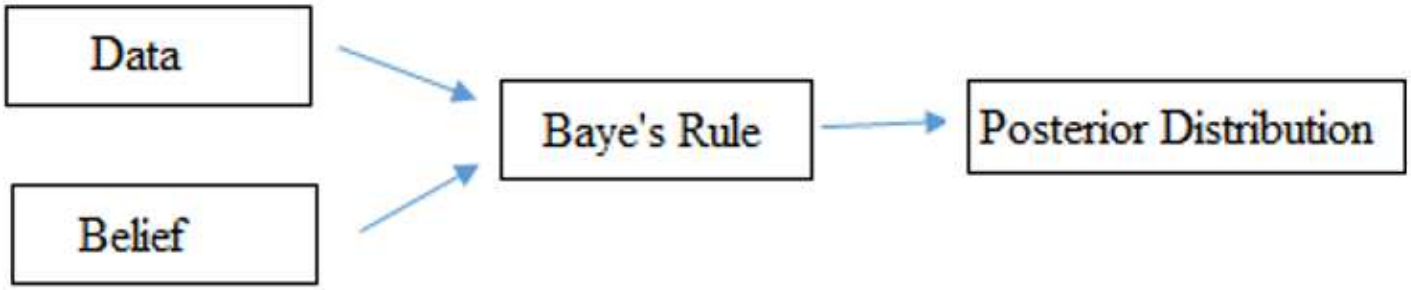

Figure 1

Schematic diagram of Bayesian methods

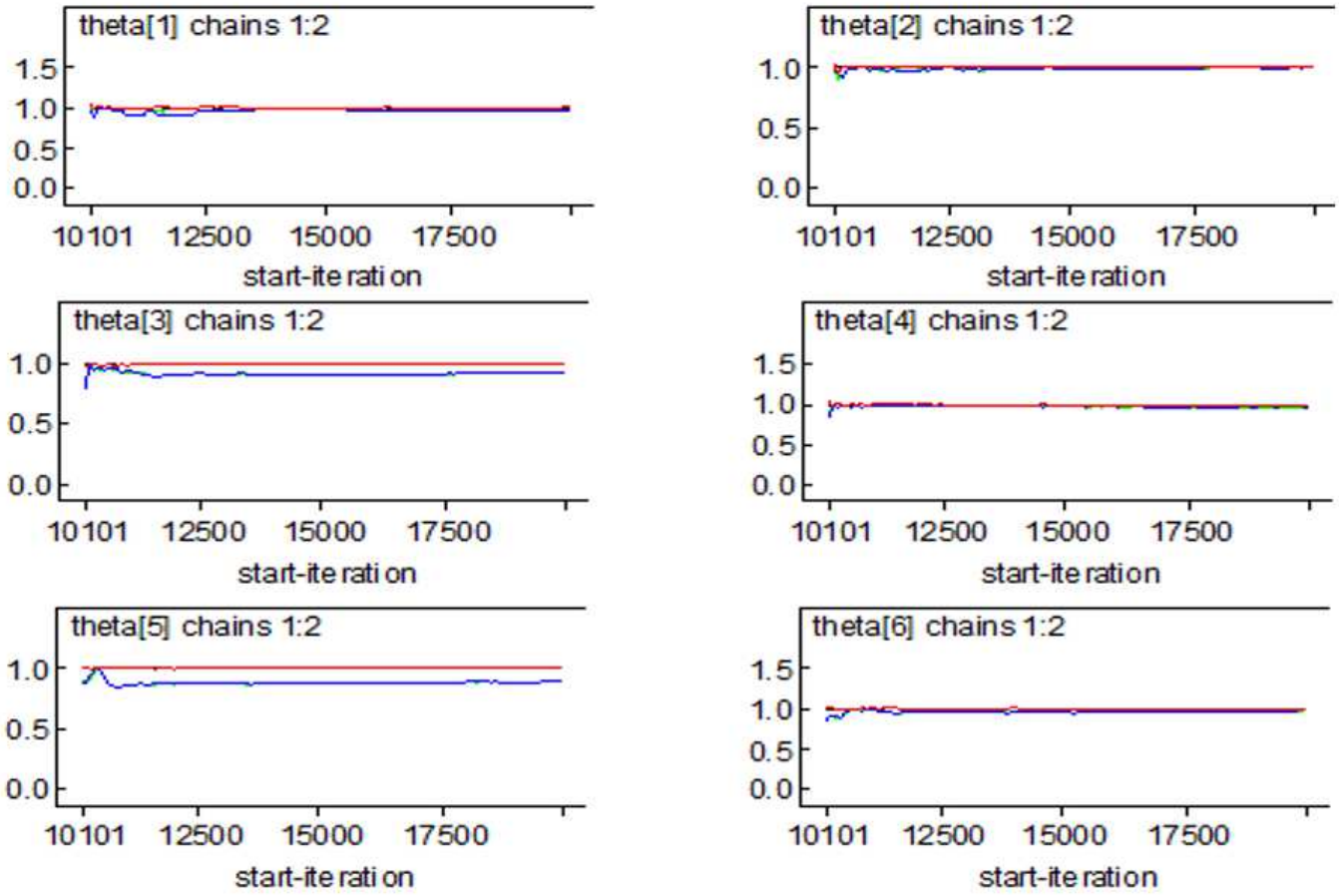

Figure 2

Brooks, Gelman and Rubin plots of the parameter theta(bgr)

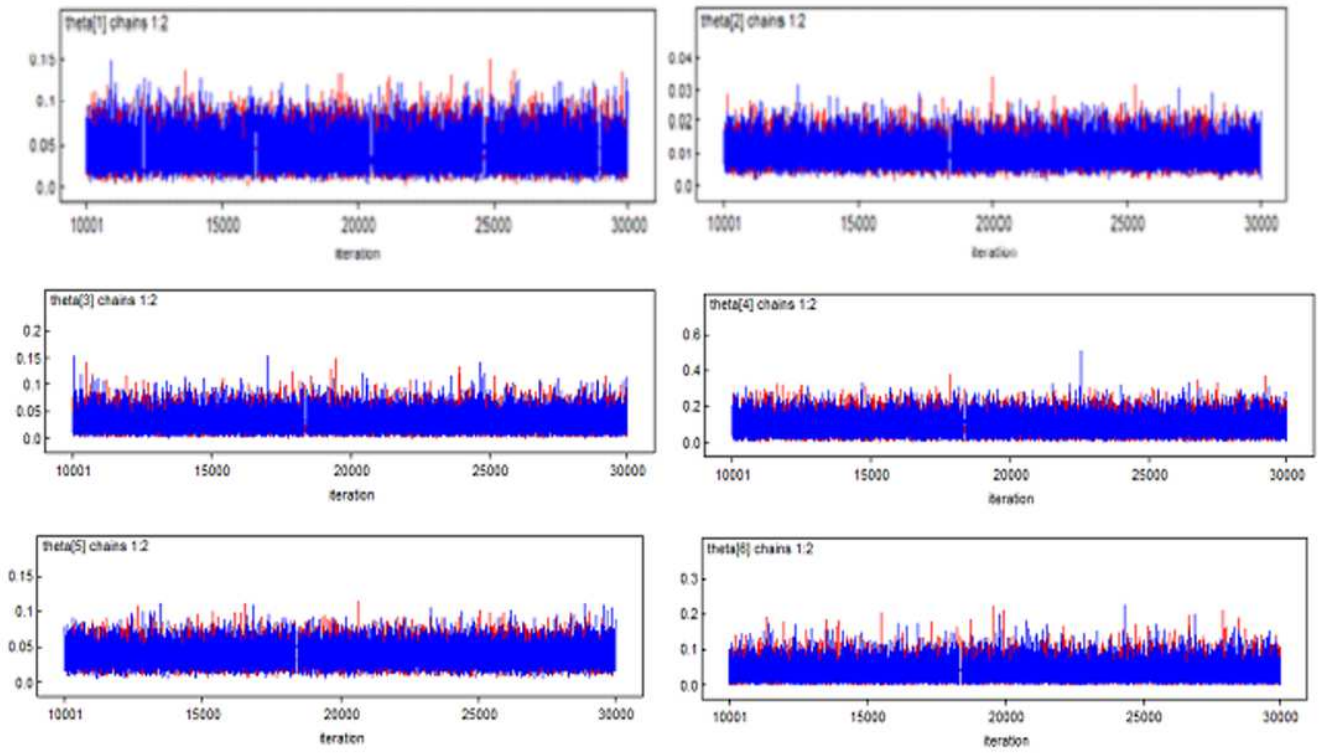

Figure 3

History plots of the parameter theta

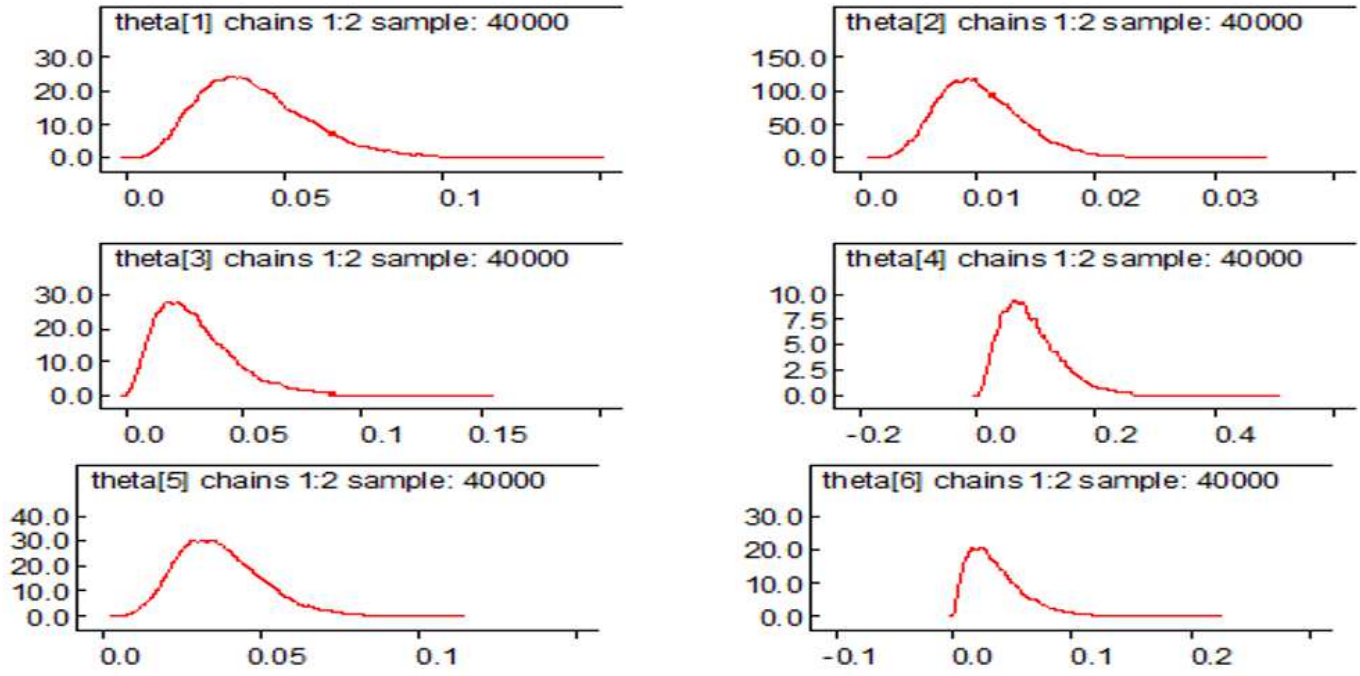

Figure 4

Density plots of theta for the first baseline regimen

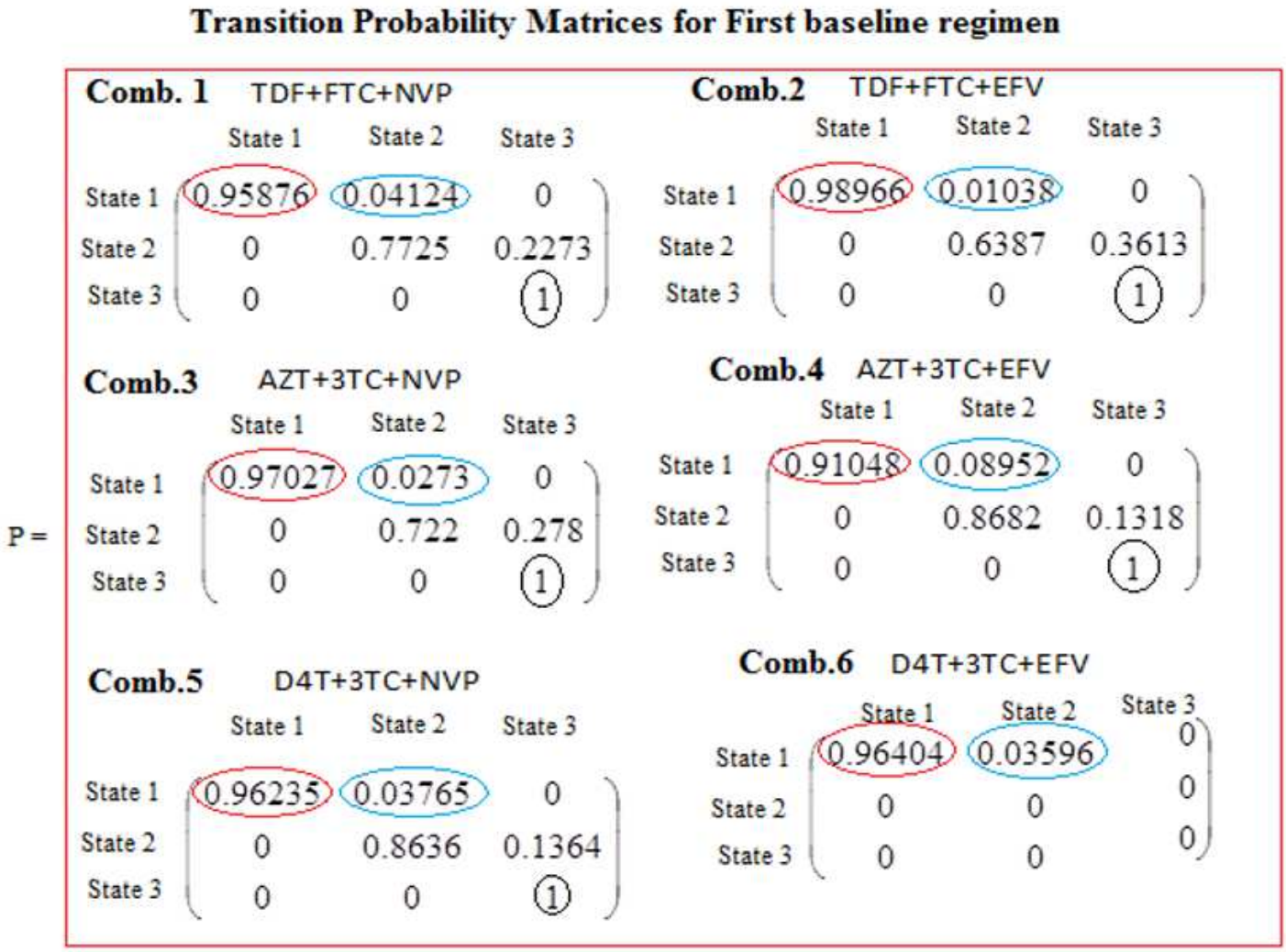

Figure 5

The transition probability matrix of ARV combinations