

# A mathematical modelling study of HIV infection in two heterosexual age groups in Kenya



E.O. Omondi\*, R.W. Mbogo, L.S. Luboobi

*Institute of Mathematical Sciences, Strathmore University, P.O Box 59857-00200, Nairobi, Kenya*

## ARTICLE INFO

### Article history:

Received 19 December 2018  
 Received in revised form 13 April 2019  
 Accepted 15 April 2019  
 Available online 21 April 2019  
 Handling Editor: Y. Shao

### Keywords:

Heterosexual transmission (HIV)  
 Basic reproduction number  
 MCMC  
 Probability distribution  
 Kruskal–Wallis test  
 Correlation

## ABSTRACT

The control of HIV demands different interventions for different age groups. In the present manuscript, we formulate and analyze a mathematical compartmental models of HIV transmission within and between two age groups in Kenya. We fitted the model to data using MCMC technique and inferred the parameters. We also estimate the reproduction numbers, namely within age group transmission and between age groups transmission basic reproduction numbers. The analysis of the data revealed that there is significant difference in mean number of new HIV infections between males and females within the two age groups. More, particularly, females are highly infected with HIV as compared to their male counterparts. Calculation of the reproduction numbers within and between age groups provides insights into control that cannot be deduced simply from observations on the prevalence of infection. More specifically, the analysis showed that the per capita rate of HIV transmission was highest when there is interaction between young adults to adults and most HIV infections occurred in adult population. Furthermore, the sensitivity analysis demonstrated that the reproduction numbers depend mainly on the probabilities of infection. This results can be used to guide HIV interventions, condom distribution and antiretroviral therapy. Precisely, the results can be used to educate the young adults on practicing safe sex with their partners in order to contain the occurrence of new infections.

© 2019 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Kenya has the joint fourth-largest HIV epidemic in the world (alongside Mozambique and Uganda) with 1.5 million people living with HIV in 2017 (AV, 2017). In the same year, 35,000 people died from AIDS-related illnesses, while this is still high it has declined steadily from 64,000 in 2010 (WHO, 2018). Kenya's HIV epidemic is driven by sexual transmission and is generalized, meaning it affects all sections of the population including children, young people, adults, women and men (AV, 2017). In 2010, the prevalence of HIV in the female population was 6.5% and 5.6% in the male population (WHO, 2018). While the prevalence of HIV infection has considerably reduced to 6.0 in 2010 from 10% in 1996, women still continue to be disproportionately affected by the HIV infections since men often dominate sexual relationships, with women not always able to practice safer sex even when they know the risks (AV, 2017). Young women are almost twice as likely to acquire HIV as their

\* Corresponding author.

E-mail address: [eomondi@strathmore.edu](mailto:eomondi@strathmore.edu) (E.O. Omondi).

Peer review under responsibility of KeAi Communications Co., Ltd.

male counterparts. At the end of 2015 young women accounted for 33% of the total number of new infections in comparison to young men that accounted for 16% (MOH, 2016).

Antiretroviral therapy (ART) is the current standard of care for patients living with HIV infection (WHO, 2014). It has led to significant reduction in AIDS related morbidity and mortality (Chow, Leong, Chow, & Hooi, 2007; Omondi, Mbogo, & Luboobi, 2018a; Williams, 2014). The control of HIV demands different interventions for different age groups. HIV education and awareness is an important component of HIV prevention. In Kenya, 73% of young women and 82% of young men in 2014 demonstrated adequate knowledge of HIV prevention (MOH, 2016). However, incorrect perception of HIV risk, and having unprotected sexual intercourse under influence of alcohol or drugs have been cited as some of the factors that contribute to the rise in HIV infection among young people (AV, 2017). HIV testing and counselling (HTC) has become a major feature of Kenya's HIV response. Targeted community-based HIV testing, door-to-door testing campaigns, and the introduction of self-testing kits are some of the innovative approaches to HIV testing adopted in recent years (UNAIDS, 2016). By 2015, about 9.9 million had been tested. Nonetheless, there remains a significant disparity between men and women. In 2014, 53% of women had tested for HIV in the past 12 months and received their results, compared to 45% of men (KNBS, 2014). In its effort to reduce HIV infections in Kenya, the government introduced self-testing kits in May 2017, the Kenya, as part of their 'Be Self Sure' campaign as well as PrEP which uses antiretroviral drugs to protect HIV-negative people from HIV before potential exposure to the virus (UNAIDS, 2017a).

Mathematical modelling is a common tool for understanding and studying the dynamics of infectious disease, and propose mitigation measures to control disease outbreaks (Keeling & Rohani, 2011). As such a number of HIV models have been constructed and analyzed to understand the transmission dynamics of the disease. Recently, Aldila (2018) studied the HIV transmission dynamics using a compartmental model. He considered the awareness of individuals both infected and uninfected with HIV as well as ART treatment intervention. Omondi, Mbogo, and Luboobi (2018b) modelled the impact of testing, treatment and control of HIV transmission that include anti-retroviral interventions in Kenya that among the adult population. They used the basic reproduction number and both the local and global stability to understand how HIV might spread. They established that an infection might suppress with the use of combination of testing, Pre-Exposure Prophylaxis and anti-retroviral treatment intervention. In another study, Kim et al. (2014) studies HIV prevention measures including Pre-Exposure Prophylaxis on HIV incidence and established that PrEP use was more beneficial in prevention of HIV infection in South Korea. Omondi, Mbogo, and Luboobi (2018c) studied the trend of HIV transmission and treatment in Kenya. However, their model was not stratified to include gender and age. Mukandavire, Chiyaka, Garira, and Musuka (2009) modelled a sex-structured model for heterosexual transmission of HIV/AIDS with explicit incubation period and provided an in-depth and complete qualitative analysis. However, the model neglected stratification by age. In another study, Mukandavire and Garira (2007) considered heterosexual interactions of males and females using integro-differential equations with a time delay due to incubation period. While they incorporated the effects of male and female condom use as the main mode of preventing HIV infection, no real-time surveillance data to establish the trend of infection. In addition, the model was not stratified by age. To the best of our knowledge, all the above research works mentioned focused on the mathematical analysis of the models and few papers of HIV infection exist, where gender and age is modelled using real-time surveillance. It is important to note that the mentioned studies did not attempt to use inference methods to fit the models to data and estimate model parameters. While both the impact and the cost of different combinations of interventions vary, we are concerned in this paper with the population impact that can be achieved for a given reduction in the individual risk of transmission irrespective of how it is brought about. This analysis focuses on the spread of an HIV epidemic in Kenya, as described by data collected. Thus, we develop a within and between age groups model of HIV transmission. The model is mathematically analyzed, fitted to data of new cases of HIV infections in Kenya and parameters are inferred.

This paper is organized as follows: In Section 2, we develop mathematical model. In Section 3, we find the expressions for basic reproductive numbers for within and between age groups. Results are presented in Section 4. Finally, the paper ends with a conclusion in Section 5.

## 2. Methods

### 2.1. Model formulation

We consider a simple mathematical model to understand the dynamics of HIV within and between two different age categories in Kenya. In our modelling framework, it is assumed that HIV transmission is mainly through heterosexual means. The population is divided into young adults (aged 15–24) and adult (age 25 and over) sub-populations. Each sub-population is divided into susceptible individuals (S), infected individuals (I) and those who have been enrolled into treatment programme mainly ART (T). We assume that all the infected individuals are connected to ART treatment. The AIDS class is not considered in this model given that full blown AIDS patients are usually hospitalized and/or sexually inactive. It is assumed that they are not able to engage in HIV transmission activities hence do not contribute to HIV infection. The total variable population at time  $t$  is described by

$$N(t) = N_m(t) + N_f(t), \quad (1)$$

where the subscripts  $m$  and  $f$  denote male and female and the individual sex oriented population described by

$$\left. \begin{aligned} N_m(t) &= S_{dm} + S_{am} + I_{dm} + I_{am} + T_{dm} + T_{am}, \\ N_m(t) &= S_{df} + S_{af} + I_{df} + I_{af} + T_{df} + T_{af}. \end{aligned} \right\} \quad (2)$$

here,  $d$  and  $a$  represent the young adults (aged between 15 and 24 years) and the adults (aged over 25 years), respectively. Individuals move from one class to the other as their status evolve with respect to the infection. The population of the susceptible young adults is generated at the rate  $\Pi$  via maturation into adulthood or immigration of which a proportion  $\tau$  are assumed to be males and  $(1 - \tau)$  are assumed to be females. Since the current study is looking at the trend of new HIV infections within and between these two age groups, it is assumed that there is no vertical transmission. The population is reduced by young adults maturation at the rate  $\alpha$  and by natural death at the rate  $\mu_a$ . The infection rate of the young adults in both males and females is respectively, given by

$$\lambda_{dm} = \frac{\beta_1 \gamma_1 (I_{df} + \theta_1 I_{af} + \theta_2 T_{df} + \theta_3 T_{af})}{N_m} \text{ and } \lambda_{df} = \frac{\beta_2 \gamma_2 (I_{dm} + \theta_4 I_{am} + \theta_5 T_{dm} + \theta_6 T_{am})}{N_f}. \quad (3)$$

The parameters  $\beta_1$ , and  $\beta_2$  are the probabilities of HIV infection through contacts with individuals in  $I_{ij}$  and  $T_{ij}$ , and  $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ , and  $\theta_6$  are modification factors in transmission probabilities, where  $i, j$  refers to young adults and adults, respectively. The infected young adults for both males and females are connected to ART treatment and care at the rates  $\phi_1$  and  $\delta_1$ , respectively. The male and female adults acquire infection at the rates, respectively given by

$$\lambda_{am} = \frac{\beta_3 \gamma_3 (I_{df} + \eta_1 I_{af} + \eta_2 T_{df} + \eta_3 T_{af})}{N_m} \text{ and } \lambda_{af} = \frac{\beta_4 \gamma_4 (I_{dm} + \eta_4 I_{am} + \eta_5 T_{dm} + \eta_6 T_{am})}{N_f}, \quad (4)$$

where the parameters  $\beta_3$ , and  $\beta_4$  are the probabilities of HIV infection through contacts with individuals in  $I_{ij}$  and  $T_{ij}$ , respectively, and  $\eta_1, \eta_2, \eta_3, \eta_4, \eta_5$ , and  $\eta_6$  are modification factors in transmission probabilities. The infected male and female adults are connected to ART treatment and care at the respective rates given by  $\phi_2$  and  $\delta_2$ . The adult classes are reduced by the natural death rate  $\mu_a$ . The parameters  $\gamma_k$ , for  $k = 1, \dots, 4$ , are the rate at which individuals in each age category acquire sexual partners. These compartments have been schematically illustrated in Fig. 1.

Given the above descriptions and assumptions, the dynamics of HIV in the population is given by the following deterministic system of non-linear differential equations.

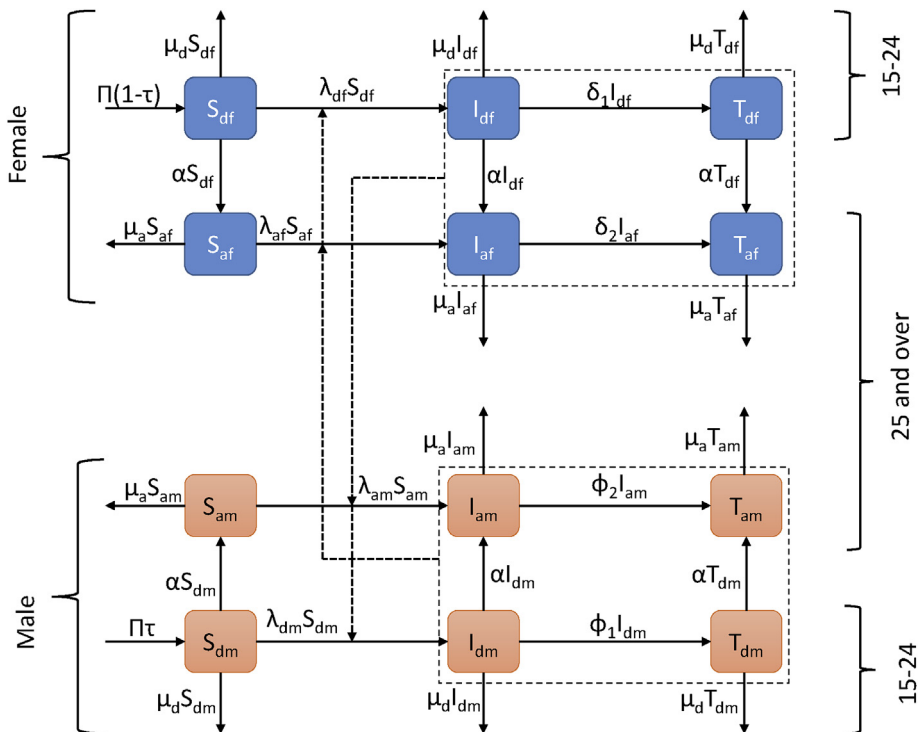


Fig. 1. Schematic diagram of HIV model in the presence of ART.

$$\left. \begin{aligned} \frac{dS_{dm}}{dt} &= \Pi\tau - \lambda_{dm}S_{dm} - (\mu_d + \alpha)S_{dm}, & \frac{dS_{df}}{dt} &= \Pi(1 - \tau) - \lambda_{df}S_{df} - (\mu_d + \alpha)S_{df}, \\ \frac{dI_{dm}}{dt} &= \lambda_{dm}S_{dm} - (\phi_1 + \alpha + \mu_d)I_{dm}, & \frac{dI_{df}}{dt} &= \lambda_{df}S_{df} - (\alpha + \delta_1 + \mu_d)I_{df}, \\ \frac{dT_{dm}}{dt} &= \phi_1 I_{dm} - (\alpha + \mu_d)T_{dm}, & \frac{dT_{df}}{dt} &= \delta_1 I_{df} - (\alpha + \mu_d)T_{df}, \\ \frac{dS_{am}}{dt} &= \alpha S_{dm} - \lambda_{am}S_{am} - \mu_a S_{am}, & \frac{dS_{af}}{dt} &= \alpha S_{df} - \lambda_{af}S_{af} - \mu_a S_{af}, \\ \frac{dI_{am}}{dt} &= \lambda_{am}S_{am} + \alpha I_{dm} - (\phi_2 + \mu_a)I_{am}, & \frac{dI_{af}}{dt} &= \lambda_{af}S_{af} + \alpha I_{df} - (\delta_2 + \mu_a)I_{af}, \\ \frac{dT_{am}}{dt} &= \phi_2 I_{am} + \alpha T_{dm} - \mu_a T_{am}, & \frac{dT_{af}}{dt} &= \delta_2 I_{af} + \alpha T_{df} - \mu_a T_{af}, \end{aligned} \right\} \tag{5}$$

subject to the following initial conditions

$$S_{ijm}(0) \geq 0, I_{ijm}(0) \geq 0, T_{ijm}(0) \geq 0, S_{ijf}(0) \geq 0, I_{ijf}(0) \geq 0, T_{ijf}(0) \geq 0, \text{ for } i, j = d, a. \tag{6}$$

### 3. Model dynamics

#### 3.1. Well-posedness of the model

In this section, we show that the system (5) is mathematically well defined and biologically feasible. The system (5) can be rewritten in the following form

$$\frac{dX}{dt} = A(X)X + F,$$

where  $X = (S_{dm}, I_{dm}, T_{dm}, S_{am}, I_{am}, T_{am}, S_{df}, I_{df}, T_{df}, S_{af}, I_{af}, T_{af})^t$ . Let  $Q_1 = \mu_d + \alpha, Q_2 = \phi_1 + \alpha + \mu_d, Q_3 = \phi_2 + \mu_a, Q_4 = \alpha + \delta_1 + \mu_d, Q_5 = \delta_2 + \mu_a$ . Thus, the matrix  $A$  is given by

$$A(X) = \begin{bmatrix} -Q_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{dm} & -Q_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_1 & -Q_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha & 0 & 0 & -Q_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha & 0 & \lambda_{am} & -Q_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha & 0 & \phi_2 & -\mu_a & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -Q_8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{df} & -Q_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta_1 & -Q_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -Q_9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{af} & -Q_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha & \delta_2 & -\mu_a \end{bmatrix}$$

and  $F = (\Pi\tau, 0, 0, 0, 0, 0, \Pi(1 - \tau), 0, 0, 0, 0, 0)^T$ . Here,  $Q_6 = (\lambda_{dm} + Q_1), Q_7 = (\lambda_{am} + \mu_a), Q_8 = (\lambda_{df} + Q_1), Q_9 = (\lambda_{af} + \mu_a)$ . It is important to note that  $A(X)$  is a Metzler matrix, that is, a matrix such that off diagonal entries non-negative, for all  $Y \in \mathbb{R}_+^{12}$ . Thus, using the fact that  $F \geq 0$ , the system (5) is positively invariant in  $\mathbb{R}_+^{12}$  (see, [Abate, Tiwari, and Sastry \(2009\)](#); [Berman and Plemmons \(1994\)](#)). This implies that any trajectory of the system (5) starting from an initial state in  $\mathbb{R}_+^{12}$  forever remains in  $\mathbb{R}_+^{12}$ . The evolution of the system (5) is described by  $\frac{dN}{dt} = \Pi - \mu N$ . Thus, solving for  $N(t)$  we get

$$N(t) \leq \frac{\Pi}{\mu} + e^{-\mu t} \left( N(0) - \frac{\Pi}{\mu} \right). \tag{7}$$

There are two possible cases in studying the behaviour of  $N(t)$  in (7). In the first case, we consider  $N(0) > \frac{\Pi}{\mu}$  so that, at time  $t = 0$ , the right-hand side (RHS) of (7) experiences the largest possible value of  $N(0)$ . That is,  $N(t) \leq N(0)$  for all time  $t \geq 0$ . In the second case, we consider  $N(0) < \frac{\Pi}{\mu}$  so that the largest possible value of the RHS of (7) approaches  $\frac{\Pi}{\mu}$  as time  $t$  approaches

infinity. Thus,  $N(t) \leq \frac{\Pi}{\mu}$  for all time  $t \geq 0$ . From these two cases, we conclude that  $N(t) \leq \max\left\{N(0), \frac{\Pi}{\mu}\right\}$  for all time  $t \geq 0$ . Therefore, we can study the system (5) in the feasible region given by

$$\Omega = \left\{ (S_{ijm}(t), I_{ijm}(t), T_{ijm}(t), S_{ijf}(t), I_{ijf}(t), T_{ijf}(t)) \in \mathbb{R}_+^{12} : N(t) \leq \max\left\{N(0), \frac{\Pi}{\mu}\right\} \right\},$$

which is positively invariant with respect to systems (5). This implies that the systems (5) is well posed epidemiologically and all the solutions starting in  $\Omega$  remain in  $\Omega$  for all  $t \geq 0$ .

### 3.2. Basic reproduction number

According to [Diekmann and Heesterbeek \(2000\)](#); [Diekmann, Heesterbeek, and Metz \(1990\)](#), the basic reproduction number commonly denoted as  $\mathcal{R}_0$  is defined as the number of secondary cases of infections arising from the introduction of a single infected individual in a wholly susceptible population. The system (5) has a unique disease-free equilibrium given by

$$\mathcal{E}_0 = \left( \frac{\Pi\tau}{Q_1}, 0, 0, \frac{\alpha\Pi\tau}{Q_1\mu_a}, 0, 0, \frac{\Pi(1-\tau)}{Q_1}, 0, 0, \frac{\alpha\Pi(1-\tau)}{Q_1\mu_a}, 0, 0 \right). \tag{8}$$

Since the model system (5) allows for free mixing of the individuals from the two stated age groups, there are two ways of the disease transmission. These are within age group transmission and between age groups transmission. We begin by finding the reproduction numbers within the age groups. For the young adults, the disease-free equilibrium is given by

$$\mathcal{E}_d = \left( \frac{\Pi\tau}{\mu_d}, 0, 0, \frac{\Pi(1-\tau)}{\mu_d}, 0, 0 \right). \tag{9}$$

Following the approach given in [Van den Driessche and Watmough \(2002\)](#), we let  $F(t)$  and  $V(t)$  be the matrices of new infections and transmission, respectively. Thus, at the disease-free equilibrium defined in (9), these matrices are respectively given by

$$F(t) = \begin{bmatrix} 0 & 0 & \beta_1\gamma_1 & \beta_1\gamma_1\theta_2 \\ 0 & 0 & 0 & 0 \\ \beta_2\gamma_2 & \beta_2\gamma_2\theta_5 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad V(t) = \begin{bmatrix} \mu_d + \phi_1 & 0 & 0 & 0 \\ -\phi_1 & \mu_d & 0 & 0 \\ 0 & 0 & \delta_1 + \mu_d & 0 \\ 0 & 0 & -\delta_1 & \mu_d \end{bmatrix}.$$

The young adults transmission reproduction number which is the spectral radius of the next-generation matrix (NGM) for the epidemic of HIV given by  $FV^{-1}(t)$  is obtained as

$$\mathcal{R}_{0d} = \sqrt{\left[ \frac{\beta_1\gamma_1(\delta_1\theta_2 + \mu_d)}{\mu_d(\mu_d + \phi_1)} \right] \left[ \frac{\beta_2\gamma_2(\mu_d + \theta_5\phi_1)}{\mu_d(\mu_d + \delta_1)} \right]}. \tag{10}$$

Similarly, the disease-free equilibrium within the adult age group is defined as

$$\mathcal{E}_a = \left( \frac{\Pi\tau}{\mu_d}, 0, 0, \frac{\Pi(1-\tau)}{\mu_d}, 0, 0 \right). \tag{11}$$

The matrices of new infections and transmission within adult (age group 25+) evaluated at disease-free in (11) are, respectively given as

$$F(t) = \begin{bmatrix} 0 & 0 & \beta_3\gamma_3\eta_1 & \beta_3\gamma_3\eta_3 \\ 0 & 0 & 0 & 0 \\ \beta_4\gamma_4\eta_4 & \beta_4\gamma_4\eta_6 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad V(t) = \begin{bmatrix} \mu_a + \phi_2 & 0 & 0 & 0 \\ -\phi_2 & \mu_a & 0 & 0 \\ 0 & 0 & \delta_2 + \mu_a & 0 \\ 0 & 0 & -\delta_2 & \mu_a \end{bmatrix}.$$

Therefore, the transmission reproduction number is given by

$$\mathcal{R}_{0a} = \sqrt{\left[ \frac{\beta_3\gamma_3(\eta_1\mu_a + \delta_2\eta_3)}{\mu_a(\mu_a + \phi_2)} \right] \left[ \frac{\beta_4\gamma_4(\eta_4\mu_a + \eta_6\phi_2)}{\mu_a(\mu_a + \delta_2)} \right]}. \tag{12}$$

If the HIV infection exists in a single age group connected to another age group through maturation, then the movement of the individuals must be reflected in the basic reproduction number. The matrices for new infection terms and the transfer terms at the disease-free equilibrium (8) are given by

$$F(t) = \begin{bmatrix} 0 & 0 & \frac{\beta_1\gamma_1\theta_1\mu_a}{\alpha + \mu_a} & \frac{\beta_1\gamma_1\theta_3\mu_a}{\alpha + \mu_a} \\ 0 & 0 & 0 & 0 \\ \frac{\alpha\beta_4\gamma_4}{\alpha + \mu_a} & \frac{\alpha\beta_4\gamma_4\eta_5}{\alpha + \mu_a} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad V(t) = \begin{bmatrix} Q_2 & 0 & 0 & 0 \\ -\phi_1 & Q_1 & 0 & 0 \\ 0 & 0 & Q_5 & 0 \\ 0 & 0 & -\delta_2 & \mu_a \end{bmatrix}$$

Thus, the basic reproduction number between the male young adults and the female adults (aged 25 + years) is given by

$$\mathcal{R}_{0mdfa} = \sqrt{\left[ \frac{\alpha\beta_4\gamma_4(\eta_5\phi_1 + Q_1)}{Q_1Q_2(\mu_a + \alpha)} \right] \left[ \frac{\beta_1\gamma_1(\theta_1\mu_a + \delta_2\theta_3)}{Q_5(\mu_a + \alpha)} \right]}. \quad (13)$$

On the other hand, the matrices for new infection terms and the transfer terms at the disease-free equilibrium in (8) for the female young adults and male adults are given by

$$F = \begin{bmatrix} 0 & 0 & \frac{\alpha\beta_3\gamma_3}{\alpha + \mu_a} & \frac{\alpha\beta_3\gamma_3\eta_2}{\alpha + \mu_a} \\ 0 & 0 & 0 & 0 \\ \frac{\beta_2\gamma_2\theta_4\mu_a}{\alpha + \mu_a} & \frac{\beta_2\gamma_2\theta_6\mu_a}{\alpha + \mu_a} & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} Q_3 & 0 & 0 & 0 \\ -\phi_2 & \mu_a & 0 & 0 \\ 0 & 0 & Q_4 & 0 \\ 0 & 0 & -\delta_1 & Q_1 \end{bmatrix}.$$

Hence, the basic reproduction number between the female young adults and the male adults is given by

$$\mathcal{R}_{0fdma} = \sqrt{\left[ \frac{\beta_2\gamma_2(\theta_4\mu_a + \theta_6\phi_2)}{Q_3(\mu_a + \alpha)} \right] \left[ \frac{\alpha\beta_3\gamma_3(\delta_1\eta_2 + Q_1)}{Q_1Q_4(\mu_a + \alpha)} \right]}. \quad (14)$$

The basic reproduction number,  $\mathcal{R}_0$ , of the between the age groups of the system (5) is given as the maximum of the between age groups specific reproduction numbers. Thus, we have

$$\mathcal{R}_0 = \max\{\mathcal{R}_{0mdfa}, \mathcal{R}_{0fdma}\}.$$

It is important to note that due to mathematical intractability we are unable to explicitly express the basic reproduction number of the whole system without splitting the model into within and between age groups. To understand the trend of HIV infection, we perform data analysis and present results in section 4.

## 4. Results

### 4.1. Epidemiological data and ethical considerations

To study the extent and trend of HIV infection, we analyze the confirmed cases of new infections in Kenya from January 2011 to September 2018. The data analyzed was routinely collected on a monthly basis and retrieved from Kenya Health information System available at KHIS. Only variables of interest were pulled out to excel spreadsheet and thereafter analyzed in **R**. The data analyzed is publicly available. Thus, the datasets used in our study were de-identified and fully anonymized in advance, and the analysis of publicly available data without identity information does not require ethical approval.

### 4.2. Parameter inference and estimation

The natural death rate was estimated to be  $\mu_d = 0.0013$ ,  $\mu_a = 0.00128$  based on the life expectancy in Kenya (WH, 2018). The young adults maturation at the rate  $\alpha = 0.0083$ . The rate at young adults acquire sexual partners is assumed to be 3, that is,  $\gamma_1 = \gamma_2 = 3$ , while that of the adults has been assumed to be 2 ( $\gamma_3 = \gamma_4 = 2$ ). Initiating and staying on treatment is particularly problematic for young adults. In 2014, it was estimated that only 34,800 out of 141,000 young adults with known HIV positive status were on ART (AV, 2017). Thus, the rates at which adolescents (males and females) are connected to ART

treatment are assumed to be  $\phi_1 = 0.24$  and  $\delta_1 = 0.28$ , respectively. On the other hand, based on the estimates from AV (2017), the rates at which the male and female adults are connected to ART treatment is assumed to be  $\phi_2 = 0.58$  and  $\delta_2 = 0.68$ , respectively. Table 1 gives the description of the parameters and the initial conditions estimates used in this work. The initial conditions for  $S_{dm}, S_{df}, S_{am}$  and  $S_{af}$  are estimated from Kenya demographics profile of both 2010 and 2018 (see KD (2018)) while other initial conditions for  $I_{dm}, T_{dm}, I_{df}, T_{fm}, I_{am}, T_{am}, I_{af}$  and  $T_{af}$  are estimated based on the retrieved data that is used in curve fitting.

The unknown parameters, that is,  $\beta_1, \beta_2, \beta_3, \beta_4, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \eta_1, \eta_2, \eta_3, \eta_4, \eta_5$  and  $\eta_6$ , were estimated on the basis of the available data as earlier described. Bayesian approach that is implemented to the Markov Chain Monte Carlo (MCMC) technique is used in parameter estimation. We minimize the sum of the squared error between the model and data, which is given by

$$SS(\hat{\vartheta}) = \sum_{i=1}^n (Y_h - P(t_h, \hat{\vartheta}))^2, \tag{15}$$

where

$$P(t_h, \hat{\vartheta}) = P(t, \hat{\vartheta}) = \int_{t-1}^t p(\lambda_{dm}S_{dm} + \lambda_{df}S_{df} + \lambda_{am}S_{am} + \lambda_{af}S_{af}) dt,$$

which is the number of new HIV cases of infection for each age group. Note that, there are  $D$  independent observations from the dataset that represent the number of new HIV cases of infection at the  $h$ th month, for  $h = 1, 2, 3, \dots, D$ . Now considering  $\varepsilon$  is the error of fit, which follows an independent Gaussian distribution having unknown variance  $\sigma^2$ , then it follows from (15) that

$$Y_h = P(t_h, \hat{\vartheta}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I_{ij}^* \sigma^2),$$

with  $*$  referring to males and females,  $i$  and  $j$  remain as earlier defined. We assume an independent Gaussian prior specification for the unknown parameters  $\hat{\vartheta}$ , that is,  $\vartheta_r \sim \mathcal{N}(\nu_r, \psi_r^2)$ , where  $r = 1, 2, 3, \dots, D$ . Furthermore, it is assumed that the inverse of the error variance follows a Gaussian distribution as prior taking the following form

$$v(\sigma^{-2}) \sim \Gamma\left(\frac{x_0}{2}, \frac{x_0 S_0^2}{2}\right).$$

here,  $x_0$  and  $S_0^2$  respectively, give the prior accuracy and prior mean of  $\sigma^2$ . Considering the conditional conjugacy property of Gamma distribution (see, Sardar, Sasmal, and Chattopadhyay (2016)), the conditional distribution of  $v(\sigma^{-2} | Y, \hat{\vartheta})$  is also a Gamma distribution with

**Table 1**  
Description of the parameters and the initial conditions estimates for the system (5). The parameters are given per month.

Par/var	Range	Value	Source	Par	Range	Value	Source
$S_{dm}(0)$	4,148,153–4,552,448	Est.	KD (2018)	$\beta_1$	0.0–1.0	Est.	
$I_{dm}(0)$	0–8,000	Est.		$\beta_2$	0.0–1.0	Est.	
$T_{dm}(0)$	0–6,000	Est.		$\beta_3$	0.0–1.0	Est.	
$S_{df}(0)$	4,147,896–4,567,894	Est.	KD (2018)	$\beta_4$	0.0–1.0	Est.	
$I_{df}(0)$	0–11,000	Est.		$\theta_1$	0.0–1.0	Est.	
$T_{fm}(0)$	0–8,000	Est.		$\theta_2$	0.0–1.0	Est.	
$S_{am}(0)$	8,460,138–9,641,107	Est.	KD (2018)	$\theta_3$	0.0–1.0	Est.	
$I_{am}(0)$	0–11,000	Est.		$\theta_4$	0.0–1.0	Est.	
$T_{am}(0)$	0–9,000	Est.		$\theta_5$	0.0–1.0	Est.	
$S_{af}(0)$	8,624,799–9,799,146	Est.	KD (2018)	$\theta_6$	0.0–1.0	Est.	
$I_{af}(0)$	0–17,000	Est.		$\eta_1$	0.0–1.0	Est.	
$T_{af}(0)$	0–14,000	Est.		$\eta_2$	0.0–1.0	Est.	
$\Pi$	40,000–85,000	44,000	KD (2018)	$\eta_3$	0.0–1.0	Est.	
$\tau$	0.0–1.0	0.48	WH (2018)	$\eta_4$	0.0–1.0	Est.	
$\mu_d$	0.0011–0.0017	0.0013	WH (2018)	$\eta_5$	0.0–1.0	Est.	
$\mu_a$	0.0011–0.0017	0.00128	WH (2018)	$\eta_6$	0.0–1.0	Est.	
$\gamma_1$	1–4	3	Assumed	$\gamma_2$	1–4	3	Assumed
$\gamma_3$	1–4	2	Assumed	$\gamma_4$	1–4	2	Assumed
$\phi_1$	0.0–1.0	0.24	AV (2017)	$\phi_2$	0.0–1.0	0.58	AV (2017)
$\delta_1$	0.0–1.0	0.28	AV (2017)	$\delta_2$	0.0–1.0	0.68	AV (2017)
$\alpha$		0.0083					

$$v(\sigma^{-2} | Y, \hat{\vartheta}) = \Gamma\left(\frac{x_0 + R}{2}, \frac{x_0 S_0^2 + SS(\hat{\vartheta})}{2}\right).$$

The above property makes it possible to sample and update  $\sigma^{-2}$  within each Metropolis Hastings simulation step for the other parameters. Since an independent Gaussian prior specification for  $\hat{\vartheta}$  is assumed, the prior sum of squares for  $\hat{\vartheta}$  is given by

$$SS_{\text{pri}}(\hat{\vartheta}) = \sum_{h=1}^D \left(\frac{\vartheta_h - \nu_h}{\psi_h}\right)^2.$$

For a fixed value of  $\sigma^2$ , the posterior distribution of  $\hat{\vartheta}$  is given by

$$v(\hat{\vartheta} | Y, \sigma^2) \propto \exp\left[-\frac{1}{2}\left(\frac{SS(\hat{\vartheta})}{\sigma^2} + SS_{\text{pri}}(\hat{\vartheta})\right)\right],$$

with the posterior ratio needed in the Metropolis-Hastings acceptable probability given as

$$\frac{v(\hat{\vartheta}^1 | Y, \sigma^2)}{v(\hat{\vartheta}^2 | Y, \sigma^2)} = \exp\left[-\frac{1}{2}\left(\left(\frac{SS(\hat{\vartheta}^1)}{\sigma^2} - \frac{SS(\hat{\vartheta}^2)}{\sigma^2}\right) + \frac{1}{2}(SS_{\text{pri}}(\hat{\vartheta}^2) + SS_{\text{pri}}(\hat{\vartheta}^1))\right)\right].$$

The modCost, modFit and modMCMC routine in package FME package (A flexible modelling environment for inverse modelling, sensitivity, identifiability and Monte Carlo Analysis) in R is used to estimate the unknown  $\vartheta$  for the model. A R code is used in which, the unknown parameter values are given a lower bound and an upper bound from which the set of parameter values that produce the best fit are obtained. The parameter estimates and other results arising from the model fitting to data are given in Section 4.4.

### 4.3. Statistical analysis

#### 4.3.1. The basic description of data

In this section, we carry out simple descriptive statistical analysis of the dataset and results presented in Table 2.

The mean number of new HIV infections in the male young adults age group is 1336.9 (95% Confidence Intervals (CI), 1114.0,1559.8) while the average number of new infections in the females of the same age group is 3164.7 (95% Confidence Intervals (CI), 2775.3,3554.1), for the period from January 2011 to September 2018. The average number of new infections in male and female adults are given by 5319.2 (95% Confidence Intervals (CI), 4817.9,5820.5) and 7692.5 (95% Confidence Intervals (CI), 6951.9,8433.4), respectively. Overall, the mean number of HIV infections in males is 3328.1 while that in females is 5428.6. It can be seen that females in both age categories are disproportionately affected with HIV more than the males. In order to establish the extent of variation in the mean number of cases of infections in the two age categories along the gender line, an error bar is plotted and presented in Fig. 2. It can be seen that the non-overlapping error bars may be significantly different. This implies that further test is required to indicate the nature of differences means. Thus, in section 4.3.2, we carry out probability distribution test in order to choose an appropriate test to establish the mean differences.

**Table 2**

Descriptive characteristics of the dataset retrieved for the duration spanning from January 2011 to September 2018.

Age group	Male				Female			
	Mean	SD	SE	95% CI <sup>a</sup>	Mean	SD	SE	95% CI <sup>a</sup>
15–24 years	1336.9	1082.4	112.2	[1114.0,1559.8]	3164.7	1891.0	196.1	[2775.3,3554.1]
25 + years	5319.2	2434.3	252.4	[4817.9,5820.5]	7692.5	3597.4	373.0	[6951.9,8433.4]
d <sup>b</sup> and a <sup>c</sup>	3328.1	2741.5	201.0	[2931.5,3724.7]	5428.6	3656.1	268.1	[4899.7,5957.5]

<sup>a</sup> 95% Confidence Interval.

<sup>b</sup> The young adults aged 15–24 years.

<sup>c</sup> The adults aged 25 and over years.



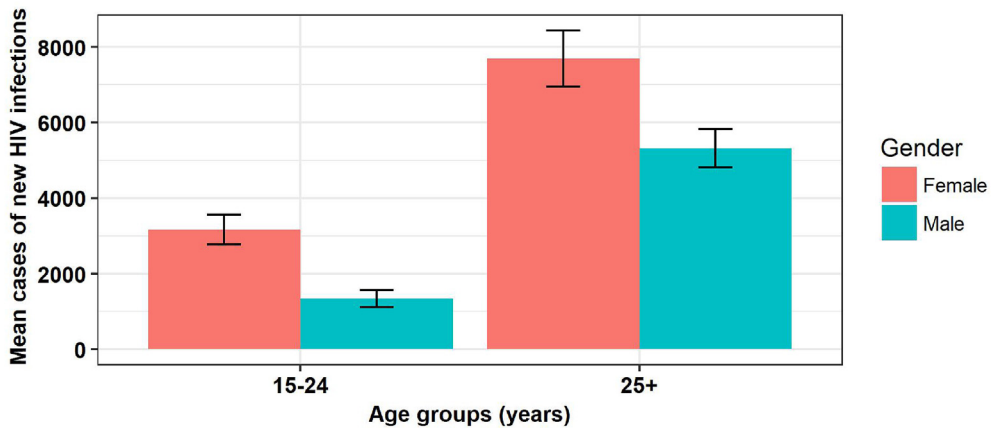


Fig. 2. A graph showing the distribution of the average number of new HIV infections in two age groups for males and females. Error bars are 95% confidence intervals.

4.3.2. The probability distribution of the data

The probability distribution of the given dataset plays an important role in determining which tests between parametric and non-parametric to conduct. There are various methods used to test for the probability distribution of a given dataset. The methods can be to test for normality or any other distribution. For normality tests, methods used include kolmogorov-smirnov, Anderson Darling, Shapiro Wilk and Lilliefors test (Shapiro & Wilk, 1965). In this study, we use the Shapiro Wilk test. This is the most powerful test when compared to the Anderson Darling, Kolmogorov-Smirnov and Lilliefors tests (Razali et al., 2011). The test statistics proposed in Shapiro and Wilk (1965) is given by

$$W = \frac{(a'y)^2}{S^2} = \frac{\left(\sum_{q=1}^n a_q y_q\right)^2}{\sum_{q=1}^n (y_q - \bar{y})^2},$$

where  $a'$  are set of weights given by

$$a' = (a_1, \dots, a_n) = \frac{m'V^{-1}}{\sqrt{(m'V^{-1}V^{-1}m)}}.$$

here,  $y_q$ , for  $q = 1, 2, \dots, n$ , is the  $i$ th order statistics whose similarity scores are sorted in either descending or ascending order,  $\bar{y}$  is the sample mean similarity score,  $m = (m_1, \dots, m_n)'$  are the first moments of the order statistics which are independent and identically normally distributed random variables,  $S^2$  is the estimator for the population variance  $\sigma^2$  and  $V$  is the covariance matrix of the order statistics. The dataset is assumed not to follow a normal distribution when the test statistics  $W$  is small, that is,  $0 < \frac{na^2}{n-1} \leq W \leq 1$  or when p-value  $< \alpha$ , the significance level. Otherwise the dataset follows a normal distribution.

Table 3 shows results from Shapiro Wilk test for normality of the dataset. The test was carried out at alpha level equal to 0.05, that is, at 95% Confidence Interval. Given that the p-value for each age category for males and females is less than 0.05, then the null hypothesis that the data are normally distributed is rejected. Thus, there is no enough evidence to assume that the data follows a normal distribution. Fig. 3 shows density plot to visualise the distribution of data. This chart uses kernel smoothing to plot values, allowing for smoother distributions by smoothing out the noise. The peaks of a density plot help

Table 3  
Shapiro Wilk test for normality of the dataset.

Age group	Male		Female	
	W	P-value	W	P-value
15–24 years	0.8585	0.0000	0.9549	0.0028
25 + years	0.9692	0.0268	0.9600	0.0061

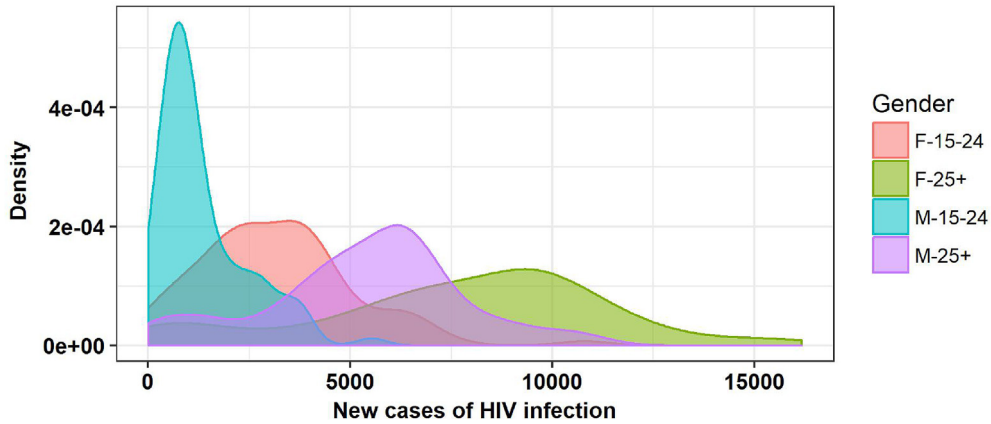


Fig. 3. Density distribution of the dataset.

display where values are concentrated over the interval. It can be easily seen that the data are positively skewed. Since the results show that the data does not follow a normal distribution, we conduct Friedman test, a non-parametric test, to establish if there exists any significant differences in mean number of HIV infections between the two age groups for the males and females.

4.3.3. *Kruskal–Wallis test*

Kruskal–Wallis’s test is a non-parametric method for testing the equality of several independent samples. It is useful in analyzing experimental data from completely randomized designs (Kruskal & Wallis, 1952). To compute the Kruskal–Wallis test statistic, all the observations are first ranked in ascending order where the smallest observation takes rank 1 and the largest observation takes rank  $N$ . The sum and average of the ranks of the observations pertaining to each sample are obtained next. If the sample effects are equal, then the average ranks are expected to be the same and if there is any difference then that is due to sampling fluctuations. The Kruskal–Wallis test statistic is based on the assessment of the differences among the average ranks. That is, let  $R_{ij}$  be the rank of  $y_{ij}$ ,  $i = 1, 2, 3, \dots, b$ ;  $j = 1, 2, \dots, t$  where  $b$  refers to the samples (treatments) and the  $i$ th treatment is replicated  $t_1$  times,  $i = 1, 2, \dots, b$ ,  $R_i = \sum_{j=1}^t R_{ij}$  be the sum of the ranks of the observations pertaining to the  $i$ th treatment,  $\bar{R}_i = \frac{R_i}{t_i}$  be the average of the ranks of the observations pertaining to the  $i$ th treatment, and  $\bar{R}$  be the mean of all the  $\bar{R}_i$ . The Kruskal–Wallis test statistic is then given by

$$H = \frac{12}{N(N+1)} \sum_{i=1}^b t_i (\bar{R}_i - \bar{R})^2 \sim \chi_{b-1}^2. \tag{16}$$

Since  $\sum_{i=1}^b R_i = \frac{N(N+1)}{2}$ , it follows that  $\bar{R} = \frac{N+1}{2}$ . Thus, expression (16) reduces to

$$H = \frac{12}{N(N+1)} \sum_{i=1}^b \frac{R_i^2}{t_i} - 3(N+1) \sim \chi_{b-1}^2. \tag{17}$$

Note that the coefficient  $\frac{12}{N(N+1)}$  is known as a suitable normalization factor (see, Manoukian (1986)). The expressions in (16) and (17) are computed if there are no ties in the observations. In the event there are ties, each observation is given the mean of the ranks for which it is tied. The Kruskal–Wallis statistics in (17) is then divided by the correction factor given by

$$cf = 1 - \frac{\sum_{i=1}^k (m_i^3 - m_i)}{N^3 - N},$$

where  $m_i$  refers to the number of ties in  $i$ th group of  $k$  tied groups. Hence, the corrected Kruskal–Wallis test statistic for ties is expressed as

$$H^m = \frac{12}{N(N+1)} \sum_{i=1}^b \frac{R_i^2}{t_i} - \frac{3N(N+1)^2(N-1)}{N(N^2-1) - \sum_{i=1}^k (m_i^3 - m_i)} \sim \chi_{b-1}^2.$$

It is important to note that the correction factor is included when there are ties to increase the value of the test statistics so as to make the results more significant. Furthermore, the Kruskal–Wallis test statistic has a chi-square distribution with  $(b - 1)$  degrees of freedom under the null hypothesis. The test results obtained in **R** are given as:

$$\text{Kruskal – Wallis chi – squared} = 180.11, \text{ df} = 3, \text{ p – value} < 2.2e - 16.$$

The results give  $\chi_{3,\alpha=0.05}^2 = 180.11$  and  $\text{p-value} < 0.05$ , the level of significance. There is very strong evidence to suggest a significant difference in HIV infection between at least one pair of the groups. Since there is a significant difference in HIV infections as the results suggest, a post-hoc analysis is performed to determine which group of the individuals differ from each other in HIV infections. We use Nemenyi test which is appropriate for groups with equal number of observations as in our case (Zar, 2010). The results are presented in Table 4. Since all the p-values are less than 0.05, the level of significance, there are significant differences in HIV infections between the groups.

#### 4.4. Model fitting

The results in Fig. 4 clearly show that the model fits well with the available data points. It is important to observe that the cases of infection peaked in the year 2013. The results show that there was a rise in HIV infection between 2011 and 2013, followed by a significant slow down in the occurrences of new cases of HIV infection. In Fig. 5, we make a comparison of new cases of HIV infection for the two groups. Our results are indicative of a long-term fall in cases of HIV infection in which there is a significant decline in the cases of infection by 2030. However, it can be clearly seen that the occurrence of new cases of HIV infection is more prominent in the adult population as compared to the young adults’ population. The most important observation is there is high number of cases of HIV infection amongst the female adults (aged 25 and over) in comparison to the remaining groups. It is known that women in this group are disproportionately affected by the HIV infections since it is men often dominate sexual relationships leaving women with no ability to always practice safer sex despite the known risks involved (AV, 2017). The results show that new cases of HIV infection amongst the young male adults would be contained by 2025 while that of their female counterparts is likely to be contained after 2030 should the current interventions against HIV in Kenya be maintained. Tables 5–7 give the estimated variable values, estimated parameter values and the transmission reproduction numbers, respectively. The computation of the reproduction numbers within and between age groups in Table 7 provides insights into control that cannot be deduced simply from observations on the prevalence of infection. More specifically, the analysis showed that the per capita rate of HIV transmission was highest when there is interaction between young adults to adults and most HIV infections occurred in adult population.

##### 4.4.1. Sensitivity analysis

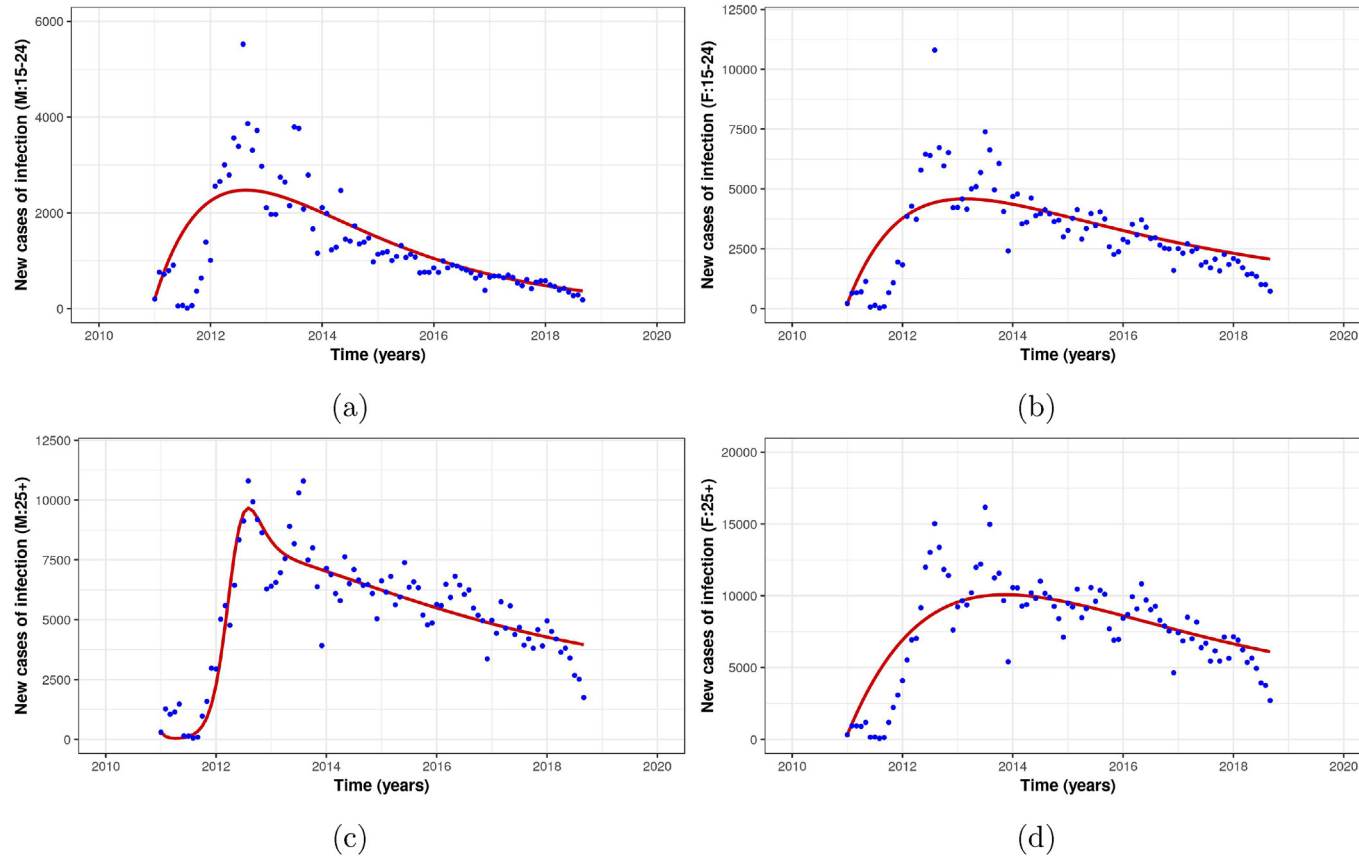
Sensitivity analysis is introduced to study the strength of the basic reproduction numbers as listed in Table 7 for the model parameters. Here, we perform sensitivity analysis to examine the model’s response to parameter variation within a wider range in the parameter space. Following the work by Marino, Hogue, Ray, and Kirschner (2008), partial rank correlation coefficients (PRCC) between the basic reproduction number  $R_0$  and each parameter are derived from 1,000 runs of the Latin hypercube sampling (LHS) method (Stein, 1987). The parameters are assumed to be random variables with uniform distributions with their mean value listed in Tables 1 and 6. Tornado plots for the normalised sensitivity index for different parameters are given in Fig. 6.

If the sensitivity index is positive, then the reproduction number increases along with increasing value of the parameter. On the other hand, if the sensitivity index is negative, then reproduction decreasing with the increasing value of the parameter. Fig. 6a and b are produced assuming that the HIV infection is localised only the young adults (15–24 years) and adults (15 + years) age groups respectively. From the figures, the parameters related to the probabilities of HIV transmission

**Table 4**

Pairwise comparisons using Tukey and Kramer (Nemenyi) test. F-15-24 and M-15-24 means the female and male young adults while F-25 + and M-25 + means the female and male adults, respectively. The lower triangles of the matrices respectively contain the  $\chi^2$  and p-values of the pairwise comparisons.

	$\chi^2$ output				P-value		
	F-15-24	F-25+	M-15-24		F-15-24	F-25+	M-15-24
F-25+	10.978			F-25+	0.0000		
M-15-24	6.828	17.806		M-15-24	0.00001	0.0000	
M-25+	6.377	4.601	13.205	M-25+	0.00004	0.006	0.0000



**Fig. 4.** Model system (5) fitted to data for the reported new cases of HIV infection. 4a shows the model fitted to the data for the young male adults (aged 15–24 years) while 4b shows the model fitted to data for the young female adults (aged 15–24 years). On the other hand 4c shows the model fitted to the data for the male adults (aged 25 + years) while 4d shows the model fitted to data for the female adults (aged 25 + years). The blue dots indicate the actual data and the red line indicates the model fit to the data. All the fitted curves are done with 95% confidence limits.

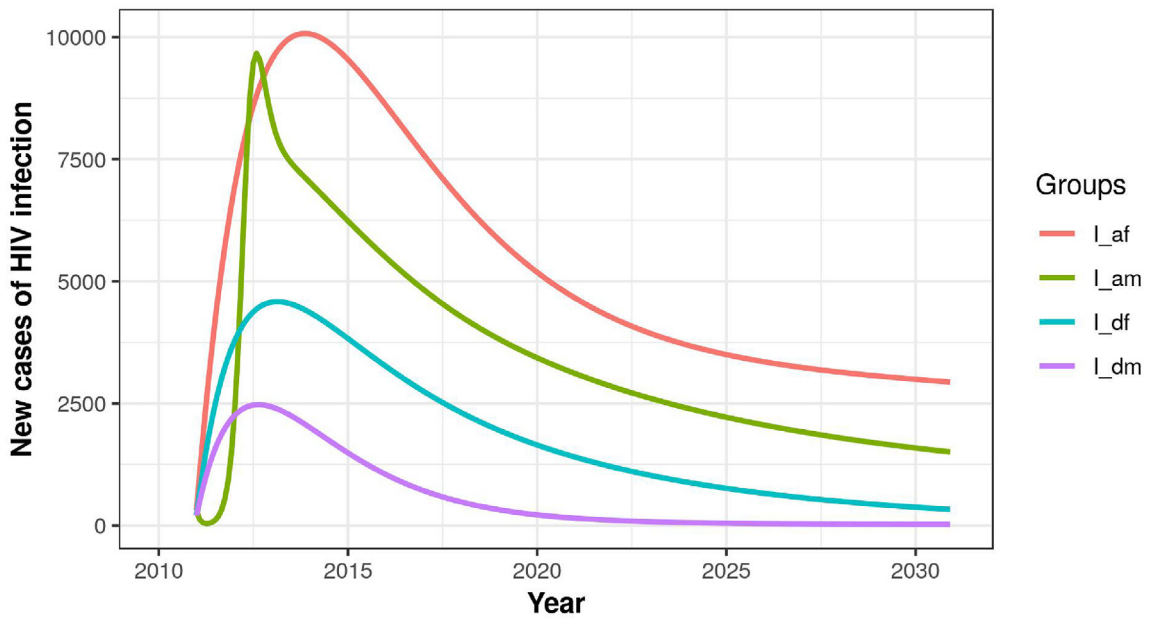


Fig. 5. The projected cases of new HIV infection within the age groups as fitted in Fig. 4.

Table 5

Estimated variable values from the model fitting to data for the period January 2011 to September 2018.

	Male			Female		
	Mean	SE	95% CI <sup>a</sup>	Mean	SE	95% CI <sup>a</sup>
$S_d$	4326140	1665	[4325114, 4327166]	4333384	548.6	[4332309, 4334459]
$I_d$	180	0.4845	[179.05, 180.95]	191	1.766	[187.54, 194.46]
$T_d$	105	0.4625	[104.09, 105.91]	126	0.2417	[125.53, 126.47]
$S_a$	9011930	1418	[9009150, 9014710]	9312839	1417	[9310062, 9315616]
$I_a$	665	1.671	[661.73, 668.28]	370	0.9102	[368.22, 371.78]
$T_a$	144	0.4261	[143.16, 144.84]	333	0.6121	[331.8, 334.2]

<sup>a</sup> 95% Confidence Interval.

Table 6

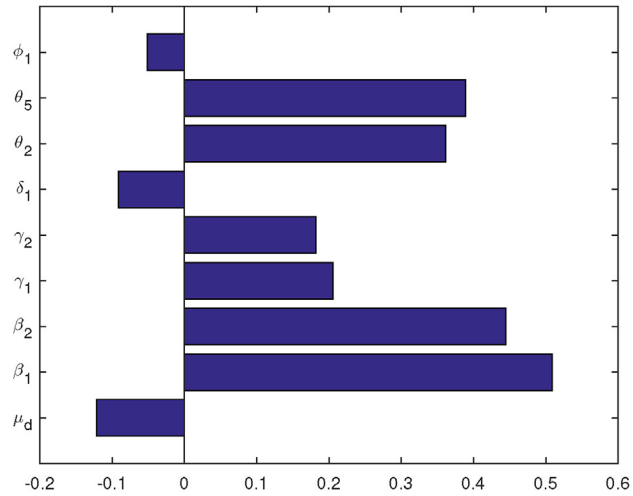
Estimated parameter values of the system (5) obtained from model fitting for the period January 2011 to September 2018.

Par	Mean	SE	95% CI	Par	Mean	SE	95% CI
$\beta_1$	0.3743	7.8e-4	[0.3728, 0.3758]	$\beta_2$	0.401e-3	5.0e-6	[4.0e-3, 4.02e-3]
$\beta_3$	4.2e-5	5.33e-8	[4.23e-5, 4.25e-5]	$\beta_4$	0.7451	0.0015	[0.7421, 0.7481]
$\theta_1$	0.1698	2.86e-4	[0.1693, 0.1704]	$\theta_2$	2.76e-4	4.57e-7	[2.7e-4, 2.8e-4]
$\theta_3$	1.282e-5	8.2e-9	[1.28e-5, 1.29e-5]	$\theta_4$	0.0422	1.301e-4	[0.0419, 0.0425]
$\theta_5$	0.0248	7.531e-5	[0.0246, 0.0250]	$\theta_6$	0.2195	2.712e-4	[0.2189, 0.2202]
$\eta_1$	0.0432	1.167e-4	[0.0429, 0.0434]	$\eta_2$	0.6256	1.0704e-3	[0.6235, 0.6277]
$\eta_3$	0.0680	1.078e-4	[0.0678, 0.0683]	$\eta_4$	5.181e-4	1.181e-6	[5.15e-4, 5.2e-4]
$\eta_5$	5.494e-3	1.735e-5	[5.46e-3, 5.53e-3]	$\eta_6$	0.2169	2.392e-4	[0.2165, 0.2174]

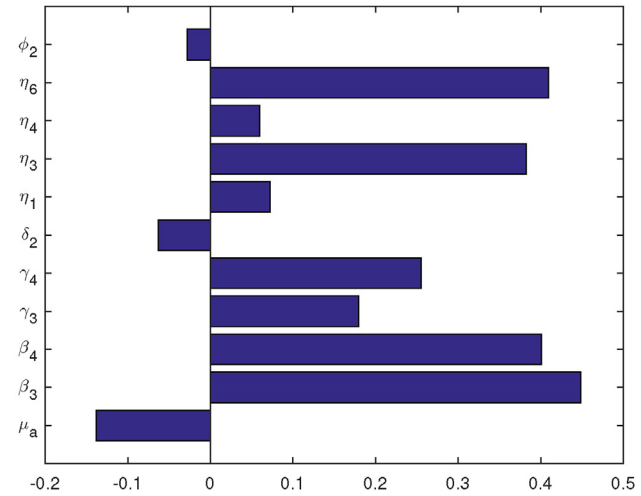
Table 7

Estimation of young adults transmission reproduction number  $\mathcal{R}_{0d}$ , adults transmission reproduction number  $\mathcal{R}_{0a}$ , basic reproduction number between the male young adults and the female adults  $\mathcal{R}_{0mdfa}$ , basic reproduction number between the female young adults and the male adults  $\mathcal{R}_{0fdma}$  and the system (5) basic reproduction number  $\mathcal{R}_0$ .

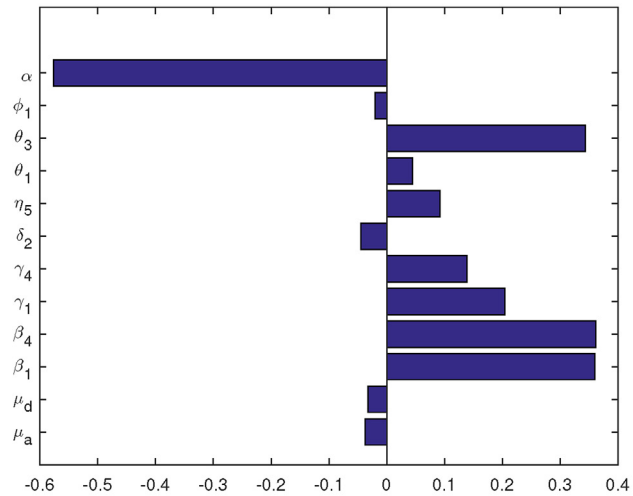
Statistics	$\mathcal{R}_{0d}$	$\mathcal{R}_{0a}$	$\mathcal{R}_{0mdfa}$	$\mathcal{R}_{0fdma}$
Mean	1.135	1.921	2.432	2.432
Std. error	0.000035	0.00014	0.00089	0.00089
95% Confidence Interval	1.131–1.139	1.901–1.941	2.397–2.467	2.397–2.467



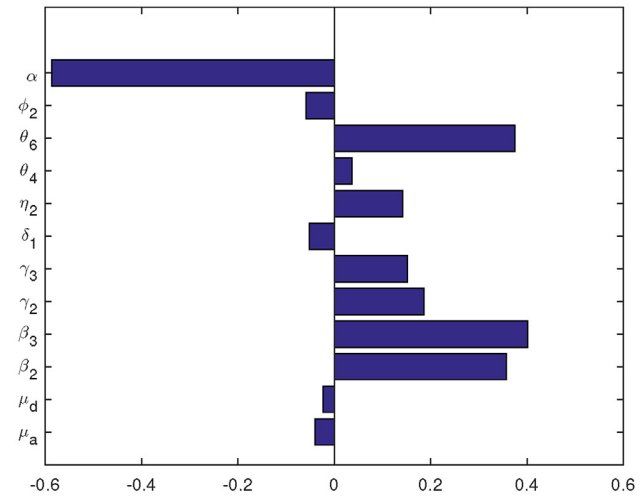
(a)



(b)



(c)



(d)

**Fig. 6.** Tornado plots showing PRCCs for the different parameter values. 6a and 6b are produced assuming that the HIV infection is localised only the young adults (15–24 years) and adults (15 + years) age groups respectively. On the other hand 6c and 6d are produced assuming that there is interaction between young male adults (15–24 years) and adult females (15 + years) and young female adults (15–24 years) and male adults (15 + years), respectively.

have reasonably significant PRCCs and cannot be ignored. The parameters  $\phi_1$ ,  $\delta_1$ ,  $\phi_2$  and  $\delta_2$  have the lowest PRCCs with respect to the corresponding disease thresholds. However, their direction of influence is clearly visible. In this regard, since no effort toward reducing disease spread is rendered insignificant, any action that increases the number of individuals under ART treatment reduces the infection. Fig. 6c and d are produced assuming that there is interaction between young male adults (15–24 years) and adult females (15 + years) and young female adults (15–24 years) and male adults (15 + years), respectively. It is also seen that probabilities of HIV transmission have the potential of making the epidemic worse if increased while parameters related to treatment of infected individuals into ART have the potential of reducing infections.

## 5. Conclusion

The HIV epidemic has been evolving in Kenya since the detection of the first case in 1984. Kenya has the fourth highest number of HIV infection globally alongside South Africa and Nigeria (UNAIDS, 2015). The number of people living with HIV was estimated to be 1.6 million in the year 2015 with 36,000 deaths resulting from AIDS-related illness (OPTIONS, 2016; UNAIDS, 2015). In its efforts to reduce HIV spread, HIV testing and counselling has been adopted through targeted community-based testing and door-to-door testing initiatives and use of preventive measures such as condoms encouraged. In 2016, an estimated 64% of the people living with HIV were receiving antiretroviral treatment (ART) of whom 51% were virally suppressed (UNAIDS, 2017b). Furthermore, according to UNAIDS (2017b), 73% of men and 55% of women used a condom the last time they engaged in sex with a non-marital as well as non-cohabiting partner. Despite the intense and aggressive interventions against HIV, there is still growing number of new infections in Kenya especially amongst the young adults. Thus, in this work, we attempt to model the trend of new HIV infections in Kenya, for which a considerable amount of data is available. A deterministic model for HIV dynamics within and between age groups that takes into consideration the sexual orientation of individuals is presented. Vital mathematical characteristics of the model have been presented. These include the invariant region of biological significance, the age group specific basic transmission numbers and inter age group specific basic transmission numbers. MCMC method has been used to estimate the parameter values based on the available data. The basic descriptive and inferential statistics of the data have been computed and presented. Our analysis of the data shows that females in both age categories are disproportionately affected with HIV more than the males. This is supported by the Kruskal-Wallis results which are indicative of very strong evidence that there exist significant differences in HIV infections between the groups.

The model was then fitted to on the new cases of HIV infections with the objective of using the model parameters that give the best fit to examine the trend of HIV infection. It has been established that the occurrence of new cases of HIV infection is more prominent in the adult population as compared to the young adults' population. It is important to note that there is high number of cases of HIV infection amongst the female adults (aged 25 and over). This can be attributed to the fact that men often dominate sexual relationships leaving women with no ability to always practice safer sex despite the known risks involved. The results show that new cases of HIV infection amongst the young male adults would be contained by 2025 while that of their female counterparts is likely to be contained after 2030 should the current interventions against HIV in Kenya be maintained. Furthermore, computation of the reproduction numbers within and between age groups provides insights into control that cannot be deduced simply from observations on the prevalence of infection. More specifically, the analysis showed that the per capita rate of HIV transmission was highest when there is interaction between young adults to adults and most HIV infections occurred in adult population.

Sensitivity of parameters was also considered. The results demonstrate that the transmission probabilities and treatment rates have the greatest impacts on the reproduction numbers. This suggests that control of HIV pivots around transmission prevention programs. Programs aimed at individuals at high risks of HIV infection that encourage them to use preventive measures such as condoms and PrEP will be particularly effective. Furthermore, enrolling more infected individuals on ART treatment would be ideal in reducing the cases of new infections for it is known that it helps in suppressing the viral load in the body thus limiting further HIV infections. It is thus critical to devote more resources to education on HIV preventive measures and treatment programs that are especially targeted to both the susceptible and infected individuals.

The model considered in this paper is consistent with the dynamics of HIV infection in Kenya and it has some lucid limitations. In fact, lack of sufficient data on the number of HIV patients enrolled in ART treatment and care limited the numerical analysis and interpretation. This work has only considered the new cases of HIV infections. It is well known that the goodness of fit measures the discrepancy between observed data and values expected from the model. In this work, no goodness of fit tests were performed. However, we relied on the MCMC method for the model fitting. We argue that MCMC method of fitting models to data provides useful insights into how the model can be linked to data despite the challenge of using statistical tools to test the goodness of fit of the model.

## Author contributions

All authors conceived and developed the study. All authors read and approved the final version of the manuscript.

## Disclosure statement

The authors declare that there are no competing interests regarding the publication of this paper.

## Funding

The authors received no direct funding for this research.

## Acknowledgment

The authors acknowledge, with thanks, the support of Strathmore University, Institute of Mathematical Sciences for the production of this manuscript.

## References

- Abate, A., Tiwari, A., & Sastry, S. (2009). Box invariance in biologically-inspired dynamical systems. *Automatica*, 45(7), 1601–1610.
- Aldila, D. (2018). Mathematical model for HIV spreads control program with ART treatment. In *Journal of physics: Conference series* (vol. 974, p. 012035). IOP Publishing.
- AV. (2017). *Avert. Global information and education on HIV and AIDS: HIV and AIDS in Kenya*. Retrieved June, 2018 from <https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/kenya>.
- Berman, A., & Plemmons, R. J. (1994). *Nonnegative matrices in the mathematical sciences* (vol. 9). Philadelphia: SIAM.
- Chow, Y., Leong, C., Chow, H., & Hooi, L. (2007). Lactic acidosis in hiv patients receiving highly active antiretroviral therapy. *Medical Journal of Malaysia*, 62(1), 78.
- Diekmann, O., & Heesterbeek, J. A. P. (2000). *Mathematical epidemiology of infectious diseases: Model building, analysis and interpretation* (vol. 5). John Wiley & Sons.
- Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. (1990). On the definition and the computation of the basic reproduction ratio  $r_0$  in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4), 365–382.
- KD. (2018). *Kenya demographics profile 2018*. Retrieved May, 2018 from [https://www.indexmundi.com/kenya/demographics\\_profile.html](https://www.indexmundi.com/kenya/demographics_profile.html).
- Keeling, M. J., & Rohani, P. (2011). *Modeling infectious diseases in humans and animals*. Princeton University Press.
- Kim, S. B., Yoon, M., Ku, N. S., Kim, M. H., Song, J. E., Ahn, J. Y., et al. (2014). Mathematical modeling of HIV prevention measures including pre-exposure prophylaxis on HIV incidence in South Korea. *PLoS One*, 9(3), e90080.
- KNBS. (2014). *Kenya demographic and health survey 2014*. Retrieved May, 2018 from <https://dhsprogram.com/pubs/pdf/FR308/FR308.pdf>.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583–621.
- Manoukian, E. B. (1986). *Mathematical nonparametric statistics*.
- Marino, S., Hogue, I. B., Ray, C. J., & Kirschner, D. E. (2008). A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*, 254(1), 178–196.
- MOH. (2016). *Kenyan ministry of health/national AIDS control council. Kenya AIDS response progress report 2016*. Retrieved May, 2018 from [https://nacc.or.ke/wp-content/uploads/2016/11/Kenya-AIDS-Progress-Report\\_web.pdf](https://nacc.or.ke/wp-content/uploads/2016/11/Kenya-AIDS-Progress-Report_web.pdf).
- Mukandavire, Z., Chiyaka, C., Garira, W., & Musuka, G. (2009). Mathematical analysis of a sex-structured HIV/AIDS model with a discrete time delay. *Nonlinear Analysis: Theory, Methods & Applications*, 71(3–4), 1082–1093.
- Mukandavire, Z., & Garira, W. (2007). Age and sex structured model for assessing the demographic impact of mother-to-child transmission of HIV/AIDS. *Bulletin of Mathematical Biology*, 69(6), 2061–2092.
- Omondi, E., Mbogo, R., & Luboobi, L. (2018a). Mathematical analysis of sex-structured population model of hiv infection in Kenya. *Letters in Biomathematics*, 5(1), 174–194.
- Omondi, E., Mbogo, R., & Luboobi, L. (2018b). Mathematical modelling of the impact of testing, treatment and control of HIV transmission in Kenya. *Cogent Mathematics & Statistics*, 1475590.
- Omondi, E., Mbogo, R., & Luboobi, L. (2018c). Modelling the trend of HIV transmission and treatment in Kenya. *International Journal of Applied and Computational Mathematics*, 4(5), 123.
- OPTIONS. (2016). *OPTIONS country situation analysis interim findings: Kenya*. FSG in partnership with LVCT Health. [http://www.prepwatch.org/wp-content/uploads/2016/05/Situation\\_Analysis\\_Kenya.pdf](http://www.prepwatch.org/wp-content/uploads/2016/05/Situation_Analysis_Kenya.pdf) Accessed August 2016.
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, Kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Sardar, T., Sasmal, S. K., & Chattopadhyay, J. (2016). Estimating dengue type reproduction numbers for two provinces of Sri Lanka during the period 2013–14. *Virulence*, 7(2), 187–200.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611.
- Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2), 143–151.
- UNAIDS. (2015). *UNAIDS. HIV and AIDS estimates*. <http://www.unaids.org/en/regionscountries/countries/kenya> Accessed April 2017.
- UNAIDS. (2016). *Prevention gap report*. Retrieved May, 2018 from [http://www.unaids.org/sites/default/files/media\\_asset/2016-prevention-gap-report\\_en.pdf](http://www.unaids.org/sites/default/files/media_asset/2016-prevention-gap-report_en.pdf).
- UNAIDS. (2017a). *Global information and education on HIV and AIDS*. Retrieved June, 2018 from [https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/kenya#footnote41\\_8o0cqle](https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/kenya#footnote41_8o0cqle).
- UNAIDS. (2017b). *UNAIDS Fact sheet. Fact sheet – latest statistics on the status of the AIDS epidemic*. <http://www.unaids.org/en/resources/fact-sheet> Accessed April 2017.
- Van den Driessche, P., & Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180(1), 29–48.
- WH. (2018). *World health rankings. Live longer live better*. Retrieved May, 2018 from <http://www.worldlifeexpectancy.com/kenya-life-expectancy>.
- WHO. (2014). *March 2014 supplement to the 2013 consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: Recommendations for a public health approach*. Geneva: World Health Organization.
- WHO. (2018). *WHO. HIV/AIDS-Global health observatory (GHO) data*. Retrieved October, 2018 from <http://www.who.int/gho/hiv/en/>.
- Williams, B. G. (2014). *Optimizing control of HIV in Kenya*. arXiv:1407.7801.
- Zar, J. H. (2010). *Biostatistical analysis* (5th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.