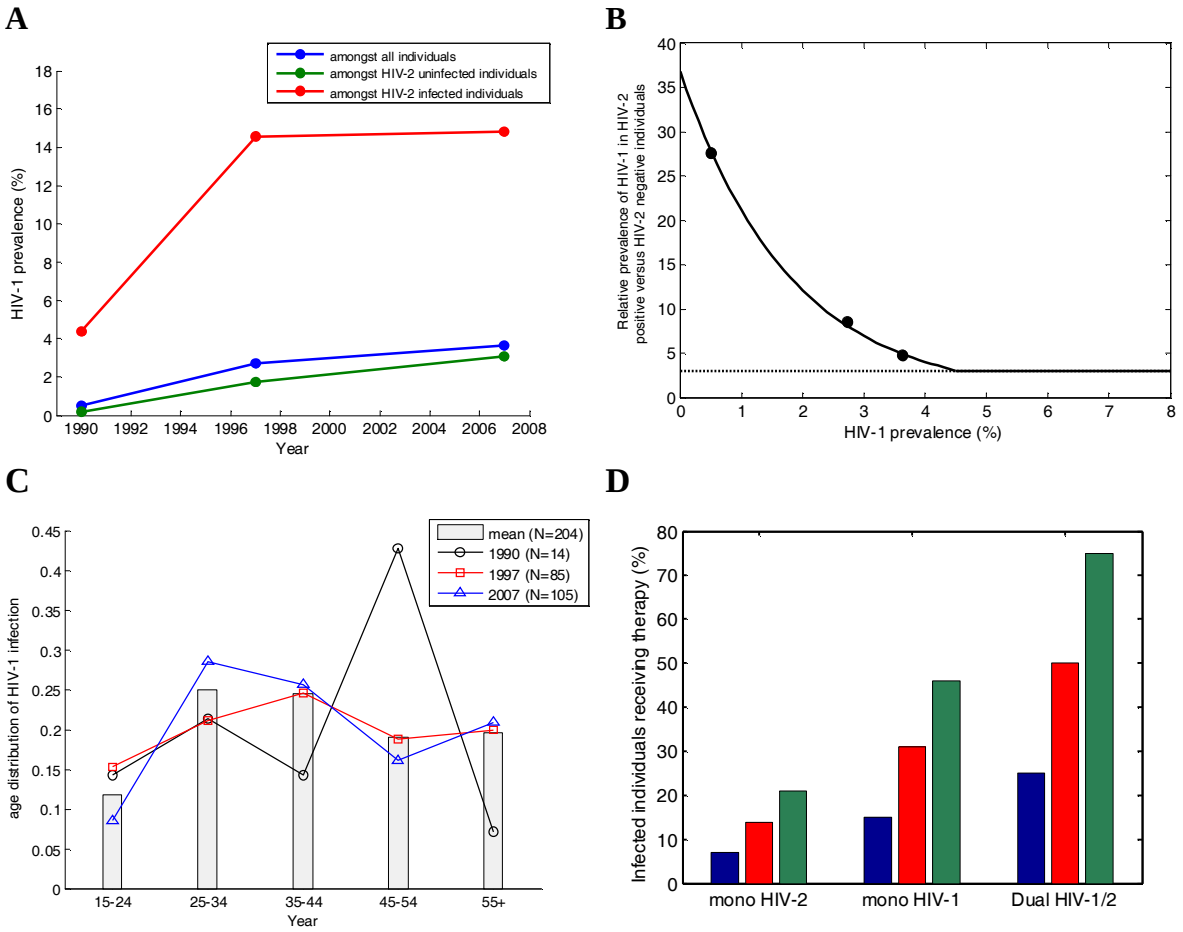
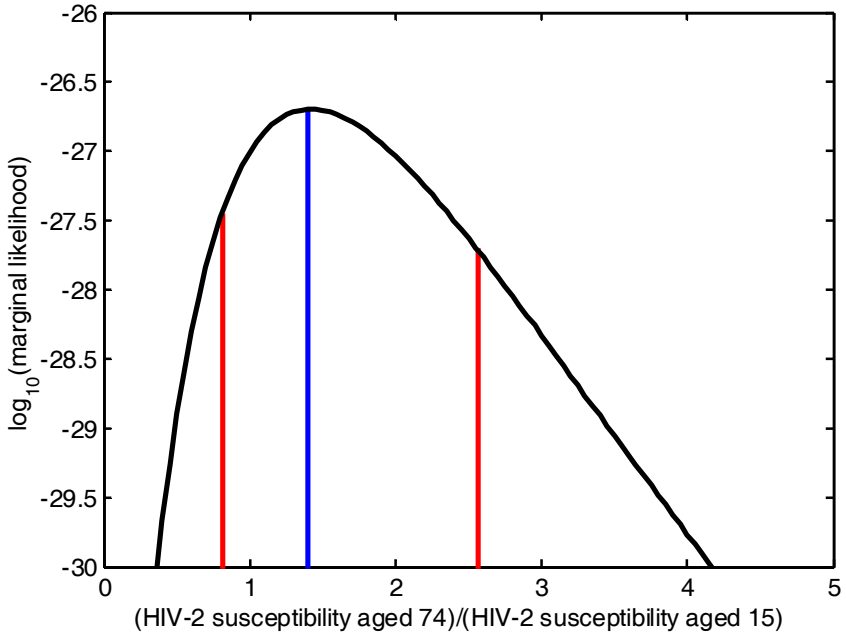


Figure S1. The observed and inferred distribution of HIV-1 and therapy amongst the population in Caió.

A) The observed prevalence of HIV-1 in Caió amongst HIV-2 infected individuals (red), HIV-2-uninfected individuals (green) and all individuals (blue; see g_t table S2), respectively. These data demonstrate that HIV-1 prevalence is higher amongst HIV-2 infected than HIV-2 uninfected hosts. B) This figure shows that the relative prevalence of HIV-1 in HIV-2 infected versus HIV-2 uninfected individuals decreases as HIV-1 prevalence increases. Black circles demonstrate data and the black line represents the relationship assumed in our model simulations. This relationship was devised by finding the maximum of 3 (see dashed line) and a fitted exponent curve $10^{24.1g+1.57}$ (see $u(g)$ table S2). C) The age distribution of HIV-1 infections, in 1990 (black), 1997 (red), 2007 (blue) and averaged across all years (grey bars). This demonstrates that the age distribution of HIV-1 infections has remained approximately constant over this period. For this reason, our model simulations have assumed that the age distribution of HIV-1 infections across the five age brackets (α_κ for $\kappa = 1:5$, table S2) is independent of time and given by the mean values demonstrated here (grey bars). D) Red bars show the best estimates of the prevalence of therapy beyond 2007 in mono HIV-2, mono HIV-1 and dual (HIV-1/2) infected individuals (see τ_t, $\tilde{\tau}_t$ and $\hat{\tau}_t$ respectively in table S2). Blue and green bars represent the therapy prevalence in the three host types used in sensitivity analysis. Blue bars assume that therapy prevalence is 50% lower that the best estimate. Green bars assume that therapy prevalence is 50% higher than the best estimate.

A



B

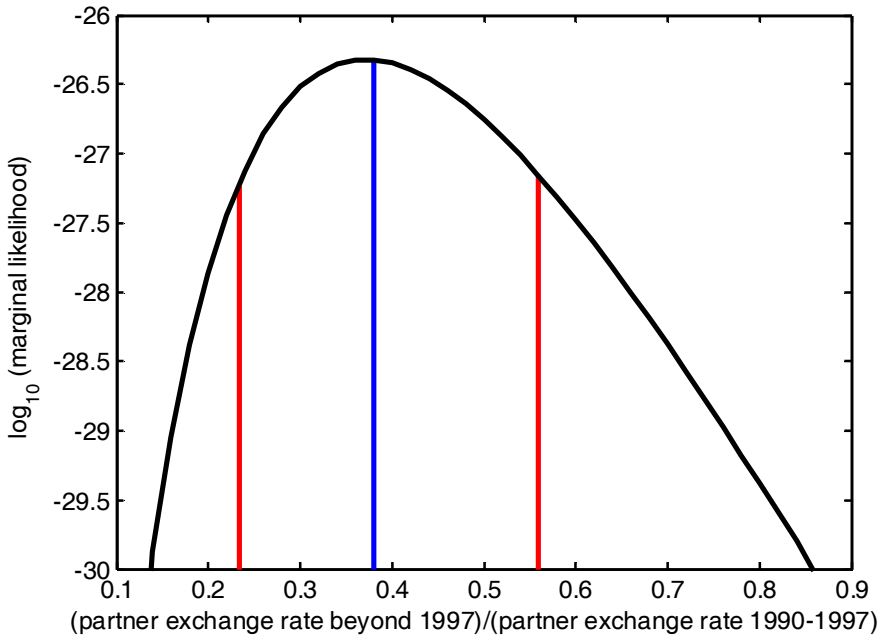
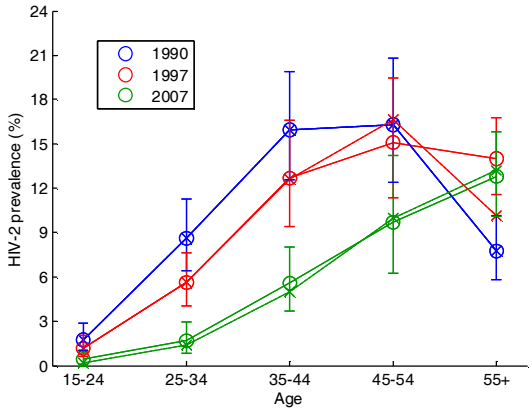


Figure S2. Marginal likelihood distributions demonstrating certainty in model parameters defining A) how susceptibility to infection varies with age and B) how the partner exchange rate varies over time.

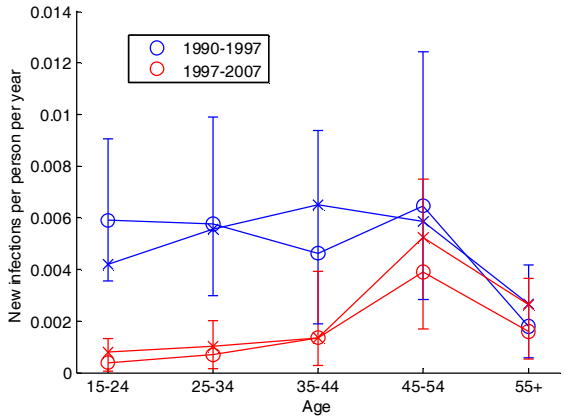
The logarithm of the marginal likelihood for A) the ratio of susceptibility of infection of hosts aged 74 to hosts aged 15 (β_{74}/β_{15}) and B) the ratio of the average rate of partner exchange beyond 1997 to between 1990 and 1997 ($\tilde{c}_{1998}/\tilde{c}_{1990}$). The marginal maximum likelihood estimator is represented by a blue line. 95% confidence intervals are represented by red lines. These metrics were calculated numerically and shown to converge on the presented values

(i.e. do not change with increasing parameter range or increasing sampling density). The results in A) demonstrate that susceptibility to infection most likely increases with age a little, with hosts aged 74 being 1.3 times more susceptible than hosts aged 15. However, confidence in this estimate (95% CI: 0.81-2.6) is such that we cannot rule out the possibility that susceptibility to infection decreases with age, is independent of age or increases more dramatically with age. The results in B) demonstrate support for risk behaviour being lower during 1997-2007 compared to during 1990-1997. We model this as a reduction in the overall rate of partner exchange. The marginal maximum likelihood estimator indicates that the rate of partner exchange is 0.38 times as fast during 1997-2007 compared to during 1990-1997. Confidence in this estimate (95% CI: 0.23-0.56) indicates that risk behaviour has decreased during this later period. For both parameters the estimate at the global maximum is equal (to 2 significant figures) to the marginal maximum likelihood estimator. Details of how the likelihood was calculated are provided in the supplementary text. Estimation assumes independence of the data.

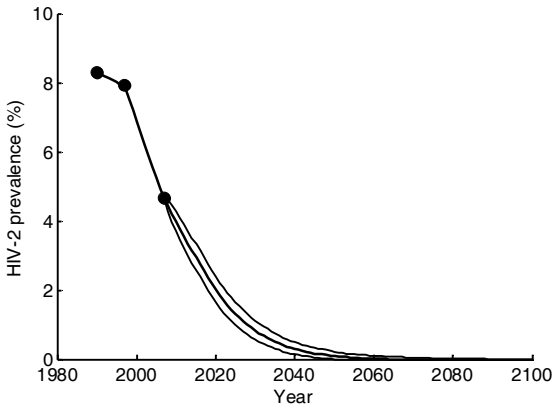
A


1990
1997
2007
HIV-2 prevalence (%)
Age
15-24
25-34
35-44
45-54
55+

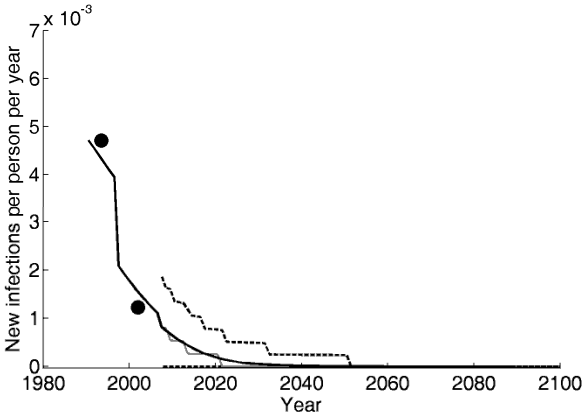
B


1990-1997
1997-2007
New infections per person per year
Age
15-24
25-34
35-44
45-54
55+

C


HIV-2 prevalence (%)
Year

D


x 10-3
New infections per person per year
Year

E

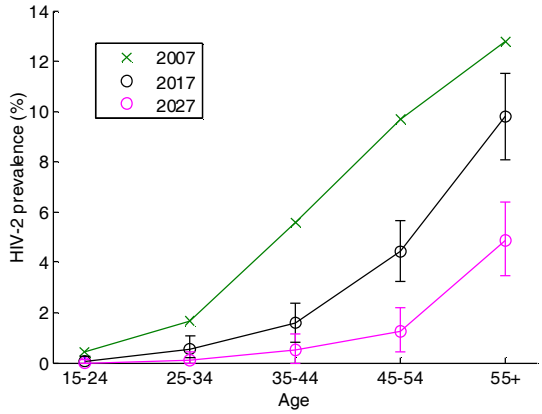
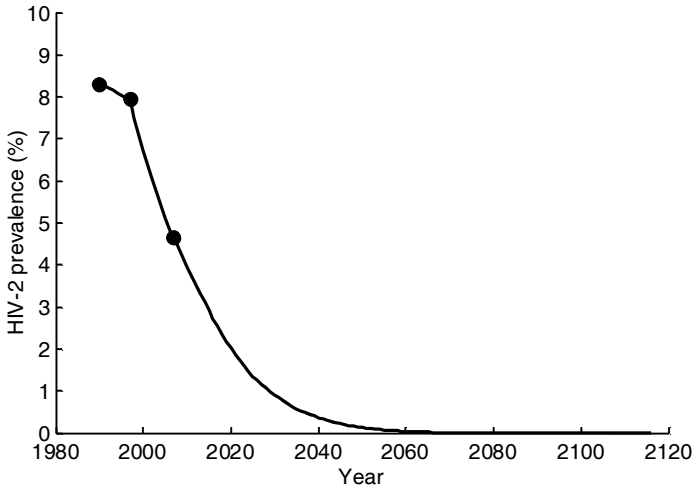

2007
2017
2027
HIV-2 prevalence (%)
Age
15-24
25-34
35-44
45-54
55+

Figure S3. Observations and predictions of HIV-2 incidence and prevalence using an alternative model in which a reduction in contact rate beyond 1997 applies only to younger cohorts

In these figures model predictions are made using an alternative model in which a reduction in contact rate beyond 1997 applies only to individuals born after 1952, i.e. younger than 45 years in 1997. For these model simulations, the average rate of partner exchange for individuals aged *a* at time *t* is defined as $c_{a,t} = \overline{c}_a \hat{c}_{a,t}$, where $\overline{c}_a$ is defined as for the simple model (Table S2) and $\hat{c}_{a,t} = 0.24$ (estimated) when $t \geq 1998$ and $t - a \geq 1952$ and $\hat{c}_{a,t} = 1$ (fixed), otherwise. The transmission probability per partnership (β_a) is assumed to be independent of age $\beta_a = 1$ (for all a). A) The prevalence of HIV-2 stratified by age in 1990 (blue), 1997 (red) and 2007 (green). The optimal model fit (dashed lines and crosses) closely reproduce these data (solid lines and circles) and lies within the (univariate) 95% confidence intervals surrounding the data. B) The yearly incidence per person of HIV-2 stratified by age in the periods 1990-1997 (blue) and 1997-2007 (red). Notice that this model (dashed lines and crosses) fits the incidence data (solid lines and circles) better than the simpler model (see Figure 1C). C) Model predictions (black line) and observations (circles) of the prevalence of HIV-2 amongst individuals over the age of 15 years. Between 1990 and 1997 the deterministic prediction is presented. Beyond 2007 the median (solid line) and 95% confidence interval (dashed lines) of 5000 stochastic predictions are presented. This optimum model fit suggests that HIV-2 prevalence will drop below 0.1% by 2047 (95% CI: 2040-2057). Extinction is predicted to occur in 2064 (2050-2089). D) The yearly incidence per person of HIV-2 in Caió amongst individuals aged over 15 years. The yearly incidence estimated during two periods (1990-1997 and 1997-2007) from data in Caió are plotted at the midpoints of the period (circles). Between 1990 and 2007 the deterministic model predictions (solid black line) of yearly incidence are shown. Beyond 2007 the mean of 5000 stochastic simulations is represented by a solid black line. Beyond 2007 the median of the simulations is presented by a solid grey line and 95% confidence intervals are represented by dashed lines. Note that the lower 95% interval is zero for all years. New infections are predicted to cease by 2036 (2022-2063) E) Model predictions of the median prevalence of HIV-2 amongst different age groups in 2017 and 2027 are estimated from 5000 simulations. 95% confidence intervals are also presented. For comparison, the 2007 prevalence data are also presented.

A



B

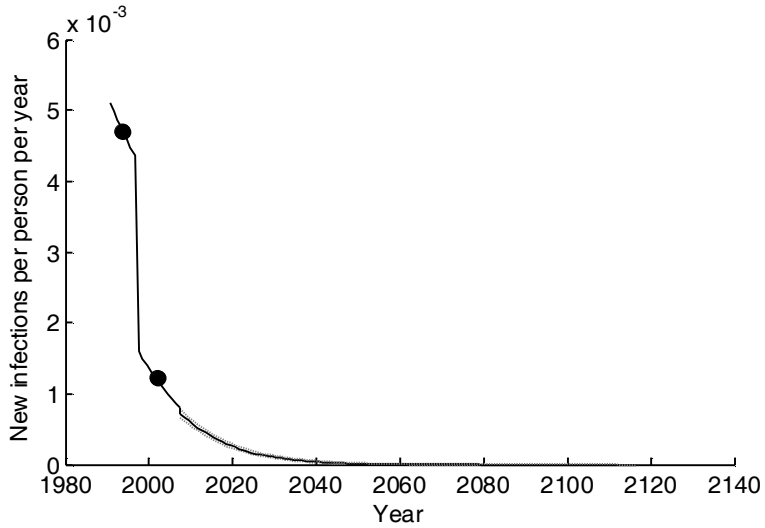


Figure S4. Model predictions of prevalence and incidence in a population the size of Guinea Bissau.

Model predictions of prevalence (A) and incidence (B) of HIV-2 are shown for a population the size of Guinea Bissau (approximately 1 million adults) using the model parameters estimated from Caió. Average incidence and prevalence estimates are the same as for the smaller model, however the stochastic confidence intervals are closer to the median. In B), the 95% CIs are represented as dashed grey lines. In A) the 95% CIs are not noticeably different to the median and are not visible. The estimated date of extinction is 2123 (95% CI: 2106-2154). New infections are predicted to cease in 2097 (2081-2130).

Table S1. A description of the model variables

Symbol	Description
a	Age
t	Time
i	HIV-1 infection status (0=uninfected, 1=infected)
j	Treatment status (0=untreated, 1=treated)
$S_{a,t}$	The number of HIV-2 uninfected hosts aged a at time t
$I_{a,t}$	The number of HIV-2 infected hosts aged a at time t

Table S2. A description of model parameters and auxiliary variables. Parameter values used in the model simulations are also listed.

Symbol	Description	Value used in optimal simulation
B	Population birth rate (years^{-1})	152 (estimated[†])
β_a	Transmission probability per partnership between an untreated HIV-2 positive individual and a susceptible host aged a	1.06×10^{-2} a=15:84 (independent of age) Sensitivity analysis: a=15:84 $10^{-5}(-3.56a+1173)$ (lower 95% CI: $\beta_{74}/\beta_{15}=0.8$) $10^{-5}(6.44a+833)$ (best estimate: $\beta_{74}/\beta_{15}=1.4$) $10^{-4}(1.85a+41)$ (upper 95% CI: $\beta_{74}/\beta_{15}=2.6$)
$c_{a,t}$	Average rate of partner exchange of hosts aged a, at time t (years^{-1})	This parameter is formulated to be the product of two parameters dependent solely upon a and t, respectively: $c_{a,t}=\bar{c}_a\tilde{c}_t$, whereby: $\bar{c}_a=\begin{cases}2.6 & a=15:19 \quad [34]\\ 3.7 & a=20:24 \quad [34]\\ 3.4 & a=25:29 \quad [34]\\ 2.5 & a=30:39 \quad [34]\\ 2.1 & a=40:49 \quad [34]\\ 1.9 & a=50:54 \quad \text{Inferred from [34]}\\ 1.6 & a=55:59 \quad \text{Inferred from [34]}\\ 1.3 & a=60:64 \quad \text{Inferred from [34]}\\ 1.0 & a=65:69 \quad \text{Inferred from [34]}\\ 0.7 & a=70:74 \quad \text{Inferred from [34]}\\ 0 & a=75:84\end{cases}$ $\tilde{c}_t=\begin{cases}1 & t=1990:1997 \quad \text{(fixed)}\\ 0.38 & t\geq1998 \quad \text{(estimated*)}\end{cases}$ Confidence intervals $\tilde{c}_t=0.23$ (lower 95% CI) $\tilde{c}_t=0.56$ (upper 95% CI)
$\gamma_{a,t}$	The net removal rate of HIV-2 uninfected hosts (years^{-1}). This is the fraction of HIV-2 uninfected hosts (including HIV-1 singly infected hosts) aged a who leave the population through death or emigration during year t.	$\gamma_{a,t}=\mu_a+r_{a,t}^{1,0}\bar{\mu}_a^0+r_{a,t}^{1,1}\bar{\mu}_a^1$
$\tilde{\gamma}_{a,t}$	The net removal rate of HIV-2 infected hosts (years^{-1}). This is the fraction of HIV-2 infected hosts (including dually HIV-1/2 infected hosts) aged a who leave the population through death or emigration during year t.	$\tilde{\gamma}_{a,t}=\mu_a+\tilde{r}_t^{0,0}\hat{\mu}_a^0+\tilde{r}_t^{0,1}\hat{\mu}_a^1+\tilde{r}_t^{1,0}\bar{\mu}_a^0+\tilde{r}_t^{1,1}\bar{\mu}_a^1$

μ_a	The net removal rate of HIV (1 or 2) uninfected hosts (years^{-1}).	0.024 a=15:24 (estimated†) 0.050 a=25:34 (estimated†) 0.010 a=35:44 (estimated†) 0.030 a=45:54 (estimated†) 0.001 a=55:69 (estimated†) 0.076 a=70:74 [9] 0.123 a=75:84 [9]
$\bar{\mu}_a^j$	The HIV-1 (mono or dual) related mortality rate (years^{-1}) of individuals with treatment status j.	$\bar{\mu}_a^0 = \begin{cases} 0.077 & a = 15:34 \\ 0.077 & a = 35:64 \\ 0.180 & a = 65:84 \end{cases}$ (untreated) inferred from [9] $\bar{\mu}_a^1 = 0.53 \times \bar{\mu}_a^0 \quad a = 15:34 \quad j = 1$ (treated) inferred from $[30,31]$ Sensitivity analysis: • $\bar{\mu}_a^j \to 0.5\bar{\mu}_a^j$ (50% slower) • $\bar{\mu}_a^j \to 1.5\bar{\mu}_a^j$ (50% faster)
$\hat{\mu}_a^j$	The HIV-2-related mortality rate (years^{-1}) of individuals with treatment status j.	$\hat{\mu}_a^0 = \begin{cases} 0.026 & a = 15:34 \\ 0.031 & a = 35:64 \\ 0.009 & a = 65:84 \end{cases}$ (untreated) [9,10] $\hat{\mu}_a^1 = 0.13 \times \hat{\mu}_a^0 \quad a = 15:84$ (treated) inferred from [31] Sensitivity analysis: • $\hat{\mu}_a^j \to 0.5\hat{\mu}_a^j$ (50% slower) • $\hat{\mu}_a^j \to 1.5\hat{\mu}_a^j$ (50% faster)
δ	Parameter defining the degree of age-structure in partnerships	1/3 (this corresponds to the age difference between sexual partners being less than 20 years for 81% of partnerships)$^{+}$ [23] Sensitivity analysis: • 0.14 (60% of partnerships with <20 year age gap) • 0.50 (100% of partnerships with <20 year age gap)
$r_{a,t}^{i,j}$	The fraction of HIV-2 uninfected hosts, aged a, at time t with HIV-1 infection status i and treatment status j	These values are set to ensure that • The prevalence of HIV-1 varies over time according to figure S1A (blue line). • The age distribution of HIV-1 infections is equal to the median age distribution shown in figure S1C (grey bars). • The relative prevalence of HIV-1 in HIV-2 infected versus HIV-2 uninfected hosts changes with HIV-1 prevalence as given by the fitted curve in figure S1B. • The prevalence of therapy in HIV-1 mono, HIV-2 mono and HIV 1/2 dual infected individuals is given by the rates provided in figure S1D. $r_{a,t}^{0,0} = 1 - \chi_{a,t} \qquad r_{a,t}^{0,1} = 0$ $r_{a,t}^{1,0} = (1-\tilde{\tau}_t)\chi_{a,t} \qquad r_{a,t}^{1,1} = \tilde{\tau}_t \chi_{a,t}$
$\tilde{r}_t^{i,j}$	The fraction of HIV-2 infected hosts at time t, with HIV-1 infection status i and treatment status j	$\tilde{r}_t^{0,0} = (1-\tau_t)(1-\tilde{\chi}_t) \qquad \tilde{r}_t^{0,1} = \tau_t(1-\tilde{\chi}_t)$ $\tilde{r}_t^{1,0} = (1-\hat{\tau}_t)\tilde{\chi}_t \qquad \tilde{r}_t^{1,1} = \hat{\tau}_t\tilde{\chi}_t$

$\chi_{a,t}$	Fraction of HIV-2 negative individuals who are infected with HIV-1 aged *a* at time *t*.	$\dfrac{\alpha_k g_t \sum_{a=15:74}\left(I_{a,t}+S_{a,t}\right)-\tilde{\chi}_t\sum_{a\in A_k} I_{a,t}}{\sum_{a\in A_k} S_{a,t}}$
$\tilde{\chi}_t$	Fraction of HIV-2 positive individuals who are also infected with HIV-1 at time *t*.	$\dfrac{g_t u(g_t)\sum_{a=15:74}\left(I_{a,t}+S_{a,t}\right)}{\sum_{a=15:74}\left(u(g_t)I_{a,t}+S_{a,t}\right)}$
g_t	Fraction of the total population infected with HIV-1 at time *t*.	$0.005+0.022\,(t-1990)/7 \quad t=1990:1996$ $0.027+0.009\,(t-1997)/10 \quad t=1997:2006$ $0.036 \quad t\geq 2007$ (see figure S1 A: blue line) Sensitivity analysis: $0.036-0.011(t-2007)/15 \; t=2007:2021$; $0.025 \; t\geq 2022$ declines by 30% $0.036+0.018(t-2007)/15 \; t=2007:2021$; $0.054 \; t\geq 2022$ increases by 50%
α_κ	Fraction of all HIV-1 infections in age bracket A_κ	$0.118 \quad \kappa=1 \quad (A_\kappa=[15:24])$ $0.250 \quad \kappa=2 \quad (A_\kappa=[25:34])$ $0.245 \quad \kappa=3 \quad (A_\kappa=[35:44])$ $0.191 \quad \kappa=4 \quad (A_\kappa=[45:54])$ $0.196 \quad \kappa=5 \quad (A_\kappa=[55:74])$ (see figure S1 C: grey bars)
$u(g)$	The ratio between the prevalence of HIV-1 in HIV-2 positive versus HIV-2 negative individuals as a function of the HIV-1 prevalence in the total population.	$\max\left(3,10^{24.1g+1.57}\right)$ (see figure S1 B)
τ_t	Fraction of mono HIV-2 infected patients with therapy at time *t*.	$0 \quad$ for $t<2007$ $0.14 \quad$ for $t\geq 2007$ (Table S3) Sensitivity analysis for t ≥ 2007 • 0.07 (50% lower than estimate) • 0.21 (50% higher than estimate)

$\tilde{\tau}_t$	Fraction of mono HIV-1 (mono or dual) infected patients with therapy at time *t*.	0 for $t < 2007$ 0.31 for $t \geq 2007$ (Table S3) Sensitivity analysis for t ≥ 2007 • 0.15 (50% lower than estimate) • 0.46 (50% higher than estimate)
$\hat{\tau}_t$	Fraction of dual HIV-1/2 infected patients with therapy at time *t*.	0 for $t < 2007$ 0.50 for $t \geq 2007$ (Table S3) Sensitivity analysis for t ≥ 2007 • 0.25 (50% lower than estimate) • 0.75 (50% higher than estimate)
π	Effectiveness of drug in reducing infectiousness of HIV-2 infected individuals and susceptibility to HIV-2	0.79 Based upon the observation that 79% of individuals alive after 36 months of therapy had undetectable viremia [31]. Sensitivity analysis • 0.5 • 1
d	Relative number of contacts made by HIV-1 singly infected individuals HIV-uninfected individuals verses HIV-2 uninfected individuals	4 Estimated from the relative numbers HIV-2 infections in HIV-1 positive versus HIV-1 negative individuals in 2007 [4]. Sensitivity analysis • 1 • 8
$\varpi_{a,t}$	Treatment induced, relative susceptibility to HIV-2 of individuals aged *a* at time *t*.	$\frac{r_{a,t}^{0,0} + d(r_{a,t}^{1,0} + r_{a,t}^{1,1}(1-\pi))}{r_{a,t}^{0,0} + d(r_{a,t}^{1,0} + r_{a,t}^{1,1})}$
$\lambda_{a,t}$	The fraction of susceptible hosts aged *a* who become infected per year at time *t*.	

$$\lambda_{a,t} = \beta_a \varpi_{a,t} c_{a,t} \left(\delta \frac{\sum_{b=15+20u}^{\min(74,34+20u)} (1-\pi\tilde{r}_t^{\bullet 1}) c_{b,t} I_{b,t}}{\sum_{b=15+20u}^{\min(74,34+20u)} c_{b,t}\left(I_{b,t}+S_{b,t}\right)} + \delta \frac{\sum_{b=\max(15,5+20m)}^{24+20m} (1-\pi\tilde{r}_t^{\bullet 1}) c_{b,t} I_{b,t}}{\sum_{b=\max(15,5+20m)}^{24+20m} c_{b,t}\left(I_{b,t}+S_{b,t}\right)} + (1-2\delta) \frac{\sum_{b=15}^{74} (1-\pi\tilde{r}_t^{\bullet 1}) c_{b,t} I_{b,t}}{\sum_{b=15}^{74} c_{b,t}\left(I_{b,t}+S_{b,t}\right)} \right)$$

where $\tilde{r}_b^{\bullet 1} = \tilde{r}_b^{0,1} + \tilde{r}_b^{1,1}$, $u = \text{floor}\left((a-15)/20\right)$ and $m = \text{floor}\left((a-5)/20\right)$

*Estimated by fitting the model to the age stratified incidence and prevalence data. †Estimated by scaling the age distribution of individuals in the model to equal the age distribution observed in the surveys [4]. +Although this metric was based upon data from married couples in rural Guinea Bissau [23], where at least one of the married couple was HIV-2 infected, we used these data because they represent the best source of information on age differences in sexual partnerships in the region. Furthermore, age difference was not a risk factor for HIV-2 in that study suggesting that the data are a valid representation of sexual partnerships amongst the whole population.

Table S3. Approximate fraction of HIV infected on therapy

Infection status	Infected in cohort and on therapy in 2014	Infected in 2014 (model prediction)	Approximate % of individuals on therapy
HIV-2 (mono)	14	97	14%
HIV-1 (mono)	38	124	31%
Dual HIV-1/2	10	20	50%

Precise data relating to the fraction of patients on therapy over time were not available. However it is known that therapy in Caió started in 2007, the same year as the last sero-survey. Information that we have on therapy uptake is available only for patients who were tested positive in one of the sero surveys and were receiving therapy in 2014 through the programme linked to the survey (column 2). A rough estimate of therapy uptake in 2014 is calculated by dividing the numbers receiving therapy in 2014 by the numbers, predicted by our model to be infected in 2014 (column 3). In our model projections we assume that therapy rates stay fixed at these rates beyond 2007. We note that the true therapy rates in 2014 may be a little higher than our estimates because they assume that only individuals who were included in one of the surveys and tested positive for HIV would be receiving treatment in Caió in 2014. Some individuals receiving therapy in 2014 may have been diagnosed since 2007 or diagnosed in individuals not included in any of the surveys. However, we expect the number of such individuals to be very low because the sero-surveys were the primary driver of diagnosis and treatment of HIV in the region around this time. It is also noteworthy that although access to antiretroviral therapy is gradually improving across Africa, no further sero-survey is planned for Caió. The percentage of HIV infected individuals who get diagnosed in the region may therefore decline in the future. To understand how uncertainty in these estimates may affect our model predictions, sensitivity analysis, assuming therapy prevalence 50% higher or lower than our best estimates after 2007, was performed (see table S4).

Table S4. Sensitivity analysis of model parameters

Univariate sensitivity analysis	Value used in primary model simulations	Value used in sensitivity analysis	Median year of extinction	Median date of last new infection	Median date when prevalence reaches 0.1%	Log likelihood
Ratio of susceptibility of infection of hosts aged 74 to hosts aged 15 years (β_{74}/β_{15})	1	0.8 (95% CI)	2068	2043	2050	-25.0
		1.4 (best fit)	2068	2043	2050	-24.4
		2.6 (95% CI)	2067	2043	2050	-25.5
Disease-related death rates ($\bar{\mu}_a^j, \hat{\mu}_a^j, \tilde{\mu}_a^j$)	See table S3	50% slower	2075	2047	2056	-24.7
		50% faster	2062	2041	2046	-28.0
Sexual partnerships with <20 years age gap	81%	60%	2072	2045	2052	-27.3
		100%	2063	2040	2048	-27.5
HIV-1 prevalence beyond 2027 (g_{2027})	3.6%	2.5% (30% lower)	2067	2043	2049	-24.6
		4.7% (50% higher)	2068	2043	2051	-24.6
Therapy prevalence beyond 2007 in mono HIV-2 (τ_{2007}), mono HIV-1 ($\tilde{\tau}_{2007}$) and dual HIV-1/2 ($\hat{\tau}_{2007}$) infected individuals (%)	14%, 31% 50%	7, 15, 25% (50% lower)	2068	2044	2050	-24.6
		21, 46, 75% (50% higher)	2068	2042	2050	-24.6
Relative number of contacts made by HIV-1 infected verses HIV-2 uninfected individuals (d)	4	1	2069	2044	2050	-24.6
		8	2067	2042	2049	-24.6
Effectiveness of drug in reducing infectiousness of HIV-2 infected individuals and susceptibility to HIV-2 (π)	0.79	0.5	2070	2045	2051	-24.6
		1	2067	2042	2050	-24.6
Primary model		See column 2	2068	2043	2050	-24.6
Multivariate lower limit		See parameters in red	2057	2037	2043	-35.0
Multivariate upper limit		See parameters in blue	2084	2054	2060	-27.3

All of the parameters in our model are derived from data, however, HIV related death rates ($\bar{\mu}_a^j, \hat{\mu}_a^j$ and $\tilde{\mu}_a^j$) vary somewhat across data sources and data on the extent to which contacts are segregated by age (determined by parameter δ) in Guinea-Bissau are limited. The ratio of susceptibility to infection of hosts aged 74 to hosts aged 15 was fitted to the age-stratified incidence and prevalence data, but certainty in this parameter was shown to be low. HIV-1 prevalence in Guinea-Bissau beyond 2007 is unknown and therapy rates are approximated from imperfect data. This table demonstrates how changes to the parameters defining these processes influence our results. Changes to only one of the parameters (univariate analysis) are explored in the first 5 rows. For comparison, results of the primary model are shown in row 6. The parameter combinations yielding the most disparate multivariate results are shown in column 3. The limits of multivariate analysis are shown in rows 7 and 8. For each parameter change, the model was refitted to the data ($\log_{10}$ likelihood in column 7). These simulations demonstrate that such parameter changes do not change our qualitative prediction that HIV-2 is going extinct in Caió. Regarding the quantitative predictions, changes to only one of these parameters varies the median date of extinction by 8 years at most. Changes to all parameters affect the median date of extinction by 16 years at most. It is noteworthy that the fit of the model to the data is reduced compared to the optimal fit, by the alternative parameter sets investigated here, providing support for the parameter values that we have used.

Supplementary Text. Additional details about the mathematical model of the spread of HIV-2 in Caió

S1. Model equations for the deterministic form of the model

Difference equations:

$$S_{15,t+1} = B \quad \{1\}$$

$$I_{15,t+1} = 0 \quad \{2\}$$

$$S_{a+1,t+1} = \left(1-\lambda_{a,t}\right)\left(1-\gamma_{a,t}\right)S_{a,t} \quad \{3\}$$

$$I_{a+1,t+1} = \left(1-\gamma_{a,t}\right)\lambda_{a,t}S_{a,t} + \left(1-\tilde{\gamma}_{a,t}\right)I_{a,t} \quad \{4\}$$

S2. Initial conditions of the model

Based upon the results of the 1990 census of Caió, we assumed in our simulations that there are 4000 adults (aged 15 and over) in the population. In addition the initial conditions (year 1990) of the model were fixed such that the distribution of hosts stratified according to age and infection status exactly matched that observed in the 1990 HIV survey.

S3. Modelling host turnover, competition with HIV-1 and the impact of therapy in the population.

In our model the population undergoes host turnover. New hosts are born into the youngest age group (age 15) with a birth rate, *B*, of 152 per year, and hosts of all ages can leave the population. Net removal rates vary with age, HIV-2 infection status and time and account not only for death, but also emigration and immigration. Furthermore they account for the impact of competitive exclusion of HIV-2 by HIV-1 through enhanced mortality of HIV-1/2 dually infected hosts (i.e. preferential removal of high risk individuals). The impact of therapy on removal rates is also included.

During year *t*, a fraction, $\gamma_{a,t}$, of HIV-2 uninfected hosts aged *a* are assumed to leave the population. These rates are formalised as the weighted sum ($\gamma_{a,t} = \mu_a + r_{a,t}^{1,0}\overline{\mu}_a^0 + r_{a,t}^{1,1}\overline{\mu}_a^1$) of the removal rates of HIV uninfected hosts (μ_a), untreated singly HIV-1 infected hosts ($\mu_a + \overline{\mu}_a^0$) and treated singly HIV-1 infected hosts ($\mu_a + \overline{\mu}_a^1$). The weightings, $r_{a,t}^{i,j}$, represent the fraction of HIV-2 uninfected hosts, aged *a*, at time *t* with HIV-1 infection status *i* (0= uninfected, 1=infected) and HIV-1 treatment status *j* (0=untreated, 1=treated). During year *t*, a fraction, $\tilde{\gamma}_{a,t}$, of HIV-2 infected hosts aged *a* are assumed to leave the population. These rates are formalised as the weighted sum ($\tilde{\gamma}_{a,t} = \mu_a + \tilde{r}_t^{0,0}\hat{\mu}_a^0 + \tilde{r}_t^{0,1}\hat{\mu}_a^1 + \tilde{r}_t^{1,0}\overline{\mu}_a^0 + \tilde{r}_t^{1,1}\overline{\mu}_a^1$) of the removal rate of untreated ($\mu_a + \hat{\mu}_a^0$) and treated ($\mu_a + \hat{\mu}_a^1$) singly HIV-2 infected hosts and untreated ($\mu_a + \overline{\mu}_a^0$) and treated ($\mu_a + \overline{\mu}_a^1$) dually HIV-1/2 infected individuals. The weightings, $\tilde{r}_t^{i,j}$ represent the fraction of HIV-2 infected hosts, at time *t* with HIV-1 infection status *i* and treatment status *j*.

HIV-2 related ($\hat{\mu}_a^0$) and HIV-1 related ($\bar{\mu}_a^0$) mortality rates of untreated individuals were estimated from Guinea Bissau [9,10]. HIV-2 related ($\hat{\mu}_a^1$) and HIV-1 (mono or dual) related ($\bar{\mu}_a^1$) mortality rates in treated individuals were estimated from elsewhere in West Africa [30,31]. Based upon the findings of others [9], dual HIV-1/2 related mortality rates were assumed to equal those of HIV-1 related mortality rates.

Because each of the three HIV surveys conducted in Caió included the majority of the adult population, the age distribution of the model population beyond 1990 was fitted to the distribution observed in the surveys. Because of the initial conditions, the age structure in the model in 1990 was also equal to that observed in the 1990 HIV survey (31%, 20%, 15%, 12% and 23% for the age groups 15-24, 25-34, 35-44, 45-54 and ≥55). This was achieved by fitting appropriate HIV-uninfected net removal rates (μ_a) for the different age categories. These net removal rates implicitly take account of all age-dependent immigration, emigration and death. We assume that individuals over the age of 70 will be settled, that is they do not immigrate or emigrate, and the net removal rates are equal to death rates provided elsewhere [9]. Notice that the fitted net removal rates are relatively low in individuals aged 55:69 years indicating that immigration into Caió may be prevalent in this age group. This is in keeping with the fact that many individuals, especially men, move away from rural regions into urban regions for work when they are young, but return to their rural home as they are approaching old age.

The weightings $r_{a,t}^{i,j}$ and $\tilde{r}_{a,t}^{i,j}$ are estimated by defining parameters τ_t and $\tilde{\tau}_t$, and auxiliary variables $\chi_{a,t}$ and $\tilde{\chi}_t$.

$$\begin{array}{llll} r_{a,t}^{0,0} = 1-\chi_{a,t} & r_{a,t}^{0,1} = 0 & r_{a,t}^{1,0} = (1-\tilde{\tau}_t)\chi_{a,t} & r_{a,t}^{1,1} = \tilde{\tau}_t\chi_{a,t} \\ \tilde{r}_t^{0,0} = (1-\tau_t)(1-\tilde{\chi}_t) & \tilde{r}_t^{0,1} = \tau_t(1-\tilde{\chi}_t) & \tilde{r}_t^{1,0} = (1-\hat{\tau}_t)\tilde{\chi}_t & \tilde{r}_t^{1,1} = \hat{\tau}_t\tilde{\chi}_t \end{array} \quad \{5\text{-}12\}$$

τ_t, $\tilde{\tau}_t$, and $\hat{\tau}_t$ define the fraction of mono HIV-2, mono HIV-1 and dually infected individuals with therapy at time *t*. Estimates for these parameters are listed in tables S2 and S3 and support for them are provided in the legend for table S3. $\chi_{a,t}$ defines the fraction of HIV-2 negative individuals who are infected with HIV-1 aged *a* at time *t*. $\tilde{\chi}_t$ defines the fraction of HIV-2 positive individuals who are also infected with HIV-1 at time *t*. These auxiliary variables were estimated using the data provided in figure S1 and formulated using additional parameters g_t, α_k and $u(g)$.

Between 1990 and 2007 the total prevalence of HIV-1 (g_t) was estimated to equal the prevalence estimated from data from Caió by interpolation (Figure S1 A blue line). This representation reveals that the total prevalence of HIV-1 increased dramatically between 1990 and 1997 and continued to increase, but at a slower pace between 1997 and 2007. For our primary estimates we therefore assumed that the trajectory of HIV-1 prevalence flattens off beyond 2007. In sensitivity analysis we also evaluated our model under the assumption that prevalence declined by 30% or increased by 50% year 2027 – 20 years beyond the last survey. Because data from 1990, 1997 and 2007 indicated that the age distribution of HIV-1 infections has remained approximately constant over time (Figure S1 C, grey bars) we assumed that it was fixed over time (α_k, see table below for more details). As might be

expected through increased risk behaviours of some individuals in the population, HIV-1 prevalence was higher amongst HIV-2 infected hosts (Figure S1 B) than HIV-2 uninfected hosts (Figure S1 A). To reflect this we fixed the ratio between the prevalence of HIV-1 in HIV-2 positive versus HIV-2 negative individuals as a function of the HIV-1 prevalence in the total population (Figure S1B). This relationship was devised by finding the maximum of 3 (dashed lines) and a fitted exponent curve to the available data points. All of the auxiliary variables and parameters used for defining $r_{a,t}^{i,j}$ and $\tilde{r}_t^{i,j}$ are described explicitly in table S2.

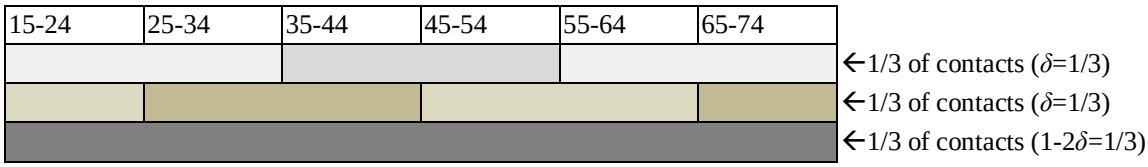
S4. Modelling the spread of infection amongst an age structured population

In the absence of therapy, susceptibility to infection is assumed to vary with age and is governed by the parameter, β_a, defined as the probability per partnership between an infected individual and a susceptible individual aged *a*. This parameter is scaled by the parameter $\varpi_{a,t}$ defined as the therapy induced relative susceptibility to HIV-2. Therapy in HIV-1 singly infected individuals is assumed to have a preventive effect on infection with HIV-2. Therapy effectiveness in preventing infection is defined by the proportion π. $\varpi_{a,t}$ is formulated to account for this effect in combination with the fact that partnerships are more likely to be made with high risk individuals who in turn are more likely to be HIV-1 positive and therefore, are also more likely to be receiving therapy. As an approximate way to account for this, the partner exchange rate of singly HIV-1 infected individuals is assumed to be *d* times as large as the partner exchange rate of HIV (1 and 2) uninfected individuals.

$$\varpi_{a,t} = \frac{r_{a,t}^{0,0} + d(r_{a,t}^{1,0} + r_{a,t}^{1,1}(1-\pi))}{r_{a,t}^{0,0} + d(r_{a,t}^{1,0} + r_{a,t}^{1,1})} \qquad \{13\}$$

Sexual partnerships between different age groups are formulated as the sum of convex combinations of contacts within and among groups [24-26]. The diagram below demonstrates the mixing pattern. For example the 35-44 age category has one third of contacts distributed amongst the same and the next highest age bracket (age 45-54; see row 1), one third of contacts distributed amongst the same and the next lowest age bracket (age 25-34; see row 2), and one third distributed amongst the whole population (see row 3). This means that, on average, the age difference between sexual partners is less than 20 years for 81% of partnerships. There is flexibility in the model to allow the average rate of sexual partner exchange, $c_{a,t}$, to vary with age and time. In practise we formulate this parameter as the product of two parameters that are dependent only upon age and time, respectively ($c_{a,t} = \bar{c}_a \tilde{c}_t$). Thus, relative rates of partner exchange across ages remains fixed over time, but the overall rate of partner exchange can vary with time. Individuals aged 75-84 years are assumed to have no sexual partnerships ($\bar{c}_a = 0$) and are therefore excluded from the contact pattern box figure shown below. Therapy is assumed to reduce the infectiousness of treated individuals. Treatment effectiveness in reducing infectiousness is also defined by π. We assume that the partner exchange rate of individuals with single HIV-2 infection and dual HIV-1/2 infection are equal.

Contact patterns within and between different age categories.



S5. Fitting the model to the incidence and prevalence data

A maximum of only three parameters were estimated by fitting the difference equation model (equations 1-4) to the age-stratified prevalence (Figure 1B) and incidence (Figure 1C) data. Two of these define how susceptibility to infection varies with age. The transmission probability per partnership between and infected individual and a susceptible individual aged a (β_a) is assumed to change linearly with age but this metric is free to vary for the youngest (aged 15) and oldest individuals who are assumed to be sexually active (aged 74) in the population. The remaining parameter ($\tilde{c}_t$ for $t \geq 1998$) defines the relative rate of partner exchange from 1998 onwards, as compared to during the period 1990 to 1997 (where $\tilde{c}_t = 1$). To estimate these three parameters, a likelihood function was devised to allow us to fit the model both to the prevalence data from 1997 and 2007 and amongst each age category (15-24, 25-34, 35-44, 45-54, ≥55) and to the incidence data across both observational time periods (1990-1997 and 1997-2007) and amongst each age category. Note that the initial conditions of the model were fixed to equal those observed in the 1990 data, ensuring that the model fits that data perfectly at that time. The likelihood function is formulated as the product of binomial errors surrounding the fraction of hosts of each age category with HIV-2 in 1990 and 2007 and the fraction of hosts of each age category with new infections during each of the two periods (1990-1997 and 1997-2007). Thus, for simplicity the incidence and prevalence data are assumed to be independent. The likelihood function is shown below (equation 14). To estimate the three parameters, we ran the difference equation model across the full parameter space and identified the parameter values that gave the global maximum of the likelihood function. 95% confidence intervals were estimated. All parameter estimates are provided in Table S2. This analysis revealed limited support for age dependency in susceptibility to HIV-2. Furthermore model simulations were similar assuming that susceptibility to HIV-2 is independent of age or varies with age according to our best fit model. For these reasons, our primary simulations assumed that susceptibility to HIV-2 was independent of age. Thus the model was refitted with only the two remaining parameters fitted allowed to vary.

Likelihood function:

$$L = \prod_{x=1:5} \left(\prod_{t=1997,2007} \binom{n_{x,t}}{k_{x,t}} p_{x,t}^{\,k_{x,t}} (1-p_{x,t})^{n_{x,t}-k_{x,t}} \prod_{v=1,2} \binom{\tilde{n}_{x,v}}{\tilde{k}_{x,v}} \tilde{p}_{x,v}^{\,\tilde{k}_{x,v}} (1-\tilde{p}_{x,v})^{\tilde{n}_{x,v}-\tilde{k}_{x,v}} \right) \qquad \{14\}$$

Where the variables included in this equation are defined are follows:

t	Time (t=1997, 2007).
v	Time period (v=1:2 represents 1990-1997 and 1997-2007, respectively).
x	Age group (x=1:5, representing age groups 15-24, 25-34, 35-44, 45-54 and ≥55, respectively).
$n_{x,t}$	Total number of individuals in age group x sampled to estimate HIV-2 prevalence at time t.
$\tilde{n}_{x,v}$	Total number of individuals in age group x (at the start of the time period) sampled to estimate HIV-2 incidence during time period v.
$k_{x,t}$	The observed number of hosts in age group x infected with HIV-2 at time t.
$\tilde{k}_{x,v}$	The observed number of individuals in age group x (at the start of the time period) with new HIV-2 infections during time period v.
$p_{x,t}$	Model prediction of the HIV-2 prevalence in age group x at time t.
$\tilde{p}_{x,v}$	Model prediction of the fraction of individuals in age group x (at the start of the time period) with new HIV-2 infections during time period v.

HIV-2 prevalence was estimated from the data (Figure 1B) as the fraction of hosts infected ($k_{x,t}/n_{x,t}$). Count data (numbers infected, $k_{x,t}$, and numbers tested, $n_{x,t}$) used to estimate prevalences in 1990, 1997 and 2007 were based upon counts of all hosts sampled across the duration of that survey (1989-1991, 1996-1998 and 2006-2007). Within each survey each person contributed only once to these data. In the model, prevalence in age group x at time t ($p_{x,t}$) was estimated to be equal to the fraction of all hosts in the population who are infected at that time. Below (equation 15), this is demonstrated for the prevalence of HIV-2 amongst hosts aged 15-24 in 1997 ($p_{1,1997}$).

$$p_{1,1997} = \frac{\sum_{a=15:24} I_{a,1997}}{\sum_{a=15:24} \left(I_{a,1997} + S_{a,1997} \right)} \qquad \{15\}$$

Incidence rates are estimated from data (Figure 1C) by considering all persons present in two consecutive surveys. The rate of new infections per person per year during each time period v amongst individuals in age group x is estimated to be equal to the fraction of persons in whom new infections occurred, ($\tilde{k}_{x,v}/\tilde{n}_{x,v}$) multiplied by the duration of the time period (7 or 10 years). In the model, incidence during year t amongst individuals aged a was calculated to be the total number of new infections that occurred during that year ($\lambda_{a,t}(1-\tilde{\gamma}_{a,t})S_{a,t}$). The incidence rate during period v amongst individuals who are in age group x at the start of the time period and alive at the end of it is estimated using the model as follows. The numerator is the total number of new infections during time period v amongst individuals in age group x at the start of the time period. The denominator is the sum (over each year of the time period) of susceptible hosts who at the start of the time period are in age group x, and who did not die during that year. Model predictions of the fraction of individuals in age group x at the start of

time period v who acquire new HIV-2 infections during time period v ($\tilde{p}_{x,v}$) is estimated using the model to be the product of the incidence rate and the duration of the time period. Below, this is demonstrated for the fraction of individuals aged 15-24 years (at the start of the time period) with new HIV-2 infections between 1997 and 2007.

$$\tilde{p}_{1,1} = 10 \times \frac{\sum_{t=1997:2006} \sum_{b=15:24} \lambda_{b+t-1997,t}(1-\tilde{\gamma}_{b+t-1997,t})S_{b+t-1997,t}}{\sum_{t=1997:2006} \sum_{b=15:24} (1-\tilde{\gamma}_{b+t-1997,t})S_{b+t-1997,t}} \quad \{16\}$$

S6. Assumption of independence during model fitting

It is noteworthy that an implicit assumption of estimating the likelihood in this way is that each data point is independent. In reality, many data points will not be independent. For example, incidence will affect prevalence. Prevalence at an earlier time will also affect prevalence at a later time because many individuals will have contributed data to both samples given that the survey spanned the majority of the population. Fitting the model in this way enables us to include all data points, rather than throwing some of them out; furthermore, it should not affect our optimal estimates. However, it is expected that the assumption of independence would make confidence in our parameters estimates appear greater than it is. That is, confidence intervals would appear tighter than they are. To investigate whether the assumption of independence of incidence and prevalence data meaningfully affects our confidence intervals, we first fitted the model to only the incidence data and then separately fitted it to only the prevalence data. We calculated for each model fit the confidence intervals for 1) the partner exchange rate between 1997 and 2007 relative to between 1990 and 1997; and 2) the susceptibility to HIV-2 of individuals aged 74 years relative to individuals aged 15 years. We found the mean of the limits of the confidence intervals across these two model runs. As expected, for each ratio, the average confidence limit is larger than the confidence interval assuming independence. However the difference is sufficiently small that it does not affect what we learn from the confidence intervals. These alternative confidence intervals confirm 1) that the risk of infection was lower in the period 1997-2007 compared to 1990-1997 (CI for ratio: 0.21-71, compared to 0.23-0.53 assuming independence). They also confirm 2) that we cannot rule out the possibility that susceptibility to infection marginally decreases with age or that it increases with age more significantly than our best estimate (CI for ratio: 0.71-3.9, compared to 0.76-2.4 assuming independence).

S7. The stochastic form of the model

To formulate the stochastic version of the model we define the following additional terms:

$D_{a,t}$	The number of susceptible hosts aged a, who are removed during year t
$E_{a,t}$	The number of infected hosts aged a, who are removed during year t
$F_{a,t}$	The number of hosts aged a, who become infected during year t

$$D_{a,t} \sim \text{Binomial}\left(S_{a,t}, \gamma_{a,t}\right) \quad \{17\}$$

$$E_{a,t} \sim \text{Binomial}\left(I_{a,t}, \tilde{\gamma}_{a,t}\right) \quad \{18\}$$

$$F_{a,t} \sim \text{Binomial}\left(S_{a,t} - D_{a,t}, \lambda_{a,t}\right) \quad \{19\}$$

The following equations define the stochastic model:

$$S_{15,t+1} = B \quad \{20\}$$

$$I_{15,t+1} = 0 \quad \{21\}$$

$$S_{a+1,t+1} = S_{a,t} - F_{a,t} - D_{a,t} \quad \{22\}$$

$$I_{a+1,t+1} = I_{a,t} + F_{a,t} - E_{a,t} \quad \{23\}$$

S8. Estimating the effective reproductive number

Because of the complexity of our model, including the fact that it is age stratified, an analytic expression for the basic reproductive number for our model was not evaluated. Instead we estimated the effective basic reproductive number directly from the survey. This metric is equivalent to the basic reproductive number – defined as the average number of secondary infections caused by one primary infection in a wholly susceptible population – but without the condition that the population is wholly susceptible. Nevertheless, because the prevalence of HIV-2 in the population is relatively low, the effects of saturation should be small and the basic and effective reproductive numbers will be similar.

To estimate the effective reproductive number (*R*) between 1990 and 1997, we first estimated from the data (Figure 1B-C) [4] the total number of new infections in Caió generated between 1990 and 1997 (new infections in sample/fraction of uninfected population in sample). We then estimated from data the total number of infected (and therefore infectious) person years in Caió between 1990 and 1997. This was done by finding the mean number of infected people in the population (infected people in sample/fraction of population in sample) between 1990 and 1997 and scaling to account for 7 years. The number of infected person years was then multiplied by the average mortality rate of HIV-2 positive individuals (0.063 per person years of observation) [9] to estimate the average number of infected lifespans between 1990 and 1997. The effective reproductive number was then calculated by dividing the number of new infections between 1990 and 1997 by the number of infected lifespans between 1990 and 1997. Equivalent calculations were made to estimate the effective reproductive number between 1997 and 20077 calculations were made

$$\text{incidence}_{1990-1997} = 51 \times 25417 / 10653 = 122 \quad \{24\}$$

$$\text{incidence}_{1997-2007} = 21 \times 37620 / 11934 = 66 \quad \{25\}$$

$$\text{mean number infected}_{1990-1997} = 0.5 \times \left(230/0.734 + 247/0.754\right) = 320 \quad \{26\}$$

$$\text{mean number infected}_{1997-2007} = 0.5 \times \left(247/0.754 + 135/0.741\right) = 255 \quad \{27\}$$

$$\text{infected lifetimes}_{1990-1997} = 320 \times 7 \times 0.063 = 141 \quad \{28\}$$

$$\text{infected lifetimes}_{1997-2007} = 255 \times 10 \times 0.063 = 160 \quad \{29\}$$

$$R = \frac{\text{incidence}}{\text{infected lifetimes}} \quad \{30\}$$

$$R_{1990-1997} = \frac{122}{141} = 0.86 \quad \{31\}$$

$$R_{1997-2007} = \frac{66}{160} = 0.41 \quad \{32\}$$

These findings indicate that the effective reproductive number was less than 1 in both periods and that it more than halved in the second period (1997-2007) compared to the first period (1990-1997)