Mathematical Biosciences and Engineering

https://www.aimspress.com/journal/mbe

MBE, 23(5): 1356–1374.
DOI: 10.3934/mbe.2026050
Received: 14 December 2025
Revised: 27 March 2026
Accepted: 30 March 2026
Published: 07 April 2026

Research article

Effect of drug resistance on an HIV epidemic in heterogeneous populations

Roberto A. Saenz*

Faculty of Sciences, Universidad de Colima, Colima, Mexico

* **Correspondence:** Email: rsaenz@ucol.mx.

Abstract: The emergence of drug resistance (DR) may impair the control of the human immunodeficiency virus (HIV) epidemic. We analyze the transmission and drug-resistance dynamics of HIV using a stochastic model in combination with Monte Carlo simulations, which considers population heterogeneity, set in terms of the set-point viral load (SPVL) distribution. In our modeling framework, once a SPVL is sampled, it determines the transmission rate and survival time for each new infected individual. On the other hand, DR emergence randomly occurs with a fixed probability in an individual receiving antiretroviral therapy (ART). In addition, we implicitly assume that the SPVL serves as a proxy for the replication pressure and resistance risk, as the DR fitness-cost proportionally reduces the SPVL in our model (an assumption motivated by modeling convenience instead of an actual biological mechanism present in the DR dynamics). Finally, we simplify the dynamics by assuming no ART regimen switching in cases of treatment failure (which we assumed to be only due to DR). Our results show that for a high treatment coverage, DR causes a higher stationary infection prevalence in populations with moderate-to-high mean viral loads. Moreover, DR could produce more infections for greater ART coverage, especially when treatment is administered early after contagion; however, an earlier initiation of ART allows for the possibility of epidemic extinction. Additionally, we evaluate the effect on the epidemic dynamics of other parameters, such as the probability of DR emergence and the relative fitness of the DR strain. Overall, our analysis contributes to the understanding of drug resistance evolution in situations where ART is inadequately managed.

Keywords: drug resistance; HIV; heterogeneous population; set-point viral load; Monte Carlo simulations

1. Introduction

Despite the progress that has been made, the epidemic of human immunodeficiency virus (HIV) is still a public health concern. By the end of 2024, there were around 40.8 million people living with HIV worldwide, with an estimated 1.3 million new infections in 2024 alone [1]. However, about 5.3 million people did not know they were living with HIV in 2024, that is, only 87% of people with

HIV knew their status [1]. From 1995 to 2010, an estimated 2.5 million deaths were averted in low- and middle-income countries by implementing prevention measures and using antiretroviral therapy (ART) [2]. ART represents a crucial epidemic intervention: it slows down disease progression, increases the survival periods, and decreases the transmissibility [3]. In 2024, 77% of people living with HIV had access to ART [1]. Regional ART coverage in 2024 went from 48% in the Middle East and North Africa, to 80% in Western and Central Europe and North America [1]. The current recommendation from the World Health Organization (WHO) is that ART should be initiated for all people living with HIV, following a confirmed diagnosis and clinical assessment [4]. Therefore, a swift detection of new infections and their corresponding diagnosis are essential to anticipate a positive impact of this initiative. On the other hand, the median time from HIV infection to diagnosis in the U.S. was estimated at 39.5 months in 2018 [5], thus highlighting the challenging task that is required to meet the recommendations.

Drug-resistant (DR) viruses emerged shortly after the initial use of a single antiretroviral, in the late 1980's [6]. DR strains of the virus have the ability to replicate in the presence of drugs. The virus' fast replication rate and its lack of proofreading mechanisms favor the emergence of DR strains [7]. The main predictors for acquired drug resistance are suboptimal antiviral therapy and incomplete therapy adherence [8]. The estimated prevalence in 2024 of DR strains resistant to at least one drug was 25%, while for dual class resistance was estimated to be between 4.7 and 8.5% [9]. Most DR strains are less fit than the wild-type virus, but they can still be transmitted between individuals [10]; there are even some DR associated mutations that have no significant impact on the transmission fitness [11].

Most studies that use mathematical models for the analysis of epidemic dynamics of drug resistance assume a homogeneous population, thereby considering, at most, different risk groups [12–17]. However, it has been reported that infected individuals are heterogeneous with respect to the set-point viral loads (SPVL), the viral load around which an individual stabilizes after acute infection, and that not all populations have the same viral load distributions [18]. The SPVL of an infected individual is relevant, as it determines the individual's duration of infection and infectiousness [18]. Viral loads used to influence ART initiation [19], so that they may have had an effect on the dynamics of drug-resistant strains of the virus. Therefore, it seems pertinent to analyze an HIV epidemic that considers these interacting components. Using a mathematical modeling framework consisting of a stochastic model attached to Monte Carlo simulations, which have been previously used to study the dynamics of HIV within a host [20], we study the impact of drug resistance on an epidemic of HIV in a heterogeneous population. To consider a more accessible framework, we study the simplified case where there is a single ART regimen and no DR screening, so that ART continues to be administered even if it fails at viral suppression; this is a situation that represents an inadequate treatment system. After justifying each of the components of our model, we employ it to analyze several epidemic scenarios. We report the effect on the epidemic dynamics of factors such as ART coverage, and fitness cost and probability of emergence of a drug resistance strain.

2. Materials and methods

2.1. Modeling framework

The model consists of a stochastic model tied up to Monte Carlo simulations. The stochastic model determines the epidemic dynamics at the population level. In turn, the infectiousness and lifespan of a newly infected individual are determined by sampling the distribution of viral loads in the population

and randomly deciding whether ART is administered and drug resistance emerges, as well as other infection attributes. Importantly, we only consider one ART regimen in our model, that is, the same ART continues to be administered even in the presence of DR. The probability of a new infection depends on the combined infectiousness, or transmission profile, of all infected individuals. Although this proposed modeling framework resembles an individual-based model, we do not follow any particular individual throughout the duration of the infection.

2.1.1. Stochastic model

We implement a stochastic model based on Markov chains [21] to describe the epidemic dynamics at the population level, which consist of demographics and new infections with either drug-sensitive (DS) or drug-resistant (DR) virus. The stochastic model keeps a record of discrete changes in the population sizes of susceptible (S), DS-infected (I_{DS}), and DR-infected (I_{DR}) individuals in a sufficiently small interval $(t, t + \Delta t)$. The time step Δt (fixed throughout the simulation) is chosen such that there is at most one event that occurs during this time interval. The events that are considered in the stochastic process are as follows: birth or death (i.e., entering or leaving the pool of susceptible individuals) and the infection of a susceptible individual with a DS or a DR strain. The corresponding probabilities of occurrence in the time interval $(t, t + \Delta t)$ are as follows: $\theta\Delta t$, $\mu S\,\Delta t$, $(\lambda_{DS}(t)/N)S\,\Delta t$, and $(\lambda_{DR}(t)/N)S\,\Delta t$, respectively, where $N = S + I_{DS} + I_{DR}$ is the total population size. Infected individuals die after their infectious period (see Section 2.1.3). We assumed that each infectious individual has his or her own transmission rate (which is a function of time since the individual got infected). Therefore, the function $\lambda_i(t)/N$ denotes the force of infection ($i \in \{DS, DR\}$), which is formed by the sum of all individual transmission rate profiles of infectious individuals at a given time t. A detailed definition is given in Section 2.1.6 below. Table 1 shows the rates for the various events in the model.

Table 1. Events and corresponding probabilities for the stochastic model.

Event	Change in population size	Probability
Birth	$S \to S + 1$	$r_1 = \theta\Delta t$
Death	$S \to S - 1$	$r_2 = \mu S\,\Delta t$
Infection with DS	$S \to S - 1$, $I_{DS} \to I_{DS} + 1$	$r_3 = \frac{\lambda_{DS}(t)}{N} S\,\Delta t$
Infection with DR	$S \to S - 1$, $I_{DR} \to I_{DR} + 1$	$r_4 = \frac{\lambda_{DR}(t)}{N} S\,\Delta t$
No event		$1 - \sum_{i=1}^{4} r_i$

For all simulations, we use parameter values as outlined in Table 2. For simplicity, we began all simulations with an initial population size of N_0 individuals, 5% of which are infected, and each of these initially infected individuals is assumed to be at any point in their first two years of infection. For the computational implementation we use $\Delta t = 0.25$ days, which we found to satisfy that the probability of occurrence of at least one event, $[\theta + \mu S + (\lambda_{DS}(t) + \lambda_{DR}(t))S/N]\Delta t$, does not exceed 1 throughout the simulation.

Table 2. Parameter values used in the stochastic model.

Parameter	Definition	Value
N_0	Initial population size	2000[a]
$1/\mu$	Average period as susceptible individuals	30 y
θ	Recruitment rate of susceptible individuals	67 y^{-1}[,b]

Note: [a]For computational convenience. [b]It approximates $\mu \times N_0$, so that N_0 is the stationary population size in the absence of disease.

2.1.2. Distribution of set-point viral loads

The heterogeneity of the population is taken into account through the variability of the SPVL. Examples of such variability for two cohorts are shown in Figure 1A (adapted from [18]). The SPVL are important as they determine the between-host transmission rate and the infectious period [18]. These SPVL distributions are well described by skew-normal distributions [18]. For our analysis, we use the cohort distributions in Figure 1B, which were generated from skew-normal distributions, with means varying from $10^{3.75}$ to $10^{5.75}$ (using the parameters, other than the mean, estimated in [18] corresponding to the Zambian cohort).

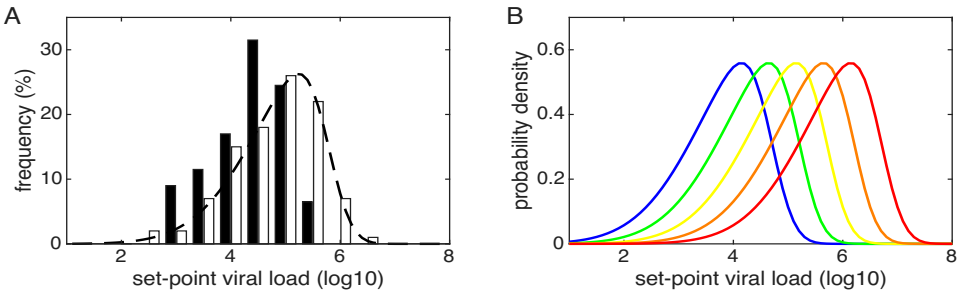

Figure 1. A) Distributions of SPVL for a Zambian (white bars) and an Amsterdam (black bars) cohorts; adapted from Fraser et al. 2007. The skew-normal distribution is fitted to the Zambian cohort (broken curve). B) Skew-normal distributions with means $10^{3.75}$ (blue), $10^{4.25}$ (green), $10^{4.75}$ (yellow), $10^{5.25}$ (orange), and $10^{5.75}$ (red). All other parameters correspond to the Zambian cohort (black broken curve).

2.1.3. Transmission rates and infectious periods

An individual infected with HIV goes through three stages of infection: primary, asymptomatic, and Acquired Immunodeficiency Syndrome (AIDS) [22]. These three stages of the infectious period are used to obtain each transmission rate profile. The duration of the primary and AIDS stages in our model are fixed to 2.9 and 9 months, respectively, according to estimates from [22]. The duration of the asymptomatic period D in years is determined by the SPVL V, which is given by the Hill's function as follows:

$$D(V) = \frac{D_{max} D_{50}^{D_k}}{V^{D_k} + D_{50}^{D_k}}, \tag{2.1}$$

where D_{max} is the maximum duration in years, D_{50} is the viral load at which the duration is half its maximum, and D_k is the steepness of the decrease in duration as a function of the viral load [18]. The values for the parameters of function $D(V)$ are shown in Table 3. After going through all three stages of infection, an infected individual will die.

The rate of infection β, that is, the transmission rate per unit time (at asymptomatic stage), is also determined by the SPVL V, using the Hill's function as follows:

$$\beta(V) = \frac{\beta_{max} V^{\beta_k}}{V^{\beta_k} + \beta_{50}^{\beta_k}}, \tag{2.2}$$

where β_{max} is the maximum infection rate per annum, β_{50} denotes the viral load at which the infectiousness is half its maximum, and β_k is the steepness of the increase in infectiousness as a function of viral load [18]. The values for the parameters of function $\beta(V)$ are shown in Table 3. We use this transmission rate at the asymptomatic stage $\beta(V)$, for a given V, to compute the rates at the primary and AIDS stages. For the whole duration of the primary and AIDS stages, the transmission rate obtained with $\beta(V)$ is multiplied by the ratios of transmission hazards between stages 13.5 and 7.2, respectively, to reflect the findings that the infectiousness in these two stages are greater than for the asymptomatic stage – and such difference is not explained by the viral load [22]. Then, the definition of the transmission profile $\rho(\tau)$ is given by the following:

$$\rho(\tau) = \begin{cases} 13.5 \cdot \beta(V), & \tau < 2.9 \\ \beta(V), & 2.9 \le \tau \le 2.9 + D(V) \\ 7.2 \cdot \beta(V), & 2.9 + D(V) < \tau \le 2.9 + D(V) + 9 \end{cases} \tag{2.3}$$

where τ is the time, in months, since the onset of infection. Figure 2A (dotted gray curve) shows an example of a transmission rate profile of an infected individual. In particular, note that the transmission rate is constant in each of its three stages (i.e., a step function), and it is greatest and shortest on its first (primary) stage.

Table 3. Parameter values used for transmission rates and infectious period functions[a].

Parameter	Definition	Value
D_{max}	Maximum infectious period	25.4 y
D_{50}	Viral load at which duration is half its maximum	3058 copies/ml
D_k	Steepness of decrease in duration	0.41
β_{max}	Maximum infection rate per annum	0.317 y^{-1}
β_{50}	Viral load at which infectiousness is half its maximum	13,938 copies/ml
β_k	Steepness of increase in infectiousness	1.02

Note: [a]As reported in [18].

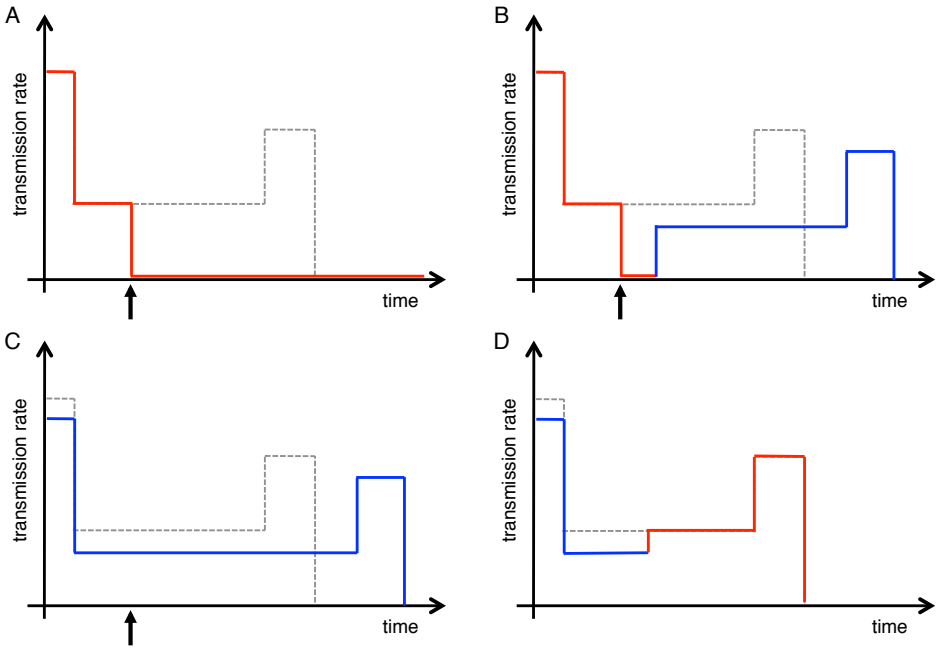

Figure 2. Graphical examples of transmission profiles. The transmission rate of an infected individual as a function of time are graphed for the DS strain (red) and the DR strain (blue) for the following cases: A) DS initial infection with successful ART; B) DS initial infection with failing ART; C) DR initial infection with ART; and D) DR initial infection without ART. Dotted grey curves represent the transmission rate of the DS strain in the absence of ART. The vertical arrow indicates the start of ART.

2.1.4. Start of ART

In an individual that is going to receive ART, we assume that ART is initiated as soon as a patient is diagnosed (aligned to the current WHO recommendation [4]). Hence, we translate the waiting time for ART initiation into the time from HIV infection to diagnosis. We evaluate two gamma probability distributions for the time from HIV infection to diagnosis (Figure 3A). The first distribution is fitted to observed data from 2018 [5]. Peruski et al. reported a median time of 39.5 months (3.29 years) from HIV infection to diagnosis, while 35.9% of people with HIV received their diagnosis within one year of infection [5]. Additionally, they reported that 1 in 4 people with HIV diagnosed in 2016 had the virus for more than 7 years before diagnosis [5]. Therefore, the data we use to fit the gamma distribution $F_1(t)$ are $F_1(1) = 0.359$, $F_1(3.29) = 0.5$, and $F_1(7) = 0.75$ (red dots in Figure 3A). In order to take the WHO recommendation for earlier diagnosis into account, we consider a second distribution for the time from infection to diagnosis, thereby assuming a median time of 0.5 years and that 75% of people are diagnosed within one year of infection. Then, the data we use to fit the gamma distribution $F_2(t)$ are $F_2(0.5) = 0.5$ and $F_2(1) = 0.75$ (blue dots in Figure 3A). For comparison purposes, in Figure 3B, we show the time it takes the $CD4^+$ count to reach less than 350 cells/mm^3, which used to be the criterion for ART initiation [19] that was found to be dependent on the viral load at baseline [23]. This time was found to range from a median of 0.7 years in those with a viral load greater than 500,000 copies/mL to a median of 4.7 years in those with less than 1000 copies/mL [23].

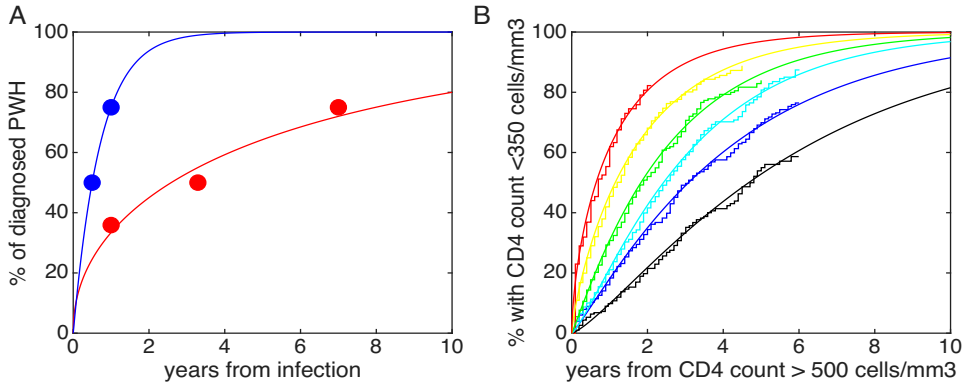

Figure 3. A) Probability distributions for time until ART starts, based on the time from HIV infection to diagnosis, with median either 3.29 (red) or 0.5 years (blue). Dots represent the assumed probabilities and the curves are the corresponding adjusted gamma distributions. B) Percentage of patients with CD4 count less than 350 cells/mm^3 over time by baseline viral load. Clinical data (stairs plot; from The UK Collaborative HIV Cohort Study Steering Committee, 2007) and corresponding adjusted gamma distribution (solid curves) for baseline viral loads: <1000 (black), 1000–10,000 (blue), 10,000–50,000 (light blue), 50,000–100,000 (green), 100,000–500,000 (yellow), and >500,000 (red) copies/mL.

2.1.5. Drug resistance

Resistance to a drug carries a fitness cost to the viral strain. We interpret this fitness cost in terms of the SPVL, similar to the concept of replication capacity [24]; however, the fitness cost could also be directly measured in terms of the transmission [11]. In our simulations, the SPVL, once sampled, is decreased by a proportion k. We take $k = 0.8$, although the replication capacity significantly varies for different mutation classes [24]. This number is consistent with the proportion obtained when we run simulations of the within-host model with parameters estimated in [25]. Our assumption that the fitness cost of a DR strain reduces its SPVL by a proportion k implies an implicit crucial role of the SPVL in DR amplification, as a higher SPVL ends up with a higher DR transmission.

Moreover, we assume that the administration of ART could only fail due to the emergence of a DR strain. DR is not the only reason why ART could fail, but ART failure is mostly caused by inadequate therapy adherence and suboptimal drug regimens, which are the same main sources for DR emergence. The failure of ART is assumed to happen with some probability p_{DR}^{emer}. For an individual that will develop DR, we assume that the probability that the DS strain has been replaced by time τ (from the moment the individual got infected) is given by the following:

$$Prob_{DS}^{rep}(\tau) = NormalCDF(\tau; \mu_s, \sigma_s), \tag{2.4}$$

that is, a normal probability distribution with parameters μ_s and σ_s, where $\mu_s = T_{ART} + LDS_{50}$. The choice of a normal probability distribution for this waiting time is solely for simplicity. T_{ART} denotes the time at which ART is initiated; thus, as previously discussed, it follows a gamma distribution (Figure 3). LDS_{50} is the average time that it takes the DS viral load to go from 50% the total viral load to under the detection level (i.e., it has been replaced by the DR strain). The standard deviation σ_s is chosen such that $Prob\{T_{ART} < \tau < \mu_s\} = 0.49$, that is, the probability that the DS strain has been replaced earlier

than the initiation of ART is small (i.e., $Prob\{\tau < T_{ART}\} = 0.01$). The later implies, by symmetry of the normal distribution, that $Prob\{\mu_s < \tau < LDS_{50}\} = 0.49$, which, in part, explains our choice of μ_s. LDS_{50} and σ_s are estimated using an extension of the stochastic model for viral dynamics within a host in [25] to include forward mutations and ART [26]. The best estimates found were $LDS_{50} = 71$ days and $\sigma_s = 30.5$ days. From the definition of $Prob_{DS}^{rep}(\tau)$, we have that the probability that an individual initially infected with a DS strain, that develops drug resistance after receiving ART, transmits the DR strain at time τ is given by the following:

$$Prob_{DR}^{trans}(\tau) = Prob_{DS}^{rep}(\tau).$$

Then, the transmission rate (as a function of time) for the DR strain is the product of the transmission rate (corresponding to a DR infection) and $Prob_{DR}^{trans}(\tau)$. Likewise, the transmission rate for the DS strain is the product of the transmission rate (corresponding to a DS infection) and $1 - Prob_{DR}^{trans}(\tau)$.

A transmitted DR strain will eventually be replaced by a DS strain if the individual does not receive ART. We take the probability that a DR strain has been replaced by a DS strain by time τ to be a gamma probability distribution fitted to the data (corresponding to thymidine analog-associated mutations, TAM) provided in [27]. The estimated parameters for such a gamma distribution are $a_{TAM} = 0.85131$ and $b_{TAM} = 3024.65$ (which gives a mean of 7 years and a median of 4.5 years for the replacement of a DR strain). Then, we take the probability that an individual initially infected with a DR strain and who does not receive ART, transmits the DR strain at time τ as follows:

$$Prob_{DR}^{trans}(\tau) = 1 - NormalCDF(\tau; \mu_r, \sigma_r) \tag{2.5}$$

where $\mu_r = T_{DR}^{rep} - LDR_{50}$. Here, $T_{DR}^{rep} \sim Gamma(a_{TAM}, b_{TAM})$, and LDR_{50} is the average time that it takes the DR viral load to go from 50% the total viral to under the detection level (i.e., it has been replaced by the DS strain). The standard deviation σ_r is chosen such that $Prob\{\mu_r < \tau < \mu_r + LDR_{50}\} = 0.49$ (i.e., approximately 0.50). We use the values $LDR_{50} = 71$ days and $\mu_r = 30.5$ days (same as for DS strain), which agree with simulations obtained from the model for viral dynamics within a host in [25]. As before, the transmission rate (as a function of time) for the DR strain is the product of the transmission rate (corresponding to a DR infection) and $Prob_{DR}^{trans}(\tau)$. Likewise, the transmission rate for the DS strain is the product of the transmission rate (corresponding to a DS infection) and $1 - Prob_{DR}^{trans}(\tau)$.

2.1.6. Transmission rate profiles and the force of infection

The force of infection for each strain, $\lambda_{DS}(t)$ or $\lambda_{DR}(t)$, as a function of time is the result of adding together the individual transmission profiles that correspond to that viral strain of all infected individuals, lined up according to their time of infection. Therefore, the DS force of infection is given by the following:

$$\lambda_{DS}(t) = \sum_{j=1}^{I} \hat{\rho}_{DS}(t; j), \tag{2.6}$$

where j represents each of the total I infected individuals, and the aligned transmission profile $\hat{\rho}_{DS}(t; j)$ for the j-th infected individual is defined as

$$\hat{\rho}_{DS}(t; j) = \begin{cases} 0, & t < T_j \\ \rho_{DS}(t - T_j; j), & T_j \le t \le T_j + D_j \\ 0, & T_j + D_j < t \end{cases} \tag{2.7}$$

which takes the j-th individual's time of infection T_j, infection period D_j, and transmission profile (from the time of infection) $\rho_{DS}(\tau; j)$ into account. The definition is analogous for $\lambda_{DR}(t)$.

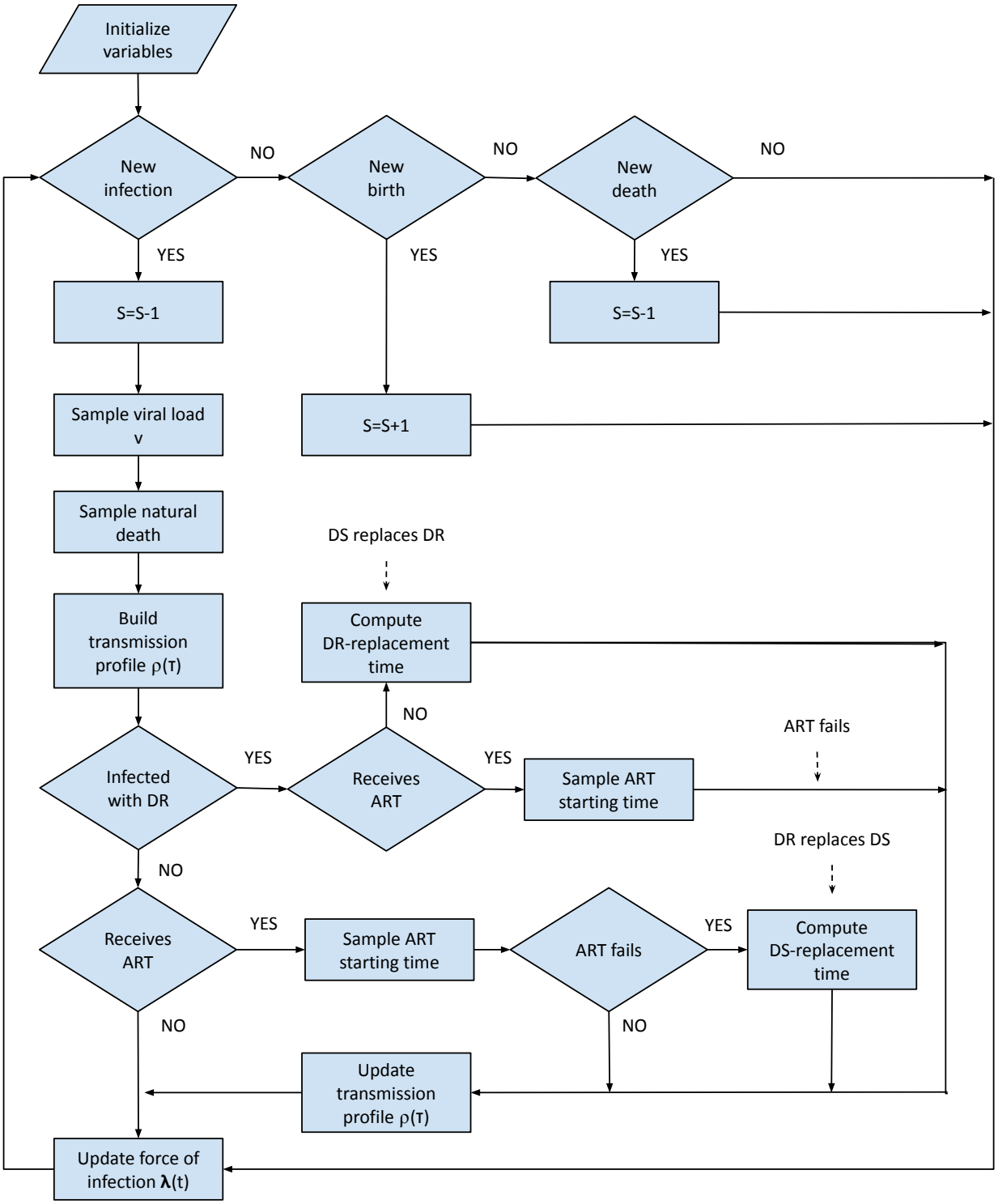

Figure 4. Flowchart describing the computational simulations.

The computational simulations are described in the flowchart diagram in Figure 4. For each new infected individual, a set-point viral load is sampled from a distribution of SPVL. Likewise, a timing of natural death is sampled in all cases from an exponential distribution with parameter μ (so it is equivalent to a natural death in the stochastic process) to ensure that an infected individual does not live longer that an uninfected individual. With this information, a first draft of the individual's transmission profile is created: the transmission rate and the duration of the asymptomatic stage are determined from Eqs 2.1 and 2.2, respectively, using the previously sampled SPVL; for the primary and AIDS stages (which last 2.9 and 9 months, respectively), the transmission rate obtained for the asymptomatic stage is multiplied by 13.5 and 7.2, respectively (Section 2.1.3). The individual transmission profile, so far, results from putting together all three stages (Figure 2A, dotted gray curve). If the new infection was with the DS strain (instead of the DR strain), then it is randomly decided with probability p_{ART} whether the individual will receive ART. If the individual will not receive ART, then the computed transmission profile is the final profile and it is added to the force of infection $\lambda_{DS}(t)$.

If, on the other hand, the individual is going to receive ART, then the onset of ART is determined from the corresponding gamma distribution (Section 2.1.4). Whether ART is successful is randomly determined with the probability p_{ART}^{succ}. If ART is indeed successful, then the viral load goes down (to negligible figures), so that the transmission rate is practically zero and the duration of infection extends to its maximum (D_{max}). Figure 2A shows a sketch of such a transmission profile. If, on the contrary, ART fails, then it is assumed that it is due to the emergence of a DR strain; thus, $p_{ART}^{succ} = 1 - p_{DR}^{emer}$. The viral load of the DR strain (the sampled SPVL decreased by the proportion k), together with the probability of transmitting the DR strain (as discussed in Section 2.1.5), determine the DR transmission rate function (the DS transmission rate includes the probability of transmitting the DS strain, i.e., one minus the DR probability). Then, the individual transmission profile is obtained by taking all the above into account, including the three stages of infection as before. Note that in this case, such a profile is composed of a DS transmission and a DR transmission function, and each of them is added to their corresponding forces of infection ($\lambda_{DS}(t)$ or $\lambda_{DR}(t)$). Moreover, as we do not assume DR testing nor ART switching, the same ART continues to be administered and the DR strain continues to be transmitted. Figure 2B illustrates such a transmission profile.

If the individual is initially infected with a DR strain and is going to receive ART, then ART is always going to fail (as there is a single ART regimen). In any case, the timing of ART initiation is determined (Section 2.1.4). While ART is absent (has not been initiated yet), there is a probability $1 - Prob_{DR}^{trans}(\tau)$ that an emergent DS strain is transmitted. This probability will be zero when ART starts and so only the DR strain is transmitted. Figure 2C shows a simplified version of this case. If a DR-infected individual does not receive ART at all, then the DR strain will eventually be replaced by the DS strain (according to the probability of transmitting the DS strain given by one minus equation 2.5). Figure 2D shows a graphical representation of this transmission profile. As before, the resulting DS and DR transmission profiles are added to their corresponding force of infection ($\lambda_{DS}(t)$ or $\lambda_{DR}(t)$).

3. Results

For all simulations (unless told otherwise), we assume a skew-normal distribution of SPVL with a mean of $10^{4.75}$ (as shown in Figure 1B). Figure 5A,B show the simulated prevalence without ART (red), and with ART (90% coverage) in the absence (green) and in the presence (yellow) of drug resistance,

for each of the two time distributions of ART initiation, respectively. For the ART initiation with a median of 3.29 years, without ART the prevalence reaches an endemic state, which is at most halved with a 90% ART coverage, although it could also be as high as without ART (Figure 5A). When the probability of emergence of drug resistance is 0.01, having a 90% ART coverage produces an endemic state as high as the endemic state without DR (Figure 5A). The latter indicates that DR does not have a significant impact on the epidemic for this ART initiation scheme. On the other hand, if ART is administered earlier (using the time distribution with median of 0.5 years), then the epidemic does not take off when 90% of infected individuals receive ART and DR does not emerge (green region in Figure 5B), but the epidemic could eventually rise to an endemic state higher than without ART if there is a 0.01 probability of DR emergence (yellow region in Figure 5B). In other words, under these assumptions, using ART could create more infections because of DR. However, in this case, we also found the possibility that the epidemic goes extinct, which is a scenario that is absent if the therapy is administered much later. Additionally, we notice that the use of ART creates a higher variability in the simulations, that is, the plotted bands are wider for the simulations with ART than without ART (red). This is due to all the random events associated with the use of ART, such as the start of treatment and the emergence of resistance.

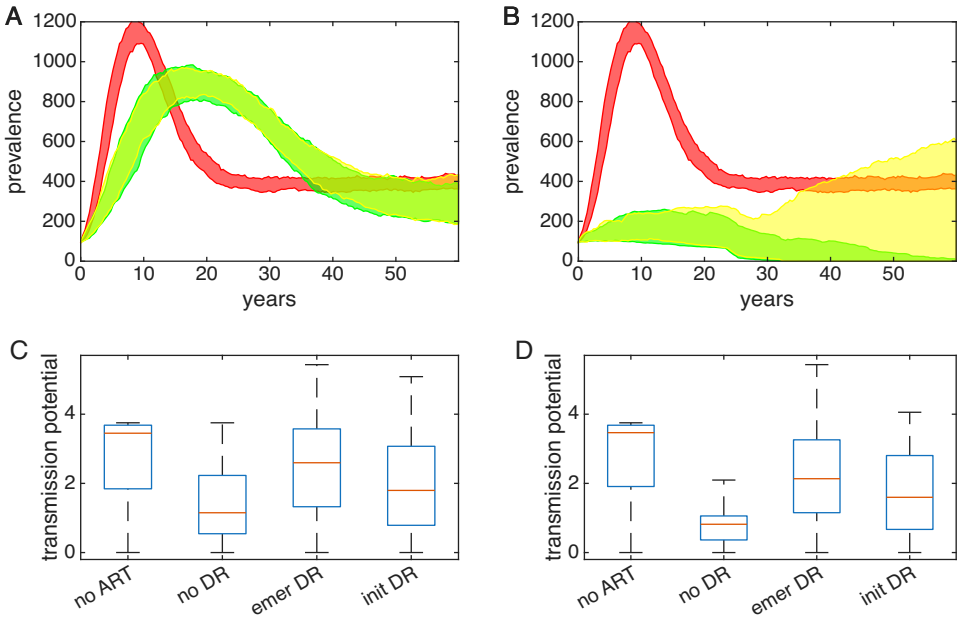

Figure 5. Prevalence without ART (red) and with ART (90% coverage) in the absence of DR (green) and with DR (yellow; $p_{DR}^{emer} = 0.01$) for a time distribution of ART initiation with median A) 3.29 years and B) 0.5 years. Each region denotes 5- to 95-percentiles of values obtained in simulations. Boxplots of transmission potential distributions for individuals without ART (no ART) and with ART without DR (no DR), with DR emergence (emer DR), and with initial infection with the DR strain (init DR) for a time distribution of ART initiation with median C) 3.29 years and D) 0.5 years; outliers are not included.

Figure 5C,D show the distributions of transmission potentials, defined as the integral of the transmission profile over the lifetime of an infected individual [18], associated with the previous cases. In particular, the transmission potential balances the effect of increasing the transmission rate for a shorter

duration when the viral load increases. We calculated these distributions of transmission potentials by sampling the SPVL distribution (a sample size of 10,000), computing the transmission profile for each sampled viral load, and determining the integral of each of these transmission functions. When ART is used, the start of ART has to be considered to compute the transmission profile for each sampled viral load. For the case where DR is present, we take two possibilities into account: when DR emerges after ART is given and when the infection starts with a DR strain. When ART is administered, there is a marked drop in transmission potential, from a median value of 3.4 (no ART) to around 1 (no DR), as shown in Figure 5C,D. This decrease in transmission potential is more significant when ART is started earlier (5D). However, if the DR strain emerges, then the transmission potential (emer DR) increases to levels similar to the case when ART is absent (although the median is only 2.4, the third quartile is almost the same and the maximum value is considerably bigger). The increase in transmission potential when the infection starts with the DR strain (init DR) is smaller than when DR emerges after ART, but is still greater and more widespread (Figure 5C,D). An earlier ART initiation only produces a slight decrease in the transmission potential when DR is present (Figure 5D). These observations on transmission potentials help to explain the resulting dynamics in Figure 5A,B.

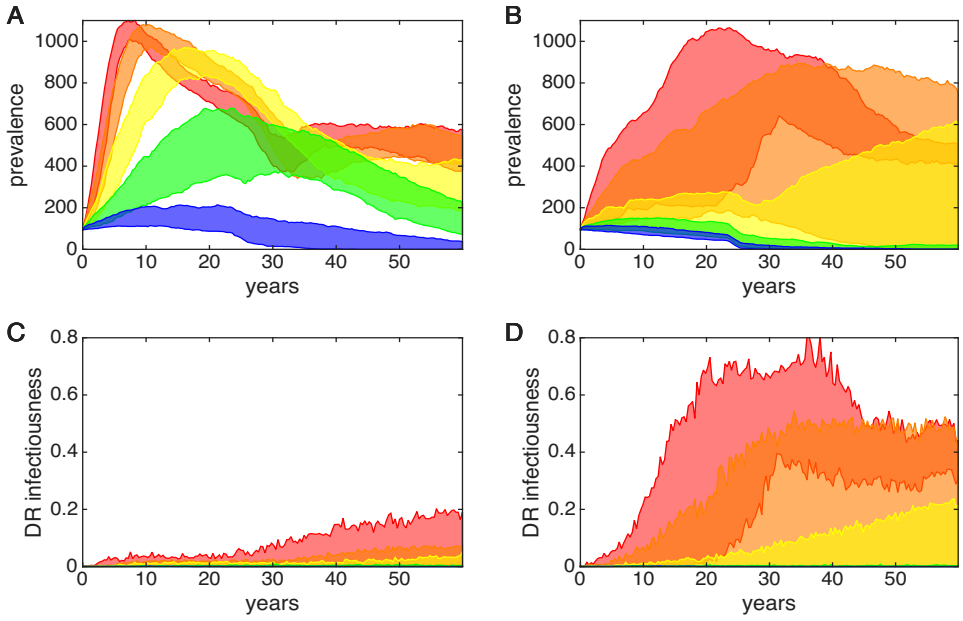

Figure 6. Prevalence for generic cohorts with skew-normal distributions for SPVL with means $10^{3.75}$ (blue), $10^{4.25}$ (green), $10^{4.75}$ (yellow), $10^{5.25}$ (orange), and $10^{5.75}$ (red), in the presence of DR for a time distribution of ART initiation with median A) 3.29 years and B) 0.5 years. Infectiousness (force of infection) due to DR strains for a time distribution of ART initiation with median C) 3.29 years and D) 0.5 years. Each region denotes 5- to 95-percentiles of values obtained in simulations.

The effect of varying the SPVL on the epidemic growth and its connection to drug resistance can be appreciated in Figure 6, where the epidemic dynamics are shown for each of the generic cohorts (with $p_{ART} = 0.90$ and $p_{DR}^{emer} = 0.01$), which correspond to each of the two time distributions of ART initiation. The epidemic does not even rise for a mean viral load of $10^{3.75}$ for any of the two ART initiation distributions (Figure 6A,B). If ART is started earlier, then the same occurs for a mean viral

load of $10^{4.25}$, and epidemic extinction is also possible for a mean of $10^{4.75}$ and $10^{5.25}$ but not for $10^{5.75}$ (Figure 6B). However, there is the possibility that prevalences for both $10^{4.75}$ and $10^{5.25}$ end up being higher than for $10^{5.75}$, which indicates a reduction in the transmission potential when the set-point of the viral loads gets too high (Figure 6B). For an ART start with a median of 3.29 years, all mean viral loads from $10^{4.25}$ to $10^{5.75}$ produce an endemic state in an increasing manner, although there is no significant difference between the last two (Figure 6A). The force of infection from the DR strain is consistently higher for populations with higher mean viral loads (Figure 6C,D), which suggests that the excess of infected individuals (as seen in Figure 6A,B) is due, in part, to the presence of DR. Moreover, an earlier initiation of ART produces a greater force of infection from the DR strain (higher levels in Figure 6D than in 6C), which emphasizes the role of drug pressure on the DR dynamics. Additionally, we explore the role of the variance of the SPVL distribution on the epidemic dynamics and find that a bigger variance (i.e., a more heterogeneous population) produces more DR infectiousness but similar levels of prevalence (figure not shown).

Moreover, we look at the effect of varying the ART coverage (p_{ART}) on the epidemic dynamics, thereby keeping the probability of DR emergence at $p_{DR}^{emer} = 0.01$ for each of the two time distributions of ART initiation. In general, increasing ART coverage delays and shortens the peak of infection prevalence (Figure 7A,B). Moreover, when ART is started with a median of 3.29 years, the steady state of prevalence stays very similar for all ART coverages, with a possible smaller prevalence for 100% ART coverage (Figure 7A). A common epidemic state is also found when ART is started earlier for 0 to 50% ART coverage, with a smaller stationary value for 75% ART coverage (Figure 7B). However, for a 100% ART coverage, there is a wider range of possible results, from epidemic extinction to an endemic state higher than for smaller ART coverages (Figure 7B).

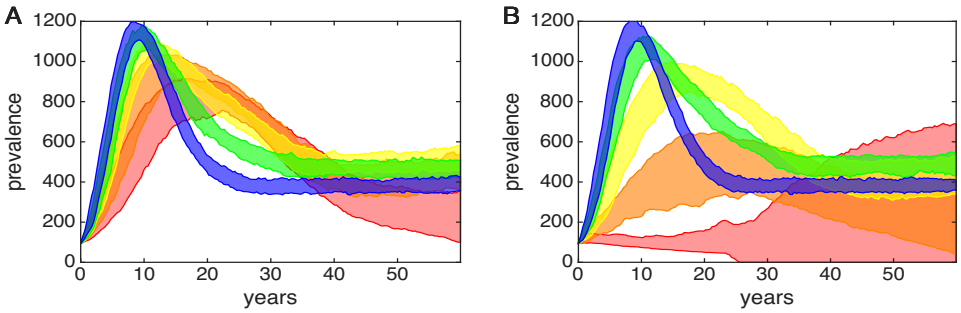

Figure 7. Prevalence with ART coverage, p_{ART}, of 0 (blue), 25 (green), 50 (yellow), 75 (orange), and 100% (red), in the presence of DR for a time distribution of ART initiation with median A) 3.29 years and B) 0.5 years. Each region denotes 5- to 95-percentiles of values obtained in simulations.

We examine the role of the probability of DR emergence, P_{DR}^{emer}, by varying it from its baseline value of 0.01 (Figure 8), for each of the two time distributions of ART initiation. For the ART initiation with a median of 3.29 years, the stationary value of the prevalence is very similar for probabilities of DR emergence of 0.01 and 0.10, and as much as half smaller for the probability of 0.001 (Figure 8A). For the case of ART initiation with an earlier median of 0.5 years, $P_{DR}^{emer} = 0.10$ always results in an epidemic state (higher than the epidemic state without ART) while $P_{DR}^{emer} = 0.001$ always drives the epidemic to extinction (Figure 8B). However, a 0.01 probability of DR emergence could end up

in epidemic extinction or could produce an endemic state as high as for $P_{DR}^{emer} = 0.10$ (Figure 8B). Therefore, shortening the waiting time for ART administration could stop the epidemic whenever the probability of DR emergence is sufficiently small.

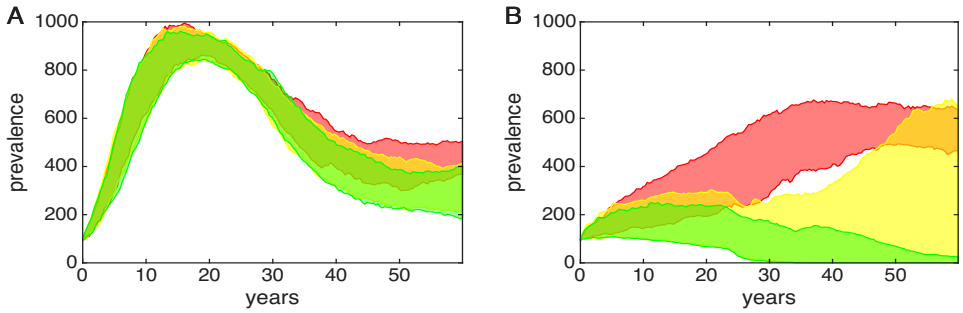

Figure 8. Prevalence with probability of DR emergence, P_{DR}^{emer}, equal to 0.001 (green), 0.01 (yellow), and 0.10 (red) for a time distribution of ART initiation with median A) 3.29 years and B) 0.5 years. Each region denotes 5- to 95-percentiles of values obtained in simulations.

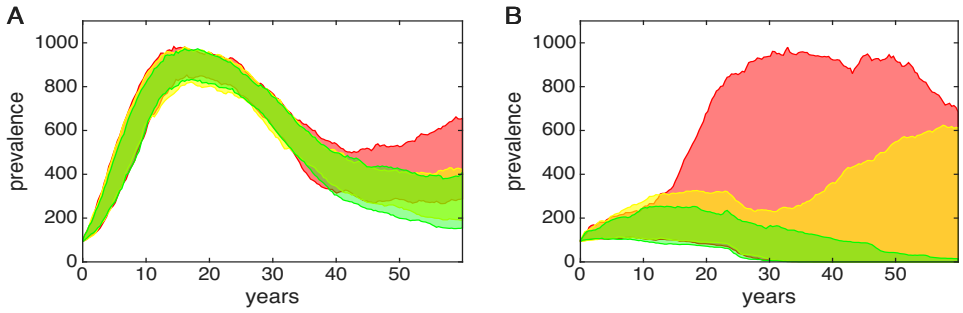

Figure 9. Prevalence for simulations with DR relative fitness of 0.7 (green), 0.8 (yellow), and 0.9 (red) in the presence of DR for a time distribution of ART initiation with median A) 3.29 years and B) 0.5 years. Each region denotes 5- to 95-percentiles of values obtained in simulations.

We evaluate the effect on the epidemic dynamics due to the relative fitness of the DR strain by varying the relative fitness parameter k between the values 0.7, 0.8 (baseline), and 0.9 (Figure 9) for each of the two time distributions of ART initiation. We consider a 90% ART coverage ($p_{ART} = 0.90$) and $P_{DR}^{emer} = 0.01$ for these simulations. For a late start of ART, the prevalence of infections reach an endemic state for all the relative fitness values in an increasing order (Figure 9A). On the other hand, if ART is initiated earlier, then the epidemic does not rise for a relative fitness of 0.7 and could be driven to extinction for both 0.8 and 0.9 (Figure 9B). However, for these two relative fitness, there is also the possibility that the epidemic reaches an endemic state higher than the prevalence without ART. Those maximum endemic states are, in fact, greater than their corresponding values for a late ART start. These results suggest that the minimum relative fitness for the DR strain that would be able to sustain an epidemic when 90% of patients receive ART has to be greater than 0.7. Furthermore, we studied the role of the speed at which the DR strain is replaced by a DS strain in the absence of treatment. For this purpose, we replaced the probability given in Eq 2.4 by a gamma probability distribution fitted to data of persistence of drug mutation classes. Of the six drug-mutation classes shown in [27], we considered

three classes: Lamivudine/emtricitabine-asociated (LEA) mutations, thymidine analog-asociated (TAA) mutations, and protease inhibitor (PI) mutations. We found no significant difference on the prevalence of infection between the three drug mutation classes, although the prevalence is slightly smaller for LEA mutations, which is associated with a smaller DR infectiousness (figure not shown).

4. Discussion

There are several sources of population heterogeneity that may have an effect on an epidemic and the DR dynamics, such as a heterogeneous infection susceptibility and a heterogeneous infectious rate. The relative impact of each on them should be evaluated. However, in this study, we considered the SPVL as the only source of population heterogeneity. The importance of the viral load is evident as it determines the transmission rate and infection period [18]. Moreover, the viral load distributions are not necessarily the same for two populations [18]. In this study, we used generic distributions built by varying their mean viral loads (based on examples reported in the literature) to show the effect they could have in the overall epidemic when DR is present. We found that the presence of DR strains could result in a similar infection prevalence between two populations with different mean viral loads, and even in a higher endemic state for a smaller mean SPVL. The latter can be explained by the viral transmission potential, because higher mean viral loads imply a shorter survival. However, populations with higher mean viral loads will end up with a higher DR infectiousness, which is, in part, due to our assumption that the SPVL of the DR strain is a proportion of the wild-type SPVL. In any case, our results indicate that considering the viral load distributions of the population under study, whenever available, would be valuable when studying the DR dynamics.

Our findings suggests that an early start of ART has the potential to stop the epidemic, even in the presence of DR and under the assumptions of no DR screening and no ART regimen switching (which amplify the DR epidemic). However, DR also creates the possibility of having more cases than without ART (although not surprising because of the aforementioned assumptions). The later could be avoided if the probability of DR emergence is sufficiently small. On the other hand, if the wait of ART initiation is longer (as the one reported in [5]), then the therapy will not be enough to eradicate the epidemic, although it may reduce it even if DR is present. Current recommendations indicate to initiate ART immediately, or as soon as possible, after HIV diagnosis in order to increase ART uptake, accelerate viral suppression for patients, and improve virologic suppression rates among people with HIV [28]. However, in high- and upper-middle-income countries, the mean time from HIV infection to diagnosis was estimated as 3 years [29]. This waiting time until diagnosis is expected to decline as the WHO advocates for an increase in HIV testing rates [4].

The effect of varying ART coverage was also studied, and we found that there could be more cases when more patients receive ART, especially when the therapy is started earlier. This pattern was also found previously with a model based on partial differential equations [14], and for models for the dynamics of influenza infections [30, 31]. As the negative effect of increasing ART coverage is likely due to the simplified assumption of not switching ART regimens (or at least stopping therapy) when DR is present, other control measures such as DR screening and alternate ART schemes should be included in the model to give a more comprehensive analysis. Such control measures could certainly be added to our modeling framework for their assessment. Additionally, in our analysis, we identified the existence of a minimum relative fitness for the DR strain that guarantees its prevalence, which of

course is expected. We could take advantage of this finding as we could determine whether a particular ART coverage could eliminate the epidemic (or drive it to minimum levels) for a given specific relative fitness from a combination of ART and a DR strain.

The modeling framework employed, thereby combining a stochastic model and Monte Carlo simulations, has several advantages from our point of view. It is similar to an individual-based model but still more efficient, as only infected (and not susceptible) individuals are tracked. Using a stochastic scheme, which is based on Markov chains, allows us to study the dynamics variability due to probabilistic events, such as the emergence of DR. Thanks to the Monte Carlo simulations, we can make use of information from observed probability distributions (i.e., data from the literature), such as distributions of SPVL, time of ART initiation, or time of DR emergence for a specific drug class. These probability distributions provide the heterogeneity of the population as well as the stochasticity of certain events. The latter determines the main difference and advantage over a deterministic age-of-infection model with a heterogeneous population. On the other hand, a clear downside of our framework is related to the computational effort needed to run the simulations, as there is an inverse relationship between the population size and the time step that guarantees a correct implementation. Furthermore, we make use of several assumptions to simplify the model. For instance, we assume that ART that fails viral suppression is only due to the emergence of a DR strain. By not including other causes of ART failure, we are indeed overestimating the benefits of therapy. A simple way to include ART failure in our model, not caused by DR, is by considering an effective ART coverage, that is, decreasing the selected ART coverage (the percentage of individuals that receive ART) by some factor (representing the percentage of treated individuals that achieve viral suppression). The model could be further improved by considering a second-line ART that could be implemented once the first-line ART fails to suppress the virus (for any reason) or DR strains resistant to a specific line of ART. Nevertheless, the previous assumptions would significantly increase the complexity of the model, which would make its implementation and analysis more demanding.

All findings shown here stress the importance of considering the dynamics of DR in heterogeneous populations of individuals infected with HIV. Perhaps more importantly, we present a simple modeling framework for heterogeneous populations that is flexible enough to be easily extended to include other factors and even other infectious diseases. Some examples of such factors are alternative criteria to start ART in a patient, the use of clinical drug-resistance tests, and the use of pre-exposure profilaxis.

Use of AI tools declaration

The authors declare they have not used artificial intelligence (AI) tools in the creation of this article.

Acknowledgments

The author would like to thank S. Bonhoeffer and D. Olmos Liceaga for valuable comments on this work.

Conflict of interest

The author declares there is no conflict of interest.

References

1. *Joint United Nations Programme on HIV/AIDS*, *Global HIV & AIDS Statistics—Fact Sheet*, 2025. Available from: https://www.unaids.org/sites/default/files/2025-07/2025_Global_HIV_Factsheet_en.pdf.
2. *Joint United Nations Programme on HIV/AIDS*, *World AIDS Day Report*, 2011. Available from: http://www.unaids.org/en/resources/documents/2011/.
3. M. S. Cohen, Y. Q. Chen, M. McCauley, T. Gamble, M. C. Hosseinipour, N. Kumarasamy, et al., Prevention of HIV-1 infection with early antiretroviral therapy, *N. Engl. J. Med.*, **365** (2011), 493–505. https://doi.org/10.1056/NEJMoa1105243
4. *World Health Organization*, *Consolidated Guidelines on HIV Prevention, Testing, Treatment, Service Delivery and Monitoring: Recommendations for A Public Health Approach*, 2021. Available from: https://www.who.int/publications/i/item/9789240031593.
5. A. H. Peruski, B. Wu, L. Linley, K. P. Delaney, E. A. DiNenno, A. S. Johnson, Time from HIV infection to diagnosis in the U.S., 2014–2018, *Am. J. Prev. Med.*, **61** (2021), 636–643. https://doi.org/10.1016/j.amepre.2021.04.015
6. F. Clavel, A. J. Hance, HIV drug resistance, *N. Engl. J. Med.*, **350** (2004), 1023–1035. https://doi.org/10.1056/NEJMra025195
7. S. Margeridon-Thermet, R. W. Shafer, Comparison of the mechanisms of drug resistance among HIV, hepatitis B, and hepatitis C, *Viruses*, **2** (2010), 2696–2739. https://doi.org/10.3390/v2122696
8. D. R. Bangsberg, E. P. Acosta, R. Gupta, D. Guzman, E. D. Riley, P. R. Harrigan, et al., Adherence-resistance relationships for protease and non-nucleoside reverse transcriptase inhibitors explained by virological fitness, *AIDS*, **20** (2006), 223–231. https://doi.org/10.1097/01.aids.0000199825.34241.49
9. R. M. Kagan, J. D. Baxter, T Kim, E. M. Marlowe, HIV-1 drug resistance trends in the era of modern antiretrovirals: 2018–2024, in *Open Forum Infectious Diseases*, Oxford University Press, US, **12** (2025), 1–9. https://doi.org/10.1093/ofid/ofaf446
10. S. Hué, R. J. Gifford, D. Dunn, E. Fernhill, D. Pillay, Demonstration of sustained drug-resistant human immunodeficiency virus type 1 lineages circulating among treatment-naive individuals, *J. Virol.*, **83** (2009), 2645–2654. https://doi.org/10.1128/JVI.01556-08
11. J. O. Wertheim, A. M. Oster, J. A. Johnson, W. M. Switzer, N. Saduvala, A. L. Hernandez, et al., Transmission fitness of drug-resistant HIV revealed in a surveillance system transmission network, *Virus Evol.*, **3** (2017), 1–12. https://doi.org/10.1093/ve/vex008
12. R. F. Baggaley, G. P. Garnett, N. M. Ferguson, Modelling the impact of antiretroviral use in resource-poor settings, *PLoS Med.*, **3** (2006), e124. https://doi.org/10.1371/journal.pmed.0030124
13. S. Blower, E. Bodine, J. Kahn, W. McFarland, The antiretroviral rollout and drug-resistant HIV in Africa: insights from empirical data and theoretical models, *AIDS*, **19** (2005), 1–14.
14. R. A. Saenz, S. Bonhoeffer, Nested model reveals potential amplification of an HIV epidemic due to drug resistance, *Epidemics*, **5** (2013), 34–43. https://doi.org/10.1016/j.epidem.2012.11.002

15. M. Shen, Y. Xiao, L. Rong, L. A. Meyers, S. E. Bellan, Early antiretroviral therapy and potent second-line drugs could decrease HIV incidence of drug resistance, *Proc. R. Soc. B*, **284** (2017), 20170525. https://doi.org/10.1098/rspb.2017.0525

16. V. Supervie, M. Barret, J. S. Kahn, G. Musuka, T. L. Moeti, L. Busang, et al., Modeling dynamic interactions between pre-exposure prophilaxis interventions and treatment programs: Predicting HIV transmission and resistance, *Sci. Rep.*, **1** (2011), 185. https://doi.org/10.1038/srep00185

17. B. G. Wagner, S. Blower, Universal access to HIV treatment versus universal 'test and treat': Transmission, drug resistance and treatment costs, *PLoS One*, **7** (2012), e41212. https://doi.org/10.1371/journal.pone.0041212

18. C. Fraser, T. D. Hollingsworth, R. Chapman, F. de Wolf, W. P. Hanage, Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis, *Proc. Natl. Acad. Sci.*, **104** (2007), 17441–17446. https://doi.org/10.1073/pnas.0708559104

19. *When To Start Consortium*, Timing of initiation of antiretroviral therapy in AIDS-free HIV-1-infected patients: A collaborative analysis of 18 HIV cohort studies, *Lancet*, **373** (2009), 1352–1363. https://doi.org/10.1016/S0140-6736(09)60612-7

20. J. M. Heffernan, L. M. Wahl, Monte Carlo estimates of natural variation in HIV infection, *J. Theor. Biol.*, **236** (2005), 137–153. https://doi.org/10.1016/j.jtbi.2005.03.002

21. L. J. S. Allen, A. M. Burgin, Comparison of deterministic and stochastic SIS and SIR models in discrete time, *Math. Biosci.*, **163** (2000), 1–33. https://doi.org/10.1016/S0025-5564(99)00047-4

22. T. D. Hollingsworth, R. M. Anderson, C. Fraser, HIV-1 transmission, by stage of infection, *J. Infect. Dis.*, **198** (2008), 687–693. https://doi.org/10.1086/590501

23. *The UK Collaborative HIV Cohort (CHIC) Study Steering Committee*, HIV diagnosis at CD4 count above 500 cells/mm^3 and progression to below 350 cells/mm^3 without antiretroviral therapy, *J. Acquir. Inmune Defic. Syndr.*, **46** (2007), 275–278. https://doi.org/10.1097/QAI.0b013e3181514441

24. J. Martinez-Picado, M. A. Martinez, HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: A review from the clinic and ex vivo, *Virus Res.*, **134** (2008), 104–123. https://doi.org/10.1016/j.virusres.2007.12.021

25. J. M. Kitayimbwa, J. Y. T. Mugisha, R. A. Saenz, Estimation of the HIV-1 backward mutation rate from transmitted drug-resistant strains, *Theor. Popul. Biol.*, **112** (2016), 33–42. https://doi.org/10.1016/j.tpb.2016.08.001

26. J. M. Kitayimbwa, J. Y. T. Mugisha, R. A. Saenz, The role of backward mutations on the within-host dynamics of HIV-1, *J. Math. Biol.*, **67** (2013), 1111–1139. https://doi.org/10.1007/s00285-012-0581-2

27. V. Jain, M. C. Sucupira, P. Bacchetti, W. Hartogensis, R. S. Diaz, E. G. Kallas, et al., Differential persistence of transmitted HIV-1 drug resistance mutation classes, *J. Infect. Dis.*, **203** (2011), 1174–1181. https://doi.org/10.1093/infdis/jiq167

28. *Panel on Antiretroviral Guidelines for Adults and Adolescents*, *Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV, Department of Health and Human Services*, 2025. Available from: https://clinicalinfo.hiv.gov/en/guidelines/adult-and-adolescent-arv. Accessed: February, 2026.

29. S. O. Gbadamosi, M. J. Trepka, R. Dawit, R. Jebai, D. M. Sheehan, A systematic review and meta-analysis to estimate the time from HIV infection to diagnosis for people with HIV, *AIDS Rev.*, **24** (2022), 32–40. https://doi.org/10.24875/AIDSRev.21000007

30. M. Lipsitch, T. Cohen, M. Murray, B. R. Levin, Antiviral resistance and the control of pandemic influenza, *PLoS Med.*, **4** (2007), e15. https://doi.org/10.1371/journal.pmed.0040015

31. S. M. Moghadas, C. S. Bowman, G. Rost, J. Wu, Population-wide emergence of antiviral resistance during pandemic influenza, *PLoS One*, **3** (2008), e1839. https://doi.org/10.1371/journal.pone.0001839

AIMS Press