Impact of antiretroviral therapy on HIV-1 transmission dynamics

Bezemer, D.O.

Citation for published version (APA):
27 years of the HIV epidemic amongst men having sex with men in the Netherlands: an in depth mathematical model-based analysis

Daniela Bezemer¹, Frank de Wolf², Maarten C. Boerlijst³, Ard van Sighem¹, T. Deirdre Hollingsworth⁴, and Christophe Fraser⁴

Abstract

Background: There has been increasing concern about a resurgent epidemic of HIV-1 amongst men having sex with men in the Netherlands, which has parallels with similar epidemics now occurring in many other countries.

Methods: A transmission model applicable to HIV-1 epidemics, including the use of antiretroviral therapy, is presented in a set of ordinary differential equations. The model is fitted by maximum likelihood to national HIV-1 and AIDS diagnosis data from 1980-2006, estimating parameters on average changes in unsafe sex and time to diagnosis. Robustness is studied with a detailed univariate sensitivity analysis, and a range of hypothetical scenarios are explored for the past and next decade.

Results: With a reproduction number around the epidemic threshold one, the HIV-1 epidemic amongst men having sex with men in the Netherlands is still not under control. Scenario analysis showed that in the absence of limiting infectiousness in treated patients, the epidemic could have been more than double its current size. Ninety percent of new HIV transmissions is estimated to take place before diagnosis. Decreasing time from infection to diagnosis, which was 2.5 years on average in 2006, can prevent many future infections.

Conclusions: Sexual risk-behaviour of men having sex with men who are not aware of their infection is the most likely factor driving this epidemic.

¹HIV Monitoring Foundation, Amsterdam, the Netherlands; ²Department of Infectious Disease Epidemiology, and ³Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, the Netherlands; ⁴Medical Research Council Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College London, UK; ⁵Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, the Netherlands;
Background

Despite the success of combination antiretroviral treatment (cART), the HIV-1 epidemic has been increasing amongst men having sex with men (MSM) in many industrialised countries, including the Netherlands [1-3]. Increases in syphilis and gonorrhoea diagnoses have also been documented in populations of MSM in several developed countries [4-6]. Surveys of risk behaviour have noticed increases in the number of casual partners and decreases in reliable condom use [5-10]. It is not clear what the best interventions are to get the HIV-1 epidemic under control.

Earlier mathematical modelling studies showed that an increase in risk behaviour has the potential to counterbalance the beneficial effect of cART [11-19]. Models have varied substantially with respect to whether transmission occurred from treated, diagnosed or undiagnosed individuals. With a recently developed mathematical model, which explicitly takes into account that the majority of people on cART are successfully treated and maintain very low HIV RNA levels, and are thus presumed largely uninfected, we found evidence that increased risk behaviour amongst undiagnosed individuals may have counterbalanced the beneficial effect of cART among MSM in the Netherlands [1]. Other models looking at the impact of cART treatment have found that treatment of the HIV infected population in an advanced stage of disease progression alone might not halt epidemic spread but that expanded and earlier access to cART can reduce the growth of the epidemic [19-21]. Besides pre-exposure prophylaxis of high risk HIV- negative MSM [22], and testing for acute primary infection, more routine health care interventions such as earlier treatment and more frequent testing are still not fully exploited [23]. Here, we apply our mathematical model to study the impact of cART use and – efficacy, risk-behaviour, and time from infection to diagnosis in a well documented epidemic among MSM.

First, we compare predictions based on our earlier analysis carried out with data till 2003 with an updated dataset covering the period 2004-2006, and we subsequently update our analysis. We present a detailed univariate sensitivity analysis which highlights some of the dependencies under the assumptions and structure of our model. We present data on CD4 count at diagnosis as a verification for the estimated changing patterns of time from infection to diagnosis. We complete with counterfactual hypothetical scenarios to explore the impact of changes in cART use, behaviour change and time to diagnosis. These scenario analyses were applied retrospectively to the past decade, and in addition to the coming decade.

Our development and application of a model framework for the interpretation of longitudinal surveillance data from the Netherlands may be adapted to similar situations in other settings (for other risk groups and other countries). In the appendix we present the mathematical details.
of the model, present derivations of key quantities such as the reproduction number [24] and the time to diagnosis, and of the maximum likelihood framework used to fit the model.

**Model and data description**

The model code was formulated in a set of ordinary differential equations implemented in the program Berkeley Madonna [25]. The model describes the rate of new HIV-1 infections as a function of the number of currently infected individuals, and individuals entering the country with infections acquired abroad. The framework captures the natural process of disease progression, as well as the effects of therapy. Key parameters relate to the relative changes in infectiousness: relative rate of risky sex; stage of infection; diagnosis, as diagnosed individuals might take less risks [26]; and cART. Parameters on cART use and failure were obtained from the ATHENA national observational cohort [27]. From this cohort we found that most people are long-term successfully treated on cART, with no detectable viral load. We separated this group from another smaller group that is only temporally successful before failing. This is different from most other models which use a constant rate of failure applicable to all patients, an assumption which only adequately describes the impact of ineffective monotherapy. We assume that people with no detectable viral load are not infectious, an assumption based on a study with discordant heterosexual couples [28], that is currently being investigated for discordant couples with access to cART. We fitted our model simultaneously to longitudinal data on annual HIV-1 diagnoses [29] and annual new AIDS cases [29-31] among MSM in the Netherlands in order to estimate the average changes over calendar time in risk-behaviour $\beta(t)$ and time to diagnosis needed to explain these data. From our parameter estimates the reproduction number $R(t)$ can be calculated. The analysis was stratified into five distinct historical intervals. Parameter values are summarized in Table 1. The protocols for handling the different sources of surveillance data, and for assigning cases with uncertain status, were described in an earlier paper [1]. Model details are in the appendix at the end of this paper.
### Table 1. Descriptions and values of Model Parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Denotation</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1/\alpha^*$</td>
<td>0.24</td>
<td>Duration (years) of Primary Infection</td>
<td>[28]</td>
</tr>
<tr>
<td>$1/\alpha$</td>
<td>1.89</td>
<td>Duration (years) of each disease stage $s$ ($s = 1,...,5$)</td>
<td>Maximum likelihood</td>
</tr>
<tr>
<td>$\rho^p$</td>
<td>2.77</td>
<td>Relative infectiousness during Primary Infection</td>
<td>[28]</td>
</tr>
<tr>
<td>$\rho^s, (s = 1,…,4)$</td>
<td>0.11</td>
<td>Relative infectiousness of asymptomatic disease stage $s$ ($s = 1,…,4$)</td>
<td>[28]</td>
</tr>
<tr>
<td>$\rho^5$</td>
<td>0.35</td>
<td>Relative infectiousness of disease stage 5 (AIDS)</td>
<td>[28]</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.5</td>
<td>Reduction in Relative infectiousness after diagnosis of each disease stages $s$</td>
<td>[26, 41]</td>
</tr>
<tr>
<td>$\delta^s (t), (s = 1,…,4)$</td>
<td>If time &lt; 1996 then 0 else estimated in four time intervals: (1984-1995); (1996-1999); (2000-2003); (2004-2006)</td>
<td>Rate per year at which people are diagnosed when in stage $s$ This version of the model does not account for diagnosis during primary stage. Estimated by fitting model to data</td>
<td></td>
</tr>
<tr>
<td>$\delta^5 (t)$</td>
<td>1/12</td>
<td>Rate per year at which people are diagnosed when in stage 5 (AIDS)</td>
<td>Assumption</td>
</tr>
</tbody>
</table>
| $A(t)$ | \[
\begin{align*}
A_0 & \text{ if } 1980 < t < 1984 \\
A_1 & \text{ if } t > 1984
\end{align*}
\] | Imported cases Estimated by fitting model to data |
| $f^p$ | 1 | Proportion of imported cases which are in Primary Infection | Assumption |
| $\gamma^s, (s = 1,…,4)$ | If time < 1996 then 0 else if time < 1998 then 0.7 \cdot (time-1996)/2 else if time >1998 then 0.7 | Rate per year of starting treatment and suppressing viral load at stages when cART naïve | [29] |
| $\gamma^5$ | If time < 1995 then 0 else if time < 1996 then 1 else then 2 | Rate per year of starting treatment and suppressing viral load at stage 5 (AIDS) when cART naïve | [29] |
| $\tau^s, (s = 1,…,5)$ | 0.5 | Fraction that fails treatment at stage $s$ | [29] |
| $\psi^s, (s = 1,…,5, f = 1,2)$ | 1.4 | Rate per year of starting subsequent treatment and suppressing viral load at stage $s$, failing round $f+1$ | [29] |
| $\kappa^s, (s = 1,…,5, f = 1,…,3)$ | 0.5 | Rate per year of failing treatment at stage $s$, treatment round $f$ | [29] |
| $\mu$ | 0 (short study period) | Basic death rate when no disease progression i.e. when endurably or temporally successfully treated | [42] |
Results

Model fit
We start by describing the best fit of our baseline model to the national surveillance data and what this reveals in terms of changing patterns of incidence, diagnosis and risk behaviour. Figures 1a and b, show the model fit to the HIV and AIDS diagnosis data. Figure 1c shows the best fit parameters, with a similar risk behaviour over the past six years, and a further decreasing time to diagnosis to 2.5 years on average in 2006. The risk behaviour rate $\beta(t)$ is standardised by setting it equal to 1.0 for untreated undiagnosed individuals in the asymptomatic stage of infection during the first phase of the epidemic (1980-1983), so that all other values are measured relative to this [1]. Figure 1d shows the resulting reproduction number is still around one, i.e. its epidemic threshold. The resulting prediction in the absolute numbers of new infections have been steadily increasing since 2000, as shown in figure 1e. A total of 620 MSM are estimated to be infected in 2006, which is close to the estimated 777 infections when the epidemic was at its peak around 1983. In figure 1f, the known and estimated unknown prevalence is shown. The percentage of the undiagnosed of the total infected population has decreased to 24%, but only so due to an increase in survival of the diagnosed population. In absolute numbers around 1600 MSM were undiagnosed at the end of 2006, estimated to be responsible for 90% of new HIV transmissions.

Update of analysis to include the period 2004-2007
Figure 2a shows that the number of new diagnoses in the period 2004–2006 are within the prediction interval of the model fit in our previous iteration which focused on data till 2003 [1]. Figure 2b shows the prediction interval of the model fit till 2006. The number of diagnoses in 2007 are within the prediction interval, but will likely be higher as there is a delay in data availability. This year was not included in the model fit as testing policy changed to an opting-out strategy at the STI clinic in Amsterdam.

Sensitivity Analysis
Because our baseline model was built on assumptions which underpinned the inferences drawn from the epidemic surveillance data, we explored the impact of varying assumptions in both the input parameters, interpretation of incomplete data, and model structure, refitting our model each time an assumption was varied. The results are shown in table 2. It shows the best model fit results and how these depend on parameters and assumptions in the model after refitting the model with implemented changes. The results on the key outcomes $\beta(t)$ and $R(t)$ appear to be very robust to a wide range of model variants. In particular, the model results were consistent till 2004 when assumptions about the relative infectiousness of disease stages, about the effect of diagnosis on behaviour, and about the time from diagnosis to start of therapy were varied. To examine whether the results could be due to delays in starting therapy over calendar time,
Figure 1. This whole figure is an update of previously published figures [1].

a. Black lines are best model fit to filled symbols of number of annual diagnosis separated for infected in the Netherlands in dots, and infected abroad in triangles, 2007 is incomplete data not included in fit. 

b. Black lines are best model fit to filled symbols of annual number of new AIDS cases in blue dots, and simultaneous HIV and AIDS diagnosis in red triangles, 2007 is incomplete data not included in fit. 

c. Fitting parameters with confidence intervals, on left axis in blue is the risk behaviour rate, in red on the right axis is the average time from infection to diagnosis. The risk behaviour rate $\beta(t)$ is standardised by setting it equal to 1.0 for untreated undiagnosed individuals in the asymptomatic stage of infection during the first phase of the epidemic (1980-1983), so that all other values are measured relative to this [1]. 

d. Resulting estimated Reproduction number, $R(t)$ over the whole study period, including the confidence interval. 

e. Estimated annual new infections over the whole study period, including the confidence interval. 

f. Blue dots is the living diagnosed MSM population, and the blue line the model estimation, The red line is the model estimation for the undiagnosed HIV infected MSM population, and the dotted red line is the percentage undiagnosed of the total prevalence.
Figure 2. **a.** Model fit to total annual diagnosis in black filled dots till 2003, including prediction interval, and data update in black circles. **b.** Model fit to total annual diagnosis till 2006, including prediction interval, and incomplete diagnosis in 2007.
we considered a model with a switch in policy after 2000 where individuals were no longer treated until the last stages of infection (model #29 in Table 2), but our conclusions remained unaffected.

The assumption which has largest impact on our predictions relates to the case where the risk behaviour rate $\beta(t)$ was assumed not to have increased in the cART era (model #19), but this model does not fit the data. If we assume (somewhat implausibly, and also resulting in a much worse model fit) that all infections with unknown source of infection are acquired outside of the Netherlands, then the epidemic would have been mainly driven by import and $R(t)$ within the Netherlands would be smaller than one (model #9). Estimates of $R(t)$ are higher (and model fit better) if all ‘unknown’ cases are assumed to be the result of transmission within the Netherlands (model #8). None of the other variations studied in the sensitivity analysis considerably altered the model predictions after a refit to the data. In all model variants, $R(t)$ for 2000-2004 is estimated to be near or above the critical threshold ($R(t)=1$), thus implying uncontrolled epidemic spread, with estimates of the reproduction number ranging between 0.95 and 1.33, depending on the scenario under consideration.

One fit is done using data from the literature on the transmission probability for receptive anal intercourse instead of relative infectiousness, resulting in $\beta(t)$ giving an indication of the average number of unsafe sex-acts per year between infectious- and negative MSM (model #20).

The data update on the period 2005–2006 is shown in model #42 to #59. It shows that adjustments to the data due to a delay in monitoring did not change our previous results. Whether $R(t)$ is just below or above the threshold one in the last four years of study depends largely on how effective cART really was and on how the time periods were defined. However, the confidence interval and all sensitivity analysis show that under any set of assumptions, it is still very close to one. Under all assumptions risk-behaviour has stayed the same over the past decade or even increased a little. The average time from infection to diagnosis has decreased under all assumptions, to 2.46 in the best fit.
Table 2. Uncertainty and sensitivity analysis. Main results after refitting the model under different parameter values then in table 1 or to alternative datasets.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Fit</td>
<td>2.39; 0.89; 0.76; 1.04</td>
<td>1.30; 0.56; 0.66; 0.93</td>
<td>7.88; 3.71; 3.16; 2.90</td>
<td>58.98</td>
<td>24.0</td>
</tr>
<tr>
<td>Lower 95% confidence limit**</td>
<td>2.17; 0.85; 0.70; 0.98</td>
<td>1.18; 0.53; 0.60; 0.87</td>
<td>***; 3.49; 3.00; 2.84</td>
<td>60.90</td>
<td>****</td>
</tr>
<tr>
<td>Upper 95% confidence limit**</td>
<td>2.76; 0.93; 0.86; 1.09</td>
<td>1.50; 0.58; 0.71; 0.97</td>
<td>***; 3.97; 3.41; 3.03</td>
<td>60.90</td>
<td>****</td>
</tr>
<tr>
<td>1 Infectiousness primary infection 50% smaller</td>
<td>3.37; 0.91; 0.72; 1.10</td>
<td>2.23; 0.71; 0.88; 1.38</td>
<td>7.88; 3.72; 3.16; 2.89</td>
<td>57.41</td>
<td>24.0</td>
</tr>
<tr>
<td>2 Infectiousness primary infection 75% smaller</td>
<td>4.70; 0.91; 0.69; 1.15</td>
<td>3.50; 0.83; 1.06; 1.84</td>
<td>7.88; 3.72; 3.16; 2.89</td>
<td>55.04</td>
<td>23.9</td>
</tr>
<tr>
<td>3 No transmission from primary phase</td>
<td>10.15; 0.92; 0.65; 1.26</td>
<td>8.63; 0.97; 1.33; 2.73</td>
<td>7.88; 3.75; 3.17; 2.89</td>
<td>49.00</td>
<td>23.9</td>
</tr>
<tr>
<td>4 No reduction in risk behaviour after diagnosis</td>
<td>2.93; 0.94; 0.71; 1.09</td>
<td>1.36; 0.43; 0.54; 0.84</td>
<td>7.88; 3.67; 3.17; 2.89</td>
<td>66.74</td>
<td>23.9</td>
</tr>
<tr>
<td>5 Ceasing risk behaviour after diagnosis</td>
<td>1.77; 0.81; 0.84; 0.98</td>
<td>1.16; 0.75; 0.85; 1.02</td>
<td>7.88; 3.90; 3.10; 2.92</td>
<td>52.02</td>
<td>24.1</td>
</tr>
<tr>
<td>6 No reduction in risk behaviour after diagnosis in cART era</td>
<td>2.39; 0.89; 0.72; 1.09</td>
<td>1.30; 0.56; 0.55; 0.84</td>
<td>7.88; 3.71; 3.17; 2.89</td>
<td>58.84</td>
<td>23.9</td>
</tr>
<tr>
<td>7 All diagnosed HIV infections as local</td>
<td>3.49; 1.0; 0.91; 1.17</td>
<td>1.89; 0.62; 0.79; 1.04</td>
<td>7.88; 3.92; 3.16; 2.88</td>
<td>49.04</td>
<td>25.2</td>
</tr>
<tr>
<td>8 Unknown country infection with local</td>
<td>2.62; 0.95; 0.80; 1.09</td>
<td>1.42; 0.59; 0.69; 0.97</td>
<td>7.88; 3.77; 3.19; 2.86</td>
<td>53.46</td>
<td>24.2</td>
</tr>
<tr>
<td>9 Unknown country infection with import</td>
<td>1.15; 0.63; 0.57; 0.82</td>
<td>0.62; 0.39; 0.50; 0.72</td>
<td>7.88; 3.57; 3.07; 3.01</td>
<td>103.48</td>
<td>24.4</td>
</tr>
<tr>
<td>10 10% more AIDS cases pre-cART</td>
<td>2.41; 0.87; 0.74; 1.05</td>
<td>1.30; 0.54; 0.61; 0.93</td>
<td>7.88; 3.83; 3.19; 2.89</td>
<td>76.26</td>
<td>24.0</td>
</tr>
<tr>
<td>11 10% more AIDS cases at diagnosis post-cART</td>
<td>2.39; 0.90; 0.75; 1.05</td>
<td>1.30; 0.56; 0.65; 0.92</td>
<td>7.88; 3.80; 3.29; 3.03</td>
<td>58.58</td>
<td>25.1</td>
</tr>
<tr>
<td>12 Added diagnosis data on people with unknown transmission route to locally acquired cases</td>
<td>2.41; 0.93; 0.82; 1.10</td>
<td>1.21; 0.58; 0.70; 0.94</td>
<td>7.88; 3.74; 3.27; 3.22</td>
<td>61.68</td>
<td>28.2</td>
</tr>
</tbody>
</table>
### Impact of Antiretroviral Therapy on HIV-1 Transmission Dynamics

<p>|   | Added diagnosis data on people with unknown transmission route to imported cases | All import divided among 5 disease stages, none in primary stage | Time to diagnosis at AIDS stage: 1 week | Time to diagnosis at AIDS stage same as time to diagnosis in earlier stages | Fixed time to diagnosis | Diagnosis rate unchanged after implementation of cART | Behaviour unchanged after implementation of cART | Transmission probability per sex act (instead of relative infectivity): primary infection and AIDS 0.183; chronic 0.014 [43] | Full cART implementation in 2000 (instead of 1998) | 0% of AIDS cases treated in 1995 | Full cART implementation in 1996, and 0% of AIDS cases treated in 1995 | 75% failing treatment each round | 75% failing treatment each round before 2000 and 25% thereafter | 10 rounds of treatment | No cART in first disease stage |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 13 | 2.45; 0.81; 0.69; 0.95 | 3.00; 0.96; 0.90; 1.04 | 2.4; 0.90; 0.76; 1.04 | 2.08; 0.74; 0.65; 1.15 | 2.40; 0.89; 0.67; 0.99 | 2.34; 0.91; 0.80; 1.07 | 2.18; 0.99; 0.75; 0.70 | 4.27; 1.05; 0.61; 1.33 | 2.39; 0.89; 0.74; 1.04 | 2.39; 0.89; 0.74; 1.05 | 2.39; 0.89; 0.75; 1.05 | 2.39; 0.89; 0.83; 1.13 | 2.39; 0.89; 0.73; 1.01 | 2.39; 0.89; 0.77; 1.05 |
|   | 1.33; 0.50; 0.59; 0.83 | 1.62; 0.61; 0.82; 0.97 | 1.32; 0.56; 0.66; 0.93 | 0.96; 0.39; 0.39; 0.72 | 1.31; 0.55; 0.59; 0.96 | 1.27; 0.58; 0.68; 0.91 | 1.19; 0.62; 0.62; 0.62 | 12.95; 3.56; 4.51; 10.06 | 1.30; 0.56; 0.64; 0.92 | 1.30; 0.56; 0.64; 0.93 | 1.30; 0.56; 0.65; 0.93 | 1.30; 0.56; 0.64; 0.88 | 1.30; 0.56; 0.66; 0.93 | 1.30; 0.56; 0.65; 0.90 |
|   | 7.88; 3.79; 3.34; 3.13 | 7.88; 3.26; 2.64; 2.43 | 7.82; 3.66; 3.11; 2.85 | 9.69; 7.2; 6.4; 6.07 | 7.88; 4.00; 3.00; 2.00 | 7.88; 3.55; 3.69; 2.97 | 7.88; 3.55; 3.69; 2.97 | 7.88; 3.68; 3.13; 2.99 | 7.88; 3.72; 3.15; 2.90 | 7.88; 3.70; 3.18; 2.89 | 7.88; 3.71; 3.19; 2.89 | 7.88; 3.71; 3.15; 2.90 | 7.88; 3.71; 3.16; 2.90 | 7.88; 3.71; 3.16; 2.90 |
|   | 78.62 | 130.52 | 61.29 | 203.53 | 102.91 | 76.83 | 156.78 | 101.50 | 58.44 | 59.28 | 62.93 | 62.93 | 58.93 | 58.88 | 58.93 | 58.88 |
|   | 24.9 | 17.9 | 23.6 | 52.6 | 16.5 | 27.36 | 13.9 | 28.2 | 24.0 | 24.5 | 24.4 | 24.4 | 25.6 | 24.2 | 24.0 | 24.1 |
|   | 90.1 | 87.1 | 90.2 | 96.0 | 88.1 | 91.1 | 82.4 | 78.2 | 90.2 | 90.4 | 90.4 | 90.4 | 85.6 | 90.11 | 91.2 | 88.2 |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>Only cART in last two stages within 0.5 year</td>
<td>2.41; 0.91; 0.76; 1.08</td>
<td>1.31; 0.56; 0.59; 0.86</td>
<td>7.88; 3.87; 3.23; 2.91</td>
<td>102.95</td>
</tr>
<tr>
<td>29</td>
<td>After 2000 only cART in last two stages within 0.5 year</td>
<td>2.39; 0.89; 0.76; 1.14</td>
<td>1.30; 0.56; 0.66; 0.90</td>
<td>7.88; 3.71; 3.16; 2.90</td>
<td>59.47</td>
</tr>
<tr>
<td>30</td>
<td>Rate of starting subsequent treatment = 0.5</td>
<td>2.39; 0.89; 0.76; 1.04</td>
<td>1.30; 0.56; 0.64; 0.89</td>
<td>7.88; 3.72; 3.15; 2.90</td>
<td>58.72</td>
</tr>
<tr>
<td>31</td>
<td>Rate of failing treatment = 2 and Rate of starting subsequent treatment = 0.5</td>
<td>2.39; 0.89; 0.73; 1.02</td>
<td>1.30; 0.56; 0.62; 0.88</td>
<td>7.88; 3.72; 3.15; 2.90</td>
<td>58.41</td>
</tr>
<tr>
<td>32</td>
<td>Rate of failing treatment = 2 and Rate of starting subsequent treatment = 0.5 and 75% failing treatment each round and rate between stages in cART era = 1/3</td>
<td>2.43; 0.93; 0.90; 1.22</td>
<td>1.32; 0.57; 0.55; 0.74</td>
<td>7.88; 4.26; 4.56; 4.51</td>
<td>82.84</td>
</tr>
<tr>
<td>33</td>
<td>Rate of starting first treatment = 0.5</td>
<td>2.39; 0.89; 0.76; 1.04</td>
<td>1.30; 0.56; 0.64; 0.91</td>
<td>7.88; 3.71; 3.15; 2.90</td>
<td>58.52</td>
</tr>
<tr>
<td>34</td>
<td>5 time intervals: 1980-1984-1990-1995-2000-2003</td>
<td>2.26; 0.98; 0.84; 0.76; 1.06</td>
<td>1.23; 0.61; 0.52; 0.66; 0.95</td>
<td>7.88; 3.75; 3.86; 3.12; 2.81</td>
<td>56.43</td>
</tr>
<tr>
<td>35</td>
<td>Time intervals: 1980-1984-1996-1999-2004</td>
<td>2.51; 0.89; 0.68; 1.02</td>
<td>1.36; 0.56; 0.60; 0.90</td>
<td>7.9; 3.8; 3.1; 3.0</td>
<td>58.40</td>
</tr>
<tr>
<td>36</td>
<td>Time intervals: 1980-1984-1996-2001-2004</td>
<td>2.39; 0.89; 0.80; 1.03</td>
<td>1.30; 0.56; 0.69; 0.94</td>
<td>7.88; 3.70; 3.19; 2.67</td>
<td>60.40</td>
</tr>
<tr>
<td>37</td>
<td>Rate between stages in cART era = 1/3</td>
<td>2.42; 0.93; 0.83; 1.13</td>
<td>1.32; 0.57; 0.61; 0.84</td>
<td>7.88; 4.25; 4.59; 4.48</td>
<td>85.14</td>
</tr>
<tr>
<td>38</td>
<td>Rate between stages in all eras = 1/2</td>
<td>2.34; 0.89; 0.74; 1.06</td>
<td>1.22; 0.53; 0.63; 0.92</td>
<td>8.32; 4.13; 3.36; 3.02</td>
<td>58.02</td>
</tr>
<tr>
<td>39</td>
<td>Rate of failing treatment = 2</td>
<td>2.39; 0.89; 0.73; 1.03</td>
<td>1.30; 0.56; 0.64; 0.91</td>
<td>7.88; 3.71; 3.16; 2.90</td>
<td>58.72</td>
</tr>
<tr>
<td>40</td>
<td>Death rate = 0.03</td>
<td>2.39; 0.89; 0.74; 1.04</td>
<td>1.30; 0.56; 0.66; 0.93</td>
<td>7.88; 3.71; 3.16; 2.90</td>
<td>58.99</td>
</tr>
<tr>
<td>41</td>
<td>4th interval: 2000-2004</td>
<td>2.38; 0.90; 0.74; 1.05</td>
<td>1.29; 0.56; 0.64; 0.94</td>
<td>7.88; 3.72; 3.14; 2.80</td>
<td>61.46</td>
</tr>
<tr>
<td>Data update</td>
<td>period</td>
<td>$R(t)$ respectively in:</td>
<td>$\beta(t)$ respectively in:</td>
<td>time to diagnosis respectively in:</td>
<td>Dev/2</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>42</td>
<td>4th period: 2000-2003; 5th period 2004-2006</td>
<td>2.37; 0.89; 0.75; 1.01; 1.00</td>
<td>1.28; 0.56; 0.65; 0.90; 0.93</td>
<td>7.88; 3.64; 3.14; 2.83; 2.47</td>
<td>67.92</td>
</tr>
<tr>
<td></td>
<td>Lower 95% confidence limit**</td>
<td>2.10; 0.86; 0.70; 0.97; 0.91</td>
<td>1.14; 0.53; 0.60; 0.87; 0.84</td>
<td>***; 3.41; 2.96; 2.63; 2.21</td>
<td>69.84</td>
</tr>
<tr>
<td></td>
<td>Upper 95% confidence limit**</td>
<td>2.56; 0.92; 0.81; 1.06; 1.05</td>
<td>1.39; 0.58; 0.71; 0.96; 0.97</td>
<td>***; 3.88; 3.25; 3.03; 2.73</td>
<td>69.84</td>
</tr>
<tr>
<td>43</td>
<td>1st period: 1980 – 1984</td>
<td>1.92; 0.86; 0.77; 1.02; 0.99</td>
<td>1.04; 0.53; 0.67; 0.91; 0.92</td>
<td>7.88; 3.79; 3.12; 2.82; 2.43</td>
<td>45.74</td>
</tr>
<tr>
<td>44</td>
<td>4th period: 2000-2002; 5th period 2003-2006</td>
<td>2.33; 0.89; 0.75; 1.01; 1.03</td>
<td>1.26; 0.56; 0.65; 0.92; 0.94</td>
<td>7.88; 3.62; 3.13; 2.77; 2.65</td>
<td>71.70</td>
</tr>
<tr>
<td>45</td>
<td>4th period: 2000-2006</td>
<td>2.33; 0.89; 0.74; 1.03</td>
<td>1.26; 0.56; 0.64; 0.93</td>
<td>7.88; 3.63; 3.10; 2.72</td>
<td>72.15</td>
</tr>
<tr>
<td>46</td>
<td>4th period: 2000-2005</td>
<td>2.32; 0.90; 0.74; 1.05</td>
<td>1.26; 0.56; 0.64; 0.94</td>
<td>7.88; 3.63; 3.11; 2.81</td>
<td>65.45</td>
</tr>
<tr>
<td>47</td>
<td>4th period: 2000-2004</td>
<td>2.32; 0.90; 0.74; 1.05</td>
<td>1.26; 0.56; 0.64; 0.94</td>
<td>7.88; 3.63; 3.11; 2.81</td>
<td>65.45</td>
</tr>
<tr>
<td>48</td>
<td>4th period: 2000-2003</td>
<td>2.32; 0.90; 0.74; 1.07</td>
<td>1.26; 0.56; 0.65; 0.96</td>
<td>7.88; 3.63; 3.10; 2.83</td>
<td>62.06</td>
</tr>
<tr>
<td>49</td>
<td>Only 25% reduction in risk behaviour after diagnosis in 5th period 2004-2006</td>
<td>2.39; 0.89; 0.76; 1.01; 1.03</td>
<td>1.30; 0.56; 0.66; 0.90; 0.89</td>
<td>7.88; 3.64; 3.13; 2.83; 2.46</td>
<td>67.66</td>
</tr>
<tr>
<td>50</td>
<td>4th period: 2000-2003; 5th period 2004-2007</td>
<td>2.32; 0.89; 0.75; 1.01; 0.99</td>
<td>1.26; 0.56; 0.66; 0.91; 0.92</td>
<td>7.88; 3.62; 3.12; 2.82; 2.46</td>
<td>67.63</td>
</tr>
<tr>
<td>51</td>
<td>4th period: 2000 – 2007</td>
<td>2.27; 0.90; 0.73; 1.03</td>
<td>1.23; 0.56; 0.64; 0.93</td>
<td>7.88; 3.63; 3.10; 2.72</td>
<td>72.27</td>
</tr>
<tr>
<td>52</td>
<td>$\gamma = 1.5$ from 2004; $\tau = 0.35$ from 2004; 4th period: 2000-2003; 5th period 2000-2004,</td>
<td>2.30; 0.90; 0.77; 1.01; 0.93</td>
<td>1.25; 0.56; 0.67; 0.91; 0.94</td>
<td>7.88; 3.63; 3.14; 2.83; 2.47</td>
<td>69.70</td>
</tr>
<tr>
<td>53</td>
<td>$\gamma = 1.5$ from 2004; $\tau = 0.35$ from 2000; 4th period: 2000-2003; Only 25% reduction in risk behaviour after diagnosis in 5th period 2004-2006</td>
<td>2.33; 0.90; 0.74; 0.98; 0.95</td>
<td>1.26; 0.56; 0.67; 0.92; 0.92</td>
<td>7.88; 3.63; 3.13; 2.82; 2.46</td>
<td>70.46</td>
</tr>
<tr>
<td>54</td>
<td>$\gamma = 1.5$ from 2003; $\tau = 0.35$ from 2003; 4th period 2000-2002; 5th period 2003-2006</td>
<td>2.32; 0.90; 0.76; 1.01; 0.97</td>
<td>1.26; 0.56; 0.66; 0.92; 0.96</td>
<td>7.88; 3.63; 3.12; 2.76; 2.65</td>
<td>72.37</td>
</tr>
<tr>
<td>55</td>
<td>$\gamma = 1.5$ from 2004; $\tau = 0.35$ from 2000; 4th period: 2000-2003; 5th period 2004-2006,</td>
<td>2.32; 0.90; 0.78; 1.02; 0.94</td>
<td>1.26; 0.56; 0.68; 0.92; 0.95</td>
<td>7.88; 3.63; 3.14; 2.81; 2.46</td>
<td>70.06</td>
</tr>
<tr>
<td>56</td>
<td>$\gamma = 1.5$ from 2000; $\tau = 0.35$ from 2000; 4th period 2000-2003; 5th period 2004-2006,</td>
<td>2.36; 0.90; 0.77; 0.98; 0.95</td>
<td>1.28; 0.56; 0.68; 0.95; 0.96</td>
<td>7.88; 3.68; 3.16; 2.82; 2.47</td>
<td>82.27</td>
</tr>
<tr>
<td>57</td>
<td>$\gamma = 1.5$ from 1995; $\tau = 0.35$ from 2000; 4th period 2000-2006</td>
<td>2.33; 0.89; 0.71; 0.99</td>
<td>1.27; 0.56; 0.67; 0.97</td>
<td>7.88; 3.63; 3.12; 2.72</td>
<td>74.54</td>
</tr>
</tbody>
</table>
CD4 cell count at diagnosis

Though we did not fit our model to CD4 count data, a qualitative comparison of changes in CD4 count at diagnosis is informative as a corroboration of our inferences on the role of changing patterns of diagnosis. We compared the estimated proportion of newly diagnosed patients in each disease stage as estimated by our best fit model with data on CD4 cell count at diagnosis. Therefore, we describe the five stages in our model as the following CD4 cell count intervals in cells/mm³: stage 1 = > 600; stage 2 = 400-600; stage 3 = 250-400; stage 4 = 100-250; stage 5 = 0-100. Figure 3a shows these stages plotted for MSM in ATHENA as a function of the year of diagnosis. While our model disease stages can only approximately be identified with categories based on CD4 cell count (Figure 3b,c), three observations can be made. First, we have a tendency to overestimate the extent to which individuals are diagnosed in the early stage of infection. Second, apart from this mismatch, our model reproduces satisfactorily the observed temporal trends, and third, the recent increase in the proportion of newly diagnosed individuals with high CD4 cell counts corroborates our model’s inferences in interpreting recent increases in annual number of new HIV diagnosis as rising transmission and increased diagnosis rather than improved diagnosis of people infected many years in the past.
Figure 3. Using CD4 count at diagnosis as a surrogate of the time since infection at diagnosis. We defined five CD4 cell count intervals for every model stage of infection with average duration of 1.89 years, and show these plotted for MSM in ATHENA as a function of year of diagnosis in a, where the proportion of people diagnosed in each stage is shown ranging from earliest (lightest) to last (AIDS, darkest). These estimates are biased by the fact that that only people who survive until 1996 are included in our study. The subsequent figures (b and c) show the estimated proportion of newly diagnosed patients in each disease stage as estimated by our best fit model. b includes the same process of truncation which generates bias prior to 1996 as the data (and is thus to be compared to a), while c indicates what we estimate actually occurred in the absence of this bias.
Hypothetical scenario analysis

Based on our best model fit, we explored a number of hypothetical “what if” scenarios, based on the assumption that the relative importance of the other contributing factors remains constant. All of the numbers below should be interpreted as ballpark approximations — exact estimates are available from figure 4, but are provided in the text rounded to the nearest hundred to reduce false accuracy.

![Graph](image)

**Figure 4.** Hypothetical scenarios for past and future decade. 

- **a**, during past decade. Best fit model (thick, $R = 1.00$). Had there been no treatment with cART (long-dash, $R = 1.43$). Had there been no increase in diagnosis rate (short-dash, $R = 1.12$). Had there been no increase in risk behaviour rate (starred, $R = 0.60$). Had there been no treatment, no increases diagnosis and no increase in risk behaviour rate (thin, $R = 0.89$).
- **b**, during next decade. If nothing changes (thick, $R = 1.04$). If proportion failing each line of cART is reduced to ten percent (short dash, $R = 0.92$). If average time to diagnosis is reduced to one year (see methods) (long dash, $R = 0.86$). If the risk behaviour rate is back to pre- cART levels (thin, $R = 0.60$). All three interventions (starred, $R = 0.46$).
Firstly, we explored a number of scenarios from 1995 onwards (Figure 4A). In the absence of cART limiting infectiousness in treated patients, the epidemic under current conditions would have been much larger, with an estimated 11,300 infections arising between 1995 and 2006, instead of the estimated 4,900. If on the other hand, cART had been introduced but there had been no increases in the risk behaviour rate, the number of new infections over this period would be only 2,000. If cART had been introduced and there had been the increase in the risk behaviour rate, but no increases in the diagnosis rate, the cumulative number of new infections would be 5,800. Finally, if no changes had occurred since 1995, i.e. no cART and no increasing risk and testing behaviour, this number would be 3,900. Thus based on these model estimates, we conclude that cART has played an important role in limiting transmission, but that any gains made have been more than offset by increases in the risk behaviour rate. Had these increases not occurred in the cART era, the reproduction number \( R(t) \) would have declined to 0.6, and the epidemic would have entered in a convincing decline.

Furthermore, we explore a number of hypothetical scenarios for the coming decade (Figure 4B). If nothing changes, the epidemic will spread uncontrolled and the cumulative number of infections between 2007 and 2016 will reach 7,700. If the frequency of testing is increased such that the mean time from infection to diagnosis (and subsequent treatment) is reduced to one year, this number will be reduced to 4,406. If the quality of treatment on offer is improved, such that the fraction of patients failing each line of therapy is only ten percent, then the cumulative number of new infections will reach 5,600. If the risk behaviour rate is reduced back to pre-cART levels, this number can be reduced to 1,700. Finally if all three interventions could be successfully implemented it is further reduced to only 800 new infections. From this analysis, we conclude that reducing the risk behaviour rate has the greatest impact on the epidemic, though earlier diagnosis and treatment can also prevent almost half of the infections.

**Discussion and Conclusion**

The model we presented in this paper is shown to be a robust and valuable tool for studying HIV epidemics in times of cART. We simultaneously estimated both changes in average risk-behaviour and time to diagnosis. Using these estimates, we calculated the reproduction number, which summarizes the state of the epidemic. In all sensitivity analyses of the current situation among MSM in the Netherlands, the reproduction number \( R(t) \) is estimated to lie very close to its threshold value of one, indicating current epidemic spread driven by local transmission. Under all assumptions, the net transmission rate \( \beta(t) \) (the per-capita rate at which infectious individuals infect new people) has shown no improvement over the past decade. The time from infection to diagnosis however has steadily decreased to 2.5 years on average in 2006, indicating improvements in testing. The qualitative comparison with CD4 data underscored
that the increase in annual diagnosis is due to a recent increase in transmissions. Data on rectal
gonorrhea and recent syphilis diagnoses have the same temporal curve as the incidence curve
resulting from our model fit [4]. R(t) is lower now than when the epidemic was at its peak in the
beginning of the eighties, but as it is now taking place on a much larger scale almost as many
MSM are infected annually. Currently, around 90% of new infections are estimated to be trans-
mitted from the undiagnosed infected population, being 24% of the total infected population.
Thus, the epidemic is driven by MSM having risky sex not being aware they are infected, and
curiously the absolute magnitude of this group has remained similar over time.

That sexual behaviour has increased seems to some extent a normal reaction to the reduced
risk, and as shown in the hypothetical scenarios cART has substantially mitigated the effects
of this return to pre-AIDS era risk behaviour levels which otherwise would have caused twice
as many cases. Hypothetical scenarios further showed that, if nothing changes, twice as many
MSM will be in need of health care for being HIV-1 infected in the coming decade than at
present. The most effective way to prevent this is to decrease risk-behaviour. A rough estimate
of average number of unsafe sex acts needed for current HIV-1 spread is around ten per year.
So even though the per-sex act change of transmission might seem small, this clearly shows
that epidemic spread is possible, especially since one still cannot easily be diagnosed during the
primary stage when one is most infectious [28].

There is a lot of scope to improve the rate of diagnosis during asymptomatic infection [5, 32-34]
and the potential for diagnosing individuals during the acute stage of infection [35]. This model
analysis does not produce reliable estimates of what proportion of transmission is due to acute
(primary) infection, and what proportion occurs during asymptomatic set-point infection. This
proportion needs to be estimated by other means. As the partner change rate during the acute
stage of infection is unknown, we could not predict to what level the epidemic can be reduced
without testing during that window-period. From phylogenetic studies it seems that about 25%
of onward transmission occurs during that stage [36], implying that using classical serological
diagnosis to limit infections after the acute phase of infection could be very effective in control-
ling the epidemic. The role of contact tracing and virological diagnosis to diagnose people in
acute stage of infection should be explored.

In the Netherlands there has been a health care policy towards promoting safe sex rather then
HIV testing [33]. It would be interesting to do qualitative comparison with other countries on
CD4 counts at diagnosis to see if the different policies in the past indeed still lead to different
stages of infection at diagnosis.

A key assumption, which is biologically plausible but remains unvalidated in practice, is that
individuals who are successfully treated, and who maintain very low levels of virus in blood, do
not contribute to onwards infections at all. Individuals who fail therapy are assumed to be as infectious as untreated individuals. While it is unlikely that mathematical model-based analyses of the type used here can validate this assumption, alternative scenarios which consider the effects of relaxing this assumption in data-driven analyses should be explored in detail in future work [19, 37-39]. Previous modelling studies focussing on hypothetical scenarios have emphasised the potential that ongoing transmission during incompletely suppressive antiretroviral therapy could have in driving epidemics [11-19, 37, 38]. We cannot rule out such a contribution to the Dutch HIV epidemic, an issue that could be best addressed by empirical discordant pair studies.

In this paper we performed a univariate sensitivity analysis to explore the impact of different model assumptions. This has the benefit of being easy to understand, as the impact of every assumption can be checked for and understood separately. However a multivariate sensitivity analysis will be developed in the future to more rigorously test the dependence of the model predictions to different assumptions. The effect of the opting-out testing policy can also be evaluated in the near future. This will provide an interesting population-based test of the hypothesis that more active diagnosis and treatment programs can contribute to mitigating what is currently a re-emerging epidemic. It would also be interesting to use the model we presented in this paper to study other well monitored epidemics, and to extend this model for studying the transmission of resistance [40].

Acknowledgements

DB was supported by grant 7014 from AIDS Fund Netherlands and by a travel grant from NWO (Netherlands Organisation for Scientific Research). CF is funded by the Royal Society. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Reference List


Appendix on derivation of the model

Survival distribution
To model the natural history of infection in untreated individuals, an Erlang survival distribution was fitted to data from 130 MSM seroconverters before the cART era in the Amsterdam Cohort Studies [44]. The maximum likelihood estimate was for an Erlang distribution of degree 5 and rate 1/1.89 per year. This best fit survival distribution can be modelled using a compartmental model, with unidirectional flow through five compartments with mean stay in each of 1.89 years (Figure 5). These compartments can roughly be understood as stages of progression of disease, but the main motivation for their use is to replicate in detail the survival distribution of untreated HIV infection.

Figure 5. Best fit survival distribution to data from 130 MSM seroconverters before the cART era in the Amsterdam Cohort Studies (all data were truncated from 22 Nov 1993 when the first person received a protease inhibitor). Solid line, Kaplan-Meier distribution estimated from data and dashed line, best fit Erlang distribution.

Relative infectiousness of different stages of disease progression
The relative infectiousness of different stages was adapted from Hollingsworth et al.[28]. This was obtained by fitting a hazard model to the data from Wawer et al.[45], a study of HIV-1 transmission in serodiscordant heterosexual couples in Uganda. Infectiousness is 25-fold higher for primary infection stage, assumed to last 0.24 years, than during asymptomatic infection (model stages 1-4, mean duration 7.56 years), and 3.2-fold higher during the AIDS stage (model stage 5, mean duration 1.89 years). In the absence of better data, it was assumed here that these relative infectiousness values were intrinsic features of natural infection, and thus similar for
this population of predominantly Dutch MSM. Dependence of the modelling results on these values was explored in the sensitivity analysis.

The relative infectiousness values refer to infectiousness within HIV-1 serodiscordant partnerships. The relative contribution of different stages will depend somewhat on the frequency of partner change, per sex act transmission probabilities, and the details of the sexual network [46]. We find in the sensitivity analysis that our main conclusions are not overly dependent on the value of the relative infectiousness parameters.

**Net transmission rate**

The net transmission rate $\beta(t)$ is a time-varying function that determines the rate at which individuals infect other people. It is defined up to an arbitrary constant, and is thus a relative rate. It only makes sense to compare this rate in different time periods rather than to assign meaning to its absolute value. The model also includes modifiers to describe the effects of stage of infection, diagnosis and treatment on infectiousness, described later.

Our main conclusions are based on fitting this net transmission rate through stages of the epidemic, a method which we expect to be fairly independent of the details of the sexual partner network. Similarly, because the overall prevalence of infection remains very low throughout the observed period, we do not explicitly account for “saturation” (e.g. depletion of the susceptible “pool”). $\beta(t)$ is primarily intended as a measure of changes in risk behaviour between discordant couples, and one should note that $\beta(t)$ is a compound measure which is affected by changes in the partner change rate, by the rate and nature of risky sex acts within partnerships, by the effect of saturation of the susceptible population, by the effect of the changing prevalence of other STIs in modulating HIV transmission, by risk-management strategies such as “sero-sorting”, and even by possible secular changes in the infectiousness of the virus.
**Incidence of Infection**

Our model relates the incidence rate of infection to the prevalence of infections, stratified by disease stage and treatment state. The parameters on infectiousness are defined as relative rate modifiers for different stages of disease progression, of diagnosis and of treatment. The total rate at which people in primary infection infect others, $\text{IY}^p$, is given by

$$\text{IY}^p = \beta(t) p^p (Y^p_{\text{local}} + Y^p_{\text{import}})$$

where $p^p$ is the infectiousness of primary infection relative to other disease stages, $Y^p_{\text{import}}$ and $Y^p_{\text{local}}$ are the number of people currently in the primary infection stage, who acquired infection abroad (import cases), and in the Netherlands (local cases). After primary infection, people progress through a series of stages labeled $s = 1, \ldots, 5$. The total rate at which people before diagnosis infect others during these stages is

$$\text{IY}^s = \beta(t) p^s (Y^s_{\text{local}} + Y^s_{\text{import}})$$

where $p^s$ is the relative infectiousness of disease stage $s$, $Y^s_{\text{import}}$ and $Y^s_{\text{local}}$ are the number of people in disease stage $s$. After diagnosis, people reduce their risk behaviour by a factor $\sigma$ due to awareness of their infection status, so that the total rate at which diagnosed, untreated people infect others, $\text{IY}^s_{\text{diag}}$, is

$$\text{IY}^s_{\text{diag}} = \beta(t) \sigma p^s Y^s_{\text{diag}}$$

where $Y^s_{\text{diag}}$ is the number of diagnosed untreated people in stage $s$, ($s = 1, \ldots, 5$). People are assumed not to be infectious while they are successfully treated. For people who have failed treatment $f$ times, where $f = 1, \ldots, 3$, the total rate at which they infect others is

$$\text{IF}^s_f = \beta(t) \sigma p^s F^s_f$$

where $F^s_f$ is the number of people who have failed treatment $f$ times who are in disease stage $s$.

The total incidence rate of infection, $I(t)$, is the sum of these terms

$$I(t) = \text{IY}^p + \sum_{s=1}^{5} \text{IY}^s + \sum_{f=1}^{3} \text{IF}^s_f$$
Undiagnosed Primary Infections – Highly infectious

The rate of change in the number of locally acquired cases that are in primary infection is given by

$$\frac{dY^p_{\text{local}}}{dt} = I(t) - \alpha^p Y^p_{\text{local}}$$

where $I(t)$ is the total incidence rate of infection, $1/\alpha^p$ is the average duration of the primary infection stage, and $Y^p_{\text{local}}$ are the number of locally acquired cases that are in the primary infection stage. The rate of changes in the number of imported cases that are in primary infection is given by

$$\frac{dY^p_{\text{import}}}{dt} = A(t) f^p - \alpha^p Y^p_{\text{import}}$$

where $A(t)$ are the number of imported cases per year, $f^p$ is the fraction of imported cases that are in primary stage, and $Y^p_{\text{import}}$ are the number of imported cases in the primary infection stage.

Undiagnosed Cases Progressing Through Stages of Infection - Infectious

Changes in the number of locally acquired undiagnosed untreated people in stage $s$, ($s = 1,\ldots,5$) is given by

$$\frac{dY^s_{\text{local}}}{dt} = \alpha^s Y^s_{\text{local}} - \alpha Y^s_{\text{local}} - \delta^s(t) Y^s_{\text{local}}$$

where $\alpha$ describes the rate of progression to each subsequent stage, $\delta^s(t)$ describes the rate at which people are diagnosed when in stage $s$, ($s = 1,\ldots,5$) and is defined as a time-varying function, and $Y^s_{\text{local}}$ are the number of locally infected cases in stage $s$, ($s = 1,\ldots,5$). Changes in the number of imported undiagnosed untreated people in stage $s$, ($s = 1,\ldots,5$) is given by

$$\frac{dY^s_{\text{import}}}{dt} = A(t) \frac{1-f^p}{5} + \alpha^s Y^s_{\text{import}} - (\alpha + \delta^s(t)) Y^s_{\text{import}}$$

where $Y^s_{\text{import}}$ are the number of imported cases in stage $s$ ($s = 1,\ldots,5$).
Diagnosed Cases Progressing Through Stages of Infection – Infectious, but with reduced risk behaviour

Changes in the number of all diagnosed untreated people together in stage $s (s = 1, \ldots, 5)$ is given by

$$\frac{dY^s_{\text{diag}}}{dt} = \delta'(t) \left( Y^s_{\text{import}} + Y^s_{\text{local}} \right) - (\alpha + \gamma'(t))Y^s_{\text{diag}} \quad ; \quad (s = 1)$$

$$\frac{dY^s_{\text{diag}}}{dt} = \delta'(t) \left( Y^s_{\text{import}} + Y^s_{\text{local}} \right) + \alpha Y^{s-1}_{\text{diag}} - (\alpha + \gamma'(t))Y^s_{\text{diag}} \quad ; \quad (s = 2, \ldots, 5)$$

where $Y^s_{\text{diag}}$ are the number of diagnosed cases in stage $s, (s = 1, \ldots, 5)$, and $\gamma'(t)$ the rate per year of starting treatment and suppressing viral load at stage $s$ when therapy naive.

Temporarily Successful Treated Cases – Not Infectious

The rate of change in the number of temporarily successful treated cases for the three rounds of treatment ($f = 1, \ldots, 3$) is defined by

$$\frac{dQ^f_s}{dt} = \tau' \gamma'(t)Y^s_{\text{diag}} - (\mu + \kappa^f)Q^f_s \quad ; \quad (s = 1, \ldots, 5, f = 1)$$

$$\frac{dQ^f_s}{dt} = \tau' \gamma'(t)f^f_{s+1} - (\mu + \kappa^f)Q^f_s \quad ; \quad (s = 1, \ldots, 5, f = 2, 3)$$

where $Q^f_s$ are the number of cases in stage $s, (s = 1, \ldots, 5)$ on treatment round $f, (f = 1, \ldots, 3)$, $\tau'$ is the fraction of cases starting therapy at stage $s, (s = 1, \ldots, 5)$ that will have only a temporarily successful viral load suppression when starting treatment, $\tau'_{f+1}$ is the rate of starting treatment and obtain viral load suppression at stage of infection $s, (s = 1, \ldots, 5)$ after treatment failure $f, (f = 1, 2)$, $\kappa^f$ is the rate of failing treatment at stage $s, (s = 1, \ldots, 5)$ after treatment failure $f, (f = 1, \ldots, 3)$, and $\mu$ is the basic death rate.
Failing Cases – Infectious
The rate of change in the number of people failing treatment \( f \) times \((f = 1,...,3)\) is defined as

\[
\frac{dF_s^f}{dt} = \kappa^s Q^s_f - (\alpha^s + \tau^s) F^s_f ; \quad (s = 1; f = 1,2)
\]

\[
\frac{dF^s_f}{dt} = \kappa^s Q^s_f + \alpha F_{s-1}^s - (\alpha^s + \tau^s) F^s_f ; \quad (s = 2, ..., 5; f = 1,2)
\]

\[
\frac{dF_s^f}{dt} = \kappa^s Q^s_f - \alpha F^s_f ; \quad (s = 1; f = 3)
\]

\[
\frac{dF^s_f}{dt} = \kappa^s Q^s_f + \alpha F_{s-1}^s - \alpha F^s_f ; \quad (s = 2, ..., 5; f = 3)
\]

where \( F^s_f \) are the number of cases in stage \( s \), \((s = 1,...,5)\) failing treatment round \( f \), \((f = 1,...,3)\), people failing the third round of therapy \( (f = 3) \) have run out of treatment options and continue disease progression.

Enduringly Successfully Treated Cases – Not Infectious
The rate of change in the number of enduringly successfully treated people is defined by

\[
\frac{dT}{dt} = \sum_{s=1}^{5} \left[ (1 - \tau^s) \gamma^s(t) Y^s_{diag} + (1 - \tau^s) \gamma_1^s F^s_1 + (1 - \tau^s) \gamma_2^s F^s_2 \right] - \mu T
\]

where \( T \) is the number of cases that are enduringly successfully treated.

Calculation of the Reproduction Number
The reproduction number \( R(t) \) can be defined as a time-varying function which is the average number of people an infected at time \( t \) would infect over his whole infectious lifespan if conditions remained the same as at time \( t \) [24]. \( R(t) \) is calculated as the sum of multiplications for all infectious stages of: the average duration in each stage of infection; the probability to get in that stage of infection; the infectiousness of that stage of infection \( (\rho^s) \); and the net transmission rate \( (\beta(t)) \).

The average duration in ‘Primary Infection’ is

\[
DY^p = \frac{1}{\alpha^p}
\]
The probability to get into ‘Primary Infection’ \((PY^p)\) is 1.
The contribution to the Reproduction Number of the ‘Primary Infection’ stage is
\[ RY^p = \beta(t) \rho^p DY^p PY^p \]

The average duration in ‘Undiagnosed Cases Progressing through Stages of Infection’ is
\[ DY^s = \frac{1}{\alpha + \delta^s(t)} \quad ; \quad (s = 1, \ldots, 5) \]

The probability to get into ‘Undiagnosed Cases Progressing through Stages of Infection’ is
\[ PY^s = \begin{array}{l} \frac{\alpha}{\alpha + \delta^{s+1}(t)} PY^{s-1} \quad ; \quad (s = 2, \ldots, 5) \\ 1 \quad ; \quad (s = 1) \end{array} \]

The contribution to the Reproduction Number of the ‘Undiagnosed Cases Progressing through Stages of Infection’ is
\[ RY^s = \beta(t) \rho^s DY^s PY^s \quad ; \quad (s = 1, \ldots, 5) \]

The average duration in ‘Diagnosed Cases Progressing through Stages of Infection’ is
\[ DY^s_{\text{diag}} = \frac{1}{\alpha + \gamma^s(t)} \quad ; \quad (s = 1, \ldots, 5) \]

The probability to get into ‘Diagnosed Cases Progressing through Stages of Infection’ is
\[ PY^s_{\text{diag}} = \begin{array}{l} \frac{\delta^s(t)}{\alpha + \delta^s(t)} PY^s \quad ; \quad (s = 1) \\ \frac{\delta^s(t)}{\alpha + \delta^s(t)} PY^s + \frac{\alpha}{\alpha + \gamma^{s+1}(t)} PY^{s-1}_{\text{diag}} \quad ; \quad (s = 2, \ldots, 5) \end{array} \]

The contribution to the Reproduction Number of the ‘Diagnosed Cases Progressing through Stages of Infection’ is
\[ RY^s_{\text{diag}} = \beta(t) \sigma^s DY^s_{\text{diag}} PY^s_{\text{diag}} \quad ; \quad (s = 1, \ldots, 5) \]
The average duration in ‘Failing Cases’ is

\[ DF_s^f = \frac{1}{\alpha + t_f^s} \quad ; \quad (s = 1, \ldots, 5; f = 1, 2) \]

\[ DF_s^f = \frac{1}{\alpha} \quad ; \quad (s = 1, \ldots, 5; f = 3) \]

The probability to get into ‘Failing Cases’ is

\[
PF_s^f = \frac{\kappa_f^s}{\mu + \kappa_f^s} \tau^s \left( \gamma P_{\text{diag}} + \frac{\alpha}{\alpha + t_f^s} PF_{s-1}^f \right) \quad ; \quad (s = 1; f = 1)
\]

\[
PF_s^f = \frac{\kappa_f^s}{\mu + \kappa_f^s} \tau^s \left( \gamma(t) P_{\text{diag}} + \frac{\alpha}{\alpha + t_f^s} PF_{s-1}^f \right) \quad ; \quad (s = 2, \ldots, 5; f = 1)
\]

\[
PF_s^f = \frac{\kappa_f^s}{\mu + \kappa_f^s} \tau^s t_f^s \left( PF_{s-1}^f \right) \quad ; \quad (s = 1; f = 2)
\]

\[
PF_s^f = \frac{\kappa_f^s}{\mu + \kappa_f^s} \tau^s \left( \frac{t_f^s}{\alpha + t_f^s} PF_{s-1}^f + \frac{\alpha}{\alpha + t_f^s} PF_{s-1}^f \right) \quad ; \quad (s = 2, \ldots, 5; f = 2)
\]

\[
PF_s^f = \frac{\kappa_f^s}{\mu + \kappa_f^s} \tau^s \left( \frac{t_f^s}{\alpha + t_f^s} PF_{s-1}^f \right) \quad ; \quad (s = 1; f = 3)
\]

\[
PF_s^f = \frac{\kappa_f^s}{\mu + \kappa_f^s} \tau^s \left( \frac{t_f^s}{\alpha + t_f^s} PF_{s-1}^f + PF_{s-1}^f \right) \quad ; \quad (s = 2, \ldots, 5; f = 3)
\]

The contribution to the Reproduction Number of the ‘Failing Cases’ is

\[
RF_s^f = \beta(t) \sigma \rho^s DF_s^f \quad ; \quad (s = 1, \ldots, 5; f = 1, \ldots, 3)
\]

The total Reproduction Number \( R(t) \) is defined as the sum of these terms

\[
R(t) = RY_p^{\sum_{i=1}^5} \left[ RY_{\text{diag}} + \sum_{f=1}^3 RF_s^f \right]
\]

We verified our formula numerically by checking for correct monotonicity in all the transmission and treatment parameters, and checking that \( R(t) = 1 \) is indeed the threshold between epidemic growth and decline when there is no import [47]. In the presence of imports, the threshold separates exponential epidemic growth (when \( R(t) > 1 \)) from a steady equilibrium of outbreak (when \( R(t) < 1 \)).
For this model of HIV, with its very long infectious period, the number of cases can grow slowly for many years approaching this equilibrium even when \( R(t) < 1 \) if \( R(t) \) is close to 1, so that the threshold \( (R(t) = 1) \) does not in practical terms mark a sudden change in dynamics in the way it does for classical models of acute infections with short infections [48].

We note that in the case of HIV, the epidemic situation can change substantially within the incubation period of a single individual, so that the actual lifetime reproduction number of an individual may differ substantially from the instantaneous value we estimate. Thus, \( R(t) \) as defined is in essence a descriptor of the epidemic at a point in time, similar in spirit to the exponential growth rate [49] or the incidence to prevalence ratio [3, 50]. There are however two advantages to \( R(t) \) over these alternative measures, namely that it is sensitive to the effects of the changing generation time induced by changing diagnosis rates and cART, and it has the meaningful threshold at \( R(t) = 1 \).

**Fitting to AIDS Cases**

The accumulation of new AIDS cases is defined as

\[
\frac{d(AIDS)}{dt} = \delta^5(t)(Y_{local}^5 + Y_{import}^5) + \alpha Y_{diag}^4 + \sum_{f=1}^{3} \alpha F_f^4
\]

where the annual increase is fitted to the annual data on new AIDS cases during the pre-cART era.

In the cART era, the cumulative number of cases diagnosed with AIDS are defined as

\[
\frac{d(AIDS_{diag})}{dt} = \delta^5(t)(Y_{import}^5 + Y_{local}^5)
\]

where the annual increase is fitted to the annual data on number of cases having AIDS when first diagnosed with HIV.

**Fitting to the Observed and True Number of Diagnoses Prior to 1996**

With the widespread implementation of cART in the Netherlands, ATHENA started monitoring all HIV patients. Patients who died before 1996 are thus not included in our data on annual cases with a new HIV diagnosis. Together with the data on annual AIDS cases in the pre-cART era and the knowledge of disease progression we were however able to infer from the observed number of diagnoses the underlying real diagnosis curve, including the people who had died prior to 1996. From the model, the total cumulative number of diagnoses are defined straightforwardly as
The probability of still being alive in 1996 per stage by year is shown in Figure 6.

**Figure 6.** Probability of surviving until 1996 given disease stage per year.

The diagnosis curves for local and imported cases, excluding cases dying before 1996, and thus comparable to the observed data are then calculated from the cumulative Erlang distribution, \( \Phi_{96} \)

\[
t_{96} \equiv \begin{cases} 
(1996 - t) & \text{if } t < 1996 \\ 
0 & \text{if } t > 1996 
\end{cases}
\]

\[
\Phi_{96}(t) = \sum_{j=1}^{5} \frac{(\alpha t_{96})^{5j}}{(5-j)!} \exp(-\alpha t_{96}) ; \quad (s = 1, \ldots, 5)
\]

\[
D_{\text{obs} \text{import}} = \sum_{j=1}^{5} \Phi_{96}(t) \left( D_{\text{import}}^j(t) - D_{\text{import}}^j(t-1) \right)
\]

\[
D_{\text{obs} \text{import}} = \sum_{j=1}^{5} \Phi_{96}(t) \left( D_{\text{local}}^j(t) - D_{\text{local}}^j(t-1) \right)
\]

These curves are then fitted to the annual diagnosis cases, separated for acquired abroad and acquired locally respectively. The imported cases initiate the epidemic and are estimated in two time intervals (see Table 1).
Average Time from Infection to Diagnosis

The average time to diagnosis, $T_D$, is calculated as

$$ T_D = \frac{1}{\alpha} + \sum_{s=1}^{4} \frac{\delta_i}{\delta_i + \alpha} \left( \frac{\alpha}{\delta_i + \alpha} \right)^{s-1} \left( \frac{\delta_i}{\delta_i + \alpha} \right)^{4} \left( \frac{1}{\delta_i + \alpha} \right) $$

The average time from infection to diagnosis is calculated for those infected in the corresponding time interval, and is thus not the average time to diagnosis for those diagnosed at that time.

Maximum Likelihood Model Fitting and Parameter Estimation

All the data we fit to – diagnoses, AIDS cases, AIDS diagnoses – are numbers of individuals. To define the likelihood, we assume these are Poisson (i.e. random) distributed around a mean defined by the model. If we define $X_i$ to be the data, and $x_i(\theta)$ the model predictions, dependent on a set of parameters $\theta$, then the likelihood can be defined as

$$ L(\theta) = \prod_i \frac{(x_i(\theta))^{X_i} \exp(-x_i(\theta))}{X_i!} $$

Since some of our data are truncated, we estimate the probability $p_i$ that the actual data are included in the database using the “survival until 1996” function ($\Phi_{1996}(t)$) above. If $Y_i$ were the actual number of diagnoses at point $i$, then the probability of being included in the data is Binomial distributed with probability $p_i$ and count $Y_i$ and the expected mean for $Y_i$ is $x_i(\theta)/p_i$. The likelihood for these truncated observations is thus

$$ L(\theta) = \prod_i \sum_{Y_i \geq X_i} \left( \frac{Y_i}{p_i} \right)^{X_i} \left( 1 - p_i \right)^{Y_i - X_i} \frac{(x_i(\theta)/p_i)^{Y_i} \exp(-x_i(\theta)/p_i)}{Y_i!} $$

but this is exactly equal to the simple Poisson likelihood defined above, so that data truncation does not need to be accounted for explicitly.

The maximum value the likelihood could ever take is obtained when the model predictions $x_i(\theta)$ are exactly equal to the data values $X_i$, i.e. the likelihood for the saturated model

$$ L_{sat} = \prod_i \frac{(X_i)^{X_i} \exp(-X_i)}{X_i!} $$
In our estimations, we use for convenience the equivalent deviance measure, given by

\[ \text{Dev}(\theta) = 2 \ln \left( \frac{L_{\text{max}}}{L(\theta)} \right) \]

\[ = 2 \left( X_i (\ln(X_i) - \ln(x_i(\theta))) + x_i(\theta) - X_i \right) \]

which is minimized to find the best fit parameters. This is equivalent to maximizing the likelihood, and gives a number standardized relative to the best possible fit (which would yield \( \text{Dev} = 0 \)).

Confidence intervals are determined by the likelihood ratio method. The deviance is minimized using the custom Optimize function in Berkeley Madonna which uses the relatively robust downhill simplex optimization algorithm to find the minimum. The algorithm was started from numerous starting values to ensure both that the optimization was robust and that a global minimum had been found.