Published: November 2018

SimpactCyan 1.0: An Open-source Simulator for Individual-Based Models in HIV Epidemiology

Individual-based models in HIV epidemiology

In epidemiology, mathematical models are widely used to simulate progression, transmission, prevention and treatment of infectious diseases. Most of these models are deterministic compartmental models, simulating population averages of changes in infection status and disease stages over time. However, many infectious diseases, in particular sexually transmitted diseases (STIs), are subject to high individual heterogeneity. Unlike compartmental models simulating population averages, individual-based models (IBMs) keep track of the events that happen to each individual separately and are therefore able to take into account various sources of individual heterogeneity.

The ability to let population-level features of complex systems emerge from processes and events that happen to interacting individuals, is arguably the most important quality of IBMs. As the computational expense of IBMs has become less prohibitive thanks to multi-core processors and increased access to high-performance computers, there is a growing use of IBMs in infectious disease epidemiology. SimpactCyan is conceived as a multipurpose model-building tool to address research questions in HIV epidemiology at the intersection of network and social epidemiology, computational biology, public health and policy modelling.

SimpactCyan is by no means the first HIV-specific individual-based modelling tool to have been developed. However, with SimpactCyan we aim to overcome some of the limitations of current software for implementing IBMs in HIV epidemiology. While some modelling tools (e.g. STDSIM for simulating transmission of HIV and other STIs (1)) are not open source, other IBMs (e.g. EMOD (2)) are relatively difficult to modify. Another limitation of EMOD is that it can only be used on computers running Windows 10, Windows Server 12, Windows HPC Server 12 or CentOS 7.1. Furthermore, while it has interfaces for Matlab and Python, it does not have an R interface. NetLogo models, on the other hand, are easily modifiable (3), and can be run from within the R environment (4) but are too slow for simulating large populations over the time-scale relevant for HIV epidemiology.

With a few exceptions (e.g. the MicSim Package (5)), existing simulators implement IBMs in with fixed-time steps. For example, the state of model population gets updated every week. However, a continuous time implementation of IBMs has the advantage that it elegantly handles competing risks to multiple events. For instance, an individual may be concurrently at risk of HIV-related mortality as well as at risk of transmitting the virus to a partner. Evaluating the model in fixed time-steps may lead to the situation where both events are scheduled to have taken place between now and the next time-step. However, in reality, this is only possible if the transmission event happens first. In the continuous time model evaluation, we know exactly which of the two events is scheduled first, and logical consequences for the likelihood of subsequent events are processed along with the execution of the first event.

Furthermore, events happening after short and long time-periods can be included in a single simulation. In contrast, in a discrete time model, simulating events that occur on vastly different time-scales can be computationally inefficient. Frequently occurring events may require a small time-step, possibly leading to the occurrence of rare events being evaluated with a much higher frequency than necessary. Another limitation of existing implementations of IBMs for dynamic sexual networks is that they require ad-hoc decisions about who and in what order people “go out” to find partners and can “be found”. SimpactCyan is a simulator for event-driven IBMs in HIV epidemiology, evaluated in continuous time: the state of the system is updated each time an event happens. Furthermore, all possible relationships are considered simultaneously instead of sequentially.

The simulation algorithm

Event times, i.e. time points in the simulation at which events are scheduled to take place, are determined using the modified Next Reaction Method (mNRM) (6). The mNRM was originally designed for simulating chemical systems with time-dependent propensities and delays, but in SimpactCyan we use it to simulate how individuals are at risk of events according to hazard functions. In the mNRM algorithm, there is a core distinction between *internal* event times} and (simulated) *real-world* event times. The internal event times determine when an event will be triggered according to the event's internal clock. We can think of it as a time bomb. When the internal clock reaches a pre-set time, the event takes place (bomb goes off). How fast this clock ticks, depends on the hazard function for the event. A low hazard means that the clock will tick slowly. As a result, two events can have the same internal clock time scheduled, but the real-world times of the events may be different if the hazard functions for the events are different. Full details of the simulation algorithm and the events that can happen to individuals in models formulated with SimpactCyan can be found at: https://simpactcyan.readthedocs.io/en/latest.

Model applications

To illustrate what kind of analyses could be conducted with SimpactCyan, we present two simple examples of model applications. The first illustrates how SimpactCyan can be used to assess the impact of progressive changes to the ART eligibility criteria in Eswatini (formerly known as Swaziland). The second illustrates the use of SimpactCyan as a data-generating tool for assessing the performance of other modelling frameworks. All code and data files necessary to reproduce the examples are available at https://github.com/wdelva/SimpactCyanExamples.

In the MaxART project (7), SimpactCyan is used to estimate the likely impact of Eswatini's shift towards "Early Access to ART for All" (EAAA) on the incidence of HIV. HIV incidence is the rate at which HIV-uninfected people acquire the infection. Such HIV transmission events are scheduled each time a relationship is formed between an HIV-infected and an HIV-uninfected individual. The hazard for transmission events primarily depends on the viral load of the infected partner but can be defined in such a way to also allow for a hazard-lowering effect of multiple ongoing relationships (so-called coital dilution), as well as a hazard-increasing effect of adolescent age among women.

The viral load of an infected individual varies over time: During the chronic stage of the infection, viral concentration assumes a set-point viral load. Model parameters determine by which factors the transmission hazard should be multiplied during the initial acute stage, as well as the early and late AIDS stages. At the time of HIV acquisition, time till HIV-related death is determined, based on an early paper by Arnaout et al. (8) to allow for a negative correlation between set-point viral load and post-infection survival time.

The set-point viral load value allocated to a newly infected individual is partially determined by that of their infector, i.e. some heritability of set-point viral load is assumed (9). ART initiation affects both the expected time to HIV-related death and the infectiousness of the person on ART. As soon as ART is started, the log10 viral load is assumed to drop by a user-defined fraction, and the updated current viral load is used to recalculate the post-infection survival time. In the simulations of which key output is shown in Figure 1, we assume that upon ART initiation, the log10 viral load drops by 70%, effectively rendering the viral load "undetectable" for most ART clients. We further assumed that ART was gradually introduced around the year 2000 and that the CD4 cell count threshold for ART eligibility progressively shifted towards ever more inclusive criteria, alongside a decreasing lag-time between HIV infection and HIV diagnosis. These assumptions hold in both the "Status Quo" scenario and the "EAAA" scenario. In the EAAA scenario, however, an additional policy change is modelled: a policy of immediate access to ART for all people infected with HIV is adopted from October 2016. In the alternative scenario, the CD4 cell count threshold for ART eligibility stays at 500 cells/microliter from mid-2013 onwards. Extra intervention events are added in both scenarios to model a moderate reduction in sexual risk behaviour (rate of partner turnover) that took place in the period 2000-2003.

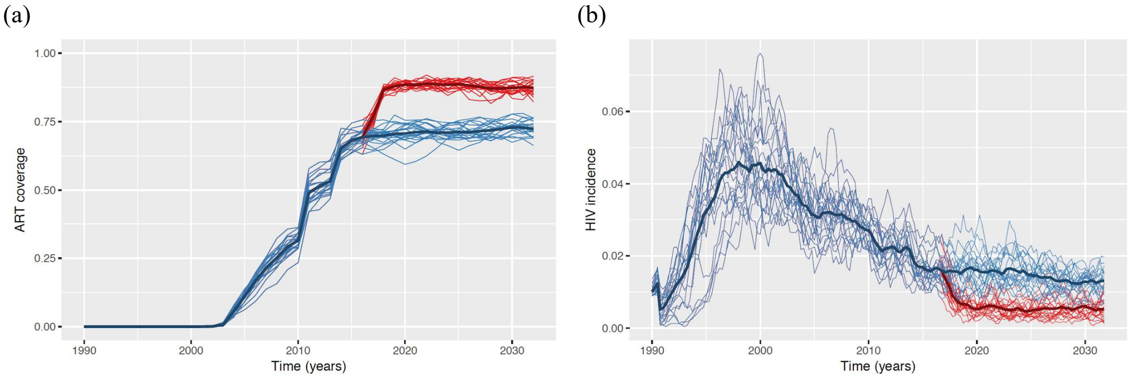

Figure 1. ART coverage (a) and HIV incidence (b) under a "Status Quo"' (blue) and "Early Access to ART for All" (red) scenario for the roll-out of a nation-wide ART programme.

The second use case illustrates how SimpactCyan can be used as a data-generating tool for benchmarking the performance of other modelling frameworks. Phylogenetic models have been used to infer properties of epidemics from reconstructed phylogenetic trees, including time-trends in HIV incidence rates (10, 11) and the age-mixing pattern in HIV transmission clusters (12). As the truth is typically unknown, it is difficult to assess the validity of these novel modelling frameworks or document their sensitivity to breaches in the models' assumptions. For instance, phylogenetic inference methods typically assume that HIV sequence data are available for most HIV-positive people and that the individuals for whom the viral genome was sequenced, are a random subset of all HIV infected people. However, in many settings, neither of these assumptions is met. In Figure 3 we illustrate the basic idea of SimpactCyan as a data-generating tool for benchmarking.

First, we simulate an emerging HIV epidemic. Panel (a) shows the cumulative HIV transmission network of the epidemic, linking all individuals who got infected with HIV by the end of the simulation. Next, assuming HIV transmission events correspond to branching points in the phylogeny, we convert the transmission network into its corresponding phylogenetic tree. Using the Seq-Gen program (13), we simulate the viral evolution along this phylogeny, assuming a generalised time-reversible substitution model (14) with a gamma-invariable mixture model for rate heterogeneity among sites. In this way, we generate synthetic HIV sequence data. Lastly, we feed these sequences into the FastTree 2 program (15) and the treedater R package (16) to reconstruct the time-resolved phylogenetic tree, shown in panel (b). If all people ever infected are included in the sequence dataset and the same molecular evolution model is used to generate the sequence data and to reconstruct the phylogenetic tree, the timing of the internal nodes in the reconstructed tree should correspond with the timing of the simulated HIV transmission events, as shown in panel (c). Now that we have established the validity of this phylogenetic inference approach under ideal circumstances, we can examine the performance of the inference method under alternative scenarios in which some of the viral sequence data are missing completely at random (MCAR), missing at random (MAR) or not at random (MNAR). Through simulation-based sensitivity analyses, we could quantify how the accuracy of epidemiological characteristics inferred by the phylodynamic method depends on the magnitude of the missing data problem and the strength of the correlations between the probability of sequences being missing and covariates such as age, indicators of sexual risk behaviour or calendar time.

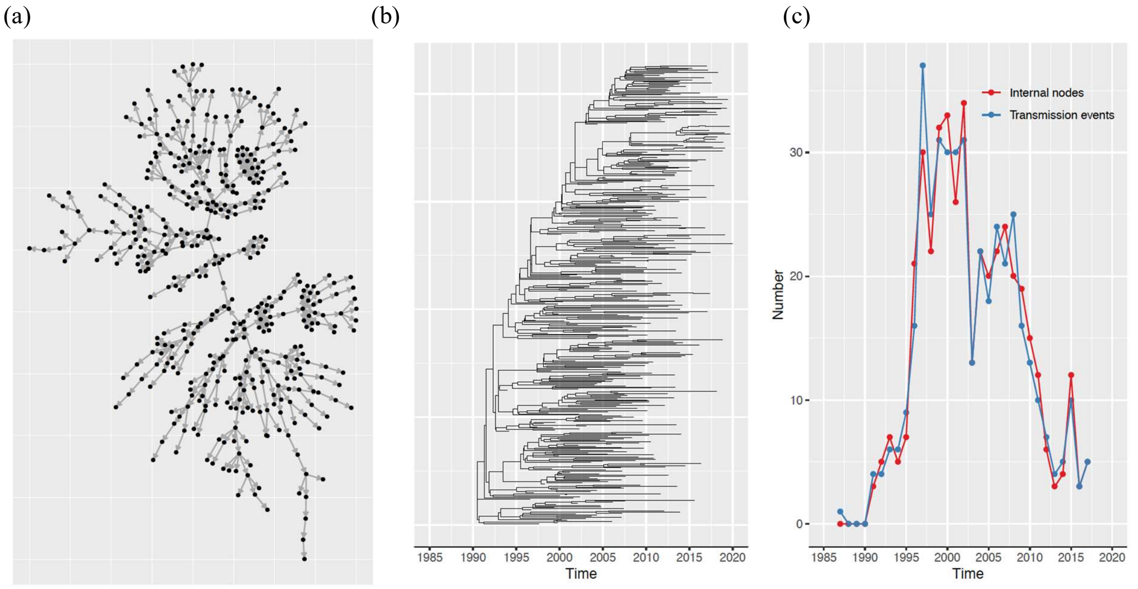

Figure 3. HIV transmission network, phylogenetic tree and timing HIV transmission events. (a) HIV transmission network. *The molecular evolution of HIV viral strains is simulated across an HIV transmission network using Seq-Gen (13).* ***(b) Time-resolved phylogenetic tree.*** *The synthetic HIV sequence data are used to reconstruct the time-resolved phylogenetic tree with FastTree 2 (15) and the treedater R package (16).* ***(c) Internal nodes in the reconstructed phylogenetic tree and HIV transmission events.*** *Under ideal circumstances of complete sampling of the transmission network and correct specification of the model for viral evolution, the timing of the internal nodes in the reconstructed phylogenetic tree (red) corresponds nearly perfectly with the timing of the simulated HIV transmission events (blue).*

Future directions

Ongoing developments of SimpactCyan include the addition of events for the transmission and treatment of other STIs such as Herpes Simplex Virus 2 (HSV-2) and Hepatitis C Virus (HCV), as well as additional events for parenteral and MTCT of HIV and co-infections, to allow studies of HIV transmission in injecting drug users (IDU) and children. We also plan to extend the software by enabling an explicit modelling of the correlation between sexual risk behaviour and health-seeking behaviour. This is in response to recent evidence to suggest that high sexual risk behaviour is associated with a lower likelihood to be aware of one's HIV infection, and a lower likelihood of being virally suppressed among people who know they are HIV positive (17).

Conceived as a flexible open-source, open access tool, rather than a proprietary asset, SimpactCyan's extensions and applications should not solely come from its original developers. Instead, we want to position this simulator as a vehicle for open science in HIV epidemiology. Therefore, others are encouraged to use it for the development of their IBMs, as the starting point for their simulation engine, as a data-generating and benchmarking tool in methodological research, or for educational purposes.

Wim Delva *- Epidemiologist, SACEMA, and Ghent University, Hasselt University and KU Leuven, Belgium. Areas of interest: statistical analysis of sexual behaviour data, stochastic and deterministic modelling of sexual network dynamics and HIV transmission.* *Wim.Delva@ugent.be*

Co-authors: ***Jori Liesenborgs, Diana M Hendrickx, Elise Kuylen, David Niyukuri, Niel Hens***

References

1. Bakker, R. et al. Stdsim: A microsimulation model for decision support in the control of hiv and other stds. Sex. Transm. Dis. 27, 652 (2000).
2. Bershteyn, A. et al. Implementation and applications of emod, an individual-based multi-disease modeling platform. Pathog. Dis. 76 (2018).
3. Kravari, K. & Bassiliades, N. A survey of agent platforms. J. Artif. Soc. Soc. Simul. 18, 11 (2015).
4. Thiele, J. C. R marries netlogo: Introduction to the rnetlogo package. J. Stat. Softw. 58 (2014).
5. Zinn, S. The micsim package of r: An entry-level toolkit for continuous-time microsimulation. Int. J. Microsimulation 7, 3–32 (2014).
6. Anderson, D. F. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. The J. Chem. Phys. 127 (2007).
7. Walsh, F. J. et al. Impact of early initiation versus national standard of care of antiretroviral therapy in Swaziland's public sector health system: study protocol for a stepped-wedge randomized trial. Trials 18, 1–10 (2017).
8. Arnaout, R. A. et al. A simple relationship between viral load and survival time in hiv-1 infection. Proc. Natl. Acad. Sci. 96 , 11549–11553 (1999).
9. Fraser, C. et al. Virulence and pathogenesis of hiv-1 infection: an evolutionary perspective. Science 343, 1243727 (2014).
10. Rasmussen, D. A., Volz, E. M. & Koelle, K. Phylodynamic Inference for Structured Epidemiological Models. PLoS Comput. Biol. 10 (2014).
11. Lewis, F., Hughes, G.J., Rambaut, A., Pozniak, A. & Leigh Brown, A. J. Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Medicine 5, 0392–0402 (2008).
12. de Oliveira, T. et al. Transmission networks and risk of HIV infection in KwaZulu-Natal, South Africa: a community-wide phylogenetic study. The Lancet HIV 3018, 1–10 (2016).
13. Rambaut, A. & Grassly, N. C. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. applications biosciences: CABIOS 13, 235–8 (1997).
14. Tavaré, S. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect. on mathematics life sciences. Vol. 17 57–86 (1986).
15. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS ONE 5 (2010).
16. Volz, E. M. & Frost, S. D. W. Scalable relaxed clock phylogenetic dating. Virus Evol. 3 (2017).
17. Huerga, H. et al. Higher risk sexual behaviour is associated with unawareness of hiv-positivity and lack of viral suppression – implications for treatment as prevention. Sci. Reports 7, 16117 (2017).