



Article

Trace the History of HIV and Predict Its Future through Genetic Sequences

Zhen Wang¹, Zhiyuan Zhang², Chen Zhang³ , Xin Jin¹, Jianjun Wu⁴, Bin Su⁴, Yuelan Shen⁴, Yuhua Ruan⁵, Hui Xing⁵ and Jie Lou^{1,*}

¹ Department of Mathematics, Shanghai University, Shanghai 200444, China

² Department of Statistics, Columbia University, New York, NY 10027, USA

³ School of Nursing, University of Rochester, Rochester, NY 14627, USA

⁴ Anhui Provincial Center for Disease Control and Prevention, Hefei 230601, China

⁵ State Key Laboratory for Infectious Disease Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China

* Correspondence: jie_lou@shu.edu.cn

Abstract: Traditional methods of quantifying epidemic spread are based on surveillance data. The most widely used surveillance data are normally incidence data from case reports and hospital records, which are normally susceptible to human error, and sometimes, they even can be seriously error-prone and incomplete when collected during a destructive epidemic. In this manuscript, we introduce a new method to study the spread of infectious disease. We gave an example of how to use this method to predict the virus spreading using the HIV gene sequences data of China. First, we applied Bayesian inference to gene sequences of two main subtypes of the HIV virus to infer the effective reproduction number (${}_C\mathcal{R}_e(t)$) to trace the history of HIV transmission. Second, a dynamic model was established to forecast the spread of HIV medication resistance in the future and also obtain its effective reproduction number (${}_M\mathcal{R}_e(t)$). Through fitting the two effective reproduction numbers obtained from the two separate ways above, some crucial parameters for the dynamic model were obtained. Simply raising the treatment rate has no impact on lowering the infection rate, according to the dynamics model research, but would instead increase the rate of medication resistance. The negative relationship between the prevalence of HIV and the survivorship of infected individuals following treatment may be to blame for this. Reducing the MSM population's number of sexual partners is a more efficient strategy to reduce transmission per the sensitivity analysis.

Keywords: HIV/AIDS; Bayesian phylogenetic method; effective reproductive number; primary drug resistance; transmitted drug resistance; dynamic model



Citation: Wang, Z.; Zhang, Z.; Zhang, C.; Jin, X.; Wu, J.; Su, B.; Shen, Y.; Ruan, Y.; Xing, H.; Lou, J. Trace the History of HIV and Predict Its Future through Genetic Sequences. *Trop. Med. Infect. Dis.* **2022**, *7*, 190. <https://doi.org/10.3390/tropicalmed7080190>

Academic Editor: Omar Bagasra

Received: 9 July 2022

Accepted: 11 August 2022

Published: 17 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

AIDS (Acquired Immunodeficiency Syndrome) is a disease caused by HIV (Human Immunodeficiency Virus) and has become a disease that seriously threatens human health in the world today. In order to alleviate the AIDS epidemic in China, the Chinese government began to implement the “Four Frees and One Care” HIV/AIDS free treatment policy in 2003 [1,2]. This policy has greatly improved the patient's survival rate and the patient's life quality. On the other hand, the high mutation rate of HIV virus has led to the generation of HIV-resistant strains [3]. The spread of drug-resistant strains causes a waste of national medical resources.

We want to know the effect of the free treatment policies, especially the drug resistance situation of HIV/AIDS by dynamic models. It is hard to avoid parameter acquisition when performing prediction with dynamics models. Some parameters are general parameters, which are common to both China and abroad. We can assign these values by searching relative literature. Some might be greatly different due to different countries or different

treatment strategies, so they need be fitted by monitoring data. However, such areas happen to lack monitoring data except for some gene sequences of HIV patients being detected before treatment. In this manuscript, we will show how to combine the gene sequences of HIV and the dynamic model to study the dynamics of HIV spreading.

HIV is an RNA virus. RNA viruses have a short reproduction time and a high mutation rate, which can easily lead to a large number of genetic mutations during the transmission process. Therefore, during the spread of HIV, relevant signals will be left in the sampled virus sequence. In recent years, researchers have begun to conduct retrospective studies on the spread of the HIV virus on the genetic level. Most of these studies are based on genetic sequences to build molecular networks of transmission using Bayesian evolution analysis [4–12]. For example, in 2008, Lewis et al. used the Bayesian framework to conduct a phylogenetic analysis of 2126 HIV gene sequences in London, and they constructed a network for HIV transmission, revealing that the local HIV/AIDS epidemic can be traced back to the late 1990s [13]. In 2018, Zhao et al. constructed a molecular network of the spread of MSM in Beijing based on the HIV gene sequence, and they proposed the need to strengthen education and intervention for people with potential high risk. These findings have important implications for the parameterization of epidemiological models and the design of intervention strategies.

Traditional dynamic models can be used to predict the spread of virus in populations [14–22], and the Bayesian evolution analysis of virus separated from the infected people can trace the historical dynamics of virus transmission [13,23,24]. However, the work connecting the gene sequences and the dynamic model has not yet been seen. This manuscript will use Anhui province as an example, combine genetic sequence molecular evolution analysis and system dynamics models to conduct an in-depth research on the HIV/AIDS epidemic in the content of the HIV-free treatment policy in China, and thus predict the drug resistance spreading in the next few years.

2. Research Methodology

2.1. Gene Sequence Analysis

To trace the history of transmission dynamics of HIV, we sampled totally 360 HIV-*pol* gene sequences (CRF 01AE (150), CRF 07BC (190), CRF 08BC (9) and CRF 5501B (11)) from 360 HIV patients who were infected with HIV through sexual behavior in Anhui Province. This data set is sampled between October 2017 and September 2018. Considering that the proportions of the CRF 08BC and CRF 5501B are too small, we only analyzed the subtypes CRF 01AE and CRF 07BC, comprising a total of 340 HIV virus gene sequences. We applied Mega for sequence alignment. The total length of the *pol* gene that we studied was 1075 bp.

The selection of a nucleotide substitution model is an important step during the phylogenetic analysis. Common nucleotide substitution models include the HKY (+ Γ + I) model, GTR (+ Γ + I) model, and so on [25]. These different substitution models determined different evolutionary distances and different phylogenetic trees [26]. Here, we used software Phylosuite to select the most suitable substitution model according to the Akaike Information Criterion (AIC) [27], where $AIC = 2k - 2\ln(L)$, in which k is the number of corresponding parameters and L is the likelihood value. Method-of-moments was used to estimate the AIC method. The model with lower AIC value is the more suitable model. After calculating the AIC score, the HKY+ Γ + I model and the uncorrelated exponential relaxed molecular clock are determined finally.

2.1.1. Bayesian Inference

The birth death skyline (BDSKY) model in BEAST2 is a random description of population change, which allows the extraction of virus transmission dynamics information from a phylogenetic tree [11]. This model is established on Bayesian inferences, which can estimate the effective reproduction number of HIV/AIDS transmission from gene sequences. The

Bayesian approach allows us to continuously update our estimate of a set of parameters, Θ , as data become available.

$$P(\Theta|data) = \frac{P(data|\Theta) \cdot P(\Theta)}{P(data)}.$$

$P(\Theta)$, the prior distribution, represents our prior estimates about the true value of Θ . $P(data|\Theta)$ is the likelihood distribution. It is also often written as $\mathcal{L}(\Theta|data)$, which means the probability of observing the data given Θ . For the method to work, it is necessary to calculate the likelihood distribution for all possible values of Θ . $P(data)$ is the model evidence, and it is the same for all possible hypotheses (values of Θ) being considered. $P(\Theta|data)$ is the posterior distribution and represents our updated estimate of the value of Θ given the observed data.

The main objective of Bayesian inference is to calculate the posterior distribution of our parameters using our prior beliefs updated with our likelihood. From the posterior distribution, we can determine the most likely values of Θ given the observed data. Since we are usually only interested in relative probabilities of different hypotheses, $P(data)$ can be left out of the calculation, and we write the model form of Bayes' theorem as

$$P(\Theta|data) \propto P(data|\Theta) \cdot P(\Theta)$$

where \propto means "proportional to".

2.1.2. The Effective Reproduction Number ${}_G\mathcal{R}_e(t)$ Inferred from Gene Sequences

Since we will study two different subscript gene sequences, we use $j = 07BC$ for the subtype CRF 07BC and $j = 01AE$ for the subtype CRF 01AE. We will deduce the propagation dynamics of the two subtypes separately. For each subtype j , the BDSKY process can conclude three parameters: transmission rate λ_j , removal rate γ_j and sampling probability ψ_j . This process allows individuals to "birth" (λ_j) or "death" (γ_j and ψ_j) at any point in time. A "birth event" corresponds to an individual's infection, and a "death event" corresponds to an infected individual becoming non-infected (i.e., removed from the infected compartment because of individual death, successful treatment, or individual behavior change). Then, the effective reproduction number $\mathcal{R}_e(t)$ derived from the gene sequences can be defined as [11]

$${}_G^j\mathcal{R}_e(t) = \lambda_j(t) / (\mu_j(t) + \psi_j(t)),$$

in which the left-hand subscript G denotes that the effective reproduction number is derived from the gene sequences. For estimating ${}_G^j\mathcal{R}_e(t)$ of subtype $j = 07BC$ or $01AE$, the Bayes' theorem that we use is

$$P({}_G^j\mathcal{R}_e(t)|gene\ sequences) \propto P(gene\ sequences|{}_G^j\mathcal{R}_e(t)) \cdot P({}_G^j\mathcal{R}_e(t)). \quad (1)$$

In order to catch the characteristics of transmission in different historical stages, we divided the period from the origin of each subtype HIV sequences to the sampling time of the last sample into a total of six segments. Then, from each subtype gene sequence of $j = 07BC$ and $01AE$, we can infer the effective reproduction number ${}_G^j\mathcal{R}_e(t)$ during each segment t , $t = 1, 2, \dots, 6$ through Equation (1). Hence, ${}_G^j\mathcal{R}_e(t)$ is a piece wise function for each of the two subtypes, which is a constant during the same time segment t . It refers to the average number of new infections caused by an infected person at a certain time point t during the outbreak. We use the Markov Chain Monte Carlo (MCMC) method to achieve Bayesian phylogeny inference.

2.2. Transmission Dynamics Model

2.2.1. A Nested Model and Its Four Submodels

To predict the prevalence of HIV drug-resistant strains, we considered the transmission dynamics of HIV/AIDS through sexual route, including gay, straight and bisexual. The whole population was divided into four subgroups: heterosexual women who only have sex with men (marked as w), heterosexual men who only have sex with women (marked with mw), homosexual men who only have sex with men (marked with mm), and bisexual men who have sex with both women and men (marked as b). The population was divided into susceptible people S , infected people I (infected but not receiving treatment), treated people T , and drug-resistant people R according to the infection. Considering that the HIV infection process has obvious stage characteristics, according to the level of CD4 cells in the patient's body, we divide the infected people into three stages: the first infection stage I_1 (CD4 count ≥ 500 cells/mm³); the second infection stage I_2 (200 cells/mm³ < CD4 counts < 500 cells/mm³); and the third stage of infection I_3 (AIDS stage, CD4 count ≤ 200 cells/mm³). Similarly, T_i and R_i , $i = 1, 2, 3$ can be explained. The details of the model can be found in Figure 1. When the superscript g in the transmission process diagram Figure 1 is taken as w, mw, mm and b , it represents the four types of subgroups. See Appendix A.1 for the corresponding dynamic model.

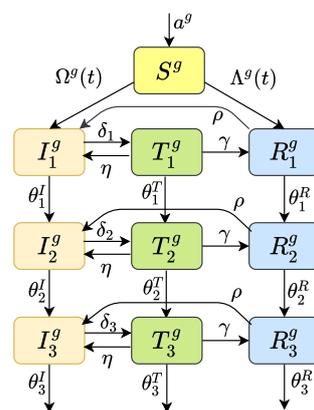


Figure 1. Transmission process diagram. The superscript g can be w, mw, mm and b , respectively, which represents the four types of subgroup populations. Parameter a is an input of susceptible individuals; $\theta_i, i = 1, 2, 3$ are the disease progression rates in infection stage 1, 2 and 3, respectively; $\delta_i, i = 1, 2, 3$ are the treatment rates of patients in stage 1, 2 and 3, respectively; η is the percentage of patients who give up HAART; γ is rate of drug resistance after HAART; ρ is the percentage of patients who are no longer resistant.

The standard treatment of HIV/AIDS in China has gone through four periods: the period before 2002 featured no treatment; the period from 2003 to 2011 was when the treatment was started when the CD4 count was less than 200 cells/mm³; the period from 2012 to 2015 was when the treatment was started when the CD4 count was less than 500 cells/mm³ and the period after 2015 was when the treatment was started soon after discovery. We call these periods I (1992–2002), II (2003–2011), III (2012–2015) and IV (2016–now), respectively.

Due to different prevention and control policies being held in different historical stages, the dynamics of HIV transmission have different stage characteristics. So, our model (Figure 1) in fact is a nested model. For the four different historical treatment periods I, II, III and IV, we can simplify the model to be four submodels by letting some parameters be zero based on its history characters; see Table 1 for the four different submodels. According to the “next-generation operator” method in the literature [28], we can obtain the basic reproduction number ${}^k_M \mathcal{R}_0$ for each submodel of corresponding historical period $k = I, II, III$ and IV (see Appendix A.2). Here, the left-hand subscript M means that the

basic reproduction number ${}^k_M\mathcal{R}_0$ was obtained from $k = I, II, III$ or IV , one of the four models respectively.

Table 1. Submodels of different historical periods and corresponding assumptions.

Submodel	Assumptions	Historical Period
I	$\delta_i = 0$ and $T_i^g = 0, R_i^g = 0, i = 1, 2, 3, g = w, mw, mm, b$	1992–2002
II	$\delta_i = 0$ and $T_i^g = 0, i = 1, 2, g = w, mw, mm, b$	2003–2011
III	$\delta_1 = 0$ and $T_1^g = 0, g = w, mw, mm, b$	2012–2015
IV	The full model	2016–now

2.2.2. The Effective Reproduction Number ${}_M\mathcal{R}_e(t)$ Inferred from the Dynamic Submodels

For a general dynamic model of an infectious disease, the basic reproduction number \mathcal{R}_0 is the expected number of infections directly generated by one case given that all individuals are equally susceptible. As the infection spreads, the susceptibility of the population decreases. The effective reproduction number, $\mathcal{R}_e(t)$, is related to the basic reproduction number \mathcal{R}_0 , by $\mathcal{R}_e(t) = \mathcal{R}_0 \cdot S(t)/N(t)$, where $S(t)/N(t)$ is the average susceptibility of the population. $\mathcal{R}_e(t)$ is often used as an indicator of the effectiveness of interventions, such as social distancing measures, to contain the spread of a virus. If $\mathcal{R}_e(t)$ is greater than 1.0, the infection is growing at an exponential rate. If $\mathcal{R}_e(t)$ is at 1.0, the spread is sustained at a linear rate. If $\mathcal{R}_e(t)$ is less than 1.0, the infection is spreading at a slower pace and will eventually die out.

Specific to our study on HIV/AIDs spreading during four different historical periods $k = I$ (1992–2002), II (2003–2011), III (2012–2015), and IV (2016–now), let us make

$${}^k_M\mathcal{R}_e(t) = {}^k_M\mathcal{R}_0 \frac{S^k(t)}{N^k(t)} \tag{2}$$

where the basic reproduction number ${}^k_M\mathcal{R}_0$ was obtained from Section 2.2.1 and $S^k(t)$ and $N^k(t)$ represent the susceptible population and the total target population of each historical period $k = I, II, III$ and IV , respectively [29].

3. Results

3.1. Tracing Back the Dynamic History of HIV/AIDS through Bayesian Inference

The MCMC converged after 20 million iterations for both subtypes. We discarded the first 2 million iterations. The phylogenetic tree and model parameters were sampled every 1000 iterations. All parameter estimates showed the effective sample size (ESS) of more than 200. Some results are shown in Table 2.

Table 2. Epidemiological parameter estimations and their 95% HPD of subtype CRF 01AE and CRF 07BC.

Subtype		MRCA	The Effective Reproduction Number $j = 01AE, 07BC$					
		Median	${}^j_G\mathcal{R}_e(1)$	${}^j_G\mathcal{R}_e(2)$	${}^j_G\mathcal{R}_e(3)$	${}^j_G\mathcal{R}_e(4)$	${}^j_G\mathcal{R}_e(5)$	${}^j_G\mathcal{R}_e(6)$
CFR01AE	95% HPD lower	1982	1.64	1.47	1.42	1.37	1.23	2.17
	95% HPD upper	1973	1.08	0.97	0.94	0.91	0.83	1.61
		1990	2.30	2.08	2.00	1.94	1.72	2.87
CFR07BC	95% HPD lower	1992	1.58	1.52	1.51	1.50	1.40	2.21
	95% HPD upper	1976	1.15	1.11	1.10	1.10	1.03	1.72
		1998	2.10	2.00	1.98	1.97	1.81	2.75

For each data set and substitution model, we analyze them with an uncorrelated exponential relaxed molecular clock and a strict clock model. In the study of virus evolution, we used nucleotide substitution models including the HKY (+ Γ + I) model and GTR (+ Γ + I) model [25]. The most appropriate combination was selected according to the Akaike Information Criterion (AIC) [27], where AIC was defined as $AIC = 2k - 2\ln(L)$, in which k is the number of corresponding parameters and L is the likelihood value. Method-of-moments was used to estimate the AIC method. Finally, the uncorrelated exponential relaxed molecular clock and HKY+ Γ + I model is determined by calculating the AIC value. The important epidemiological parameter estimation and their 95% HPD of subtype CRF 01AE and CRF 07BC are shown in Table 2.

The results in Table 2 show that the most recent common ancestor (MRCA) to the 150 CRF 01AE subtype gene sequences was around 1982, and the 95% highest posterior density (HPD) is between 1973 and 1990. Correspondingly, the MRCA to these 190 CRF 07BC subtype gene sequences was relatively late, around 1992 (95% HPD: 1976–1998). Most importantly, we obtained the effective reproduction number ${}^j_G\mathcal{R}_e(t)$ over time segment t from the two subtypes, respectively, in which $t = 1, 2, \dots, 6$ represents the time segment. The effective reproduction number ${}^j_G\mathcal{R}_e(t)$ of the two subtypes showed some similar transmission characteristics in terms of time: both of them were in a relatively stable state in the early stage, kept around 1.6 with slight fluctuations, and showed a slight downward trend from 2008 to 2012. However, things changed a lot around 2012: for some reason, the effective reproduction number increased rapidly and then remained at about 2.2.

When we study the history of HIV transmission dynamics in the entire population through the two effective reproduction numbers ${}^j_G\mathcal{R}_e(t)$, $j = 01AE, 07BC$, $j = 1, 2$, we will have to integrate them together. The two effective reproduction numbers originated at different times but ended at the same time. We took the common historical stage of them. So, the research period was from 1992 to 2018. During this common historical period, we denote the average of the effective reproduction number for the entire population as ${}_G\mathcal{R}_e(t)$, which changes along with time t .

In order to figure out the formula for ${}_G\mathcal{R}_e(t)$, we used the proportion of the number of sequences of each subtype in the total gene sequence as its weight in the effective reproduction number, since subtypes widely disseminated in the population should have more contribution for ${}_G\mathcal{R}_e(t)$. Let M_j be the sequence number of gene subtype j ($j = 01AE, 07BC$). Then, the weighted average effective reproduction number of the entire population at time t can be described by Formula (3).

$${}_G\mathcal{R}_e(t) = \frac{\sum_j {}^j_G\mathcal{R}_e(t)M_j}{\sum_j M_j}, \quad (j = 01AE, 07BC; t = 1992, 1993, \dots, 2018) \quad (3)$$

Use R software to visualize ${}_G\mathcal{R}_e(t)$; then, the result can be found in Figure 2. Figure 2 shows the median line of ${}_G\mathcal{R}_e(t)$ and its 95% HPD interval from 1992 to 2018. Before 2011, we can see that the epidemic showed a relatively stable transmission trend through sexual route. The value of ${}_G\mathcal{R}_e(t)$ remained stable and slightly decreased, with a median value of about 1.33 (95% HPD: 0.94–1.8). However, from 2011 to 2014, ${}_G\mathcal{R}_e(t)$ rose rapidly to 2.13 (95% HPD: 1.32–2.83) for some reason. After 2015, ${}_G\mathcal{R}_e(t)$ remained stable, with its median value maintained at approximately 2.20 (95% HPD: 1.67–2.85). Some explanations will be given later on the possibility of ${}_G\mathcal{R}_e(t)$ rising in this period.

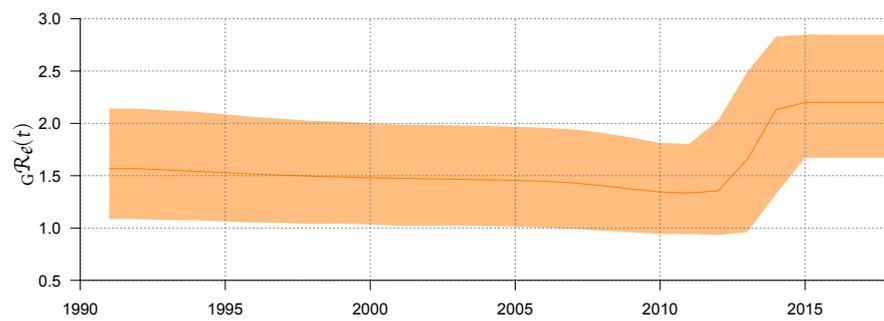


Figure 2. The weighted average of the effective reproduction number ${}_G\mathcal{R}_e(t)$ and its 95% HPD interval.

3.2. Dynamic Model Analysis: Parameter Estimations and Predicting the Transmission of HIV/AIDS

3.2.1. Parameter Estimations

As we all know, an important step in the prediction of infectious disease spreading is to obtain (or estimate from data) parameters values. For most of the general parameters we treated them as constants from references. Details of the values and references can be found in Appendix A.4. However, there are still some epidemiological related parameters that we are not sure of and need to be estimated from data.

Unlike novel coronavirus patients, HIV-infected people do not have obvious infection symptoms. Many HIV infected people do not know that they have been infected for a long time. This prevents the government from obtaining information about people living with HIV as early as possible. Moreover, the Chinese government's screening efforts for HIV-infected people are far less than those for infected novel coronavirus. As a result, recorded HIV infection data, if available, may not be very accurate and may not be suitable for parameter fitting. In Anhui province, which we studied, there is not much surveillance data on HIV/AIDS that can be used. One form of reliable data available are 340 genetic sequences of HIV-infected people taken before treatment.

In the previous subsections, we have extrapolated the dynamics of HIV transmission (${}_G\mathcal{R}_e(t)$) from 1992 to 2018 using the 340 genetic sequences. In order to use these gene data to estimate unknown parameters of the model (Figure 1, which consists of four different submodels I (1992–2002), II (2003–2011), III (2012–2015) and IV (2016–now), depending on the historical period from 2003–now), we calculated the effective reproduction number ${}_M^k\mathcal{R}_e(t)$ for each submodel $k = I, II, III$ and IV by Equation (2). Obviously, ${}_M^k\mathcal{R}_e(t)$ is functions of parameters and time t . Since the two effect reproduction numbers obtained through different methods should be the same at the same time t , i.e., ${}_M\mathcal{R}_e(t) = {}_G\mathcal{R}_e(t)$, we fit them at the same historical period. In this way, we explored the humanistic and statistical features in different historical stages. We used MCMC to realize this. The algorithm ran for 2.00 million iterations, and we adapted the proposal distribution after 2 million iterations using Geweke's method [30] to assess convergence. The fitted curves of the two effective reproduction numbers for four historical periods I, II, III and IV are shown in Figure A1 in Appendix A.4. The fitted parameters obtained from the above four periods are shown in Table 3. The value of these parameters will be referred to as baseline parameter values in the following research.

Table 3. Model parameters and the 95% HPD fitted by the two effective reproduction numbers.

Parameters	Description	Value	Fitted from Submodel
a^w	input of heterosexual women	343780	IV
a^{mw}	input of heterosexual men	358200	IV
a^{mm}	input of homosexual men	4974 (2936, 7012)	IV
a^b	input of bisexual men	4646 (2662, 6630)	IV
c^w	the average number of sexual partners per year for heterosexual women	1.9957 (0.846, 3.1454)	IV
c^{mw}	the average number of sexual partners per year for heterosexual men	2.0946 (0.8514, 3.1742)	IV
c^{mm}	the average number of sexual partners per year for homosexual men	5.4984 (2.0701, 7.8565)	IV
c^b	the average number of sexual partners per year for bisexual men	5.5765 (2.1816, 8.3334)	IV
ϵ	transmissibility ratio of drug-resistant to untreated wild strain carriers	0.43983	II
$1/\theta_1^T$	course of the disease after treatment in stage 1	7.4030 (6.8823, 8.0000)	IV
$1/\theta_2^T$	course of the disease after treatment in stage 2	10.8948 (10.0000, 12.5000)	III
$1/\theta_3^T$	course of the disease after treatment in stage 3	4.1969(3.3333, 4.7619)	II
δ_1	treatment rate of stage 1	0.50188 (0.2155, 0.7883)	IV
δ_2	treatment rate of stage 2	0.75282 (0.6086, 0.8971)	IV
δ_3	treatment rate of stage 3	0.74938 (0.6057, 0.8930)	IV
v	the percentage of male partners of bisexual men	0.62567 (0.4549, 0.7965)	IV

3.2.2. HIV/AIDS Epidemic Trend under Baseline Parameters

We first studied the dynamic process of the total number of infected people and the proportion carrying drug-resistant strains over time from 2015 to 2025 in the entire population of Anhui province under baseline parameters (Figure 3). The results show that the total number of HIV/AIDS infections during this period appears to be an upward trend, and the total number of surviving infections will increase to 33,260 by 2025 (95% confidence interval (CI): 28,096–40,090, Figure 3a). The number of new infections is also increasing and will reach 4522 (95% CI: 3366–6130) by 2025. According to the Anhui Provincial Health and Family Planning Commission, from 2017 to 2019, the cumulative reports of HIV/AIDS surviving patients whose current address is Anhui province are 18619 [31], 17183 [32] and 19604 [33], respectively. They are indicated by red dots in Figure 3a. Considering that the detection rate of HIV infections in Anhui province has not reached 100%, the number of reported cases should be less than the actual number of HIV infections. Based on the reported HIV-positive people and the HIV testing rate [31] in the corresponding year, we have reason to say that our predicted results are basically consistent with the actual situation.

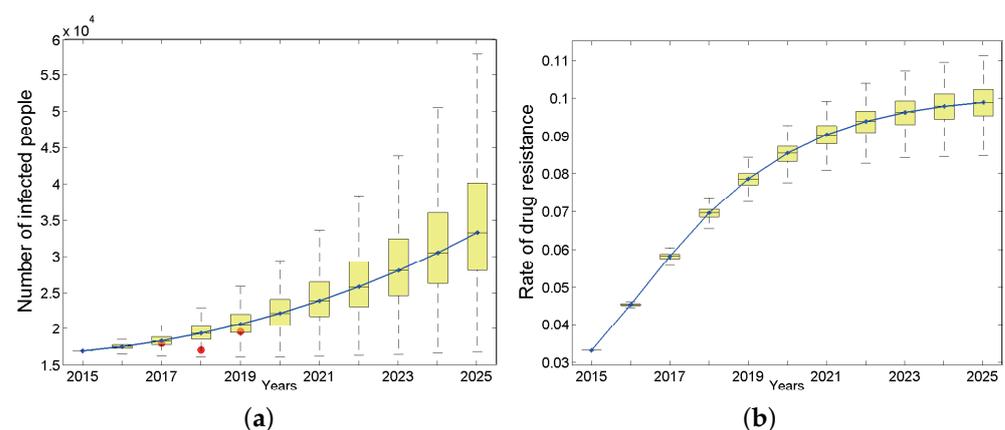


Figure 3. The number of HIV-positive people in the entire population (a) and drug resistance rate (b), where the continuous curve is the median line. Red dots in (a) are reported cases of 2017, 2018 and 2019 respectively.

The target population of our study is the general population. We studied the proportion of each population in the total new infections under baseline parameters from 2015 to 2025. The results showed that the proportion of heterosexual HIV-positive people decreased from 30.28% in 2015 to 20.43% in 2025 (95%CI: 15.92–26.09%, of which female infected

people accounted for 15.61%, while heterosexual males accounted for 4.82%). In contrast, the proportion of HIV-positive MSM (homosexual and bisexual men) has steadily increased. By 2025, 78.84% (95% CI: 68.97–88.70%) of new HIV infections will be infected through homosexual sex behavior. Among them, homosexual HIV-positive people accounted for 53.87%, and bisexual HIV-positive people accounted for 24.97%.

Regarding the subpopulations, although the HIV infection rate among heterosexual men has declined and tends to be stable, the rates in other populations have shown steady upward trends. By 2025, the HIV-positive rate among heterosexual women will reach 0.25% (95% CI: 0.21–0.28%), while the HIV-positive rate among heterosexual men will be lower, only about half of the positive rate of women (0.12%, 95% CI: 0.10–0.13%). In contrast, men who have sex with men belong to high-risk groups, and their HIV positive rate is dozens of times that of heterosexual men: the HIV infection rate of bisexual men is 2.44% (95% CI: 1.88–3.12%), while that of homosexual men is much higher, reaching 3.62% (95% CI: 2.81–4.64%) by 2025.

Considering that large-scale treatment for HIV-infected people in China has lasted for quite a long time, we are concerned that drug-resistant strains (secondary drug resistance) will appear in the patients due to poor medication compliance, which will lead to the wide spread of drug-resistant strains (primary drug resistance) in the population. Unsurprisingly, the proportion of HIV-positive people carrying drug-resistant strains in Anhui province has shown a rapid upward trend (Figure 3b) and will basically reach the highest point of 9.88% by 2025, which corresponds to 3282 (95% CI: 2862–3812) in terms of quantity. The proportion of primary drug-resistant patients in the total newly infected population will also increase over time. By 2025, the number of primary drug-resistant patients will account for 9.63% of total patients carrying drug-resistant strains (95% CI: 6.56–14.2%). In addition, the proportion of primary drug-resistant patients among total new infections first showed an upward trend, then tended to be stable each year, and finally will increase to 7.01% (95% CI: 6.61–7.44%) by 2025.

3.2.3. Influence of Intervention Measures on HIV / AIDS Epidemic Trend

In order to implement the “Healthy China 2030” program and deepen the medical and health system reformation, China’s “13th Five-Year Plan for Combating and Prevention of AIDS” in 2017 set the “90–90–90” goal. That is, the proportion of infected persons detected by testing should be above 90%; the proportion of diagnosed infected persons receiving antiviral treatment should be above 90%; the treatment success rate of infected persons receiving antiviral treatment should be above 90% [34] by 2020. In the baseline parameters, the treatment rates of the three different infection stages were 50.19% (95% CI: 21.55%, 78.83%), 75.28% (95% CI: 60.86%, 89.71%) and 74.94% (95% CI: 60.57%, 89.30%). These treatment rates are still far from the second “90%”. Next, we consider what will happen if the treatment rate increases to the target state in 2020.

The total HIV infection rate and drug resistance of Anhui province from 2020 to 2025 under the “90%” treatment rate are shown in Figure 4, which is a comparison chart between the baseline and the target situation. The results show that the total infection rate still keeps increasing (the green curve in Figure 4a) even under the “90%” treatment rate, which is only slightly lower than the baseline parameters (the red curve in Figure 4a). Even more surprising, the total drug resistance rate after intensified treatment will increase more rapidly than the baseline situation, reaching 10.92% (95% CI: 10.69–11.12%) by 2025, which is seen as the red curve in Figure 4b. In addition, the proportion of primary drug resistance in new infections has also shown an increasing trend compared with baseline parameters.

In short, when we increase the treatment rate to 90%, the total infection rate can be reduced only by 0.017% by 2025, but in turn, the total drug resistance rate and the proportion of primary drug resistance in new infections will increase 1.06% and 1.12%, respectively. What prevents ART treatment from playing an effective role in the epidemic? Is the current drug treatment effect not good enough? If the new first-line drugs can improve

the treatment effect (including reducing the infectivity and prolonging the lifespan), how much will the epidemic be improved?

To answer these questions, we assume the following two situations: (a) the new drug can only prolong the lifespan after treatment by 20% (it can survive for 24 years after extension); (b) the new drug can not only prolong the life after treatment by 20% but also reduce the infection rate by 20%. The results are shown in Figure 5. We were surprised to find that assumption (a) will make the epidemic worse instead (Figure 5a, green line). This may be because patients will have more risky sexual behaviors due to longer lifespan. However, assumption (b) will slow down the trend (Figure 5a, red line). In terms of drug resistance, the drug resistance rate will decrease slightly under assumption (a) (Figure 5b, green line) and increase under assumption (b) (Figure 5b, red line). The main reason may be that treatment only suppresses the non-resistant strains, which may instead give drug-resistant strains more space for transmission. This result suggests that prolonging the lifespan of the patients is a double-edged sword. When developing new drugs, we should not only consider how to prolong the life span of patients but also consider how to reduce the infectivity. This conclusion will continue to be discussed in the sensitivity analysis.

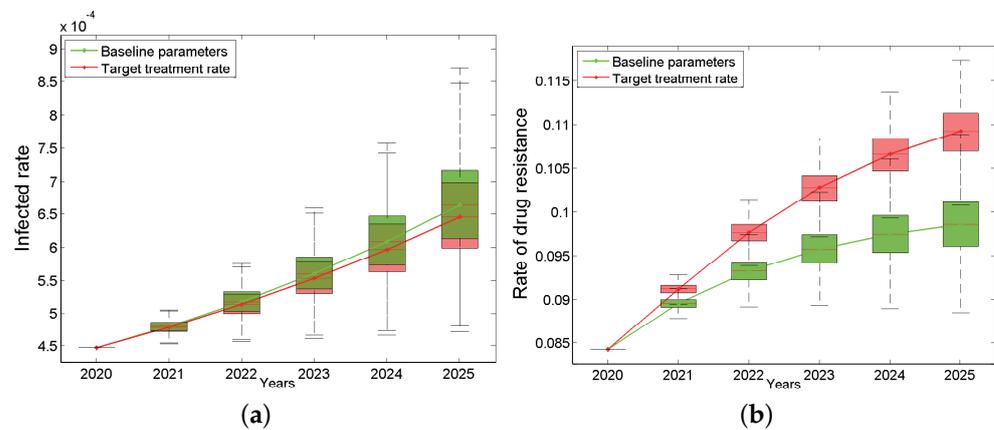


Figure 4. Impact of target treatment rates on the HIV/AIDS epidemic: the infected rate (a) and the rate of drug resistance (b). The green box plots are based on the baseline parameters, while the red ones are based on target treatment rate. The continuous green and red curves are their median lines respectively.

3.2.4. Sensitivity Analysis

In previous studies, we unexpectedly found that merely increasing the treatment rate to the target rate (90% treatment rate) could not lead to a significant improvement in HIV epidemic (Figure 5a). In order to find the reason, we did some sensitivity analysis. We find that only when shortening each infectious stage's life span to less than 25% of its baseline values can the epidemic be controlled. Details can be found in Appendix A.5. In the process of new drug development, if we cannot effectively reduce the infection rate while continuing to pursue prolonging the life-span of patients after treatment, the epidemic will keep rising as a consequence. We speculate that this may be one of the reasons that the effective reproduction \mathcal{R}_e^g has been increasing since 2012 (Figure 2). In other words, it is difficult to control the epidemic only by developing new drugs in a short period. So, what are the key factors to control the HIV/AIDS epidemic?

To answer this question, we conducted a sensitivity analysis on the number of sexual partners in different populations. An interesting fact is that changing the number of heterosexual people's sexual partners has little impact on the epidemic, and it almost can be negligible while comparing with the two subpopulations of homosexual sexual behavior who contributed a significant effect to the epidemic. As long as the number of bisexual men's annual sex partners is controlled below 6, and meanwhile, the number of homosexual

men's annual sex partners is controlled below 4, the epidemic can be controlled. Details can be found in Appendix A.6.

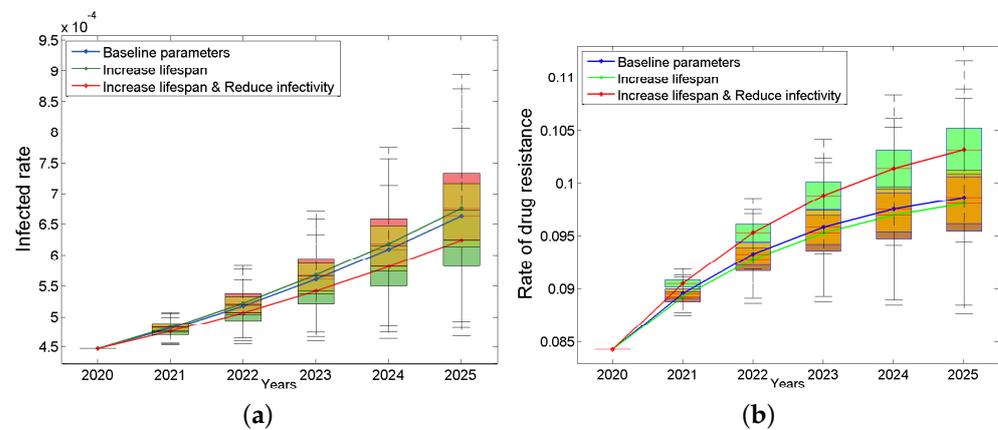


Figure 5. Impacts of improving the treatment effect on the infected rate (a) and the rate of drug resistance (b). The blue box plots are based on the baseline parameters, the green ones are based on increase lifespan, and red ones are based on both increase lifespan and reduce infectivity. The continuous blue, green and red curves are their median lines respectively.

4. Discussion and Conclusions

Generally speaking, predictive models for infectious diseases generally cannot avoid the problem of parameter estimation. After all, not all parameter values are easy to obtain. Even if some foreign literature mentions the value range of certain parameters, there may still exist differences due to different races. To some degree, parameter values determine the credibility of predictions. Facing the complex history of HIV treatment in China, different historical stages may have different values even for the same parameter. We innovatively carried out data mining on virus gene sequence detected from the infected people, so as to fit some parameter values which are difficult to obtain by social surveys or laboratory tests, making our prediction results more reliable. However, other non-Bayesian research methods are also good choices for the mathematical modeling of infectious diseases with a large number of reported data [35].

The idea of combining gene sequences with dynamic models came from the absence of monitoring data when doing model fitting. Unlike novel coronavirus patients, HIV-infected people do not have obvious infection symptoms. Many HIV-infected people do not know that they have been infected for a long time. This prevents the government from obtaining information about people living with HIV as early as possible. Moreover, the Chinese government's screening efforts for HIV-infected people are far less than those for individuals infected with novel coronavirus. As a result, recorded HIV infection data, if available, may not be very accurate and may not be suitable for parameter fitting. In Anhui province, which we studied, there is not much surveillance data on HIV/AIDS that can be used. The only reliable data available for use are 340 genetic sequences of HIV-infected people taken before treatment. HIV is an RNA virus. This gives it the ability to mutate faster, and these mutations show up in the entire genetic sequence of the virus from generation to generation. In other words, the evolutionary history of HIV is hidden in its genetic sequences. Since both gene sequences and dynamic models can describe the transmission dynamics of viruses, it should be possible to combine them for the purpose of learning from each other. This is our motivation to finish this manuscript.

What is more, most dynamic models generally only consider heterosexuality and homosexuality when classifying the target population, and they will not further subdivide homosexual people. Our study divided the MSM population into homosexual men and bisexual men to analyze the importance of bisexual men in the spread of HIV/AIDS. These bisexual men may be biological and psychological bisexuals; also, it may be that homosexuals are forced to marry women under the pressure of public opinion to cover

up their particular sexual orientation. We generally think that such people are a bridge that links HIV/AIDS from high-risk groups to the general population. Thus, cutting off this bridge will slow down the HIV epidemic. However, the result of our model shows an opposite result: if bisexual men with the unchanged number of sexual partners disconnect from men, the epidemic will slow down a lot; otherwise, it will lead to an increase in the epidemic. For more details, please read Appendix A.7.

In “China’s 13th Five-Year Plan for Containment and Prevention of AIDS”, China advocates that HIV/AIDS prevention and treatment should achieve three 90%: the proportion of infected people and patients who have been diagnosed and know about their infection status should reach 90% or more; the proportion of infected people and patients who meet the treatment conditions receiving antiviral treatment should be more than 90%; the success treatment rate of infected people and patients receiving antiviral treatment should be more than 90%. These three 90% involve screening, treatment and drug development. According to our findings, if no more effective drugs are available, simply increasing the treatment rate will cause a slight decrease in the number of new infections and the number of people with primary drug resistance. Moreover, the epidemic is still on the rise; that is to say, the increase of treatment rate cannot effectively control the epidemic, and the prevention and control effect is not ideal. A similar phenomenon was also found in the study [36] by Lou Jie et al. A plausible explanation is that after treatment, the life expectancy of HIV-infected patients is extended, but the current drug’s effect is not so good at reducing the infection rate, thus creating a greater possibility of transmission instead. In the meantime, free treatments of HIV have not yet covered the entire course of patients, so once drug resistance develops, the treatment led to failure.

Therefore, after the implementation of the “90–90–90” strategic measures, continuous improvements of the treatment rate did not bring the HIV epidemic under control but inversely caused more drug resistance. The Chinese government needs to intervene in various aspects of different groups of people in combination with other perspectives in order to curb the current domestic AIDS epidemic. The sensitivity analysis on the two critical parameters, infection rate after treatment and progress rate of disease, indicates neither one is the most critical factor to impact the epidemic. It would be easier to control the epidemic by reducing the number of sexual partners of the two subpopulations in the MSM group through publicity and other means. In addition, it is worth mentioning that considering the absolute leading role of homosexual people in the epidemic at the present stage, we should tilt most of the resources to this group so that we can benefit from spending money wisely.

In short, gene sequences can tell us the history of HIV/AIDS spreading and the dynamic model can tell us its future. Combining the two together is an innovative approach, especially for epidemics where reliable surveillance data are lacking. In addition, we believe that this new method is not only suitable for HIV but also for other RNA viruses, such as novel coronavirus. Of course, our method is not perfect, since the influencing factors considered by the Bayesian model are relatively simple. However, at the very least, genetic sequence data can be used as a supplement to macro surveillance data for epidemic prediction.

Author Contributions: Z.W. and J.L. conceptualized the study and wrote the manuscript. Z.W., X.J. and Z.Z. analyzed the data and presented the results. Z.Z. and C.Z. was involved in manuscript integration. J.L. provided the epidemiological significance of the study, while Y.R. facilitated access to the data used in this study. C.Z., H.X., J.W., B.S. and Y.S. contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This work has received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by The Institutional Review Board of National Center for AIDS/STD Control and Prevention (NCAIDS) of the China Center for Disease Control and Prevention (China CDC) (protocol code X140617334, approval date: 7 December 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All relevant data are within the paper. Additional data cannot be shared publicly based on regulations of the National Center for AIDS/STD Control and Prevention (NCAIDS) of the China Center for Disease Control and Prevention (China CDC). These data are available from the the Institutional Review Boards of the NCAIDS for researchers who meet the criteria for access to confidential data (Tel.: 58900922, Fax: 58900920, E-mail: office606@chinaaids.cn).

Acknowledgments: The authors would like to thank the Chinese Center for Disease Control and Prevention for approving accessing to data and other valuable documents used in the write-up of this manuscript. The manuscript preprint has been presented.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. The Dynamic model

$$\begin{cases} \dot{S}^g = a^g - \mu S^g - \Omega^g(t) - \Lambda^g(t) \\ \dot{I}_1^g = \Omega^g(t) - (\theta_1^I + \delta_1 + \mu) I_1^g + \eta T_1^g + \rho R_1^g \\ \dot{I}_2^g = \theta_1^I I_1^g - (\theta_2^I + \delta_2 + \mu) I_2^g + \eta T_2^g + \rho R_2^g \\ \dot{I}_3^g = \theta_2^I I_2^g - (\theta_3^I + \delta_3 + \mu) I_3^g + \eta T_3^g + \rho R_3^g \\ \dot{T}_1^g = \delta_1 I_1^g - (\eta + \theta_1^T + \gamma + \mu) T_1^g \\ \dot{T}_2^g = \delta_2 I_2^g + \theta_1^T T_1^g - (\eta + \theta_2^T + \gamma + \mu) T_2^g \\ \dot{T}_3^g = \delta_3 I_3^g + \theta_2^T T_2^g - (\eta + \theta_3^T + \gamma + \mu) T_3^g \\ \dot{R}_1^g = \Lambda^g(t) + \gamma T_1^g - (\theta_1^R + \rho + \mu) R_1^g \\ \dot{R}_2^g = \theta_1^R R_1^g + \gamma T_2^g - (\theta_2^R + \rho + \mu) R_2^g \\ \dot{R}_3^g = \theta_2^R R_2^g + \gamma T_3^g - (\theta_3^R + \rho + \mu) R_3^g \end{cases} \tag{A1}$$

where $\Omega^g(t)$ and $\Lambda^g(t)$ ($g = w, mw, mm, b$) denote the number of newly infected people at the time of t at subgroup g infected with wild HIV strains and resistant HIV strains, respectively.

$$\begin{cases} \Omega^w(t) = \frac{c^w S^w \sum_{i=1}^3 [\beta_i^I (I_i^{mw} + I_i^b) + \beta_i^T (T_i^{mw} + T_i^b)]}{N^{mw} + N^b} \\ \Omega^{mw}(t) = \frac{c^{mw} S^{mw} k \sum_{i=1}^3 (\beta_i^I I_i^w + \beta_i^T T_i^w)}{N^w} \\ \Omega^{mm}(t) = \frac{c^{mm} S^{mm} \sum_{i=1}^3 [\beta_i^I (I_i^{mm} + I_i^b) + \beta_i^T (T_i^{mm} + T_i^b)]}{N^{mm} + N^b} \\ \Omega^b(t) = \frac{vc^b S^b \sum_{i=1}^3 [\beta_i^I (I_i^{mm} + I_i^b) + \beta_i^T (T_i^{mm} + T_i^b)]}{N^{mm} + N^b} + \frac{(1-v)c^b S^b k \sum_{i=1}^3 (\beta_i^I I_i^w + \beta_i^T T_i^b)}{N^w} \\ \Lambda^w(t) = \frac{c^w S^w \sum_{i=1}^3 \epsilon \beta_i^I (R_i^{mw} + R_i^b)}{N^{mw} + N^b} \\ \Lambda^{mw}(t) = \frac{c^{mw} S^{mw} k \sum_{i=1}^3 \epsilon \beta_i^I R_i^w}{N^w} \\ \Lambda^{mm}(t) = \frac{c^{mm} S^{mm} \sum_{i=1}^3 \epsilon \beta_i^I (R_i^{mm} + R_i^b)}{N^{mm} + N^b} \\ \Lambda^b(t) = \frac{vc^b S^b \sum_{i=1}^3 \epsilon \beta_i^I (R_i^{mm} + R_i^b)}{N^{mm} + N^b} + \frac{(1-v)c^b S^b k \sum_{i=1}^3 \epsilon \beta_i^I R_i^w}{N^w} \end{cases} \tag{A2}$$

and

$$\begin{cases} N^w = S^w + I_1^w + I_2^w + I_3^w + T_1^w + T_2^w + T_3^w + R_1^w + R_2^w + R_2^w \\ N^{mw} = S^{mw} + I_1^{mw} + I_2^{mw} + I_3^{mw} + T_1^{mw} + T_2^{mw} + T_3^{mw} + R_1^{mw} + R_2^{mw} + R_2^{mw} \\ N^{mm} = S^{mm} + I_1^{mm} + I_2^{mm} + I_3^{mm} + T_1^{mm} + T_2^{mm} + T_3^{mm} + R_1^{mm} + R_2^{mm} + R_2^{mm} \\ N^b = S^b + I_1^b + I_2^b + I_3^b + T_1^b + T_2^b + T_3^b + R_1^b + R_2^b + R_2^b. \end{cases} \quad (A3)$$

Appendix A.2. The Basic Reproduction Number R_0^d

Here, we only show how to calculate the most complex basic reproduction number in the IV period, $R_0^d(IV)$. The R_0^d of the other stages (I, II, III) can be simplified from $R_0^d(IV)$.

We use the "regeneration matrix" method to find the basic reproduction number $R_0^d(IV)$ of the model (A1); the process is as follows:

Step 1:

Construct matrices f_0 and v_0 , where f_0 is a column vector of $n \times 1$, representing a new infection in each infection state, and v_0 is a column vector of $n \times 1$, representing that each infected state has been transported (not newly infected), where n is the number of infected states.

$$f_0^s = (\Omega^s, 0, 0, 0, 0, 0, \Lambda^s, 0, 0)' \quad (A4)$$

then

$$f_0 = \begin{pmatrix} f_0^w \\ f_0^{mw} \\ f_0^{mm} \\ f_0^b \end{pmatrix} \quad (A5)$$

$$v_0^s = \begin{pmatrix} -(\theta_1^I + \delta_1 + \mu) * I_1^s + \eta * T_1^s + \rho * R_1^s \\ \theta_1^I * I_1^s - (\theta_2^I + \delta_2 + \mu) * I_2^s + \eta * T_2^s + \rho * R_2^s \\ \theta_2^I * I_2^s - (\theta_3^I + \delta_3 + \mu) * I_3^s + \eta * T_3^s + \rho * R_3^s \\ \delta_1 * I_1^s - (\eta + \theta_1^T + \gamma + \mu) * T_1^s \\ \delta_2 * I_2^s + \theta_1^T * T_1^s - (\eta + \theta_2^T + \gamma + \mu) * T_2^s \\ \delta_3 * I_3^s + \theta_2^T * T_2^s - (\eta + \theta_3^T + \gamma + \mu) * T_3^s \\ \gamma * T_1^s - (\theta_1^R + \rho + \mu) * R_1^s \\ \theta_1^R * R_1^s + \gamma * T_2^s - (\theta_2^R + \rho + \mu) * R_2^s \\ \theta_2^R * R_2^s + \gamma * T_3^s - (\theta_3^R + \rho + \mu) * R_3^s \end{pmatrix} \quad (A6)$$

then

$$v_0 = \begin{pmatrix} v_0^w \\ v_0^{mw} \\ v_0^{mm} \\ v_0^b \end{pmatrix} \quad (A7)$$

Step 2:

Construct the matrix F and V , and calculate the value of the Jacobian matrix of f_0 and v_0 at the disease-free equilibrium, respectively.

We can easily find the disease-free equilibrium of the model (A1) (disease-free equilibrium) E_0 is

$$\left(\frac{a^w}{\mu}, 0, 0, 0, 0, 0, 0, 0, 0, \frac{a^{mw}}{\mu}, 0, 0, 0, 0, 0, 0, 0, 0, 0, \frac{a^{mm}}{\mu}, 0, 0, 0, 0, 0, 0, 0, 0, 0, \frac{a^b}{\mu}, 0, 0, 0, 0, 0, 0, 0, 0\right) \quad (A8)$$

The matrix F, V is the Jacobian matrix of f, v . Now, the matrix F, V is represented by a block matrix of 9×9 . So, we can obtain F_{E_0}, V_{E_0} as follows.

$$F_{E_0} = \begin{pmatrix} 0 & F_{12} & 0 & F_{14} \\ F_{21} & 0 & 0 & 0 \\ F_{31} & 0 & F_{33} & F_{34} \\ F_{41} & 0 & F_{43} & F_{44} \end{pmatrix}$$

$$V_0 = \begin{pmatrix} -\delta^1 - \mu - \theta_1^I & 0 & 0 & \eta & 0 & 0 & \rho & 0 & 0 \\ \theta_1^I & -\delta_2 - \mu - \theta_2^I & 0 & 0 & \eta & 0 & 0 & \rho & 0 \\ 0 & \theta_2^I & -\delta_3 - \mu - \theta_3^I & 0 & 0 & \eta & 0 & 0 & \rho \\ \delta_1 & 0 & 0 & -\eta - \gamma - \mu - \theta_1^T & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta_2 & 0 & \theta_1^T & -\eta - \gamma - \mu - \theta_2^T & 0 & 0 & 0 & 0 \\ 0 & 0 & \delta_3 & 0 & \theta_2^T & -\eta - \gamma - \mu - \theta_3^T & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma & 0 & 0 & -\mu - \rho - \theta_1^R & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma & 0 & \theta_1^R & -\mu - \rho - \theta_2^R & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma & 0 & \theta_2^R & -\mu - \rho - \theta_3^R \end{pmatrix}$$

$$V_{E_0} = \begin{pmatrix} V_0 & 0 & 0 & 0 \\ 0 & V_0 & 0 & 0 \\ 0 & 0 & V_0 & 0 \\ 0 & 0 & 0 & V_0 \end{pmatrix}$$

Step 3:

Obtain basic reproduction number $M R_0(I V) = \rho(F_{E_0} V_{E_0}^{-1})$, where ρ represents the maximum eigenvalue.

Appendix A.3. Initial Population

According to the statistical results of all groups of people in Anhui in 2015, the initial value is shown in Table A1.

Table A1. The initial population in 2015.

Stage	Population	Heterosexual Women	Heterosexual Men	Homosexual Men	Bisexual Men
S		$S^w = 24,347,402$	$S^{mw} = 24,682,596$	$S^{mm} = 449,877$	$S^b = 330,210$
I ₁		$I_1^w = 1200$	$I_1^{mw} = 1216$	$I_1^{mm} = 1857$	$I_1^b = 1365$
T ₁		$T_1^w = 0$	$T_1^{mw} = 0$	$T_1^{mm} = 0$	$T_1^b = 0$
R ₁		$R_1^w = 5$	$R_1^{mw} = 5$	$R_1^{mm} = 5$	$R_1^b = 5$
I ₂		$I_2^w = 335$	$I_2^{mw} = 340$	$I_2^{mm} = 527$	$I_2^b = 389$
T ₂		$T_2^w = 810$	$T_2^{mw} = 820$	$T_2^{mm} = 1240$	$T_2^b = 925$
R ₂		$R_2^w = 60$	$R_2^{mw} = 61$	$R_2^{mm} = 85$	$R_2^b = 67$
I ₃		$I_3^w = 335$	$I_3^{mw} = 340$	$I_3^{mm} = 527$	$I_3^b = 389$
T ₃		$T_3^w = 810$	$T_3^{mw} = 820$	$T_3^{mm} = 1240$	$T_3^b = 925$
R ₃		$R_3^w = 60$	$R_3^{mw} = 61$	$R_3^{mm} = 85$	$R_3^b = 67$

Appendix A.4. Fitted Parameters

Details of the values and references can be found in Tables A2–A4. The details of other fitting parameters are shown in Table A5. The fitted curves of the two effective reproduction numbers $C R_e(t)$ and $M R_e(t)$ for four historical periods I, II, III and IV are shown in Figure A1.

Table A2. Infection rate: β .

Parameters	Description	Value (95%CI)	Source
β_1^I	infection rate of wild strain carriers in stage 1	0.0428 (0.0160, 0.6590)	[15,37]
β_2^I	infection rate of wild strain carriers in stage 2	0.0500 (0.0200, 0.0700)	[37]
β_3^I	infection rate of wild strain carriers in stage 3	0.1000 (0.0500, 0.1400)	[37]
β_1^T	infection rate of wild strain carriers after HAART in stage 1	0.0282 (0.0106, 0.4349)	[14]
β_2^T	infection rate of wild strain carriers after HAART in stage 2	0.0330 (0.0132, 0.0462)	[14]
β_3^T	infection rate of wild strain carriers after HAART in stage 3	0.0120 (0.0060, 0.0168)	[14,38]
β_1^R	infection rate of drug-resistant carriers in stage 1	$\epsilon \beta_1^I$	
β_2^R	infection rate of drug-resistant carriers in stage 2	$\epsilon \beta_2^I$	
β_3^R	infection rate of drug-resistant carriers in stage 3	$\epsilon \beta_3^I$	
ϵ	transmissibility ratio of drug-resistant to untreated wild strain carriers	fit	

Table A3. Disease progression rate: θ .

Parameters	Description	Value: Years	Source
$1/\theta_1^I$	course of wild strain carriers in stage 1	4	[14]
$1/\theta_2^I$	course of wild strain carriers in stage 2	6	[14]
$1/\theta_3^I$	course of wild strain carriers in stage 3	2	[14]
$1/\theta_1^R$	course of drug-resistant carriers in stage 1	4	[14,39,40]
$1/\theta_2^R$	course of drug-resistant carriers in stage 2	6	[14,39,40]
$1/\theta_3^R$	course of drug-resistant carriers in stage 3	2	[14,39,40]

Table A4. Other parameters.

Parameters	Description	Value	Source
η	the percentage of patients who give up HAART	0.0602	[41,42]
γ	rate of drug resistance after HAART	0.0545	[42,43]
ρ	the percentage of patients who are no longer resistant	0.2687	[42,44]
μ	natural mortality rate	0.0131	[45]
k	ratio of female to other and male to other transmission	0.5263	[36]

Table A5. Other fitted parameters.

Period	Parameter	Value/Function
I	a^w	509832
	$a^{mw}(t)$	$542781 - a^{mm}(t) - a^b(t)$
	$a^{mm}(t)$	$1432.6 \times t + 10527$
	$a^b(t)$	$1607 \times t + 10431$
	$c^{mm}(t)$	$-0.036928 \times t + 2.8484$
	$c^b(t)$	$-0.040123 \times t + 3.1024$
	c^w	1.5228
	c^{mw}	1.5659
	v	0.70368
II	a^w	613031
	$a^{mw}(t)$	$685421 - a^{mm}(t) - a^b(t)$
	$a^{mm}(t)$	$1622.4 \times t + 12382$
	$a^b(t)$	$1398.7 \times t + 10591$
	$c^{mm}(t)$	$-0.042873 \times t + 3.4122$
	$c^b(t)$	$0.092656 \times t + 3.4909$
	c^w	1.4985
	c^{mw}	1.5066
	$v(t)$	$-0.0098732 \times t + 0.59836$
III	$\delta_3(t)$	$0.03733 \times t + 0.35124$
	a^w	449161
	$a^{mw}(t)$	$493770 - a^{mm}(t) - a^b(t)$
	$a^{mm}(t)$	$1455.8 \times t + 5545.3$
	$a^b(t)$	$2079.9 \times t + 5093.9$
	$c^{mm}(t)$	$1/(0.61892 + \exp(-2.6515 \times t + 3.1585)) + 3.5015$
	$c^b(t)$	$1/(0.54199 + \exp(-2.7309 \times t + 3.2935)) + 3.4814$
	c^w	1.5362
	c^{mw}	1.5631
$v(t)$	$0.07004 \times t + 0.51912$	
$\delta_2(t)$	$1/(1.6826 + \exp(-1.809 \times t + 0.53915)) + 0.2192$	
$\delta_3(t)$	$1/(1.6585 + \exp(-1.8287 \times t + 0.55639)) + 0.19956$	

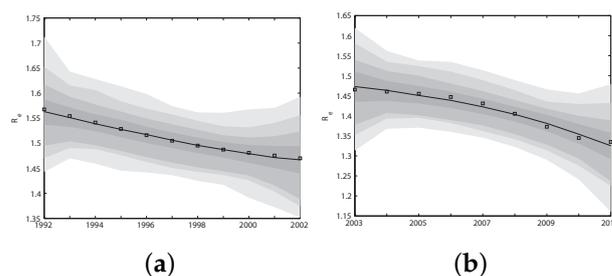


Figure A1. Cont.

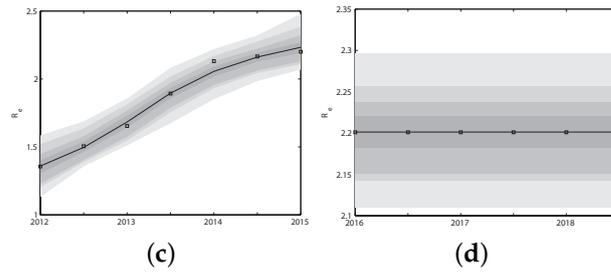


Figure A1. The fitted curves of the two effective reproduction numbers ${}_G\mathcal{R}_e(t)$ and ${}_M\mathcal{R}_e(t)$ for four historical periods I (a), II (b), III (c) and IV (d), respectively.

Appendix A.5. Sensitivity Analysis on Therapeutic Effect

There are two parameters related to the therapeutic effect: the post-treatment transmission rate (β^T) and the course of disease ($1/\theta^T$). Next, we will discuss the effect of changes in these parameters on the effective reproduction number ${}_M R_0(IV)$. We keep other parameters unchanged during this process. The sensitivity analysis results are shown in Figure A2b.

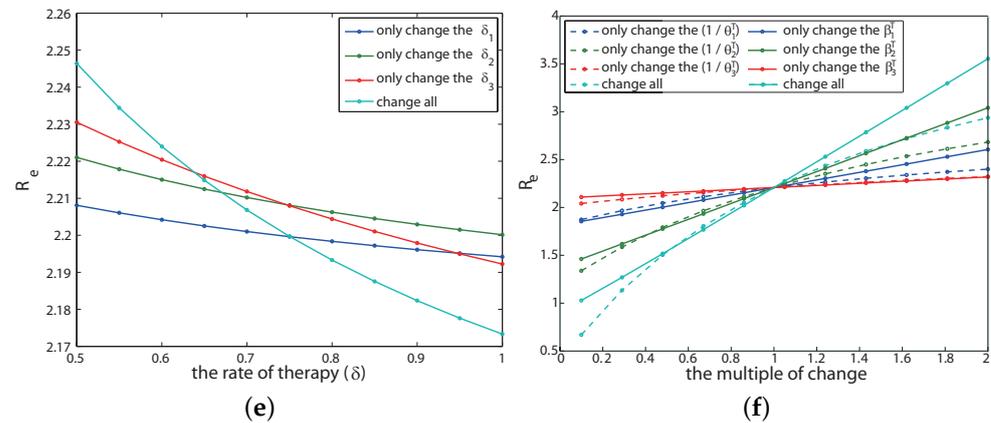


Figure A2. Sensitivity analysis of treatment rate (a), infection rate after treatment (solid line in (b)) and course of the disease after treatment (dotted line in (b)) to $R_0^d(IV)$.

If only changing the infection rate after treatment at each infection stage, that is, reducing or increasing β^T , the result is shown by the solid line in the Figure A2b. The abscissa is the multiple of the infection rate after the current baseline treatment (indicated by p , that is, the infection rate β^T becomes $p * \beta^T$ after each stage of treatment), and the ordinate is the corresponding ${}_M R_0(IV)$ value. The result shows that under the condition that other parameters remain unchanged, only when the infection rate of each infection stage after treatment simultaneously reduces to about 10% of the baseline infection rate after treatment β^T can ${}_M R_0(IV)$ drop below 1 (cyan solid line). Moreover, among the three infection stages, the infection rate in the second stage is the most sensitive to ${}_M R_0(IV)$ (solid green line), while the infection rate after treatment in the third stage is the least sensitive to ${}_M R_0(IV)$ (solid red line). This is because the course of the second stage is the longest among the three stages, and the patients in the third stage of infection have weaker sexual ability.

If only changing the course of the disease after treatment at each infection stage ($1/\theta^T$), the result is shown as the dotted line in Figure A2b, where the abscissa is p , and it is the value of a multiple of the course parameter value after current baseline treatment. We are surprised to find that under the condition that other parameters remain unchanged, ${}_M R_0(IV)$ is positively correlated with the course of disease after treatment, that is, the longer the course of disease after treatment, the higher the ${}_M R_0(IV)$ will be. Moreover, only when the course of the disease after each stage of treatment is reduced to less than 25%

of its baseline parameter value at the same time can ${}_M R_0(IV)$ drop below 1 (cyan dotted line). The order of sensitivity to ${}_M R_0(IV)$ in the course of the disease after treatment at different infection stages is consistent with the sensitivity to drug infectivity.

Appendix A.6. Sensitivity Analysis on the Number of Sexual Partners

If other parameters remain unchanged, and only the number of annual sexual partners of various groups of people is changed, the result is shown in Figure A3a, where the abscissa is the multiple of the current parameter change, and the ordinate is the corresponding ${}_M R_0(IV)$ value. The greater the slope of the straight line, the greater the influence of this parameter is on ${}_M R_0(IV)$. The results show that the disease transmission can be effectively controlled when the annual number of sexual partners in all populations is reduced to 0.4 times the current number of sexual partners (cyan line). The order of the influence of the number of different groups annual sexual partners on ${}_M R_0(IV)$ is: homosexual male (green line) > bisexual male (red line) > heterosexual people (the blue line, see the illustration on the Figure A3a). An interesting fact is that changing the number of heterosexual people's sexual partners has little impact on the spread of HIV/AIDS (blue line): almost negligible. The two types of people with homosexual sexual behavior (green line and red line) contributed a significant effect.

We further analyzed two types of MSM population and explored the impact of the number of sexual partners of bisexual men (c^b) and the number of heterosexual men (c^{mm}) on ${}_M R_0(IV)$ (Figure A3b).

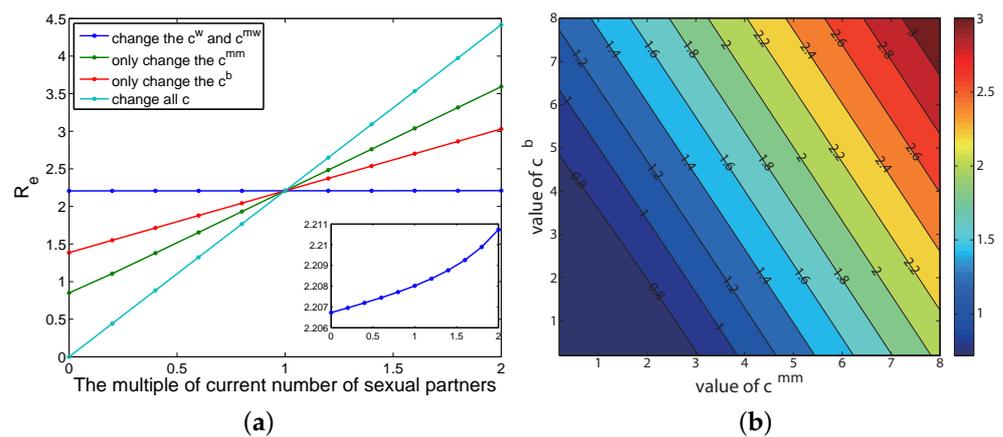


Figure A3. Sensitivity analysis of sexual partners to the reproduction number. (a) Only the number of annual sexual partners of various groups of people is changed and other parameters remain unchanged. (b) The impact of the number of sexual partners of bisexual men and the number of heterosexual men on the reproduction number.

Appendix A.7. Influence of Bisexual Men's Sexual Preference on HIV/AIDS Prevalence

Traditionally, we believe that bisexual men play a bridge role in HIV transmission because they are not only involved in the transmission between homosexual groups but also in the transmission between heterosexual people. We would like to know what impact bisexual men would have on the HIV epidemic if they become purely homosexual or heterosexual. Figure A4a is the curve of the number of HIV-positive patients in the whole population of Anhui province from 2020 to 2025 over time after changing the nature of bisexual males. The results show that under the assumption that bisexual men are the pure MSM population, the total number of HIV/AIDS infections in Anhui province has shown a rapid upward trend, far exceeding the baseline status, and the number of surviving infections will increase to 38,171 by 2025 (95% CI: 34,694–42,078, the red curve in Figure A4a), and the number of new infections will also increase to 6333 accordingly. However, the resistance rate is lower than the baseline status and will decrease to 9.41% by 2025 (95% CI: 8.83–9.88%, the red line in Figure A4b). On the contrary, if you disconnect

from male sex and become purely heterosexual, the number of HIV infections will be significantly lower than the baseline status by 2025, it will be as low as 26,957 by 2025 (95% CI: 25,890–28,633, Figure A4a green curve), and the number of new infections will reduce to 2465. Meanwhile, its resistance rate will exceed the baseline status, increasing to 10.59% (95% CI: 10.14–10.93%) (the green curve in Figure A4b) in 2025.

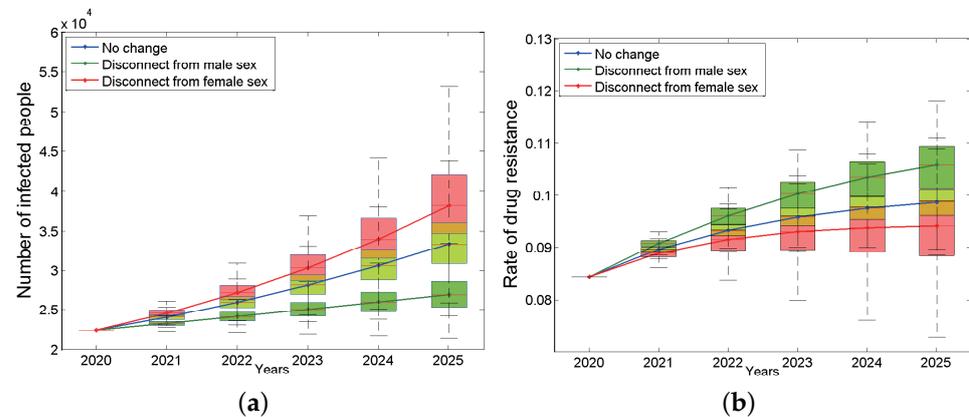


Figure A4. The impact of changing the sexual preferences of bisexual men on the on the number of infected people (a) and the rate of drug resistance (b). The blue box plots are the situation of no change, the green ones are the situation that disconnect from male sex, and the red ones are the situation that disconnect from female sex. The continuous blue, green and red curves are their median lines respectively.

Appendix A.8. Prevention and Control Strategy Optimization Analysis

For the current prevention and control measures, the spread of HIV in Anhui province has not been well controlled. Three important measures are now considered to be strengthened, and the new infection rate and the rate of drug resistance in 2021–2025 are numerically simulated (see the results in Tables A6 and A7).

The measures of reinforcement are to increase the treatment rate, reduce the withdrawal rate, and reduce the number of sexual partners. The specific measures are as follows:

- (1). The treatment rate is increased by 10%;
- (2). The treatment rate is increased by 20%;
- (3). The withdrawal rate is reduced by 10%;
- (4). The withdrawal rate is reduced by 20%;
- (5). The number of sexual partners is reduced by 10%;
- (6). The number of sexual partners is reduced by 20%;
- (7). (1)(3)(5) combination;
- (8). (2)(4)(6) combination.

From Tables A6 and A7, it is not difficult to see that increasing prevention and control efforts can reduce the population of new infections, and at the same time, it will also reduce the primary drug resistance population. The greater the prevention and control efforts are, the more the new infections and primary drug resistance population reduce. Among the three measures, the best effect is to reduce the number of sexual partners. Combining the three measures (cases 7, 8) compared with one measure is more effective in controlling both new infections and primary drug resistance population. For the prevention and control of HIV in Anhui, China, comprehensive measures can not only control new infections but also reduce the population of primary drug resistance, thus being a wise choice.

Table A6. New infection population under different measures.

Measures	Years				
	2021	2022	2023	2024	2025
Baseline parameter (95%CI)	3194 (2829, 3560)	3499 (3031, 3970)	3831 (3257, 4434)	4207 (3511, 4955)	4619 (3780, 5541)
(1) The treatment rate is increased by 10% (95%CI)	3150 (2787, 3510)	3432 (2981, 3884)	3752 (3193, 4325)	4102 (3424, 4820)	4497 (3670, 5378)
(2) The treatment rate is increased by 20% (95%CI)	3007 (2678, 3360)	3258 (2855, 3714)	3554 (3049, 4122)	3875 (3269, 4585)	4234 (3502, 5110)
(3) The withdrawal rate is reduced by 10% (95%CI)	3128 (2787, 3500)	3416 (2982, 3893)	3738 (3201, 4346)	4094 (3450, 4861)	4495 (3708, 5437)
(4) The withdrawal rate is reduced by 20% (95%CI)	3069 (2740, 3448)	3336 (2983, 3937)	3656 (3143, 4279)	4002 (3381, 4781)	4392 (3628, 5344)
(5) The number of sexual partners is reduced by 10% (95%CI)	2781 (2474, 3113)	2991 (2613, 3406)	3223 (2768, 3740)	3474 (2940, 4106)	3755 (3123, 4502)
(6) The number of sexual partners is reduced by 20% (95%CI)	2326 (2062, 2601)	2456 (2142, 2788)	2602 (2232, 2999)	2757 (2332, 3228)	2925 (2435, 3477)
(1)(3)(5) combination (95%CI)	2780 (2502, 3145)	2927 (2608, 3424)	3153 (2766, 2744)	3401 (2931, 4094)	3677 (3104, 4494)
(2)(4)(6) combination (95%CI)	2309 (2056, 2583)	2416 (2119, 2743)	2540 (2200, 2929)	2680 (2291, 3138)	2831 (2384, 3373)

Table A7. Primary drug resistance population under different measures.

Measures	Years				
	2021	2022	2023	2024	2025
Baseline parameter (95%CI)	209 (188, 230)	235 (209, 261)	262 (231, 294)	292 (253, 331)	323 (276, 371)
(1) The treatment rate is increased by 10% (95%CI)	208 (187, 229)	234 (209, 260)	261 (231, 295)	291 (255, 333)	321 (277, 374)
(2) The treatment rate is increased by 20% (95%CI)	201 (181, 222)	228 (204, 255)	257 (227, 189)	287 (251, 327)	318 (274, 368)
(3) The withdrawal rate is reduced by 10% (95%CI)	206 (185, 227)	231 (207, 258)	258 (229, 291)	287 (251, 327)	318 (274, 367)
(4) The withdrawal rate is reduced by 20% (95%CI)	202 (183, 224)	228 (205, 255)	254 (226, 288)	283 (249, 325)	313 (272, 365)
(5) The number of sexual partners is reduced by 10% (95%CI)	184 (168, 199)	205 (183, 228)	227 (201, 256)	250 (218, 284)	273 (236, 315)
(6) The number of sexual partners is reduced by 20% (95%CI)	155 (139, 172)	172 (153, 192)	188 (166, 212)	205 (179, 232)	221 (190, 254)
(1)(3)(5) combination (95%CI)	176 (157, 194)	197 (175, 219)	219 (192, 246)	241 (209, 275)	264 (226, 305)
(2)(4)(6) combination (95%CI)	154 (137, 174)	170 (151, 190)	184 (164, 210)	201 (178, 234)	219 (189, 250)

References

- Sun, Z.Y.; Wang, K.; Fan, H. Progress in the development and control of drug resistance in highly active antiretroviral therapy (HAART) for HIV/AIDS. *Jiangsu Prev. Med.* **2019**, *30*, 159–162.
- Zhang, F.J.; Pan, J.; Yu, L.; Wen, Y.; Zhao, Y. Current progress of China's free ART program. *Cell Res.* **2005**, *15*, 877. [[CrossRef](#)]
- Franois, C.; Allan, J.H. HIV Drug Resistance. *N. Engl. J. Med.* **2004**, *350*, 1023–1035.
- Grenfell, B.T.; Pybus, O.G.; Gog, J.R.; Wood, J.L.; Daly, J.M.; Mumford, J.A.; Holmes, E.C. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **2004**, *303*, 327–332. [[CrossRef](#)]
- Ragonnet-Cronin, M.; Jackson, C.; Bradley-Stewart, A.; Aitken, C.; McAuley, A.; Palmateer, N.; Gunson, R.; Goldberg, D.; Milosevic, C.; Brown, A.J.L. Recent and Rapid Transmission of HIV Among People Who Inject Drugs in Scotland Revealed Through Phylogenetic Analysis. *J. Infect. Dis.* **2018**, *217*, 1875–1882. [[CrossRef](#)]
- Villabona-Arenas, C.J.; de Oliveira, J.L.; de Sousa-Capra, C.; Balarini, K.; da Fonseca, C.R.T.P.; de Andrade Zanotto, P.M. Epidemiological dynamics of an urban dengue 4 outbreak in Sao Paulo, Brazil. *PeerJ* **2016**, *4*, e1892. [[CrossRef](#)]
- Samuel, A.; Sbaiébastien, L.; Carmen Lía, M.; Jessica, L.A. Quantifying the epidemic spread of Ebola virus (EBOV) in Sierra Leone using phylodynamics. *Virulence* **2014**, *5*, 825–827.
- Drummond, A.J.; Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **2007**, *7*, 214. [[CrossRef](#)]
- Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [[CrossRef](#)]

10. Hastings, W.K. Monte Carlo Sampling methods Using Markov chains and their Applications. *Biometrika* **1970**, *57*, 97–109. [[CrossRef](#)]
11. Stadler, T.; Kuhnert, D.; Bonhoeffer, S.; Drummond, A.J. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 228–233. [[CrossRef](#)] [[PubMed](#)]
12. Lou, J.; Wang, Z.; Jin, X. Traceback the epidemic dynamics of COVID-19. *J. Shanghai Norm. Univ. Nat. Sci.* **2020**, *27*, 1138–1143.
13. Lewis, F.; Hughes, G.J.; Rambaut, A.; Pozniak, A.; Leigh Brown, A.J. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med.* **2008**, *5*, e50. [[CrossRef](#)] [[PubMed](#)]
14. Wu, J. A Dynamic Model Was Established to Evaluate the Effect of AIDS Prevention and Treatment in China. Ph.D. Thesis, Chinese Center for Disease Control and Prevention, Beijing, China, 2014.
15. Smith, R.J.; Okano, J.T.; Kahn, J.S.; Bodine, E.N.; Blower, S. Evolutionary dynamics of complex networks of HIV drug-resistant strains: The case of San Francisco. *Science* **2010**, *327*, 697–701. [[CrossRef](#)]
16. Lou, J.; Wu, J.; Chen, L.; Ruan, Y.; Shao, Y. A sex-role-preference model for HIV transmission among men who have sex with men in China. *BMC Public Health* **2009**, *9*, 145–147. [[CrossRef](#)]
17. Baggaley, R.; Ferguson, N.; Garnett, G. The epidemiological impact of antiretroviral use predicted by mathematical models: A review. *Emerg. Themes Epidemiol.* **2005**, *2*, 1742–7622. [[CrossRef](#)]
18. Peng, Z.; Chen, F. Research progress of HIV/AIDS transmission dynamics model. *Chin. Health Stat.* **2011**, *28*, 730–734.
19. Zhou, Y. Modeling and prediction of HIV in China: transmission rates structured by infection ages. *Math. Biosci. Eng.* **2008**, *5*, 403.
20. James, M.; Hyman, L.J.; Stanley, E.A. Modeling the impact of random screening and contact tracing in reducing the spread of HIV. *Math. Biosci.* **2003**, *181*, 17–54.
21. Hsieh, Y.H.; Sheu, S.P. The effects of density-dependent treatment and behavior change on the dynamics of HIV transmission. *Math. Biol.* **2001**, *43*, 69–80. [[CrossRef](#)]
22. Kwon, Y.M.; Yeun, E.J.; Kim, H.Y.; Youn, M.S.; Cho, J.Y.; Lee, H.J. Application of the Transtheoretical Model to Identify Aspects Influencing Condom Use Among Korean College Students. *West. J. Nurs. Res.* **2008**, *8*, 1–14. [[CrossRef](#)]
23. Zhao, S.; Feng, Y.; Xin, R.; Li, Z.; Hao, M.; Liao, L.; Ruan, Y.; Lu, H.; Shao, Y.; Xing, H. Molecular transmission network was used to study the transmission characteristics of HIV-1 virus among male sex workers in Beijing. *HIV/AIDS China* **2018**, *24*, 241–245. Available online: https://xueshu.baidu.com/usercenter/paper/show?paperid=2868d899e9e688ccca32c92bd993207c&site=xueshu_se (accessed on 15 December 2019).
24. Zeng, H.; Li, T.; Wang, Y.; Sun, B.; Yang, R. The Epidemic Dynamics of Four Major Lineages of HIV-1 CRF01_AE Strains After Their Introduction into China. *Aids Res. Hum. Retrovir.* **2016**, *32*, 5. [[CrossRef](#)] [[PubMed](#)]
25. Drummond, A.J.; Rambaut, A.; Shapiro, B.; Pybus, O.G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **2005**, *22*, 1185–1192. [[CrossRef](#)] [[PubMed](#)]
26. Zhang, D.; Gao, F.; Jakovlić, I.; Zou, H.; Zhang, J.; Li, W.X.; Wang, G.T. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* **2020**, *20*, 348–355. [[CrossRef](#)] [[PubMed](#)]
27. Akaike, H.A. New Look at the Statistical Model Identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
28. Diekmann, O.; Heesterbeek, J.A.P.; Metz, J.A.J. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **1990**, *28*, 365. [[CrossRef](#)]
29. Castillo-Chavez, C.; Blower, S.; van den Driessche, P.; Kirschner, D.; Yakubu, A. *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 229–250.
30. Haario, H.; Laine, M.; Mira, A.; Saksman, E. DRAM: Efficient adaptive MCMC. *Stat. Comput.* **2006**, *16*, 339–354. [[CrossRef](#)]
31. Souhu. How Widespread Is HIV in 2019? Available online: https://www.sohu.com/a/354674894_52725019 (accessed on 20 December 2019).
32. Zhang, D.; Wu, J.; Zhang, Y.; Shen, Y. Genetic characterization of HIV-1 epidemic in Anhui Province, China. *Viol. J.* **2020**, *17*, 1. [[CrossRef](#)]
33. People’s Daily Online: Anhui Announced the Statutory Report of Infectious Diseases in 2018. Available online: <http://ah.people.com.cn/n2/2019/0307/c35\8266--32716708.html> (accessed on 7 March 2019).
34. Guo, Q. Implementation path of “Healthy China 2030” Plan. *Health Res.* **2016**, *36*, 601–604.
35. Bartolomeo, N.; Trerotoli, P.; Serio, G. Short-term forecast in the early stage of the COVID-19 outbreak in Italy. Application of a weighted and cumulative average daily growth rate to an exponential decay model. *Infect. Dis. Model.* **2021**, *6*, 212–221. [[CrossRef](#)] [[PubMed](#)]
36. Duriux-Smith, A.; Goodman, J.T. Comparison of female to male and male to female transmission of HIV in 563 stable couples. *Br. Med. J.* **1992**, *304*, 809.
37. Krebs, E.; Enns, B.; Wang, L.; Zang, X.; Panagiotoglou, D.; Del Rio, C.; Dombrowski, J.; Feaster, D.J.; Golden, M.; Granich, R.; et al. Developing a dynamic HIV transmission model for 6 US cities: An evidence synthesis. *PLoS ONE* **2019**, *14*, 217559. [[CrossRef](#)] [[PubMed](#)]
38. Anglemyer, A.; Rutherford, G.W.; Horvath, T.; Baggaley, R.C.; Egger, M.; Siegfried, N. Antiretroviral therapy for prevention of HIV transmission in HIV-discordant couples. *Cochrane Database Syst. Rev.* **2013**, *4*. [[CrossRef](#)]
39. Vardavas, R.; Blower, S. The emergence of HIV transmitted resistance in Botswana: When will the WHO detection threshold be exceeded? *PLoS ONE* **2007**, *2*, e152. [[CrossRef](#)]

40. Johnson, L.F.; Dorrington, R.E. Modelling the demographic impact of HIV/AIDS in South Africa and the likely impact of interventions. *Demogr. Res.* **2006**, *14*, 541–574. [[CrossRef](#)]
41. Chinese Center for Disease Control and Prevention STD AIDS Prevention and Control Center. *Statistics on the Number of Adults Receiving Antiviral Treatment for HIV/AIDS in 2018*; Chinese Center for Disease Control and Prevention STD AIDS Prevention and Control Center: Beijing, China, 2019.
42. Zhang, F.; Haberer, J.E.; Wang, Y.; Zhao, Y.; Ma, Y.; Zhao, D.; Yu, L.; Goosby, E.P. Study on drug resistance of HIV patients discontinued from antiviral therapy in some regions of China. *J. Trop. Med.* **2018**, *12*, 18.
43. Liang, X.; Peng, X. Meta-analysis of drug resistance after antiretroviral therapy in HIV-infected and AIDS patients in China. *Int. J. Epidemiol. Infect. Dis.* **2016**, *43*, 197–202.
44. Liu, L.; Zuo, Z.B.; Liao, L.J.; Liang, S.; Ma, Y.; Wu, G.; Liang, S.; Tian, S.; He, J.; Shao, Y.; et al. The drug resistance in HIV/AIDS patients who had stopped ART in 2016. *J. Trop. Med.* **2018**, *18*, 1613–1634.
45. The State Council: The 13th Five-Year Health and Health Plan. 2016. Available online: <https://wenku.baidu.com/view/574e021384c24028915f804d2b160b4e767f81a4.html> (accessed on 15 December 2019).