The $SILD$ model for the HIV expansion: Anticipating a fast transition from a concentrated to a generalized epidemic in Madagascar

David Alonso and Xavier Vallès

Blanes, Barcelona, Catalonia

Abstract

Keywords: Science, Publication, Complicated

1. Introduction

Madagascar lies in the east coast of Africa, close to Mozambique and South-Africa, both or them are high burden HIV countries. By contrast, Madagascar show a very low prevalence of HIV in the general population, alongside a high prevalence among vulnerable populations: Men who have sex with Men (MSM), Sex workers (SW) and Injected Drug Users (IDU). This is an astonishing concentrated epidemic profile, particularly given that Madagascar shows a widespread presence of most of the so-called risk factors associated to high prevalence countries: early onset of sexual intercourse, specially among girls, high degree of sexual concurrency, common presence of transactional sex and low or inconsistent use of barrier methods, gender inequality and general poverty. As a result, it has been reported a high prevalence and incidence of other sexual transmitted diseases (STD) which in turn are a risk factor for HIV acquisition. These observations are reassured by the low prevalence of HIV among women in antenatal clinics and the low rate of TB/HIV coinfection and the scarcity of HIV diagnostics among general population reported by NGOs and health workers, except among risky groups, which among them MSM and IDU have been largely neglected. Since the early 90s, previous articles have foreseen a close tipping point towards a generalized epidemic, that has not occurred to date, even if the general trend is an increase of HIV prevalence among key populations.

Madagascar is one of the few countries in sub-Saharan Africa with a PIB decrease, following a political crisis between 2009-2013. Therefore, the health system is weak and rather inefficient, specially in rural areas, and so is the response to HIV. The basic indicators of the services cascade are very poor, with xxx, xxx and xxxx. Furthermore HIV is not perceived as a health priority in a country, overwhelmed by other health problems: malaria, TB, malnutrition, maternal and child health. Regarding this situation, it makes sense to examine the explanatory hypothesis to this situation and in general the controversial explanations that may modulate HIV epidemics (concurrency, the role of male circumcision, etc.), and, in turn, to draft reliable scenarios in order to foresee the future evolution of HIV epidemic in Madagascar. Madagascar may be an interesting model to explore the transition from a concentrated to a generalized epidemic. This may help to stress the efforts to avoid such transition, in a context that may produce a perfect health crisis storm.

2. Materials and Methods

2.1. The SILD Model

We represent the temporal dynamics of disease spread by a set of ordinary differential equations [1]. This system represents the progression of the disease as a consequence of sexual encounters between infectious and non-infected individuals. The whole population is divided in a set of groups. Male population is considered as a single group, while female population is subdivided into four groups: two groups according to sexual activity (sexual workers and rest of women), where each of them is in turn subdivided into young and adult females. Both males and females are recruited into the population as fully susceptible individuals represented by the subscript S in the dynamic variables —see Eqs (1)-(8). These recruitment rates represent the number of males and females per unit time that reach sexual maturity at any given time. As these individuals encounter infectious sexual partners, they can get the virus and then transition to a first infectious stage (see subscript I in the equations below). This stage is short, but highly infectious. Then, after a certain time, controlled by rate γ, individuals enter a second infectious stage (see subscript L). At this stage individuals can still infect through sexual contacts, but with lower probability. They can remain in this *latent* stage, without showing disease signs, for years until the outset of severe disease symptoms following the breakdown of their immune system.

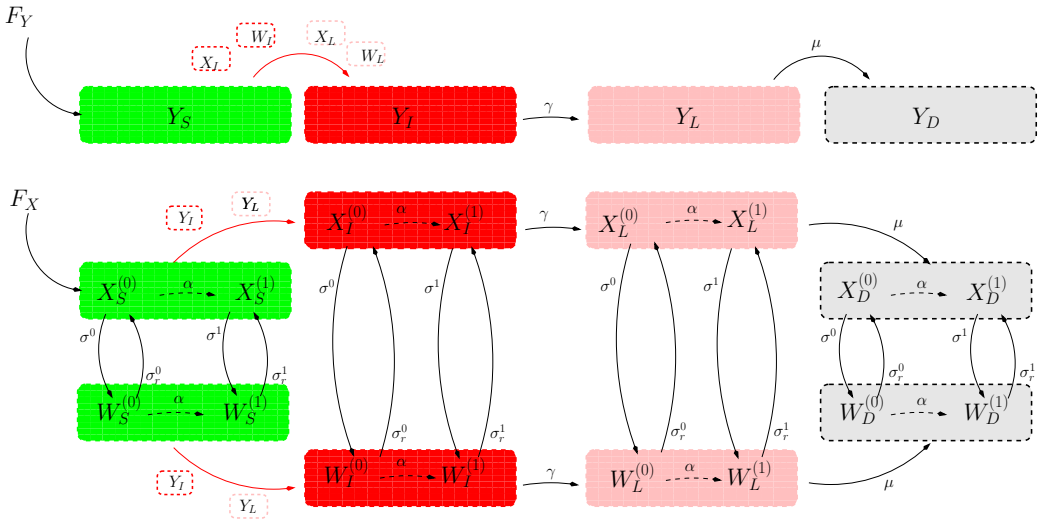

Figure 1: Graphic representation of women and men subpopulations progressing through the different stages of the disease since they acquire the infection from infectious males (Y_I or Y_L) and women ($\hat{X}_I$, $\hat{X}_L$), respectively. At any stage of the disease, women can become sexual workers, at rate σ, or reverse that condition, at rate σ_r. The *aging* rate α controls the transition from young to adult women. All stages, both in men and women, are subject to mortality. For simplicity, arrows representing this fatal transition are not shown.

When this occurs, individuals enter a terminal stage (see subscript D) where most disease-induced mortality concentrates (see Fig 1).

The female-male coupled system is then organized into two separate submodels: the male and the female submodel. These separate submodels are coupled through the force of infection or transmission rate, this is, the per capita rate at which male (or female) susceptible individuals acquire the infection through sexual contancts from female (or male) infectious individuals. This transmission rate takes into account the number of sexual contacts per year of an average male (or female) individuals, β_Y (or β_X), a distinct probability of transmission from infectious females to males (or from infectious males to females), p_{YX} (or p_{XY}), and the effective population fractions of infectious females (or males), x (or y). Notice also that basal female contact rate, β_X, is corrected by a factor $\eta > 1$ for sexual workers to indicate the higher contact rates that characterize this activity.

The male subsystem reads:

$$\begin{aligned}
\frac{dY_S}{dt} &= F_Y - \beta_Y\, p_{YX}\, x\, Y_S - \delta\, Y_S \\
\frac{dY_I}{dt} &= \beta_Y\, p_{YX}\, x\, Y_S - \delta\, Y_I - \gamma\, Y_I \\
\frac{dY_L}{dt} &= \gamma\, Y_I - \mu\, Y_L - \delta\, Y_L \\
\frac{dY_D}{dt} &= \mu\, Y_L - (1+m)\,\delta\, Y_D
\end{aligned} \tag{1}$$

where the total effective fraction of infectious females is a weighted average of the effective infectious fractions including young, x_0, and adult women, x_1:

$$x = f_0\, x_0 + (1 - f_0)\, x_1 \tag{2}$$

where f_0 is the fraction of total sexual encounters a male has with young females. Effective infectious fractions of females should account for differential transmission of the disease from females either in the highly infectious group or the latent group:

$$x_0 = \frac{f_W\,(\chi\, W_I^{(0)} + W_L^{(0)}) + (1 - f_W)(\chi\, X_I^{(0)} + X_L^{(0)})}{N_f} \tag{3}$$

$$x_1 = \frac{f_W\,(\chi\, W_I^{(1)} + W_L^{(1)}) + (1 - f_W)(\chi\, X_I^{(1)} + X_L^{(1)})}{N_f} \tag{4}$$

where f_W if the fraction of male sexual contacts with women who are sexual workers, and N_f is the total female population.

The female subsystem accounts for the progression of the disease in the four female groups, which all follow the same four-equation scheme. Equations for young females that are not non-sexual workers are given by:

$$\begin{aligned}
\frac{dX_S^{(0)}}{dt} &= F_X - \beta_X\, p_{XY}\, y\, X_S^{(0)} - \delta\, X_S^{(0)} - \alpha\, X_S^{(0)} - \sigma^0\, X_S^{(0)} + \sigma_r^0\, W_S^{(0)} \\
\frac{dX_I^{(0)}}{dt} &= \beta_X\, p_{XY}\, y\, X_S^{(0)} - \delta\, X_I^{(0)} - \alpha\, X_I^{(0)} - \gamma\, X_I^{(0)} - \sigma^0\, X_I^{(0)} + \sigma_r^0\, W_I^{(0)} \\
\frac{dX_L^{(0)}}{dt} &= \gamma\, X_I^{(0)} - \mu\, X_L^{(0)} - \delta\, X_L^{(0)} - \alpha\, X_L^{(0)} - \sigma^0\, X_L^{(0)} + \sigma_r^0\, W_L^{(0)} \\
\frac{dX_D^{(0)}}{dt} &= \mu\, X_L^{(0)} - (1+m)\,\delta\, X_D^{(0)} - \alpha\, X_D^{(0)} - \sigma^0\, X_D^{(0)} + \sigma_r^0\, W_D^{(0)}
\end{aligned} \tag{5}$$

The four basic equations for old/adult non-sexual-worker females are:

$$\begin{aligned}
\frac{dX_S^{(1)}}{dt} &= -\beta_X\, p_{XY}\, y\, X_S^{(1)} - \delta\, X_S^{(1)} + \alpha\, X_S^{(0)} - \sigma^1\, X_S^{(1)} + \sigma_r^1\, W_S^{(1)} \\
\frac{dX_I^{(1)}}{dt} &= \beta_X\, p_{XY}\, y\, X_S^{(1)} - \delta\, X_I^{(1)} + \alpha\, X_I^{(0)} - \gamma\, X_I^{(0)} - \sigma^1\, X_I^{(1)} + \sigma_r^1\, W_I^{(1)} \\
\frac{dX_L^{(1)}}{dt} &= \gamma\, X_I^{(1)} - \mu\, X_L^{(1)} - \delta\, X_L^{(1)} + \alpha\, X_L^{(0)} - \sigma^1\, X_L^{(1)} + \sigma_r^1\, W_L^{(1)} \\
\frac{dX_D^{(1)}}{dt} &= \mu\, X_L^{(1)} - (1+m)\,\delta\, X_D^{(1)} + \alpha\, X_D^{(0)} - \sigma^1\, X_D^{(1)} + \sigma_r^1\, W_D^{(1)} \qquad (6)
\end{aligned}$$

Likewise, the four equations for young, sexual-worker females read:

$$\begin{aligned}
\frac{dW_S^{(0)}}{dt} &= -\beta_X\, p_{XY}\, (1+\eta)\, y\, W_S^{(0)} - \delta\, W_S^{(0)} - \alpha\, W_S^{(0)} - \sigma_r^0\, W_S^{(0)} + \sigma^0\, X_S^{(0)} \\
\frac{dW_I^{(0)}}{dt} &= \beta_X\, p_{XY}\, (1+\eta)\, y\, W_S^{(0)} - \delta\, W_I^{(0)} - \alpha\, W_I^{(0)} - \gamma\, W_I^{(0)} - \sigma_r^0\, W_I^{(0)} + \sigma^0\, X_I^{(0)} \\
\frac{dW_L^{(0)}}{dt} &= \gamma\, W_I^{(0)} - \mu\, W_L^{(0)} - \delta\, W_L^{(0)} - \alpha\, W_L^{(0)} - \sigma_r^0\, W_L^{(0)} + \sigma^0\, X_L^{(0)} \\
\frac{dW_D^{(0)}}{dt} &= \mu\, W_L^{(0)} - (1+m)\,\delta\, W_D^{(0)} - \alpha\, W_D^{(0)} - \sigma_r^0\, W_D^{(0)} + \sigma^0\, X_D^{(0)} \qquad (7)
\end{aligned}$$

Finally, those equations for old/adult sexual-worker females are represented by:

$$\begin{aligned}
\frac{dW_S^{(1)}}{dt} &= -\beta_X\, p_{XY}\, (1+\eta)\, y\, W_S^{(1)} - \delta\, W_S^{(1)} + \alpha\, W_S^{(0)} - \sigma_r^1\, W_S^{(1)} + \sigma^1\, X_S^{(1)} \\
\frac{dW_I^{(1)}}{dt} &= \beta_X\, p_{XY}\, (1+\eta)\, y\, W_S^{(1)} - \delta\, W_I^{(1)} + \alpha\, W_I^{(0)} - \gamma\, W_I^{(1)} - \sigma_r^1\, W_I^{(1)} + \sigma^1\, X_I^{(1)} \\
\frac{dW_L^{(1)}}{dt} &= \gamma\, W_I^{(1)} - \mu\, W_L^{(1)} - \delta\, W_L^{(1)} + \alpha\, W_L^{(0)} - \sigma_r^1\, W_L^{(1)} + \sigma^1\, X_L^{(1)} \\
\frac{dW_D^{(1)}}{dt} &= \mu\, W_L^{(1)} - (1+m)\,\delta\, W_D^{(1)} + \alpha\, W_D^{(0)} - \sigma_r^1\, W_D^{(1)} + \sigma^1\, X_D^{(1)} \qquad (8)
\end{aligned}$$

Total effective infectious fractions, x an y, are the link between female and male disease dynamics. As you see in Eqs (5)-(8), the force of infection of females includes a factor, the total effective fraction of infectious males, y, which is given by:

$$y = \frac{\chi Y_I + Y_L}{N_m} \tag{9}$$

where N_m is the total male population. Total populations, N_m and N_f, also change dynamically. They are written in terms of sums over the different disease stages and grups:

$$\begin{aligned} N_m &= Y_S + Y_I + Y_L + Y_D \\ N_f &= \hat{X}_S + \hat{X}_I + \hat{X}_L + \hat{X}_D \end{aligned} \tag{10}$$

where

$$\begin{aligned} \hat{X}_a &= X_a^{(0)} + X_a^{(1)} + W_a^{(0)} + W_a^{(1)} \\ a &\in \{S, I, L, D\} \end{aligned} \tag{11}$$

The rest of symbols appearing in the equations are explained in Table 2.1.

2.2. Demography

2.2.1. Model assumptions

The SILD model extremely simplifies population demography. In the absence of disease transmission, and provided demographic parameters are kept constant, women and men subpopulations are assumed to evolve in time according to the following system:

$$\begin{aligned} \frac{dY}{dt} &= F_Y - \delta Y \\ \frac{dX^{(0)}}{dt} &= F_X - \sigma^0 X^{(0)} + \sigma_r^0 W^{(0)} - \alpha X^{(0)} - \delta X^{(0)} \\ \frac{dW^{(0)}}{dt} &= \sigma^0 X^{(0)} - \sigma_r^0 W^{(0)} - \alpha W^{(0)} - \delta W^{(0)} \\ \frac{dX^{(1)}}{dt} &= \alpha X^{(0)} + \sigma_r^0 W^{(1)} - \sigma^0 X^{(1)} - \delta X^{(1)} \\ \frac{dW^{(1)}}{dt} &= \alpha W^{(0)} + \sigma^0 X^{(1)} - \sigma_r^0 W^{(1)} - \delta W^{(1)} \end{aligned} \tag{12}$$

Model Parameter	**Symbol**	V_0	V_1	V
Demographic parameters				
No of males entering sexual age per year	F_Y	0.00	$5\ 10^4$	$2.5\ 10^4$
No of females entering sexual age per year	F_X	0.00	$5\ 10^4$	$2.5\ 10^4$
Natural mortility percapita rate	δ	0.01	0.05	0.02
Maturation rate into the adult women stage	α	0.07	0.10	0.07
Transition rate into the sexual worker stage (young females)	σ^0	0.00	0.50	0.10
Reversal rate back from the sexual worker stage (young females)	σ_r^0	0.00	0.50	0.50
Transition rate into the sexual worker stage (adult females)	σ^1	0.00	0.50	0.10
Reversal rate back from the sexual worker stage (adult females)	σ_r^1	0.00	0.50	0.50
Disease transmission parameters				
Total male sexual contact rate	β_Y	0.00	600.00	50.00
Total female sexual contact rate	β_X	0.00	600.00	10.00
Male-to-Female transmission probability	p_{XY}	0.00	1.00	0.03
Female-to-Male transmission probability	p_{YX}	0.00	1.00	0.01
Recovery rate from I into L stage	γ	6.00	24.00	12.00
Transition rate from L into D stage	μ	0.05	0.20	0.10
Relative disease-induced increase mortality factor	m	0.00	99.00	9.00
Contact rate relative increase factor for sexual workers	η	0.00	9.00	49.00
Number of latent substages during L phase	n	1.00	10.00	1.00
Relative transmission increase probability factor during acute infectious phase, I	χ	0.00	9.00	9.00
Fraction of male sexual contacts with female sexual workers	f_W	0.00	0.50	0.01
Fraction of male sexual contacts with young females	f_0	0.00	0.50	0.70

Table 1: Model Parameters for the $SILD$ model. All rates are given in year^{-1}. V_0 and V_1 define reasonable parameter ranges for each parameter value, V.

2.2.2. Madagascar Demographic Data

In order to add realism to our projections, we considered real Madagascar demographic data from governmental and institutional sources (mainly US government and the World Health Organization)[1].

For each year, between 2000 and 2016, four parameters were estimated directly from data: annual *per capita* mortality rates for males and females, and total number of males (F_Y) and females (F_X) reaching sexual maturity every year. Since F_X and F_Y depend on female overall fecundity from previous years, which is not explicitly considered in the SILD model, they were instead estimated to capture total population temporal evolution in agreement with the populations for the most important cities in Madagascar (see Table 2.2.3). Since we only had demographic data at the national level, no further differences between cities were considered. In particular, age-dependent *per capita* average rates were assumed the same across the eleven cities under study. Further details on the estimation of overall mortality rates (δ) and recruitment rates, F_X and F_Y, per year are given in appendix Appendix B.

The rates controlling female population distribution into the non-sexual and sexual worker groups (σ^0, σ_r^0, σ_1, and σ_r^1) were also adjusted to yield ratios comparable to real data, which considerably change from city to city (see Table 2.2.3).

Averages ages at which males and females are considered fully sexually active were chosen 15 and 17 years, respectively. The *aging* rate, α, is not a true demographic parameter, but defines the average time females are considered *young* by males in terms of sexual choice. In our projections, this parameter was considered to be about 15 years, this is, females are *young* when they are about between 15 and 30 years old. There is an undetermined variance around those parameter values reflecting individual heterogeneity. This is beyond the scope of the SILD model we present here.

2.2.3. Realistic Model Projections

Demographic trends are reflected in parameter values changing from year to year. When we let parameters annually change according to data, we obtain a realistic demographic trajectory. Predictions for disease spread in the different groups can then be generated by numerically integrating the

[1] http://apps.who.int/gho/data/node.home and https://www.cia.gov/library/publications/the-world-factbook/fields/2018.html

City	Inhabitants	Sexual Worker Population		HIV Prevalence		
		Minimum	Maximum	%	CI	
Antananarivo	1300000	28925	35021	1.8	0.4	4
Antsiranana	115015	2978	6812	2.9	0.8	6.5
Mahajanga	220629	2290	5637	22.7	13.9	31.2
Toamasina	274667	4289	12336	5.4	2.9	8.1
Fianarantsoa	190318	3415	7795	1.2	0	3.3
Toliary	156710	4190	11367	2.7	0.7	5.1
Taolagnaro	46000	1392	5059	4.1	0.7	9.6
Moramanga	282600	1648	2807	0.6	0	5.7
Antsirabe	238478	3677	9328	0	0	0
Morondava	123739	1648	3672	7	3.2	11.3
Nosy Be	73010	7268	14830	9.5	4.5	15.3

Table 2: Total and sexual worker populations for the main cities in Madagascar. Prevalence percentages within the worker populations are shown along with confidence interval (CI). Are these data from 2017? Source: ?

system (Eqs (1), (5)-(8)) considering that demographic parameters are no longer constant but change annually.

Initial population values correspond to the year 2000. Sex ratio is considered 1:1 (ref). Disease initial prevalence for sexual workers is set to match the data we have for each city in 200X (see Table 2.2.3). For simplicity, disease prevalences for the rest of groups are set to zero at the initial year. This gives a conservative lower bound of disease expansion within each city. A model projection is calculated from 2016 up to year 2030 taking into account the uncertainty we have in estimating each model parameter, and assuming demographic age-specific *per capita* rates follow two scenarios: either (A) they do not change on average significantly or (B) they follow the overall trend of the period 2000-2016.

3. Results

4. Discussion

Appendix A. The duration of the second infectious stage, L

After acquiring the infection individuals enter the first infectious stage, I. This first phase is highly contagious. Then, after a period of time of the order of months, individuals enter a second long phase of lower infectivity. The length of this phase is variable but it can last about 10 years. In order to take into account not only the average duration of this phase ($1/\mu$), but also its variance across individuals in the population, we take the approach of [2], which considers the L phase subdivided in n subcategories through which individuals transient up to eventually entering the terminal D stage.

$$\begin{aligned} \frac{dY_1}{dt} &= \gamma\, Y_I - \mu\, n\, Y_1 - \delta\, Y_1 \\ \cdots &= \cdots \\ \frac{dY_n}{dt} &= \mu\, n\, Y_{n-1} - \delta\, Y_n - \mu\, n\, Y_n \end{aligned} \tag{A.1}$$

When we define $Y_L \equiv \sum_{i=1}^{n} Y_i$ as the total male population in this 2nd stage, L, then notice that, if we take $n = 1$, then $Y_L = Y_1$, and the system above collapses into a single equation —see Eqs (1)— which is:

$$\frac{dY_L}{dt} = \gamma\, Y_I - \mu\, Y_L - \delta\, Y_L \tag{A.2}$$

Appendix B. Estimation of δ, F_X, and F_Y from demographic data

In the absence of disease, population dynamics in SILD model depends only on two sex-specific recruitment rates, F_X and F_Y, and two mortality rates, δ_X and δ_Y. From equation (12), it is easy to show that population change can be encapsulated only in a two-equation system:

$$\begin{aligned} \frac{dY}{dt} &= F_Y - \delta_Y\, Y \\ \frac{dX}{dt} &= F_X - \delta_X\, X \end{aligned} \tag{B.1}$$

This is non-age-structured model tracking the temporal evolution of female and male total populations. However, both fertility, which determines recruitment rates F_X and F_Y, and mortality are age-specific processes. In spite of this simplified demographic assumption, our goal here is to show that one can reasonably estimate non-age-specific model rates from age-specific data, and capture, in this way, year-to-year realistic variability and average trends.

Appendix B.1. Averge mortality rates

Human mortality is clearly an age-dependent process. However, SILD model assumes that mortality is constant, which involves that survival curves are negative exponentials, and then individual average life-span can be represented by the inverse of the mortality rate, $1/\delta$. A reasonable way to take into account the year-to-year variability and a slight sex-dependency in mortality rates is using the life-expectancy at birth L_0, from compiled life-table data, as the best estimate of a time-dependent $1/\delta$. Specifically, we consider:

$$\begin{aligned} \delta_Y(t) &= \frac{1}{L_0^Y(t)} \\ \delta_X(t) &= \frac{1}{L_0^X(t)} \end{aligned} \tag{B.2}$$

where $L_0^Y(t)$ and $L_0^X(t)$ are life table values for life-expectancies at birth of males and females, respectively, at year t (from 2000 to 2016).

Appendix B.2. Recruitment rates

SILD model F_Y and F_X parameters represent the number of males and females, respectively, starting a fully active sexual life per unit time. Let a_Y and a_X the entering age for active sexual life for males and females, respectively. If we knew the age distribution every year, the number of males and females passing from age $a_X - 1$ to a_X, and from $a_Y - 1$ to a_Y, respectively, in year t would be simply the empirical time-dependent parameters, $F_X(t)$ and $F_Y(t)$, SILD model requires. However, such a degree of detail is very difficult to gather in demographic studies. In the absence of detailed demographic data, F_Y and F_X at time t can be instead estimated from the total number of males and women born in year $t - a_Y$ and $t - a_X$, respectively, along with sex- and age-specific mortality rates. Such data have been compiled and made publicly available by the WHO and other institutions on a quinquennal basis for every country in the world at least since 2000.

Let $B_Y(t)$ and $B_X(t)$ be the total number of males and women born in year t. As we show below, these two numbers can be in turn estimated from average *per capita* fertilities and total populations in a given year t. Since the two cohorts $B_Y(t - a_Y)$ and $B_X(t - a_X)$ will suffer from age- and sex-dependent mortality up to year t, the parameters $F_Y(t)$ and $F_X(t)$ can be estimated by calculating the number of survivors until year t for those males

and women born in years $t - a_Y$ and $t - a_X$, respectively. Therefore, the parameters of interest can be estimated as:

$$\begin{aligned} F_X(t) &= B_X(t-a_X)\, s_X^{(0)}(t-a_X)\, s_X^{(1)}(t-a_X+1)\ldots s_X^{(a_X-1)}(t-1) \\ F_Y(t) &= B_Y(t-a_Y)\, s_Y^{(0)}(t-a_Y)\, s_Y^{(1)}(t-a_Y+1)\ldots s_Y^{(a_Y-1)}(t-1) \end{aligned} \quad \text{(B.3)}$$

where $s_X^{(a)}(t)$ is the survival probability during year t of individuals of age a. For instance, $s_X^{(0)}(t)$ is survival probability of new-born females during year t, who will have their one-year birthdays, on average, in year $t+1$.

Eq. (B.3) shows that the estimation of the recruitment rates F_Y and F_X requires estimating, first, survival probabilities and, second, number of new-born individuals for a given year. Let us proceed in order.

Appendix B.2.1. Survival Probabilities

In typical life tables, age-dependent mortality rates are associated to finite time intervals of a given number of years, n, usually 5 years. They are defined as:

$$m_a^n(t) \equiv \frac{D_a^n(t)}{\langle N_a^n(t)\rangle} \quad \text{(B.4)}$$

where $D_a^n(t)$ is the number of deaths in the age group between age a and age $a+n-1$ during year t and $\langle N_a^n(t)\rangle$ is the average number of individuals in this age group in year t. Survival probabilities can be estimated from this statistic under the assumption of constant mortality within each age group. The survivors at the end of a year are related to the initial number of individuals at the beginning of that year for each age group through an exponential decay:

$$N_a^n(t+1) = N_a^n(t)\exp\left(-\delta_a\right) \quad \text{(B.5)}$$

where δ_a is an age-specific instantaneous mortality rate assumed constant for all individuals within the same age group, which means, individuals between age a and age $a+n-1$ experience the same probability of death. In addition, $\langle N_a^n(t)\rangle$ can be approximated the average population in that age group:

$$\langle N_a^n(t)\rangle = \frac{N_a^n(t+1) + N_a^n(t)}{2} \quad \text{(B.6)}$$

By introducing Eqs (B.5) and (B.6) into Eq. B.4, the age-dependent mortality rate of age group a at year t, can be rewritten as:

$$m_a^n(t) = \frac{1-\exp\left(-\delta_a\right)}{1+\exp\left(-\delta_a\right)} \quad \text{(B.7)}$$

By inverting the last equation, instantaneous time-dependent mortality rates corresponding to each age group a and year t, δ_a, can be estimated from empirical age-dependent mortality rate tables, $m_a^n(t)$, as typically given in demographic surveys.

Once instantaneous death rates are estimated for each sex, year, and age group, all survival probabilities appearing in Eq (B.3) can be also estimated. Notice that a survival probability for a one-year time interval is defined as:

$$s^{(a)}(t) \equiv \frac{N_a^n(t+1) - N_a^n(t)}{N_a^n(t)} \tag{B.8}$$

which, using Eq. (B.5), leads to:

$$s^{(a)}(t) = 1 - \exp\left(-\delta_a(t)\right) \tag{B.9}$$

where the year-to-year dependency of $\delta_a(t))$ is here explicitly written, as it results from the inversion of Eq. (B.7).

Appendix B.2.2. Number of Newborns at year t

All that is now left it is the estimation of the quantities $B_Y(t)$ and $B_X(t)$, the number of male and female new-borns in the population at year t. Again detailed demographic monitoring of populations would ideally register these two quantities every year. In the absence of this direct information, we can calculate these estimates from total fertility rates, total female population, and sex ratio at birth:

$$\begin{aligned} B_X(t) &= (1-f)\, F(t)\, X^{(r)}(t) \\ B_Y(t) &= f\, F(t)\, X^{(r)}(t) \end{aligned} \tag{B.10}$$

where f is the ratio of males to females (*sez ratio*), $F(t)$ is a average *per capita* fertility rate (per year), this is, number of children an average female has per unit time, and $X^{(r)}(t)$ is the total number of reproductive females in the population at time t. For instance, Madagascar sex ratio at birth is about 1.02, while adult sex ratio is about 1.00. This means we can always estimate the total female population by dividing by two total population. The average fertility rate can be estimate from the total fertility rate. this is, the average number of children born to an average female during her whole reproductive life if she follows the birthing pattern being experienced by the overall population at year t.

Assuming that female reproductive age spans from age $a_X^{(0)}$ up to $a_X^{(1)}$, both rates can be estimated from total annual births at each age:

$$F(t) = \frac{\sum_{j=a_X^{(0)}}^{a_X^{(1)}} B^{(j)}(t)}{\sum_{j=a_X^{(0)}}^{a_X^{(1)}} X^{(j)}(t)} \tag{B.11}$$

$$b(t) = \sum_{j=a_X^{(0)}}^{a_X^{(1)}} \frac{B^{(j)}(t)}{X^{(j)}(t)} \tag{B.12}$$

where $B^{(j)}(t)$ is the annual number of births from females at age j. We assume that the previous age of entrance to active sexual life matches $a_X^{(0)}$. Notice also that $F(t)$ and $b(t)$ are not expressed in the same units. While $F(t)$ is a *per capita* annual rate, $b(t)$ is the female *per capita* fertility during her whole fertile period, this is, between ages $a_X^{(0)}$ and $a_X^{(1)}$. We can express $b(t)$ as a *per year* rate by dividing it by $(a_X^{(1)} - a_X^{(0)})$. Still the relation between these two rates is not that obvious. Since total fertilities, $b(t)$ are the ones usually provided in typical demographic studies, we estimate the annual *per capita* fertility rate, $F(t)$, from the total rate, $b(t)$, in the following way:

$$F(t) = r(t)\,\frac{b(t)}{(a_X^{(1)} - a_X^{(0)})} \tag{B.13}$$

where $r(t)$ can be calculated from Eqs (B.11) and (B.12). For our purposes, we have made the simplifying assumption that the parameter $r(t)$ is roughly constant through time.

To sum up, the SILD model parameters, F_X and F_Y, have been estimated from simple demographic data by the following equations:

$$F_X(t) = (1-f)\,r\,\frac{b(t-a_X)}{(a_X^{(1)} - a_X^{(0)})}\,\frac{N(t-a_X)}{2} s_X^{(0)}(t-a_X)\,s_X^{(1)}(t-a_X+1)\ldots s_X^{(a_X-1)}(t-1)$$

$$F_Y(t) = f\,r\,\frac{b(t-a_Y)}{(a_X^{(1)} - a_X^{(0)})}\,\frac{N(t-a_Y)}{2} s_Y^{(0)}(t-a_Y)\,s_Y^{(1)}(t-a_Y+1)\ldots s_Y^{(a_Y-1)}(t-1)$$

where $N(t)$ is total population at year t, f is the sex ratio at birth, and r is a roughly constant parameter.

In practice, r should be regarded as a free model parameter used to scale the temporal evolution of each population to approximate the real population trajectory for each city in the best way possible. This is achieved through a fitting procedure. Heuristically, the r parameter could account also for some extra population growth due to immigration from rural areas.

Appendix B.2.3. Fitting Procedure

Given demographic data, F_X and F_Y can be regarded as two functions of time, t, and a free parameter r. This can be expressed mathematically as:

$$\begin{aligned} F_X &= F_X(t, r|\mathbf{D}) \\ F_Y &= F_Y(t, r|\mathbf{D}) \end{aligned}$$

where $\mathbf{D}$ is the full set of demographic data, this is, sex-ratio at birth, ages a_X, $a_X^{(1)}$, age-dependent mortality rates, total fertility rates, and total populations for each year t.

For each city, we know the real population trajectory, $N_e(t)$, for $t = 2000$ to 2016. Note the subscript e for *empirical*. Through numerical integration of the ODE system (B.1), we can generate a model prediction for that trajectory, $N(t, r)$, for $t = 2000$ to 2016, taking as initial condition $N_e(2000)$. The value of the free parameter r is determined by minimizing the mean squared difference between empirical and model predicted population values, this is, by minimizing the following objective function:

$$G(r|\mathbf{D}) = \frac{\sqrt{\sum_{t=2000}^{2016}(N_e(t) - N(t, r))^2}}{16} \tag{B.14}$$

[1] R. M. Anderson, R. M. May, Infectious Diseases of Humans. Dynamics and Control, Oxford University Press, Oxford, 1991.

[2] A. L. Lloyd, Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics 60 (2001) 59–71.

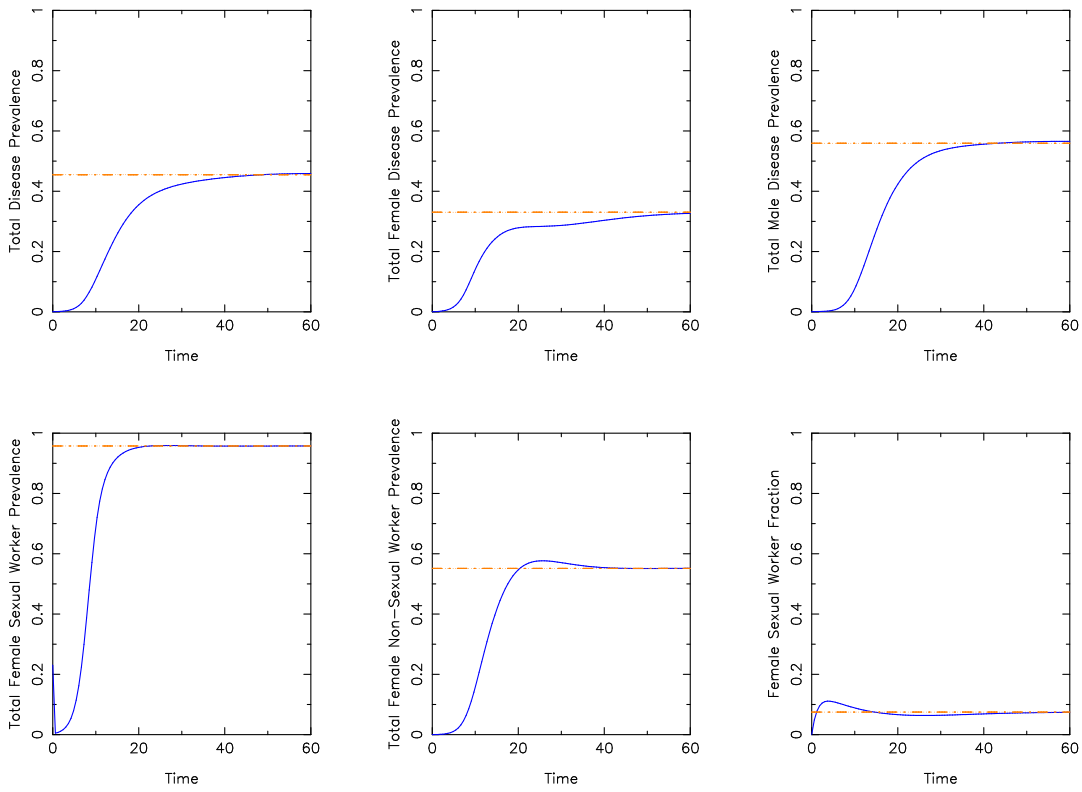

Figure 2: Temporal evolution of model HIV prevalences in men and women. Total disease prevalence is calculated with respect to the total population. Total male and female prevalences are given in relation to the total male and female populations, respectively. Total female sexual worker prevalence is the disease prevalence within the sexual worker population. Finally, the disease prevalence in the rest of women is also calculated within this group. Time in years. Horizontal lines correspond to the stationary state of the system when model parameter values are kept constant. Parameter values are given in Table 2.1.