Effects of Contact Structure on the Transient Evolution of HIV Virulence

Sang Woo Park[1] Benjamin M. Bolker[1,2,3,*]

1 Department of Mathematics & Statistics, McMaster University, Hamilton, Ontario, Canada

2 Department of Biology, McMaster University, Hamilton, Ontario, Canada

3 Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada

* bolker@mcmaster.ca

Abstract

Early in an epidemic, high densities of susceptible hosts select for relatively high parasite virulence; later in the epidemic, lower susceptible densities select for lower virulence. Thus over the course of a typical epidemic the average virulence of parasite strains increases initially, peaks partway through the epidemic, then declines again. However, precise quantitative outcomes, such as the peak virulence reached and its timing, may depend sensitively on epidemiological details. Fraser *et al.* proposed a model for the eco-evolutionary dynamics of HIV that incorporates the tradeoffs between transmission and virulence (mediated by set-point viral load, SPVL) and their heritability between hosts. Their models use implicit equations to capture the effects of partnership dynamics that are at the core of epidemics of sexually transmitted diseases. Our models combine HIV virulence tradeoffs with a range of contact models, modeling partnership formation and dissolution and allowing for individuals to transmit disease outside of partnerships. We assess summary statistics such as the peak virulence (corresponding to the maximum value of population mean $\log_{10}$ SPVL achieved throughout the epidemic) across models for a range of parameters applicable to the HIV

epidemic in sub-Saharan Africa. Although virulence trajectories are broadly similar across models, the timing and magnitude of the virulence peak vary considerably. Previously developed implicit models predicted lower virulence and slower progression at the peak (a maximum of 3.5 $\log_{10}$ SPVL) compared both to more realistic models and to simple random-mixing models with no partnership structure at all (both with a maximum of $\approx 4.7 \log_{10}$ SPVL). In this range of models, the simplest random-mixing structure best approximates the most realistic model; this surprising outcome occurs because the dominance of extra-pair contact in the realistic model swamps the effects of partnership structure.

Author Summary

Pathogens such as HIV can evolve rapidly when the environment changes. One important aspect of a pathogen's environment is the probability that an infectious contact (a sneeze for a respiratory disease, or an unprotected sex act for a sexually transmitted disease) encounters an uninfected person and thus has a chance to transmit the pathogen. As an epidemic grows the number of uninfected people shrinks, changing evolutionary pressures on the pathogen. While researchers have used models to explore pathogen evolution during epidemics, their models usually neglect important processes such as people entering and leaving sexual partnerships. We compared several evolutionary models for HIV that include partnership dynamics as well as sexual contact outside of stable partnerships. Models of intermediate complexity predicted lower virulence midway through the epidemic (a minimum of 15 years to progress to AIDS) than either more realistic models or simple models with no partnership structure (both with a minimum of ≈ 7.25 years to progress to AIDS), because random sexual contacts tended to wash out the effects of stable partnerships. Researchers trying to predict the evolution of pathogens must try to understand the implications of their modeling choices; models of intermediate complexity may not produce intermediate conclusions.

Introduction

The evolution of pathogen virulence has both theoretical and, potentially, practical importance. In general, evolutionary theory suggests that pathogens with higher reproduction numbers ($\mathcal{R}_0$) — that is, the number of secondary infections caused by a single infected host over the course of its infectious period — will tend to increase in prevalence relative to strains with lower reproductive ratios. Pathogens can increase their reproduction numbers either by increasing their transmission rate, the rate (per infected host) at which they infect new hosts, or by decreasing their clearance or disease-induced mortality rate, the rate at which hosts recover or die from disease. The *trade-off theory* [1] postulates that transmission and disease-induced mortality rates are both driven by the rate at which the pathogen exploits host resources for within-host reproduction, and that pathogen evolution will thus strike a balance between the pathogen's rate of transmission to new hosts and its rate of killing its host (or of provoking the host's immune system to eliminate it). Some biologists have criticized the tradeoff theory [2,3], but others have successfully applied it to a variety of host-pathogen systems [4–7].

Fraser *et al.* have applied these ideas to HIV, showing that the virus appears to satisfy the prerequisites of the tradeoff theory. The *set-point viral load* (SPVL: i.e., the characteristic virus load measured in blood during the intermediate stage of infection) is a measurable proxy for the rate of HIV within-host reproduction. Higher viral loads are correlated with shorter progression times (higher virulence). Studies of discordant partnerships — stable sexual partnerships with one infected and one uninfected partner — have shown that SPVL is (1) correlated with faster transmission (the time to infection of the uninfected partner is shorter when their partner has a higher SPVL) and (2) heritable (when the uninfected member of the couple does become infected, their SPVL is similar to their partner's). Furthermore, the rate of transmission increases with increasing SPVL (and hence with decreasing progression time) but with a decreasing slope as progression time decreases, fulfilling the requirements of the tradeoff theory [8]. Subsequent studies [10–12] used these data to parameterize mechanistic models of HIV virulence evolution, suggesting that HIV invading a novel population would initially evolve increased virulence, peaking after approximately 100-200 years and then

declining slightly to a long-stable virulence level.   32

The work of Shirreff *et al.* [10], and particularly the predicted transient peak in HIV   33
virulence midway through the epidemic, highlights the importance of interactions   34
between epidemiological and evolutionary factors [13, 14]. However, despite these   35
studies' attention to detail at the individual or physiological level, the population-level   36
contact structures used in these models are relatively simple.   37

As we discuss in detail below, existing models of HIV eco-evolutionary dynamics   38
either use implicit models that incorporate the average effects of within-couple sexual   39
contact — without representing the explicit dynamics of partnership formation and   40
dissolution or accounting for extra-partnership contact — or use an agent-based   41
formulation with parameters that effectively lead to random mixing among infected and   42
uninfected individuals **may not be true any more, Herbeck may have fixed this**   43
in their revised versions. Here we explore the effects of incorporating *explicit*   44
contact structure in eco-evolutionary models.   45

Because our main goal is to explore how conclusions about virulence evolution   46
depend on the way in which contact structure is modeled, we consider a series of models   47
with increasing levels of complexity in the contact structure, but simplify some of the   48
other epidemiological processes (such as the within-host life history of HIV). We   49
evaluate our models across a wide range of parameters, using a Latin hypercube design;   50
for each model run, we compute a set of metrics that summarize the evolutionary   51
trajectory of SPVL over the course of the epidemic.   52

Materials and Methods   53

Infection dynamics   54

Our models explicitly track the evolution of the distribution of $\log_{10}$ SPVL (which we   55
denote as α) rather than the rate of progression to AIDS itself (hereafter we use   56
"virulence" to denote $\log_{10}$ SPVL). We use a single-stage model of HIV that assumes   57
constant infectivity over the course of an exponentially distributed infectious period.   58
This assumption contrasts with Shirreff *et al.*'s previous model which explicitly tracked   59
three stages of HIV infection (primary, asymptomatic, and AIDS) and used a more   60

realistic Weibull-distributed infectious period; we show below that our results are not   61

overly sensitive to this simplification. Shirreff *et al.* allowed the transmission rate and   62

infection duration to depend on virus load only during the asymptomatic stage. In   63

order to adapt their parameterization to our single-stage model, we used their estimates   64

to compute the SPVL-dependent transmission rate and duration during the   65

asymptomatic stage and then derived the overall duration of the infectious period as the   66

sum of the three stage durations and the average transmission rate as the   67

duration-weighted average of the three stage-specific transmission rates. Thus the   68

within-couple transmission rate, β, for our models is given by: 69

$$\beta(\alpha) = \frac{D_P \beta_P + D_A(\alpha)\beta_A(\alpha) + D_D \beta_D}{D_P + D_A(\alpha) + D_D}, \tag{1}$$

where the duration of infection (D_P and D_D) and rate of transmission (β_P and β_D) of   70

the Primary and Disease stages of infection are independent of the host's SPVL (Table 1   71

gives definitions, units, and values for all parameters). Following Shirreff *et al.*, the   72

duration of infection (D_A) and rate of transmission (β_A) for the Asymptomatic stage   73

are Hill functions of the SPVL: 74

$$D_A(\alpha) = \frac{D_{\max} D_{50}^{D_k}}{V_\alpha^{D_k} + D_{50}^{D_k}},$$
$$\beta_A(\alpha) = \frac{\beta_{\max} V_\alpha^{\beta_k}}{V_\alpha^{\beta_k} + \beta_{50}^{\beta_k}}, \tag{2}$$

where $V_\alpha = 10^\alpha$.   75

The uncoupled and extra-couple transmission rates (i.e., the rates of transmission   76

among people outside of a stable partnership, or between people inside of a stable   77

partnership and people other than their partner) are scaled by multiplying the   78

within-couple transmission rate β by the contact ratios c_u/c_w and c_e/c_w (see Appendix   79

S1). Simplifying the model of HIV pathogenesis from three stages to a single stage   80

could affect our conclusions about the evolution of virulence (e.g. Kretzschmar and   81

Dietz [17] show that pair formation dynamics and multiple stages of infectivity have   82

interactive effects on $\mathcal{R}_0$). However, our simplified model produces results that are   83

qualitatively similar to those of Shirreff *et al.*'s [10] model; when our model is calibrated   84

Table 1. Parameter ranges/values. Parameters with fixed values are kept constant throughout simulations while other parameter values are taken from Latin hypercube samples using ranges specified in the table. Values of c and ρ are doubled from those given by Champredon *et al.* because we keep track of individuals in the model, while they keep track of couples. Starred (*) parameters (used in Fig 1), and descriptions of Hill function coefficients, are taken from [10].

Notation	Description	Range/Value	Source
ρ	Partnership formation rate	1/10-2/5 per year	[15]
c	Partnership dissolution rate	1/15-1/5 (1.25*) per year	[15]
c_u/c_w	Relative contact rate for uncoupled transmission	1/5-5	Assumption
c_e/c_w	Relative contact rate extra-couple	0.01-1	[15]
β_P	Rate of transmission during primary infection	1.31-5.09 (2.76*) per year	[18]
β_D	Rate of transmission during high transmission disease stage	0.413-1.28 (0.76*) per year	[18]
D_P	Duration of primary infection	1.23/12-6/12 (0.25*) years	[18]
D_D	Duration of high transmission disease stage	4.81/12-14/12 (0.75*) years	[18]
β_{max}	Maximum rate of transmission during asymptomatic stage	0.317 per year	[10]
β_{50}	SPVL at which infectiousness is half maximum	13938 copies per ml	[10]
β_k	Hill coefficient: steepness of increase in infectiousness as a function of SPVL	1.02	[10]
D_{max}	Duration of primary infection	25.4 years	[10]
D_{50}	SPVL at which duration of asymptomatic infection is half maximum	3058 copies per ml	[10]
D_k	Hill coefficient: steepness of decrease in duration as a function of SPVL	0.41	[10]
σ_M	Mutation standard deviation of $\log_{10}$ SPVL	0.12	[10]
α_{min}	Minimum $\log_{10}$ SPVL	2	[10]
α_{max}	Maximum $\log_{10}$ SPVL	7	[10]
n	Number of strains	21 (51*)	Assumption
μ	Mean number of non-cohabiting sexual partners	0.103 - 1.206	[19]
κ	Squared coefficient of variation of number of non-cohabiting sexual partners	0.01 - 100	Assumption

to have a similar initial epidemic growth rate r, the peak $\log_{10}$ SPVL occurs at the same time ($\approx$ 200 years) but slightly higher (4.6 $\log_{10}$ SPVL vs. 4.3 $\log_{10}$ SPVL: Fig 1).
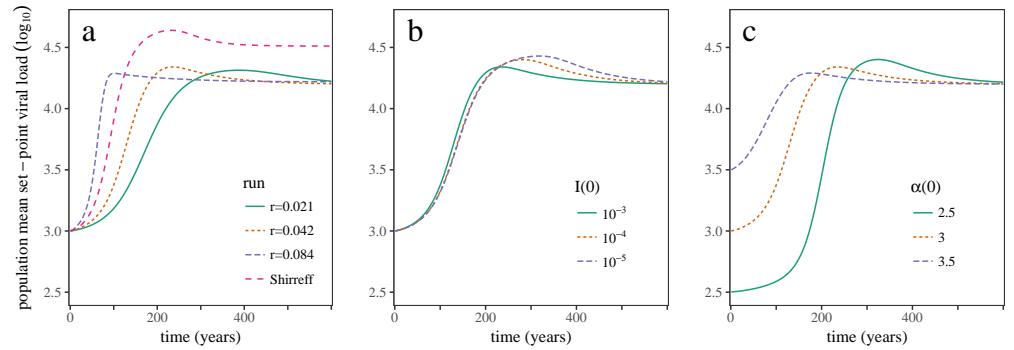
Fig 1. Baseline dynamics. Time series of mean population $\log_{10}$ SPVL using baseline parameters. Each initial condition was varied independently in order to test their effects on transient dynamics of virulence evolution. Note that the initial condition has negligible effect on equilibrium virulence. (a) Contrast between the three-stage Shirreff model and the single-stage model calibrated to varying initial exponential growth rates, r. (b) Effects of varying initial infectious density $I(0)$. (c) Effects of varying initial mean virulence $\alpha(0)$. The middle $r = 0.042$ (orange, dotted) curve in panel (a), calibrated to match the epidemic dynamics of Shirreff *et al.*'s three-stage model [10], shows that our simplified single-stage HIV model can produce similar SPVL trajectories to the original three-stage model; the other r values are chosen to show the effects of doubling or halving the initial epidemic growth rate. The eco-evolutionary dynamics are qualitatively similar across a range of initial conditions (panels b,c).

Mutation

Over the course of infection, mutation occurs within the host. However, we follow Shirreff *et al.* in assuming that SPVL of the strain transmitted by an infected individual is determined by the SPVL at the time of infection and is not further affected by within-host mutation. Instead, the mutational effect is modeled as occurring in a single step at the time of transmission. First, the distribution of $\log_{10}$ SPVL is discretized into a vector:

$$\alpha_i = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min})\frac{i-1}{n-1} \qquad i = 1, 2, 3, \ldots n. \tag{3}$$

We have experimented with varying degrees of discretization in the strain distribution (i.e., values of n); in our model runs comparing results with Shirreff *et al.* [10] (Fig 1) we use $n = 51$ (i.e. a bin width of 0.1 $\log_{10}$ SPVL for α), but reducing n to 21 (bin width = 0.25 $\log_{10}$ SPVL) makes little difference; we use this coarser grid for all other simulations reported.

We construct an $n \times n$ mutational matrix, M — which is multiplied with the transmission term — so that M_{ij} is the probability that a newly infected individual will have $\log_{10}$ SPVL of α_j given that the infector has $\log_{10}$ SPVL of α_i. Finally, the

probabilities are normalized so that each row sums to 1:   102

$$M_{ij} = \frac{\Phi(\alpha_j + d/2; i) - \Phi(\alpha_j - d/2; i)}{\Phi(\alpha_{\max} + d/2; i) - \Phi(\alpha_{\min} - d/2; i)}, \tag{4}$$

where $\Phi(x; i)$ is the Gaussian cumulative distribution function with mean α_i and   103
variance of σ_M^2, and $d = (\alpha_{\max} - \alpha_{\min})/(n-1)$. Unlike Shirreff *et al.*, who allowed for   104
variation in the expressed phenotype (SPVL) of each genotype, we use a one-to-one   105
genotype-phenotype map. Thus there is a single value for transmission rate and for   106
progression rate corresponding to each SPVL compartment in the model: 107

$$\begin{aligned} \beta_i &= \beta(\alpha_i), \\ \lambda_i &= \frac{1}{D_P + D_A(\alpha_i) + D_D}. \end{aligned} \tag{5}$$

Contact structure and partnership dynamics   108

We developed seven multi-strain evolutionary models covering a gamut including   109
Champredon *et al.*'s relatively realistic [15] and Shirreff *et al.*'s relatively simple [10]   110
contact structures, each of which is based on different assumptions regarding contact   111
structure and partnership dynamics. Specifically, we focus on the effects of the   112
assumptions of (1) instantaneous vs. non-instantaneous partnership formation; (2) zero   113
vs. positive extra-partnership sexual contact and transmission; and (3) homogeneous vs.   114
heterogeneous levels of sexual activity on the evolution of mean $\log_{10}$ SPVL.   115

Our first four models consider explicit partnership dynamics [15]. The first assumes   116
non-instantaneous partnership formation (i.e. individuals spend some time uncoupled,   117
outside of partnerships) and consist of five states that are classified by infection status   118
and partnership status; single (uncoupled) susceptible individuals (S), single infected   119
individuals (I), concordant negative (susceptible-susceptible) couples (SS), discordant   120
(susceptible-infected) couples (SI), and concordant positive (infected-infected) couples   121
(II). The rates of pair formation are based on the numbers of uncoupled susceptible   122
and infected individuals and the pair-formation rate; partnerships can either dissolve   123
into singletons or be transformed into other types of partnerships by infection of one   124
partner. This model also includes extra-partnership contact with both uncoupled   125
individuals and individuals in other partnerships (we denote it "pairform+epc"), so that   126

susceptible uncoupled individuals and susceptible partners in any type of partnership can be infected by infected uncoupled individuals or infected partners in any type of partnership.

Specifically, single individuals (S and I) acquire partners at a rate ρ, and partnerships dissolve at a rate c. Infected individuals in a discordant partnership infect their susceptible partner at a rate β (within-couple transmission rate) and susceptible individuals outside the partnership at a rate c_e (extra-couple transmission rate). Infected individuals in seropositive (II) partnerships can also infect any susceptible individual at rate c_e. Likewise, single infected individuals (I) can infect any susceptible individuals (single individuals S, or susceptible members of SS or SI partnerships) at a rate c_u through uncoupled mixing. This parameterization follows Champredon *et al.*; we have adapted some of the details of their model to a multi-strain scenario, so that we track (for example) a matrix II_{ij} that records the number of concordant, HIV-positive partnerships in which the two partners have $\log_{10}$ SPVL of α_i and α_j.

Our second model ("pairform") only considers within-couple transmission, in which case infection can only occur within a serodiscordant partnership; that is, we set c_e and c_u to zero.

Our third and fourth models, which are intended to bridge the gap between models with fully explicit pair-formation dynamics and the simpler, implicit models used by Shirreff *et al.* [10], assume instantaneous partnership formation ("instswitch"). The compartmental structure thus omits the single states S and I, comprising only the three partnered states: SS, SI, and II. Like the first two models, this pair of models differs in their inclusion of extra-pair contact: the third model ("instswitch+epc") includes extra-partnership contact (now only with individuals in other partnerships, since uncoupled individuals do not exist in this model) while the fourth ("instswitch") only considers within-couple transmission. Although these models can also be implemented by setting the partnership formation rate of the explicit partnership models to a high value (and we have tested that both methods in fact produce same results), we model instantaneous partnership formation models independently in order to avoid scaling of partnership formation rate during model calibration affecting the virulence trajectory.

The fifth and sixth models represent extreme simplifications of sexual partnership dynamics. One ("implicit") is an implicit serial monogamy model based on the

epidemiological model used by Shirreff *et al.* [10]. It is actually a random mixing model 159
that explicitly tracks only the total number of susceptible and infected individuals. 160
However, to reflect the effect of partnership structure, it uses an adjusted transmission 161
rate derived from an approximation of the basic reproduction number of a serial 162
monogamy model with instantaneous pair formation [18]. The second model of this pair 163
("random") is a simple random-mixing model. 164

Lastly, we incorporate heterogeneity in sexual activity into the models. Individuals 165
are divided into different risk groups based on their level of sexual activity; we scale all 166
aspects of sexual activity, assuming that sexual activity level in both within- and 167
extra-couple contacts is directly proportional to number of non-cohabiting (extra-couple 168
and uncoupled) partners per year [19] (see Appendix S1). We assume random 169
activity-weighted mixing between risk groups [20]. (In the main text we focus on the 170
model with non-instantaneous pair formation, extra-partnership contact 171
("pairform+epc") and heterogeneous sexual activity, which we denote as "hetero"; 172
Figure S4 in Appendix S2 presents results on the effect of adding heterogeneity to other 173
model variants.) While this model lacks some important elements, such as 174
age-structured mixing patterns, needed for realistic models of HIV transmission in 175
sub-Saharan Africa, it represents a first step toward assessing the effects of 176
epidemiological complexity. As even the models shown here push the limits of 177
compartmental-based models (the heterogeneity model comprises 24530 coupled 178
ordinary differential equations), adding further complexity will probably require a shift 179
to an agent-based model framework, as well as considerable effort in model 180
calibration [11, 21]. 181

For simplicity (and following Shirreff *et al.*), all of our base models use an SIS 182
(susceptible-infected-susceptible) formulation, where there is no natural mortality (and 183
no explicit introduction of newly sexually active individuals into the susceptible pool). 184
Individuals who die from AIDS are immediately replaced by an individual in the 185
uncoupled-susceptible compartment. While admittedly unrealistic, this approach is 186
reasonable given that (1) the natural mortality rate is low relative to the 187
epidemiological dynamics and (2) the infectious period is long, so that the overall rates 188
of disease-induced mortality and recruitment of newly sexually active individuals would 189
roughly balance. To check the importance of this assumption, we also built models with 190

vital dynamics where individuals dying from AIDS are removed from the population, 191
with constant recruitment rates and constant low *per capita* natural mortality rates; 192
this additional structure had only minor effects on the results. 193

Initial conditions 194

Choosing the initial conditions for the simulations is something of a challenge. In the 195
most common theoretical modeling contexts researchers are primarily interested in 196
equilibria (or other long-term dynamical attractors such as limit cycles) and are 197
exploring models that have a single stable attractor, so initial conditions can be ignored 198
as long as we run models for long enough. In eco-evolutionary dynamics, however, the 199
initial conditions do affect our conclusions. We have no empirical information that 200
would justify a particular quantitative choice of the fraction infected and the mean and 201
variance of the distribution of SPVL at the point when the pandemic strain of HIV-1 202
entered the human population; in any case, the level of realism of our model does not 203
support such a detailed consideration of the early dynamics of HIV. In most cases, we 204
started with an initial $\log_{10}$ SPVL of 3.0, to match the value used by Shirreff *et al.*. 205
Shirreff *et al.* use an initial $I(0) = 10^{-3}$; we show results for this value in Fig 1. Because 206
we calibrated parameters based on the initial epidemic growth rate (see "Simulation 207
runs" below), we set $I(0)$ to 10^{-4} to ensure that the exponential growth phase lasted 208
long enough for reliable estimation of the initial growth rate. 209

   We ran one small set of sensitivity analyses with baseline values of the 210
epidemiological parameters (Table 1) to test the effects of the initial conditions. 211
Appendix S1 provides further details on initial conditions, such as the initial 212
distribution of SPVL around the mean and the distribution of initial infected density 213
across single people and different partnership types. 214

Latin hypercube sampling 215

Despite considerable effort [15, 18], the parameters determining the rate and structure of 216
sexual partnership change and contact are still very uncertain; this led Champredon *et* 217
al. [15] to adopt a Latin hypercube sampling (LHS) strategy [22] that evaluates model 218
outcomes over a range of parameter values. In order to make sure that our comparisons 219

among models apply across the entire space of reasonable parameter values, and in

order to evaluate the differential sensitivity of different model structures to parameter

values, we follow a similar protocol and perform LHS over a parameter set including

both the early- and late-stage transmission and duration parameters (β_P, D_P, β_D, D_D)

and contact/partnership parameters (ρ, c, c_u/c_w, and c_e/c_w). For the heterogeneity

model, the mean and squared coefficient of variation (CV) for the number of

non-cohabiting partners are sampled as well. We do not allow for uncertainties in

parameters that are directly related to the evolutionary process ($\beta_{\max}$, β_{50}, β_k, $D_{\max}$,

D_{50}, D_k, σ_M), instead using Shirreff et $al.$'s point estimates throughout [10].

Latin hypercube sampling is done as in Champredon et $al.$ [15]. For each parameter,

z, its range is divided into $N = 1000$ equal intervals on a log scale:

$$z_i = \exp\left(\log(z_{\min}) + [\log(z_{\max}) - \log(z_{\min})]\frac{i-1}{N-1}\right) \qquad i = 1, 2, 3, \ldots, N. \qquad (6)$$

Random permutations of these vectors form columns in a sample parameter matrix;

each row contains a different parameter set that is used for one simulation run.

Table 1 gives the ranges of the model parameters used for LHS. Parameter ranges

regarding contact and partnership dynamics (ρ, c, and c_e/c_w) are taken from

Champredon et $al.$ [15], whereas those regarding infection (β_P, D_P, β_D, and D_D) are

taken from Hollingsworth et $al.$ [18]. The remaining parameters are taken from Shirreff

et $al.$ [10].

One parameter in our model, the ratio of uncoupled to within-couple transmission

c_u/c_w, is needed to more flexibly contrast uncoupled and extra-couple transmission

dynamics within multi-strain models (see Appendix S1); it appears neither in either

Shirreff et $al.$ nor Champredon et $al.$'s models, so we need to pick a reasonable range for

it. Champredon et $al.$ [15] assume that the effective within-couple contact rate and

effective uncoupled contact rate have the same range of 0.05 - 0.25. Given Champredon

et $al.$'s parameter range, the possible maximum and minimum values of c_u/c_w are 5 and

1/5. Therefore, we use 1/5-5 as the range for the parameter c_u/c_w. Although this adds

more uncertainty to the parameter c_u — Champredon et $al.$'s range implies a 5-fold

difference whereas ours gives a 25-fold difference — we consider the wider range

appropriate, as little is not much known about the uncoupled transmission rate.

Two parameters, mean and the squared coefficient of variation (CV) of number of non-cohabiting partners, are sampled for heterogeneity in sexual activity. To allow for a wide range of uncertainty, range for the mean number of non-cohabiting partners was taken from unmarried men, as that was the group with the largest variability [19]. Omori *et al.* [19] give a very wide range for the coefficient of variation (≈ 0 - 20, corresponding to squared CV range of 0-400): we narrowed this range for CV^2 to 0.01-100. At the bottom end of the range, estimating that a group behaves perfectly homogeneously ($CV = 0$) is likely to be a sampling artifact; at the upper end, the estimate is also likely to be noisy because of the low mean value among married females (who have the largest range of CV). We assume that the number of non-cohabiting partners follows a Gamma distribution.

Simulation runs

One of the most difficult parts of model comparison is finding parameter sets that are commensurate with many different model structures. For the most part, our models are too complex to easily derive analytical correspondences among them. Given a numerical criterion, such as r (initial exponential growth rate) or $\mathcal{R}_0$ (intrinsic reproductive number), we can adjust one or more parameters by brute force to ensure that all of the models match according to that criterion. While $\mathcal{R}_0$ is often considered the most fundamental property of an epidemic, and might thus seem to be a natural matching criterion, here we focus on matching the initial growth rate r for several reasons. First, our primary interest is in the transient evolutionary dynamics of virulence, which are more strongly affected by r than $\mathcal{R}_0$. Second, r is more directly observable in real epidemics; r can be estimated by fitting an exponential curve to the initial incidence or prevalence curves [23], while $\mathcal{R}_0$ typically requires either (1) knowledge of *all* epidemic parameters or (2) calculations based on r and knowledge of the serial interval or generation interval of the disease [24]. Thus, we scale parameters so that every run has the same initial exponential growth rate in disease incidence.

In order to allow for all models to have equal initial exponential growth rate, r, we need to pick a parameter, s, such that $\lim_{s \to 0} r(s) = 0$ and $\lim_{s \to \infty} r(s) = \infty$. As adjusting either partnership change rate (i.e. partnership formation and dissolution

rate) or transmission rate fails this requirement for some of our models, we scaled partnership change rate and dissolution rate by the same factor of γ: $\beta_{\mathrm{adj}} = \gamma\beta_{\mathrm{base}}$, $c_{\mathrm{adj}} = \gamma c_{\mathrm{base}}$, $\rho_{\mathrm{adj}} = \gamma\rho_{\mathrm{base}}$. Since transmission rate is adjusted by the scale of γ, uncoupled and extra-couple transmission rates are adjusted as well. For the instantaneous-switching and implicit models, none of which track single individuals, only the transmission rate and partnership dissolution rate (in this case equivalent to the partnership change rate) are adjusted.

We run each model for each of 1000 parameter sets chosen by Latin hypercube sampling, with fixed starting conditions of mean $\log_{10}$ SPVL of 3 and epidemic size of 10^{-4}. After each run, initial exponential growth rate is calculated. Then, parameters are scaled so that the initial exponential growth rate is scaled to 0.04, a value that approximates the growth rates of Shirreff *et al.*'s original models. For calibration purposes, we run each model for only 500 years (full simulations are run for 4000 years), which is always long enough to capture the exponential growth phase of the model. We use a 4/5 order Runge-Kutta method (`ode45` from the `deSolve` package [25]) for all simulations. (For the heterogeneous model, approximately 10% of the samples failed due to numerical instability.)

Although each disease strain's core characteristic is its SPVL, the SPVL has one-to-one correspondences (based on eq. 2) with both the expected time to progression to AIDS and with the rate (probability per unit time) of HIV transmission. Because the time to progression (measured in years) is easier to interpret than SPVL (measured in $\log_{10}$ SPVL units), we summarize the virulence trajectories for each model run in terms of time to progression rather than SPVL. Because the time to progression is inversely related to SPVL (increasing SPVL decreases the time to progression), the time to progression is technically measuring inverse virulence rather than virulence (we did not think that reporting virulence as the rate of progression to AIDS, in units of years^{-1}, would help interpretability). For each model we derive the following summary statistics: minimum expected time to progression; time at which this minimum occurs (corresponding to peak virulence — this is also the time at which the maximum rate of progression, maximum SPVL, and maximum transmission rate occur); equilibrium time to progression; and the ratio of progression time at its minimum to the equilibrium value. Equilibrium progression time is calculated after 4000 years of simulated time.

Although most simulations reach equilibrium much earlier, we set our time horizon at a much later date as some simulation runs have slow rate of evolution depending on the parameter set and model assumptions.

Knowing the minimum progression time, timing of the minimum progression time/peak virulence, and equilibrium progression time provide sufficient detail to identify the overall shape of the virulence trajectory. In particular, knowing the timing of the peak virulence (how many years into the epidemic the virulence peaks) can help epidemiologists guess whether the virulence of an emerging pathogen is likely (1) to have peaked early, possibly even before the pathogen is detected spreading in the population, and decline over the remaining course of the epidemic; (2) to increase, peak, and decline over the foreseeable future; or (3) to increase very slowly, peaking only in the far future. To the extent that our simplistic model for HIV reflects reality, we would take the peak time of 150-300 years (Fig 1c) to mean that, in the absence of treatment, the epidemic would probably still be increasing in virulence.

Results

Our simplifications of Shirreff *et al.*'s model [10] reproduce its qualitative behaviour — in particular, its predictions of virulence dynamics — reasonably well. As we change r from 0.042 to 0.084 (the former value matching the initial rate of increase in prevalence in Shirreff *et al.*'s full model) the initial trajectory of increasing virulence brackets the rate from the original model (Fig 1a). However, our model produces lower peak virulence (≈ 4.3 vs. ≈ 4.6 $\log_{10}$ SPVL) and equilibrium virulence (≈ 4.25 vs. ≈ 4.5 $\log_{10}$ SPVL) than Shirreff's, even for matching initial incidence trajectories (i.e., $r = 0.042$ year^{-1}).

Changing the initial infectious density ($I(0)$), while it produces the expected changes in the initial epidemic trajectory (Supplementary material), has little effect on the virulence trajectory. Decreasing $I(0)$ makes SPVL peak slightly later and higher, because it allows a longer exponential-growth phase before the transition to endemic dynamics (Fig 1b). Decreasing the initial SPVL also leads to progressively later, higher peaks in SPVL (Fig 1c). In this case the delay in the peak is more pronounced than for low $I(0)$, because the rate of SPVL increase is eventually limited by the mutation rate.

The peak SPVL is actually larger for a lower starting SPVL, presumably because lower 341
SPVL also allows for a longer epidemic phase. However, the peaks are similar across the 342
entire range of initial conditions, because even in the most limited (high-$I(0)$, high-$\alpha(0)$) 343
cases HIV is able to evolve most of the way to the SPVL that would be optimal in an 344
indefinitely prolonged growth phase. 345

Across the entire range of parameters covered by the LHS analysis, all of the classes 346
of models we considered produce qualitatively similar virulence trajectories, which we 347
quantify in terms of population mean $\log_{10}$ SPVL (Fig 2: higher population mean $\log_{10}$ 348
SPVL corresponds to higher virulence). Although the speed of virulence evolution 349
varies, leading to wide variation in the peak SPVL (ranging from 3 to 5.5) because HIV 350
can evolve farther toward its exponential-phase optimum before the transition to the 351
endemic phase, virulence peaks in all models between 200 and 300 years. 352
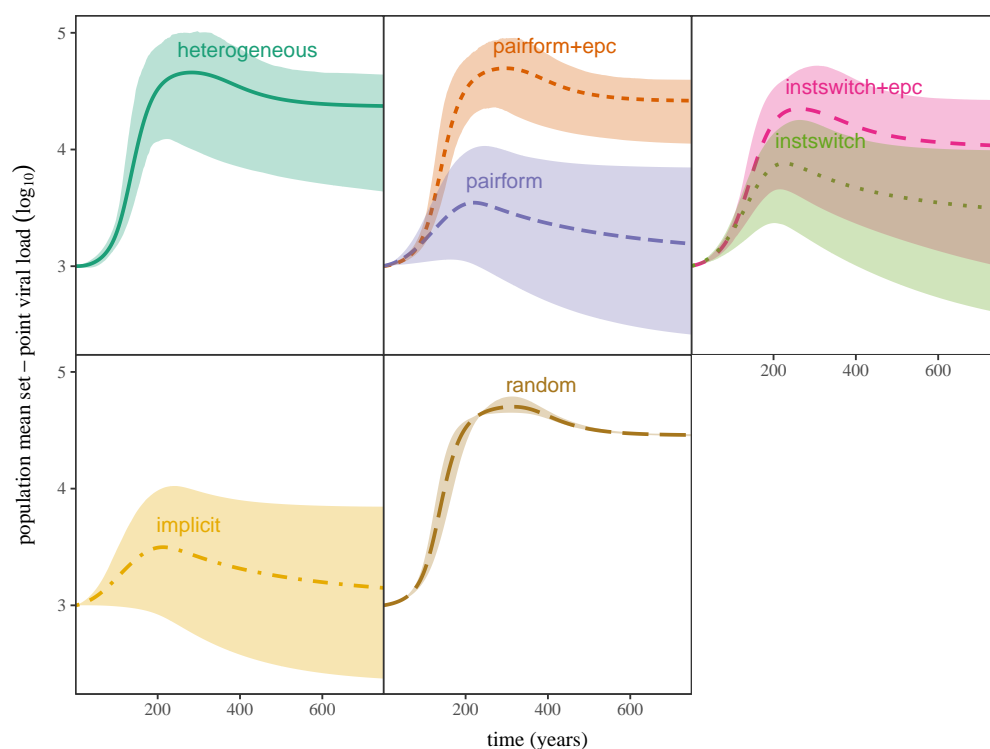
Fig 2. Envelopes of virulence trajectories (population mean $\log_{10}$ SPVL) under all models. Models arranged in order of complexity of contact structure, from "random" (least complex) to "hetero" (most complex). All models were run until $t = 4000$ years using parameters from 1000 Latin hypercube samples to illustrate the range of dynamics each model can exhibit. Envelopes contain the middle 95% of trajectories (i.e. we select all points between the 0.025 and 0.975 quantiles for each model at each year), while lines within the envelopes show mean trajectories. Truncated series (up to year 800) are shown here.

Our chosen summary statistics (peak time, maximum mean $\log_{10}$ SPVL, equilibrium   353

mean $\log_{10}$ SPVL, and minimum mean progression time) all vary considerably across   354

models (Fig 3). We first consider the models of intermediate realism: implicit,   355

instantaneous-switching with and without extra-pair contact, and pair formation   356

without extra-pair contact. Some parameter sets for these models lead to low   357

equilibrium virulence (2.3-3 $\log_{10}$ SPVL). For these data sets, virulence may either   358

increase from its initial value, reaching an early peak (≈ 200 years) between 3 and 4   359

$\log_{10}$ SPVLand then declining to a lower equilibrium value, or in extreme cases   360

virulence may decline immediately, leading to a peak virulence (as we have defined it)   361

equal to the starting value of $\alpha(0) = 3$ $\log_{10}$ SPVLat $t = 0$. (Fig 4). At the opposite   362

extreme, parameter sets that produce high equilibrium virulence (4.7 $\log_{10}$ SPVL) also   363

produce late peaks (> 200 years) and high peak virulence (5.6 $\log_{10}$ SPVL).   364
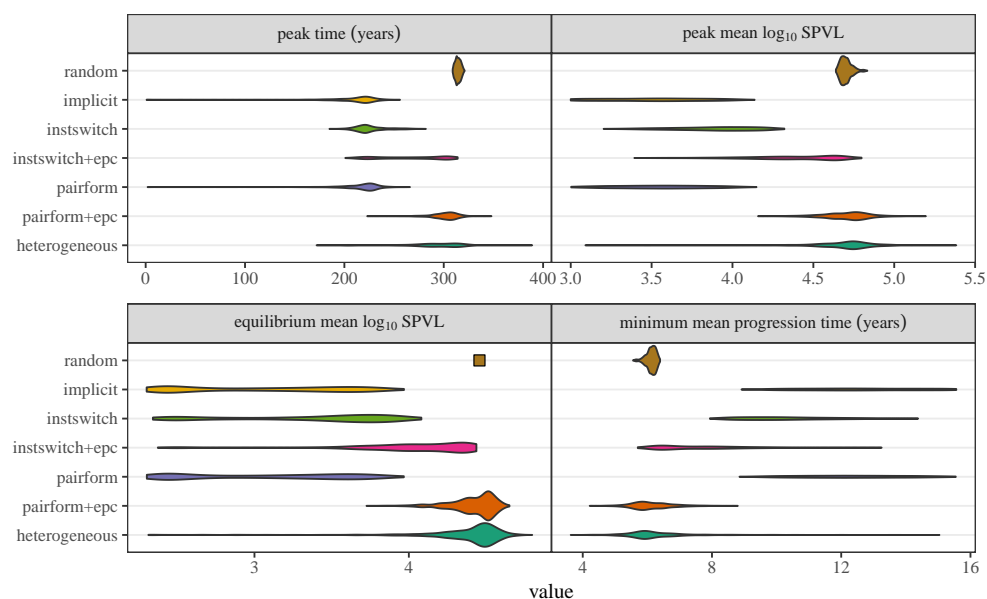

Fig 3. Univariate distributions of summary statistics. Models arranged in order of complexity, as in Fig 2. For every simulation, three summary statistics were obtained: peak time, maximum mean $\log_{10}$ SPVL, and equilibrium mean $\log_{10}$ SPVL. Peak time is the time at which the SPVL reaches its maximum for each simulation run (maximum mean $\log_{10}$ SPVL). Equilibrium mean $\log_{10}$ SPVL is taken by taking the mean $\log_{10}$ SPVL value at $t = 4000$. Minimum mean progression time is obtained by applying the Hill function to maximum mean $\log_{10}$ SPVL of each run. Because the distribution of equilibrium mean $\log_{10}$ SPVL (lower left panel) for the random mixing model is very narrow, it has been replaced by a point in order to preserve the vertical axis scaling.

The most striking aspect of the univariate comparisons in Fig 3, (and the bivariate   365

comparisons in Fig 4) is the similarity between the results of the least (random mixing)   366

Fig 4. Pairs plot: bivariate relationships among summary statistics for each model structure. Three summary statistics obtained from each simulation have been plotted against each other to visualize the relationship among the summary statistics and to allow for clearer comparison between models. Surprisingly, the implicit model, an approximation for instantaneous partnership formation model (instswitch), shows an almost identical trend with a model that has pair formation dynamics (pairform). Dashed line in equilibrium vs. peak virulence plot shows 1:1 line, where equilibrium and peak virulence are equal. 100 values were sampled from each model to allow for clearer distinction between the models

and the most complex (pair formation with extra-pair contact, pairform+epc with heterogeneity) models. The random-mixing model has the lowest variability, because it is unaffected by uncertainty in pair formation and extra-pair contact parameters, but otherwise the virulence dynamics of these three extreme models are remarkably similar. This phenomenon is driven by the strong effects of extra-pair contact in the model with explicit pair formation and extra-pair contact ("pairform+epc" in Figs 2-5). When

367
368
369
370
371
372

individuals spend time uncoupled between partnerships, and when these single
individuals can transmit disease to coupled individuals, the resulting unstructured
mixing overwhelms the effect of structured mixing within partnerships, leading to
mixing that is effectively close to random. Once unstructured mixing is strong, adding
realistic heterogeneity of mixing to the model has little effect other than increasing the
variability in the outcomes.

These differences are practically as well as scientifically important. The
random-mixing, pairform+epc, and heterogeneous models all predict high population
mean $\log_{10}$ SPVL at the virulence peak (median (95% CI) = 4.7 (4.65-4.79), 4.72
(4.37-4.96), 4.72 (4.09-5.03)). In contrast, the implicit model predicts a much lower peak
$\log_{10}$ SPVL value: 3.52 (3-4.02) years. In terms of mean time to progression to AIDS,
The random-mixing, pairform+epc, and heterogeneous models predict rapid progression
to AIDS at the virulence peak (median/95% CI = 6.1 (5.7-6.3), 6.02 (5.04-7.7), 6.03
(4.8-9.2)), while the implicit model predicts minimum progression times about twice as
long (12.5 (9.6-15.6) years). The corresponding differences in mean within-couple
transmission probability at the peak are even more extreme, about a fourfold difference:
0.249 (0.24-0.26), 0.252 (0.19-0.28), and 0.252 (0.15-0.28) per year for the random and
pairform+epc models vs. 0.059 (0.02-0.13) per year for the implicit model (see
Appendix S2 for plots showing univariate summaries of expected progression time to
AIDS and transmission probability).

The bivariate relationships (Fig 4) help distinguish the results of different models
with similar univariate dynamical summaries. While the relationship between
equilibrium $\log_{10}$ SPVL and peak time is similar for all model structures (top left
panel), the other relationships show more variation. In particular, the implicit and
pair-formation (without extra-pair contact) are very similar to each other, but distinct
from the other models. We still do not have a convincing explanation for this
distinction; we would have expected the implicit model to be most similar to the the
instantaneous-switching model without extra-pair contact, which most closely matches
its derivation. However, we note that the implicit model derivation is based on defining
the force of infection to match a scaled version of $\mathcal{R}_0$, and as such would be expected to
match the equilibrium behaviour but not necessarily the epidemic-phase behaviour of a
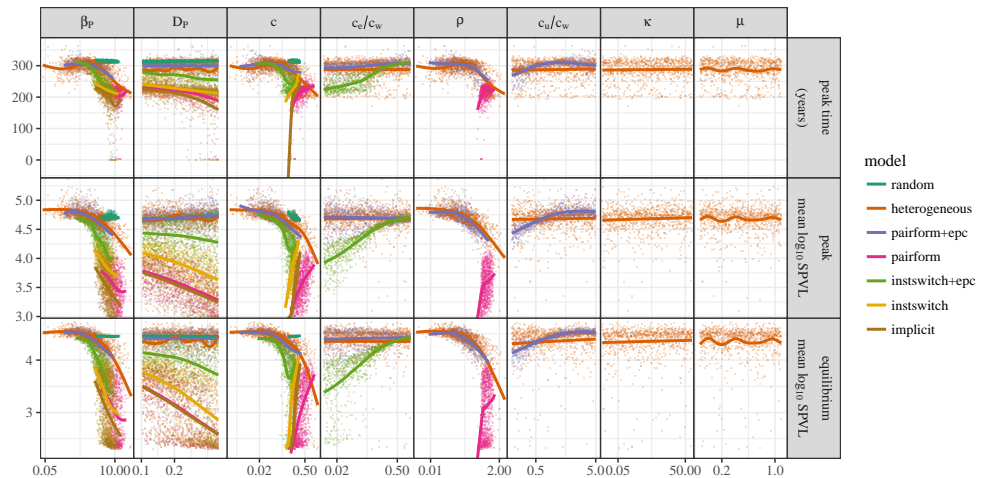model with explicit partnership dynamics.

Fig 5. Sensitivity plot. For most parameters in the Latin hypercube samples, the figure shows the distribution (points) and trend (smooth line) of the summary statistic as a function of the parameter value, some of which have adjusted to achieve the standard initial epidemic growth rate (see Methods).

Opposing trends in summary statistics with respect to partnership dissolution rate, c, (third column) between the model with and without extra-pair contact illustrates that similar model with different assumptions may yield opposing predictions.

Finally, the sensitivity plot (Fig 5) shows the effects of each parameter on the summary statistics. The most notable difference can be observed by comparing the scaled parameters (e.g. β_P, β_D, c, ρ) with the unscaled parameters (e.g. D_P, D_D, c_e/c_w, c_u/c_w, κ, μ); note that the effects of β_D and D_D are not shown in Fig 5 as their patterns are so similar to β_P and D_P, respectively.

One of the distinguishing features of the scaled parameters (in comparison to the unscaled parameters) is that the range of the parameters is compressed (relative to other models) for models without extra-pair contact as those models require parameters to be scaled by a large value in order to achieve a given r of 0.04. Models with extra-pair contact show wide range parameters as they can display a wide range of dynamics depending on c_e/c_w (as well as c_u/c_w for models with uncoupled mixing) and thus require wide range of scaling factors to achieve the target r. Parameter range for random-mixing model is severely compressed because latin hypercube sampling does not affect the prevalence trajectory very much, and scaling factor does not vary much across parameters.

For parameters involved in partnership turnover (c and ρ), main distinction is between models with and without extra-pair contact. With extra-pair contact, we get

gradual decrease in peak time, maximum $\log_{10}$ SPVL, and equilibrium $\log_{10}$ SPVL with 422
increasing turnover rates; Otherwise we get increase in summary statistics. Although we 423
have not verified this idea analytically, we speculate that increase in turnover rates 424
diminishes the effect of epc (equivalently increasing relative importance of 425
within-partnership contact), therefore selecting for lower $\log_{10}$ SPVL. For models 426
without epc, increased rates of partnership change decrease the level of structured 427
mixing (mimicing epc models), resulting in selection for higher $\log_{10}$ SPVL. 428

We also note that the implicit model and the instantaneous partnership formation 429
model show similar patterns in scaled parameters. In fact, we can see that the effect of 430
partnership dissolution rate, c, on equilibrium $\log_{10}$ SPVL of the two models are almost 431
identical (in Fig 4, they have distinguishing features). Lastly, we note that increase in 432
transmission rates (β_P and β_D) has a decreasing effect on summary statistics for all 433
models except for random mixing model. 434

Surprisingly, once calibration is taken into account, the unscaled parameters have 435
little to no effect overall. Increase in duration (D_P, D_D) in the primary and disease 436
stages generally decrease the equilibrium virulence, peak virulence, and peak time, 437
although the models with uncoupled mixing and random mixing model have high, 438
relatively constant values with respect to these parameters. The ratio of extra-pair to 439
within-pair contact (c_e/c_w) affects summary statistics in the instantaneous-switching 440
+epc model, but not the pair-formation+epc model (probably because the uncoupled 441
individuals present in the pair-formation+epc model make extra-pair contact by coupled 442
individuals less important). Similarly, increasing the ratio of uncoupled to within-pair 443
contact, c_u/c_w, increases peak and equilibrium $\log_{10}$ SPVL and delays peak time of the 444
pair-formation+epc model but has almost no effect on the heterogeneous model. 445
Neither the uncoupled contact rate nor the mean (μ) or CV^2 of the number of 446
non-cohabiting sexual partners has much systematic effect in the heterogeneous model. 447
discussion of effect of heterogeneity and vital dynamics here (referring 448
to supp material) 449

Discussion

All models must simplify the world. Many constraints — among them data availability, computation time, and code complexity — drive the need for parsimony, with different constraints applying in different contexts. The critical question that modelers must ask is whether the simplified model gives adequate answers, or whether the simplifications lead to qualitative or quantitative errors. This question is especially important for modelers who are hoping that their conclusions will guide management decisions.

In the particular example of HIV virulence eco-evolutionary dynamics and the complexity of contact structures we reach the slightly ironic conclusion that the effort put into building a more realistic model essentially cancels out, putting us back where we started when used a naive random-mixing contact model. However, we are not quite back where we started, as the complex models lead to wider, presumably more realistic confidence intervals on the predictions. In general, unstructured mixing — whether occurring through purely random mixing, or through extra-pair contact and contact among people outside of stable partnerships — tends to drive faster virulence evolution, leading to higher peak virulence and lower times to progression at the peak time.

Taking further steps to make the model even more realistic might add further structure, making the random-mixing model predictions less accurate. For example, our model forms partnerships randomly, and assumes that extra-pair contact is randomly mixing across the population; one could instead model extra-pair contact as arising from multiple concurrent partnerships (some, such as contact with sex workers, of very short duration) and/or more structured partnership formation (by age, ethnicity, or behaviour group). The effects of other realistic complications such as explicit modeling of two sexes (both in contact structure and differential transmission probabilities), temporal and spatial variation in epidemic processes, or presence of genetic variation in hosts are harder to predict.

Parameterization is one of the biggest challenges of epidemiological modeling. In addition to following Champredon *et al.* [15] by doing Latin hypercube sampling across a wide range of epidemiological parameters, we calibrated each set of parameters to the same initial epidemic growth rate, chosen to match the results of previous models [10]. Previous models in this area have drawn their parameters from cohort studies from the

1990s [18, 26] rather than doing any explicit calibration to epidemic curves, but they give reasonable order-of-magnitude growth rates (≈ 0.04 year^{-1}) for the early stages of the HIV epidemic (although considerably lower than estimates of $\approx 0.07 - 0.1$ year^{-1} based on population genetic reconstructions [27]). However, our reason for calibrating was not to match any specific observed epidemic, but rather to make sure that we were making meaningful comparisons across a range of models with radically different contact structures, and hence involving different interpretations of the same quantitative parameters. For example, in models with instantaneous switching the partnership dissolution rate c is identical to the partnership formation rate; in models with explicit partnership formation, the partnership formation rate is also c at equilibrium, but might vary over the course of an epidemic. It is not obvious whether models with equal parameters but different structures should be directly compared; calibration solves this problem.

More generally, any model that wants to be taken seriously for management and forecasting purposes should be calibrated to *all* available data, using informative priors to incorporate both realistic distributions of uncertainty for all parameters from independent measurements [28] and calibration from population-level observations of epidemic trajectories. Such a procedure would also be an improvement on the common — although not universal — practice, which we have followed here, of assessing uncertainty over uniform ranges rather than using distributions that allow more continuous variation in support over the range of a parameter.

Researchers have documented that HIV virulence and set-point viral load are changing, on time scales comparable to those portrayed here (e.g., compare Fig 2 to Herbeck *et al.*'s estimated rate of change of 1.3 $\log_{10}$ SPVL per century [95% CI -0.1 to 3] [29]), and have begun to build relatively realistic models that attempt to describe how interventions such as mass antiretroviral therapy (ART) can be expected to change the trajectory of virulence evolution [12, 30, 31]. While these efforts are well-intentioned, we caution that structural details that are currently omitted from these models could significantly change their conclusions.

Acknowledgements 510

Supporting Information 515

Appendix S1: Model Details 516

Appendix S2: Dynamics of Transmission and Virulence 517

References

1. Alizon S, Hurford A, Mideo N, van Baalen M. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. J Evol Biol. 2009;22:245–259. doi:10.1111/j.1420-9101.2008.01658.x.

2. Ebert D, Bull JJ. Challenging the trade-off model for the evolution of virulence: is virulence management feasible? Trends Microbiol. 2003;11(1):15–20.

3. Alizon S, Michalakis Y. Adaptive virulence evolution: the good old fitness-based approach. Trends in Ecology & Evolution. 2015;30(5):248–254. doi:10.1016/j.tree.2015.02.009.

4. Dwyer G, Levin SA, Buttel L. A simulation model of the population dynamics and evolution of myxomatosis. Ecol Monog. 1990;60:423–447.

5. Mackinnon MJ, Read AF. Genetic relationships between parasite virulence and transmission in the rodent malaria *Plasmodium chabaudi*. Evolution. 1999; p. 689–703.

6. Jensen KH, Little T, Skorping A, Ebert D. Empirical support for optimal virulence in a castrating parasite. PLoS Biol. 2006;4(7):e197.

7. De Roode JC, Yates AJ, Altizer S. Virulence-transmission trade-offs and population divergence in virulence in a naturally occurring butterfly parasite. Proceedings of the National Academy of Sciences. 2008;105(21):7489–7494.

8. Fraser C, Hollingsworth TD, Chapman R, de Wolf F, Hanage WP. Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. PNAS. 2007;104:17441–17446.

9. Fraser C, Lythgoe K, Leventhal GE, Shirreff G, Hollingsworth TD, Alizon S, et al. Virulence and Pathogenesis of HIV-1 Infection: An Evolutionary Perspective. Science. 2014;343(6177):1243727. doi:10.1126/science.1243727.

10. Shirreff G, Pellis L, Laeyendecker O, Fraser C. Transmission Selects for HIV-1 Strains of Intermediate Virulence: A Modelling Approach. PLoS Computational Biology. 2011;7(10):e1002185. doi:10.1371/journal.pcbi.1002185.

11. Herbeck JT, Mittler JE, Gottlieb GS, Mullins JI. An HIV Epidemic Model Based on Viral Load Dynamics: Value in Assessing Empirical Trends in HIV Virulence and Community Viral Load. PLoS Comput Biol. 2014;10(6):e1003673.

12. Herbeck JT, Mittler JE, Gottlieb GS, Goodreau SM, Murphy JT, Cori A, et al. Evolution of HIV virulence in response to widespread scale up of antiretroviral therapy: a modeling study. Virus Evolution. 2016;2(2):vew028. doi:10.1093/ve/vew028.

13. Day T, Proulx SR. A General Theory for the Evolutionary Dynamics of Virulence. The American Naturalist. 2004;163(4):E40–E63. doi:10.1086/382548.

14. Alizon S. The Price equation framework to study disease within-host evolution. Journal of Evolutionary Biology. 2009;22(5):1123–1132. doi:10.1111/j.1420-9101.2009.01726.x.

15. Champredon D, Bellan S, Dushoff J. HIV Sexual Transmission Is Predominantly Driven by Single Individuals Rather than Discordant Couples: A Model-Based Approach. PLoS ONE. 2013;8(12):e82906. doi:10.1371/journal.pone.0082906.

16. Dietz K, Hadeler KP. Epidemiological models for sexually transmitted diseases. Journal of Mathematical Biology. 1988;26(1):1–25.

17. Kretzschmar M, Dietz K. The effect of pair formation and variable infectivity on the spread of an infection without recovery. Mathematical Biosciences. 1998;148(1):83–113.

18. Hollingsworth TD, Anderson RM, Fraser C. HIV-1 Transmission, by Stage of Infection. Journal of Infectious Diseases. 2008;198(5):687–693. doi:10.1086/590501.

19. Omori R, Chemaitelly H, Abu-Raddad LJ. Dynamics of non-cohabiting sex partnering in sub-Saharan Africa: a modelling study with implications for HIV transmission. Sexually Transmitted Infections. 2015;doi:10.1136/sextrans-2014-051925.

20. May RM, Anderson RM. The Transmission Dynamics of Human Immunodeficiency Virus (HIV). Philosophical Transactions of the Royal Society of London Series B, Biological Sciences. 1988;321(1207):565–607.

21. Delva W, Leventhal GE, Helleringer S. Connecting the dots: network data and models in HIV epidemiology. AIDS. 2016;30(13):2009–2020. doi:10.1097/QAD.0000000000001184.

22. Blower SM, Hartel D, Dowlatabadi H, Anderson RM, May RM. Drugs, Sex and HIV: A Mathematical Model for New York City. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 1991;331(1260):171–187. doi:10.1098/rstb.1991.0006.

23. Ma J, Dushoff J, Bolker BM, Earn DJD. Estimating Initial Epidemic Growth Rates. Bulletin of Mathematical Biology. 2014;76(1):245–260. doi:10.1007/s11538-013-9918-2.

24. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. Proceedings of the Royal Society of London B: Biological Sciences. 2007;274(1609):599–604. doi:10.1098/rspb.2006.3754.

25. Soetaert K, Petzoldt T, Setzer RW. Solving Differential Equations in R: Package deSolve. Journal of Statistical Software. 2010;33(9):1–25. doi:10.18637/jss.v033.i09.

26. Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, Laeyendecker O, et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. Journal of Infectious Diseases. 2005;191(9):1403–1409.

27. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. The early spread and epidemic ignition of HIV-1 in human populations. Science (New York, NY). 2014;346(6205):56–61. doi:10.1126/science.1256739.

28. Elderd BD, Dukic VM, Dwyer G. Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. Proceedings

of the National Academy of Sciences. 2006;103(42):15693 –15697. doi:10.1073/pnas.0600816103.

29. Herbeck JT, Müller V, Maust BS, Ledergerber B, Torti C, Di Giambenedetto S, et al. Is the virulence of HIV changing? A meta-analysis of trends in prognostic markers of HIV disease progression and transmission. AIDS (London, England). 2012;26(2):193–205. doi:10.1097/QAD.0b013e32834db418.

30. Payne R, Muenchhoff M, Mann J, Roberts HE, Matthews P, Adland E, et al. Impact of HLA-driven HIV adaptation on virulence in populations of high HIV seroprevalence. Proceedings of the National Academy of Sciences. 2014;111(50):E5393–E5400. doi:10.1073/pnas.1413339111.

31. Roberts HE, Goulder PJ, McLean AR. The impact of antiretroviral therapy on population-level virulence evolution of HIV-1. Journal of The Royal Society Interface. 2015;12(113):20150888.