Evolutionary Dynamics of Complex Networks of HIV Drug-Resistant Strains: The Case of San Francisco

Robert J. Smith?,[1]*† Justin T. Okano,[1]† James S. Kahn,[2] Erin N. Bodine,[1]‡ Sally Blower[1]§

Over the past two decades, HIV resistance to antiretroviral drugs (ARVs) has risen to high levels in the wealthier countries of the world, which are able to afford widespread treatment. We have gained insights into the evolution and transmission dynamics of ARV resistance by designing a biologically complex multistrain network model. With this model, we traced the evolutionary history of ARV resistance in San Francisco and predict its future dynamics. By using classification and regression trees, we identified the key immunologic, virologic, and treatment factors that increase ARV resistance. Our modeling shows that 60% of the currently circulating ARV-resistant strains in San Francisco are capable of causing self-sustaining epidemics, because each individual infected with one of these strains can cause, on average, more than one new resistant infection. It is possible that a new wave of ARV-resistant strains that pose a substantial threat to global public health is emerging.

HIV resistance to antiretroviral drugs (ARVs) is causing serious clinical and public health problems throughout the United States and Europe. HIV strains began to acquire resistance in 1987 when ARVs were introduced as therapies for HIV-infected individuals (1). Since then, a multitude of drug-resistant strains have evolved that differ considerably in their susceptibility to three major classes of ARVs: nucleoside reverse-transcriptase inhibitors (NRTIs), non-nucleoside reverse-transcriptase inhibitors (NNRTIs), and protease inhibitors (PIs). These drug-resistant strains are now being transmitted to individuals who have never received ARVs; that is, transmitted drug resistance (TDR) has arisen. TDR is reported to range between 8 and 22% in many HIV-infected communities in resource-rich countries, and if it continues to increase, the effectiveness of therapeutic regimens, as well as efforts to control the HIV pandemic, will be compromised. We have developed a theoretical model (the amplification cascade model) to help understand and predict the evolutionary dynamics of complex transmission networks composed of multiple ARV-resistant strains. We calibrated and parameterized the model to represent the HIV epidemic in San Francisco in the community of men who have sex with men (MSM), where TDR is already high (~13%) (2). The model was able to reproduce the observed dynamics and evolution of transmitted resistance in this city over the past 20 years. We used the model first to predict the future evolutionary dynamics of TDR. Next, we determined whether any of the currently circulating ARV-resistant strains are capable of generating self-sustaining epidemics. Third, we identified the key drivers that generate high levels of TDR. We also discuss here the implications of our results for resource-constrained countries where ARV treatment programs are being rolled out.

All of the published HIV transmission models of ARV resistance are based on simple biological assumptions and can track only one resistant strain (3–8). Our amplification cascade model captures biological complexity by generating a dynamic network composed of multiple ARV-resistant strains. We modeled the multistrain network in San Francisco by classifying ARV-resistant strains into seven categories; each category was defined based on the specific class of drugs to which the strain was resistant (NRTIs, NNRTIs, or PIs) and the level of resistance (single-, dual-, or triple-class) (Fig. 1A and fig. S1). Single-class resistance was to NRTIs, NNRTIs, or PIs. Dual-class resistance was to NRTIs and NNRTIs, NRTIs and PIs, or NRTIs and PIs. Triple-class resistance was to all three. Each class of ARVs contains several drugs (table S1) (9). In our modeling framework, if a strain is classified as resistant to a certain class of ARVs, then the strain is resistant to at least one drug in that class.

We modeled treatment effects by specifying treatment regimens and then assessing the effects of these regimens on infectivity and the probability of developing resistance. In the model, treated individuals receive a regimen to which their virus is sensitive; hence, we assume that treated individuals achieve either complete or partial viral suppression. We consider patients who achieve complete viral suppression to be noninfectious and incapable of developing resistance. Patients who achieve only partial viral suppression retain some degree of infectivity and are capable of developing resistant strains. When individuals experience treatment failure (which is usually determined by viral rebound), they can be switched to new drugs either in the same class or in a new class. For example, if a patient (in the model) is on a regimen containing zidovudine (NRTI), lamivudine (NRTI), and nelfinavir (PI) and develops resistance to nelfinavir, he could be switched to another PI (for example, indinavir). The model includes a matrix that specifies the rates at which strains develop resistance; therefore, strains are directly linked through the acquisition and amplification of resistance.

In the model, resistant and wild-type strains are assumed to compete to transmit HIV to uninfected at-risk MSM. These competitive interactions are mediated through strain-specific infectivity: The greater the infectivity, the higher the probability that the strain will be transmitted. We ascribe a competitive advantage to wild-type strains by assuming that they are always more infectious than the resistant strains. Furthermore, based on available competitive-fitness assays, replication-capacity assays, and patterns of developed resistance, we assume that the NNRTI-resistant strains are more transmissible than the NRTI-resistant strains, which, in turn, are more transmissible than the PI-resistant strains (10, 11). In addition, we assume, based on the available data, that the transmissibility of virus strains decreases as the number of classes of resistance increases (12). Once an individual becomes infected with a wild-type or resistant strain, the model tracks viral dynamics, and consequently infectivity, through four stages of disease progression: (i) primary infection; (ii) not yet eligible for ARVs (that is, CD4 count > 350 cells/μl); (iii) eligible for ARVs (CD4 ≤ 350 cells/μl) but not currently undergoing ARV treatment; and (iv) ARV treatment. The 33 equations that specify the model, as well as a more detailed description of the structure, are given in (9). Parameter estimates are discussed in section 2 of (9), tables S2 to S11, and fig. S2. The model can be extended to include any number of additional drug classes, such as integrase inhibitors, co-receptor blockers, and fusion inhibitors, as they are introduced into new therapeutic regimens.

Before making predictions, we used the model (coupled with an uncertainty analysis) to reconstruct the evolution and transmission dynamics of the network of ARV-resistant strains (13). We calibrated the model, using Monte Carlo filtering techniques, to match the epidemiological conditions in San Francisco in 1987 when ARVs were first introduced [section 3 of (9)] (table S12). By the late 1980s, almost half of the MSM community was infected with HIV (14, 15). After calibration, we used the model to simulate the evolutionary dynamics, from 1987 to 2008, of a network of ~4000 resistant strains, where each strain differed in drug susceptibility and infectivity. The history of ARV therapy in San Francisco can be divided into four eras spanning two decades

[1]Center for Biomedical Modeling, Semel Institute of Neuroscience & Human Behavior, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90024, USA. [2]Department of Medicine AIDS Division, University of California San Francisco, San Francisco, CA 94110, USA.

*Present address: Department of Mathematics and Faculty of Medicine, University of Ottawa, Ottawa, K1N 6N5, Ontario, Canada.
†These authors contributed equally to this work.
‡Present address: Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA.
§To whom correspondence should be addressed. E-mail: sblower@mednet.ucla.edu
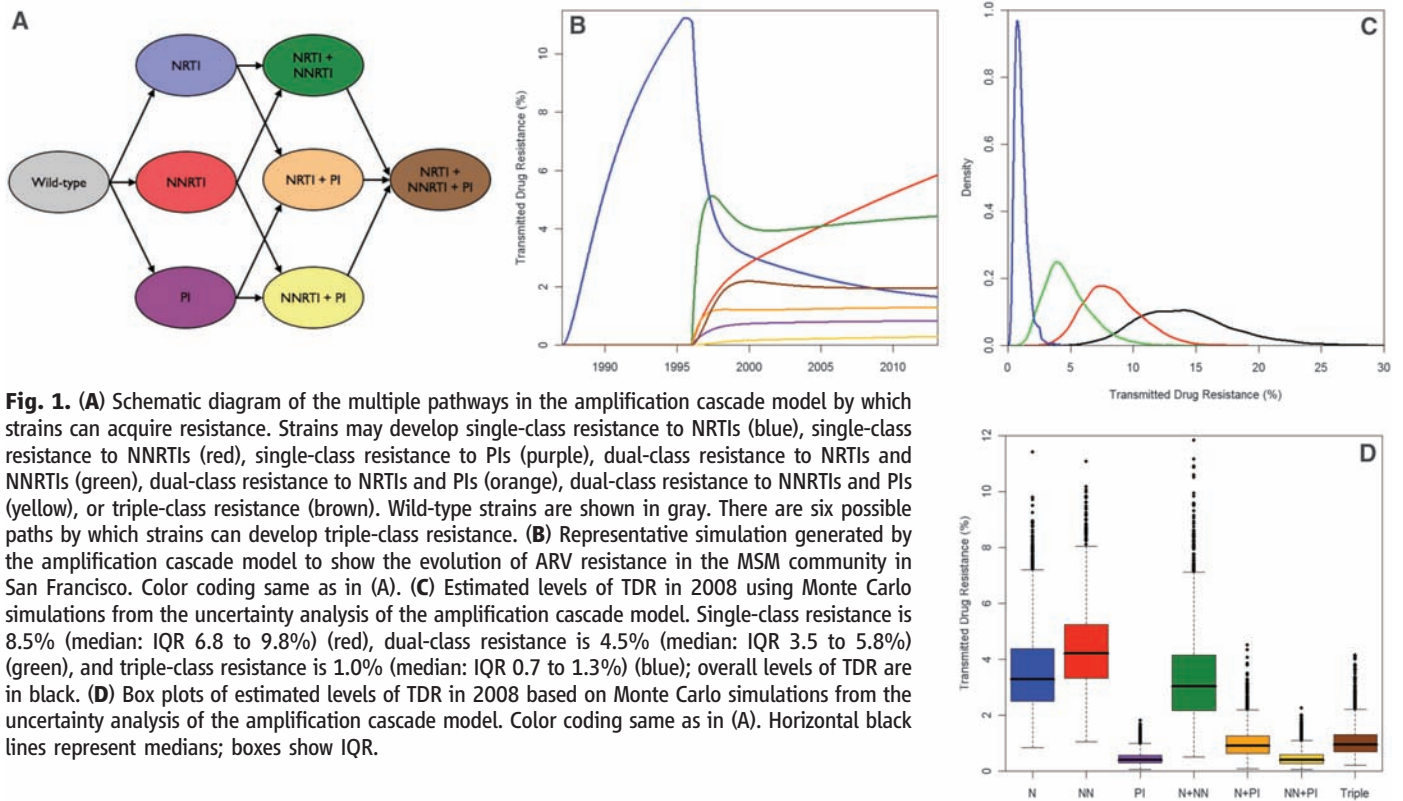
Fig. 1. (**A**) Schematic diagram of the multiple pathways in the amplification cascade model by which strains can acquire resistance. Strains may develop single-class resistance to NRTIs (blue), single-class resistance to NNRTIs (red), single-class resistance to PIs (purple), dual-class resistance to NRTIs and NNRTIs (green), dual-class resistance to NRTIs and PIs (orange), dual-class resistance to NNRTIs and PIs (yellow), or triple-class resistance (brown). Wild-type strains are shown in gray. There are six possible paths by which strains can develop triple-class resistance. (**B**) Representative simulation generated by the amplification cascade model to show the evolution of ARV resistance in the MSM community in San Francisco. Color coding same as in (A). (**C**) Estimated levels of TDR in 2008 using Monte Carlo simulations from the uncertainty analysis of the amplification cascade model. Single-class resistance is 8.5% (median: IQR 6.8 to 9.8%) (red), dual-class resistance is 4.5% (median: IQR 3.5 to 5.8%) (green), and triple-class resistance is 1.0% (median: IQR 0.7 to 1.3%) (blue); overall levels of TDR are in black. (**D**) Box plots of estimated levels of TDR in 2008 based on Monte Carlo simulations from the uncertainty analysis of the amplification cascade model. Color coding same as in (A). Horizontal black lines represent medians; boxes show IQR.

(9, 16) (fig. S3 and table S1). Different regimens were used in each era. We modeled the specific regimens that were available in each era by using data on the proportion of patients achieving viral suppression (tables S6 to S9), degree of reduction in viral load in partially virally suppressed patients (tables S2 and S5), rate of development of resistance in treated patients (tables S6 to S9), and treatment-induced increase in survival time (table S10) (9). Because usage of ARVs has increased over the past two decades, we modeled era-specific treatment rates (table S4) (9).

The model reproduced and explained the observed evolutionary dynamics of the network of ARV-resistant strains over the four treatment eras (Fig. 1B). The first era began in 1987 when AZT (azidothymidine, an NRTI) was introduced as a monotherapy. AZT was used by a high proportion (36 to 68%) of MSM in San Francisco (17–19). Single-class resistance to NRTIs arose quickly (1), because they were ineffective at suppressing viral loads (19). In 1992, the second era began when dual therapies (based on two NRTIs) were introduced. These therapies were substantially more effective than monotherapies and achieved 30 to 60% viral suppression (20, 21). Single-class resistance to NRTIs decreased, but dual-class resistance quickly developed, because many individuals had previously developed resistance to AZT. In 1996, the third era [early highly active antiretroviral therapy (HAART)] began when NNRTIs and PIs were used in triple-therapy regimens. Resistance to PIs was slow to emerge and has only risen to low levels, because multiple mutations are necessary to develop resistance to

most drugs in this class (22). By 2001, more effective triple therapies (characterized by dual PIs combined with NRTIs) were developed, marking the beginning of the fourth era (modern HAART). During this recent era, the overall level of TDR appears to have stabilized (2); the model-generated network also exhibits this behavior (Fig. 1B). Recent empirical data from San Francisco indicate that transmission of single-class resistance is high, that of dual-class is moderate, and that of triple-class is low. In addition, studies indicate that transmission of NNRTI resistance is greater than that of NRTI resistance, which is greater than transmission of PI resistance. The model-generated transmission network shows these same patterns (Fig. 1, C and D). Our modeling estimates the overall level of TDR in 2008 to be 14% [median: interquartile range (IQR) 11.4 to 16.5%] (Fig. 1C), which is in extremely close agreement with empirically derived estimates of 13 to 16% (2).
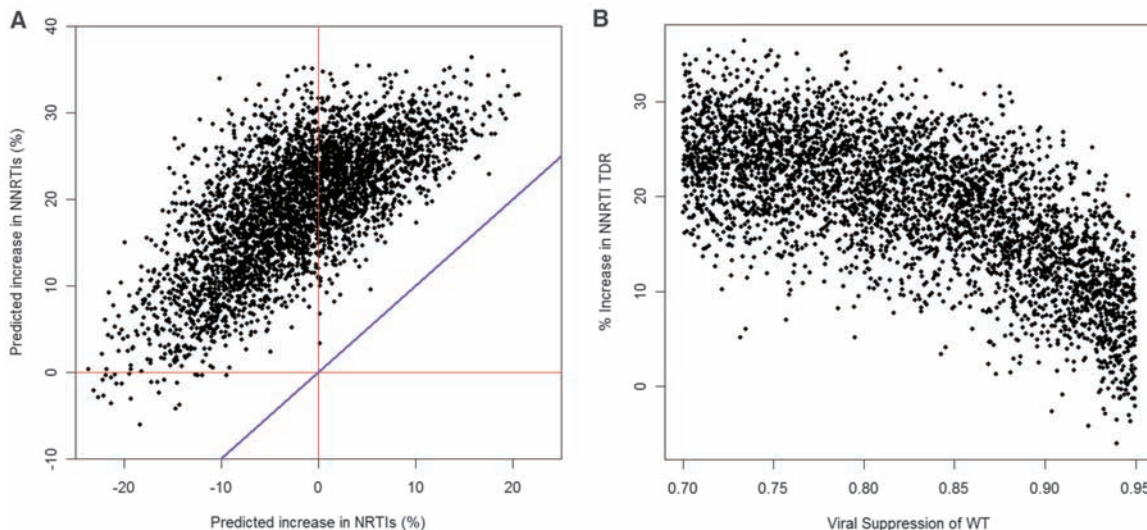
After reconstructing the historical epidemiology up to 2008, we simulated the amplification cascade model for 5 more years to predict the levels of TDR in 2013. Our simulations revealed that resistance to single-class NRTIs and PIs will remain at current levels, but NNRTI resistance will increase (Fig. 2A). Regression analysis determined that the degree of increase in NNRTI resistance will depend ($P < 0.05$) on the proportion of patients who are infected with wild-type strains and are being treated with a regimen of two NRTIs and one NNRTI and who achieve viral suppression (Fig. 2B). This proportion depends on the efficacy of the regimen and adherence to it; thus, if only 70% are virally suppressed, NNRTI re-

sistance could increase by more than 30% (Fig. 2B). This increase is predicted to be mainly due to transmission from untreated individuals infected with NNRTI-resistant strains who are in either the acute or chronic stage of infection.

The value of a strain's control reproduction number R_c specifies the average number, based on the probability that the individual is treated, of secondary HIV infections that an individual generates during their entire infectious period. R_c is a measure of a strain's transmission potential. A strain is capable of generating a self-sustaining epidemic if $R_c > 1$. The R_cs of the currently circulating ARV-resistant strains in San Francisco vary considerably (Fig. 3A). However, strains fall into three mutually exclusive groups (Fig. 3B) [section 4 of (9)]. Almost a quarter (24%) of the strains (Fig. 3B) cause less than one new infection ($R_c < 1$) and will eventually be eliminated (blue). Although other strains (Fig. 3B) also cause, on average, less than one new infection ($R_c < 1$), they will continue to be transmitted, because they evolve greater levels of resistance (green). We estimated that 60% of resistant strains have an $R_c > 1$ (Fig. 3B; red). Approximately 75% of these resistant strains have single-class resistance to NNRTIs, and 20% have dual-class resistance to NNRTIs and NRTIs. Although all have the potential to cause self-sustaining epidemics of resistance, they are all less infectious than the wild-type strains in San Francisco (Fig. 3C).

Similar trends for TDR to those observed in San Francisco and those predicted by our model have been documented in other cities in the

Fig. 2. (A) Predictions showing that transmission of strains that are resistant to NNRTIs will increase in San Francisco over the next 5 years. Predictions were made using Monte Carlo simulations from the uncertainty analysis of the amplification cascade model. Red lines show no increase in NNRTIs or NRTIs over the next 5 years. Blue line indicates an equal increase of NNRTIs and NRTIs over the next 5 years. **(B)** Predicted increase in the level of transmitted NNRTI resistance in San Francisco over the next 5 years as a function of the proportion of patients (who are infected with wild-type strains and are being treated with a regimen of two NRTIs and one NNRTI) who achieve viral suppression. Predictions were made using Monte Carlo simulations from the uncertainty analysis of the amplification cascade model.

United States and Europe that have analogous histories of ARV therapy. Potentially NNRTI-resistant strains similar to those we have identified in San Francisco may be increasing elsewhere. Although the NNRTI-resistant strains that we have identified are causing the rising wave of NNRTI resistance, they are unlikely to lead to self-sustaining epidemics in San Francisco or other communities in resource-rich countries, because new drugs will continue to become available. However our results may have important implications for HIV treatment programs in resource-constrained countries, where second-line regimens are not generally available. NNRTI-resistant strains are already evolving in many of these countries, because their first-line regimens are based on two NRTIs plus one NNRTI. Our current predictions have been obtained by modeling the evolution of resistance in individuals infected with subtype B strains. Subtype B accounts for ~12% of worldwide infections (and persons with subtype B are the most ARV-experienced), but 50% of prevalent HIV infections and 47% of all new HIV infections worldwide are caused by subtype C (*23*). Although information is limited, preliminary data suggest that treatment response and resistance patterns for subtype C are similar to those of subtype B (*24*). These data suggest that our results are likely to be generalizable to an epidemic of HIV-1 resistance among individuals infected with HIV-1 subtype C, and NNRTI-resistant strains with $R_c > 1$ could emerge in resource-constrained countries. If the R_c of the wild-type strains is reduced below one, as could occur by using a universal testing and treatment strategy (*25*), self-sustaining epidemics of NNRTI-resistant strains could arise (Fig. 3B and fig. S5) [section 5 of (*9*)].

Current levels of TDR, as well as the biological composition of the complex multistrain network, have emerged from two decades of treatment. To identify the key drivers of ARV resist-
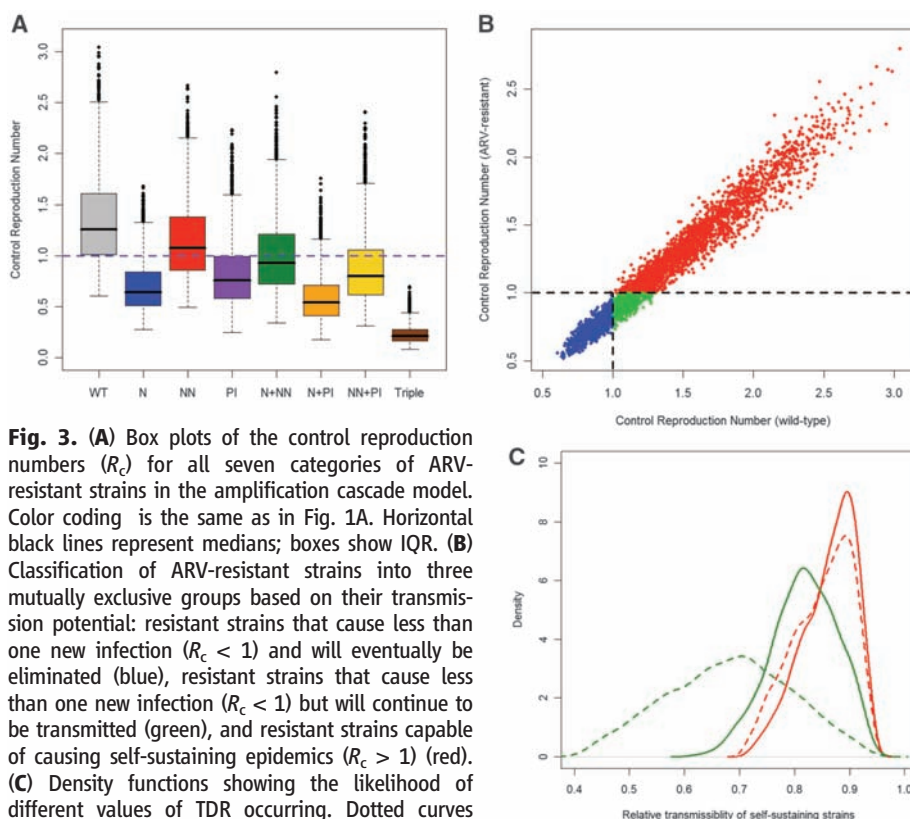
Fig. 3. (A) Box plots of the control reproduction numbers (R_c) for all seven categories of ARV-resistant strains in the amplification cascade model. Color coding is the same as in Fig. 1A. Horizontal black lines represent medians; boxes show IQR. **(B)** Classification of ARV-resistant strains into three mutually exclusive groups based on their transmission potential: resistant strains that cause less than one new infection ($R_c < 1$) and will eventually be eliminated (blue), resistant strains that cause less than one new infection ($R_c < 1$) but will continue to be transmitted (green), and resistant strains capable of causing self-sustaining epidemics ($R_c > 1$) (red). **(C)** Density functions showing the likelihood of different values of TDR occurring. Dotted curves show density functions for the relative transmissibility for all of the strains with single-class resistance to NNRTIs (red: median 86%, IQR 81 to 89%) and dual-class resistance to NRTIs and NNRTIs (green: median 69%, IQR 60 to 76%) that are circulating in the current network in San Francisco. Solid curves show density functions for the relative transmissibility of NNRTI-resistant strains with $R_c < 1$: single-class NNRTIs (red: median 87%, IQR 83 to 90%) and dual-class NRTIs and NNRTIs (green: median 82%, IQR 78 to 86%). Transmissibility is defined relative to the wild type.

ance, we constructed classification and regression trees (CART) (*26*) using the 20-year data set (1987 to 2008) that was generated during the uncertainty analysis of the amplification cascade model. To build trees, we used the model's estimated level of TDR for 2008 as the response variable and the model's 50 parameters as predictor variables [section 6 of (*9*)]. The optimal tree revealed the hidden hierarchical structure of the data (Fig. 4). Key drivers of TDR are the predictor variables with

the highest importance scores (IS) (table S13) (9). The most important driver (IS = 100) is the average time (at the population level) it takes for CD4 cell counts in infected individuals to fall below 350 cells/μl (v^{-1}) (Fig. 4). TDR was significantly higher (>15%) when CD4 counts fell to this threshold within ~6 years than when counts fell more gradually (Fig. 4 and fig. S6A) (9). This occurred because faster immunological deterioration led to increased treatment rates and accelerated the acquisition of resistance; hence, TDR increased as v^{-1} decreased.

A high proportion of the transmission of wild-type strains over the past 20 years has occurred from asymptomatic individuals with a CD4 count > 350 cells/μl (fig. S6B) (9). Consequently, α_1^H, the infectiousness of strains in asymptomatic individuals, has been the second key driver of TDR (IS = 73) (Fig. 4); infectivity is defined in terms of the probability of transmitting HIV per sex act. These results can be understood in terms of classical competition theory (27): The most infectious wild-type strains had the greatest advantage over resistant strains and hence caused the lowest levels of TDR. A recent review of empirical estimates of the transmission probability per sex act indicates that

α_1^H is likely to be greater than 0.0024 (28). The tree (Fig. 4) reveals that if wild-type strains had been less infectious (specifically, $\alpha_1^H \leq 0.0024$), it would have been very likely (probability 0.71) that TDR in San Francisco would be even higher than the current level (fig. S6C) (9).

We found α_1^T, the infectiousness of strains under treatment pressure, to be the third key driver of TDR (IS = 60) (Fig. 4). This driver represents the probability that an individual who is receiving current ARV regimens transmits HIV during one sex act. In contrast to our previous finding for α_1^H, TDR was significantly higher (>15%) when wild-type strains were more infectious ($\alpha_1^T > 0.0015$) than when they were less infectious ($\alpha_1^T \leq 0.0015$) (Fig. 4). This paradoxical result cannot be understood in terms of classical competition theory (27). It occurred because the effect of evolution on network dynamics was greater than that of competition. Under treatment pressure, the most infectious wild-type strains ($\alpha_1^T > 0.0015$) tended to evolve into the most infectious resistant strains; α_1^T only had a minor effect on competition, because treated individuals were relatively unimportant in transmitting wild-type strains (fig. S6D) (9). The value of α_1^T can be translated into viral load

(fig. S2) [section 2 of (9)]; a value of 0.0015 corresponds to a viral load of 20,000 copies/ml. Effective therapies used in recent years have reduced viral loads in patients infected with wild-type strains to well below 20,000 copies/ml (29), indicating that α_1^T is (and was) significantly less than 0.0015. Given these effective treatments, our tree shows it is highly unlikely (probability 0.22) that TDR in San Francisco could have risen to more than 15% by 2008 (Fig. 4).

Our CART analysis also identified four other parameters that are important drivers of TDR, including the relative transmissibility of strains with single-class resistance to NRTIs (λ_2) (IS = 51), the degree of viral suppression in patients who are infected with wild-type strains and not completely virologically suppressed (γ_1) (IS = 45), the relative transmissibility of strains with dual-class resistance to NRTIs and NNRTIs (λ_5) (IS = 40), and finally the degree of viral suppression in patients who are infected with strains that have single-class resistance to NNRTIs and are not completely virologically suppressed (γ_3) (IS = 39). None of the 43 other predictor variables was found to be important (IS < 30). The tree shows that TDR has remained below 15% because of specific immunologic, virologic, and treatment factors operating in San Francisco (Fig. 4).

The amplification cascade model can be recalibrated and reparameterized to assess the dynamics of networks of ARV-resistant strains of HIV in any setting where ARVs are available. We have applied it to San Francisco. We have shown that a complex network of HIV strains has arisen in this city due to two decades of sequential selection for resistance; first with single agents, then dual agents, and, more recently, a combination of multiple-class agents. By designing a biologically complex multistrain network model, we have obtained important insights into the otherwise hidden dynamics of drug-resistant strains of HIV. We have identified the key immunological, virological, and treatment variables, as well as the hierarchical interactions among these variables, which have had a key role in driving resistance. Our results have shown that effective treatments have prevented TDR from increasing to greater than 15% in San Francisco. However, our modeling shows that the network is continuing to evolve. We found that the majority of the resistant strains currently being transmitted in this city are capable of causing self-sustaining epidemics, and we have estimated that an individual with an NNRTI-resistant strain can cause, on average, more than one new infection. We predict that a wave of NNRTI-resistant strains will emerge over the next 5 years in San Francisco due to transmission from untreated individuals. Our results also have implications for resource-constrained countries where first-line regimens are based on NNRTIs. If the resistant strains we have identified in our analyses evolve in these countries, they could substantially compromise HIV treatment programs. Consequently, current-
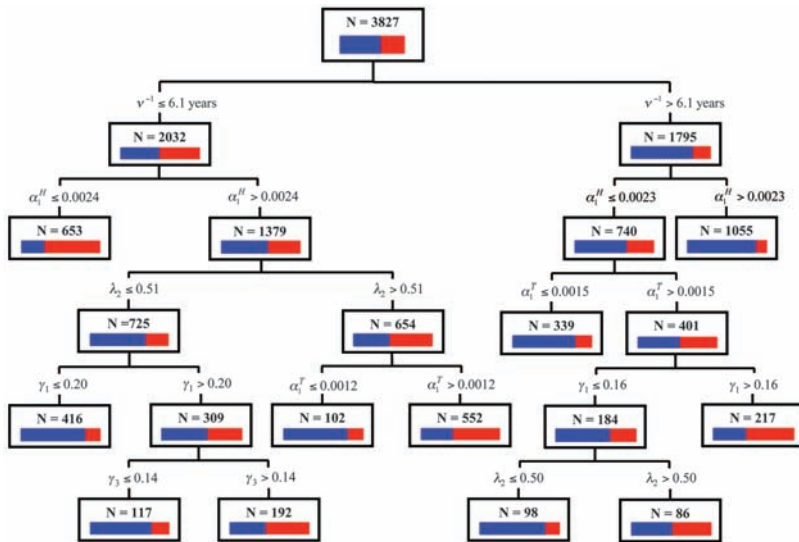
Fig. 4. A pruned version of the optimal tree. The root node contains data from the 3827 filtered Monte Carlo simulations that were generated by the amplification cascade model; filtered simulations are after model calibration [section 3 of (9)]. Inside each node is the total number of simulations it contains (N), as well as the distribution of the response variable TDR. Low levels of TDR (<15%) are blue, whereas high levels of TDR (>15%) are red. The most important variable (IS = 100) is v^{-1}, the average time (at the population level) it takes for CD4 cell counts in infected individuals to fall below 350 cells/μl. The variable α_1^H reflects the degree of infectivity of wild-type strains during the asymptomatic stage of infection, where infectivity is specified as the probability of transmitting HIV during one sex act. The variable α_1^T represents the probability that an individual receiving a current ARV regimen transmits HIV during one sex act. The remaining variables are as follows: the transmissibility of strains (relative to the wild type) with single-class resistance to NRTIs (λ_2), the degree of viral suppression in patients who are infected with wild-type strain and are not completely virologically suppressed (γ_1), and the degree of viral suppression in patients who are infected with strains that have single-class resistance to NNRTIs and are not completely virologically suppressed (γ_3). Because the pruned tree is a subtree of the optimal tree, not every variable deemed important appears in it. The optimal tree has 84% predictive power in correctly identifying which simulations will generate high levels of TDR and 82% predictive power in correctly identifying which simulations will generate low levels of TDR.

ly circulating NNRTI-resistant strains in San Francisco pose a great and immediate threat to global public health.

References and Notes

1. B. A. Larder, G. Darby, D. D. Richman, *Science* **243**, 1731 (1989).
2. H. M. Truong *et al.*, *AIDS* **20**, 2193 (2006).
3. S. M. Blower, H. B. Gershengorn, R. M. Grant, *Science* **287**, 650 (2000).
4. J. Goudsmit *et al.*, *AIDS* **15**, 2293 (2001).
5. A. Phillips, *Nat. Med.* **7**, 993 (2001).
6. E. Tchetgen, E. H. Kaplan, G. H. Friedland, *J. Acquir. Immune Defic. Syndr.* **26**, 118 (2001).
7. R. Vardavas, S. Blower, J. J. Miranda, *PLoS ONE* **2**, e152 (2007).
8. G. S. Zaric, M. L. Brandeau, A. M. Bayoumi, D. K. Owens, *Simulation* **71**, 262 (1998).
9. Materials and methods are available as supporting material on *Science* Online.
10. C. L. Booth, A. M. Geretti, *J. Antimicrob. Chemother.* **59**, 1047 (2007).
11. J. Martinez-Picado, M. A. Martínez, *Virus Res.* **134**, 104 (2008).
12. L. Ross, N. Parkin, R. Lanier, *AIDS Res. Hum. Retroviruses* **24**, 617 (2008).
13. S. Blower, H. Dowlatabadi, *Int. Stat. Rev.* **62**, 229 (1994).
14. W. Lang *et al.*, *JAMA* **257**, 326 (1987).
15. W. Winkelstein Jr. *et al.*, *JAMA* **257**, 321 (1987).
16. J. K. Louie, L. C. Hsu, D. H. Osmond, M. H. Katz, S. K. Schwarcz, *J. Infect. Dis.* **186**, 1023 (2002).
17. W. Lang, D. Osmond, K. Page-Bodkin, A. Moss, W. Winkelstein Jr., *J. Acquir. Immune Defic. Syndr.* **6**, 191 (1993).
18. W. Lang *et al.*, *J. Acquir. Immune Defic. Syndr.* **4**, 713 (1991).
19. J. C. Schmit *et al.*, *AIDS* **12**, 2007 (1998).
20. J. J. Eron *et al.*, *N. Engl. J. Med.* **333**, 1662 (1995).
21. C. Katlama *et al.*, *JAMA* **276**, 118 (1996).
22. D. R. Kuritzkes, *AIDS Patient Care STDS* **18**, 259 (2004).
23. B. S. Taylor, M. E. Sobieszczyk, F. E. McCutchan, S. M. Hammer, *N. Engl. J. Med.* **358**, 1590 (2008).
24. A. J. Kandathil *et al.*, *Indian J. Med. Microbiol.* **27**, 231 (2009).
25. R. M. Granich, C. F. Gilks, C. Dye, K. M. De Cock, B. G. Williams, *Lancet* **373**, 48 (2009).
26. L. Breiman, J. H. Freidman, R. A. Olshen, C. J. Stone, *Classification and Regression Trees* (Chapman & Hall, Boca Raton, FL, 1984).
27. C. Darwin, *The Origin of Species* (Signet, London, new ed. 1, 2003).
28. K. A. Powers, C. Poole, A. E. Pettifor, M. S. Cohen, *Lancet Infect. Dis.* **8**, 553 (2008).
29. S. M. Hammer *et al.*, *JAMA* **300**, 555 (2008).
30. R.J.S., J.T.O., E.N.B., and S.B. acknowledge the financial support of the National Institute of Allergy and Infectious Diseases (NIAID) (RO1 AI041935). R.J.S. is supported by a Natural Sciences and Engineering Research Council of Canada discovery grant, an Early Researcher award, and funding from Mathematics of Information Technology and Complex Systems. In addition, S.B. acknowledges the John Simon Guggenheim Foundation, the National Academies Keck Foundation, and the Semel Institute for Neuroscience & Human Behavior. J.S.K. acknowledges NIAID grants P30-AI27763, NCRR K24RR024369, and AHRQ R18-HS017784. We thank R. Breban, D. Freimer, J. Freimer, N. Freimer, N. Jewell, E. Kajita, T. Pylko, V. Supervie, and R. Vardavas for useful discussions throughout the course of this research.

Optimal Localization by Pointing Off Axis

Yossi Yovel,[1] Ben Falk,[2] Cynthia F. Moss,[2] Nachum Ulanovsky[1]*

Is centering a stimulus in the field of view an optimal strategy to localize and track it? We demonstrated, through experimental and computational studies, that the answer is no. We trained echolocating Egyptian fruit bats to localize a target in complete darkness, and we measured the directional aim of their sonar clicks. The bats did not center the sonar beam on the target, but instead pointed it off axis, accurately directing the maximum slope ("edge") of the beam onto the target. Information-theoretic calculations showed that using the maximum slope is optimal for localizing the target, at the cost of detection. We propose that the tradeoff between detection (optimized at stimulus peak) and localization (optimized at maximum slope) is fundamental to spatial localization and tracking accomplished through hearing, olfaction, and vision.

Most sensory systems allow some active control over the information acquired from the environment (*1–6*). Nowhere is this more evident than in echolocating bats (*4, 7–10*), which control many aspects of their sonar signal design (*4, 7, 9, 11–16*) and use returning echoes to orient and forage in the dark (*4, 7–16*). We trained Egyptian fruit bats to fly in a large flight room and land on a spherical target while relying exclusively on sonar (*17*). The bats' three-dimensional (3D) position was measured with two infrared cameras, and the shape and direction of their sonar beam pattern were measured with a 20-microphone array (*17*) (Fig. 1, A to D, and movie S1).

At the beginning of each trial, the target was randomly repositioned. Subsequently, the bat searched for the target, approached it, and landed on it, either by a straight flight or a curved trajectory (Fig. 1C and fig. S1). Unlike micro-bats (microchiropteran bats), which emit laryngeal tonal calls, Egyptian fruit bats are megabats (megachiropteran bats) that produce very short (50- to 100-μs) impulse-like tongue clicks, with frequencies centered at 30 to 35 kHz (fig. S2). While flying, bats typically emitted pairs of clicks, with an ~20-ms interval within the click pair and an ~100-ms interval between the pairs (Fig. 1A and fig. S3) (*18, 19*). The bats pointed their sonar beam toward the left or the right, in an alternating manner as follows: left→right→100-ms interval→right→left (Fig. 1D and movie S1).

We observed two different phases of behavior. During the first stage, the bats did not necessarily lock their click pairs onto the target, and the directions of clicks were widely distributed (the "unlocked" phase). At the final stage, the bats directed their sonar clicks so that the vector average of the pair of clicks pointed toward the target with accuracy better than 30°

(*17*). We refer to this as the "locked" phase (Figs. 1E, arrows, and 2A, top, and fig. S1C). During this phase, 0.5 s before landing, 80% of the click pairs were locked with accuracy better than 15° (Fig. 2A, bottom, gray lines). In 10% of the trials, the bats locked onto the target with average accuracy better than 5°. The left-right orientation of the clicks in the locked phase implies that the bats did not direct the maximum intensity of the click toward the target, contradicting the common notion that bats steer their sonar beam in order to maximize the signal-to-noise ratio (SNR) of the echoes (*13, 20*).

Another possible strategy would be for the bats to direct the maximal slope of the beam's emission curve toward the target, because this would maximize changes in reflected echo energy that result from changes in the relative position of the bat and the target. Plotting the directional span of the beams between the right and left maximum slope (green lines in Fig. 1, E and F, and fig. S1, C and D) showed that the bats consistently placed the maximum slope of their beams onto the target (Fig. 1F and fig. S1D; the top and bottom of the green lines are close to direction 0°). Next, we examined the population distribution of the directions of the beams' maximum intensity and maximum slope (Fig. 2, B and C, top two rows). Before locking, the bats directed their sonar beams over a wide range of angles, spanning >100° around the target (Fig. 2B, top). After locking, however, they clearly directed their beam so that the maximum slope of the intensity curve of the beam, and not its peak, was on the target (Fig. 2C, middle row). All six bats exhibited this behavior (fig. S4).

When the maximum slope of the beam is directed toward an object, any motion of the object relative to the bat will result in the largest possible change in echo intensity. The sign of the energy change (positive or negative) corresponds to the

[1]Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Department of Psychology, Institute for Systems Research and Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD 20742, USA.

*To whom correspondence should be addressed. E-mail: nachum.ulanovsky@weizmann.ac.il

Supporting Online Material for

Evolutionary Dynamics of Complex Networks of HIV Drug-Resistant Strains: The Case of San Francisco

Robert J. Smith?, Justin T. Okano, James S. Kahn, Erin N. Bodine, Sally Blower*

*To whom correspondence should be addressed. E-mail: sblower@mednet.ucla.edu

This PDF file includes:

<u>Table of Contents</u>

Section 1: Amplification Cascade Model: Structure & Equations

All published HIV transmission models of ARV-resistance are based on simple biological assumptions and can only track one resistant strain (*S1-S8*). These models were useful predictive tools for ARV resistance in the early years of treatment, but are now inadequate for predicting the complex dynamics of the multitude of ARV-resistant strains circulating in resource-rich countries. Therefore a new generation of predictive models is needed. We have designed the first of this generation of mathematical models: the Amplification Model. Our innovative model captures a high degree of biological complexity by generating by generating a dynamic network composed of multiple strains of HIV, both wild type and ARV resistant.

In the model, susceptible/uninfected individuals (S) can become infected with any of the wild-type or ARV-resistant strains in the network. Once an individual becomes infected, they progress through four stages: (i) primary infection (P), (ii) not yet eligible for ARVs (i.e., CD4 count > 350 cells/microL) (H), (iii) eligible for ARVs (i.e., CD4 ≤ 350 cells/microL) but not on treatment (Y), and (iv) treated (T). The ARV treatment drugs are documented in Table S1. Viral load is modeled (see Section 2 and Table S2) such that P has the highest viral load, H a low viral load, Y a higher viral load than the previous state and T the lowest viral load. *N* represents the total number of individuals in the sexually active community.

Resistant strains in the network are grouped into seven categories which are defined based upon the specific class of drugs they are resistant to (Nucleoside Reverse Transcriptase Inhibitors (NRTIs), non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) and Protease Inhibitors (PIs)) and their level of resistance (single, dual or triple class). Categories are denoted by subscript i: wild-type strains $(i=1)$, single-class resistance to NRTIs $(i=2)$, single-class resistance to NNRTIs $(i=3)$, single-class resistance to PIs $(i=4)$, dual-class resistance to NRTIs and NNRTIs $(i=5)$, dual-class resistance to NRTIs and PIs $(i=6)$, dual-class resistance to NNRTIs and PIs $(i=7)$, and triple-class resistance to NRTIs, NNRTIs and PIs $(i=8)$.

We model treatment effects by specifying treatment regimens and then assessing the effects of these regimens on both infectivity and the probability of developing resistance. In our modeling, treated individuals generally receive a regimen to which their virus is sensitive; hence, we assume treated individuals either achieve complete or partial viral suppression. In our model, patients who achieve complete viral suppression become uninfectious and are incapable of developing resistance. We represent the percentage of patients who achieve viral suppression by (1 - γ_i). Patients who only achieve partial viral suppression retain some degree of infectivity (which is based on the degree of treatment-induced reduction in viral load) and are capable of developing resistant strains.

The amplification cascade model can be extended to include any number of additional drug classes, such as integrase inhibitors, co-receptor blockers and fusion inhibitors which are gradually being introduced into new therapeutic regimens. These classes are currently used almost exclusively among patients who have failed other classes of ARVs.

A flow diagram of the complete model is shown in Fig. S1. The model consists of 33 ordinary differential equations, but using matrix notation (matrices in bold), the model can be written as:

$$\dot{S} = \Omega - \mu_0 S - \frac{cS}{N} \sum_{i=1}^{8} \left(\beta_i^P P_i + \beta_i^H H_i + \beta_i^Y Y_i + \gamma_i \beta_i^T T_i \right)$$

$$\dot{\mathbf{P}} = \frac{cS}{N} \left(\boldsymbol{\beta}^P \mathbf{P} + \boldsymbol{\beta}^H \mathbf{H} + \boldsymbol{\beta}^Y \mathbf{Y} + \boldsymbol{\Gamma} \boldsymbol{\beta}^T \mathbf{T} \right) - \mu_0 \mathbf{P} - \boldsymbol{\theta} \mathbf{P}$$

$$\dot{\mathbf{H}} = \boldsymbol{\theta} \mathbf{P} - \mu_0 \mathbf{H} - \mathbf{v} \mathbf{H} \qquad\qquad (1)$$

$$\dot{\mathbf{Y}} = \mathbf{v} \mathbf{H} - \boldsymbol{\rho} \mathbf{Y} - \boldsymbol{\mu}^Y \mathbf{Y} + \boldsymbol{\omega} \mathbf{T}$$

$$\dot{\mathbf{T}} = \boldsymbol{\rho} \mathbf{Y} - \boldsymbol{\omega} \mathbf{T} - \boldsymbol{\mu}^T \mathbf{T} - \mathbf{K} \mathbf{T}$$

where $N = S + \sum_{i=1}^{8} (P_i + H_i + Y_i + T_i)$.

The variable state vectors are

$$\mathbf{P} = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \\ P_6 \\ P_7 \\ P_8 \end{bmatrix}, \ \mathbf{H} = \begin{bmatrix} H_1 \\ H_2 \\ H_3 \\ H_4 \\ H_5 \\ H_6 \\ H_7 \\ H_8 \end{bmatrix}, \ \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ Y_8 \end{bmatrix}, \ \mathbf{T} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \\ T_7 \\ T_8 \end{bmatrix},$$

The parameter matrices are

$$\boldsymbol{\beta}^P = diag\left(\beta_i^P\right), \ \boldsymbol{\beta}^H = diag\left(\beta_i^H\right), \ \boldsymbol{\beta}^Y = diag\left(\beta_i^Y\right), \ \boldsymbol{\beta}^T = diag\left(\beta_i^T\right), \ \boldsymbol{\Gamma} = diag\left(\gamma_i\right)$$

$$\boldsymbol{\theta} = diag\left(\theta_i\right), \ \mathbf{v} = diag\left(v_i\right), \ \boldsymbol{\rho} = diag\left(\rho_i\right), \ \boldsymbol{\omega} = diag\left(\omega_i\right), \ \boldsymbol{\mu}^Y = diag\left(\mu_i^Y\right), \ \boldsymbol{\mu}^T = diag\left(\mu_i^T\right),$$

where $i \in \{1, 2, 3, 4, 5, 6, 7, 8\}$.

Here, Ω is the rate at which susceptible/uninfected individuals join the sexually active community; $1/\mu_0$ is the average time period for acquiring new sex partners; c is the average number of new sex partners per individual per year; β_i^P is the per-partnership probability of an individual in the primary stage of infection transmitting strains in category i; β_i^H is the per-partnership probability of an infected individual with a CD4 count greater than 350 cells/microL transmitting strains in category i; β_i^Y is the per-partnership probability of an infected untreated individual with a CD4 count less than or equal to 350 cells/microL transmitting strains in category i; β_i^T is the per-partnership probability of an individual on treatment transmitting strains in category i; $1/\theta_i$ is the average duration of primary infection; $1/v_i$ is

the average time that an infected individual's CD4 counts remain above 350 cells/microL; $1/\rho_i$ is the average time an individual spends in the treatment-eligible stage; $1/\mu_i^Y$ is the average survival time of an individual (infected with strains in category i) who has a CD4 count less than or equal to 350 cells/microL and is not on treatment; ω_i is the rate at which individuals go off treatment; and $1/\mu_i^T$ is the average survival time of an individual (infected with strains in category i) on treatment.

The **K** matrix specifies the degree of acquisition and amplification of resistance from one state to the next, accounting for the fact that the order in which resistance develops is important.

$$\mathbf{K} = \begin{bmatrix} \gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\gamma_1 k_{12} & \gamma_2 k_{25} + \gamma_2 k_{26} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\gamma_1 k_{13} & 0 & \gamma_3 k_{35} + \gamma_3 k_{37} & 0 & 0 & 0 & 0 & 0 \\ -\gamma_1 k_{14} & 0 & 0 & \gamma_4 k_{46} + \gamma_4 k_{47} & 0 & 0 & 0 & 0 \\ 0 & -\gamma_2 k_{25} & -\gamma_3 k_{35} & 0 & \gamma_5 k_{58} & 0 & 0 & 0 \\ 0 & -\gamma_2 k_{26} & 0 & -\gamma_4 k_{46} & 0 & \gamma_6 k_{68} & 0 & 0 \\ 0 & 0 & -\gamma_3 k_{37} & -\gamma_4 k_{47} & 0 & 0 & \gamma_7 k_{78} & 0 \\ 0 & 0 & 0 & 0 & -\gamma_5 k_{58} & -\gamma_6 k_{68} & -\gamma_7 k_{78} & 0 \end{bmatrix}$$

Here, $k_{ii'}$ is the per-capita rate at which individuals in treatment class T_i develop resistance and move to treatment class $T_{i'}$, where $i, i' \in \{1, 2, 3, 4, 5, 6, 7, 8 \mid i < i'\}$. An equivalent, but more physically meaningful measure is $1 - e^{k_{ii'}}$, which represents the proportion of individuals in treatment class T_i who develop resistance per year.

The parameter γ_i represents the proportion of treated individuals who are resistant to strains in category i and who are not virally suppressed.

Section 2: Parameter Estimates for Monte Carlo Simulations

We estimated all model parameters from virologic, epidemiological or clinical data (Tables S2-S11).

We calculated per-act and per-partner transmission probabilities based on viral load data which was a function of the stage of HIV infection. We used a relationship between viral load and the per-act transmission probability α_i^j (Equation 1) that Smith and Blower (*S9*) derived based on data from references (*S10*) and (*S11*).

$$\frac{\alpha_i^j(v)}{\alpha_i^j(w)} = 2.45^{\log_{10}(v/w)} \qquad\qquad (2)$$

This relationship, depicted in Fig. S2, enables us to calculate the probability of transmission for an individual in disease stage $j \in \{P,H,Y,T\}$ and resistance category i. Equation 2 compares the viral load of interest, v, with a baseline viral load, w. We used data from Gray *et al.*, which compares viral load to per-act transmission probability (*S11*), to calculate a baseline. Since a range of viral loads were given, we used the largest viral load from each range for 25-29 year olds. To test the baseline, we set $v = 38,500$, $w = 12,500$ and $\alpha_i^j(w) = 0.0018$. Applying Equation 2 gives $\alpha_i^j(v) = 0.0028$, which is close to the actual per-act transmission probability of 0.0026 observed in the Gray *et al.* study. This suggests that these data can be used as a reasonable baseline for Equation 2.

We used a weighting factor λ_i to ensure that resistant strains were less transmissible (per act) than wild-type strains and included sampling constraints so that transmissibility decreased as the number of classes of resistance increased (i.e., $\alpha_1^P > \alpha_2^P, \alpha_3^P, \alpha_4^P > \alpha_5^P, \alpha_6^P, \alpha_7^P > \alpha_8^P$ and similar for α_i^H etc). Based on available competitive fitness assays, replication capacity assays and patterns of developed resistance, we assumed NNRTI-resistant strains were more transmissible than NRTI-resistant strains, which in turn were more transmissible than PI-resistant strains (*S12-S17*). The values for λ_i are given in Table S3. We then used these estimates to determine the per-partnership transmission probability β_i^j as follows:

$$\beta_i^j = 1 - \left[1 - \alpha_i^j(w) \right]^{n/c}$$

for an individual in disease stage $j \in \{P,H,Y,T\}$ transmitting strains in category i. Here, n is the average number of sex acts per year, c (as defined previously) is the average number of new sex partners per individual per year and n/c is therefore the number of sex acts per partnership. For our Monte Carlo simulations, we sampled c from a triangular distribution ($1 < c < 5$, with a peak at 1) and n/c from a triangular distribution ($5 < n/c < 60$, with a peak at 10).

Probability density functions (pdfs) for estimates of per-act and per-partnership transmission probabilities for individuals infected with wild-type strains are given in Table S2. To estimate parameters, we sampled with constraints so that the per-partnership probabilities satisfied the following relationships: $\beta_i^H < \beta_i^P$, $\beta_i^H < \beta_i^Y$ and $\beta_i^T \leq \beta_i^Y$. These constraints ensured that: (i) individuals had the highest viral load (and therefore transmissibility) during primary infection (ii) viral load (and transmissibility) decreased after the primary infection stage, (iii) viral load (and transmissibility) increased as the individuals CD4 counts fell below 350 cells/microL and (iv) viral load decreased under treatment (except in the first treatment era of monotherapy).

In San Francisco, monotherapy began in 1987 with the NRTI Zidovudine (AZT). At this time, almost half of the MSM community was estimated to be infected with HIV (*S18-S20*). In the late eighties ARV resistance arose quickly (*S21*) because AZT was ineffective in suppressing viral loads (*S22*) and was very widely used (*S23-S26*). In 1992, dual therapy became available, featuring treatment regimens of two NRTIs at a time (for example, Zidovudine (then AZT) and Didanosine). Dual therapies were substantially more effective than monotherapies and achieved 30-60% viral suppression (*S27-S30*).

In 1996, Highly Active Antiretroviral Therapy (HAART) became available with the arrival of two new classes of drugs: NNRTIs and PIs. Due to the different efficacy of treatment regimens, we subdivided the HAART era into two: Early HAART (1996-2000) and Modern HAART (2001-2008). The treatment eras reflect the approval and rollout of new medications. Fig. S3 shows a detailed timeline of the introduction of drugs used to treat HIV infection; these drugs are also categorized under each treatment era in Table S1. The specific treatment regimens that we model are given in Tables S6 (for Era 1), S7 (for Era 2), S8 (for Era 3) and S9 (for the current era of HAART: Era 4).

For each era, we modeled the impact of the specific regimens that were available by varying: (i) the percentage of treatment-eligible individuals who began treatment per year ($100 \cdot (1 - e^{-\rho_i})$), (ii) the per-act and per-partner transmissibility (α_i^j and β_i^j), (iii) the percentage of patients who were virally suppressed ($100 \cdot (1 - \gamma_i)$), (iv) the percentage of patients who acquired resistance ($100 \cdot (1 - e^{k_{ii'}})$), and (v) the average time spent on treatment ($1/\mu_i^T$). Pdfs for these parameter estimates for the regimens used in each era are given in Tables S2-S11. Notably, resistance rates vary considerably for different classes of ARVs. The rate at which a strain develops resistance to a drug depends on the genetic barrier (i.e., the number of mutations needed to acquire resistance to that specific drug). For example, single point mutations lead to resistance for some NRTIs (e.g., lamivudine and emtricitabine); however, multiple mutations are necessary to develop resistance to most PIs.

For each era, we computed $1/\mu_i^T$, the average time spent in the treatment stage for individuals infected with strains in category i (Table S11). To make these computations, we first derived Equation 3.

$$\frac{1}{\mu_i^Y} + \frac{1}{\mu^g} = \frac{1}{\mu_i^Y + \rho_i \left(1 - \dfrac{\omega}{\mu_i^T + \omega + \gamma_i \sum k_{ii'}}\right)} + \frac{1}{\mu_i^T} \tag{3}$$

Here, $1/\mu_i^Y$ is the average survival time of an individual (infected with strains in category i) who is untreated but has a CD4 count less than or equal to 350 cells/microL. $1/\mu^g$ represents the additional life years gained due to treatment ($g \in \{a, b, c, d\}$ where a represents wild-type, b represents single-class resistance, c represents dual-class resistance and d represents triple-class resistance). These parameters were estimated in (*S31*) and appear in Table S10. It should be noted that $1/\mu_i^Y$ + $1/\mu^g$ represents the average survival time from when a treated individual's CD4 count falls below 350 cells/microL. We then estimated the average time an infected individual spent in the treatment-eligible stage (i.e., with a CD4 count less than 350 cells/microL) before they received treatment. Finally, using Equation 3, we calculated the average time spent on treatment ($1/\mu_i^T$) (Table S11).

The remaining parameter estimates that specify the natural history of HIV infection were obtained from reference (*S32*) and are given in Table S10.

To reconstruct the evolution and transmission dynamics of the network of ARV-resistant strains from 1987 to 2008 (that we present in the main text) we analyzed the amplification cascade model using a time-dependent uncertainty analysis, (*S33, S34*).

Section 3: Model Calibration using Monte Carlo filtering

We calibrated our model using Monte Carlo filtering techniques. We began this calibration process by deriving an analytical expression for the Basic Reproduction Number (R_0); R_0 is the average number of secondary infections caused by a single infectious individual in a wholly susceptible population in the absence of treatment. It is a threshold condition that determines whether the pathogen will die out or become endemic.

We derived R_0 from the amplification cascade model using the following method (*S35, S36*):

Step 1:	Construct matrices $\mathbb{F}$ and $\mathbb{V}$ where $\mathbb{F}$ is an $n \times 1$ column vector representing the new infections in each infected state and $\mathbb{V}$ is an $n \times 1$ column vector representing the transported (non-new infections) in each infected state and where n is the number of infected states.
Step 2:	Construct matrices $\mathbf{F}$ and $\mathbf{V}$, the Jacobian matrices of $\mathbb{F}$ and $\mathbb{V}$, respectively, evaluated at the disease-free equilibrium.
Step 3:	The basic reproduction number R_0 is the maximum eigenvalue of $\mathbf{F} \cdot (-\mathbf{V})^{-1}$.

In Step 1, we derived

$$\mathbb{F} = \begin{bmatrix} \dfrac{cS}{N} \left(\beta_1^P P_1 + \beta_1^H H_1 + \beta_1^Y Y_1 \right) \\ 0 \\ 0 \end{bmatrix}$$

where the first row represents new infections in primary infection, the second row represents new infections in individuals with a CD4 count greater than 350 cells/microL and the third row represents new infections in individuals with a CD4 count less than or equal to 350 cells/microL. Furthermore,

$$\mathbb{V} = \begin{bmatrix} -\mu_0 P_1 - \theta_1 P_1 \\ \theta_1 P_1 - \mu_0 H_1 - v_1 H_1 \\ v_1 H_1 - \mu^Y Y_1 \end{bmatrix}$$

where the first row represents infections transported to primary infection, the second row represents infections transported to individuals with a CD4 count greater than 350 cells/microL and the third row represents infections transported to individuals with a CD4 count less than or equal to 350 cells/microL.

In Step 2, we calculated the Jacobian matrices of $\mathbb{F}$ and $\mathbb{V}$ evaluated at the disease-free equilibrium, i.e. $(S,0,0,0)$, and found that

$$\mathbf{F} = \begin{bmatrix} c\beta_1^P & c\beta_1^H & c\beta_1^Y \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{V} = \begin{bmatrix} -\mu_0 - \theta_1 & 0 & 0 \\ \theta_1 & -\mu_0 - \nu_1 & 0 \\ 0 & \nu_1 & -\mu_1^Y \end{bmatrix}.$$

Lastly, in Step 3, we found that

$$\mathbf{F} \cdot (-\mathbf{V})^{-1} = \begin{bmatrix} \dfrac{c\left(\beta_1^P \mu_1^Y (\mu_0 + \nu_1) + \beta_1^H \theta_1 \mu_1^Y + \beta_1^Y \theta_1 \nu_1\right)}{\mu_1^Y (\mu_0 + \theta_1)\mu_1^Y (\mu_0 + \nu_1)} & \dfrac{c\left(\beta_1^H \theta_1 \mu_1^Y + \beta_1^Y \theta_1 \nu_1\right)}{\mu_1^Y (\mu_0 + \nu_1)} & \dfrac{c\beta_1^Y}{\mu_1^Y} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Eigenvalues of this matrix are $\left\{0, 0, \dfrac{c\left(\beta_1^P \mu_1^Y (\mu_0 + \nu_1) + \beta_1^H \theta_1 \mu_1^Y + \beta_1^Y \theta_1 \nu_1\right)}{\mu_1^Y (\mu_0 + \theta_1)\mu_1^Y (\mu_0 + \nu_1)}\right\}$, the greatest of which can be

rewritten as

$$R_0 = \frac{c\beta_1^P}{\mu_0 + \theta_1} + \frac{c\beta_1^H}{\mu_0 + \nu_1} \cdot \frac{\theta_1}{\mu_0 + \theta_1} + \frac{c\beta_1^Y}{\mu_1^Y} \cdot \frac{\theta_1}{\mu_0 + \theta_1} \cdot \frac{\nu_1}{\mu_0 + \nu_1} \tag{4}$$

This expression can be understood biologically as the sum of the following three R_0's:

$$R_0^P = \frac{c\beta_1^P}{\mu_0 + \theta_1}, \quad R_0^H = \frac{c\beta_1^H}{\mu_0 + \nu_1} \cdot \frac{\theta_1}{\mu_0 + \theta_1} \quad \text{and} \quad R_0^Y = \frac{c\beta_1^Y}{\mu_1^Y} \cdot \frac{\theta_1}{\mu_0 + \theta_1} \cdot \frac{\nu_1}{\mu_0 + \nu_1}$$

$R_0^P = \dfrac{c\beta_1^P}{\mu_0 + \theta_1}$ represents the number of infections $\left(c\beta_1^P\right)$ an individual causes throughout the duration

$\left(1/\left(\mu_0 + \theta_1\right)\right)$ of $\left(P_1\right)$ (i.e., the primary infection stage).

$R_0^H = \dfrac{c\beta_1^H}{\mu_0 + \nu_1} \cdot \dfrac{\theta_1}{\mu_0 + \theta_1}$ represents the number of infections $\left(c\beta_1^H\right)$ the individual causes throughout the

duration $\left(1/\left(\mu_0 + \nu_1\right)\right)$ of stage $\left(H_1\right)$ (i.e., after primary infection but before CD4 counts have fallen

below 350 cells/microL) multiplied by the probability $\left(\dfrac{\theta_1}{\mu_0 + \theta_1}\right)$ they survive to the H_1 stage.

9

$R_0^Y = \dfrac{c\beta_1^Y}{\mu_1^Y} \cdot \dfrac{\theta_1}{\mu_0+\theta_1} \cdot \dfrac{v_1}{\mu_0+v_1}$ represents the number of infections $\left(c\beta_1^Y\right)$ the individual causes throughout the duration $\left(1/\mu_1^Y\right)$ of stage $\left(Y_1\right)$ (i.e., when they are untreated but treatment eligible) multiplied by the probability $\left(\dfrac{\theta_1}{\mu_0+\theta_1}\right)$ they survive to the H_1 stage and then the probability $\left(\dfrac{v_1}{\mu_0+v_1}\right)$ they survive to the Y_1 stage.

We then used the expression that we had derived for R_0 to calculate the values of the R_0's for wild-type strains in San Francisco in the pre-treatment era. We made these calculations using Equation 4, Latin Hypercube Sampling (*S33*) and pdfs for the model's parameter estimates given in Tables S2-S11; each pdf was sampled 10,000 times. This procedure resulted in 10,000 parameter sets, which led to a median estimate of R_0 of 0.84 (Inter-Quartile-Range (IQR) 0.49 to 1.43).

We then filtered these parameter sets in order to calibrate the model to match the prevalence of HIV in the MSM community in San Francisco in 1987 when treatment was first introduced. Prevalence in this city in the late eighties has been estimated to be as high as 50% (*S18-S20*); this corresponds to an R_0 of 2.0. We used the value of R_0 as a filtering criterion and determined how many of the 10,000 parameter sets generated an R_0 value between 1 and 3. This Monte Carlo filtering procedure reduced our parameter sets from 10,000 to 3,827; after filtering, the median R_0 value was 1.5 (Inter-Quartile-Range (IQR) 1.2 to 2.0).

We used Kolmogorov-Smirnov tests to compare the distributions of the model's parameters before and after filtering. Four of the parameter ranges were statistically different; for these, see Table S12.

The 3,827 filtered parameter sets were used to conduct the Monte Carlo simulations for our time-dependent uncertainty analysis and historical reconstructions (*S33*). For our Monte Carlo simulations, we modeled monotherapy for 5 years (1987-1991), dual therapy for 4 years (1992-1995), early HAART for 5 years (1996-2000) and modern HAART thereafter.

Section 4: A classification system for ARV-resistant strains

We derived a classification system for the resistant strains in the network by deriving an analytical expression for the Control Reproduction Number, R_c. The quantity R_c is a measure of the average number of secondary HIV infections an individual generates during their entire infectious period; it is calculated as a weighted average based on the probability the individual is treated. We used the same methods to derive R_c from the amplification cascade model that we had used previously to derive R_0 (*S35, S36*). We calculated a Control Reproduction Number (R_c) for each of the seven categories of ARV resistance included in the model, as well as for wild-type strains. Categories are denoted by subscript i: wild-type strains $(i=1)$, single-class resistance to NRTIs $(i=2)$, single-class resistance to NNRTIs $(i=3)$, single-class resistance to PIs $(i=4)$, dual-class resistance to NRTIs and NNRTIs $(i=5)$, dual-class resistance to NRTIs and PIs $(i=6)$, dual-class resistance to NNRTIs and PIs $(i=7)$ and triple-class resistance to NRTIs, NNRTIs and PIs $(i=8)$.

$$R_c^1 = \frac{c\left(\beta_1^H \theta_1 + \beta_1^P\left(\mu_0 + v_1\right)\right)}{\left(\mu_0 + \theta_1\right)\left(\mu_0 + v_1\right)} + \frac{c\theta_1 v_1\left(\beta_1^Y\left(\omega_1 + \gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^T\right) + \gamma_1 \beta_1^T \rho_1\right)}{\left(\mu_0 + \theta_1\right)\left(\mu_0 + v_1\right)\left(\mu_1^Y \omega_1 + \left(\gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^T\right)\left(\mu_1^Y + \rho_1\right)\right)},$$
(5)

$$R_c^2 = \frac{c\left(\beta_2^H \theta_2 + \beta_2^P\left(\mu_0 + v_2\right)\right)}{\left(\mu_0 + \theta_2\right)\left(\mu_0 + v_2\right)} + \frac{c\theta_2 v_2\left(\beta_2^Y\left(\omega_2 + \gamma_2 k_{25} + \gamma_2 k_{26} + \mu_2^T\right) + \gamma_2 \beta_2^T \rho_2\right)}{\left(\mu_0 + \theta_2\right)\left(\mu_0 + v_2\right)\left(\mu_2^Y \omega_2 + \left(\gamma_2 k_{25} + \gamma_2 k_{26} + \mu_2^T\right)\left(\mu_2^Y + \rho_2\right)\right)}$$
(6)

$$R_c^3 = \frac{c\left(\beta_3^H \theta_3 + \beta_3^P\left(\mu_0 + v_3\right)\right)}{\left(\mu_0 + \theta_3\right)\left(\mu_0 + v_3\right)} + \frac{c\theta_3 v_3\left(\beta_3^Y\left(\omega_3 + \gamma_3 k_{35} + \gamma_3 k_{37} + \mu_3^T\right) + \gamma_3 \beta_3^T \rho_3\right)}{\left(\mu_0 + \theta_3\right)\left(\mu_0 + v_3\right)\left(\mu_3^Y \omega_3 + \left(\gamma_3 k_{35} + \gamma_3 k_{37} + \mu_3^T\right)\left(\mu_3^Y + \rho_3\right)\right)}$$
(7)

$$R_c^4 = \frac{c\left(\beta_4^H \theta_4 + \beta_4^P\left(\mu_0 + v_4\right)\right)}{\left(\mu_0 + \theta_4\right)\left(\mu_0 + v_4\right)} + \frac{c\theta_4 v_4\left(\beta_4^Y\left(\omega_4 + \gamma_4 k_{46} + \gamma_4 k_{47} + \mu_4^T\right) + \gamma_4 \beta_4^T \rho_4\right)}{\left(\mu_0 + \theta_4\right)\left(\mu_0 + v_4\right)\left(\mu_4^Y \omega_4 + \left(\gamma_4 k_{46} + \gamma_4 k_{47} + \mu_4^T\right)\left(\mu_4^Y + \rho_4\right)\right)}$$
(8)

$$R_c^5 = \frac{c\left(\beta_5^H \theta_5 + \beta_5^P\left(\mu_0 + v_5\right)\right)}{\left(\mu_0 + \theta_5\right)\left(\mu_0 + v_5\right)} + \frac{c\theta_5 v_5\left(\beta_5^Y\left(\omega + \gamma_5 k_{58} + \mu_5^T\right) + \gamma_5 \beta_5^T \rho_5\right)}{\left(\mu_0 + \theta_5\right)\left(\mu_0 + v_5\right)\left(\mu_5^Y \omega_5 + \left(\gamma_5 k_{58} + \mu_5^T\right)\left(\mu_5^Y + \rho_5\right)\right)}$$
(9)

$$R_c^6 = \frac{c\left(\beta_6^H \theta_6 + \beta_6^P\left(\mu_0 + v_6\right)\right)}{\left(\mu_0 + \theta_6\right)\left(\mu_0 + v_6\right)} + \frac{c\theta_6 v_6\left(\beta_6^Y\left(\omega_6 + \gamma_6 k_{68} + \mu_6^T\right) + \gamma_6 \beta_6^T \rho_6\right)}{\left(\mu_0 + \theta_6\right)\left(\mu_0 + v_6\right)\left(\mu_6^Y \omega_6 + \left(\gamma_6 k_{68} + \mu_6^T\right)\left(\mu_6^Y + \rho_6\right)\right)}$$
(10)

$$R_c^7 = \frac{c\left(\beta_7^H \theta_7 + \beta_7^P\left(\mu_0 + v_7\right)\right)}{\left(\mu_0 + \theta_7\right)\left(\mu_0 + v_7\right)} + \frac{c\theta_7 v_7\left(\beta_7^Y\left(\omega_7 + \gamma_7 k_{78} + \mu_7^T\right) + \gamma_7 \beta_7^T \rho_7\right)}{\left(\mu_0 + \theta_7\right)\left(\mu_0 + v_7\right)\left(\mu_7^Y \omega_7 + \left(\gamma_7 k_{78} + \mu_7^T\right)\left(\mu_7^Y + \rho_7\right)\right)}$$
(11)

$$R_c^8 = \frac{c\left(\beta_8^H \theta_8 + \beta_8^P\left(\mu_0 + v_8\right)\right)}{\left(\mu_0 + \theta_8\right)\left(\mu_0 + v_8\right)} + \frac{c\theta_8 v_8\left(\beta_8^Y\left(\omega_8 + \mu_8^T\right) + \gamma_8 \beta_8^T \rho_8\right)}{\left(\mu_0 + \theta_8\right)\left(\mu_0 + v_8\right)\left(\mu_8^Y \omega_8 + \mu_8^T\left(\mu_8^Y + \rho_8\right)\right)}$$
(12)

The Control Reproduction Number (R_c) for the entire system is defined as $R_c = \max(R_c^i)$.

Each of the R_c's shown in Equations 5-12 can be rewritten in a similar fashion to R_0 in Equation 4 and thus be understood in biological terms. For example, R_c^1 can be rewritten as:

$$R_c^{\,1} = R_c^{\,P_1} + R_c^{\,H_1} + R_c^{\,Y_1} + R_c^{\,T_1}, \text{ where}$$

$$R_c^{\,P_1} = \frac{c\beta_1^{\,P}}{\mu_0 + \theta_1}$$

$$R_c^{\,H_1} = \frac{c\beta_1^{\,H}}{\mu_0 + v_1} \cdot \frac{\theta_1}{\mu_0 + \theta_1}$$

(13)

$$R_c^{\,Y_1} = \frac{c\beta_1^{\,Y}}{\mu_1^{\,Y} + \rho_1 \frac{\gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^{\,T}}{\omega_1 + \gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^{\,T}}} \cdot \frac{\theta_1}{\mu_0 + \theta_1} \cdot \frac{v_1}{\mu_0 + v_1}$$

$$R_c^{\,T_1} = \frac{c\gamma_1 \beta_1^{\,T}}{\gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^{\,T}} \cdot \frac{\theta_1}{\mu_0 + \theta_1} \cdot \frac{v_1}{\mu_0 + v_1} \cdot \frac{\rho_1 \frac{\gamma_1 k_{12} + \gamma_1 k_{13} + k_{14} + \mu_1^{\,T}}{\omega_1 + \gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^{\,T}}}{\mu_1^{\,Y} + \rho_1 \frac{\gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^{\,T}}{\omega_1 + \gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^{\,T}}}$$

The terms here are very similar to the ones in Equation 4; however, the duration of time an individual spends in state (Y_1) is more complex, due to the fact that an individual can go on and off treatment. The term $\rho_1 \frac{\gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^{\,T}}{\omega_1 + \gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^{\,T}}$ can be interpreted as the probability that an individual moves from state Y_1 to state T_1 times the probability that the individual does not return to state Y_1. It should be noted that if there is no backflow from state T_1 to state Y_1 (i.e. if $\omega_1 = 0$) then $\rho_1 \frac{\gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^{\,T}}{\omega_1 + \gamma_1 k_{12} + \gamma_1 k_{13} + \gamma_1 k_{14} + \mu_1^{\,T}} = \rho_1$.

We used the expressions in Equations 5 through 12 to calculate the values of the R_c's for the currently circulating wild-type and ARV-resistant strains in San Francisco. We made these calculations using the Monte Carlo filtered parameter sets, Latin Hypercube Sampling (*S33*) and pdfs for the model's treatment parameters given in Tables S2-S11.

We then classified the resistant strains in the network based on their transmission potential (Fig. 3B in the main text). We identified three groups: (i) where the R_c of the wild type and the ARV-resistant strain is less than one (blue data in Fig. 3B in main text) indicating the strains will eventually be eliminated; (ii) where the R_c of the wild-type strain is greater than one but the R_c of the resistant strain is less than one, indicating that the resistant strain is only being sustained through acquired and/or amplified resistance (green data in Fig. 3B in main text), and (iii) where both the R_c of the wild-type strain and the R_c of the resistant strain is greater than one, indicating that the resistant strain is self-sustaining (red data in Fig. 3B in main text).

This analysis also showed us that treatment has significantly reduced the severity of the HIV epidemic in the MSM community in San Francisco over the past twenty years. Before treatment, the value of the Reproduction Number (i.e., the Basic Reproduction Number, R_0) was 1.5 (median: IQR 1.2 to 2.0); after treatment was introduced, the Reproduction Number (i.e., the Control Reproduction Number, R_c) was reduced to 1.2 (median: IQR 1.0 to 1.6).

Section 5: Evaluating the impact of NNRTI-resistant strains in resource-constrained countries

Although information is limited, preliminary data suggest that treatment response and resistance patterns for subtype C are similar to subtype B (*S37-S42*).

In resource-constrained countries where ARVs are not widely available, the value of the Basic Reproduction Numbers (R_0's) of the wild-type strains and the value of the R_0's for the NNRTI-resistant strains will determine the effect of the competitive interaction among strains. If wild-type strains are more infectious than NNRTI-resistant strains, they will generate (on average) more secondary infections than resistant strains and will therefore out-compete them (Fig. S4).

However, in resource-constrained countries where ARVs are widely available (but second-line regimens are not) the outcome of the competitive interaction will depend upon the values of the R_0's of the NNRTI-resistant strains and the value of the R_c's of the wild-type strains. The NNRTI-resistant strains will out-compete the wild-type strains if their R_0's are greater than the R_c's of the wild-type strains. The more effective the treatment regimens and the greater the coverage level of ARVs, the lower the value of the R_c's of the wild-type strains; hence, the more likely the NNRTI-resistant strains will out-compete the wild-type strains. Fig. S5 shows the outcome of the competitive dynamics if a high proportion of infected individuals with a CD4 count < 350 cells/mL receive ARVs. Under these conditions, the majority of the NNRTI-resistant strains could out-compete the wild-type strains (Fig. S5). Many resource-constrained countries are striving to increase access to ARVs; recently, the World Health Organization has recommended that all HIV-infected individuals should receive ARVs (*S43*). If this occurs, the R_c's of wild-type strains are predicted to fall below one (*S43*), but the R_0's of the NNRTI-resistant strains will remain above one. Under these conditions, a paradoxical result is likely to occur in resource-constrained countries: instead of eliminating the HIV pandemic the NNRTI-resistant strains are likely to out-compete wild-type strains and cause self-sustaining epidemics.

Section 6: CART Methods and Results

CART is a nonlinear, nonparametric regression methodology that uses recursive partitioning and pruning to grow trees and produce accurate predictors for model dynamics (*S44*). It also provides insights into the structure of data sets and enables visualization of the key determinants of complex systems. To build trees, we used the model's estimated level of TDR for 2008 as the response variable and the model's 50 parameters as predictor variables. We classified simulations (i.e., data) into two groups based on the level of TDR that they generated for 2008: greater than 15% (high TDR) and less than 15% (low TDR). Starting from the root node of the tree, data were split into two daughter nodes by the predictor variables that produced the most internally homogenous nodes for levels of TDR in 2008. The nodes that are not split into daughter nodes are called terminal nodes and it is in these nodes that simulations are classified as being high or low TDR. Predictor selection and cutoff values (listed below each node being split) were determined by computing node purity for every potential predictor split. Node purity was calculated for every possible splitting value of every predictor using the entropy/impurity expression P*Log(P)+(1-P)Log(1-P), where P is the probability of a simulation resulting in high (>15%) levels of TDR. We determined the split that led to the greatest homogeneity of the daughter nodes by maximizing the reduction of impurity when going from the parent node to the daughter nodes.

We grew the tree by repeating this process until the potential splits had been exhausted (i.e., the tree was saturated) (*S44*). This process led to a large tree with terminal nodes containing very few simulations, so we pruned this saturated tree to achieve an optimal tree (Fig. S6). Pruning allowed us to simplify our model, increasing interpretability while sacrificing as little predictive power as possible (*S44*). The resulting tree revealed the most important variables (out of the 50 variables used in the analysis) for predicting, and therefore generating, TDR (Fig. 4 in the main text: red represents the number of simulations with TDR > 15% and blue represents the number of simulations with TDR < 15%). The optimal tree has 84% predictive power in correctly identifying which simulations will generate high levels of TDR and 82% predictive power in correctly identifying which simulations will generate low levels of TDR. Our pruned CART tree (Fig. 4 in the main text) shows the conditional hierarchical relationships among the six most important drivers; the unpruned tree would show the effects of the seventh driver and the other predictor variables.

The Importance Scores (IS) generated from our CART analysis (shown as a tree in Fig. 4 in the main text) are given in Table S13. The value of the IS depends upon the number of times that the predictor variable appears in the tree, as well as its position in the tree. Fifty predictor variables were included in the CART analysis; only predictor variables with an IS greater than 25 are listed. The fifty predictor variables were all of the model's parameters used in the Monte Carlo simulations. The value of the IS depends upon the number of times that the predictor variable appears in the tree, as well as its position in the tree. Values for the IS from the CART analysis are given in Table S13.

It should be noted that λ_5 (the relative transmissibility of strains with dual-class resistance to NRTIs and NNRTIs) is important in our analysis (IS = 40) but does not appear in the tree in Fig. 4 of the main text. This is because the presented tree is a pruned version of the optimal tree; all presented Importance Scores are calculated from the optimal tree.

As discussed in the main text, the most important drivers of TDR and their hierarchical relationships are shown in the optimal tree (Fig. 4 in main text). As the tree shows, the relationship between the drivers of TDR can be fairly complex. For example, λ_2 (the relative transmissibility of strains with single-class resistance to NRTIs) has an IS of 51. The importance of the value of λ_2 in driving the level of TDR depends on the values of other predictor variables. If $1/\nu$ < 6.1 years and α_1^H < 0.002, TDR is likely to be greater than 15% regardless of the value of λ_2 (Fig. 4 in main text). However, if $1/\nu$ < 6.1 and α_1^H > 0.002, TDR is likely to be greater than 15% if λ_2 > 0.51 but is likely to be less than 15% if λ_2 < 0.51 (Fig. 4 in main text). The predictor variable λ_2 can also be important in explaining high levels of TDR when $1/\nu$ > 6.1 years; however, it can be seen that now four different conditions must hold (Fig. 4 in main text).

Table S1: ARV treatments

Treatment	Era 1 1987-92	Era 2 1992-96	Era 3 1996-2001	Era 4 2001-08
ARV Class (medication name) [date of FDA approval]				
nRTI (zidovudine) [03/19/1987]	X	X	X	X
nRTI (didanosine) [09/10/1991]		X	X	X
nRTI (zalcitabine) [06/19/1992]		X	X	X
nRTI (stavudine) [06/24/1994]		X	X	X
nRTI (lamivudine) [11/17/1995]		X	X	X
PI (saquinavir mesylate) [12/06/1995]			X	X
PI (ritonavir) [03/01/1996]			X	X
PI (indinavir) [03/13/1996]			X	X
nnRTI (nevirapine) [06/21/1996]			X	X
PI (nelfinavir mesylate) [04/14/1997]			X	X
nnRTI (delavirdine) [04/21/1997]			X	X
nRTI (zidovudine+lamivudine) [09/27/1997]			X	X
nnRTI (efavirenz) [09/17/1998]			X	X
nRTI (abacavir) [12/17/1998]			X	X
PI (amprenavir) [04/15/1999]			X	X
PI (lopinavir+ritonavir) [09/15/2000]			X	X
nRTI (zidovudine+lamivudine+abacavir) [11/14/2000]			X	X
nRTI (tenofovir) [10/26/2001]			X	X
Fusion Inhibitor (enfuvirtide) [03/13/2003]#				X
PI (atazanavir) [06/20/2003]				X
nRTI (emtricitabine) [07/02/2003]				X
PI (fosampreavir calcium) [10/20/2003]				X
nRTI (abacavir+lamivudine) [08/02/2004]				X
nRTI (tenofovir+emtricitabine) [08/02/2004]				X
PI (tipranavir) [06/22/2005]				X
PI (darunavir) [06/23/2006]				X
nRTI+nnRTI (efavirenz+emtricitabine+ tenofovir) [07/12/2006]				X
CCR5 antagonist (maraviroc) [06/08/2007]#				X
Integrase Inhibitor (raltegravir) [12/10/2007]#				X
nnRTI (etravirine) [01/18/2008]				X

not in the present model

Table S2: Viral load and transmission probabilities for individuals infected with wild-type strains. Viral load is high during primary infection, low after primary infection (when the CD4 count is greater than 350 cells/microL) and high once the CD4 count drops below 350 cells/microL. During monotherapy, when the available drugs were not as effective as current regimens, the average viral load on treatment was modeled as being the same as it was for untreated individuals with CD4 counts less than or equal to 350 cells/microL. Viral load ranges were estimated from empirical studies (*S27, S32, S45-S50*). Estimates for average viral transmissibility per partnership (β_1^j) were calculated as described in Section 2.

Infection stage	Average viral load (v) (copies/mL)	Average viral transmissibility per act (α_1^j)	Average viral transmissibility per partnership (β_1^j)
P_1 Primary infection	Uniform pdf 50,000-5,000,000	Uniform pdf 0.0031-0.0185	0.016-0.659 median 0.235
H_1 Infected (after primary infection, CD4 count > 350 cells/microL)	Triangular pdf 10,000-50,000 (peak 20,000)	0.0017-0.0031	0.008-0.164 median 0.049
Y_1 Infected, CD4 count ≤ 350 cells/microL	Triangular pdf 35,000-100,000 (peak 60,000)	0.0027-0.0040	0.015-0.204 median 0.069
T_1 On treatment (Era 1)	Triangular pdf 35,000-100,000 (peak 60,000)	0.0027-0.0040	0.015-0.204 median 0.069

Table S3: Resistance categories in the amplification cascade model. Resistant strains in the network are grouped into seven categories which are defined based upon their level of resistance (single, dual or triple class) and the specific class of drugs they are resistant to (Nucleoside Reverse Transcriptase Inhibitors (NRTIs), non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) and Protease Inhibitors (PIs)). There is *in-vitro* and *in-vivo* evidence that transmitted resistant strains persist and that the resistant strains will influence the laboratory response to antiretroviral treatment (*S15, S51-S55*).

Resistance Category	Per-act transmissibility relative to wild-type strains (λ_i)
Single-class resistance to NRTIs	0.40 – 0.85 (peak 0.6)
Single-class resistance to NNRTIs	0.70 – 0.95 (peak 0.9)
Single-class resistance to PIs	0.30 – 0.90 (peak 0.6)
Dual-class resistance to NRTIs and NNRTIs	0.40 – 0.95 (peak 0.7)
Dual-class resistance to NRTIs and PIs	0.20 – 0.65 (peak 0.35)
Dual-class resistance to NNRTIs and PIs	0.40 – 0.95 (peak 0.55)
Triple-class resistance to NRTIs, NNRTIs and PIs	0.10 – 0.20 (peak 0.1)

Table S4: Average rates of individuals going on and off treatment in each of the four treatment eras. ρ_i is the model parameter representing the average rate at which infected individuals with CD4 ≤ 350 cells/microL begin treatment per year. This parameter was calculated from the percentage of infected individuals with CD4 ≤ 350 cells/microL who begin treatment each year (i.e., from $100 \cdot (1 - e^{-\rho_i})$). The average rate at which treated individuals stop treatment per year, ω_i, was calculated in a similar manner.

Parameter	Definition	Era	Range
$100 \cdot (1 - e^{-\rho_i})$	% of infected individuals with CD4 ≤ 350 cells/microL who begin treatment per year	1	40-60%
		2	50-70%
		3,4	60-80%
$100 \cdot (1 - e^{-\omega_i})$	% on treatment who stop treatment per year	1-4	0-15%

Table S5: Viral load and transmission probabilities while on treatment in eras 2, 3 and 4.
Estimates for average viral transmissibility per partnership (β_1^T) were calculated as described in Section 2. Viral loads are not treated individuals who are only partially suppressed.

Infection stage	Average viral load while in stage (v) (copies/mL)	Average viral transmissibility per act (α_1^T)	Average viral transmissibility per partnership (β_1^T)
T_1 On treatment (Era 2)	Not sampled	Uniform 0.0013-0.0026	0.007-0.140 median 0.039
T_1 On treatment (Era 3)	Not sampled	Uniform 0.0011-0.0023	0.006-0.123 median 0.034
T_1 On treatment (Era 4)	Triangular 75-20,000 (peak 5,500)	0.0003-0.0022 median 0.0015	0.002-0.112 median 0.030

Table S6: Treatment parameters during the era of monotherapy (1987-1991) Estimates of clinical parameters were obtained from reference (*S32*).

Strains causing infection	Drug classes included in treatment regimen	% of patients on regimen	% of patients virally suppressed $100 \cdot (1 - \gamma_i)$	% of patients not virally suppressed who develop resistance, $100 \cdot (1 - e^{k_{ii'}})$
wild-type	1 NRTI (AZT)	100%	0-5%	to NRTIs: 90-100%

Table S7: Treatment parameters during the era of dual therapy (1992-1995) During this era, patients could have been treated with several different regimens, so the percentage of patients on each regimen (column 3) was variable(*S56, S57*). Effectively virally suppressed individuals did not develop resistance (*S58-S60*). Once on a specific treatment regimen, a certain percentage of patients developed resistance to different classes of ARVs (*S61, S62*).

Strains causing infection	Drug classes included in treatment regimen	% of patients on regimen	% of patients virally suppressed $100 \cdot (1-\gamma_i)$	% of patients not virally suppressed who develop resistance, $100 \cdot (1-e^{k_{ii'}})$
wild-type	2 NRTIs	50-70%	30-60%	to NRTIs: 70-100%
	1 NRTI	30-50%	30-60%	to NRTIs: 90-100%
single-class resistance to NRTI	2 NRTIs	50-70%	5-20%	to NRTIs: 70-100%
	1 NRTI	30-50%	5-20%	to NRTIs: 90-100%

Table S8: Treatment parameters during the era of triple therapy (1996-2000) Early HAART During this era, patients could have been treated with several different regimens, so the percentage of patients on each regimen (column 3) was variable(*S56, S57*). Effectively virally suppressed individuals did not develop resistance (*S58-S60*). Once on a specific treatment regimen, a certain percentage of patients developed resistance to different classes of ARVs (*S61, S62*).

Strains causing infection	Drug classes included in treatment regimen	% of patients on regimen	% of patients virally suppressed $100 \cdot (1-\gamma_i)$	% of patients not virally suppressed who develop resistance, $100 \cdot (1-e^{k_{ii'}})$
wild-type	2 NRTIs and 1 NNRTI	60%	55-75%	to NRTIs: 15-30%
				to NNRTIs: 60-80%
	2 NRTIs and 1 PI	40%	55-75%	to NRTIs: 15-30%
				to PIs: 15-30%
single-class resistance to NRTIs	2 other NRTIs and 1 NNRTI	30%	50-75%	to NNRTIs: 60-80%
	2 other NRTIs and 1 PI	60%	50-75%	to PIs: 30-40%
	1 NNRTI and 1 PI	10%	50-75%	to NNRTIs: 15-40%
				to PIs: 15-30%
single-class resistance to NNRTIs	2 NRTIs and 1 PI	100%	55-75%	to NRTIs: 15-30%
				to PIs: 15-30%
single-class resistance to PIs	2 NRTIs and a NNRTI	50%	40%-75%	to NRTIs: 15-40%
				to NNRTIs: 10-25%
	2 NRTIs and different PIs	50%	30-60%	to NRTIs: 30-60%
dual-class resistance to NRTIs and NNRTIs	2 NRTIs and 1 PI	100%	0-5%	to PIs: 15-35% (peak 30%)
dual-class resistance	2 NRTIs, PI and NNRTI	100%	0-5%	to NNRTIs: 20-40% (peak 30%)

to NRTIs and PIs				
dual-class resistance to NNRTIs and PIs	2 NRTIs, PIs and NNRTI	100%	0-1%	to NRTIs: 40-60% (peak 60%)
triple-class resistance	1 NRTI, 1 PI, 1 NNRTI	100%	0-1%	Not Applicable

Table S9: Treatment parameters during the era of Triple Therapy (2000-2008) Modern Era HAART During this era, patients could have been treated with several different regimens, so the percentage of patients on each regimen (column 3) was variable(*S56, S57*). Effectively virally suppressed individuals did not develop resistance (*S58-S60*). Once on a specific treatment regimen, a certain percentage of patients developed resistance to different classes of ARVs (*S61, S62*).

Strains causing infection	Drug classes included in treatment regimen	% of patients on regimen	% of patients virally suppressed $100 \cdot (1 - \gamma_i)$	% of patients not virally suppressed who develop resistance, $100 \cdot (1 - e^{k_{ii'}})$
wild-type	2 NRTIs and 1 NNRTI	50%	65-85%	to NRTIs: 15-30%
				to NNRTIs: 60-80%
	2 NRTIs and boosted or dual PIs	50%	70-95%	to NRTIs: 15-30%
				to PIs: 5-25%
single-class resistance to NRTIs	2 other NRTIs and 1 NNRTI	30%	50-75%	to NNRTIs: 40-60%
	2 other NRTIs and boosted or dual PI	60%	60-80%	to PIs: 30-40%
	1 NNRTI and boosted or dual PI	10%	50-75%	to NNRTIs: 15-40%
				to PIs: 15-30%
single-class resistance to NNRTIs	2 NRTIs and boosted or dual PI	100%	65-85%	to NRTIs: 15-30%
			70-95%	to PIs: 5-25%
single-class resistance to PIs	2 NRTIs and a NNRTI	50%	65-85%	to NRTIs: 15-40%
				to NNRTIs: 10-25%
	2 NRTIs and different PIs in boosted or dual therapy	50%	40-70%	to NRTIs: 30-60%
dual-class resistance to NRTIs and NNRTIs	2 different NRTIs and boosted or dual PI	100%	15-30%	to PIs: 15-35% (peak 30%)

dual-class resistance to NRTIs and PIs	2 NRTIs, boosted or dual PI and NNRTI	100%	5-20%	to NNRTIs: 20-40% (peak 30%)
dual-class resistance to NNRTIs and PIs	2 NRTIs, boosted or dual PIs and NNRTI	100%	5-20%	to NRTIs: 40-60% (peak 60%)
triple-class resistance	1 NRTI, 1 NNRTI and 1PI	100%	0-1%	Not Applicable

Table S10: Parameter ranges for the average time spent in different stages of the amplification cascade model The additional survival time was estimated using reference (*S31*).

Symbol	Definition	Era 1 Monotherapy	Era 2 Dual therapy	Era 3 Early HAART	Era 4 Late HAART
$1/\mu_0$	Average time (in years) spent acquiring new sex partners	25-40 (peak=30)	25-40 (peak=30)	25-40 (peak=30)	25-40 (peak=30)
$1/\theta$	Average time (in days) of primary infection	15-55 days	15-55 days	15-55 days	15-55 days
$1/\nu$	Average time (in years) before becoming treatment eligible (i.e., with a CD4 count > 350 cells/microL)	5-7	5-7	5-7	5-7
$1/\mu_i^Y$	Average time (in years) with a CD4 count ≤ 350 cells/microL	1-4	1-4	3-7	3-7
$1/\mu^a$	Average additional time (in years) gained through treatment for individuals infected with wild-type strains	1-3	1-4	4-10	8-15
$1/\mu^b$	Average additional time (in years) gained through treatment for individuals infected with strains with single-class resistance	0.04-1	0.04-1	2-5	2-5
$1/\mu^c$	Average additional time (in years) gained through treatment for individuals infected with strains with dual-class resistance	N/A	N/A	2-6	2-6
$1/\mu^d$	Average additional time (in years) gained through treatment for individuals infected with strains with triple-class resistance	N/A	N/A	0.04-1	0.04-1

Table S11: Average time (in years) spent in the treated stage stratified by era and resistance category The average time spend in the treated stage was calculated as described in Section 2.

| Treatment Era	Strains causing infection	First Quartile	Median	Third Quartile
Era 1	wild type	3.0	3.7	4.4
	single-class resistance to NRTIs	1.4	2.0	2.7
Era 2	wild type	5.5	6.5	7.5
	single-class resistance to NRTIs	3.5	4.3	5.2
Era 3	wild type	9.6	11.1	12.7
	single-class resistance to NRTIs	6.6	7.6	8.6
	single-class resistance to NNRTIs	6.6	7.6	8.6
	single-class resistance to PIs	6.6	7.6	8.6
	dual-class resistance to NRTIs and NNRTIs	7.0	8.2	9.4
	dual-class resistance to NRTIs and PIs	7.1	8.2	9.4
	dual-class resistance to NNRTIs and PIs	7.1	8.2	9.3
	Triple-class resistance	3.7	4.6	5.5
Era 4	wild type	13.8	15.5	17.3
	single-class resistance to NRTIs	6.5	7.5	8.6
	single-class resistance to NNRTIs	6.5	7.5	8.5
	single-class resistance to PIs	6.6	7.6	8.5
	dual-class resistance to NRTIs and NNRTIs	7.0	8.1	9.3
	dual-class resistance to NRTIs and PIs	7.0	8.2	9.3
	dual-class resistance to NNRTIs and PIs	7.1	8.2	9.3
	Triple-class resistance	3.7	4.6	5.5

Table S12: Parameters with distributions that changed significantly after Monte Carlo filtering

	Data	1Q	Median	Mean	3Q
c	Filtered	2.1	2.7	2.8	3.4
	Unfiltered	1.5	2.2	2.3	3.0
n	Filtered	62.8	78.4	84.6	101.8
	Unfiltered	24.9	43.5	54.6	73.4
α_1^H	Filtered	0.00218	0.00240	0.00241	0.00263
	Unfiltered	0.00216	0.00238	0.00239	0.00262
$1/\mu_i^Y$ (Era 1)	Filtered	1.9	2.7	2.6	3.4
	Unfiltered	1.8	2.5	2.5	3.3

Table S13: Importance Scores (IS) calculated from CART analysis

Predictor Variable	Definition	IS
$1/\nu$	average rate (at the population level) at which CD4 cell counts fall to below 350 cells/microL	100
α_1^H	per-act probability that individuals (with CD4 counts > 350 cells/microL) transmit wild-type strains	73
α_1^T	per-act probability that treated individuals transmit wild-type strains	60
λ_2	relative transmissibility (per act) of strains with single-class resistance to NRTIs	51
γ_1	% of patients (infected with wild-type strains) not virally suppressed	45
λ_5	relative transmissibility (per act) of strains with dual-class resistance to NRTIs and NNRTIs	40
γ_3	% of patients (infected with strains with single-class resistance to NNRTIs) not virally suppressed	39
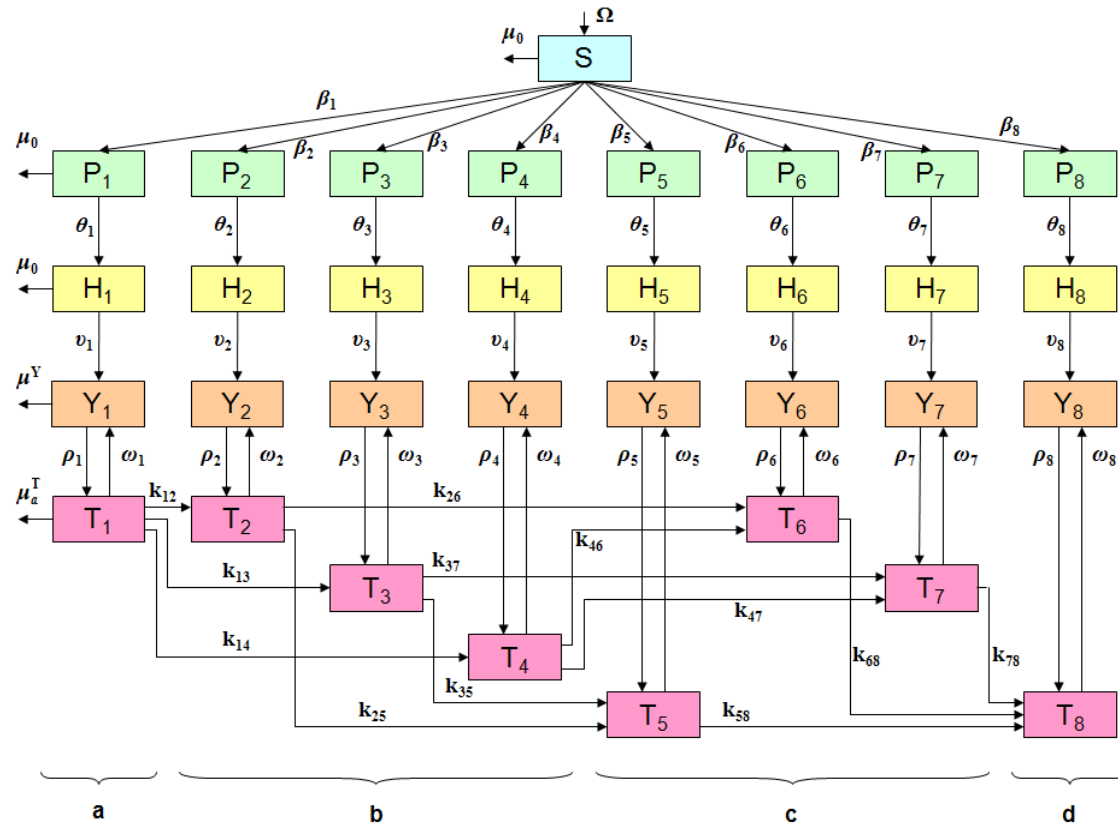
Fig. S1: The full 33 equation model, showing susceptible individuals (S), individuals with primary infection (P), individuals who are not yet eligible for treatment (i.e., with CD4 counts > 350 cells/microL) (H), individuals who are treatment eligible (i.e., with CD4 ≤ 350 cells/microL) but not on treatment (Y) and individuals receiving treatment (T). Resistant strains in the network are grouped into seven categories which are defined based upon their level of resistance (single, dual or triple class) and the specific class of drugs they are resistant to (Nucleoside Reverse Transcriptase Inhibitors (NRTIs), non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) and Protease Inhibitors (PIs)). Categories are denoted by subscript i: wild-type strains $(i=1)$, single-class resistance to NRTIs $(i=2)$, single-class resistance to NNRTIs $(i=3)$, single-class resistance to PIs $(i=4)$, dual-class resistance to NRTIs and NNRTIs $(i=5)$, dual-class resistance to NRTIs and PIs $(i=6)$, dual-class resistance to NNRTIs and PIs $(i=7)$ and triple-class resistance to NRTIs, NNRTIs and PIs $(i=8)$.
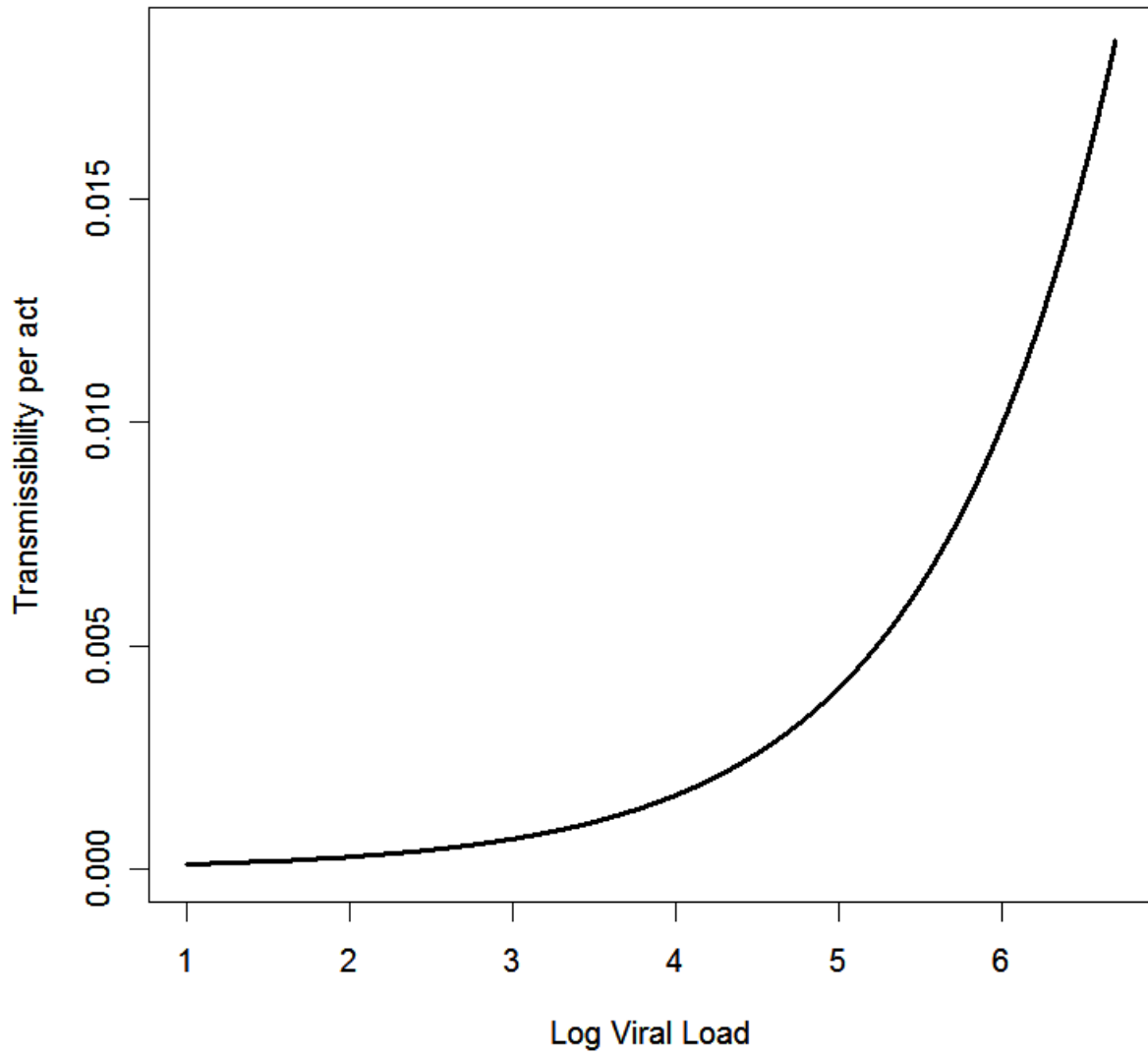
Fig. S2: The relationship between transmissibility per act and viral load (on a log scale).
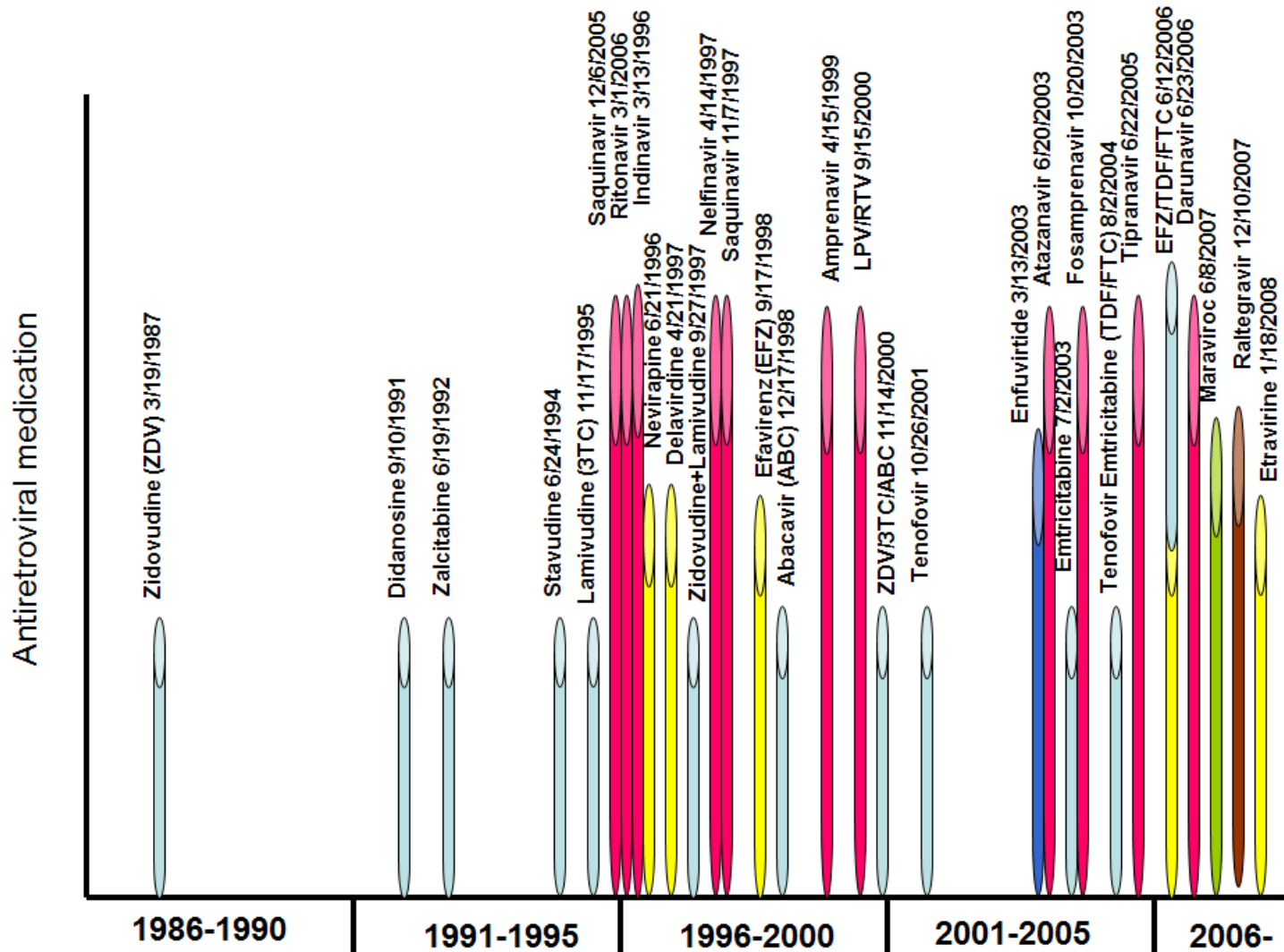
Fig. S3: Detailed timeline of drug introduction in San Francisco during the four eras of treatment. Color is used to indicate drug classes: light blue/gray for NRTIs, yellow for NNRTIs, red for PIs, blue for fusion inhibitors, green for entry (CCR5 receptor) inhibitors and brown for integrase inhibitors.
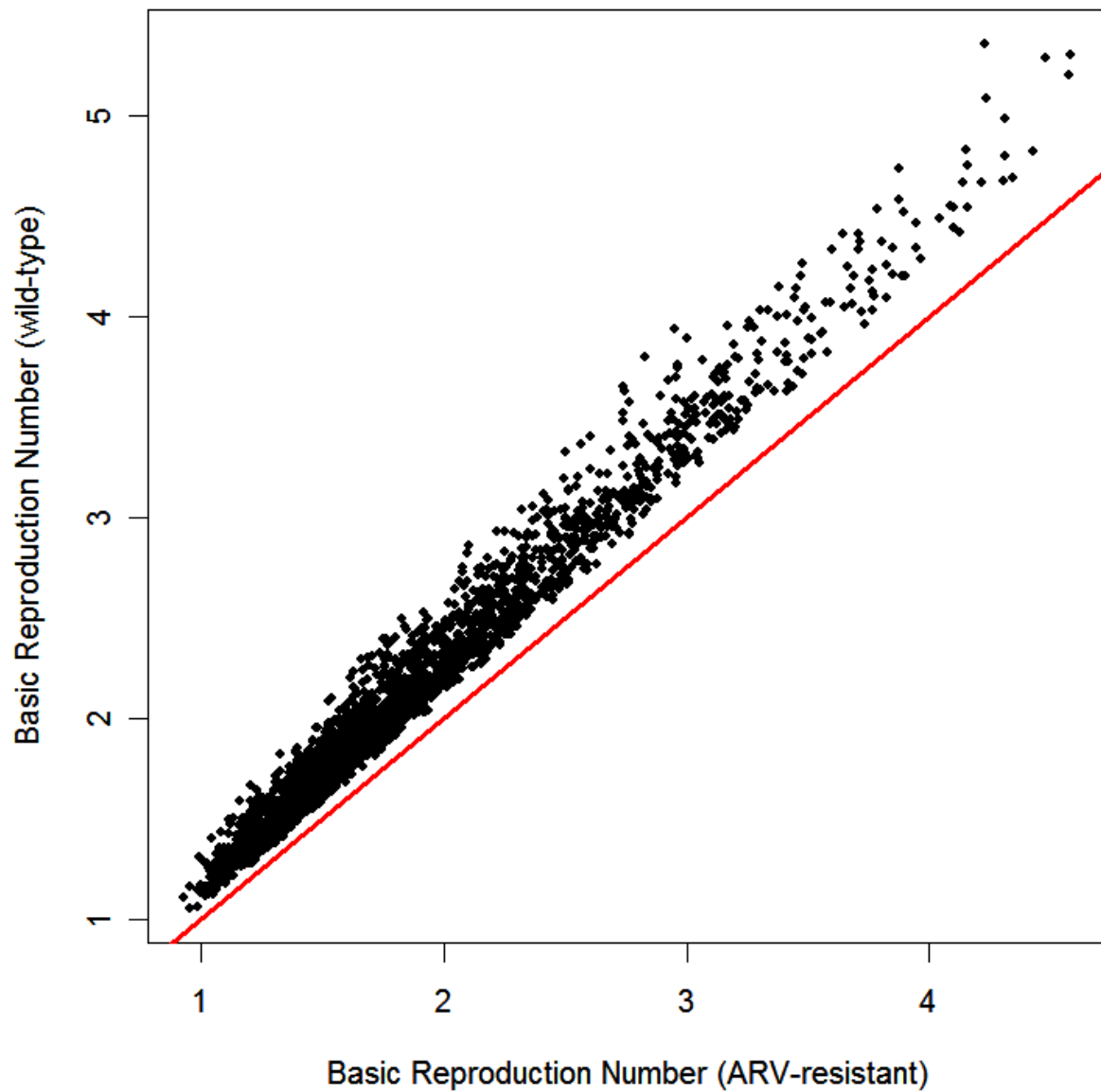
Fig. S4: The Basic Reproduction Number of wild-type strains relative to those of NNRTI-resistant strains that have the transmission potential to cause self-sustaining epidemics.
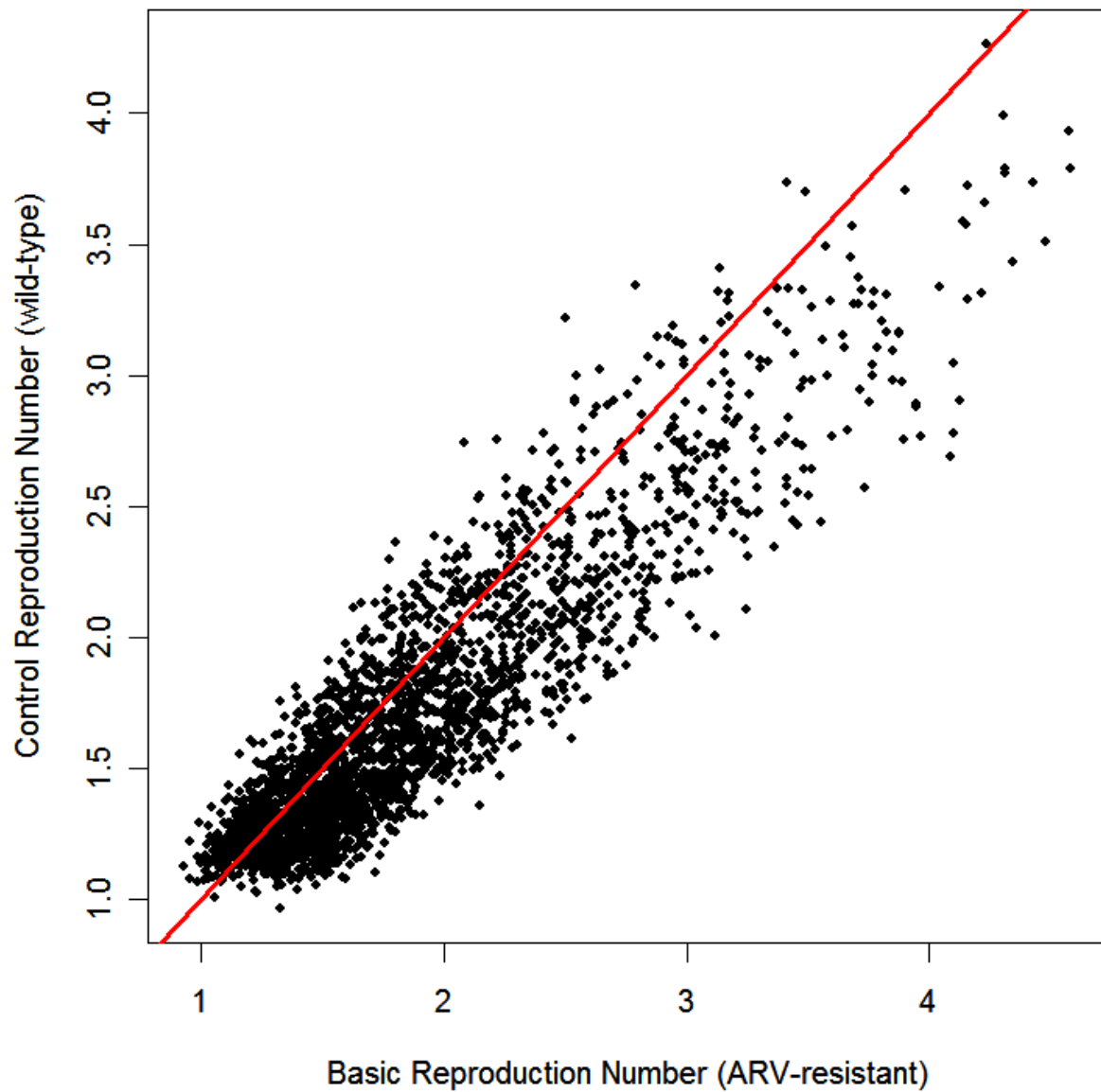
Fig. S5: The Control Reproduction Number of wild-type strains relative to the Basic Reproduction Number of NNRTI-resistant strains that have the transmission potential to cause self-sustaining epidemics.
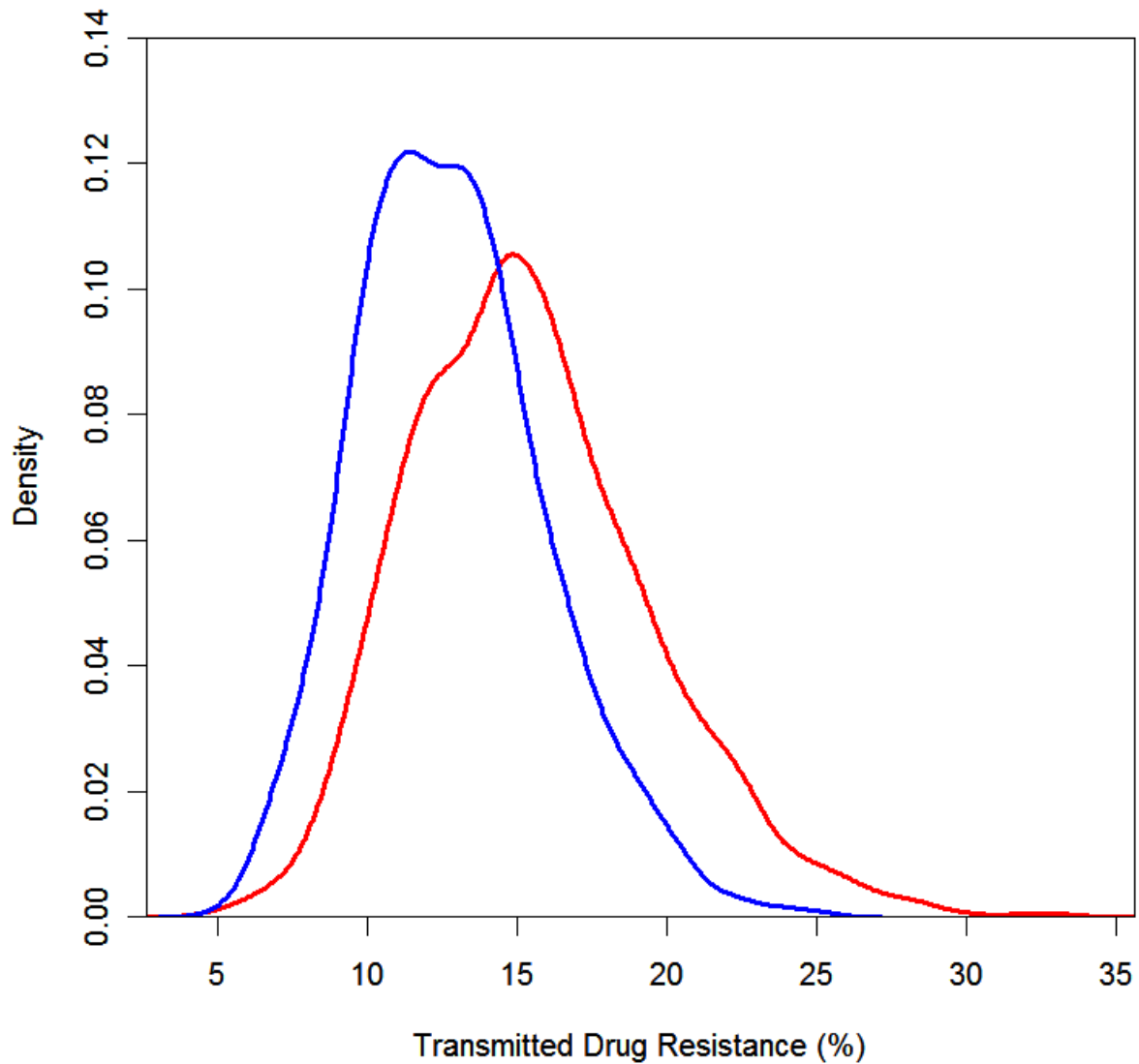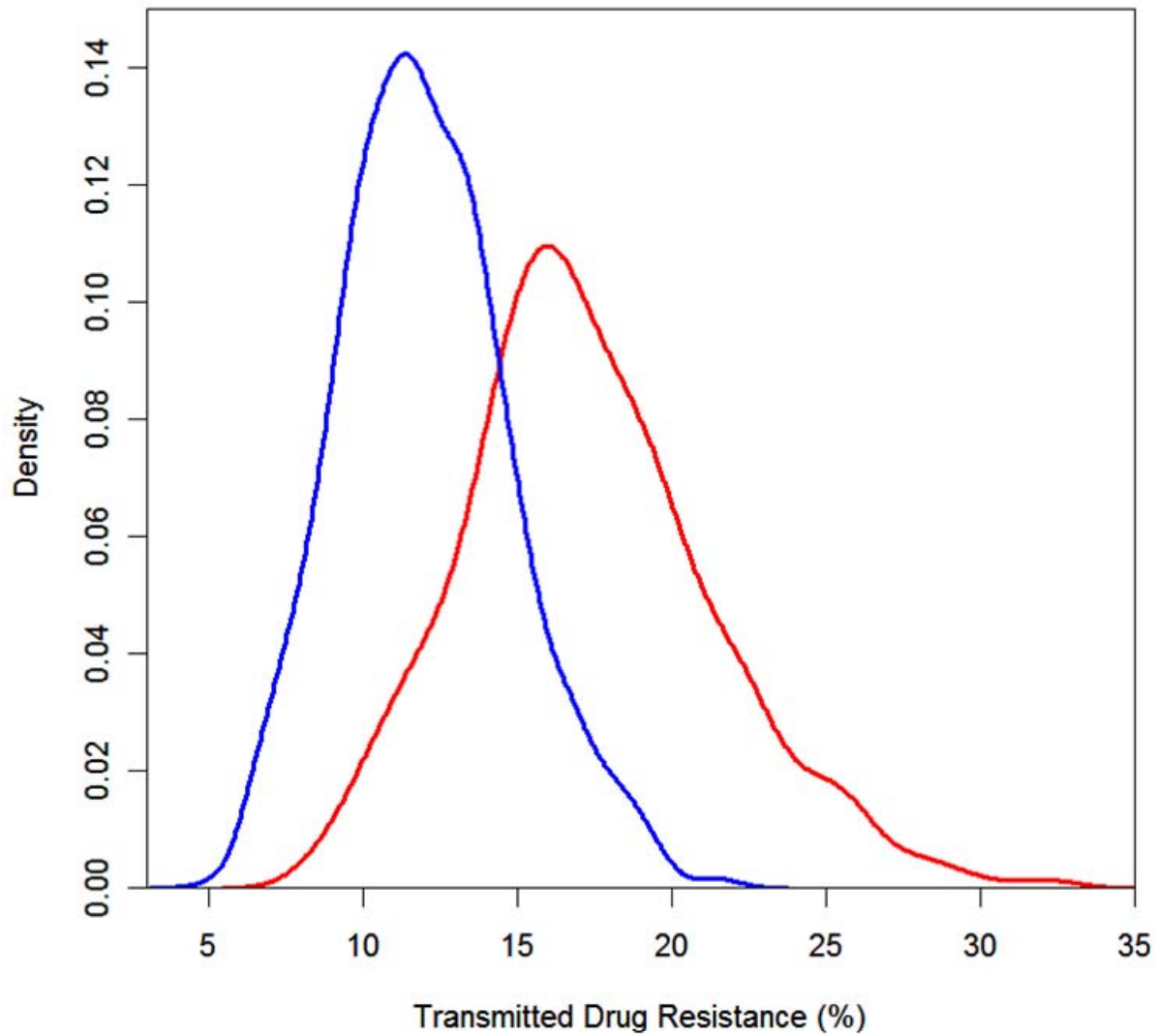
Fig. S6: (A) Density plots illustrate how the most important variable (v^{-1}, the average time, at the population level, it takes for CD4 cell counts in infected individuals to fall below 350 cells/microL) affects levels of TDR in 2008. Red data show results from our CART analysis from simulations where $v^{-1} \leq$ 6.1 years; blue data show results from simulations where $v^{-1} >$ 6.1 years.
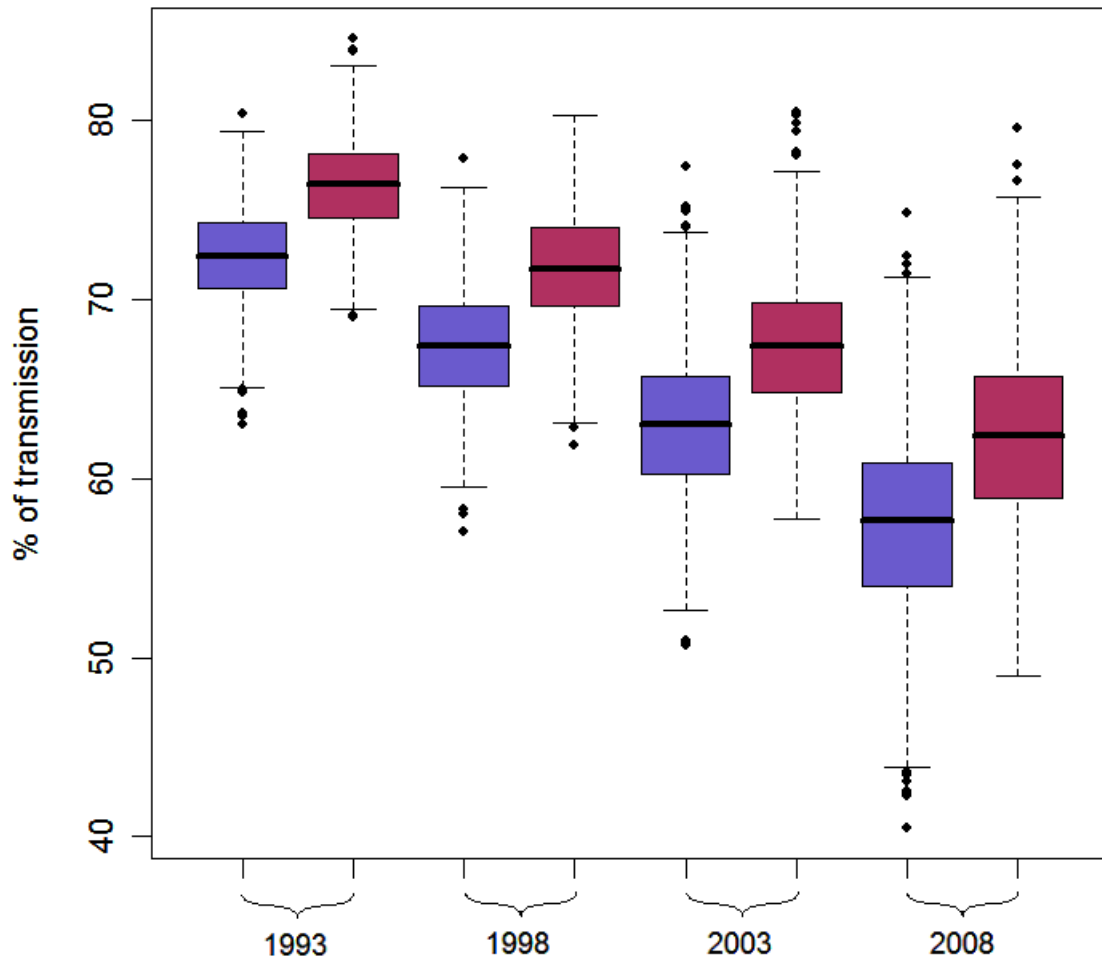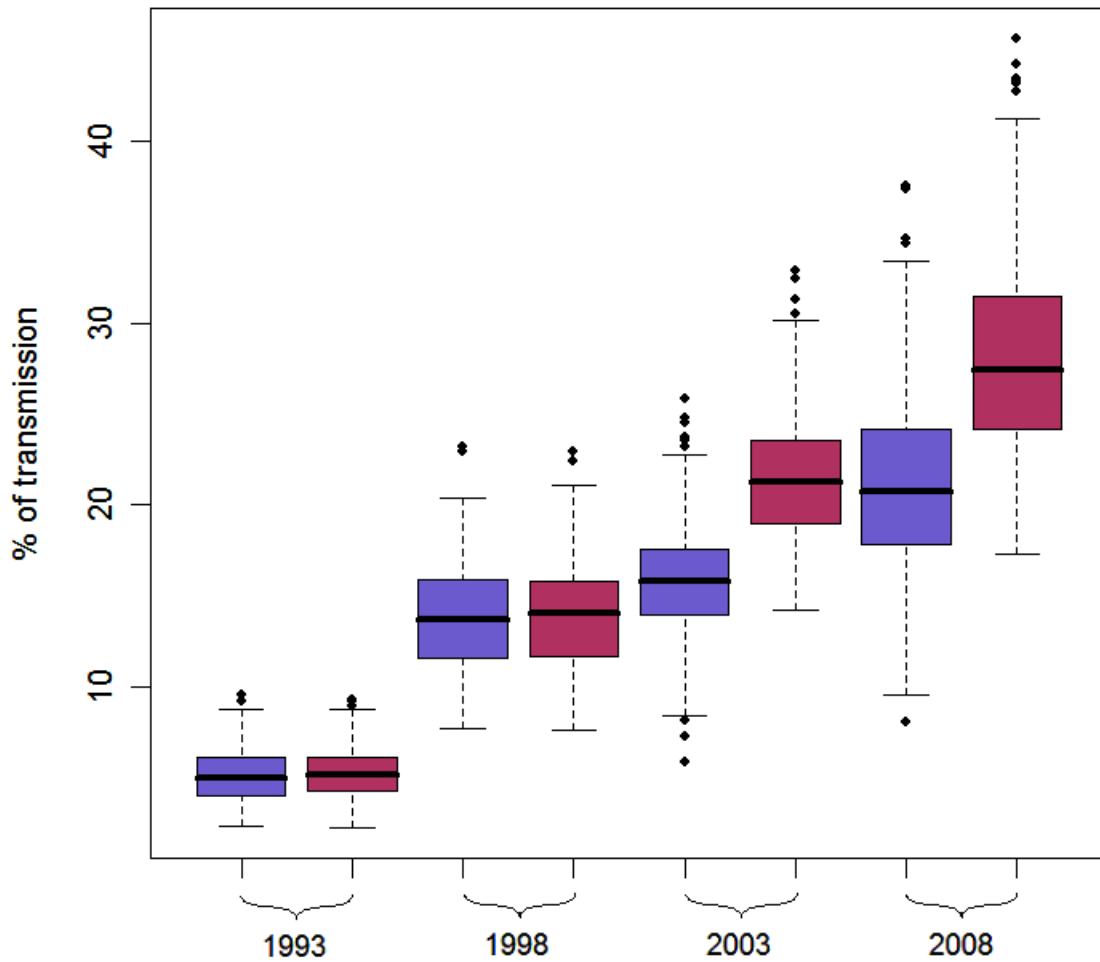
Fig. S6: (B) Boxplots based on the Monte Carlo simulations from the uncertainty analysis showing the contribution to transmission of individuals infected with wild-type strains and with a CD4 count > 350 cells/microL. All data are from simulations in the CART analysis with $v^{-1} \leq 6.1$ years (left side of the tree in Fig. 4 in the main text); however, a similar effect is seen for simulations with $v^{-1} > 6.1$ years (right side of the tree in Fig. 4 in the main text). Pink data show results from simulations where $\alpha_1^H > 0.0024$; blue data show results from simulations where $\alpha_1^H \leq 0.0024$. The variable α_1^H reflects the degree of infectivity of wild-type strains during the period when the infected individual has a CD4 count > 350 cells/microL; infectivity is specified as the probability of transmitting HIV during one sex act.

Fig. S6: **(C)** Density plots illustrate how hierarchical relationships among the two most important predictor variables determine the level of TDR in 2008. High levels of TDR (red data; median 17%) are likely if $v^{-1} \leq 6.1$ years and $\alpha_1^H \leq 0.0024$. If $v^{-1} > 6.1$ years and $\alpha_1^H > 0.0023$, then low levels of TDR (blue data; median 12%) are likely.

Fig. S6: **(D)** Boxplots based on the Monte Carlo simulations from the uncertainty analysis show the contribution to transmission of individuals who are infected with wild-type strains and on treatment. All data are from simulations in the CART analysis that had a $v^{-1} > 6.1$ years and $\alpha_1^H > 0.0023$. Pink data show results from simulations where $\alpha_1^T > 0.0015$; blue data shows results from simulations where $\alpha_1^T \leq 0.0015$. The variable α_1^T represents the probability that an individual receiving ARVs during the current era of Modern HAART transmits HIV during one sex act.

References

S1.   S. M. Blower, H. B. Gershengorn, R. M. Grant, *Science* **287**, 650 (2000).
S2.   J. Goudsmit *et al.*, *AIDS* **15**, 2293 (2001).
S3.   A. Phillips, *Nat Med* **7**, 993 (2001).
S4.   E. Tchetgen, E. H. Kaplan, G. H. Friedland, *J Acquir Immune Defic Syndr* **26**, 118 (2001).
S5.   R. Vardavas, S. Blower, *PLoS ONE* **2**, e152 (2007).
S6.   J. X. Velasco-Hernandez, H. B. Gershengorn, S. M. Blower, *Lancet Infect Dis* **2**, 487 (2002).
S7.   J. Wu, P. Yan, C. Archibald, *BMC Public Health* **7**, 300 (2007).
S8.   G. S. Zaric, A. Bayoumi, M. L. Brandeau, D. K. Owens, *Simulation* **71**, 262 (1998).
S9.   R. J. Smith, S. M. Blower, *Lancet Infect Dis* **4**, 636 (2004).
S10.   T. C. Quinn *et al.*, *N Engl J Med* **342**, 921 (2000).
S11.   R. H. Gray *et al.*, *Lancet* **357**, 1149 (2001).
S12.   D. R. Bangsberg *et al.*, *AIDS* **20**, 223 (2006).
S13.   J. D. Barbour *et al.*, *J Infect Dis* **190**, 251 (2004).
S14.   J. D. Barbour *et al.*, paper presented at the 12th International HIV Drug Resistance Workshop, Los Cabos, Mexico, 2003.
S15.   C. L. Booth, A. M. Geretti, *J Antimicrob Chemother* **59**, 1047 (2007).
S16.   M. S. Hirsch *et al.*, *Clin Infect Dis* **47**, 266 (2008).
S17.   J. Martinez-Picado, M. A. Martinez, *Virus Res* **134**, 104 (2008).
S18.   W. Lang *et al.*, *JAMA* **257**, 326 (1987).
S19.   W. Winkelstein, Jr. *et al.*, *JAMA* **257**, 321 (1987).
S20.   W. Winkelstein, Jr. *et al.*, *Am J Public Health* **77**, 685 (1987).
S21.   B. A. Larder, G. Darby, D. D. Richman, *Science* **243**, 1731 (1989).
S22.   J. C. Schmit *et al.*, *AIDS* **12**, 2007 (1998).
S23.   S. D. Holmberg *et al.*, *AIDS* **7**, 699 (1993).
S24.   W. Lang, D. Osmond, K. Page-Bodkin, A. Moss, W. Winkelstein, Jr., *J Acquir Immune Defic Syndr* **6**, 191 (1993).
S25.   W. Lang *et al.*, *J Acquir Immune Defic Syndr* **4**, 713 (1991).
S26.   P. S. Rosenberg *et al.*, *J Acquir Immune Defic Syndr* **4**, 392 (1991).
S27.   J. J. Eron *et al.*, *N Engl J Med* **333**, 1662 (1995).
S28.   C. Katlama *et al.*, *JAMA* **276**, 118 (1996).
S29.   D. R. Kuritzkes *et al.*, *AIDS* **13**, 685 (1999).
S30.   D. R. Kuritzkes *et al.*, *AIDS* **10**, 975 (1996).
S31.   V. D. Lima *et al.*, *AIDS* **21**, 685 (2007).
S32.   J. O. Kahn, B. D. Walker, *N Engl J Med* **339**, 33 (1998).
S33.   S. Blower, H. Dowlatabadi, *International Statistical Review* **2**, 229 (1994).
S34.   M. A. Sanchez, S. M. Blower, *Am J Epidemiol* **145**, 1127 (1997).
S35.   P. van den Driessche, J. Watmough, *Math Biosci* **180**, 29 (2002).
S36.   O. Diekmann, J. A. Heesterbeek, J. A. Metz, *J Math Biol* **28**, 365 (1990).
S37.   C. Orrell *et al.*, *Antivir Ther* **14**, 523 (2009).
S38.   M. C. Hosseinipour *et al.*, *AIDS* **23**, 1127 (2009).
S39.   A. Jlizi *et al.*, *Arch Virol* **153**, 1103 (2008).
S40.   V. Pillay *et al.*, *AIDS Res Hum Retroviruses* **24**, 1449 (2008).
S41.   A. Soria *et al.*, *Antivir Ther* **14**, 339 (2009).
S42.   A. J. Kandathil *et al.*, *Indian J Med Microbiol* **27**, 231 (2009).
S43.   R. M. Granich, C. F. Gilks, C. Dye, K. M. De Cock, B. G. Williams, *Lancet* **373**, 48 (2009).

S44.   L. Breiman, J. H. Freidman, R. A. Olshen, C. J. Stone, *Classification and Regression Trees* (Chapman & Hall, Boca Raton, 1984).

S45.   R. M. Gulick *et al.*, *N Engl J Med* **350**, 1850 (2004).

S46.   S. M. Hammer *et al.*, *N Engl J Med* **335**, 1081 (1996).

S47.   S. M. Hammer *et al.*, *N Engl J Med* **337**, 725 (1997).

S48.   D. A. Katzenstein *et al.*, *N Engl J Med* **335**, 1091 (1996).

S49.   J. S. Montaner *et al.*, *JAMA* **279**, 930 (1998).

S50.   F. J. Palella, Jr. *et al.*, *N Engl J Med* **338**, 853 (1998).

S51.   H. L. Devereux, V. C. Emery, M. A. Johnson, C. Loveday, *J Med Virol* **65**, 218 (2001).

S52.   J. A. Johnson *et al.*, *PLoS Med* **5**, e158 (2008).

S53.   C. E. Koval, C. Dykes, J. Wang, L. M. Demeter, *Virology* **353**, 184 (2006).

S54.   G. Bleiber, M. Munoz, A. Ciuffi, P. Meylan, A. Telenti, *J Virol* **75**, 3291 (2001).

S55.   A. Shet *et al.*, *J Acquir Immune Defic Syndr* **41**, 439 (2006).

S56.   F. C. Lampe *et al.*, *Arch Intern Med* **167**, 692 (2007).

S57.   M. F. Shapiro *et al.*, *JAMA* **281**, 2305 (1999).

S58.   B. E. Landon *et al.*, *Ann Intern Med* **140**, 887 (2004).

S59.   R. D. Moore, D. Stanton, R. Gopalan, R. E. Chaisson, *N Engl J Med* **330**, 763 (1994).

S60.   M. D. Stein *et al.*, *J Gen Intern Med* **6**, 35 (1991).

S61.   D. Dunn *et al.*, *Antivir Ther* **13**, 771 (2008).

S62.   V. von Wyl *et al.*, *Clin Infect Dis* **48**, 979 (2009).