The EMOD-HIV v0.7 Model in Detail
Electronic Supplementary Material 1

Anna Bershteyn*, Daniel J. Klein*, Philip A. Eckhoff

Intellectual Ventures Laboratory, 1555 132nd Ave. NE, Bellevue, WA 98005
*A. Bershteyn and D.J. Klein contributed equally to this work
Correspondence: abershteyn@intven.com

1 Implementation

EMOD-HIV v0.7 is an individual-based stochastic simulation of sexual and vertical HIV transmission implemented in C++. Population-level outputs are calibrated to South Africa at a national level. Sexual mixing parameters are based on data from rural KwaZulu-Natal.

It is part of a suite of agent-based microsimulations developed at Intellectual Ventures Laboratory, which also includes malaria [1], polio, and tuberculosis. The models and their accompanying software infrastructure are designed for disease modelers, researchers, epidemiologists, and public health professionals seeking to simulate infectious disease conditions and evaluate the effectiveness of different eradication or mitigation approaches.

The EMOD-HIV model propagates forward in time using a combination of discrete events and time steps. Discrete events, implemented using C++11 lambda expressions, allow for one or more function calls to be scheduled for execution at a future time in the simulation.

Geographic distribution of the population was not explicitly modeled in v0.7, but an evolving network of human relationships was formed based on pair formation rules. Though capable of modeling homosexual partnerships and commercial partnerships (in which providers are distinguished from patrons), v0.7 utilized only heterosexual relationships due to their importance in the modern sub-Saharan Africa transmission setting [2]. Thus, HIV transmission could only result from heterosexual intercourse or mother-to-child vertical transmission in v0.7. What follows is a detailed walk-through of the model components, shown diagrammatically in Figure S1.

2 Demographics

Vital dynamics within the EMOD-HIV model are derived from fertility and mortality tables that are passed to the model as input. In v0.7, these values came from the Actuarial Society of South Africa (ASSA) AIDS and Demographic Model [3], using linear interpolation between age-specific data points from ASSA to construct a cumulative probability distribution function (CDF) of death date from a person's date of birth. Upon instantiation, individual agents in the model sample stochastically from this CDF using an inverse transform of the distribution.

Female agents similarly sample the age at next childbirth, if any, upon instantiation and birth of a child. Pregnancy is not linked to relationship status in v0.7, although newly born individuals are linked to a mother. The fertility rate changes by simulation year and female age. HIV impacts fertility in an age-dependent manner, with older women experiencing a greater reduction in fertility [4]. Available estimates from ASSA range from 1980 to 2020. Values outside of this range were chosen by clamping - i.e., choosing the nearest value within the range. Clamping was also used when necessary to determine the non-AIDS mortality rate, which varies by gender, age, and simulation year.

3 Individual properties

Individuals within the simulation have a variety of properties, represented by continuous or discrete state variables. Some properties are static throughout life, and others dynamically change through the course of the simulation in response to aging or to simulation events such as infection. Included in these properties are dynamic objects for HIV infection and relationships, to be described shortly. Static properties assigned upon instantiation (simulation initial-
Figure S1: The EMOD-HIV model is designed to separate components related to the contact network (Network) from within-host biology of HIV infection (Intrahost). Both of these build upon a generic “Person” object with properties that are more broadly applicable in epidemiological models (e.g., age, gender, and vital statistics). This Person object is specialized into “STI Person,” containing attributes related to behavior within the contact network and potentially applicable to other STIs, and a further specialized “HIV Person” object containing attributes that are specific to the biology of HIV. Interventions can modify factors in both modules that influence transmission, such as the rates of condom use or use of antiretroviral therapy. In v0.7, each HIV-discordant interaction is modeled explicitly, and the transmission probability is calculated based on behavioral factors (type of sexual contact, condom use, etc.) and biological factors (stage of HIV disease, co-infections, ART, etc.) after interventions are applied.

The algorithm used to form pairs among relationship-seeking individuals is called the pair formation algorithm (PFA). Each of the three relationship types uses the same PFA algorithm, but operates on different data about the age distribution and age gaps within partnerships. A mathematical description of the PFA has been recently published [10].

The PFA takes as input a matrix, called the matrix of partnership age distributions (MOPAD), in which the entry at position \((i, j)\) is the probability of a male in age bin \(i\) pairing with a female in age bin \(j\). After entering a PFA, individuals are organized into gender-specific queues, as shown in the diagram in Figure S2. There is a single queue for all males, and for females, there are separate queues for each 2.5-year age bin. Every day (transitory and informal) or 14 days (marital), relationships are formed among all those who have accumulated in the queues, starting from male at the head of the single male queue. A female partner is selected based on age and availability. Specifically, a male of age \(m\) samples a partner...
Figure S2: A diagram of the Pair Formation Algorithm (PFA). Males form a single queue, while females are sorted into queues according to their age. Males and females accumulate in the order they entered and remain in the queues until the PFA is processed. Processing takes place once per day in the transitory and informal PFAs, and once every 14 days in the marital PFA. During processing, the individual at the head of the male queue selects a partner based on the MOPAD-derived conditional probability weighting. The paired individuals exit the algorithm. After processing is complete, only individuals with no available partners of the chosen age are left in queues for the next round of processing.

The pair formation algorithm can produce relationships distributed according to the MOPAD only if equal numbers of males and females enter the algorithm, and if their age distributions match the respective marginals of the MOPAD. To ensure that this occurs, (exponential) rates at which individuals in each age bin and gender enter for each relationship type are adjusted daily. The MOPAD dictates only the relative number of relationships formed between pairs of different ages, but not the absolute number of relationships formed by the PFA. This total throughput of relationships formed is set such that the expected number of individuals seeking a relationship of a particular type, after rate adjustment to meet the MOPAD, matches the number of males and females that would have sought that relationship had the rates not been adjusted for the MOPAD. These unadjusted rates are listed in Table 1. Thus, the rate adjustment changes the age distributions of the individuals seeking relationships, but not the total number of each type of relationship formed.

This adaptive daily rate control allows the model to automatically discover the rates of relationship entry that are consistent with the MOPAD. It is conceivable, however, that events causing large demographic shifts might change the true age distribution of partnerships. For example, when comparing simulations with universal HIV treatment versus no treatment, it is conceivable that the demographic influence of AIDS mortality could cause the age distribution of newly formed partnerships to diverge from the MOPAD. This is because the MOPAD would have been based on data collected prior to the demographic change.

Therefore, adaptive rate control is enabled only for an initial burn-in period of 15 years. After the burn-in period, the entrance rates remain fixed at their final values, so that the age distribution of relationships formed by the PFA can change as the simulation progresses.

The calculation of the total relationship formation rate (Table 1) and adjustment of individual relationship entry rates are based only on individuals eligible

<table>
<thead>
<tr>
<th>Duration</th>
<th>Rate (per year)</th>
<th>Mean (per year)</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitory</td>
<td>1/2</td>
<td>1/0.9</td>
<td>1.2</td>
</tr>
<tr>
<td>Informal</td>
<td>1/1.9</td>
<td>1/1.8</td>
<td>2.0</td>
</tr>
<tr>
<td>Marital</td>
<td>1/30</td>
<td>1/20</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Table 1: Base exponential relationship formation rates and Weibull dissolution rate parameters.

3
to form a new relationship. The minimum age to enter transitory or informal partnerships is the individual’s debut age. Marriages additionally require a minimum age of 15 years.

On average, young individuals will tend to enter transitory partnerships, while older individuals will tend to enter marital partnerships due to the different age distributions in the MOPADs. Once in a relationship, the previously described Boolean concurrency flags determine if an individual is allowed to add additional (concurrent) relationships, after a waiting period of 60 days. For example, if the transitory concurrency flag is true, the individual is allowed to add additional relationships of any type as long as their only ongoing relationships are of transitory type.

The maximum number of simultaneous relationships is restricted to three transitory, two informal, and one marital. The formation of a marital partnership, which has the lowest probability of permitting concurrency and the longest duration, frequently prevents individuals from taking on additional partnerships. Combined with the increased rates of entry into shorter transitory and informal partnerships, this leads to increased concurrency earlier in life and declining concurrency later in life, although there is no explicit age-dependence of concurrency in the model.

6 Relationship Properties

In addition to person-specific properties, EMOD-HIV also supports relationship-specific properties, such as condom usage probability.

The condom usage probability is determined at the time a relationship is created, and held at this value until the relationship ends. For each relationship type, the condom usage probability $P(t)$ depends on the simulation time $t$ as follows:

$$P(t) = \frac{h}{1 + e^{-R(t-t_0)}} + l.$$  

The values of parameters $l$, $h$, $t_0$, and $r$ for each relationship type are listed in Table 2.

The relationship type-dependence of these parameters accounts for lower condom usage rates in longer-term relationships, while the time-dependence accounts for reported increases in usage over time. Self-reported condom usage probabilities vary greatly. Because of their uncertainty, these values were tuned in the process of model calibration. The values in Table 2 come from calibrating parameters that were initially based on data from the 2002, 2005, and 2008 HSRC national surveys [2, 7, 8].

When a relationship forms, its breakup time is sampled from a PFA-type-dependent Weibull distribution. The mean and shape parameters of these distributions are listed in Table 1. Note that relationships can terminate early if either partner dies, in which case the surviving partner acts equivalently to someone who has become eligible through a relationship breakup.

7 Coital Frequency and Dilution

The base time between coital acts is exponentially distributed with a mean of three days [12]. For individuals in two, three, or four-plus concurrent partnerships, the model additionally accounts for coital dilution, i.e., a negative correlation between the number of simultaneous partnerships and the frequency of coital acts within these partnerships [13, 14]. While some studies suggest a very strong effect of coital dilution [15, 16], others find little evidence for this phenomenon [17, 18]. EMOD-HIV v0.7 uses a relatively modest dilution level: the act frequency per relationship for an individual belonging to two, three, or four (or more) concurrent relationships is multiplied by 0.75, 0.60, and 0.45, respectively.

8 Stages of HIV Infection

We assume three stages of untreated HIV infection: acute, latent, and AIDS. In v0.7, the durations of the acute and AIDS stages are 2.9 and 9 months, respectively [19]. The survival time with HIV is sampled stochastically based on the individual’s age at the time of infection (discussed below), and the duration of the latent phase is the survival time minus the duration of the acute and AIDS stages. In rare cases when the survival time is shorter than 11.9 months, the AIDS phase is shortened, and in even rarer cases when survival is shorter than 2.9 months, the remaining acute phase is shortened to fit within the survival duration.

9 HIV Transmission

The probability of HIV transmission within a discordant relationship is calculated on a per-act basis and varies by disease stage. The per-act transmission probabilities for the acute, latent, and AIDS stage are 0.034, 0.0013, and 0.0094, respectively. The relative magnitude of these values comes from the per-time
transmission rates measured in HIV-discordant couples in Uganda [19]. These probabilities have been multiplied by the inverse act frequency and increased by 50% as a result of the model calibration process. Note that these transmission probabilities apply in the absence of any interventions or other cofactors. The per-act transmission probability is multiplied by applicable act-dependent factors, which are described in Table 3. There is no gender bias in transmission other than the male circumcision cofactor, which reduces the circumcised male’s HIV acquisition probability by 60% [20, 21, 22].

Mother-to-child transmission (MTCT) is enabled in v0.7 using the linkage between the mother and offspring. The baseline rate of MTCT is 1 in 3 births [23, 24, 25]. Antiretroviral therapy (ART) reduces the transmission rate to 1 in 100 births [26], except in the final 9 months prior to death due to ART failure, when the baseline rate is used. Additionally, some mothers not on ART receive single-dose Nevirapine prophylaxis (sdNVP), which reduces the MTCT rate to 1 in 10 births [27, 28, 29]. Historical trends in usage of these options are configurable. In v0.7, sdNVP was set to a constant rate of 45% per birth.

10 Survival with HIV Infection

HIV prognosis is both heterogeneous and highly age-dependent [30, 31, 32], and is therefore sampled from an age-dependent distribution, shown in Figure S3. For individuals over the age of 15, the distribution is Weibull-distributed with shape parameter \( \beta \) set to 2 and an age-variable scale parameter \( \alpha \) (in years) as follows [30]:

\[
\lambda = 21.182 - 0.2717a. \quad (3)
\]

For those under 15, the distribution is a convex combination of two Weibull distributions,

\[
\lambda_1 = 1.515 + 1.039a \quad (4)
\]
\[
\kappa_1 = 0.97 + 0.0687a \quad (5)
\]
\[
\lambda_2 = 10.0 + 0.474a \quad (6)
\]
\[
\kappa_2 = 5.39 - 0.226a, \quad (7)
\]

distinguishing infant mortality and later child mortality [33]. The parameters \( \lambda_1 \) and \( \kappa_1 \) are used 65% of the time.

CD4+ T-cell count is time-dependent based on the fraction remaining of the individual’s prognosis \( f \), calculated by dividing the time since infection by the total HIV prognosis. It drops to 594 cells/\( \mu \)L immediately upon infection to represent very rapid CD4 T-cell depletion (and partial rebound) during the acute phase (the detailed timecourse of which is not modeled here), and decreases down to 59 cells/\( \mu \)L at death as follows:

\[
CD4(\text{cells/}\mu\text{L}) = (24.363 - 16.672f)^2. \quad (8)
\]

11 Timecourse of a Simulation

Simulation begins in year 1960 to allow ample time for relationships to burn in. During the period between 1960 and 1975, the relationship formation rates for each gender and relationship type are updated daily using the relationship flow algorithm. Adjustment of PFA entry rates is terminated in 1975, and the rates are fixed at their 1975 values for the remainder of the simulation.

Infections are seeded in year 1980. The seeding process infects 5% of the population according to the age and gender distribution reported in the 2008 HSRC survey [8]. This initial prevalence is higher than values generally accepted for 1980 in South Africa; however, reproducing the historical trend of HIV prevalence would require information about HIV risk during the early epidemic, about which little is known.

Rather than arbitrarily modeling one of many possible modes of behavior change to match early epidemic trends, EMOD-HIV v0.7 did not model these early-epidemic conditions. Instead, the epidemic was initialized at an inflated prevalence to account for the relatively lower-risk individuals who are initially infected in the model.

12 Model Calibration

The EMOD-HIV v0.7 calibration prioritized data from the 2002, 2005, and 2008 HSRC surveys [2, 7, 8]
Table 3: Cofactors used in determining the HIV transmission probability.

<table>
<thead>
<tr>
<th>Cofactor</th>
<th>Transmission Probability Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male circumcised (F→M transmission only)</td>
<td>0.4</td>
</tr>
<tr>
<td>STI present</td>
<td>4.0</td>
</tr>
<tr>
<td>Condom used</td>
<td>0.1</td>
</tr>
<tr>
<td>Anal intercourse (infected receptive female partner)</td>
<td>4.0</td>
</tr>
<tr>
<td>Anal intercourse (infected insertive male partner)</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure S3: The untreated HIV prognosis of an individual is determined at the time of infection using this age-dependent distribution. Before age 15, a sum of two weighted Weibull distributions is used to represent the typically bimodal survival time of children. Beyond age 15, survival time is Weibull-distributed, with older individuals having a shorter average survival time. Note that untreated HIV-infected individuals die at the end of their HIV prognosis or their normal lifespan (determined by the demographics module), whichever occurs earlier.

with less weight given to antenatal prevalence data from 1990 to 2009 [34] because of potential biases in extrapolating prevalence from the antenatal client population to the general population or sub-populations modeling in EMOD-HIV.

Parameter values were initialized to values reported in national surveys, even when these values have high uncertainty and likelihood of survey bias. Then, a coordinate descent with golden section search was used to incrementally adjust parameter values, one at a time, so as to minimize a squared error loss function. This function penalized differences between the simulation output and HSRC prevalence data by age and gender, and ANC prevalence as compared to simulated females aged 25 to 30. Parameters that are known with the lowest certainty were prioritized for adjustment. A limitation of this approach was its inability to automatically resolve co-varying parameters.

The parameters that were subject to the most careful calibration were those to which the epidemic trajectory was highly sensitive, but for which available data is insufficient to know their value with reasonable certainty. These include the absolute per-contact transmission rate, the rates of condom use by year in each relationship type, and the probability of concurrency in each relationship type. Calibration to population and HIV prevalence by age, gender, and year are shown in Figure S4. Improvements to the calibration are ongoing and a major component of future work with this model.

In comparison with other HIV models, EMOD-HIV produces similar baseline trends such as population, HIV prevalence, and HIV incidence. This is as expected, since models in the systematic comparison were asked to represent the HIV epidemic in South Africa based on the best available national-level data [38]. The population-level outputs from EMOD-HIV are summarized in Figure S5. Early HIV prevalence (Figure S5a) is initialized at a higher level and climbs more slowly than is generally accepted during the early epidemic, because EMOD-HIV does
Figure S4: Age distribution of the population and of HIV prevalence in the population in 2002, 2005, and 2008. Male statistics are shown in blue above the axis, and female statistics are in red above the axis. Faded lines show outputs from fifty EMOD-HIV simulations. In the population graphs on the left, the dotted, dashed, and solid dark lines show demographic estimates from Statistics South Africa [35], the World Bank [36], and the US Census Bureau [37]. In the prevalence graphs on the right, dark lines show estimates from the 2002, 2005, and 2008 HSRC surveys that measured HIV prevalence in the general population [2, 7, 8].

not attempt to model the behavioral patterns of the early epidemic period, about which little is known.

The drop in incidence (Figure S5b) is driven largely by the increase in condom usage in all relationship types, especially transitory and informal relationships, which reduces per-contact transmission probabilities as described in the previous section. Not only was increasing condom use consistent with behavioral survey data, but it was also necessary for prevalence to climb quickly and then level off, consistent with national-level prevalence data.

The proportion of transmissions that occur during the acute stage also declines over time, as shown in Figure S5d. Acute-stage transmissions rely mainly on transitory and informal partnerships that exhibit higher concurrency. These partnerships reach higher levels of condom usage rates than marriages, and the increased condom usage tends to prevent transmission or delay it until later stages of disease.

References
Figure S5: Population-level outputs from the EMOD-HIV model calibrated to a national-level South African epidemic. Faded lines show individual traces from each of 20 simulations with different random number seeds, solid lines show the average of these simulations, and dashed lines show one standard deviation from the mean. (a) HIV prevalence and (b) HIV incidence among adult males (blue) and females (red). Incidence was calculated as a moving average across one year of infections per uninfected sexually active person-year. (c) Population size (black, scale bar to the right), and number of infected individuals (scale bar to the left): total (purple), with CD4 count \(<350\) cells/\(\mu\)L (teal), and with CD4 count \(<200\) cells/\(\mu\)L (orange), in a simulation initialized with 75,000 individuals. Number infected does not drop off as steeply as prevalence due to population growth. (d) The proportion of transmissions occurring during the acute (red), latent (green), or late (blue) stage of disease.


