bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Comparison of two individual-based models for simulating HIV epidemics in a population with HSV-2 using as case study Yaoundé-Cameroon, 1989-1998

Diana M Hendrickx[a,1,2], João Dinis Sousa[b,c,2], Pieter JK Libin[b,d], Wim Delva[a,b,e,f,g], Jori Liesenborgs[h], Niel Hens[a,i], Viktor Müller[j] and Anne-Mieke Vandamme[b,c]

[a] Center for Statistics, Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium
[b] KU Leuven – University of Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Clinical and Epidemiological Virology, B-3000, Leuven, Belgium
[c] Global Health and Tropical Medicine (GHTM), Unidade de Microbiologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, Portugal
[d] Artificial Intelligence Lab, Department of Computer Science, Vrije Universiteit Brussel, 1050 Brussels, Belgium
[e] The South African Department of Science and Technology-National Research Foundation (DST-NRF) Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa
[f] Department of Global Health, Faculty of Medicine and Health, Stellenbosch University, Stellenbosch, South Africa
[g] International Centre for Reproductive Health, Ghent University, Ghent, Belgium
[h] Expertise Centre for Digital Media, Hasselt University – tUL, Diepenbeek, Belgium
[i] Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium
[j] Institute of Biology, Eötvös Loránd University, Budapest, Hungary

Abstract

Mathematical modelling has been widely applied to better understand the transmission, treatment and prevention of infectious diseases. The comparison of different models is crucial for providing robust evidence for policy-makers because differences in model properties can influence their prediction.
In this study, two individual-based models for simulating HIV epidemics and sexually transmitted infection (STI) co-factor effects, i.e. models developed with the Simpact Cyan 1.0 and StepSyn 1.0 frameworks, were compared. Simpact Cyan 1.0 uses a continuous-time implementation and is updated each time an event happens, while StepSyn 1.0 uses discrete daily or weekly time-steps and updates population, sexual links, and STI states at each time-step. Furthermore, there are differences in how stochastic processes are described, how individuals enter and leave the population, and in the formation and break-up of ties in the sexual network. We evaluated how these differences would affect survival and HIV prevalence curves during the course of a heterosexual HIV epidemic with Herpes simplex virus-2 as contributing STI, using as case study Yaoundé (Cameroon), 1989-1998. To allow such direct comparison, neither modelling framework used its full potential. For each model, 100 simulations were performed with parameters calibrated to HIV prevalence data from Yaoundé between 1989-1998. The median time profile for both models matched the data equally well, but a slight difference in variability among simulations could be observed. This can be explained by differences in model initialization and the generation of the sexual network. Moreover, our example shows that similar HIV prevalence curves can be simulated using a different combination of sexual network and HIV transmission parameters.
We conclude that it is important to carefully consider differences in model properties, here in particular their initialization and the generation of sexual networks. We strongly recommend using different models to assess the robustness of evidence provided by mathematical modelling.

Keywords: model comparison, individual-based modeling, HIV, sexually transmitted infection co-factor, prevalence curves

1. Introduction

Mathematical modelling has been widely applied to better understand the transmission, treatment and prevention of infectious diseases. The role of mathematical models in understanding the dynamics of the human immunodeficiency virus (HIV) epidemic has been recently reviewed by Geffen & Welte [1]. They present examples of HIV models that were used for estimating the size of the epidemic in specific subpopulations such as the Black population in South Africa and for estimating the impact of interventions such as the use of anti-retrovirals and condoms to reduce HIV and AIDS.

[1] Corresponding author at: Center for Statistics, Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium
E-mail address: diana.hendrickx@uhasselt.be
[2] Authors contributed equally.

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

When considering mathematical models, two commonly used types of implementation can be distinguished: compartmental and individual-based models (IBMs). While compartmental models simulate population counts, IBMs (also called agent-based models or micro-simulation models) keep track of the history of each individual in the population separately.

Recent applications of compartmental HIV models include testing the effect of different assumptions for HIV dynamics on the predicted impact of anti-retroviral therapy (ART) in men-having-sex-with-men (MSM) [2], and the study of the influence of concurrent partnerships on HIV dynamics [3]. IBMs have been recently applied to assess the influence of pre-exposure prophylaxis (PrEP) in MSM [4], to evaluate the long-term effect of early ART initiation [5] and to understand the factors underlying the emergence of HIV in humans [6].

In contrast to compartmental models, IBMs allow a high degree of individual heterogeneity [7] (e.g. regarding (sexual) risk behaviour, age demography and individual response to treatment). This is an advantage when heterogeneity matters for the particular question/process that is studied. Because individual heterogeneity is inherent in transmission, prevention and treatment of HIV and other sexually transmitted infections (STI), IBMs are particularly suitable to estimate the most beneficial intervention for specific individuals. Furthermore, IBMs allow for an explicit modelling of the sexual relationships that jointly form the sexual network over which STI/HIV infections are transmitted [8].

Model comparisons are crucial for providing robust evidence for decision-making in public health and policy purposes [9]. To assess uncertainty, it is necessary to study how differences in model properties influence their prediction. Eaton *et al*. compared ten mathematical models aimed at studying HIV prevalence, incidence and ART [10]. However, eight of these models were compartmental, and the majority of these models did not include the co-factor effects of other STIs. The remaining two models were discrete time IBMs.

This study presents and compares two IBM frameworks for simulating HIV transmission dynamics with STI co-factor effects. The purpose of the comparison was to assess how the differences between the two modelling frameworks affect the simulation of an HIV epidemic influenced by an STI co-factor. Model options that were absent in one of the frameworks had to be turned off in the other framework to allow direct comparison. For this purpose we chose the HIV epidemic in Yaoundé (Cameroon) and we will refer to the models built in the two frameworks as Simpact Yaoundé and StepSyn Yaoundé. As representative use cases, we compare survival curves and fit both models to HIV prevalence time series from Yaoundé.

2. Methods

2.1. Simpact Cyan 1.0 modelling framework

Simpact Cyan 1.0[3] [11] is a freely available framework for developing IBMs to simulate HIV transmission, progression and treatment. Models developed with Simpact Cyan 1.0 are event-driven, which means that the models are not updated at fixed time-intervals, but at every time an event happens, making Simpact Cyan 1.0 a continuous-time simulation modelling framework. Events occur as a result of stochastic event processes, described by hazard functions. The timing of an event is determined using the modified next reaction method (mNRM) [12].

To initialize a model, a number of individuals are generated. The age of each person is drawn from an age distribution based on a population pyramid, and when the age is larger or equal to the sexual debut age, a person is marked as sexually active.

In Simpact Cyan 1.0, the user can specify which events are possible in the simulation, depending on the research question he or she wants to solve. The possible events are: heterosexual and homosexual relationship formation and dissolution, conception and birth, HIV transmission, AIDS and non-AIDS mortality, HIV diagnosis and treatment, HIV treatment dropout, and STI co-factor transmission events. Furthermore, events describing the natural history of HIV are implemented in Simpact Cyan 1.0. For each event, the form and the parameters of the hazard function can be modified flexibly, depending on the research question one wants to answer.

Simpact Cyan 1.0 is implemented in C++ with R [13] and Python [14] interfaces.

2.2. StepSyn 1.0 modelling framework

StepSyn 1.0 is an IBM modelling framework that simulates epidemics of STIs, including HIV. It focuses on the epidemiological synergy and interactions between HIV and other STIs. Time is divided into fixed one-day or one-week intervals. In this study, we use only weekly time steps to speed up computational time. Formation and dissolution of sexual relationships, and STI and HIV transmission are modelled as stochastic processes. For each

[3] https://simpactcyan.readthedocs.io/en/latest/index.html

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

STI, life history is explicitly modelled, including several stages, recurrences, genital ulcers, and urethral discharge symptoms. The effects of these symptoms on the HIV transmission probability are explicitly modelled. The parameters for the timing and duration of the STI stages and symptoms, and the co-factor effects of genital ulcers and discharge on HIV transmission are based on a literature review. To enable comparison of the two models, in the present study, StepSyn is parameterized to track only HIV and Herpes simplex virus type 2 (HSV-2), and the latter virus is modelled without explicit stages or recurrences, and with the probability of transmission and co-factor effect on HIV both constant and not dependent on recurrences. Heterosexual relationships can include marital and short-term links, and contacts between commercial sex workers (CSW) and clients. In the present study CSWs are not included. Individuals vary in their tendency to form short-term relationships, so that their number of non-marital partners within a year would follow a power law distribution. The parameters of the latter were derived from behavioural data gathered in the 4 Cities Study [15,16]. StepSyn is implemented in R [13].

2.3. Key model differences

The Simpact Cyan 1.0 and StepSyn 1.0 modelling frameworks were developed separately. Similarities and differences between the two modelling frameworks are outlined in Table 1. Model options that were absent in one of the frameworks were turned off in the other framework to allow comparison (see Table 2). In Simpact 1.0 model options were turned off in the parameter configuration file (R file calling the C++ source code), while in StepSyn both changes in the parameter files and R source code were necessary.

Simpact Cyan 1.0 is freely available. StepSyn is not yet freely available but will become so upon publication of a manuscript that describes the modelling framework into more detail (manuscript in preparation). Models developed with Simpact Cyan 1.0 and StepSyn 1.0 are both individual-based and both include HIV natural history. While StepSyn models can also include the natural history of other STIs, Simpact Cyan 1.0 models are restricted to a generic, constant STI co-factor effect on HIV. Within the StepSyn 1.0 framework, five different STIs (other than HIV) can be simulated but for the purpose of the present study only HSV-2 and HIV were simulated (Table 2). Furthermore, StepSyn Yaoundé was ran without modelling HSV-2 recurrences, considering only the chronic stage as infective and using a constant co-factor effect on HIV.

Models developed with Simpact Cyan 1.0 and StepSyn 1.0 are both accessible from R, but they differ in the way they are implemented. While the source code for StepSyn is R code, Simpact Cyan 1.0 is implemented in C++ with interfaces to Python and R. In this way, Simpact Cyan 1.0 combines the computational efficiency of C++ with the user-friendliness of R and Python.

Simpact Cyan 1.0 and StepSyn 1.0 also differ in the way the state of the model system is updated and, as a consequence, in the way stochastic processes are described. StepSyn implements discrete time steps of either one day or one week (one week in this study) and updates population, sexual links, and STI states at each time step. Stochastic processes are implemented through R random distribution functions which are called in each time step. Simpact Cyan 1.0 implements a continuous time model and the state of the model system is updated each time an event happens, and therefore requires the use of hazard functions. The hazard at a certain time point is defined as the event rate at that particular time point, conditional on model states such as still being alive at that time point.

An advantage of the continuous time implementation is that events that happen on different time scales can be integrated into a single simulation. A disadvantage is that model parameters in the literature are mostly described as probabilities, and as a consequence have to be converted to hazards. Another drawback is that continuous time models are computationally more complex, and therefore require a longer simulation time.

Another difference between both modelling frameworks is the way individuals can enter and leave the population. In Simpact Cyan 1.0, individuals can enter the population by birth, and leave the population by non-AIDS or AIDS mortality. The mortality events are age-dependent. StepSyn 1.0 is not age-structured; individuals enter the population by immigration or sexual debut and leave the population by emigration, AIDS- or non-AIDS age-independent mortality. In this study, the age-dependent parameters of Simpact Cyan 1.0 were set to 0, no new individuals enter the population, and individuals only leave the population by AIDS-mortality. In this study, StepSyn parameters are set so that immigration, emigration, sexual debut, and background mortality are set to 0, so that individuals only leave the population by AIDS mortality.

An extra option in Simpact Cyan 1.0 that is not implemented in StepSyn 1.0 is pregnancy. In Simpact Cyan 1.0, women are pregnant between a conception event and a birth event. This means that after a simulation run, one can measure how many women were pregnant at any given point in time in the simulation.

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Comparison of properties of models developed with Simpact Cyan 1.0 and StepSyn 1.0 (general).

Model property	Simpact Cyan 1.0	StepSyn 1.0
Simulation framework	Individual-based model	Individual-based model
Implementation language	C++ with R and Python interfaces	R
Source code freely available	Yes	After publication of the manuscript describing the modelling framework in detail
Update of the state of the model system	Event-driven: the state of the model system is updated each time an event happens	Fixed time steps of either one day or one week, with creation and dissolution of sexual relationships and STI and HIV transmission
Description of stochastic processes	By hazard functions	By random distribution functions called in each time step
Age-structured	Yes	No
Entering the population	Birth	Sexual debut, immigration
Leaving the population	AIDS- and non-AIDS mortality	AIDS- and non-AIDS mortality, emigration
Formation and break-up of relationships within the sexual network	Formation and dissolution event. Timing of events sampled from probability distribution resulting from carefully chosen hazard functions.	Each time step, using an individual-specific probability of forming a new relationship
Individual variation in sexual behaviour	In the formation hazard, the eagerness of a person to form a relationship and the preferred age gap are used, to allow for individual variation in sexual behaviour. Both values are drawn from user-specified probability distributions.	Number of short term relationships follows a power law distribution
HIV natural history	Yes	Yes
Other STIs as co-factor for HIV infection	Yes (implemented as a simplified STI transmission hazard)	Yes (syphilis, chancroid, gonorrhoea, chlamydia, HSV-2)
Natural history – other STIs	No	Yes, with genital ulcers and discharge explicitly modelled
Pregnancy	Yes, as the period between a conception event and a birth event.	No
Diagnosis / treatment of HIV	Yes	No
HIV seeding	A fraction to specify the probability of each person in the group to be a seeder, or an amount to specify the amount of seeders. Fraction or amount is taken from the number of people in the population.	Either a random binomial for HIV seeding (seeding varies between simulations), or a fixed fraction to specify the fraction of seeders; this fraction is taken from the number of males and the number of females separately.
Changing simulation settings during the simulation	Yes, through one or more intervention events.	No, the simulation setting cannot be changed during the simulation.
Computational time for 10,000 simulations*	~35 hours	~5 hours (using weekly steps)

*on 100 cores of the VSC (Xeon E5-2680v2 CPUs 2.8 GHz, 25 MB level 3 cache). Population size ~ 9000; simulation time = 35 years; parameters: see Table S1

Formation and break-up of relationships in the sexual network are implemented differently in StepSyn 1.0 and Simpact Cyan 1.0. In Simpact Cyan 1.0, formation and break-up of relationships happens through formation and dissolution events. The timing of these events is sampled from a probability distribution that emerges as a result of the hazard function that was specified. In StepSyn, formation and break-up of relationships is performed each time step (week in the current study), using an individual-specific probability of forming a new relationship.
An extra option in Simpact Cyan 1.0 that is not currently implemented in StepSyn 1.0 is the simulation of diagnosis and treatment, including treatment dropout, and the possibility to introduce treatment during the simulation through a so-called intervention event.

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 2. Model properties adapted in this study to allow comparison between Simpact Cyan 1.0 and StepSyn 1.0.

Model property	Simpact Cyan 1.0	StepSyn 1.0
Age-structured	No (all age-related parameters are equal to 0)	No
Entering the population	No new individuals entering the population during the simulation. All individuals are sexually active.	No new individuals entering the population during the simulation.
Leaving the population	AIDS mortality only	AIDS mortality only
Other STIs as co-factor for HIV infection	Simplified STI (HSV-2) hazard	HSV-2 (simplified HSV-2 transmission probability, not dependent on recurrences)
Natural history – other STIs	No	No
Pregnancy	No (no birth events occur during the simulation)	No
Diagnosis / treatment of HIV	No	No
HIV seeding for fitting parameters	Fixed amount of seeders	Fixed fraction of seeders
HIV seeding for generating 100 simulations with a fixed set of parameters	Amount of seeders. To vary the HIV seeding among simulations, the amount of seeders is sampled from a binomial distribution.	Random binomial for HIV seeding, seed varies between simulations
Changing simulation settings during the simulation	No	No

2.4. Data - city of Yaoundé

2.4.1. Data available for generating the sexual network in StepSyn 1.0

Since StepSyn 1.0 models explore behavioural differences between individuals, behavioural data on the distribution of the number of non-marital partners in the last 12 months were gathered for Yaoundé during the development of the StepSyn 1.0 modelling framework. These data were obtained from the surveys of the 4 Cities Study, and provided by the Institute of Tropical Medicine (ITM) in Antwerp, Belgium (Anne Buvé, unpublished database). These data were used to estimate the parameters of a power law distribution to generate the sexual network for the StepSyn Yaoundé model (see paragraph 2.5). This work was performed prior to the comparison study.

2.4.2. Data available for comparison of Simpact Cyan 1.0 with StepSyn 1.0

For the comparison of the two models, demographic data for 1997, and HIV prevalence data for 1989–1998 related to the African city of Yaoundé (Cameroon), are used. The adult male population for that year was estimated at 387,398, in the 4 Cities Study [17]. Based on the surveys made by the same study, 34.5% of men and 44.2% of women were married; and 7.2% of married men were polygamous [16]. Assuming most of the latter had 2 wives, we estimate the number of women as $387{,}398 * 0.345 * (1+0.072) / 0.442 = 324{,}152$, implying an adult sex ratio (M:F) of 1000:836.74.

Since one of our aims here is to compare HIV prevalence curves resulting from fitting HIV transmission parameters of our two models to data, we gathered HIV-1 prevalence data for Yaoundé for the period 1989–1998. The only prevalence data available was of pregnant women, for whom prevalence increased from 0.7% in 1989 to 5.5% in 1998 (US Census Bureau, 2001, HIV/AIDS Profile, Cameroon, HIV/AIDS Surveillance Database [18]) as shown in Table 3.

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 3: HIV-1 prevalence data for Yaoundé's pregnant women. Source: US Census Bureau, 2001, HIV/AIDS Profile, Cameroon, HIV/AIDS Surveillance Database [18].

Year	HIV-1 prevalence (%)
1989	0.71
1990	1.32
1991	2.10
1992	1.91
1993	1.30
1994	3.00
1995	2.72
1996	4.81
1998	5.51

Because StepSyn 1.0 does not implement pregnant women as a separate category of the population, and we wanted to compare model simulations for males and females separately, the prevalence rates for pregnant women were converted to prevalence rates for males and females as follows. In 1997, during the 4 Cities Study, HIV-1 prevalence was 4.1% in men and 7.8% in women [19]. The prevalence for pregnant women in 1997 was estimated by fitting a smoothing spline through the data from Table 3, using the smooth.spline function of the stats package in R, and evaluating the spline at the time point corresponding with 1997 (see Figure S1). We obtained a prevalence of 4.778 % for pregnant women in 1997, and calculated the prevalence ratio men/pregnant women (resp. women/pregnant women) by dividing 4.1 (resp. 7.8) by 4.778. The prevalence ratios (0.858 for men and 1.632 for women) were multiplied with the data from Table 3 to obtain HIV prevalence for men and women separately for all years between 1989 and 1998 (see Table 4).

Table 4. Estimated HIV-1 prevalence data for Yaoundé's men and women.

Year	HIV-1 prev (%) men	HIV-1 prev (%) women
1989	0.609	1.159
1990	1.133	2.154
1991	1.802	3.427
1992	1.639	3.117
1993	1.115	2.122
1994	2.574	4.896
1995	2.334	4.439
1996	4.127	7.850
1997	4.100	7.800
1998	4.728	8.992

2.5. StepSyn 1.0 sexual network

Power law distributions (separately for men and women) were fitted to the data on the reported number of partners in the last 12 months taken from the 4 Cities Study. Each individual is assigned a preferred degree (number of partners within a period of 12 months) drawn from the appropriate distribution. The individual-specific preferred degrees (PD_i) are translated into probabilities of forming a new short term relationship per week (PF_i) using the formula $PF_i = PD_i / (MD + 52)$, with MD being the mean duration of short term relationships in weeks. With these parameters, in each week, the total male demand for new relationships is higher than the female one, partly because of the male-biased sex ratio, and partly because of female underreporting of number of partners in the survey that serves as basis for the distribution used. The extra male demand is uniformly distributed between the females, a method that is acceptable for modelling purposes [20]. The number of weekly sex acts in married couples and in short-term relationships is calculated by applying two different Poisson distributions, so that their means would be consistent with the values found in the 4 Cities Study for Yaoundé [15,16].

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2.6. Simpact 1.0 sexual network

In Simpact Cyan 1.0, formation and dissolution of relationships in the sexual network are simulated as discrete events. For man-woman pairs of sexually active persons, formation events are scheduled. When a formation event is triggered, a sexual relationship is established, and a dissolution event is scheduled. When the dissolution event is triggered, the relationship ceases to exist.
Formation and dissolution events are described by hazards. In the description of the hazards, parameters can be included describing dependence of relationship formation and dissolution on the number of relationships the person already has, the age of the person and the preferred age gap between the partners.
In this comparison study, the default settings of Simpact Cyan 1.0 for the formation and dissolution hazard were applied (the developers of Simpact Cyan 1.0 did not have access to the behavioural data described in paragraph 2.4.1).
Because Simpact Cyan 1.0 keeps track of the history of each individual, we are able to reconstruct the sexual network from the model output. The network can be reconstructed at a given time point by including all relationships a person has at that time point for each person in the population. Furthermore, a cumulative network over a certain period of interest can be reconstructed, including all relationships between individuals in the population during that period (e.g. all relationships people had during the past 12 months).

2.7. Converting transmission probability parameters to hazard parameters

While in the majority of the literature, and also in StepSyn 1.0, transmission parameters are described as probabilities, the parameters of Simpact Cyan 1.0 are described in terms of hazards. Transmission probability parameters were converted to hazard parameters using the following formula [21]:

$$F(t) = 1 - \exp\left(-\int_0^t \lambda(x)dx\right)$$

where $F(t)$ is the cumulative distribution function and $\lambda(x)$ is the hazard function.
An example of conversion of a transmission probability parameter to a hazard parameter is presented in the Supplementary Material.

2.8. Parameter fitting methodology

For four of the HIV transmission parameters in StepSyn Yaoundé, there exists a corresponding set of parameters in Simpact Yaoundé (see Table 5). These four HIV transmission parameters for both models (see Table 7) were fitted to the data in Table 4 by applying an iterative active learning approach [22] using the procedure described in [23] and minimizing the sum of squared relative errors [24] to determine model performance. The remaining model parameters were drawn from the literature.

Table 5. Relationship between HIV transmission parameters in Simpact Cyan 1.0 and StepSyn 1.0. Full name of the parameters in the models: 1) Simpact Cyan 1.0: hiv.a: hivtransmission.param.a; hiv.f1: hivtransmission.param.f1; hiv.e1: hivtransmission.param.e1; hiv.e2: hivtransmission.param.e2. 2) StepSyn 1.0: baseline: hiv.tr.prob.m2f.baseline; fmratio: hiv.tr.prob.baseline.f.m.ratio; hsv2.index: hiv.tr.prob.baseline.mult.hsv2.chronic.index; hsv2.exposed: hiv.tr.prob.baseline.mult.hsv2.chronic.exposed.

Parameter	Simpact Cyan 1.0 (hazard)	StepSyn 1.0 (probability)
transmission parameter M → F	hiv.a + hiv.f1	baseline
transmission parameter F → M	hiv.a	baseline x fmratio
transmission parameter F → M in case the HIV-infected person has HSV-2	hiv.a + hiv.e1	baseline x fmratio x hsv2.index
transmission parameter F → M in case the HIV-uninfected person has HSV-2	hiv.a + hiv.e2	baseline x fmratio x hsv2.exposed

For each of the four HIV transmission parameters, values were drawn from a uniform distribution, using the ranges from Table S1. These uniform distributions were used as initial (prior) probability distributions for the parameters in the parameter fitting procedure. We applied Latin Hypercube Sampling (LHS) [25] to select 10,000 parameter sets.

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In brief, the parameter fitting procedure consists of studying the subset of model simulations that corresponds to the top 1% of lowest values of the sum of squared relative errors. First, a parameter wise comparison between the density of the initial probability distribution of the parameters and the density of the probability distribution of the subset of parameters corresponding with the top 1% solutions is conducted in order to determine which parameters are highly influenced by the data. Second, classification trees and generalized additive models are applied to determine which patterns of parameter vectors characterize the subspace of the top 1% solutions. Finally, we apply the Maximal Information Coefficient (MIC) [26] to determine associations between the parameters in the subspace of the top 1%.
Based on the results of the analyses above, we can narrow the solution space and repeat the steps described above several times.
A more detailed description of the parameter fitting methodology is available in the Supplementary Material.
The R-scripts that have been used for fitting HIV transmission parameters are available from GitHub[4].

3. Results

3.1. Survival curves

To compare how both models simulate the progression of AIDS to death, we simulated survival curves for both models, setting the HIV transmission parameters to zero and using the literature parameter values in Table 6. For each model, 100 simulations were performed by initializing the models with approximately 2% of HIV-infected people, and assuming no HIV transmission. The results show that the two models are able to simulate similar survival curves (see Figure 1).

Table 6. Parameter values used in Simpact Yaoundé and StepSyn Yaoundé for simulating the HIV survival curves.

Parameter	Simpact Cyan 1.0	StepSyn 1.0	Reference
simulation time(years)	population.simtime = 25	years.sim = 25	---
sex ratio	birth.boygirlratio = 0.54 (proportion males)	5:4.184 (M:F)	[16,17]
duration acute stage	chronicstage.acutestagetime = 12/52 years	hiv.strain1.acute.dur.min <- 12 weeks hiv.strain1.acute.dur.max <- 12 weeks	[27]
time of survival	from a distribution file, including a frequency table for 100 samples drawn from the distribution of the survival time defined by the parameters for StepSyn Yaoundé	hiv.strain1.aids.progr.mean.time <- 468 weeks hiv.strain1.aids.progr.start.time <- 104 weeks AIDS might start after 2 years mean time to AIDS ~9 years (only ~1% of infected people develop AIDS in the first 2 years)	[28,29]
death from AIDS after 1 year	aidsstage.start = 1	hiv.strain1.aids.dur <- 52 weeks (sources say it is between 9 and 15 months)	[28,30]
HIV seeding	an amount of seeders, drawn from a binomial distribution with a probability of 0.02 (2%)	drawn from a binomial distribution with a probability of 0.02 (2%).	---

[4] https://github.com/dmhendrickx/Scripts_comparison_Simpact_StepSyn

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

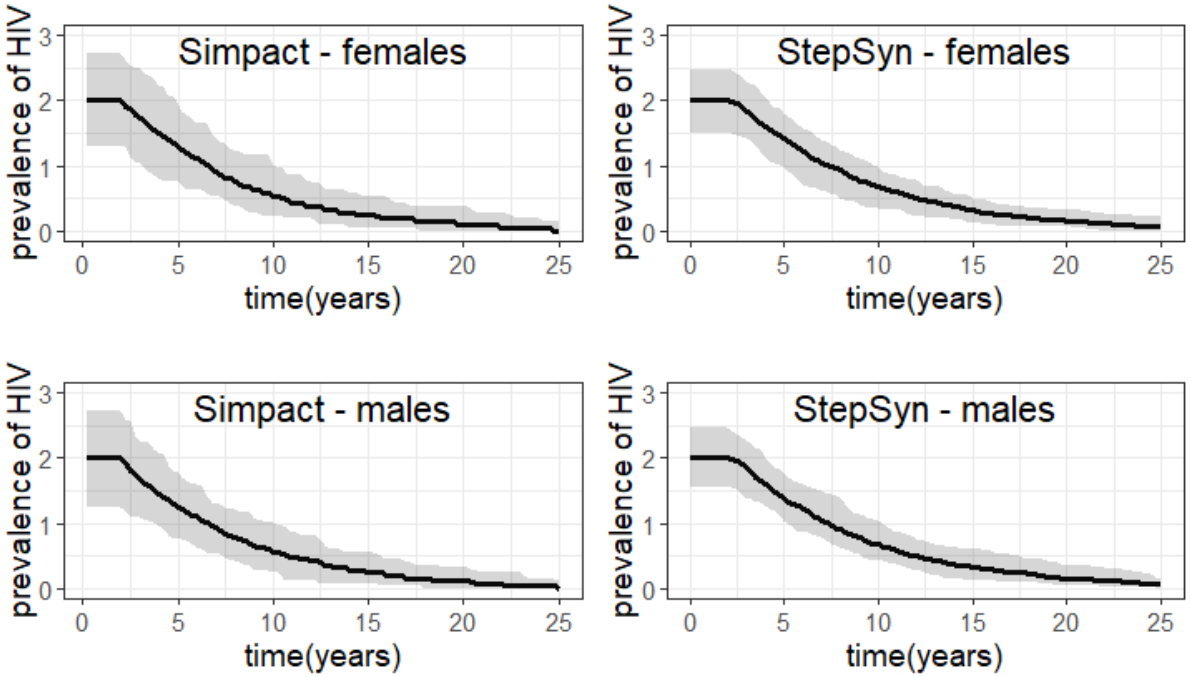

Fig. 1. HIV survival curves for Simpact Yaoundé (left) and StepSyn Yaoundé (right) in absence of transmission and changes in population size except death by AIDS– median HIV prevalence (in %) of 100 simulations and 100% percentile (shaded area).

3.2. HIV prevalence curves

Both models were used to simulate HIV prevalence in Yaoundé between 1989 and 1998.

Before fitting the HIV parameters, we adapted the parameters for the STI (HSV-2) co-factor. In major epidemiological reviews of HSV-2 [31-33] the only data about Yaoundé that is mentioned is the data collected by Buvé *et al.* [34] in the 4 Cities Study, in which the HSV-2 seroprevalence was 50% in females. Thus, we do not have information about the temporal trends of HSV-2 in Yaoundé. However, since the prevalence in Africa seems to have slightly declined between 2003 and 2012 [31] we assumed that the measured prevalence in Yaoundé, in 1997 [34], reflects an epidemic not far from its peak. Accordingly, we adapted the parameters so that the seroprevalence of this virus first increases and afterwards stabilizes at approximately 50% for females (see Figure 2), corresponding to the HSV-2 prevalence in 1997 described by Buvé *et al.* [34]. The corresponding parameters for the STI co-factor are presented in Table 8. An overview of the fitted HIV transmission parameters can be found in Table 7.

For each model, Figure 3 compares the median (with 100% percentile) of 100 simulations with the fitted parameters to the estimated HIV-1 prevalence data for Yaoundé's men and women in Table 4. Figure 4 presents the median (with 100% percentile) HIV prevalence rate over time during the period 1980–2005. For both simulation models, the median of the 100 simulations matches the data equally well. Furthermore, model predictions for the HIV prevalence rate in the period 1998–2005 are similar for both models.

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 7. Fitted HIV transmission parameters for StepSyn Yaoundé and Simpact Yaoundé (see paragraph 2.7. for parameter fitting methodology). The corresponding parameter value in the other model, calculated with the formula in paragraph 2.7, is presented between brackets. Full name of the parameters in the models: 1) Simpact Cyan 1.0: hiv.a: hivtransmission.param.a; hiv.f1: hivtransmission.param.f1; hiv.e1: hivtransmission.param.e1; hiv.e2: hivtransmission.param.e2. 2) StepSyn 1.0: baseline: hiv.tr.prob.m2f.baseline; fmratio: hiv.tr.prob.baseline.f.m.ratio; hsv2.index: hiv.tr.prob.baseline.mult.hsv2.chronic.index; hsv2.exposed: hiv.tr.prob.baseline.mult.hsv2.chronic.exposed.

Parameter	Simpact Yaoundé transmission hazard (*corresponding transmission probability*)	StepSyn Yaoundé transmission probability (*corresponding transmission hazard*)
HIV transmission baseline	hiv.a = -3.01 (F to M) (*baseline = 0.0035 M to F*)	baseline = 0.0045 (M to F) (*hiv.a = -2.66 F to M*)
gender ratio	hiv.f1 = 1.31 (M to F) (*fmratio = 0.27 F to M*)	fmratio = 0.30 (F to M) (*hiv.f1 = 1.21 M to F*)
influence of HSV-2 in HIV-infected person (index)	hiv.e1 = 0.89 (*hsv2.index = 2.43*)	hsv2.index = 1.37 (*hiv.e1 = 0.32*)
influence of HSV-2 in HIV-uninfected person (exposed)	hiv.e2 = 1.37 (*hsv2.exposed = 3.92*)	hsv2.exposed = 2.64 (*hiv.e2 = 0.97*)

Table 8. Other parameters used for simulating HIV prevalence curves.

Parameter	Simpact Yaoundé (transmission hazard)	StepSyn Yaoundé (transmission probability)	Reference
HSV-2 transmission - baseline	person.hsv2.a.dist.fixed.value = -2.550	hsv2.tr.prob.chronic.m2f = 0.003	corresponds to a HSV-2 prevalence of ~50% in females in 1997 [34]
HSV-2 transmission – gender ratio	hsv2transmission.hazard.c = 0.694	hsv2.tr.prob.f.m.ratio = 0.5	[37,38]
HIV seed in 1980	for fitting: hivseed.amount equal to 0.5% of the population 100 simulations: seed drawn from a binomial distribution with probability = 0.005	for fitting: fixed fraction of 0.5% of the population 100 simulations: seed drawn from a binomial distribution with probability = 0.005	leads to a HIV prevalence of 0.6% for males and 1.2% for females in 1989 (values in Table 4)

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

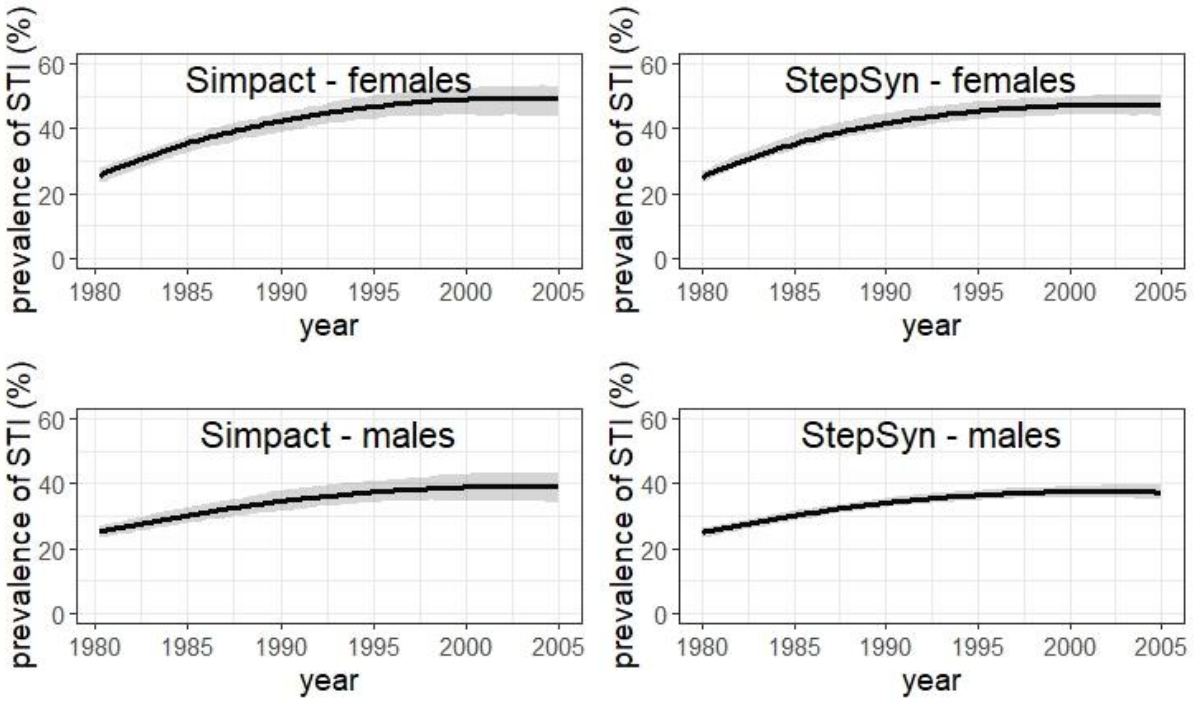

Fig. 2. Prevalence curves for STI (HSV-2) co-factor for the period 1980–2005. Median STI prevalence (in %) of 100 simulations and 100% percentile (shaded area). Left: Simpact Yaoundé; right: StepSyn Yaoundé.

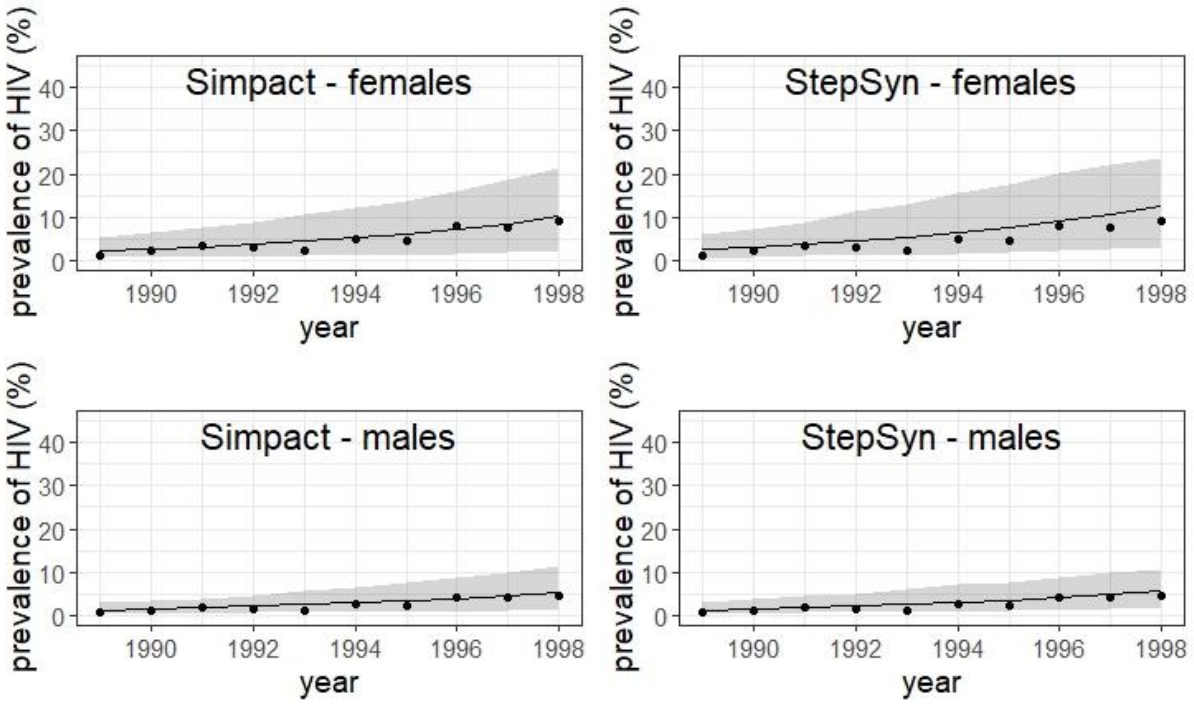

Fig. 3. Median HIV prevalence (in %) of 100 simulations with the fitted parameters (solid line) and 100% percentile (shaded area) for the period 1989–1998 (9–18 years into the simulation). The points represent the estimated HIV-1 prevalence data for Yaoundé's men and women in Table 4. Left: Simpact Yaoundé; right: StepSyn Yaoundé.

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

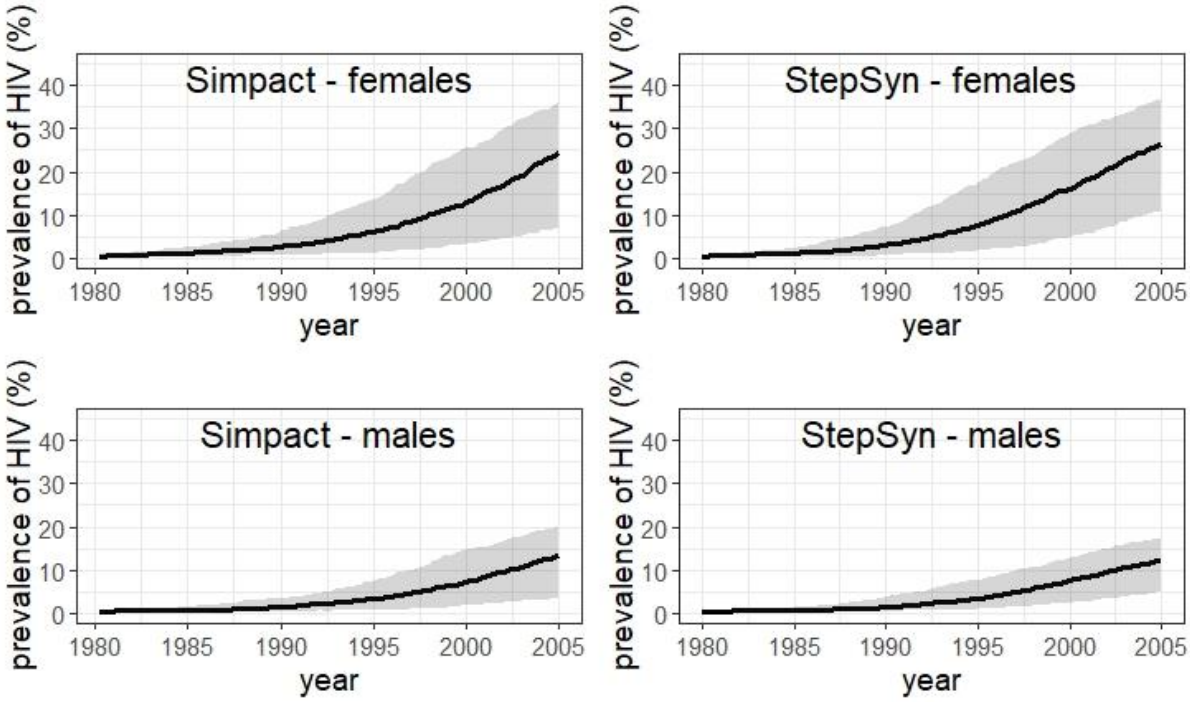

Fig. 4. Prevalence curves for HIV from 1980 to 2005. Median HIV prevalence (in %) of 100 simulations and 100% percentile (shaded area). Left: Simpact Yaoundé; right: StepSyn Yaoundé.

4. Discussion

In this study, two agent-based models were compared based on their model properties and their ability to simulate survival curves and the course of the HIV epidemic in Yaoundé, Cameroon. Model features that were absent in one of the frameworks were turned off in the other one to allow comparison using a comparable set of parameters. The two models generate similar survival curves for the same initial conditions (equal HIV seeds) (Figure 1), but the survival curves for Simpact Cyan 1.0 show more variation than the ones for StepSyn 1.0. This can be explained by the fact that the percentage HIV seeds is taken randomly from the whole population for Simpact Cyan 1.0, and consequently the percentage HIV seeds for males and females separately can vary. In contrast, for StepSyn 1.0, the percentage of HIV seeds for males and females is the same (see Figure 5, which shows 100 survival curves for a constant HIV seed of 2%).

The slightly larger variation observed in the HSV-2 and HIV prevalence curves in Figure 2 and 4 respectively can be explained in the same way.

When calibrating the models using the same data, the STI (HSV-2) co-factor parameters are estimated higher for Simpact Cyan 1.0 than for StepSyn 1.0, while the baseline parameter is estimated lower (see Table 7). This can be explained by the difference in how the sexual network was generated. When comparing the distribution of the number of partners for both models (see Figure 6), it can be observed that in StepSyn Yaoundé some individuals have a very high promiscuity (> 10 concurrent partnerships), which is caused by using a power law distribution for generating the sexual network. In Simpact Yaoundé, the maximum number of concurrent partnerships is lower (max. 6 partners for the example in Figure 6). In summary, the example of the HIV epidemic in Yaoundé shows that similar HIV prevalence curves can be simulated using different combinations of sexual network and HIV transmission parameters.

Both modelling frameworks have several advantages and limitations. Simpact Cyan 1.0 is implemented in C++, a programming language with better performance in terms of speed compared to R. The C++ code is linked to an R (and Python) user interface, so that people can use Simpact Cyan 1.0 without knowledge of C++ and take advantage of the strengths of R (e.g. user-friendly for performing data analysis and visualization) [39].

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

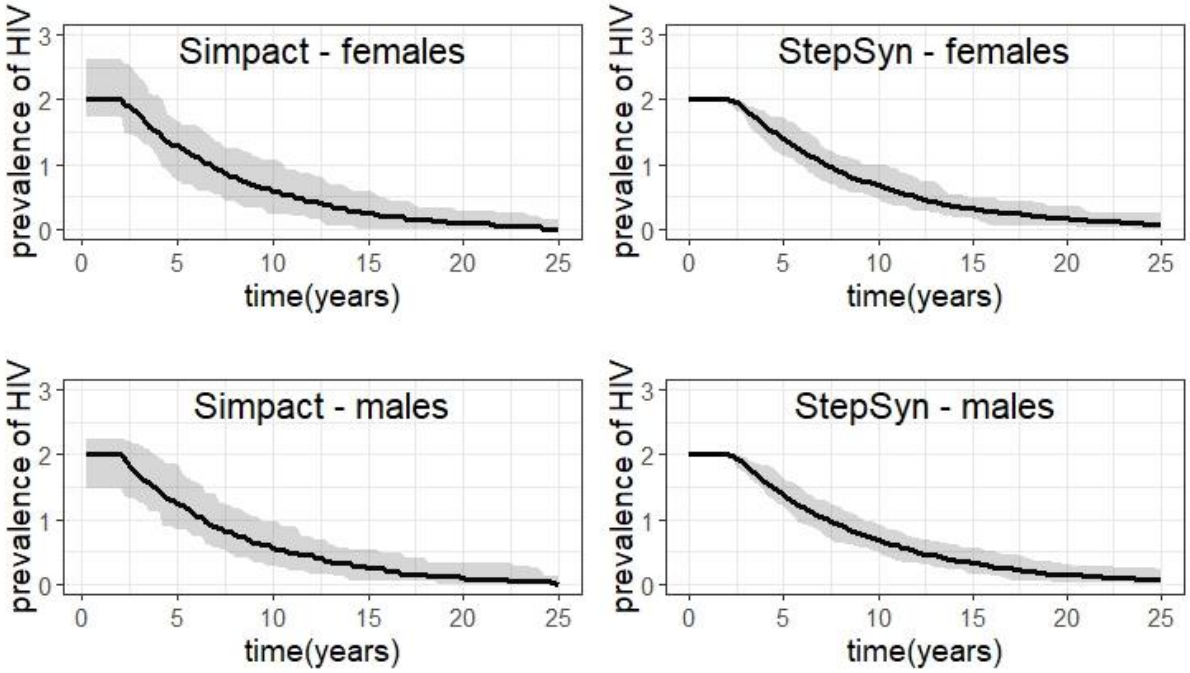

Fig. 5. HIV survival curves for Simpact Yaoundé (left) and StepSyn Yaoundé (right) in absence of transmission and changes in population size except death by AIDS, and in case of a constant HIV seed of 2% – median HIV prevalence (in %) of 100 simulations and 100% percentile (shaded area).

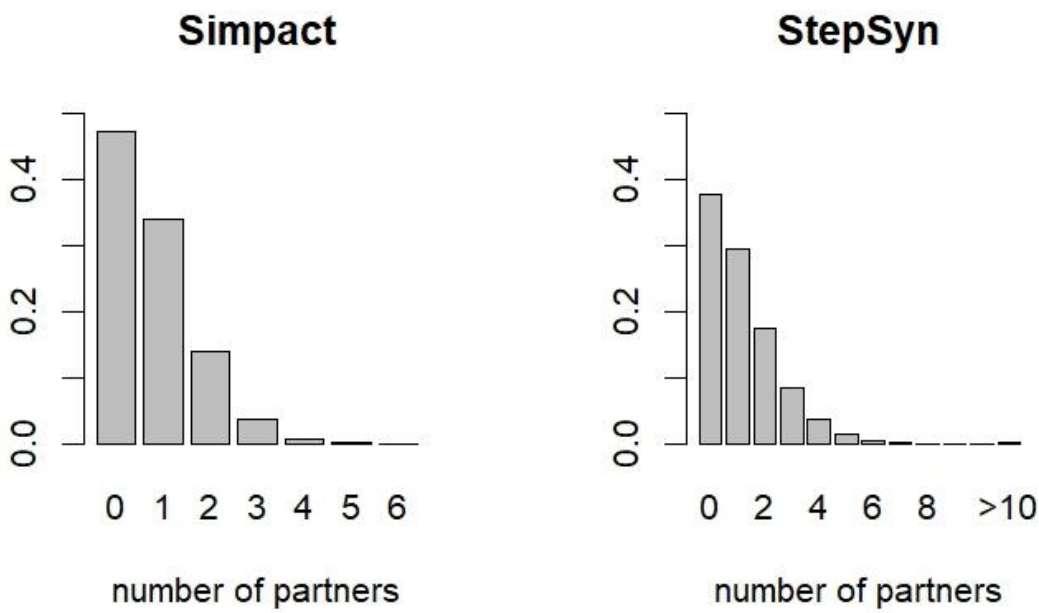

Fig. 6. Distribution of the number of partners for the whole population at the start of the HIV epidemic (HIV seeding) for a simulation with Simpact Yaoundé (left) and a simulation with StepSyn Yaoundé (right).

Despite being programmed in a language with better performance, a model developed with the current version of Simpact Cyan 1.0 is as expected, being a continuous-time event-driven model, slower than a model developed with StepSyn 1.0, being a discrete time model (see Table 1), because Simpact Cyan 1.0 has to be updated more regularly and because of its higher mathematical complexity. In addition, if StepSyn 1.0 is ran with daily instead of weekly steps, it will be considerably slower (only weekly steps were ran in this study). As next steps in the

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

further development of Simpact Cyan 1.0, a refactoring exercise will be performed (among others: selection of output to keep track of, using more optimal packages, modifying how relationships are scheduled in the core of the program) likely closing the gap in computing time between both models. Note however, that using parallel computing, the computing times for both models do not necessarily present a bottleneck.

Simpact Cyan 1.0 is a continuous-time modelling framework, which has the advantage that sexual behaviour and disease progression and transmission can be described more realistically. However, discrete-time models, like those developed with StepSyn 1.0, are easier to handle in terms of parameter estimation, because empirical data for fitting are only available at discrete time-intervals, and are often expressed in terms of probabilities and durations, instead of hazard functions [40].

Models in Simpact Cyan 1.0 can be stratified by age, while in StepSyn 1.0 this option is not available. However, in this study all age-related options in Simpact Cyan 1.0 were turned off. Ignoring age-heterogeneity can among others lead to underestimation of the initial growth rate of the HIV epidemic [41].

Both modelling frameworks include STI co-factor effects, which are known to strongly influence HIV infectivity [42]. Apart from STI co-factor effects, StepSyn 1.0 models can also include the natural history of several STIs. While StepSyn 1.0 mainly focuses on the natural history and transmission of STIs, Simpact Cyan 1.0 has also features to evaluate the effect of interventions, like treatment with antiretroviral therapy.

5. Conclusion

In this study, we compared two individual-based models for simulating HIV transmission dynamics and STI co-factor effects. The models were used to simulate survival curves and HIV prevalence over time in Yaoundé (Cameroon). When calibrated using the same HIV prevalence data, simulations performed with both models fitted the data equally well. However, survival curves and HIV or STI prevalence curves simulated over a larger period showed more variation between simulations for Simpact Cyan 1.0 than for StepSyn 1.0. Differences in the initialization of the models and the generation of the sexual network explain the observed differences in the variance of the output and the estimated parameters for the two models. We strongly recommend using different models to assess the robustness of evidence provided by mathematical modelling.

Acknowledgements

The research conducted by DMH and NH in this study was funded by the Fonds Wetenschappelijk Onderzoek - Vlaanderen (Research Foundation – Flanders; FWO, http://www.fwo.be/en/) (Grant agreements G0E8416N and G0B2317N).

The research done by JDS and AMV in this study has been supported in part by grant G.0692.14 and G0B2317N, funded by the FWO, Belgium.

PL was supported by a PhD grant of the FWO (1S31916N) .

WD was supported by a postdoctoral followship from FWO (12L5816N).

Research done by VM in this study has been supported by the ELTE Institutional Excellence Program (1783-3/2018/FEKUTSRAT) of the Hungarian Ministry of Human Capacities. VM was also supported by the ÚNKP-18-4 New National Excellence Program of the Hungarian Ministry of Human Capacities and by a Bolyai János Research Fellowship of the Hungarian Academy of Sciences.

The authors gratefully acknowledge support from the FWO Scientific Research Community on Network Statistics for Sexually Transmitted Diseases Epidemiology.

The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government – department EWI.

Appendix A. Supplementary material

see Supplementary_material.pdf

References

[1] Geffen N, Welte A. Modelling the human immunodeficiency virus (HIV) epidemic: A review of the substance and role of models in South Africa. South Afr J HIV Med. 2018;19(1):756.

[2] Sucharitakul K, Boily MC, Dimitrov D, Mitchell KM. Influence of model assumptions about HIV disease progression after initiating or stopping treatment on estimates of infections and deaths averted by scaling up antiretroviral therapy. PLoS One. 2018;13(3):e0194220.

[3] Gurski KF, Hoffman KA. Influence of concurrency, partner choice, and viral suppression on racial disparity in the prevalence of HIV infected women. Math Biosci. 2016;282:91-108.

[4] Kasaie P, Pennington J, Shah MS, Berry SA, German D, Flynn CP, et al. The Impact of Preexposure Prophylaxis Among Men Who Have Sex With Men: An Individual-Based Model. Jaids-Journal of Acquired Immune Deficiency Syndromes. 2017;75(2):175-83.

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[5] Sun XD, Xiao YN, Tang SY, Peng ZH, Wu JH, Wang N. Early HAART Initiation May Not Reduce Actual Reproduction Number and Prevalence of MSM Infection: Perspectives from Coupled within- and between-Host Modelling Studies of Chinese MSM Populations. Plos One. 2016;11(3).
[6] de Sousa JD, Muller V, Lemey P, Vandamme AM. High GUD incidence in the early 20 century created a particularly permissive time window for the origin and initial spread of epidemic HIV strains. PLoS One. 2010;5(4):e9936.
[7] Willem L, Verelst F, Bilcke J, Hens N, Beutels P. Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015). BMC Infect Dis. 2017;17(1):612.
[8] Johnson LF, Geffen N. A Comparison of Two Mathematical Modeling Frameworks for Evaluating Sexually Transmitted Infection Epidemiology. Sex Transm Dis. 2016;43(3):139-46.
[9] Hollingsworth TD, Medley GF. Learning from multi-model comparisons: Collaboration leads to insights, but limitations remain. Epidemics. 2017;18:1-3.
[10] Eaton JW, Bacaer N, Bershteyn A, Cambiano V, Cori A, Dorrington RE, et al. Assessment of epidemic projections using recent HIV survey data in South Africa: a validation analysis of ten mathematical models of HIV epidemiology in the antiretroviral therapy era. Lancet Global Health. 2015;3(10):E598-E608.
[11] Liesenborgs J, Hendrickx DM, Kuylen E, Niyukuri D, Hens N, Delva W. SimpactCyan 1.0: An Open-source Simulator for Individual-Based Models in HIV Epidemiology with R and Python Interfaces. bioRxiv. 2018:440834.
[12] Anderson DF. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. J Chem Phys. 2007;127(21):214107.
[13] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
[14] Python Core Team. Python: A dynamic, open source programming language. Python Software Foundation. URL https://www.python.org/
[15] Buve A, Lagarde E, Carael M, Rutenberg N, Ferry B, Glynn JR, et al. Interpreting sexual behaviour data: validity issues in the multicentre study on factors determining the differential spread of HIV in four African cities. AIDS. 2001;15 Suppl 4:S117-26.
[16] Ferry B, Carael M, Buve A, Auvert B, Laourou M, Kanhonou L, et al. Comparison of key parameters of sexual behaviour in four African urban populations with different levels of HIV infection. AIDS. 2001;15 Suppl 4:S41-50.
[17] Morison L, Weiss HA, Buve A, Carael M, Abega SC, Kaona F, et al. Commercial sex and the spread of HIV in four cities in sub-Saharan Africa. AIDS. 2001;15 Suppl 4:S61-9.
[18] U.S. Census Bureau. HIV/AIDS surveillance data base. [Internet]. US Census Bureau. 2001 [cited April 7, 2018]. Available from: https://www.census.gov/programs-surveys/international-programs/about/hiv.html.
[19] Buve A, Carael M, Hayes RJ, Auvert B, Ferry B, Robinson NJ, et al. Multicentre study on factors determining differences in rate of spread of HIV in sub-Saharan Africa: methods and prevalence of HIV infection. AIDS. 2001;15 Suppl 4:S5-14.
[20] Morris M. Telling Tails Explain the Discrepancy in Sexual Partner Reports. Nature. 1993;365(6445):437-40.
[21] Lawless JF. Statistical models and methods for lifetime data: John Wiley & Sons; 2011.
[22] Willem L, Stijven S, Vladislavleva E, Broeckhove J, Beutels P, Hens N. Active Learning to Understand Infectious Disease Models and Improve Policy Making. Plos Computational Biology. 2014;10(4).
[23] Castro Sanchez AY, Aerts M, Shkedy Z, Vickerman P, Faggiano F, Salamina G, et al. A mathematical model for HIV and hepatitis C co-infection and its assessment from a statistical perspective. Epidemics. 2013;5(1):56-66.
[24] Chen KN, Guo SJ, Lin YY, Ying ZL. Least Absolute Relative Error Estimation. Journal of the American Statistical Association. 2010;105(491):1104-12.
[25] Stein M. Large sample properties of simulations using Latin hypercube sampling. Technometrics. 1987;29(2):143-51.
[26] Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, et al. Detecting Novel Associations in Large Data Sets. Science. 2011;334(6062):1518-24.
[27] Hollingsworth TD, Anderson RM, Fraser C. HIV-1 transmission, by stage of infection. J Infect Dis. 2008;198(5):687-93.
[28] Jaffar S, Grant AD, Whitworth J, Smith PG, Whittle H. The natural history of HIV-1 and HIV-2 infections in adults in Africa: a literature review. Bull World Health Organ. 2004;82(6):462-9.
[29] Rutherford GW, Lifson AR, Hessol NA, Darrow WW, O'Malley PM, Buchbinder SP, et al. Course of HIV-I infection in a cohort of homosexual and bisexual men: an 11 year follow up study. BMJ. 1990;301(6762):1183-8.
[30] Mellors JW, Rinaldo CR, Jr., Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. Science. 1996;272(5265):1167-70.
[31] Looker KJ, Magaret AS, Turner KM, Vickerman P, Gottlieb SL, Newman LM. Global estimates of prevalent and incident herpes simplex virus type 2 infections in 2012. PLoS One. 2015;10(1):e114989.
[32] O'Farrell N. Increasing prevalence of genital herpes in developing countries: implications for heterosexual HIV transmission and STI control programmes. Sex Transm Infect. 1999;75(6):377-84.
[33] Weiss H. Epidemiology of herpes simplex virus type 2 infection in the developing world. Herpes. 2004;11 Suppl 1:24A-35A.
[34] Buve A, Carael M, Hayes RJ, Auvert B, Ferry B, Robinson NJ, et al. The multicentre study on factors determining the differential spread of HIV in four African cities: summary and conclusions. Aids. 2001;15:S127-S31.
[35] Foss AM, Vickerman PT, Chalabi Z, Mayaud P, Alary M, Watts CH. Dynamic modeling of herpes simplex virus type-2 (HSV-2) transmission: issues in structural uncertainty. Bull Math Biol. 2009;71(3):720-49.
[36] Garnett GP. The geographical and temporal evolution of sexually transmitted disease epidemics. Sex Transm Infect. 2002;78 Suppl 1:i14-9.
[37] Korenromp EL, Bakker R, de Vlas SJ, Gray RH, Wawer MJ, Serwadda D, et al. HIV dynamics and behaviour change as determinants of the impact of sexually transmitted disease treatment on HIV transmission in the context of the Rakai trial. Aids. 2002;16(16):2209-18.
[38] Orroth KK, Freeman EE, Bakker R, Buve A, Glynn JR, Boily MC, et al. Understanding the differences between contrasting HIV epidemics in east and west Africa: results from a simulation model of the Four Cities Study. Sexually Transmitted Infections. 2007;83:I5-I16.
[39] Eddelbuettel D, Sanderson C. RcppArmadillo: Accelerating R with high-performance C plus plus linear algebra. Computational Statistics & Data Analysis. 2014;71:1054-63.
[40] Ng J, Orav EJ. A Generalized Chain Binomial Model with Application to Hiv-Infection. Mathematical Biosciences. 1990;101(1):99-119.
[41] Gupta S, Anderson RM. Age-dependent sexual behaviour amongst male homosexuals and the transmission dynamics of HIV-1. Scand J Infect Dis Suppl. 1990;69:187-97.

bioRxiv preprint doi: https://doi.org/10.1101/637389; this version posted May 17, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[42] Powers KA, Poole C, Pettifor AE, Cohen MS. Rethinking the heterosexual infectivity of HIV-1: a systematic review and meta-analysis. Lancet Infect Dis. 2008;8(9):553-63.