

RESEARCH ARTICLE

Estimating epidemiological parameters of a stochastic differential model of HIV dynamics using hierarchical Bayesian statistics

Renee Dale^{1,2*}, BeiBei Guo¹

1 Department of Experimental Statistics, Louisiana State University, Baton Rouge, Louisiana, United States of America, **2** Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, United States of America

* rdale1@lsu.edu



OPEN ACCESS

Citation: Dale R, Guo B (2018) Estimating epidemiological parameters of a stochastic differential model of HIV dynamics using hierarchical Bayesian statistics. PLoS ONE 13(7): e0200126. <https://doi.org/10.1371/journal.pone.0200126>

Editor: Dimitrios Paraskevis, National and Kapodistrian University of Athens, GREECE

Received: November 16, 2017

Accepted: April 29, 2018

Published: July 25, 2018

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was partially funded by National Institute of Health (nih.gov) R01 AI12125903. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

Abstract

Current estimates of the HIV epidemic indicate a decrease in the incidence of the disease in the undiagnosed subpopulation over the past 10 years. However, a lack of access to care has not been considered when modeling the population. Populations at high risk for contracting HIV are twice as likely to lack access to reliable medical care. In this paper, we consider three contributors to the HIV population dynamics: at-risk population exhaustion, lack of access to care, and usage of anti-retroviral therapy (ART) by diagnosed individuals. An extant problem in the mathematical study of this system is deriving parameter estimates due to a portion of the population being unobserved. We approach this problem by looking at the proportional change in the infected subpopulations. We obtain conservative estimates for the proportional change of the infected subpopulations using hierarchical Bayesian statistics. The estimated proportional change is used to derive epidemic parameter estimates for a system of stochastic differential equations (SDEs). Model fit is quantified to determine the best parametric explanation for the observed dynamics in the infected subpopulations. Parameter estimates derived using these methods produce simulations that closely follow the dynamics observed in the data, as well as values that are generally in agreement with prior understanding of transmission and diagnosis rates. Simulations suggest that the undiagnosed population may be larger than currently estimated without significantly affecting the population dynamics.

Introduction

The human immunodeficiency virus (HIV) progresses in three stages. The first stage lasts approximately 3 months and individuals in this stage are approximately 10 to 25 times more effective at transmitting the disease [1–3]. The chronic stage can last from 5–10 years without medication [4]. This is followed by acquired immunodeficiency syndrome (AIDS) and death shortly thereafter [1–3]. Individuals with HIV may go many years without diagnosis, during which time they may expose uninfected individuals to HIV. Efforts to improve the diagnosis

Competing interests: The authors have declared that no competing interests exist.

rate include educational programs, as an individual's perceived risk was shown to be highly correlated with the individual obtaining multiple HIV tests [5–8]. Several studies have found a 50% reduction in risky behaviors after diagnosis, including safer sex practices, reduction in partner number, and medications that reduce viral load [9, 10]. However, diagnosis events resulting in behavior modification are not thought to be sufficient to eradicate the epidemic [9, 11].

After diagnosis, infected individuals have the opportunity to take anti-retroviral therapy (ART) that reduces their viral load and retards the progression of the disease, providing an improvement in life expectancy [12, 13]. The earlier that ART is received the higher the reduction in transmission events, particularly if obtained during the initial stage of HIV [1–3]. ART therapies could eradicate the epidemic in a population with high prevalence of infection even without the additional effect of behavioral changes [11]. Mathematical models estimate that the HIV epidemic could be reduced to less than 1% of the population infected (elimination phase) with universal testing and by providing ART consistently to newly diagnosed individuals [14]. However, issues with adherence and resistance are well documented in the literature [15–20]. Patients tend to report their adherence as much higher than it actually is, but studies indicate that even low adherence may be sufficient for control of the epidemic [10, 15]. Transmission is rare for individuals on ART, even with relatively high plasma HIV concentrations [21, 22].

The largest barrier to eradication of the epidemic is lack of access to care, including diagnostic services and ART costs or prescriptions. A lack in access to care could create pockets of undiagnosed individuals while the overall trend appears to be a reduction in the size of the epidemic [23]. Various studies report between 50–96% of diagnosed individuals in the U.S. rely on public medical care to obtain their ART medications [10, 20, 24–26]. Access to care remains critical, but this has not been considered when modeling the dynamics of the epidemic [5, 6].

Estimates using CD4 levels of newly diagnosed individuals suggest that the undiagnosed population is decreasing between 2005–2013 in the US [4, 27]. CD4 levels can be used to estimate the progression of HIV [28]. We consider three possible causes for this decrease including exhaustion of the at-risk population. The size of the at-risk population is not easy to estimate since it depends on behavior. High risk populations include individuals in poverty and men who have sex with men (MSM) [4]. This is particularly critical in the southern U.S., where individuals tend to be poor and lack access to medical care [4, 23]. As the at-risk population decreases the number of diagnoses will also decrease, which will cause the estimated number of undiagnosed individuals to decrease.

An additional possibility is that the reduction in number of diagnoses is due to individuals lacking access to care. HIV is over-represented in impoverished populations where access to diagnosis and treatment may be more difficult to obtain. In this case, the number of newly diagnosed individuals is not representative of the number of undiagnosed individuals, and the estimates will be inaccurate. Finally, the usage of anti-retroviral therapies reduces the viral load and transmission potential of infected individuals.

The difficulty in studying this system mathematically lies in parameter estimates. A minimal model of this system requires at least three parameters: transmission of the disease, diagnosis of the disease, and death due to the disease. Since knowledge about the undiagnosed population is restricted to those who have been diagnosed, estimates of these parameters are generally forced to assume that this population is representative of the whole.

In this work we use coupled statistical and mathematical methodology to study the relationships between the three hypothesized causes and their respective population dynamics. We use hierarchical Bayesian statistics to get estimates for the size of the infected populations and their proportional changes across the years. These estimates are used to calculate epidemiological

parameters for a system of stochastic differential equations. Currently we are not aware of any similar methods in the literature. Such a problem is challenging as the proportional change across the populations is a hyperparameter controlling the yearly proportions, which each have their own statistical model. This results in a large model that must be studied numerically.

The resulting simulations give insight into the implications of the estimated undiagnosed population on epidemiological parameters. Our model suggests that the undiagnosed population may be larger than current estimates while recovering population dynamics. The best recovery of the dynamics observed in the existing data occurs when the increase in the diagnosed population due to diagnosis is greater than the decrease in the undiagnosed population. We hope this study will help inform future efforts to improve the situation of infected individuals and prevent future outbreaks.

Materials and methods

Bayesian statistics

A Markov model where p_t centered at qp_{t-1} is used to estimate the proportional change in the infected populations over time, where p_t is the proportion in the current year and q is the proportional change. These random variables are estimated using Bayesian statistics.

The sampling model is $x_t \sim Bin(n_t, p_t)$, where n_t is population size in the current year. The random variable q is taken as a hyperparameter for p_t . The random variables to be estimated for each infected subpopulation are q and p_t , where $t = 2005, \dots, 2013$. We estimate the random variables of undiagnosed and diagnosed subpopulations independently.

Prior. The prior for the proportional change q is a gamma distribution.

$$\pi(q) \propto q^{\alpha-1} e^{-\beta q}$$

The parameters were chosen so that the prior distribution was centered at the arithmetic estimates of q obtained from the CDC [4]. The arithmetic estimates were obtained by calculating:

$$\frac{1}{n} \sum_{i=2}^n \frac{p_i}{p_{i-1}}$$

The arithmetic estimate for the undiagnosed q (q_u) is 0.979, and for the diagnosed q (q_d) 1.025. The priors used were $GAM(9.79, 10)$ and $GAM(10.25, 10)$ so they were centered at 1.

The prior for the random variable p_t , the undiagnosed proportion, is a beta distribution centered at the previous proportion times q . The parameters of the beta distribution are $\alpha = 0.1n_{t-1} \times qp_{t-1}$, and $\beta = 0.1n_{t-1} - \alpha$.

$$\pi(p_t) \propto p_t^{\alpha-1} (1 - p_t)^{\beta-1}$$

In the case where $t = 1$, the previous undiagnosed proportion is taken to be the expert opinion of 20%, and the diagnosed proportion to be $1 - p_0(\text{undiagnosed})$ [29]. The prior for the diagnosed population is formulated in the same way. Population sizes were considered in units of thousands.

Likelihood. The likelihood is a binomial likelihood, representing the chance of selecting an undiagnosed or diagnosed individual at random from the total infected population. For a given year t , the proportion of undiagnosed individuals depends on the total number of individuals:

$$\mathcal{L}(p_t | x_t, n_t) \propto p_t^{x_t} (1 - p_t)^{n_t - x_t}$$

where x_t is the total number of undiagnosed or diagnosed individuals, and n_t is the total number of infected individuals. The likelihood across all the years is the product of each year's likelihood.

$$\mathcal{L}(p_1, p_2, \dots, p_9 | x_1, x_2, \dots, x_9, n_1, n_2, \dots, n_9) \propto \prod_{t=1}^9 p_t^{x_t} (1 - p_t)^{n_t - x_t}$$

The likelihood for the diagnosed population was formulated in the same way.

Posterior. The joint posterior distribution is proportional to the priors multiplied by the likelihoods for all 9 years:

$$f(p_1, p_2, \dots, p_9, q) \propto \pi(q) \times \prod_{t=1}^9 \pi(p_t | q, p_{t-1}) \times \mathcal{L}(p_1 | x_{t=1}) \times \mathcal{L}(p_2 | x_{t=2}) \times \dots \times \mathcal{L}(p_9 | x_{t=9}) \\ \propto q^{z-1} e^{-\beta q} \times \prod_{t=1}^9 \left(p_t^{z-1} (1 - p_t)^{\beta-1} \times p_t^{x_t} (1 - p_t)^{n_t - x_t} \right)$$

The posterior full conditional of p_t for $t = 2005, \dots, 2012$ is:

$$f(p_t | q, p_{t-1}, p_{t+1}) \propto \prod_{t=1}^9 \left(\mathcal{L}(p_t | x_t) \times \pi(p_t | q, p_{t-1}) \times \pi(p_{t+1} | q, p_t) \right)$$

The posterior full conditional of 2013, the 9th year, is:

$$f(p_9 | x_9, p_8, q) \propto \mathcal{L}(p_9 | x_{t=9}) \times \pi(p_9 | q, p_8)$$

The full conditional of q does not have a closed form. The forms of the diagnosed random variables are the same. Random variable estimates were obtained using Metropolis-Hastings nested within a Gibbs sampler over 100,000 iterations with R version 3.3.3 [30]. The proposal distribution was a truncated normal distribution since the hyperparameter q cannot be negative, using package `rmutil` [31], centered at the previous value of the parameter. Proportions 1 through 9 were sampled consecutively, followed by hyperparameter q . The trace plots quickly converged within 200 iterations, and the first 2000 samples were removed. Code is provided in [S3](#) and [S4](#) Files.

Stochastic differential equations

The hyperparameter q was estimated to be 0.982 for the undiagnosed population and 1.036 for the diagnosed population. These were used as a boundary to solve for the epidemiological parameters in a simple stochastic differential model.

$$dU = (q_u - 1)Udt + d\omega_u dt \\ dD = (q_d - 1)Ddt + d\omega_d dt$$

where U is the undiagnosed and D is the diagnosed populations, and $d\omega_t \sim \text{Nor}(0, \sigma)$ is Brownian white noise with units $\sqrt{(t)}$. The variance σ is chosen to be 10% of the size of the population.

The simplest model is constructed describing the dynamics of the infected subpopulations. The values of parameters transmission (τ), diagnosis (δ), and death (ϵ) are calculated using the constraining q :

$$dU = (\tau(U + D) - \delta U - \epsilon U)dt \tag{1}$$

$$\cong (1 - q_u)Udt + d\omega_u dt \tag{2}$$

$$= -0.018Udt + d\omega_u dt \tag{3}$$

$$dD = (\delta U - \epsilon D)dt \tag{4}$$

$$\cong (1 - q_d)Ddt + d\omega_t dt \tag{5}$$

$$= 0.036Ddt + d\omega_t dt \tag{6}$$

We consider the parameters pseudo-steady state, and use the 2005 population sizes to estimate them. In addition, we assume that the general population are at steady state, and consider only the increased death rate due to infection ϵ as 0.01 [4]. The diagnosis rate δ is estimated by:

$$\delta U = q_d D + \epsilon D = \frac{0.046 D_{2005}}{U_{2005}} = 0.165$$

and the transmission rate τ is estimated by:

$$\tau(U + D) = (1 - q_u)U + \delta U + \epsilon U$$

$$\tau = \frac{0.157 U_{2005}}{U_{2005} + D_{2005}} = 0.034$$

Due to the magnitude of the scale of this system we assume that all events will happen, and the source of the stochasticity is primarily reporting issues. Tau leap algorithm was used to preform the stochastic simulations. A time step of 1 year was selected, and the population at time $t+1$ is the numerical solution of the population at time t and random noise from a $NOR(0, \sigma)$, where σ is 10% of the population at time t_0 with units \sqrt{t} . The initial conditions for the infected populations were sampled from the posterior distributions obtained by the Bayesian estimates. Calculations were performed in Matlab [32] and code is available upon request.

The diagnosis rate was calculated using data from [27]. The at-risk population is estimated as twice the national average rate of self-identified MSM among adults. The mortality rate increase due to HIV was estimated using data from [4]. All calculations, including the effective parameter rates, are provided in [S1 File](#).

Results

Bayesian model

Bayesian estimates for the proportions of diagnosed or undiagnosed individuals was obtained concurrently with the estimated proportional change. The prior distribution was chosen to be a beta for the proportions and a gamma for the proportional changes. The likelihood function was a binomial, representing the chance of randomly selecting a diagnosed or undiagnosed individual from a pool of infected individuals. The posterior did not have a closed form. Due to the symmetry of the posterior samples we summarize them using their mean and variance. The posterior means of the proportions for both undiagnosed and diagnosed estimates are very close to the original data ([Fig 1](#)) [27]. The posterior mean of q_u is 0.982, and q_d is 1.036. This means that 98% of the undiagnosed population is preserved from year to year, or is dropping by about 2% per year. Similarly, the diagnosed population is increasing by 3.6% per year. Posterior means and variances are given in [Table 1](#).

Stochastic differential model

The Bayesian estimates of the proportional change in the diagnosed and undiagnosed population from 2005 to 2013 were used to determine the epidemiological parameters for a system of

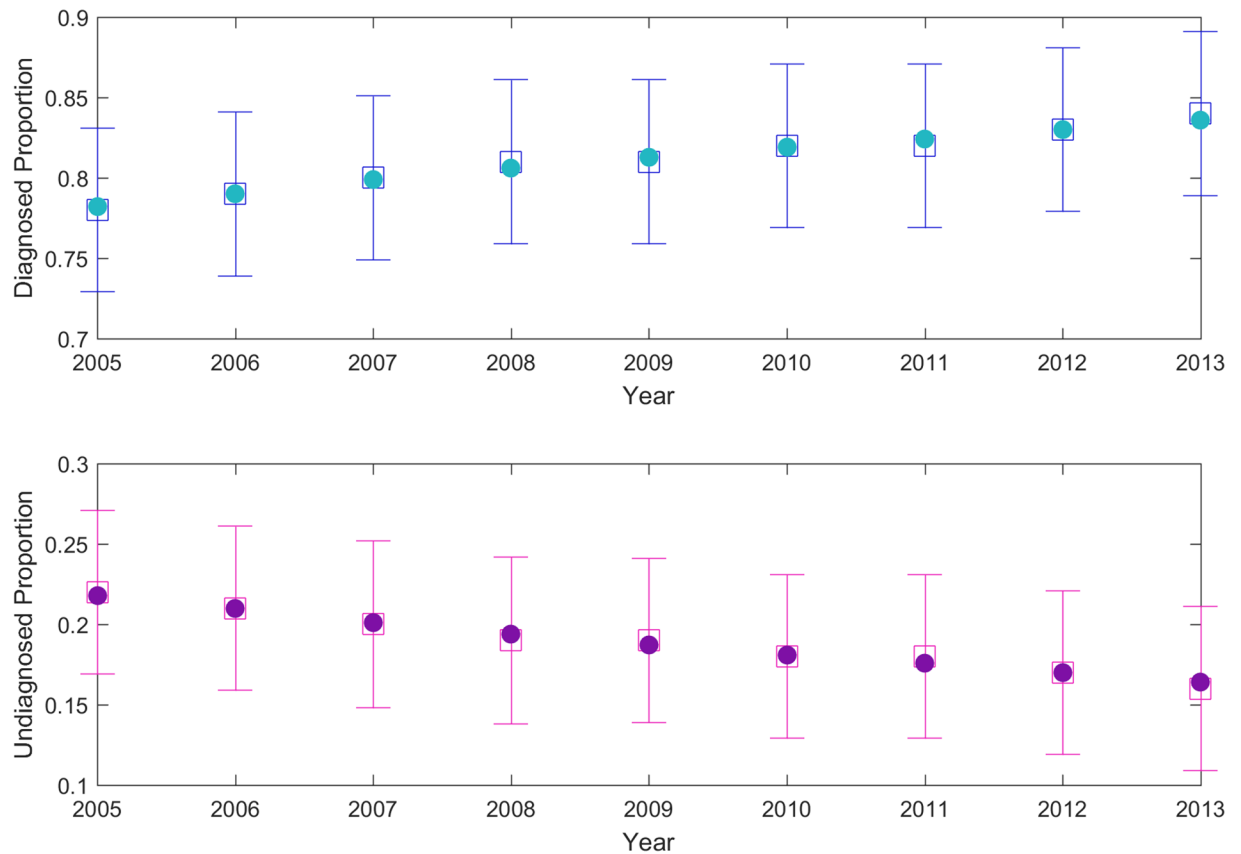


Fig 1. Posterior information obtained from hierarchical Bayesian statistics. Bayesian estimates are shown as hollow squares with error bars showing standard deviations. Estimated proportion of diagnosed (pink) and undiagnosed (blue) populations recover the estimated proportions (circles) [27].

<https://doi.org/10.1371/journal.pone.0200126.g001>

stochastic differential equations. The parameters transmission (τ), diagnosis (δ), and death (ϵ) were calculated using the proportional changes in the respective population.

$$dU = (q_u - 1)Udt + d\omega_t dt = (\tau(U + D) - \delta U - \epsilon U)dt \tag{7}$$

$$dD = (q_d - 1)Ddt + d\omega_t dt = (\delta U - \epsilon D)dt \tag{8}$$

Table 1. Summary statistics of the posterior distribution. The parameter p represents the estimated size of the proportion in that year. The hyperparameter q represents the estimated proportional change of the population across all years.

Parameters	Diagnosed		Undiagnosed		
	Mean	Variance	Parameters	Mean	Variance
p_{2005}	0.78	0.0026	p_{2005}	0.22	0.0026
p_{2006}	0.79	0.0026	p_{2006}	0.21	0.0026
p_{2007}	0.80	0.0026	p_{2007}	0.20	0.0027
p_{2008}	0.81	0.0026	p_{2008}	0.19	0.0027
p_{2009}	0.81	0.0026	p_{2009}	0.19	0.0026
p_{2010}	0.82	0.0026	p_{2010}	0.18	0.0026
p_{2011}	0.82	0.0026	p_{2011}	0.18	0.0026
p_{2012}	0.83	0.0026	p_{2012}	0.17	0.0026
p_{2013}	0.84	0.0026	p_{2013}	0.16	0.0026
q_d	1.036	0.2355	q_u	0.982	0.0939

<https://doi.org/10.1371/journal.pone.0200126.t001>

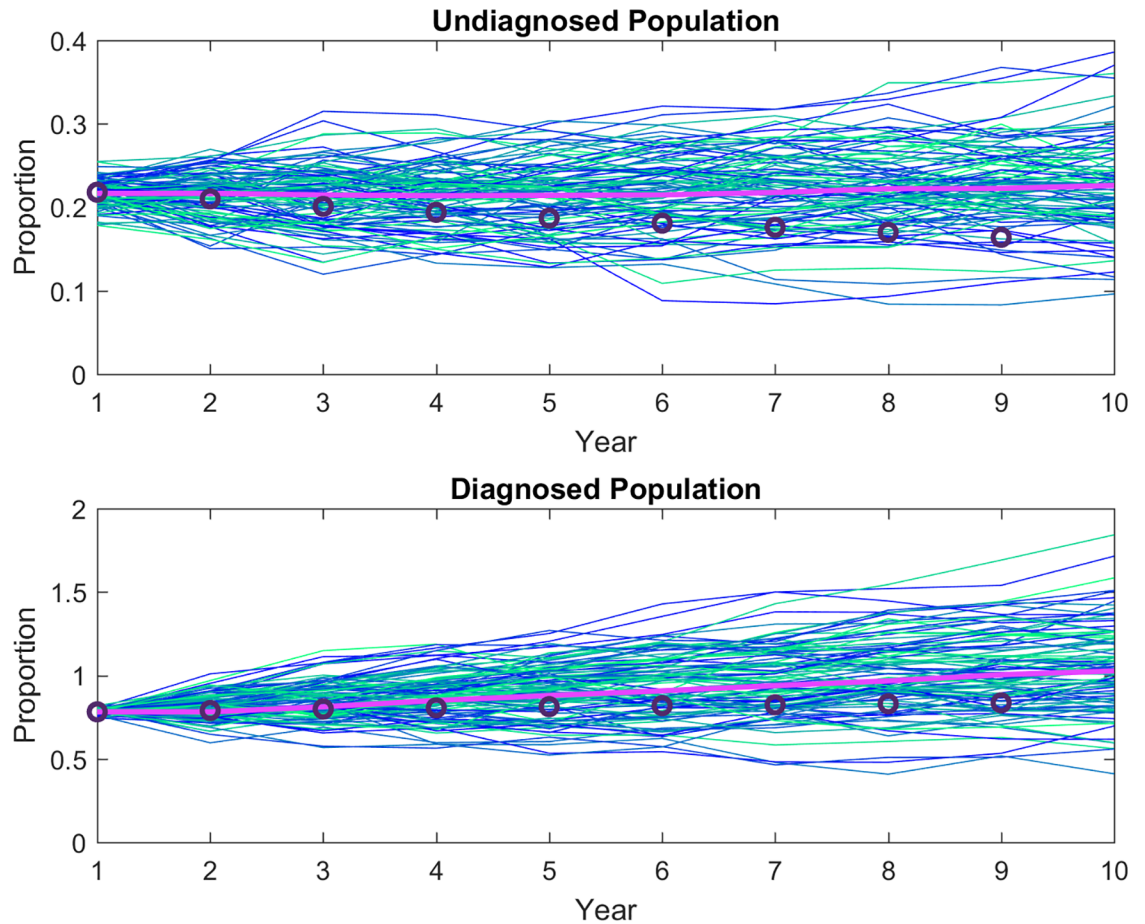


Fig 2. Method validation. The method was tested by simulating with the epidemiological parameters calculated using the Bayesian estimates of the proportional changes as constraints. The mean of 100 stochastic simulations (pink line) is compared with the data (circles). Proportions are relative to initial proportion.

<https://doi.org/10.1371/journal.pone.0200126.g002>

where U is the undiagnosed and D is the diagnosed populations, and $dw_t \sim Nor(0, \sigma)$ is the noise term. These base estimates fit the data very well (Fig 2).

Exhaustion of Susceptibles. In the case where the at-risk population is not much larger than the infected population, the transmission is dependent on the size of both populations. We estimate the at-risk population size as a fraction of the total infected population:

$$S = fT$$

Then this is substituted into the model. The transmission term becomes

$$\tau TS \cong \tau f T^2$$

This gives an effective increase of $f\tau$ in the transmission rate (Table 2). This increase causes the simulations to fail to recover the diagnosed and undiagnosed population dynamics, although the at-risk population does decrease significantly (Fig 3). This result is intuitive since the infection rate is increased, but the diagnosis rate is not representative of this rate.

Lack of access to care. Lack of access to care may be conceptualized as pockets of undiagnosed individuals who are not being diagnosed. To capture this, we consider the diagnosis rate

Table 2. Transmission and diagnosis rates are different under the different hypotheses. Average likelihood across both populations and all years (Fig 6, S2 File).

Model	Transmission Rate	Diagnosis Rate	Likelihood
Base model	$\tau(U + D)$	δU	0.87
Exhaustion of Susceptibles (ES)	$\tau f(U + D)^2$	δU	0.16
Anti-retroviral Therapies (ART)	$\tau(U + 0.04D)$	δU	0.72
Lack of Access to Care (LAC)	$\tau(U + D)$	δ_0	0.87
ES and ART	$\tau f(U + 0.04D)^2$	δU	0.52
ES and LAC	$\tau f(U + D)^2$	δ_0	0.54
ART and LAC	$\tau(U + 0.04D)$	δ_0	0.70
ES, LAC, and ART	$\tau f(U + 0.04D)^2$	δ_0	0.58

<https://doi.org/10.1371/journal.pone.0200126.t002>

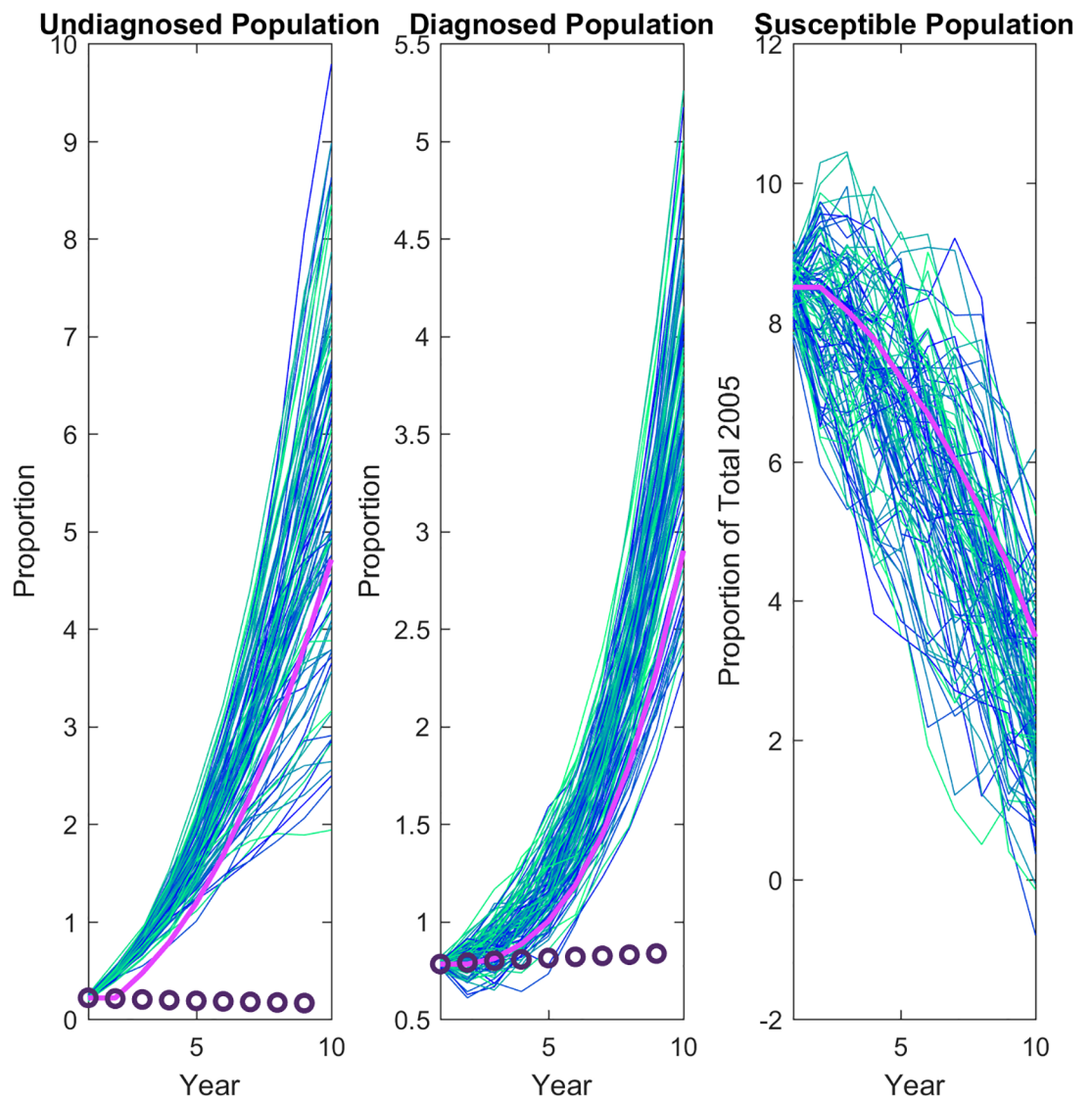


Fig 3. Exhaustion of susceptibles. Transmission of the disease is altered to reflect the impact of the size of the at-risk population. The mean of 100 stochastic simulations (pink line) is compared to the data (circles). Proportions are relative to initial proportion.

<https://doi.org/10.1371/journal.pone.0200126.g003>

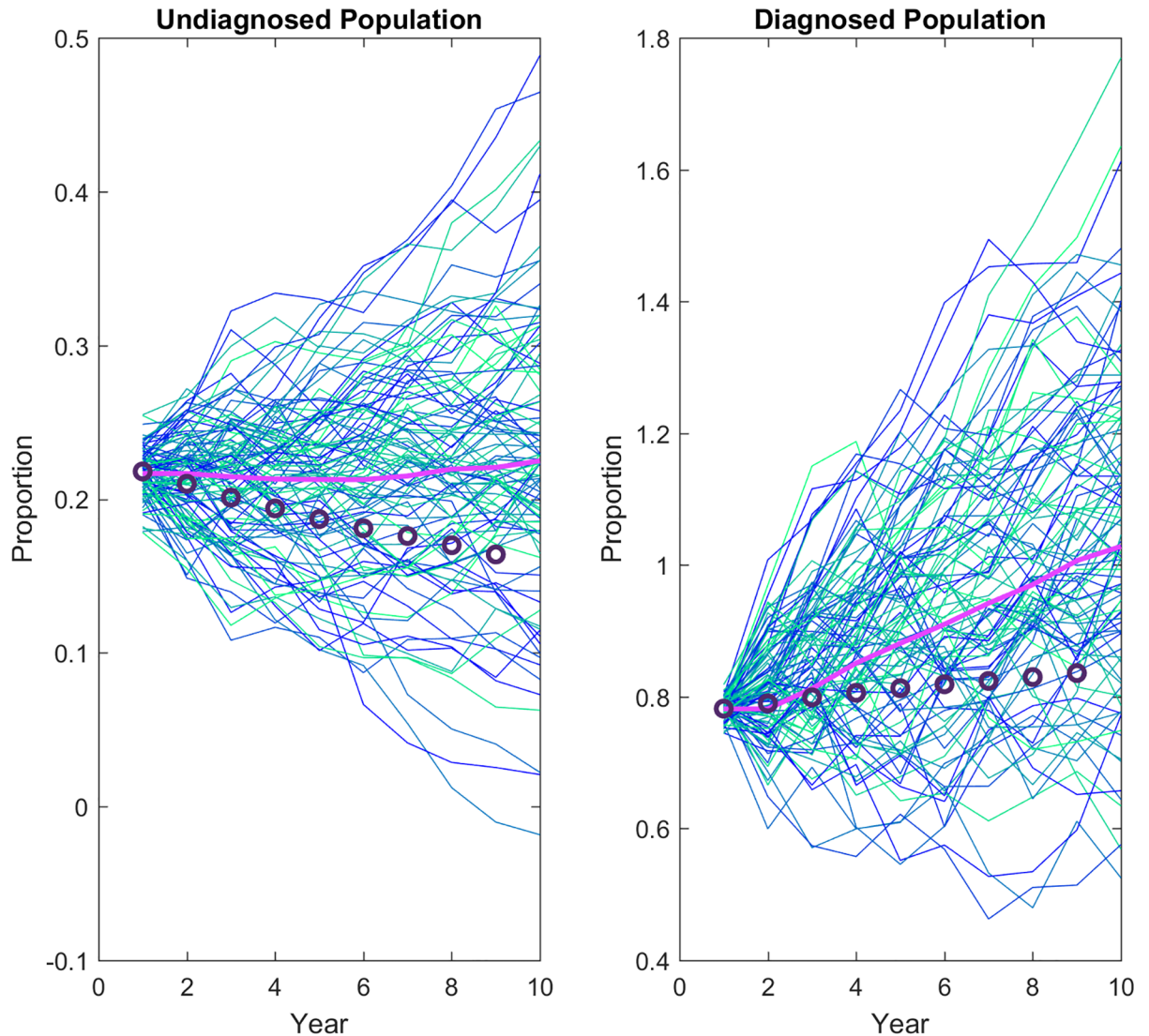


Fig 4. Lack of access to care. The effect of undiagnosed individuals lacking access to care affects the rate of diagnosis of the undiagnosed individuals. The diagnosis rate is held constant to reflect this scenario. The mean of 100 stochastic simulations (pink line) is compared with the data (circles). Proportions are relative to initial proportion.

<https://doi.org/10.1371/journal.pone.0200126.g004>

to be independent of the size of the undiagnosed population. The diagnosis rate is estimated as:

$$\delta U = q_a D + \epsilon D = \frac{0.046 D_{2005}}{U_{2005}} \cdot U_{2005} = 0.036$$

The resulting equation for the undiagnosed subpopulation then becomes:

$$dU = (\tau(U + D) - \delta_0 - \epsilon U)dt$$

where δ_0 is 0.036. This large reduction in the diagnosis rate recovers the population dynamics well (Fig 4).

ART usage. Since ART results in a viral load that has low chance of infecting a at-risk individual, we removed these individuals from the pool of infected individuals able to transmit

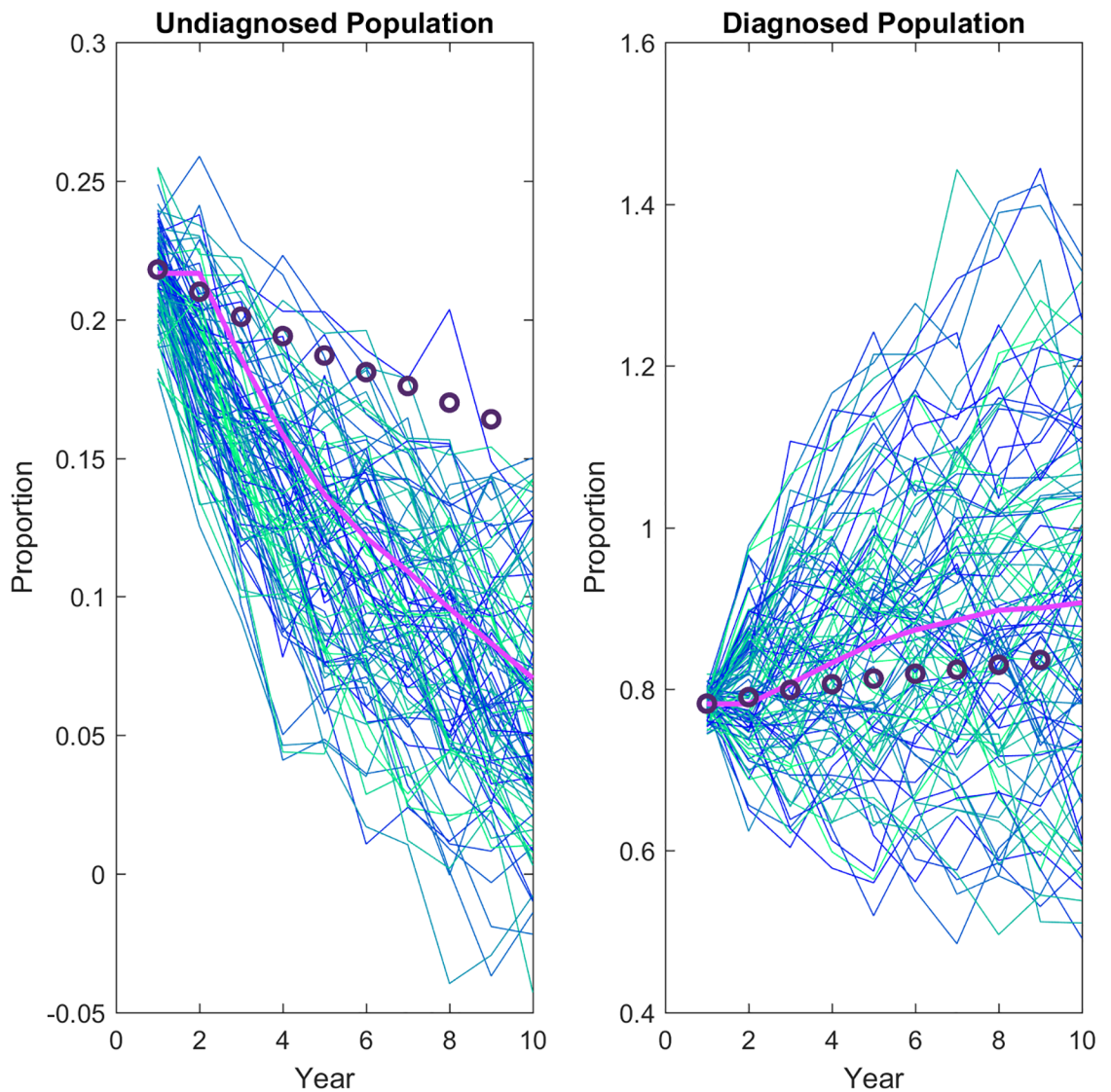


Fig 5. Anti-retroviral therapy usage. To reflect the high levels of ART prescription and usage reported by diagnosed individuals this percentage is removed from the pool of diagnosed individuals able to transmit the disease. The mean of 100 stochastic simulations (pink line) is compared to the data (circles). Proportions are relative to initial proportion.

<https://doi.org/10.1371/journal.pone.0200126.g005>

the disease. Since 96% of diagnosed individuals reported taking anti-retroviral therapies in a previous study, the transmission term was modified as follows [10].

$$\tau(U + (1 - 0.96)D)$$

Variable or poor adherence on the part of diagnosed individuals is ignored due to the body of literature indicating that large benefit is gained from even poor adherence [15, 16]. This gives good recovery of both subpopulation dynamics and agrees best with both undiagnosed and diagnosed estimates (Fig 5).

Multiple causes

Since it seems likely that most or all of these scenarios affect the infected population simultaneously, we analyze all their possible combinations (S1 Fig). The parameters were altered as

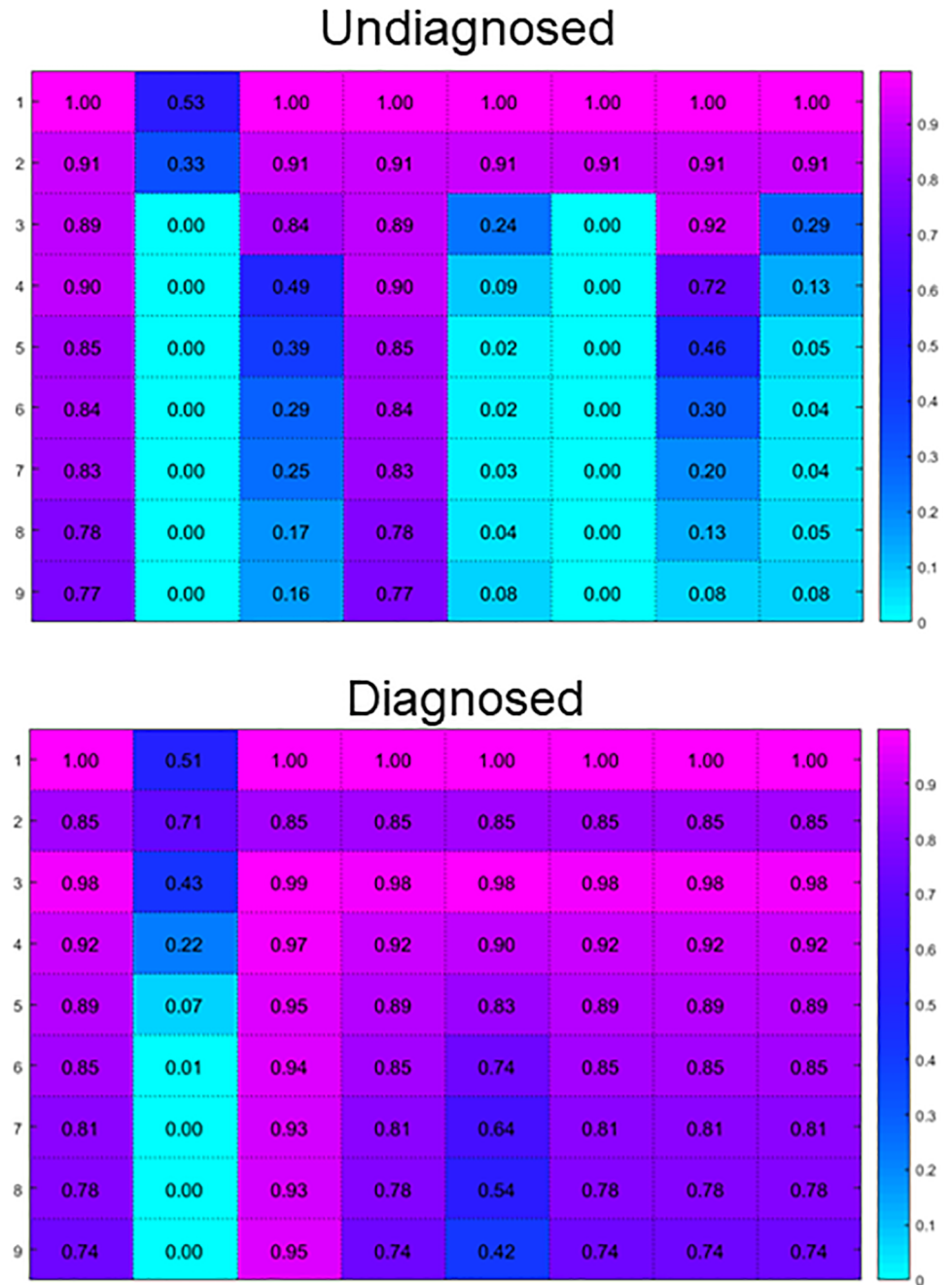


Fig 6. Model fit was quantified by calculating the relative likelihood of observing the data within the simulations. A higher likelihood is represented by a hotter color. From left to right: Base model, Exhaustion of Susceptibles (ES), Lack of Access to Care (LAC), Anti-Retroviral Therapy usage (ART), ES and LAC, ES and ART, LAC and ART, and ES, LAC, and ART. Details provided in [S2 File](#).

<https://doi.org/10.1371/journal.pone.0200126.g006>

described in [Table 2](#). To determine the best cause, we quantify the goodness of fit by determining the relative likelihood of observing the data given the mean and variance of the stochastic simulations. These probabilities are given in [Fig 6](#) with numerical details in [S2 File](#), as well as the average probability over the 9 years.

Lack of access to care, ART usage, or their combined resulted in the best recovery of the data for both the undiagnosed and diagnosed populations. Under the ART scenario, the diagnosed population has been reduced by 96%, resulting in a dynamic reduction in the transmission rate. We originally estimate the transmission rate to be 3.4% of the total infected population. This is close to the literature estimate of around 4–6% of infected individuals transmitting HIV to an at-risk individual [29, 33]. With the majority of the diagnosed population removed, the effective transmission rate is much lower. Under the LAC scenario, there is a constant diagnosis rate. This represents a yearly reduction in the undiagnosed population and increase in the diagnosed population of 3.6%. This means the diagnosed population grows faster than the undiagnosed population is reduced. Although lack of access to care in the undiagnosed population would mean the data are inaccurate, in our case the best fitting model is consistent for both subpopulations.

Discussion

We were able to obtain conservative estimates of the proportional changes in the diagnosed and undiagnosed HIV-infected populations using hierarchical Bayesian statistics. Our estimates suggest that the proportion of infected individuals who are undiagnosed is decreasing by approximately 2.8% each year from 2005 to 2013, while the proportion of diagnosed individuals is increasing by approximately 3.6%. We used the proportional change as constraints on a system of stochastic differential equations. This allowed us to estimate the transmission and diagnosis rates. We were able to recover reasonable parameter estimates and population dynamics using this methodology. To learn more about the cause of the decrease in the undiagnosed population, we considered some scenarios that would affect the epidemiological parameters: exhaustion of the at-risk population, lack of access to care, and reduction in viral load by anti-retroviral therapy.

We were able to recover the diagnosed population dynamics when we altered the parameters to reflect these scenarios with the exception of including exhaustion of susceptibles. Modeling transmission as a function of the size of the at-risk population caused the size of the infected populations to increase rapidly. In the other scenarios some interesting dynamics could be observed in the undiagnosed population. Lack of access to care was simulated by considering diagnosis rate a constant unaffected by the size of the undiagnosed population. This resulted in an improvement in the likelihood of observing the data (Fig 6, S2 File). Anti-retroviral therapy usage also improved the overall recovery, but this effect was weaker for the undiagnosed population dynamics. Although the undiagnosed population size is dependent on the quality of the data available on the diagnosed population of that year, these results indicate that the scenarios that maximizes the probability of observing the diagnosed population also maximizes the probability of observing the diagnosed population estimates.

The observed results suggest that lack of access to care and ART usage contribute to the infected population dynamics. This is not unexpected. Many individuals with HIV are reported to lack access to care [23, 34]. In areas with high poverty rates the death rate of infected individuals is much higher than that of the general population [4, 35]. In 2017 the New York Times reported groups of untreated individuals in the deep south dying due to their lack of access to care [23]. The southern US in particular suffers from high rates of STDs and lack of access to care that exacerbates the spread of HIV [36, 37]. In 2007 the number of uninsured individuals had risen by 60% since 2003, with those living in poverty over-represented [38]. Although the Affordable Care Act increased the number of individuals eligible for Medicaid, physician retention issues that continue to plague the majority of rural areas are exacerbated with increasing numbers of Medicaid patients [39, 40].

Both models and studies have shown that providing ART to infected individuals in the early stages of HIV reduces transmission events and frequency of death due to AIDS [11, 14, 19–22]. Even poor adherence may be enough to control or eradicate the epidemic and increase quality of life for infected individuals [16–18]. Greater effort must be made to ensure these populations have access to life-saving treatments.

Supporting information

S1 File. Calculations. Calculations for stochastic modeling using data obtained from [4, 27]. (XLSX)

S2 File. Model fit. Details of the model fit probabilities for each year and simulation. (PDF)

S3 File. Matlab code. Code used to run simulations and generate heatmap. (M)

S4 File. R Code—Diagnosed. Code used to run Bayesian simulations. (R)

S5 File. R Code—Undiagnosed. Code used to run Bayesian simulations. (R)

S1 Fig. Combined cases. Stochastic simulations for the combined cases. (A) Exhaustion of Susceptibles and lack of access to care. (B) Exhaustion of Susceptibles and ART. (C) Lack of access to care and ART. (D) Exhaustion of Susceptibles, lack of access to care, and ART. (PDF)

Author Contributions

Conceptualization: Renee Dale, BeiBei Guo.

Data curation: Renee Dale.

Formal analysis: Renee Dale.

Methodology: Renee Dale, BeiBei Guo.

Supervision: BeiBei Guo.

Visualization: Renee Dale.

Writing – original draft: Renee Dale.

Writing – review & editing: BeiBei Guo.

References

1. Rasmussen D.A., Volz E.M., Koelle K., 2014. Phylodynamic inference for structured epidemiological models. *PLoS computational biology* 10, e1003570. <https://doi.org/10.1371/journal.pcbi.1003570> PMID: 24743590
2. Cohen M.S., Shaw G.M., McMichael A.J., Haynes B.F., 2011. Acute HIV-1 Infection. *New England Journal of Medicine* 364, 1943–1954. <https://doi.org/10.1056/NEJMr1011874> PMID: 21591946
3. Volz E.M., Ionides E., Romero-Severson E.O., Brandt M.-G., Mokotoff E., Koopman J.S., 2013. HIV-1 Transmission during Early Infection in Men Who Have Sex with Men: A Phylodynamic Analysis. *PLoS Medicine* 10, e1001568. <https://doi.org/10.1371/journal.pmed.1001568> PMID: 24339751
4. Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2014. HIV

- Surveillance Supplemental Report 2016;21(No. 4). <http://www.cdc.gov/hiv/library/reports/surveillance/>. Published July 2016. Accessed 3/14/2017.
6. Davey D.J., Beymer M., Roberts C.P., Bolan R.K., Klausner J.D., 2017. Differences in Risk Behavior and Demographic Factors Between Men Who Have Sex With Men With Acute and Nonacute Human Immunodeficiency Virus Infection in a Community-Based Testing Program in Los Angeles. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 74, e97–e103. <https://doi.org/10.1097/QAI.0000000000001233>
 5. Stein J.A., Nyamathi A., 2000. Gender differences in behavioral and psychosocial predictors of HIV testing and return for test results in a high-risk population. *AIDS Care* 12, 345–356. <https://doi.org/10.1080/09540120050043007>
 7. Cabrera S., Hernández, Sepúlveda, Valverde, Domínguez-Gil Hurlé A., Iglesias, 2013. Impact of a pharmaceutical care program on clinical evolution and antiretroviral treatment adherence: a 5-year study. *Patient Preference and Adherence* 729. <https://doi.org/10.2147/PPA.S47519>
 8. Amico K.R., Barta W., Konkle-Parker D.J., Fisher J.D., Cornman D.H., Shuper P.A., et al. 2009. The Information—Motivation—Behavioral Skills Model of ART Adherence in a Deep South HIV+ Clinic Sample. *AIDS and Behavior* 13, 66–75. <https://doi.org/10.1007/s10461-007-9311-y> PMID: 17876697
 9. Marks G., Crepaz N., Senterfitt J.W., Janssen R.S., 2005. Meta-analysis of high-risk sexual behavior in persons aware and unaware they are infected with HIV in the United States: implications for HIV prevention programs. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 39, 446–453. <https://doi.org/10.1097/01.qai.0000151079.33935.79> PMID: 16010168
 10. Centers for Disease Control and Prevention. Behavioral and Clinical Characteristics of Persons Receiving Medical Care for HIV Infection—Medical Monitoring Project, United States, 2014 Cycle (June 2014–May 2015). HIV Surveillance Special Report 17.
 11. Velasco-Hernandez J.X., Gershengorn H.B., Blower S.M., 2002. Could widespread use of combination antiretroviral therapy eradicate HIV epidemics? *The Lancet infectious diseases* 2, 487–493. [https://doi.org/10.1016/S1473-3099\(02\)00346-8](https://doi.org/10.1016/S1473-3099(02)00346-8) PMID: 12150848
 12. Reynolds S.J., Makumbi F., Nakigozi G., Kagaayi J., Gray R.H., Wawer M., et al. 2011. HIV-1 transmission among HIV-1 discordant couples before and after the introduction of antiretroviral therapy: *AIDS* 25, 473–477. <https://doi.org/10.1097/QAD.0b013e3283437c2b> PMID: 21160416
 13. Nakagawa F., Lodwick R.K., Smith C.J., Smith R., Cambiano V., Lundgren J.D., et al. 2012. Projected life expectancy of people with HIV according to timing of diagnosis: *AIDS* 26, 335–343. <https://doi.org/10.1097/QAD.0b013e32834dcec9> PMID: 22089374
 14. Granich R.M., Gilks C.F., Dye C., De Cock K.M., Williams B.G., 2009. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. *The Lancet* 373, 48–57.
 15. Bartlett J. A. (2002). Addressing the challenges of adherence. *JAIDS*, 29(S1), S2–S10. PMID: 11832696
 16. Bangsberg D.R., Deeks S.G., 2002. Is average adherence to HIV antiretroviral therapy enough? *Journal of general internal medicine* 17, 812–813. <https://doi.org/10.1046/j.1525-1497.2002.20812.x> PMID: 12390559
 17. Bangsberg D. R., Acosta E. P., Gupta R., Guzman D., Riley E. D., Harrogon P. R., et al. 2006. Adherence resistance relationships for protease and non-nucleoside reverse transcriptase inhibitors explained by virological fitness. *AIDS*, 20(2), 223–231. <https://doi.org/10.1097/01.aids.0000199825.34241.49> PMID: 16511415
 18. Paterson D. L., Swindells S., Mohr J., Brester M., Vergis E. N., Aquier C., et al. 2000. Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. *Annals of Internal Medicine*, 133, 21–30 PMID: 10877736
 19. Schneider M.F., Gange S.J., Williams C.M., Anastos K., Greenblatt R.M., Kingsley L., et al. 2005. Patterns of the hazard of death after AIDS through the evolution of antiretroviral therapy: 1984–2004. *Aids* 19, 2009–2018. <https://doi.org/10.1097/01.aids.0000189864.90053.22> PMID: 16260908
 20. Sayles J.N., Wong M.D., Kinsler J.J., Martins D., Cunningham W.E., 2009. The Association of Stigma with Self-Reported Access to Medical Care and Antiretroviral Therapy Adherence in Persons Living with HIV/AIDS. *Journal of General Internal Medicine* 24, 1101–1108. <https://doi.org/10.1007/s11606-009-1068-8> PMID: 19653047
 21. Cohen M.S., Chen Y.Q., McCauley M., Gamble T., Hosseinipour M.C., Kumarasamy N., et al. 2011. Prevention of HIV-1 infection with early antiretroviral therapy. *New England journal of medicine* 365, 493–505. <https://doi.org/10.1056/NEJMoa1105243> PMID: 21767103
 22. Montaner J.S., Hogg R., Wood E., Kerr T., others, 2006. The case for expanding access to highly active antiretroviral therapy to curb the growth of the HIV epidemic. *The Lancet* 368, 531. [https://doi.org/10.1016/S0140-6736\(06\)69162-9](https://doi.org/10.1016/S0140-6736(06)69162-9)

23. America's Hidden H.I.V. Epidemic [Internet]. [cited 2017 Jun 23]. Available from: <https://www.nytimes.com/2017/06/06/magazine/americas-hidden-hiv-epidemic.html?nytmobile=0>
24. Kinsler J.J., Wong M.D., Sayles J.N., Davis C., Cunningham W.E., 2007. The Effect of Perceived Stigma from a Health Care Provider on Access to Care Among a Low-Income HIV-Positive Population. *AIDS Patient Care and STDs* 21, 584–592. <https://doi.org/10.1089/apc.2006.0202> PMID: 17711383
25. Lanoy E., Mary-Krause M., Tattevin P., Perbost I., Poizot-Martin I., Dupont C., et al. 2007. Frequency, determinants and consequences of delayed access to care for HIV infection in France. *Antiviral therapy* 12, 89. PMID: 17503752
26. Simoni J.M., Huh D., Wilson I.B., Shen J., Goggin K., Reynolds N.R., et al. 2012. Racial/Ethnic Disparities in ART Adherence in the United States: Findings From the MACH14 Study. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 60, 466–472. <https://doi.org/10.1097/QAI.0b013e31825db0bd> PMID: 22595873
27. Song R., Hall H.I., Green T.A., Szwarcwald C.L., Pantazis N., 2017. Using CD4 Data to Estimate HIV Incidence, Prevalence, and Percent of Undiagnosed Infections in the United States. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 74, 3–9. <https://doi.org/10.1097/QAI.0000000000001151> PMID: 27509244
28. Vlahov D., Graham N., Hoover D., Flynn C., Bartlett J.G., Margolick J.B., et al. 1998. Prognostic indicators for AIDS and infectious disease death in HIV-infected injection drug users: plasma viral load and CD4+ cell count. *JAMA* 279, 35–40. <https://doi.org/10.1001/jama.279.1.35> PMID: 9424041
29. Pinkerton S.D., 2012. HIV Transmission Rate Modeling: A Primer, Review, and Extension. *AIDS and Behavior* 16, 791–796. <https://doi.org/10.1007/s10461-011-0042-8> PMID: 21928097
30. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
31. Bruce Swihart and Jim Lindsey (2017). rmutl: Utilities for Nonlinear Regression and Repeated Measurements Models. R package version 1.1.0. <https://CRAN.R-project.org/package=rmutl>
32. MATLAB R2016b, The MathWorks, Inc., Natick, Massachusetts, United States.
33. Chen Y., Dale R., He H., Le Q.-A.T., 2017. Posterior Estimates of Dynamic Constants in HIV Transmission Modeling. *Computational and Mathematical Methods in Medicine* 2017, 1–8. <https://doi.org/10.1155/2017/1093045>
34. Valdiserri R.O., Forsyth A.D., Yakovchenko V., Koh H.K., 2013. Measuring what matters: development of standard HIV core indicators across the US Department of Health and Human Services. SAGE Publications Sage CA: Los Angeles, CA.
35. Michigan Department of Health and Human Services. Annual HIV Surveillance Report, City of Detroit. July 2015. http://www.michigan.gov/documents/mdch/Detroit_496828_7.pdf
36. Reif S., Geonnotti K.L., Whetten K., 2006. HIV Infection and AIDS in the Deep South. *American Journal of Public Health* 96, 970–973. <https://doi.org/10.2105/AJPH.2005.063149> PMID: 16670228
37. Sherman, S.G., and Carolina, N. 2008. Southern States Manifesto: update 2008. HIV/AIDS and Sexually Transmitted Diseases in the South. Available at: <https://southernaidscoalition.org/wp-content/uploads/2017/03/Southern-States-Manifesto-20081.pdf>
38. Schoen C., Collins S.R., Kriss J.L., and Doty M.M. 2008. How Many Are Underinsured? Trends Among U.S. Adults, 2003 And 2007. *Health Affairs* 27, w298–w309. <https://doi.org/10.1377/hlthaff.27.4.w298> PMID: 18544591
39. Douthit N., Kiv S., Dwolatzky T., and Biswas S. 2015. Exposing some important barriers to health care access in the rural USA. *Public Health* 129, 611–620. <https://doi.org/10.1016/j.puhe.2015.04.001> PMID: 26025176
40. MacDowell M., Glasser M., Fitts M., Nielsen K., and Hunsaker M. 2013. A national view of rural health workforce issues in the USA. 16.