A Tale of Two Futures: HIV and Antiretroviral Therapy in San Francisco

S. M. Blower, 1* H. B. Gershengorn, 1 R. M. Grant 2

The effect of antiretroviral therapy (ART) in preventing human immunodeficiency virus (HIV) infections and averting acquired immunodeficiency syndrome (AIDS) deaths in the San Francisco gay community over the next 10 years was predicted. A transmission model was coupled with a statistical approach that enabled inclusion of a high degree of uncertainty in the potential treatment effects of ART (in terms of infectivity and survival), increase in risky behavior, and rate of emergence of drug resistance. Increasing the usage of ART in San Francisco would decrease the AIDS death rate and could substantially reduce the incidence rate.

Currently, 30% of the San Francisco gay community are HIV-infected (1). About 30% of these HIV-infected men are taking combination ART (2); these three or more drug regimens include recently developed protease inhibitors, nonnucleoside reverse transcriptase inhibitors, or both. Part of the recent decrease in the San Francisco AIDS death rate (3) could be attributable to the effect of ART, as ART decreases disease progression rates (4). However, because treated individuals are likely to retain some degree of infectivity, it is possible that ART could lead to an increase in the infection rate (5). Furthermore, drug-resistant HIV strains (that are less responsive to therapy) have emerged (6), and risky behavior has begun to increase in San Francisco (7). Therefore, whether the epidemic-level effects of ART will be beneficial or detrimental is unclear.

To predict (with a degree of uncertainty) the effectiveness of ART in the San Francisco gay community, we developed and analyzed a mathematical model. Our model includes the potential effects of ART on the transmission dynamics of both drug-sensitive and drug-resistant HIV strains. It is specified by five ordinary differential equations (8) (Fig. 1) and allows for drug-resistant strains (that differ in their infectivity and disease progression rates from drug-sensitive strains) to emerge during treatment and to be sexually transmitted (6). Acquired resistance develops because of a variety of factors (8); we model the aggregate effect of all these factors by a single parameter. We model the potential treatment effects of ART by assuming that ART [by reducing viral load (9)] increases average survival time and reduces infectivity, and that drug-resistant strains will be less responsive to therapy than drug-sensitive strains (6). Treatment (in our model) has three outcomes. A patient can respond to ART and remain as a nonprogressor for a specified amount of time, experience clinical failure and death without developing drug resistance (9), or virologically fail treatment and develop drug resistance (10). Individuals can go on and off ART, and drug-resistant infections can revert to drug-sensitive infections if the selective pressure of treatment is removed (11) (Fig. 1).

We predicted the effectiveness of a high usage of ART over the next 10 years in the San Francisco gay community by analyzing our model with time-dependent uncertainty analyses (12, 13). Effectiveness was predicted in terms of the cumulative number of HIV infections prevented and the cumulative number of AIDS deaths averted (14). The San Francisco epidemic has been well studied, and the values of several of the parameters necessary for prediction are known (15); however, the values of other parameters are less certain. Hence, we conducted two uncertainty analyses (an optimistic and a pessimistic analysis) on the basis of different assumptions regarding the rate of increase in risky behavior and the rate of emergence of drug resistance. Both analyses included a high degree of uncertainty in the potential treatment effects of ART (on increasing survival and reducing infectivity). For the optimistic analysis we assumed that the rate of emergence of resistance would remain at a constant, fairly low value [only 10% of cases would acquire resistance per year (16)], and that risk behavior would not increase. For the pessimistic analysis we assumed that the rate of emergence of resistance could substantially increase [10 to 60% of cases could acquire resistance per year (17)], and that risk behavior could increase from almost no increase to a doubling (17).

For each uncertainty analysis we used our

1Department of Medicine, University of California San Francisco, 513 Parnassus, Box 0414, San Francisco, CA 94143–0414, USA. 2Gladstone Institute of Virology and Immunology, University of California, San Francisco, San Francisco, CA 94147–9100, USA.

*To whom correspondence should be addressed. E-mail: sally@hfs.ucsf.edu
model and Latin hypercube sampling (LHS), a type of stratified Monte Carlo sampling (18); LHS has been described elsewhere (12). To make predictions, we assigned each uncertain parameter a probability density function (pdf); the pdf reflected either the uncertainty in the value of the parameter, or the degree to which the parameter could vary if it was being used as an “experimental variable” (12). We used the usage rate of ART \( F_s \) as an “experimental variable” and predicted the effect of increasing usage rates that ranged from treating 50 to 90% of HIV-infected men; currently, only \( \approx 50\% \) of HIV-infected gay men in San Francisco take ART (2, 19).

We used a uniform pdf (range 0.5 to 0.9) to specify the uncertainty in \( F_s \), where \( F_s \) represents the fraction of drug-sensitive cases.

Fig. 1. Flow diagram of the transmission dynamics of an HIV epidemic in the presence of combination antiretroviral therapy (ART); for model equations see (8). The model keeps track of the temporal dynamics of five groups: susceptible individuals \( X \), untreated individuals infected with either drug-sensitive \( Y_{S_U} \) or drug-resistant strains \( Y_{R_U} \), and ART-treated individuals infected with either drug-sensitive \( Y_{S_T} \) or drug-resistant strains \( Y_{R_T} \). The parameter’s subscript specifies whether the infection is drug-sensitive \( S \) or drug-resistant \( R \); the superscript identifies whether the individuals are treated with ART \( T \) or untreated \( U \). Parameter definitions are as follows: \( \pi = \) rate at which gay men join the sexually active community; \( 1/\mu = \) average time during which a gay man acquires new sex partners; \( c = \) average number of new receptive anal sex partners per year; \( p = \) probability of a drug-resistant case (relative to a drug-sensitive case) transmitting drug-sensitive viruses; \( 1/q = \) average time for an untreated drug-resistant infection to revert to a drug-sensitive infection; \( r = \) rate of emergence of resistance due to acquired resistance; \( g = \) proportion of cases that give up ART per year; and \( \nu = \) average disease progression rate. \( \lambda \) specifies the per capita force of infection for drug-sensitive \( (\lambda_s) \) and drug-resistant \( (\lambda_R) \) HIV; \( \lambda_s \) and \( \lambda_R \) are calculated from Eqs. 6 and 7, respectively (8), and are a function of the number of infected people at any particular time \( (Y_{S_U}, Y_{R_U}, Y_{S_T}, \text{and } Y_{R_T}) \) and the infectiousness (as specified by the transmissibility coefficients \( b_{S_U}, b_{R_U}, b_{S_T}, \text{and } b_{R_T} \)) of each of the four types of infected people.

Fig. 2. Results of the two time-dependent uncertainty analyses on the effectiveness of ART on the San Francisco HIV epidemic. For each graph, every 6 months, the 1000 simulations are plotted as a box-plot; these plots show the median value (horizontal red line), upper and lower quartiles, and the outlier cutoffs. In (A) and (B), effectiveness is calculated in terms of the cumulative number of AIDS deaths averted (14); in (C) and (D), effectiveness is calculated in terms of the cumulative number of new HIV infections prevented (14). (A) and (C) show the optimistic predictions; (B) and (D) show the pessimistic predictions. ART decreases the death rate and reduces the incidence rate; these epidemic-level effects balance, hence the prevalence of infection remains fairly stable under both optimistic and pessimistic assumptions (29). However, after 10 years, a fairly high proportion (median value 42%) of the prevalent infections are drug-resistant (under pessimistic assumptions), whereas under optimistic assumptions, a substantially lower proportion (median value 26%) are drug-resistant (29).
treated with ART (20). We used LHS to sample 1000 values of $F_S$ from this pdf; hence, each of the 1000 predictions in each uncertainty analysis had a different ART usage rate (20).

By specifying pdfs, we included uncertainty in our estimates for the average duration of treatment, and the treatment effects of ART (in terms of survival and infectivity). Because ART has only been widely available for 2-3 years, it is not known how long patients will remain on ART. We assumed that from 1 to 25% (with a most likely value of 5%) of drug-sensitive patients would give up ART per year (21). Zidovudine monotherapy, by reducing viral load, decreases infectivity for both vertical and sexual transmission (22). Viral suppression with ART is 10-100-fold more effective than viral suppression with Zidovudine (9). Thus, to construct a pdf we assumed that ART could cause a 2-100-fold reduction in infectivity (23). Many patients on ART who have maintained a virological response have not clinically deteriorated, and clinical benefits of ART can exceed the duration of the virological response (24). However, the data on the survival effects of ART are from clinical studies with short-term follow-up. Hence, we included a high degree of uncertainty in estimating the potential longer-term survival effect of ART. We assumed that the average survival time for ART patients who remain virologically suppressed could be 1.5 to 3 times greater than the average survival time for untreated drug-sensitive individuals (25).

Drug-resistant HIV strains have only recently emerged; thus, their behavior with respect to transmission and pathogenesis is relatively unknown. For our uncertainty analyses, we assumed that drug-resistant strains would be attenuated to some uncertain degree in comparison with drug-sensitive strains (26). We also included uncertainty in the degree to which individuals infected with drug-resistant strains would transmit drug-sensitive strains (27) and in the rate at which untreated drug-resistant infections would revert to drug-sensitive infections (27). Because drug-resistant strains are less responsive to therapy (6), we modeled a differential treatment response to ART (28). We assumed that ART would be less effective in decreasing disease progression rates and in reducing infectiousness in drug-resistant infections (28).

Our uncertainty predictions for the effectiveness of ART in terms of the cumulative number of AIDS deaths averted are shown in Fig. 2. Under optimistic assumptions the cumulative AIDS death rate was always reduced; hence, effectiveness was always positive (Fig. 2A). Under pessimistic assumptions, certain conditions led to an increase in the cumulative AIDS death rate over time; hence, effectiveness was sometimes negative (Fig. 2B). However, both analyses predicted that the most likely outcome would be a substantial reduction in the cumulative AIDS death rate (Fig. 2). After 10 years, 33% (median value under optimistic assumptions) (Fig. 2A) and 18% (median value under pessimistic assumptions) (Fig. 2B) of AIDS deaths had been averted.

The two uncertainty analyses generated very different predictions for the effectiveness of ART in terms of the cumulative number of HIV infections prevented. Under optimistic assumptions, effectiveness began at zero and increased with time; ART substantially reduced the number of new infections (Fig. 2C). After 10 years of ART, 40% (median value under optimistic assumptions) (Fig. 2C) of new HIV infections were prevented. Under pessimistic assumptions, the cumulative number of new HIV infections increased (Fig. 2C). After 10 years of ART, 40% (median value under pessimistic assumptions) (Fig. 2C) of new HIV infections were prevented. Under pessimistic assumptions, effectiveness began at zero and increased with time; ART substantially reduced the number of new infections (Fig. 2C). After 10 years of ART, 40% (median value under optimistic assumptions) (Fig. 2C) of new HIV infections were prevented. Under pessimistic assumptions, the cumulative number of new HIV infections increased (Fig. 2C). After 10 years of ART, 40% (median value under pessimistic assumptions) (Fig. 2C) of new HIV infections were prevented.

To identify the key factors that determined the effectiveness of ART, we performed time-dependent sensitivity analyses (12, 18). We used the 10-year prediction of inpatient from the uncertainty analyses to calculate time-dependent sensitivity coefficients; a partial rank correlation coefficient (PRCC) was calculated annually for each parameter. Key factors that determined effectiveness (as identified by the value of their PRCC) are shown in Table 1 for years 1, 5, and 10. Not surprisingly, one of the two key factors that increased the effectiveness of ART (in terms of decreasing the death rate) was the degree to which ART increased survival (1/TA_2S) (Table 1). The longer ART-treated individuals survived, the lower the cumulative death rate; under optimistic assumptions, this effect was relatively constant over time. Under pessimistic assumptions, this effect waned over time as the increased levels of risk behavior that initially caused a rise in the incidence rate were translated into an increasing death rate. The most important key factor in increasing effectiveness was the usage rate of ART (F_T) (Table 1). Increasing usage rates of ART significantly reduced the AIDS death rate and prevented a substantial number of new infections—under the optimistic assumptions and, perhaps surprisingly, even under the pessimistic assumptions. This conclusion is also shown by the unadjusted data generated by the uncertainty analyses (Fig. 3). Under optimistic or pessimistic assumptions, a high usage of ART substantially de-

### Table 1. Time-dependent sensitivity coefficients (PRCCs) for key parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Optimistic</th>
<th>Pessimistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction of drug-sensitive cases</td>
<td>0.99</td>
<td>0.97</td>
</tr>
<tr>
<td>Transmissibility coefficient of drug-sensitive, treated infection ($\beta_s^t$)</td>
<td>$-0.18$</td>
<td>$-0.10$</td>
</tr>
<tr>
<td>Average survival time in drug-sensitive, treated individuals ($1/TA_s^t$)</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>Increase in risk behavior ($l$)</td>
<td>$-0.86$</td>
<td>$-0.89$</td>
</tr>
<tr>
<td>Fraction of drug-resistant cases treated ($F_R^t$)</td>
<td>1.00</td>
<td>0.87</td>
</tr>
<tr>
<td>Transmissibility coefficient of drug-sensitive, treated infection ($\beta_r^t$)</td>
<td>$-0.96$</td>
<td>$-0.51$</td>
</tr>
<tr>
<td>Transmissibility coefficient of drug-resistant, treated infection ($\beta_r^t$)</td>
<td>$-0.13$</td>
<td>$-0.47$</td>
</tr>
<tr>
<td>Increase in risk behavior ($l$)</td>
<td>$-0.99$</td>
<td>$-0.94$</td>
</tr>
</tbody>
</table>

*Partial rank correlation coefficients (PRCCs) were calculated by using the 1000 values (generated by LHS) for each parameter included in the time-dependent uncertainty analyses and the 1000 predicted values of effectiveness. The cumulative number of AIDS deaths averted and new HIV infections prevented (Table 1) generated by the model; PRCCs were calculated at each year in the 10-year time period. A parameter was identified as a key factor in determining effectiveness if the absolute value of the PRCC (at any year in the 10-year period) was greater than 0.5.
increased the death rate (Fig. 3A) and prevented a significant number of new infections (Fig. 3B).

Not surprisingly, effectiveness decreased as both the infectiousness of treated drug-sensitive (βs) and drug-resistant patients (βr) increased (Table 1). Hence, if infectiousness was reduced (either by increasing condom usage in treated patients or by developing more effective drugs for viral suppression), the effectiveness of ART would substantially increase. High rates of emergence of drug-resistant strains would result in a fairly high prevalence of drug-resistant infections (29), but our sensitivity results revealed that even these high rates of emergence of resistance would not significantly affect either the death rate or the incidence rate (30). The most important key factor that decreased the effectiveness of ART (both in terms of the number of deaths averted and the number of infections prevented) was the degree of increase in risk behavior (Table 1). Even under pessimistic assumptions, a high usage of ART decreased the incidence rate (Fig. 3B); however, an increase in risky behavior of only 10% was enough to counterbalance the benefits of ART (Fig. 3C). Greater increases in risk behavior resulted in the incidence rate increasing, and hence effectiveness becoming negative (Fig. 3C).

Since 1996–97 (when ART became readily available), the AIDS death rate in San Francisco has decreased (3). Our predictions show that a decreased death rate is to be expected under both optimistic and pessimistic assumptions. However, our optimistic and pessimistic assumptions lead to very divergent incidence predictions. Our results show that the higher the usage of ART, the greater the number of infections that will be prevented (Fig. 3B). Because the current usage rate of ART in the San Francisco gay community is ~50% (2, 19), the usage rate should be increased. Recently, increases in risky behavior in the gay community have been reported (7). Our results show that a high usage of ART could counterbalance the effect of increasing levels of risky behavior and prevent a substantial number of new infections. Our results imply that the incidence rate in San Francisco will first rise (to a level that will be determined by the degree of increase in risky behavior) and will then fall (to a level that will be determined by the degree of usage of ART). Significant efforts should be made to prevent risk behavior increasing because even small increases will overtake the effect of ART on reducing the incidence rate (Fig. 3C). To maximize the effectiveness of ART, 13 treatment programs should be combined with effective behavioral intervention programs.

Mathematical models can be used as health policy tools to guide public health decisions (31). However, a model is always an abstraction of reality and never a mirror of reality. Our model reflects current biomedical understanding; we applied Occam’s razor to capture the essential processes of the transmission dynamics of drug-sensitive and drug-resistant strains in San Francisco. As biomedical knowledge accumulates our model can be made more complex, and as data accumulate we can reduce the uncertainty in our predictions. We have presented only short-term predictions because it is likely that more effective drugs and drug regimens will be developed over time; however, our model can also be used to evaluate the longer term consequences of ART (32). Here, we have used parameters that are specific for San Francisco, but our methodology can be applied to evaluate the impact of more widespread usage of ART in other HIV-infected communities.

References and Notes
2. R. Stall et al., in preparation.
The flow diagram for the model is shown in Fig. 1; state variables and parameters are defined in the legend. The model is specified by the following equations:

\[ \frac{dX}{dt} = \pi - X(c_a + \lambda_a) + \mu \]  

\[ \frac{dY_i}{dt} = Xc_i + Y_i q - Y_i g_i + \epsilon_i(s_i + \nu_i) + v_i + p_i(Y_i - Y_i^{1/2}) \]  

\[ \frac{dY_r}{dt} = Y_r q - Y_r g_r + \epsilon_r(s_r + \nu_r) + v_r + p_r(Y_r - Y_r^{1/2}) \]  

Individuals infected with drug-resistant strains (\( Y_i \) and \( Y_r \)) can be dually infected with drug-sensitive strains. The model assumes that drug-resistant strains can be transmitted from a uniform pdf (range 0 to 1.0). We modeled uncertainty in the rate of emergence of drug resistance due to acquired resistance (\( r \)) by using LHS to sample 1000 values from a uniform pdf (min = 0.1, max = 0.6) (12, 13). We modeled uncertainty in the degree of increase in risky behavior (\( b \)) by using LHS to sample 1000 values from a uniform pdf (range 0.0 to 1.0). The average level of risky behavior for each of the 10 simulations (at time zero) was then calculated as \( 1 + \beta \).

15. Recently reported data from large-scale probability surveys of the San Francisco gay community (1, 2) indicate that the HIV prevalence is 30%; because ART has only been available for a few years, it would not have had time to affect prevalence. At time zero in our time-dependent simulations, the prevalence predictions generated by our model (in the absence of ART) for our 1000 simulations matched the reported prevalence data of 30%. To generate these initial predictions, we used parameter values that were specific for the HIV epidemic in the San Francisco gay community (S. M. Blower and A. R. McLean, Science 265, 1451 (1994)). We used the following \( b = 0.0 \), \( c = 0.0 \), and \( \epsilon = 0.0 \), with \( \lambda_a = 0.0 \), each for 12 simulations (the same degree for each simulation) for the average survival time of untreated drug-resistant cases (\( 1/\lambda_a \)) by using sampling constraints during LHS to ensure that \( 1/\lambda_a = 1/\lambda_a^{1/2} \approx 1/2 \).

16. We assumed that \( \alpha = 0.9 \); this value reflects the current optimal antiretroviral therapy in clinical trials where only 10% of the cases on ART develop resistance per year (4).

17. Virolologic failure has been observed in up to 40% of subjects enrolled in ART, even after careful selection of study subjects (4). We modeled uncertainty in the rate of emergence of drug resistance due to acquired resistance (\( r \)) by using LHS to sample 1000 values from a uniform pdf (min = 0.1, max = 0.6) (12, 13). We modeled uncertainty in the degree of increase in risky behavior (\( b \)) by using LHS to sample 1000 values from a uniform pdf (range 0.0 to 1.0). The average level of risky behavior for each of the 10 simulations (at time zero) was then calculated as \( 1 + \beta \).


19. L. Pollak and J. Catania (UCSF), personal communication; data for the study reported in (2) were collected from November 1996 to February 1998.

20. For each of the 1000 sampled values of \( F_r \), the per capita effective treatment rate of drug-resistant cases (\( f_r \)) was calculated from \( F_r = f_r \) (range 0.0 to 1.0). The average level of effective treatment for each of the 10 simulations (at time zero) was then calculated as \( 1 + \beta_r \).

21. For \( g_r \), we used a uniform random regular distribution (min = 0.01, max = 0.25, PDF = 0.05). We assumed that drug-resistant patients would be twice as likely to give up ART as drug-sensitive patients; thus, for \( g_r \), we used a pdf with a triangular distribution (min = 0.02, max = 0.5, peak at 0.05).


23. We modeled uncertainty by multiplying the infectivity of an untreated individual \( \beta_i \) [where \( \beta_i = 0.1 \); from R. M. Grant, J. A. Wiley, W. Winkelstein, J. Infect. Dis. 189, 183 (1998)] with a multiplier (\( \alpha_i \)) that was sampled 1000 times using LHS from a uniform pdf (range 0.0 to 1.0). Values of \( \beta_i \) (range 0.0 to 1.0) were then calculated from \( \beta_i = \alpha_i \beta_i \) [the calculation of \( \beta_i \) is described in (26)]. This procedure ensured that \( \beta_i > \beta_i \). Because \( \beta_i > \beta_i \) (22) and \( \beta_i > \beta_i \) (26), then in each simulation \( \beta_i > \beta_i > \beta_i > \beta_i \).

24. These data are available on Science Online at www.sciencemag.org/feature/data/10442874.htm.

25. Over the 10-year time period the value of the PRCC for effectiveness and the parameter \( \alpha \) (which specifies the rate at which drug resistant strains emerge) remained between zero and 0.2.


27. We modeled differential response to therapy by using two assumptions. First, we reduced the per capita effective treatment rate of drug-resistant strains (in comparison to that of drug-sensitive strains) by an efficacy parameter (\( e \)); where \( 1 \leq e \leq 0 \) (thus, the per capita effective treatment rate of drug-resistant strains equals \( e \)). Second, we assumed that ART could increase survival (\( 1/\lambda_a^{1/2} \)) (to some uncertain degree) in drug-resistant patients, but that this effect would be reduced in drug-sensitive cases in comparison to drug-resistant cases. We operationalized this assumption by using sampling constraints during LHS. We assumed that the average survival times could range from 12 to 36 weeks for both treated (\( 1/\lambda_a^{1/2} \)) and untreated (\( 1/\lambda_a^{1/2} \)) drug-resistant patients. Sampling constraints ensured that for each of the 1000 simulations (in each analysis), \( 1/\lambda_a > 1/\lambda_a^{1/2} \) to model uncertainty in the effect of ART on reducing infectiousness in treated drug-resistant cases (\( \beta_i \)). We used LHS to sample 1000 values of a multiplier (\( \alpha_i \)) from a uniform pdf (range 0.0 to 1.0). We also calculated from \( \beta_i = \alpha_i \beta_i \) for \( \beta_i > \beta_i \). [The calculation of \( \beta_i \) is described in (26)]. This procedure ensured that \( \beta_i > \beta_i \). Because \( \beta_i > \beta_i \) (23) and \( \beta_i > \beta_i \) (26), then in each simulation \( \beta_i > \beta_i > \beta_i > \beta_i \).