Appendix

Modelling the HIV epidemic among men who have sex with men in the United Kingdom: quantifying the contributions to HIV transmission to better inform prevention initiatives

Rev: March 18, 2015

Table of contents

1. HIV transmission model

1.1 Overview of the HIV transmission model

A mathematical model of the HIV transmission dynamics among men who have sex with men (MSM) aged 15-64 in the United Kingdom was developed based on a system of differential equations. We included in the model a number of behavioural and biological heterogeneities to simulate the epidemic over the 20-year time course from 2001 to 2020. The simulations were started from the beginning of 2000 to allow the model to become more stable before conducting further analyses from 2001 onwards. Our model was parameterised based on available information from multiple sources including behavioural surveys, surveillance systems, and the literature. We also fitted the model to the data and external estimates using the Monte-Carlo filtering method [1]. The model was used to estimate the current epidemic situation in various aspects and predict how it would evolve in the near future. The detailed descriptions of the model are shown in this appendix.

1.2 Key assumptions

A number of assumptions were made throughout the model. The main assumptions that our model was based on are as follows:

- The model is a deterministic population-based model therefore the members of any subgroups all share the same characteristics.

- MSM get infected with HIV only through anal or oral intercourse in a sexual partnership with a man.
- The model allows only two types of sexual partnership: one-off and repeat sexual partnership (the definitions are provided in Section 1.3.4).
- Concurrency of the same partnership types (one-off and repeated) is not allowed (serial monogamy).
- Anal-sex partnerships involve only anal sexual intercourse. Similarly, oral-sex partnership involves only oral sexual intercourse.
- All HIV-positive MSM start from the first disease stage (primary HIV infection) and progress to the next stages with decreased CD4 cell counts. The model does not allow an increase in CD4 cell counts.
- Once treated, individuals remain in the treatment stage until removed from the model.
- The sensitivity and specificity of the HIV test is 100%.
- Safe sex refers to sexual intercourse with condom use. Unsafe or unprotected sex refers to sexual intercourse without condom use.
- No HIV infection from outside the UK is included.
- We modelled the HIV epidemic in the UK as a whole and there was no differentiation by geographic area.

1.3 Model structure

Structuring the model concerns stratification of the modelled population according to various characteristics that have an essential effect on the model outputs [2]. The

entire MSM populations in the UK were stratified simultaneously by MSM types, age, sexual activity level, partnership status, and HIV-related status.

1.3.1 Current and past MSM

Homosexual men in general have been considered at high risk for HIV infection specifically because of the frequent anal sexual intercourse they have. However, a considerable number of UK homosexual males have never had anal sex with men and their last sex (defined as a genital contact) with men was more than 5 years ago [3]. These were excluded from the model. Those who have had sex with men in the last 5 years are included as 'current' MSM. Those who have had sex with men but not in the last 5 years and also who last had anal sex more than 5 years ago are defined as 'past' MSM. Only current MSM acquire new partners and are at risk of HIV infection. The characteristics of past MSM resemble that of current low-activity MSM. All MSM in the model are the summation of the current and past MSM.

1.3.2 Age

The age range of the modelled population was 15-64 years. We distinguished the younger and older MSM by dividing them into two age groups of 15-34 and 35-64 years. Individuals in the younger group progress to the older group and individuals in the older group move out of the system at constant rates over time.

1.3.3 Sexual activity level

We defined two levels of sexual activity—low and high—based on the number of sexual partners per year. The low-activity MSM have the average new male sexual

partner per year of less than or equal to one. The high-activity MSM have more than one new male sexual partner per year. We chose this cut-off point to differentiate MSM who have a similar rate of partner change to the average of UK heterosexual male from those MSM at greater risk, as well as to ensure that low-activity MSM have no concurrent sexual partnerships. Together with the two age groups we formed a total of four 'classes' of MSM in our model, i.e. age group 1 – low activity, age group 1 – high activity, age group 2 – low activity, and age group 2 – high activity.

1.3.4 Partnership status

We distinguished between one-off and repeat sexual partners by defining that the one-off sexual partnership is a relationship that has only a single sexual contact which is assumed to start and end instantaneously. The repeat sexual partnership is defined as a relationship in which sexual contact occurs more than once, such as a married couple, a steady partnership, or a casual partnership that has several sexual encounters. Unlike the typical mean-field model, the partnership-based model [4, 5] allows a repeat sexual partnership to last for a finite duration of time, and only during that period can HIV transmission between partners can occur. Both individuals with and without a repeat sexual partner can have one-off sexual encounter and therefore be at risk for HIV infection.

1.3.5 Disease stage and CD4 levels

Based primarily on the CD4 cell counts, the long and variable natural history of HIV infection was divided into five stages: primary HIV infection (PHI), CD4 $\geq$ 500, 350-499, 200-349, and < 200 cells/μL. This allowed us to distinguish among disease stages

with respect to the three important factors: HIV transmission probability, HIV diagnosis rate, and the rate of initiating antiretroviral treatment (ART), determining the outcomes of the model. HIV-positive men progress from the first stage through the last without going backward. We assumed no CD4 stratification for MSM on treatment.

1.3.6 HIV diagnosis and treatment status

The five disease stages were further stratified into undiagnosed and diagnosed MSM. The latter was split into those who have never and ever been treated with ART. We noted that ART-treated MSM included all men who have ever been treated with ART regardless of their current ART status. Furthermore, the model does not allow individuals to move back to the previous stages in which no ART is administered. Taken together, there are 12 different stages of HIV infection illustrated schematically in Figure 1 in the main text.

1.4 Model equations and calculations

The model was constructed based on a set of ordinary differential equations. We denoted the numbers of single current MSM and single past MSM by Y and Z, respectively. The total numbers of single MSM, $X = Y + Z$. Let P be the numbers of pairs (not individuals) of current MSM. The total number of MSM in the model, $N = X + 2P$. The subscripts j,k,h of $X_{j,k,h}$ are denoted as age group (j=1 for age group 1, j=2 for age group 2), sexual activity group (k=1 for low-activity group, k=2 for high-activity group), and HIV infection stages ($h = 1,\ldots,12$ according to 12 HIV

stages presented in Figure 1 in the main text: 1=susceptibles, 2=undiagnosed PHI, 3=diagnosed PHI, 4=undiagnosed CD4>500, 5=diagnosed CD4>500, 6=undiagnosed CD4 350-499, 7=diagnosed CD4 350-499, 8=undiagnosed CD4 200-349, 9=diagnosed CD4 200-349, 10=undiagnosed CD4<200, 11=diagnosed CD4<200, and 12=on treatment). The superscripts m,n,r of the numbers of pairs $P_{j,k,h}^{m,n,r}$ are denoted as age group *m*, sexual activity group *n*, and HIV stages *r* of the repeat sexual partner, and that $P_{j,k,h}^{m,n,r} = P_{m,n,r}^{j,k,h}$. Our partnership-based model consists of 1,296 compartments which, following Xiridou et al [6] and Powers et al [7], can be described by the 46 equations listed in the next section. All parameters are summarised in Table S1. The time step in the model is a day (δ =1/365=0.0027). For any per-year rates, multiplying by δ will result in the corresponding rates per time-step.

1.4.1 Differential equations

We noted that $[j,k] \neq [m,n]$ in all equations below, and the $\{condition\}$ denotes a specific condition that must be satisfied to enable the corresponding term.

Current single MSM

$$\frac{dY_{j,k,1}}{dt} = \nu_{j,k}\{j=1\} - \eta_{j,k,1}\{k=1\} + \sigma_{j,k}(P_{j,k,1}^{j,k,1} + \sum_{m=1}^{2}\sum_{n=1}^{2}\sum_{r=1}^{12} P_{j,k,1}^{m,n,r}) + \alpha_{j-1}Y_{j-1,k,1}\{j=2\} - (\alpha_j + \mu_{j,1} + \rho_{j,k}^{rep} + \lambda_{Y,j,k}^{one})Y_{j,k,1} \quad (1)$$

$$\frac{dY_{j,k,2}}{dt} = \sigma_{j,k}(P_{j,k,2}^{j,k,2} + \sum_{m=1}^{2}\sum_{n=1}^{2}\sum_{r=1}^{12} P_{j,k,2}^{m,n,r}) + \alpha_{j-1}Y_{j-1,k,2}\{j=2\} + \lambda_{Y,j,k}^{one}Y_{j,k,1} - (\alpha_j + \mu_{j,2} + \gamma_2 + \phi_{j,k,2} + \rho_{j,k}^{rep})Y_{j,k,2} \quad (2)$$

$$\frac{dY_{j,k,3}}{dt} = \sigma_{j,k}(P_{j,k,3}^{j,k,3} + \sum_{m=1}^{2}\sum_{n=1}^{2}\sum_{r=1}^{12} P_{j,k,3}^{m,n,r}) + \alpha_{j-1}Y_{j-1,k,3}\{j=2\} + (1-s_k)\phi_{j,k,2}Y_{j,k,2} + s_l\phi_{j,l,2}Y_{j,l,2}\{l \neq k\} - (\alpha_j + \mu_{j,3} + \gamma_3 + \rho_{j,k}^{rep})Y_{j,k,3} \quad (3)$$

$$\frac{dY_{j,k,h}}{dt} = -\eta_{j,k,h}\{k=1\} + \sigma_{j,k}(P_{j,k,h}^{j,k,h} + \sum_{m=1}^{2}\sum_{n=1}^{2}\sum_{r=1}^{12} P_{j,k,h}^{m,n,r}) + \alpha_{j-1}Y_{j-1,k,h}\{j=2\} + \gamma_{h-2}Y_{j,k,h-2} - (\alpha_j + \mu_{j,h} + \gamma_h\{h \neq 10\} + \phi_{j,k,h} + \rho_{j,k}^{rep})Y_{j,k,h} \quad \text{for } h = 4, 6, 8, 10. \quad (4)$$

$$\frac{dY_{j,k,h}}{dt} = -\eta_{j,k,h}\{k=1\} + \sigma_{j,k}(P_{j,k,h}^{j,k,h} + \sum_{m=1}^{2}\sum_{n=1}^{2}\sum_{r=1}^{12} P_{j,k,h}^{m,n,r}) + \alpha_{j-1}Y_{j-1,k,h}\{j=2\} + (1-s_k)\phi_{j,k,h-1}Y_{j,k,h-1} + s_l\phi_{j,l,h-1}Y_{j,l,h-1}\{l \neq k\} + \gamma_{h-2}Y_{j,k,h-2} - (\alpha_j + \mu_{j,h} + \gamma_h\{h \neq 11\} + \tau_h + \rho_{j,k}^{rep})Y_{j,k,h} \quad \text{for } h = 5, 7, 9, 11. \quad (5)$$

$$\frac{dY_{j,k,12}}{dt} = -\eta_{j,k,12}\{k=1\} + \sigma_{j,k}(P_{j,k,12}^{j,k,12} + \sum_{m=1}^{2}\sum_{n=1}^{2}\sum_{r=1}^{12} P_{j,k,12}^{m,n,r}) + \alpha_{j-1}Y_{j-1,k,12}\{j=2\} + \sum_{g \in s}\tau_g Y_{j,k,g} - (\alpha_j + \mu_{j,12} + \rho_{j,k}^{rep})Y_{j,k,12} \quad \text{where } s = 5, 7, 9, 11. \quad (6)$$

Past MSM

$$\frac{dZ_{j,1,1}}{dt} = \eta_{j,1} + \alpha_{j-1} Z_{j-1,1,1}\{j = 2\} - (\alpha_j + \mu_{j,1}) Z_{j,1,1} \quad (7)$$

$$\frac{dZ_{j,1,2}}{dt} = \eta_{j,2} + \alpha_{j-1} Z_{j-1,1,2}\{j = 2\} - (\alpha_j + \mu_{j,2} + \gamma_2 + \phi_{j,1,2}) Z_{j,1,2} \quad (8)$$

$$\frac{dZ_{j,1,3}}{dt} = \eta_{j,3} + \alpha_{j-1} Z_{j-1,1,3}\{j = 2\} + \phi_{j,1,2} Z_{j,1,2} - (\alpha_j + \mu_{j,3} + \gamma_3) Z_{j,1,3} \quad (9)$$

$$\frac{dZ_{j,1,h}}{dt} = \eta_{j,h} + \alpha_{j-1} Z_{j-1,1,h}\{j = 2\} + \gamma_{h-2} Z_{j,1,h-2} - (\alpha_j + \mu_{j,h} + \gamma_h\{h \neq 10\} + \phi_{j,1,h}) Z_{j,1,h} \quad \text{for } h = 4, 6, 8, 10. \quad (10)$$

$$\frac{dZ_{j,1,h}}{dt} = \eta_{j,h} + \alpha_{j-1} Z_{j-1,1,h}\{j = 2\} + \phi_{j,1,h-1} Z_{j,1,h-1} + \gamma_{h-2} Z_{j,1,h-2} - (\alpha_j + \mu_{j,h} + \gamma_h\{h \neq 11\} + \tau_h) Z_{j,1,h} \quad \text{for } h = 5, 7, 9, 11. \quad (11)$$

$$\frac{dZ_{j,1,12}}{dt} = \eta_{j,12} + \alpha_{j-1} Z_{j-1,1,12}\{j = 2\} + \sum_{g \in s} \tau_g Z_{j,1,g} - (\alpha_j + \mu_{j,12}) Z_{j,1,12} \quad \text{where } s = 5, 7, 9, 11. \quad (12)$$

Paired MSM

Susceptible (HIV stage no.1) – Susceptible (HIV stage no.1)

$$\frac{dP_{j,k,1}^{j,k,1}}{dt} = \psi_{j,k,1}^{rep,j,k,1}\rho_{j,k}^{rep}Y_{j,k,1} + 2\alpha_{j-1}P_{j-1,k,1}^{j,k,1}\{j=2\} - (2\alpha_j + 2\mu_{j,1} + \sigma_{j,k} + 2\lambda_{P,j,k}^{one})P_{j,k,1}^{j,k,1} \tag{13}$$

$$\frac{dP_{j,k,1}^{m,n,1}}{dt} = \psi_{j,k,1}^{rep,m,n,1}\rho_{j,k}^{rep}Y_{j,k,1} + \alpha_{j-1}P_{j-1,k,1}^{m,n,1}\{j=2\} + \alpha_{m-1}P_{m-1,n,1}^{j,k,1}\{m=2\} - (\alpha_j + \alpha_m + \mu_{j,1} + \mu_{m,1} + \frac{\sigma_{j,k}+\sigma_{m,n}}{2} + \lambda_{P,j,k}^{one} + \lambda_{P,m,n}^{one})P_{j,k,1}^{m,n,1} \tag{14}$$

Susceptible (HIV stage no.1) – Infected (HIV stage no.2-12)

$$\frac{dP_{j,k,1}^{j,k,2}}{dt} = \psi_{j,k,1}^{rep,j,k,2}\rho_{j,k}^{rep}Y_{j,k,1} + 2\alpha_{j-1}P_{j-1,k,1}^{j,k,2}\{j=2\} + 2\lambda_{P,j,k}^{one}P_{j,k,1}^{j,k,1} - (2\alpha_j + \mu_{j,1} + \mu_{j,2} + \gamma_2 + \phi_{j,k,2} + \sigma_{j,k} + \lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,j,k,2})P_{j,k,1}^{j,k,2} \tag{15}$$

$$\begin{aligned}\frac{dP_{j,k,1}^{m,n,2}}{dt} &= \psi_{j,k,1}^{rep,m,n,2}\rho_{j,k}^{rep}Y_{j,k,1} + \alpha_{j-1}P_{j-1,k,1}^{m,n,2}\{j=2\} + \alpha_{m-1}P_{m-1,n,2}^{j,k,1}\{m=2\} + (\lambda_{P,j,k}^{one} + \lambda_{P,m,n}^{one})P_{j,k,1}^{m,n,1} \\ &\quad - (\alpha_j + \alpha_m + \mu_{j,1} + \mu_{m,2} + \gamma_2 + \phi_{m,n,2} + \frac{\sigma_{j,k}+\sigma_{m,n}}{2} + \lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,m,n,2})P_{j,k,1}^{m,n,2}\end{aligned} \tag{16}$$

$$\frac{dP_{j,k,1}^{j,k,3}}{dt} = \psi_{j,k,1}^{rep,j,k,3}\rho_{j,k}^{rep}Y_{j,k,1} + 2\alpha_{j-1}P_{j-1,k,1}^{j,k,3}\{j=2\} + (1-s_k)\phi_{j,k,2}P_{j,k,1}^{j,k,2} + s_l\phi_{j,l,2}P_{j,k,1}^{j,l,2}\{l\neq k\} - (2\alpha_j + \mu_{j,1} + \mu_{j,3} + \gamma_3 + \sigma_{j,k} + \lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,j,k,3})P_{j,k,1}^{j,k,3} \tag{17}$$

$$\begin{aligned}\frac{dP_{j,k,1}^{m,n,3}}{dt} &= \psi_{j,k,1}^{rep,m,n,3}\rho_{j,k}^{rep}Y_{j,k,1} + \alpha_{j-1}P_{j-1,k,1}^{m,n,3}\{j=2\} + \alpha_{m-1}P_{m-1,n,3}^{j,k,1}\{m=2\} + (1-s_n)\phi_{m,n,2}P_{j,k,1}^{m,n,2} + s_l\phi_{m,l,2}P_{j,k,1}^{m,l,2}\{l\neq n\} \\ &\quad - (\alpha_j + \alpha_m + \mu_{j,1} + \mu_{m,3} + \gamma_3 + \frac{\sigma_{j,k}+\sigma_{m,n}}{2} + \lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,m,n,3})P_{j,k,1}^{m,n,3}\end{aligned} \tag{18}$$

$$\frac{dP_{j,k,1}^{j,k,h}}{dt} = \psi_{j,k,1}^{rep,j,k,h}\rho_{j,k}^{rep}Y_{j,k,1} + 2\alpha_{j-1}P_{j-1,k,1}^{j,k,h}\{j=2\} + \gamma_{h-2}P_{j,k,1}^{j,k,h-2}$$
$$-(2\alpha_j + \mu_{j,1} + \mu_{j,h} + \gamma_h\{h\neq 10\} + \phi_{j,k,h} + \sigma_{j,k} + \lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,j,k,h})P_{j,k,1}^{j,k,h}$$

for h = 4, 6, 8, 10. (19)

$$\frac{dP_{j,k,1}^{m,n,r}}{dt} = \psi_{j,k,1}^{rep,m,n,r}\rho_{j,k}^{rep}Y_{j,k,1} + \alpha_{j-1}P_{j-1,k,1}^{m,n,r}\{j=2\} + \alpha_{m-1}P_{m-1,n,r}^{j,k,1}\{m=2\} + \gamma_{r-2}P_{j,k,1}^{m,n,r-2}$$
$$-(\alpha_j + \alpha_m + \mu_{j,1} + \mu_{m,r} + \gamma_r\{r\neq 10\} + \phi_{m,n,r} + \frac{\sigma_{j,k}+\sigma_{m,n}}{2} + \lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,m,n,r})P_{j,k,1}^{m,n,r}$$

for r = 4, 6, 8, 10. (20)

$$\frac{dP_{j,k,1}^{j,k,h}}{dt} = \psi_{j,k,1}^{rep,j,k,h}\rho_{j,k}^{rep}Y_{j,k,1} + 2\alpha_{j-1}P_{j-1,k,1}^{j,k,h}\{j=2\} + \gamma_{h-2}P_{j,k,1}^{j,k,h-2} + (1-s_k)\phi_{j,k,h-1}P_{j,k,1}^{j,k,h-1} + s_l\phi_{j,l,h-1}P_{j,k,1}^{j,l,h-1}\{l\neq k\}$$
$$-(2\alpha_j + \mu_{j,1} + \mu_{j,h} + \gamma_h\{h\neq 11\} + \tau_h + \sigma_{j,k} + \lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,j,k,h})P_{j,k,1}^{j,k,h}$$

for h = 5, 7, 9, 11. (21)

$$\frac{dP_{j,k,1}^{m,n,r}}{dt} = \psi_{j,k,1}^{rep,m,n,r}\rho_{j,k}^{rep}Y_{j,k,1} + \alpha_{j-1}P_{j-1,k,1}^{m,n,r}\{j=2\} + \alpha_{m-1}P_{m-1,n,r}^{j,k,1}\{m=2\} + \gamma_{r-2}P_{j,k,1}^{m,n,r-2} + (1-s_n)\phi_{m,n,r-1}P_{j,k,1}^{m,n,r-1} + s_l\phi_{m,l,r-1}P_{j,k,1}^{m,l,r-1}\{l\neq n\}$$
$$-(\alpha_j + \alpha_m + \mu_{j,1} + \mu_{m,r} + \gamma_r\{r\neq 11\} + \tau_r + \frac{\sigma_{j,k}+\sigma_{m,n}}{2} + \lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,m,n,r})P_{j,k,1}^{m,n,r}$$

for r = 5, 7, 9, 11. (22)

$$\frac{dP_{j,k,1}^{j,k,12}}{dt} = \psi_{j,k,1}^{rep,j,k,12}\rho_{j,k}^{rep}Y_{j,k,1} + 2\alpha_{j-1}P_{j-1,k,1}^{j,k,12}\{j=2\} + \sum_{g\in s}\tau_g P_{j,k,1}^{j,k,g} - (2\alpha_j + \mu_{j,1} + \mu_{j,12} + \sigma_{j,k} + \lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,j,k,12})P_{j,k,1}^{j,k,12}$$

where s = 5, 7, 9, 11. (23)

$$\frac{dP_{j,k,1}^{m,n,12}}{dt} = \psi_{j,k,1}^{rep,m,n,12}\rho_{j,k}^{rep}Y_{j,k,1} + \alpha_{j-1}P_{j-1,k,1}^{m,n,12}\{j=2\} + \alpha_{m-1}P_{m-1,n,10}^{j,k,1}\{m=2\} + \sum_{g\in s}\tau_g P_{j,k,1}^{m,n,g}$$
$$-(\alpha_j + \alpha_m + \mu_{j,1} + \mu_{m,12} + \frac{\sigma_{j,k}+\sigma_{m,n}}{2} + \lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,m,n,12})P_{j,k,1}^{m,n,12}$$

where s = 5, 7, 9, 11. (24)

Undiagnosed PHI (HIV stage no.2) – Infected (HIV stage no.2-12)

$$\frac{dP_{j,k,2}^{j,k,2}}{dt} = \psi_{j,k,2}^{rep,j,k,2}\rho_{j,k}^{rep}Y_{j,k,2} + 2\alpha_{j-1}P_{j-1,k,2}^{j,k,2}\{j=2\} + (\lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,j,k,2})P_{j,k,1}^{j,k,2} - (2\alpha_j + 2\mu_{j,2} + 2\gamma_2 + 2\phi_{j,k,2} + \sigma_{j,k})P_{j,k,2}^{j,k,2} \quad (25)$$

$$\frac{dP_{j,k,2}^{m,n,2}}{dt} = \psi_{j,k,2}^{rep,m,n,2}\rho_{j,k}^{rep}Y_{j,k,2} + \alpha_{j-1}P_{j-1,k,2}^{m,n,2}\{j=2\} + \alpha_{m-1}P_{m-1,n,2}^{j,k,2}\{m=2\} + (\lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,m,n,2})P_{j,k,1}^{m,n,2} + (\lambda_{P,m,n}^{one} + \lambda_{m,n}^{rep,j,k,2})P_{j,k,2}^{m,n,1}$$
$$-(\alpha_j + \alpha_m + \mu_{j,2} + \mu_{m,2} + 2\gamma_2 + \phi_{j,k,2} + \phi_{m,n,2} + \frac{\sigma_{j,k} + \sigma_{m,n}}{2})P_{j,k,2}^{m,n,2} \quad (26)$$

$$\frac{dP_{j,k,2}^{j,k,3}}{dt} = \psi_{j,k,2}^{rep,j,k,3}\rho_{j,k}^{rep}Y_{j,k,2} + 2\alpha_{j-1}P_{j-1,k,2}^{j,k,3}\{j=2\} + 2\phi_{j,k,2}P_{j,k,2}^{j,k,2} + (\lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,j,k,3})P_{j,k,1}^{j,k,3} - (2\alpha_j + \mu_{j,2} + \mu_{j,3} + \gamma_2 + \gamma_3 + \phi_{j,k,2} + \sigma_{j,k})P_{j,k,2}^{j,k,3} \quad (27)$$

$$\frac{dP_{j,k,2}^{m,n,3}}{dt} = \psi_{j,k,2}^{rep,m,n,3}\rho_{j,k}^{rep}Y_{j,k,2} + \alpha_{j-1}P_{j-1,k,2}^{m,n,3}\{j=2\} + \alpha_{m-1}P_{m-1,n,2}^{j,k,3}\{m=2\} + \phi_{m,n,2}P_{j,k,2}^{m,n,2} + (\lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,m,n,3})P_{j,k,1}^{m,n,3}$$
$$-(\alpha_j + \alpha_m + \mu_{j,2} + \mu_{m,3} + \gamma_2 + \gamma_3 + \phi_{j,k,2} + \frac{\sigma_{j,k} + \sigma_{m,n}}{2})P_{j,k,2}^{m,n,3} \quad (28)$$

$$\frac{dP_{j,k,2}^{j,k,h}}{dt} = \psi_{j,k,2}^{rep,j,k,h}\rho_{j,k}^{rep}Y_{j,k,2} + 2\alpha_{j-1}P_{j-1,k,2}^{j,k,h}\{j=2\} + \gamma_{h-2}P_{j,k,2}^{j,k,h-2} + (\lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,j,k,h})P_{j,k,1}^{j,k,h}$$
$$-(2\alpha_j + \mu_{j,2} + \mu_{j,h} + \gamma_2 + \gamma_h\{h\neq 10\} + \phi_{j,k,2} + \phi_{j,k,h} + \sigma_{j,k})P_{j,k,2}^{j,k,h} \quad \text{for } h = 4, 6, 8, 10. \quad (29)$$

$$\frac{dP_{j,k,2}^{m,n,r}}{dt} = \psi_{j,k,2}^{rep,m,n,r}\rho_{j,k}^{rep}Y_{j,k,2} + \alpha_{j-1}P_{j-1,k,2}^{m,n,r}\{j=2\} + \alpha_{m-1}P_{m-1,n,r}^{j,k,2}\{m=2\} + \gamma_{r-2}P_{j,k,2}^{m,n,r-2} + (\lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,m,n,r})P_{j,k,1}^{m,n,r}$$
$$-(\alpha_j + \alpha_m + \mu_{j,2} + \mu_{m,r} + \gamma_2 + \gamma_r\{r\neq 10\} + \phi_{j,k,2} + \phi_{j,k,r} + \frac{\sigma_{j,k} + \sigma_{m,n}}{2})P_{j,k,2}^{m,n,r} \quad \text{for } r = 4, 6, 8, 10. \quad (30)$$

$$\frac{dP_{j,k,2}^{j,k,h}}{dt} = \psi_{j,k,2}^{rep,j,k,h}\rho_{j,k}^{ste}Y_{j,k,2} + 2\alpha_{j-1}P_{j-1,k,2}^{j,k,h}\{j=2\} + \gamma_{h-2}P_{j,k,2}^{j,k,h-2} + (1-s_k)\phi_{j,k,h-1}P_{j,k,2}^{j,k,h-1} + s_l\phi_{j,l,h-1}P_{j,k,2}^{j,l,h-1}\{l\neq k\}$$
$$+ (\lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,j,k,h})P_{j,k,1}^{j,k,h} - (2\alpha_j + \mu_{j,2} + \mu_{j,h} + \gamma_2 + \gamma_h\{h\neq 11\} + \phi_{j,k,2} + \phi_{j,k,h} + \tau_h + \sigma_{j,k})P_{j,k,2}^{j,k,h}$$

for h = 5, 7, 9, 11. (31)

$$\frac{dP_{j,k,2}^{m,n,r}}{dt} = \psi_{j,k,2}^{rep,m,n,r}\rho_{j,k}^{rep}Y_{j,k,2} + \alpha_{j-1}P_{j-1,k,2}^{m,n,r}\{j=2\} + \alpha_{m-1}P_{m-1,n,r}^{j,k,2}\{m=2\} + \gamma_{r-2}P_{j,k,2}^{m,n,r-2} + (1-s_n)\phi_{m,n,r-1}P_{j,k,2}^{m,n,r-1} + s_l\phi_{m,l,r-1}P_{j,k,2}^{m,l,r-1}\{l\neq n\}$$
$$+ (\lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,m,n,r})P_{j,k,1}^{m,n,r} - (\alpha_j + \alpha_m + \mu_{j,2} + \mu_{m,r} + \gamma_2 + \gamma_r\{r\neq 11\} + \phi_{j,k,2} + \tau_r + \frac{\sigma_{j,k}+\sigma_{m,n}}{2})P_{j,k,2}^{m,n,r}$$

for r = 5, 7, 9, 11. (32)

$$\frac{dP_{j,k,2}^{j,k,12}}{dt} = \psi_{j,k,2}^{rep,j,k,12}\rho_{j,k}^{rep}Y_{j,k,2} + 2\alpha_{j-1}P_{j-1,k,2}^{j,k,12}\{j=2\} + \sum_{g\in s}\tau_g P_{j,k,2}^{j,k,g} + (\lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,j,k,12})P_{j,k,1}^{j,k,12}$$
$$- (2\alpha_j + \mu_{j,2} + \mu_{j,12} + \gamma_2 + \phi_{j,k,2} + \sigma_{j,k})P_{j,k,2}^{j,k,12}$$

where s = 5, 7, 9, 11. (33)

$$\frac{dP_{j,k,2}^{m,n,12}}{dt} = \psi_{j,k,2}^{rep,m,n,12}\rho_{j,k}^{rep}Y_{j,k,2} + \alpha_{j-1}P_{j-1,k,2}^{m,n,12}\{j=2\} + \alpha_{m-1}P_{m-1,n,12}^{j,k,2}\{m=2\} + \sum_{g\in s}\tau_g P_{j,k,2}^{m,n,g} + (\lambda_{P,j,k}^{one} + \lambda_{j,k}^{rep,m,n,12})P_{j,k,1}^{m,n,12}$$
$$- (\alpha_j + \alpha_m + \mu_{j,2} + \mu_{m,12} + \gamma_2 + \phi_{j,k,2} + \frac{\sigma_{j,k}+\sigma_{m,n}}{2})P_{j,k,2}^{m,n,12}$$

where s = 5, 7, 9, 11. (34)

Infected except U2 (HIV stage no.3-12) – Infected except U2 (HIV stage no.3-12)

Undiagnosed – Undiagnosed

$$\frac{dP_{j,k,h}^{j,k,r}}{dt} = \psi_{j,k,h}^{rep,j,k,r}\rho_{j,k}^{ste}Y_{j,k,h} + 2\alpha_{j-1}P_{j-1,k,h}^{j,k,r}\{j=2\} + \gamma_{h-2}P_{j,k,h-2}^{j,k,r} + \gamma_{r-2}P_{j,k,h}^{j,k,r-2}$$
$$-(2\alpha_j + \mu_{j,h} + \mu_{j,r} + \gamma_h\{h\neq 10\} + \gamma_r\{r\neq 10\} + \phi_{j,k,h} + \phi_{j,k,r} + \sigma_{j,k})P_{j,k,h}^{j,k,r}$$

for h = 4, 6, 8, 10, and r = 4, 6, 8, 10. (35)

$$\frac{dP_{j,k,h}^{m,n,r}}{dt} = \psi_{j,k,h}^{rep,m,n,r}\rho_{j,k}^{rep}Y_{j,k,h} + \alpha_{j-1}P_{j-1,k,h}^{m,n,r}\{j=2\} + \alpha_{m-1}P_{m-1,n,r}^{j,k,h}\{m=2\} + \gamma_{r-2}P_{j,k,h}^{m,n,r-2}$$
$$-(\alpha_j + \alpha_m + \mu_{j,h} + \mu_{m,r} + \gamma_h\{h\neq 10\} + \gamma_r\{r\neq 10\} + \phi_{j,k,h} + \phi_{j,k,r} + \frac{\sigma_{j,k}+\sigma_{m,n}}{2})P_{j,k,h}^{m,n,r}$$

for h = 4, 6, 8, 10, and r = 4, 6, 8, 10. (36)

Undiagnosed – Diagnosed

$$\frac{dP_{j,k,h}^{j,k,r}}{dt} = \psi_{j,k,h}^{rep,j,k,r}\rho_{j,k}^{rep}Y_{j,k,h} + 2\alpha_{j-1}P_{j-1,k,h}^{j,k,r}\{j=2\} + \gamma_{h-2}P_{j,k,h-2}^{j,k,r} + \gamma_{r-2}P_{j,k,h}^{j,k,r-2}\{r\neq 3\}$$
$$+(1-s_k)\phi_{j,k,r-1}P_{j,k,h}^{j,k,r-1} + s_l\phi_{j,l,r-1}P_{j,k,h}^{j,l,r-1}\{l\neq k\}$$
$$-(2\alpha_j + \mu_{j,h} + \mu_{j,r} + \gamma_h\{h\neq 10\} + \gamma_r\{r\neq 11\} + \phi_{j,k,h} + \tau_r\{r\neq 3\} + \sigma_{j,k})P_{j,k,h}^{j,k,r}$$

for h = 4, 6, 8, 10, and r = 3, 5, 7, 9, 11. (37)

$$\frac{dP_{j,k,h}^{m,n,r}}{dt} = \psi_{j,k,h}^{rep,m,n,r}\rho_{j,k}^{rep}Y_{j,k,h} + \alpha_{j-1}P_{j-1,k,h}^{m,n,r}\{j=2\} + \alpha_{m-1}P_{m-1,n,r}^{j,k,h}\{m=2\} + \gamma_{h-2}P_{j,k,h-2}^{m,n,r} + \gamma_{r-2}P_{j,k,h}^{m,n,r-2}\{r\neq 3\}$$
$$+(1-s_n)\phi_{m,n,r-1}P_{j,k,h}^{m,n,r-1} + s_l\phi_{m,l,r-1}P_{j,k,h}^{m,l,r-1}\{l\neq n\}$$
$$-(\alpha_j + \alpha_m + \mu_{j,h} + \mu_{m,r} + \gamma_h\{h\neq 10\} + \gamma_r\{r\neq 11\} + \phi_{j,k,h} + \tau_r\{r\neq 3\} + \frac{\sigma_{j,k}+\sigma_{m,n}}{2})P_{j,k,h}^{m,n,r}$$

for h = 4, 6, 8, 10, and r = 3, 5, 7, 9, 11. (38)

Undiagnosed – On treatment

$$\frac{dP_{j,k,h}^{j,k,12}}{dt} = \psi_{j,k,h}^{rep,j,k,12}\rho_{j,k}^{rep}Y_{j,k,h} + 2\alpha_{j-1}P_{j-1,k,h}^{j,k,12}\{j=2\} + \sum_{g\in s}\tau_g P_{j,k,h}^{j,k,g} - (2\alpha_j + \mu_{j,h} + \mu_{j,12} + \gamma_h\{h\neq 10\} + \phi_{j,k,h} + \sigma_{j,k})P_{j,k,h}^{j,k,12}$$

for h = 4, 6, 8, 10, and s = 5, 7, 9, 11. (39)

$$\frac{dP_{j,k,h}^{m,n,12}}{dt} = \psi_{j,k,h}^{rep,m,n,12}\rho_{j,k}^{rep}Y_{j,k,h} + \alpha_{j-1}P_{j-1,k,h}^{m,n,12}\{j=2\} + \alpha_{m-1}P_{m-1,n,12}^{j,k,h}\{m=2\} + \sum_{g\in s}\tau_g P_{j,k,h}^{m,n,g} - (\alpha_j + \alpha_m + \mu_{j,h} + \mu_{m,12} + \gamma_h\{h\neq 10\} + \phi_{j,k,h} + \frac{\sigma_{j,k}+\sigma_{m,n}}{2})P_{j,k,h}^{m,n,12}$$

for h = 4, 6, 8, 10, and s = 5, 7, 9, 11. (40)

Diagnosed – Diagnosed

$$\frac{dP_{j,k,h}^{j,k,r}}{dt} = \psi_{j,k,h}^{rep,j,k,r}\rho_{j,k}^{rep}Y_{j,k,h} + 2\alpha_{j-1}P_{j-1,k,h}^{j,k,r}\{j=2\} + \gamma_{h-2}P_{j,k,h-2}^{j,k,r}\{h\neq 3\} + \gamma_{r-2}P_{j,k,h}^{j,k,r-2}\{r\neq 3\} + (1-s_k)\phi_{j,k,h-1}P_{j,k,h-1}^{j,k,r} + s_l\phi_{j,l,h-1}P_{j,l,h-1}^{j,k,r}\{l\neq k\} + (1-s_k)\phi_{j,k,r-1}P_{j,k,h}^{j,k,r-1} + s_l\phi_{j,l,r-1}P_{j,k,h}^{j,l,r-1}\{l\neq k\} - (2\alpha_j + \mu_{j,h} + \mu_{j,r} + \gamma_h\{h\neq 11\} + \gamma_r\{r\neq 11\} + \tau_h\{h\neq 3\} + \tau_r\{r\neq 3\} + \sigma_{j,k})P_{j,k,h}^{j,k,r}$$

for h = 3, 5, 7, 9, 11, and r = 3, 5, 7, 9, 11. (41)

$$\frac{dP_{j,k,h}^{m,n,r}}{dt} = \psi_{j,k,h}^{rep,m,n,r}\rho_{j,k}^{rep}Y_{j,k,h} + \alpha_{j-1}P_{j-1,k,h}^{m,n,r}\{j=2\} + \alpha_{m-1}P_{m-1,n,r}^{j,k,h}\{m=2\} + \gamma_{h-2}P_{j,k,h-2}^{m,n,r}\{h\neq 3\} + \gamma_{r-2}P_{j,k,h}^{m,n,r-2}\{r\neq 3\} + (1-s_k)\phi_{j,k,h-1}P_{j,k,h-1}^{m,n,r} + s_l\phi_{j,l,h-1}P_{j,l,h-1}^{m,n,r}\{l\neq k\} + (1-s_n)\phi_{m,n,r-1}P_{j,k,h}^{m,n,r-1} + s_u\phi_{m,u,r-1}P_{j,k,h}^{m,u,r-1}\{u\neq n\} - (\alpha_j + \alpha_m + \mu_{j,h} + \mu_{m,r} + \gamma_h\{h\neq 11\} + \gamma_r\{r\neq 11\} + \tau_h\{h\neq 3\} + \tau_r\{r\neq 3\} + \frac{\sigma_{j,k}+\sigma_{m,n}}{2})P_{j,k,h}^{m,n,r}$$

for h = 3, 5, 7, 9, 11, and r = 3, 5, 7, 9, 11. (42)

Diagnosed – On treatment

$$\frac{dP_{j,k,h}^{j,k,12}}{dt} = \psi_{j,k,h}^{rep,j,k,12}\rho_{j,k}^{rep}Y_{j,k,h} + 2\alpha_{j-1}P_{j-1,k,h}^{j,k,12}\{j=2\} + \gamma_{h-2}P_{j,k,h-2}^{j,k,12}\{h\neq 3\}$$
$$+(1-s_k)\phi_{j,k,h-1}P_{j,k,h-1}^{j,k,12} + s_l\phi_{j,l,h-1}P_{j,l,h-1}^{j,k,12}\{l\neq k\} + \sum_{g\in s}\tau_g P_{j,k,h}^{j,k,g} \qquad \text{for } h = 3, 5, 7, 9, 11, \text{ and } s = 5, 7, 9, 11. \qquad (43)$$
$$-(2\alpha_j + \mu_{j,h} + \mu_{j,12} + \gamma_h\{h\neq 11\} + \tau_h\{h\neq 3\} + \sigma_{j,k})P_{j,k,h}^{j,k,12}$$

$$\frac{dP_{j,k,h}^{m,n,12}}{dt} = \psi_{j,k,h}^{rep,m,n,12}\rho_{j,k}^{rep}Y_{j,k,h} + \alpha_{j-1}P_{j-1,k,h}^{m,n,12}\{j=2\} + \alpha_{m-1}P_{m-1,n,12}^{j,k,h}\{m=2\} + \gamma_{h-2}P_{j,k,h-2}^{m,n,12}\{h\neq 3\}$$
$$+(1-s_k)\phi_{j,k,h-1}P_{j,k,h-1}^{m,n,12} + s_l\phi_{j,l,h-1}P_{j,l,h-1}^{m,n,12}\{l\neq k\} + \sum_{g\in s}\tau_g P_{j,k,h}^{m,n,g} \qquad \text{for } h = 3, 5, 7, 9, 11, \text{ and } s = 5, 7, 9, 11. \qquad (44)$$
$$-(\alpha_j + \alpha_m + \mu_{j,h} + \mu_{m,12} + \gamma_h\{h\neq 11\} + \tau_h\{h\neq 3\} + \frac{\sigma_{j,k}+\sigma_{m,n}}{2})P_{j,k,h}^{m,n,12}$$

On treatment – On treatment

$$\frac{dP_{j,k,12}^{j,k,12}}{dt} = \psi_{j,k,12}^{rep,j,k,12}\rho_{j,k}^{rep}Y_{j,k,12} + 2\alpha_{j-1}P_{j-1,k,12}^{j,k,12}\{j=2\} + 2\sum_{g\in s}\tau_g P_{j,k,12}^{j,k,g} - (2\alpha_j + 2\mu_{j,12} + \sigma_{j,k})P_{j,k,12}^{j,k,12} \qquad \text{for } s = 5, 7, 9, 11. \qquad (45)$$

$$\frac{dP_{j,k,12}^{m,n,12}}{dt} = \psi_{j,k,12}^{rep,m,n,12}\rho_{j,k}^{rep}Y_{j,k,12} + \alpha_{j-1}P_{j-1,k,12}^{m,n,12}\{j=2\} + \alpha_{m-1}P_{m-1,n,12}^{j,k,12}\{m=2\} + \sum_{g\in s}\tau_g P_{j,k,12}^{m,n,g} + \sum_{g\in s}\tau_g P_{m,n,12}^{j,k,g} \qquad \text{for } s = 5, 7,9, 11. \qquad (46)$$
$$-(\alpha_j + \alpha_m + \mu_{j,12} + \mu_{m,12} + \frac{\sigma_{j,k}+\sigma_{m,n}}{2})P_{j,k,12}^{m,n,12}$$

In the next sections we describe the method and computational formula used for calculating the model parameters.

1.4.2 Calculating ageing

For each time step, individuals progress from the first to the second age group, and are removed out of the model from the second age group according to the proportion of MSM at the maximum age to all MSM in each age group. The derivation of the proportion is described in Section 3.1.2. Note that per-year rates must be converted to per-time step before being used as model input.

1.4.3 Calculating mortality

The death rate, $\mu_{j,h}$, stratified by age group and HIV stage is applied directly to the numbers of MSM in each corresponding group (see Section 3.1.4 for details of mortality rate). The number of deaths is subtracted from the model every time step and used for calculating the influx of new MSM.

1.4.4 Calculating influx of new MSM

A set of new MSM enters the model every time step and becomes a part of age group 1. This influx is calculated from the summation of the number of individuals who age from age group 1, those who die while in age group 1, and growth per time step of age group 1. The formula used is as follows:

$$\nu_{1,k} = (\alpha_1 + \mu_1 + g_1)N_{1,k} \tag{47}$$

where $N_{1,k}$ is the total number of MSM in age group 1 and sexual activity group *k* and g_1 is the growth rate of age group 1. See Section 3.1.3 for derivation of the population growth rate.

1.4.5 Calculating transition from the current to past MSM

We modelled a transition from the current to past MSM based on the assumption that the ratio between past and current MSM remains fixed over time (See Section 3.7 for more detail about finding the ratio). Every time step the ratio $\frac{Z_{j,h}}{Y_{j,h}}$ is updated. If any updated ratios are less than that of the initial time step, a number of current MSM from age group *j* and HIV stage *h* will be moved to the same group of past MSM to preserve the original ratio. According to the definition of the past MSM (Section 1.3.1), only current MSM of low-activity group (See Section 3.2.1 for the definition) are legitimate for such movement. The numbers of past and current MSM of any groups remain unchanged if the updated ratio equals or exceeds that of the initial one. Our model assumed no transition from past to current MSM.

1.4.6 Calculating disease progression

HIV-positive MSM progress from one disease stage to another according to declined CD4 cell counts (Figure 1 in the main text). Multiplying the progression rate, γ_h, (Section 3.6) by the number of individuals in each disease stage ranging from PHI to CD4 200-349 cells/μL yields the number of MSM approaching the more advanced stages of HIV. Those with CD4<200 cells/μL remain in the last stage until they were either removed from the model by ageing or death, or get HIV diagnosed.

1.4.7 Calculating new HIV diagnoses

During the model year 2000–2009, the rates that HIV-positive MSM get diagnosed with HIV are updated every time step with an increase in diagnosis rate. We derived

changes in HIV diagnosis rate (Section 3.4) and then updated the time-dependent HIV diagnosis rates by:

$$\phi_{j,k,h}(t) = \phi_{j,k,h}(t-1) + \Delta\phi_{j,k,h}(t) \quad (48)$$

where $\Delta\phi_{j,k,h}(t)$ represents change in diagnosis rate from time step *t*-1 to *t*. MSM newly diagnosed with HIV each time step move from undiagnosed CD4 stages to the corresponding diagnosed stages. We also compared model outputs to the reported numbers of new diagnoses by CD4 to validate the model (Section 5).

The model also allows current MSM to change their sexual activity levels due to HIV diagnosis. The constant proportions of individuals who switch from high- to low-activity group and vice versa, $s_{j,k}$ (see Section 3.4 for parameter values), were applied directly and instantly to the number of new diagnoses in each time step to obtain the number of MSM moving between sexual activity levels. Switching between sexual activity levels was permanent and thus individuals were not able to move back to the previous sexual activity levels.

1.4.8 Calculating MSM getting ART- treated

The diagnosed MSM in any CD4 stages except the PHI are given ART treatment every time step at the different rates, τ_h, with men in more advanced stages being more likely to be treated with ART. MSM new to treatment moved to a single compartment of 'on treatment' stage that no longer stratified by CD4 cell counts. We believed that having no CD4 stratification would have only a minor effect on the model outcome since the transmission probability of ART-treated MSM is very low, particularly in developed countries. This also helped simplify our model that already

consists of a considerable number of heterogeneities. We allowed the rates of ART initiation to change according to the new guidelines for HIV treatment in the UK in 2008 [8]. The new rates were used once the model ran into the year 2009 and remain unchanged until the end of simulation. Section 3.5 summarises the derivation of ART initiating rates.

1.4.9 Calculating formation and dissolution of repeat sexual partnerships

The chance of an individual in any of the classes and stages to choose a repeat sexual partner from the same or another group is calculated based on two main components: the repeat sexual partnership formation rate and the sexual mixing preference. We adopted the odds ratio approach suggested by Goodreau and Golden [9] due to its flexibility in modelling multidimensional mixing preferences. The method also ensures that, at any modelling time steps, the number of new repeat sexual partners that MSM of group *m* acquire from group *n* is identical to that of MSM of group *n* acquire from group *m* [10].

The mixing preferences by age group, sexual activity level, and the perceived HIV serostatus are included in the model. Each type consists of 2 subgroups: age group consists of the age group 1 and 2, sexual activity level consists of the low-activity and high-activity group, and the perceived HIV serostatus consists of the groups of perceived HIV-negative (stage S, and U1 to U5) and diagnosed HIV-positive (stage D1 to D5, and T). Consequently, the three odds ratios based on 2x2 mixing matrix for each mixing preference are required. The method for deriving the odds ratio can be found in Section 3.2.7. For simplicity, each mixing preference was modelled

independently so that when all other mixing preferences are discarded, the chance of MSM selecting a new repeat sexual partner from the same or different groups matches exactly the original odds ratio of the remaining mixing preference.

At each time step, we began the calculations by finding the number of MSM who are looking for a new repeat sexual partner and stratified independently by age group, sexual activity level, and the perceived HIV serostatus. These numbers were then allocated between groups and created the following mixing matrix for each mixing preference:

	Group 1	Group 2
Group 1	*a*	*b*
Group 2	*c*	*d*

The matrix shows the number of MSM in pairs between the two groups. If the odd ratios *ad*/*bc* is greater than one the mixing is assortative, i.e. an individual is more likely to select a partner from their own group rather than from the other groups. Odds ratio of less than one represents the disassortative mixing, i.e. an individual is more likely to select a partner from the other groups rather than their own group. Odds ratio of one represents proportionate mixing, i.e. an individual selects a partner based on the proportion of group 1 and 2 to all available men. The elements of the mixing matrix are calculated by:

$$a = \rho_1^{rep} Y_1 - b \tag{49}$$

$$b = \frac{-\frac{1}{2}\rho_1^{rep} Y_1 - \frac{1}{2}\rho_2^{rep} Y_2 + \sqrt{\frac{1}{4}\left(\rho_1^{rep} Y_1 + \rho_2^{rep} Y_2\right)^2 + \rho_1^{rep} Y_1 \rho_2^{rep} Y_2 (w-1)}}{w-1}, b = c \tag{50}$$

$$d = \rho_2^{rep} Y_2 - b \tag{51}$$

where ρ_i^{rep} is the average repeat sexual partner formation rate of group *i*, Y_i is the number of current MSM in group *i*, and *w* is the odds ratio of the mixing preference.

We then combined any two of the three mixing preferences based on the assumption that the odds ratio of any type in any subgroup is identical to the same odds ratio in the overall population. In practice, we assigned the one mixing matrix to all elements of another mixing matrix and, within each of those elements of the latter matrix, calculated the mixing by the former matrix. For example, if mixing matrices by age group and sexual activity are:

Age	Group 1	Group 2
Group 1	*a* = A1	*b* = A2
Group 2	*c* = A3	*d* = A4

Activity	Group 1	Group 2
Group 1	*a* = B1	*b* = B2
Group 2	*c* = B3	*d* = B4

the sexual activity mixing matrix in an element *a* of age mixing matrix is:

	Group 1	Group 2
Group 1	*k*×B1	*k*×B2
Group 2	*k*×B3	*k*×B4

where $k = \frac{\mathrm{B1}+\mathrm{B2}+\mathrm{B3}+\mathrm{B4}}{\mathrm{A1}}$. The resulting odds ratio of this sub-matrix equals the overall age-mixing odds ratio. The calculations for other elements and for combining all three mixing preferences can be achieved using the same logic. Note the order of combining the three mixing preferences has no effect on the final mixing matrix.

The combined 8x8 mixing matrix contains the numbers of MSM in repeat sexual partnerships between MSM of 8 subgroups stratified simultaneously by two age groups, two sexual activity levels, and two perceived HIV serostatuses. Comparing the

partnership formation rates derived from the combined mixing matrix to the original parameter values revealed some inconsistencies between the two sources which necessitated further adjustments. We therefore adjusted the numbers of men required for forming repeat sexual partnerships from the eight subgroups by multiplying the mixing elements that contribute the total number of required MSM in each subgroup with the ratio between original (data-derived) and actual repeat sexual partnership formation rates. The mixing elements are adjusted using the following formula:

$$a_{j,k,p}^{m,n,q} = a'^{m,n,q}_{j,k,p} \frac{\rho_{j,k}^{rep}}{\rho'^{rep}_{j,k}} \tag{52}$$

where $a'^{m,n,q}_{j,k,p}$ is a mixing element before adjustment between men of age group *j*, sexual activity group *k*, and perceived HIV serostatus group *p* and men of the corresponding groups *m*, *n*, and *q*, respectively, $\rho_{j,k}^{rep}$ is the original (data-derived) repeat sexual partnership formation rate for men in age group *j* and sexual activity group *k*, and $\rho'^{rep}_{j,k}$ is the partnership formation rate that derived from the combined mixing matrix. In the above equation we can see that all the non-diagonal elements ($[j,k,p] \neq [m,n,q]$) contribute concurrently to two mixing subgroups; we decided to adjust the subgroup that has a larger deviation of the partnership formation rates, $\left| \frac{\rho_{j,k}^{rep}}{\rho'^{rep}_{j,k}} - 1 \right|$, at each time step. After adjustment, the repeat sexual partnership formation rates of the subgroup that has the largest deviation would be identical to that of the original rate. The deviations in other subgroups may still remain but considerably smaller than without the adjustment. In exchange for more precise partnership formation rates, all types of odds ratio in the overall populations began to shift from the original values. However, the benefit of adjustment is clear since an improvement

in accuracy of partnership formation rates after adjustment is substantial while only a minor difference in the odds ratio can be found. This is because the adjustment was aimed at subgroups with large deviations which usually are of small size and, in turn, having less effect on the odds ratio.

The final mixing matrix can now be used to obtain $\psi_{j,k,h}^{rep,m,n,r}\rho_{j,k}^{rep}Y_{j,k,h}$. Since we assumed proportionate mixing by disease stage, the numbers of repeat sexual partnerships were allocated proportionally to all pairs within each of the eight subgroups. If proportionate mixing was assumed for all mixing preferences, then $\psi_{j,k,h}^{rep,m,n,r} = \frac{Y_{m,n,r}\rho_{m,n}^{rep}}{N^{rep}}$ where $N^{rep} = \sum_{j=1}^{2}\sum_{k=1}^{2}\sum_{h=1}^{12} Y_{j,k,h}\rho_{j,k}^{rep}$ is the total number of MSM acquiring new repeat sexual partners at each time step. In order for $P_{j,k,h}^{m,n,r} = P_{m,n,r}^{j,k,h}$ to remain valid at all times, the condition $\psi_{j,k,h}^{rep,m,n,r}\rho_{j,k}^{rep}Y_{j,k,h} = \psi_{m,n,r}^{rep,j,k,h}\rho_{m,n}^{rep}Y_{m,n,r}$ must always be satisfied. This can be achieved by using the adjusted repeat sexual partnership formation rates instead of one originally derived from the data in all main model equations (1) to (6) and (13) to (46).

The breakup between repeat sexual partners occurs every time step at the rate $\sigma_{j,k}$. The two members of a separated pair move to the single state and are ready for a new repeat sexual partnership. We included an average gap period to the dissolution rate to reflect the fact that, after the end of a long relationship, individual stays single for some period of time before acquiring a new partner.

1.4.10 Calculating force of infection in repeat sexual partnerships

In the partnership-based HIV transmission model, the force of infection for a repeat sexual partnership consists of two main components: the transmission probability and the frequency of sexual acts. The HIV transmission probabilities for each of the disease stages were calculated from a function of the average viral load in that stage (Section 3.3). Combining with all related factors, the per-act HIV transmission probability from an HIV-positive man of age group *m* (*m*=1,2), sexual activity *n* (*n*=1,2) and disease stage *r*, (*r*=2,...,12) to a susceptible of age group *j* (*j*=1,2), and sexual activity *k* (*k*=1,2) was calculated by:

$$\begin{aligned}\beta_{j,k}^{ste,m,n,r} = p_{anal,j,k}^{rep}[p_{uai,j,k,r}^{rep}(p_{ins}\beta_{uiai,r}) + (1-p_{ins})\beta_{urai,r} + \varepsilon_{cdm}(1-p_{uai,j,k,r}^{rep})(p_{ins}\beta_{uiai,r}) \\ + (1-p_{ins})\beta_{urai,r})] + (1-p_{anal,j,k}^{rep})[p_{uroi,j,k}^{rep}\beta_{uroi,r}]\end{aligned} \quad (53)$$

The notations are summarised in Table 1 and 2 in the main text and Table S9.

The force of infection per time step is the product of the combined HIV per-act transmission probability and the average frequency of sexual acts per time step between MSM of class *j* and *k*, $\varphi_{j,k}$:

$$\lambda_{j,k}^{rep,m,n,r} = \varphi_{j,k}\beta_{j,k}^{rep,m,n,r} \quad (54)$$

The force of infection was calculated for all pairs between HIV-positive and HIV-negative MSM. Newly HIV-infected MSM at each time step move from HIV stage S to U1 while retain the current repeat sexual partnership.

1.4.11 Calculating formation and force of infection in one-off sexual partnerships

We used the same method as we did for the repeat sexual partnership formation to calculate the formation of one-off sexual partnership. Both single and paired MSM

acquired a new one-off sexual partner at different rates. For paired MSM, only high-activity group members can form a one-off sexual partnership while having a repeat sexual partner. Only mixing by age group and sexual activity were incorporated. The perceived HIV serostatus was accounted for differently by subsequently dividing MSM into subgroups according to unprotected sexual intercourse (UAI) with one-off sexual partners. Within each perceived HIV serostatus, we assumed proportionate mixing by HIV stage.

The mixing matrices and one-off sexual partnership formation rates were adjusted in the same manner as for the formation of repeat sexual partnership. From the final mixing matrix we derived a chance of selecting a one-off sexual partner from different age and sexual activity groups, $\psi_{j,k}^{one,m,n,r}$. This was subsequently used to formulate the force of infection in one-off sexual partnerships. We assumed that the partnerships are formed and separated instantaneously, and hence a dissolution rate is not required.

To derive the force of infection within one-off sexual partnerships, we divided susceptibles into three groups: (A) no UAI at all, (B) only have UAI with perceived HIV-negative partners, and (C) can (but not only) have UAI with diagnosed HIV-positive partners. This allows us to model serosorting in one-off sexual partnership more comprehensively. The members of group A perform only safe anal sex regardless of the perception they have of their one-off sexual partner's serostatus. Group B will only have UAI with one-off sexual partners they believe are HIV negative (serosorting). However, it is not necessary that men of group B will always have UAI with an HIV-negative one-off sexual partner. Group C can have UAI with

one-off sexual partners of any HIV serostatuses. The expression for the combined HIV transmission probability in one-off sexual partnerships that accounted for all UAI groups is given by:

$$\beta_{G,j,k}^{one,all} = p_{A,G}^{one}\beta_{G,A,j,k}^{one} + p_{B,G}^{one}\beta_{G,B,j,k}^{one} + p_{C,G}^{one}\beta_{G,C,j,k}^{one} \tag{55}$$

where $p_{A,G}^{one}$, $p_{B,G}^{one}$, and $p_{C,G}^{one}$ is the proportion of men in group A, B and C, respectively, and $G = Y, P$ denotes current single and paired MSM. The HIV transmission probabilities for each UAI group are:

$$\beta_{G,A,j,k}^{one} = \varepsilon_{cdm}\sum_{m=1}^{2}\sum_{n=1}^{2}\sum_{r=2}^{12}\psi_{j,k}^{one,m,n,r}\beta_{G,j,k}^{one} \tag{56}$$

$$\begin{aligned}\beta_{G,B,j,k}^{one} = {} & (\varepsilon_{cdm}(1-p_{uai,G,B,j,k}^{one})\sum_{m=1}^{2}\sum_{n=1}^{2}\sum_{r=2}^{12}\psi_{j,k}^{one,m,n,r}\beta_{G,j,k}^{one}) \\ & + (p_{uai,G,B,j,k}^{one}\sum_{m=1}^{2}\sum_{n=1}^{2}\sum_{r\in d}\psi_{j,k}^{one,m,n,r}\beta_{G,j,k}^{one})\end{aligned} \quad \text{where } d = 1,2,4,6,8,10 \tag{57}$$

$$\beta_{G,C,j,k}^{one} = (\varepsilon_{cdm}(1-p_{uai,G,C,j,k}^{one}) + p_{uai,G,C,j,k}^{one})\sum_{m=1}^{2}\sum_{n=1}^{2}\sum_{r=2}^{12}\psi_{j,k}^{one,m,n,r}\beta_{G,j,k}^{one} \tag{58}$$

The common term of transmission probability shared by all three groups is given by the expression:

$$\begin{aligned}\beta_{G,j,k}^{one} = {} & p_{anal,G,j,k}^{one}[p_{ins}\beta_{uiai,r}(p_{cir,j}\varepsilon_{cir} + (1-p_{cir,j})) + (1-p_{ins})\beta_{urai,r}] \\ & + (1-p_{anal,G,j,k}^{one})[p_{uroi,j,k}^{one}\beta_{uroi,r}]\end{aligned} \tag{59}$$

The force of infection per time step is the product of one-off sexual partnership formation rates and the combined HIV transmission probabilities:

$$\lambda_{G,j,k}^{one} = \rho_{G,j,k}^{one}\beta_{G,j,k}^{one,all} \tag{60}$$

The notations used here are summarised in Table 1 in the main text.

2. Data

The data from multiple behavioural surveys and HIV surveillance systems that included MSM in the UK were used in conjunction with information obtained from the literature to estimate model parameters. We adjusted and categorised data according to the MSM subgroups in the model before further analyses.

2.1 Data sources

The primary data sources for deriving model parameters are as follows.

2.1.1 The National Survey of Sexual Attitudes and Lifestyles (NATSAL)

NATSAL is a major survey of sexual attitudes and lifestyles providing detailed information about sexual and related behaviour patterns of the general UK populations aged 16-44 including MSM. This survey is conducted at ten-year intervals. The first round was initiated in 1990 [11], the second in 2000–2001 [3], and the third round in 2010–2012, the results of which will be published by 2013 [12]. The data from the 2000 survey was our primary data.

2.1.2 The Gay Men's Sexual Health Survey (GMSHS)

Initiated in 1996, GMSHS is a survey aimed primarily at gathering information on demographics, sexual behaviour, and HIV status among gay men aged 16 or above who attend community venues in London [13]. Since 2000, survey respondents had also been offered an anonymous HIV test using the OraSure device. This allows linking respondent's serostatus to the data collected from the questionnaire. We used

the data for years 2000–2008 except 2007 when the survey was not conducted. University College London, the Public Health England (PHE, previously known as the Health Protection Agency), and the Medical Research Council are responsible for the survey.

2.1.3 The London Gym Survey (GYM)

GYM is a survey on demographics, social characteristics, sexual behaviour, HIV status, and risk factors among gay men who use gyms in London [14]. With collaboration between City University London and PHE, the survey was conducted annually between 1998 and 2005 and again in 2008. We used all data that are available from 2000 to 2008.

2.1.4 The HIV/AIDS diagnoses and deaths surveillance

This is the national HIV/AIDS surveillance reporting systems established in 1982. The data have been reported on a regular basis including the number of new HIV diagnoses, laboratory tests, probable routes of infection, demographics, and epidemiological data [15]. The data are made available by PHE.

2.1.5 The CD4 surveillance systems

As a supplementary to the HIV/AIDS surveillance, the CD4 systems monitor trends in CD4 cell count at HIV diagnosis among HIV-infected adults. The CD4 cell count at HIV diagnosis is defined as the CD4 cell count closest to, and within 30 days of, the date of HIV diagnosis. The data are reported as a supplement to HIV diagnoses [16].

2.1.6 The Survey of Prevalent HIV Infections Diagnosed (SOPHID)

SOPHID is a cross-sectional survey of all individuals with diagnosed HIV infection who require HIV-related care within the National Health Service in England, Wales, and Northern Ireland within a calendar year run by the PHE. Scottish data are collected separately by Health Protection Scotland, and are included in the final UK totals. The survey began in 1995 and is conducted twice a year in London and annually outside London. The primary aim is to determine the total number of HIV prevalent patients seen for treatment and care [16].

2.2 Selecting survey respondents

This process involved selecting and categorising survey respondents to match the model structure. We began by selecting NATSAL respondents according to the definition of current and past MSM outlined earlier. Weights for the core sample and the ethnic minority boost sample were applied for all calculations. For GMSHS and GYM, there is no exact same set of questions as in NATSAL that can be used to categorise current and past MSM. We therefore assumed that GMSHS and GYM respondents were current MSM because the majority reported anal sex with men in the last year (GMSHS: 90%, GYM: 89%).

The survey respondents aged 65 or above or with missing age data were excluded. The remaining was divided into two age groups: 16–34 and 35–64. Most of the subsequent analyses were based on the two age groups except when the size of any groups was too small that we would use the overall MSM. Table S2 shows the numbers of sampled MSM from the surveys.

For data from HIV/AIDS surveillance systems including CD4 and SOPHID databases, we selected only men aged 15-64 who were reported to acquire HIV infection from sex between men. The large number of cases allowed us to divide the datasets by age groups for all data analyses.

2.3 Data adjustment

In our model, the population of interest is the entire MSM population aged 15-64 in the UK. However, the data from GMSHS and GYM were based on convenience-sampling from the MSM community in London. This may introduce bias into the data by including MSM with characteristics that are markedly different from the general MSM in the UK. Such differences can be observed by comparing the GMSHS and GYM data to the national-based NATSAL 2000 survey that used probability-sampling on a household basis. Table S3 shows, for example, that the mean number of male sexual partners in the last year of GMSHS men aged less than 35 were around 22 compared to that of 7 from NATSAL. Adjustment was necessary to improve the representativeness of survey populations. The adjustment method suggested by Reidy and Goodreau [17] was used here.

To adjust the GMSHS and GYM data, we selected only the surveys that were conducted during 1999–2001 which corresponded to the data collection timeframe of NATSAL. The data were then adjusted according to four variables: (1) age of respondent, (2) the number of male sexual partners in the last year, (3) the number of unprotected anal intercourse male partners in the last year, and (4) time of last HIV

test. We calculated a probability weight for each GMSHS and GYM respondent independently by each of the four variables. The weight was calculated as:

$$w_{i,a} = \frac{p_{i,a}}{k_{i,a}} \tag{61}$$

where $w_{i,a}$ is the weight for respondents that fall into group *i* of the adjusted variable *a*, and $p_{i,a}$ and $k_{i,a}$ are the proportions from the reference dataset (NATSAL) and the adjusted dataset (GMSHS or GYM), respectively.

By applying the weights independently to each variable, the distribution of adjusted proportions of GMSHS and GYM respondents for the variable will be identical to that of NATSAL. Applying all weights simultaneously somewhat distorted the distribution of all variables (Table S3). The problem can be avoided by replacing the marginal distribution of each variable in equation (61) with the joint distribution of all four variables. However, the small sample size of MSM in NATSAL prevented us from properly deriving such joint distributions and thus the marginal distributions were used instead.

The weights were then applied to GMSHS and GYM data of all years that were used in this study. Consequently, all model parameter values derived from these datasets were affected by the adjustment. The total weights range from 0.014 to 18.558 for GMSHS and from 0.187 to 11.984 for GYM data (Table S3). We conducted an analysis to see the effects of introducing weight limits on the derived repeat sexual partnership formation rates (Figure S1). To maintain the accuracy of the derived model parameters, the full weights were used.

3. Parameter estimation

The model parameters can be stratified into five categories: demographics, sexual behaviour parameters, biological parameters, HIV diagnosis parameters, and model initial conditions. All parameters were initially derived using either the data from surveys, surveillance systems, reports, or the literature. In addition, we described the derivation of lower and upper bounds for parameters that were also estimated through fitting the model to reported data.

3.1 Demographics

According to the demographic structure of the modelled population, we required estimates of the UK MSM population size, rate of ageing, population growth rate, and mortality rate by age groups. Table S6 summarises all derived demographic parameter values.

3.1.1 MSM population size

The data from the UK Office of National Statistics (ONS) and NATSAL were used to obtain estimates of MSM population size in the UK in 2000, the initial time point in the model. First, we calculated the proportions of MSM to all male respondents in NATSAL by regions and found that London had the highest proportion of MSM (6.3%) compared to elsewhere in England (2.5%), and outside England (1.7%). Since the age of the modelled populations (15-64) and NATSAL respondents (16-44) did not match, we derived the proportion of UK male population aged 45-64 to aged 40-44 from the 1999–2001 ONS data [18] and assumed the same proportion for UK

MSM. For MSM aged 15, we simply assumed that the proportion of MSM in males aged 15-34 was the same as that of aged 16-34. The proportions of MSM by age groups and regions can now be obtained. The analysis also provided an estimated 5-year-interval age distribution of MSM aged 15-64 (Table S4) which will be used later for calculating the mortality rates (section 3.1.4).

Combining all these estimated proportions provided the proportion of MSM aged 15-64 to the male population with the overall mean of 2.9% regardless of age groups and regions. However, the PHE estimated that 3.4% of all UK males are MSM [19]. We therefore adjusted the age group- and region-specific proportions of MSM derived above so that the total number of MSM accounted for 3.4% (not 2.9%) of male population in the UK. The adjusted proportions were then multiplied by the ONS mid-year population estimates and yielded 648,500 MSM (aged 15-34: 259,500, aged 35-64: 389,000) living in the UK in 2000. Of these and according to the proportions of current and past MSM derived from NATSAL, there were 527,800 current MSM (aged 15-34: 242,678, aged 35-64: 285,122) and 120,700 past MSM (aged 15-34: 16,822, aged 35-64: 103,878). The data used for estimating the MSM size are summarised in Table S5 and the derived parameter values are shown in Table S6. The proportion of past MSM was also used to model the flow from current to past MSM (Table S6).

3.1.2 Rate of ageing

The annual ageing rate equals the proportion of individuals at the maximum age in each age group. Since the age distribution of UK MSM populations was not available,

we assumed that of the UK male population instead. The 2000–2020 estimated proportions of men aged 34 and 64 to all men aged 15-34 and 35-64, respectively, were obtained from the ONS projections [20]. The average proportions over the entire simulation period were 5.23% and 2.61% for younger and older groups (Table S6). These proportions were held constant at all time throughout model simulations.

3.1.3 Population growth rate

We assumed that size of the two age groups changes over time at the same rate as the general UK male populations of the same age group. The growth rate was derived by averaging the annual growth rates of men over 2000–2020. The average rate was estimated at 0.51% per year for all MSM populations (Table S6). The estimated growth rate of the younger age group (0.43%) was used for calculating the influx into the model of new MSM, while the rate of the older age group (0.56%) was used for validating the size of the modelled populations over time (see Section 5 for details).

3.1.4 Mortality rate

The mortality rates for HIV-negative MSM were derived based on the mortality rates of the overall UK males. Age-specific data were obtained from the historic interim life tables for three periods (1999–2001, 2002–2004, and 2005–2007) reported by ONS [21]. We averaged the data over the three periods and converted into 5-year age-specific mortality rates to match the previously defined age distribution of MSM (Table S4). By multiplying the age distribution with the mortality rates of the corresponding age group, we had the rate by 5-year age band which were then collapsed into the two age groups using an arithmetic mean (Table S6).

For HIV-positive men, a study by UK Collaborative HIV Cohort [22] estimated that, depending on various factors, the mortality rates could increase from less than 1.1 to nearly 10 times due to HIV compared to the UK male population. The effect is more pronounced for younger age groups (age 20-44) compared to the overall. Based on this information, we arbitrarily assumed that the death rates in our first and second age group increased by 9 and 2 times, respectively, from the mortality rates of HIV-negative MSM (Table S6). We applied HIV death rates only to ART-treated MSM in the model since the majority of deaths in HIV-positive MSM in the UK occurred after they have been diagnosed and treated with ART in late disease stages [19, 23].

3.2 Sexual behaviour

The three main sources of data for deriving sexual behaviour parameters were NATASL, GMSHS, and GYM survey. Although the national-based probability-sampling NATSAL survey was our primary source, we were unable to derive all parameters based on this survey alone due to the small sample size of MSM and lack of some required information. GMSHS and GYM were used when needed. Most sexual behaviour parameters were stratified by age and sexual activity levels simultaneously unless stated otherwise. Uncertainty ranges of parameters that later fitted to the data were also derived.

3.2.1 Sexual activity level

Before deriving sexual behavioural parameters, we stratified survey respondents into two risk groups—low and high—according to the reported number of male sexual partners in the last year. We defined risk groups based on three conditions. First, the

low-activity MSM have a similar rate of partner change to the average of UK heterosexual male [3] which helped distinguish MSM with normal risk from the whole MSM population. Second, MSM in the low-activity group have no more than one sexual partner at a time. This was to ensure that a low-activity individual can never get infected from having a concurrent partnership. Third, the proportion of any risk groups must be greater than 25% because modelling sexual mixing between two groups that differ greatly in size could be problematic. The number of individuals in the smaller group may be insufficient to satisfy the demand for sexual partners from within and outside the group. Consequently, we chose one male sexual partner in the last year as a cut-off point: low-activity MSM reported one or no male partner in the last year, high-activity MSM reported two or more male partners in the last year. Observations with missing values for the variable were removed from the stratification, and hence excluded from any further calculations that must be stratified by sexual activity level. Information on the number of male sexual partners in the last year was not available from GYM. Instead, we used the reported number of male anal-sexual partners in the past 12 months and stratified GYM participants using the same cut-off point. The numbers of MSM by sexual activity level are shown in Table S7.

3.2.2 Rate of partnership formation

The rate of partnership formation was calculated from the reported number of male sexual partners in the last year categorising into repeat sexual and one-off sexual partners. GMSHS provided information that matched our previously defined definitions of partnership type (Section 1.3.4) and was hence used. We estimated that mean repeat sexual partnership formation rate of high-activity MSM is approximately

seven times higher than that of the low-activity (Table 1 in the main text). For one-off sexual partners, respondents were also stratified into those who are currently having a repeat sexual partner or not by assuming that individuals having concurrent relationships are those who reported both steady and casual partners in the last year. According to our definition of risk group, only high-activity MSM can have a one-off sexual partner while in a relationship with a repeat sexual partner. Both repeat sexual and one-off sexual partnership formation rates were included in the model fitting with the ranges derived from the 95% confidence interval of the GMSHS data. All derived rates are summarised in Table 1 in the main text.

3.2.3 Rate of repeat sexual partnership dissolution

We began by finding the gap duration between two consecutive non-concurrent repeat sexual partnerships. NATSAL provided date (in month) of the first and last sex acts for the three most recent partners. Gap length was defined as time between the first sex act with the most or second most recent partner and the last sex act with the second or third most recent partner, respectively. A gap length of 15 days was assumed for those who reported the same month for the end of last relationship and the start of new relationship. We excluded any concurrent relationships from the calculation and derived the median gap duration by sexual activity level only due to the small number of cases. The results suggested that, after pair separation, low-activity MSM tend to stay single for a much longer period than the high-activity MSM before acquiring a new repeat sexual partner (Table 1 in the main text).

With the above gap lengths, we calculated the mean duration of repeat sexual partnership as follows:

$$\sigma = \frac{1}{\rho} - G \tag{62}$$

where ρ is the repeat sexual partnership formation rate and G is the gap length. The rate of repeat sexual partnership dissolution was then derived from the inverse of repeat sexual partnership duration. The results are shown in Table 1 in the main text.

3.2.4 Proportion of anal-sex and oral-sex partnerships

In order to model HIV infection via oral and anal sex separately, our model required the proportions of repeat sexual and one-off sexual partnership that involves anal sexual intercourse. This was estimated from GMSHS data by dividing the number of male anal-sexual partners by the total number of male sexual partners in the last year of the same partnership type. Table 1 in the main text shows the point and 95% CI estimates of the mean proportions of anal-sex repeat sexual partnership of around 90% for low-activity and 70% for high-activity MSM. This seems plausible since the high-activity MSM are able to have an anal-sex relationship with a one-off sexual partner while in a repeat sexual partnership.

Based on the assumption that all sexual intercourses in non-anal sex partnerships were oral sex, subtracting the anal-sex proportion from one resulted in the proportion of oral-sex partnerships (Table 1 in the main text). We ignored all oral sexual intercourses that might occur in an anal-sex partnership because the risk of HIV transmission through oral sex is much less than that of anal sex [24].

3.2.5 Unprotected sexual intercourse

There are a number of parameters related to UAI including proportion of UAI acts, proportion of susceptibles who perform UAI acts, and proportion of sexual roles—insertive and receptive act. All must be distinguished between repeat sexual and one-off sexual partnerships. For repeat sexual partnerships, the first two parameters are the proportion of susceptibles who have UAI with a perceived HIV negative repeat sexual partner to all susceptibles who have a perceived HIV negative repeat sexual partner, and the proportion of susceptibles who have UAI with a diagnosed HIV positive repeat sexual partner to all susceptibles who have a diagnosed HIV positive repeat sexual partner. The perceived HIV negative partner included all those partners that have never been diagnosed with HIV. Data were available in the GYM survey. We selected only respondents who reported anal sex in the past 12 months and disclosed the perceived serostatus of their repeat sexual partner. Since this is the proportion of susceptibles only, we excluded those who tested HIV positive. Although some of the selected respondents may have been infected with HIV but were unaware of the infection, this was unlikely to add bias to the data since their sexual behaviour should remain unchanged. The mean estimates of the proportion were derived from dividing the number of respondents who reported UAI with perceived HIV negative regular partners in the past 3 months by the total number of those who had perceived HIV negative regular partners (Table 1 in the main text). The same method was applied to a repeat sexual partner of diagnosed HIV positive status. The inconsistency in the definitions of sexual partner used in the GYM survey (survey-participant-defined regular partner) and our study (repeat sexual partner) may result in an increased uncertainty of the derived parameters. Subsequently, we arbitrarily expanded the derived lower limit of both parameters by 30% due to the fact that UAI should be less

likely in a relationship that is more casual which is a part of the repeat sexual partnership. These two parameters were later assigned the uniform distribution and included in parameterisation by model fitting. Details are shown in Section 4.

We further assumed that, in a pair of repeat sexual partners between susceptibles and undiagnosed MSM with UAI, all sex acts are UAI of which 50% were insertive acts. This was supported by GMSHS and GYM data that suggested that, on average, MSM across age group and sexual activity level performed versatile sexual roles with their regular partner. For a pair between susceptibles and diagnosed MSM, the proportion of UAI acts was reduced by 70% according to the findings from the meta-analysis study of high-activity sexual behaviour in HIV-positive individuals in the United States [25]. The reduction in UAI acts was then added to safe sex to maintain the constant number of acts across all groups and time period. See Table 1 in the main text for summary of the parameter values.

For one-off sexual partnerships, susceptibles were categorised into: A) MSM who had no UAI with a one-off sexual partner, B) MSM who had UAI only with a perceived HIV negative one-off sexual partner, and C) MSM who had UAI with a diagnosed HIV positive one-off sexual partner. The proportions for these three groups were estimated using GMSHS data. We started from dividing perceived HIV negative respondents into those who had or had no UAI with a one-off sexual partner. The number of MSM without UAI was used for calculating the proportion of group A. For those who reported UAI with a one-off sexual partner, if the reported number of perceived HIV negative UAI one-off sexual partners was equal to the total number of

UAI one-off sexual partners, they were included in group B. And group C corresponded to those who had at least one HIV-positive one-off sexual partner. We divided all groups by their sum to obtain estimates of the proportions of men in each group (Table 1 in the main text). Similarly to one-off sexual partnership formation rate, the estimates were calculated separately for men who had and had no repeat sexual partners in the last year.

In group B and C, the proportion of UAI was derived from the mean proportion of reported number of UAI one-off sexual partners to all one-off sexual partners in the last year. For group B, perceived HIV negative partner is from susceptible or undiagnosed HIV-infected stages only. For group C, we assumed chance of choosing one-off sexual partners of any HIV stages is proportional to the number of individuals in each stage.

Similarly to the repeat sexual partnership, we assumed 50% for insertive and receptive act for both safe and unsafe sex with a one-off sexual partner of all classes. Although, studies among MSM in Australia [26] and the United States [27] suggested that in an unprotected anal intercourse with an HIV positive partner, MSM were more likely to perform insertive rather than receptive acts, while receptive acts were performed more frequently with a perceived HIV negative partner, this was not evident in the UK [28]. Given that the sexual act is one-off, assuming the proportion of sexual roles based either on the number of partners or the number of acts makes no differences.

Regarding oral sex, UROI with ejaculation was assumed the only type of oral intercourse that presents a risk of transmitting HIV. We further assumed UROI occurs in all sex acts with an oral-sex repeat sexual partner (Table 1 in the main text). This may seem an extreme case, but its effect is considered minor since the HIV transmission probability of UROI is substantially lower than that of anal sex (Table 2 in the main text). For one-off oral sex, we derived the proportion of UROI from the number of GYM men who reported UROI with a one-off sexual partner (Table 1 in the main text). For simplicity, we modelled HIV spread via one-off oral sex without categorising MSM into three groups based on serostatus of partner as previously did for the anal sex.

3.2.6 Frequency of sexual acts

The mean numbers of sex acts per week with a repeat sexual partner were estimated through the model fitting. We started by obtaining the plausible ranges of the parameter values from the number of occasions of sex with men in the last 4 weeks reported in NATSAL. We selected only the respondents who, in the last 4 weeks, had only one male partner of any types. Due to a small number of eligible cases, the derived estimates could only be stratified by age groups which led to identical ranges for both sexual activity levels in the same age groups. The sampling distribution of the parameter in the fitting process was assumed uniform and the sampled values were allowed to vary independently by both age groups and sexual activity levels. Table 1 in the main text summarises the derived estimates.

3.2.7 Sexual mixing preference

In our model, mixing preferences were stratified by age group, sexual activity level, and perceived HIV serostatus. All are modelled using the odds ratio in a 2-by-2 mixing matrix [9, 10].

NATSAL included information on mixing by age. However, the age cut-off point must be adjusted specifically for this calculation due to the small number of cases. We, therefore, used a median age of 28 as a new cut-off point for age stratification. To construct a mixing matrix by age, we divided respondents by their age and the age of their most recent partner on the first sex. For the younger group (under 28 years of age), 72% selected a partner in the same age group, while only 37% of the older group (28 years of age or more) did that. The corresponding odds ratio was 4.2 (Table 1 in the main text) which indicated that MSM are 4.2 times more likely to choose a partner of the same age group than when choosing randomly (a proportionate mixing).

With insufficient data available to inform the mixing preference by sexual activity level of MSM in the UK, we adopted the estimated odds ratio of 3.0 of male heterosexuals in London [29] (Table 1 in the main text).

The perceived HIV serostatus was divided into perceived HIV negative and diagnosed HIV positive status. The perceived negative status included non-infected and undiagnosed HIV-infected MSM. The diagnosed positive status included all MSM with diagnosed HIV. From the reported HIV status of GYM men and their regular partner, we estimated the odds ratio of 2.3 (Table 1 in the main text). Mixing by the

perceived HIV serostatus using odds ratio was only applied to a repeat sexual partner selection. We used a different method to model the effects of one-off sexual partner's serostatus and the proportion of UAI (serosorting). See Section 1.4.11 for more details of the method.

3.3 HIV transmission probability

To estimate the per-sex act probability of HIV transmission among MSM, we categorised sex acts according to the sexual types and roles existing in the model—insertive and receptive anal sex, and receptive oral sex. Each of these was divided into protected and unprotected acts according to condom use. We accounted for changes in HIV transmission due to disease progression on the basis of CD4 cell counts. The very low, but non-zero, HIV transmission probabilities of ART-treated MSM were calculated separately and allowed to change over time to reflect an increase in ART effectiveness on infectiousness reduction and improved drug adherence among HIV-infected patients. The main calculation steps were calculating the relative risks of HIV infection of the viral load of interest to the baseline viral load. The derived relative risks were then applied to the baseline transmission probability obtained from literature review to estimate the infectiousness according to viral load. Finally, the proportion of men by viral load in each CD4 stage was multiplied to the above transmission probability to provide the final per-act HIV transmission probability.

We started by analysing the viral load data from the survey of diagnosed HIV-infected MSM seen for HIV-related care at clinics in the UK [16] from 2005 to 2009. Only individuals who have never been treated with ART were selected and their reported

viral load at that clinic visit was divided into five ranges: 0-399, 400-999, 1000-9999, 10000-49999, and 50000+ copies/mL of blood plasma, and the CD4 counts at the same visit into four ranges: 0-199, 200-349, 350-499, and ≥500 cells/μL. If there were more than one record of viral load in each CD4 stage, only the latest one available was used. The proportions of men by viral load ranges in each CD4 stage were then calculated (Table S8). It is clearly seen that in the lower CD4 stages, the proportion of men with high viral load is markedly higher than in higher CD4 stages. For example, 55% of MSM with CD4<200 cells/μL had viral load of 50000+ copies/mL while there were only 23% among MSM with CD4≥500 cells/μL. We also accounted for uncertainty in these proportions by running 10,000 simulations for each CD4 stage based on multinomial distribution with the number of cases parameter equals to the number of included SOPHID respondents and the event probability parameter equals the average proportions. We then derived a median viral load to represent the five viral load ranges (Table S8). The CD4 stratification was neglected at this point because the minor effects it had on the derived median estimates.

We obtained the baseline transmission probabilities per an URAI act of 1.4% (0.2-2.5) suggested by Baggaley et al based on systematic review and meta-analysis [30]. For an UIAI act, the estimates of 0.62% (0.07-1.68) from the cohort study among homosexual men in Sydney was adopted [31]. We used per an UROI contact risk of 0.04% (0.01-0.017) suggested by Vittinghoff et al [24]. The viral load associated with the above baseline transmission probabilities was assumed 20,000 copies/mL which was derived from the median viral load of the survey participants who have never been on ART.

We then described the relationship between transmission probability and plasma viral load according to the function proposed by Smith & Blower [32], following the original notations, as:

$$\beta(v) = 2.45^{\log_{10}(v/w)}\beta(w) \quad (63)$$

where w is the initial viral load which, in this case, is the viral load of the baseline transmission probability, v is the viral load of interest, $\beta(v)$ and $\beta(w)$ are transmission probability at the corresponding viral load v and w. The lower and upper bound were derived from the corresponding limits of the baseline per-act transmission probability (Table S8). Based on these boundaries, we sampled from the beta distribution and produced 10,000 sets of per-act transmission probability by viral load for each CD4 stage. Multiplying these 10,000 samples with the previously sampled proportion of individuals by viral load in each CD4 stage, we obtained the combined sets of simulation results which were then used for deriving the median and 2.5^{th} and 97.5^{th} percentiles as lower and upper estimates of the final per-act transmission probabilities. The infectiousness of PHI stage was estimated by applying a relative risk of 9.17, suggested by Boily et al [33], to the baseline transmission probabilities. All derived estimates are summarised in Table 2 in the main text.

For HIV transmission probability of ART-treated MSM, only MSM in the UK who have ever been treated with ART were included in the calculation. We allowed a decrease over time in our derived per-act transmission probabilities. To achieve that, we fitted a linear function to the proportions of MSM by viral load, CD4 stages, and calendar year simultaneously and estimated an average changes over 2005–2009 in these proportions. The proportions before 2005 was assumed equal to that of 2005,

and for the period after 2009 we assumed the same probabilities as of 2009. The rest of the calculations were carried out in the same manner to that of ART-naïve MSM. We derived annual change rate in transmission probabilities by subtracting the linear estimates of the any calendar year by estimates of the immediate previous year. In the fitting process, a lowest ART transmission probability of 0.001% was set for all types of sex acts to prevent the zero or negative values. The derived probabilities of transmitting HIV from ART-treated MSM and the changes over time are reported in Table 2 in the main text.

The effects of circumcision were not included according to a recent study that found no evidence to support the effects of circumcision in reducing HIV transmission risk among MSM in Britain [34]. We also decided not to include the effects of STIs in our model due to a lack of evidence to support that sexually transmitted infections (STIs) are risk factors for HIV transmission among MSM [35, 36].

The average efficacy of condom at reducing transmission of HIV among MSM was taken from the meta-analysis estimates of the efficacy in reducing heterosexual HIV transmission [37-39]. We assumed the per-act efficacy of 80% and the lower and upper limit of 75-95% (Table S9). The estimates were adopted only for anal sexual intercourses. We assumed no risk of HIV infection for all types of oral sex with condom use.

3.4 HIV diagnosis rate

The rate of HIV diagnosis controls the number of MSM moving from undiagnosed to diagnosed stages in each time step. We defined time-dependent HIV diagnosis rates by assuming increasing rates over 2000–2004 according to the rise in the reported number of new diagnoses during that period [15]. The initial diagnosis rate as of the year 2000 increased by a constant rate of changes until the end of 2004 and remained unchanged until the end of the simulation.

The HIV diagnosis rates were estimated through model fitting in Section 4 based on uninformative uniform priors. Here in this section, we only derived the lower and upper estimates using the reported numbers of new diagnoses by age and CD4 cell counts from 2000–2009. The effects of sexual activity level were also taken into account in this step.

The baseline rates of diagnosis with HIV in UK MSM were calculated by dividing the reported number of new diagnoses by the estimates of undiagnosed HIV prevalence with 95% credible intervals which were available from 2001 to 2007 [40, 41]. Since the 2000 estimates of undiagnosed prevalence were not available, we used the 2001 estimates instead and derived the baseline rates of 0.15 to 0.27 per year. According to the reported median CD4 at diagnosis of UK MSM of 366 cells/μL in 2001, we assumed the above baseline diagnosis rates for HIV-positive MSM at CD4 350-499 cells/μL in our model.

We then reduced the lower and upper limits of the initial rates by 30% to account for three main uncertainties. First, the initial rates are likely to be lower if the 2000 estimates of undiagnosed prevalence were made available and used. Second, the difference in the age ranges which were 15-59 years for the reported undiagnosed prevalence and 15-64 years in our model should decrease the rates. Third, it is more likely for HIV-positive MSM at CD4 350-499 cells/μL to have lower average diagnosis rates than the baseline because the median CD4 at diagnosis of UK MSM is almost out of the lower range of the CD4 stage. Based on the fact that the diagnosis rates could vary substantially across different disease stages with a very low rate during PHI stage to a much higher rate in an advanced HIV stage [42], we allowed initial diagnosis rates to increase by 0.05–0.15 per a CD4 stage advanced in disease progression.

Data from all three behavioural surveys (NATSAL, GMSHS, and GYM) surveys showed that, regardless of age group, high-activity MSM tested for HIV considerably more frequently than low-activity men. Accordingly, we distinguished the diagnosis rates between low- and high-activity men by estimating the differences in frequency of HIV testing from the survey data. We analysed the reported time of the last HIV test from NATSAL and arbitrarily assumed that individual tests every 12 months if he reported last test in the last year, 24 months for 1-2 years ago, and 60 months for 2-5 years ago. The rates of diagnosis were calculated as an inverse of the assumed frequencies of testing. Those who have tested more than 5 years ago or never had an HIV test were assigned a rate of zero. This resulted in an average testing rate of high-activity MSM that was 2.5-fold higher than that of low-activity MSM. Conducting the

similar analysis on GMSHS and GYM data yielded estimates of 1.70 and 1.78, respectively. Furthermore, the analysis also suggested that this ratio tended to remain unchanged over time. We therefore assumed a range of 1 (no difference) through 3 for this ratio which, in the next step of model fitting, will be applied to the initial diagnosis rate to calculate diagnosis rates by sexual activity level. The derived initial diagnosis rates are summarised in Table S10.

We derived the change rates per year by conducting 10,000 simulations and uniformly drew from the interval estimates of the 2001–2004 baseline diagnosis rates. For each simulation, subtracting the rates of all calendar year by the rates of the immediate previous year provided the changes of rate over time that were further used for calculating the average changes for the period. The 2.5^{th} and 97.5^{th} percentiles of these averages were -0.014 to 0.034. According to the reported numbers of new HIV diagnoses in MSM in the UK [15], it is unlikely that the diagnosis rates had decreased in that period, and hence we replaced the negative value with a very small positive value (0.001) for the lower limit of the change rates. No differences by CD4 stages and sexual activity levels were included. The derived estimates for changes over time in diagnosis rates are shown in Table S10.

We estimated the effects of HIV diagnosis on sexual activity level from a cohort study of 98 UK MSM [43]. The findings suggested that around 66% of MSM reduced the number of casual partners in the past 12 weeks after diagnosis of PHI while only 7% reported more casual partners. We directly used these percentages as the proportions of MSM who change sexual activity level after HIV diagnosis (Table S10).

3.5 Rate of initiating HIV treatment

The rate of diagnosed HIV-positive MSM starting an HIV treatment was taken from the UK Collaborative HIV Cohort (CHIC) Study [44] which provided the per 3-month period estimates by CD4 counts that ranged from 0.005 for CD4 350-499 cells/μL to 0.95 for CD4<100 cells/μL. The rates were derived based upon the guidelines regarding time to initiate HIV treatment in the UK [8]. Starting ART at CD4≥500 cells/μL is rather unusual; therefore we made an assumption that the treatment initiation rate for MSM at CD4≥500 cells/μL is as low as 10% of those with CD4 350-499 cells/μL. For CD4 200-349, we used an average derived from the rates of CD4 200-249, 250-299, and 300-349 cells/μL. For those with CD4<200 cells/μL, an average rate of CD4 0-99 and 100-199 cells/μL was used (Table S11). Due to a small number of cases, we assumed no ART initiation during PHI stage.

3.6 Disease progression

The disease progression rate for each CD4 stage is equal to the inverse of the mean duration of that stage. We adopted the estimates provided by a study [45] that fitted a multistate Markov model to the historical CD4 data from the CASCADE database. The study estimated an average time spent in each CD4 stage that is equivalent to a total time from seroconversion to AIDS of 8.6 years [45]. For PHI stage, we assumed according to Hollingsworth et al [46] a duration of 3 months with a range from 1 to 6 months and estimated this parameter via model fitting. All derived disease stage durations are shown in Table S12.

3.7 Initial population size

The initial MSM population size in the UK of the year 2000 (Section 3.1) was stratified according to various modelled subpopulations to form the initial size of each subgroup. Range of parameters that later fitted to the data were also derived. We began with stratification by HIV serostatus – HIV negative and positive. The 2000 UK annual HIV report [47] provided the estimated number of MSM aged 15-59 living with HIV in the UK at the end of 1999 of 17,200 which corresponded to the overall HIV prevalence of approximately 3.5%. The interval estimates were not provided so we selected the lower limit of 2.5% and the upper limit of 4.5% (Table S13). In the model fitting, these interval estimates were assigned to the low-activity men, of which we sampled the HIV prevalence parameters independently by age groups from the uniform distribution.

We accounted for differences in HIV prevalence by sexual actively level by applying the multiplicative factors directly to the sampled prevalence parameters of low-activity men in the same age group to obtain the estimates of high-activity men. Analysing GMSHS data revealed that the prevalence in the high-activity group was around 1.4-fold higher than that of the low-activity. We therefore assumed the range of 1 to 2 and uniformly sampled the multiplicative factors from this range for both age groups. Altogether, we can calculate the initial HIV prevalence for each class and estimate its values through model fitting.

HIV-positive MSM were then stratified by diagnosis status. We adopted the estimates derived from the Multi-parameter Evidence Synthesis (MPES) method of the

proportion of undiagnosed MSM aged 15-44 in the UK in 2001, 0.39 (0.32-0.47) [41], for low-activity MSM in age group 1, and the MPES estimates of MSM aged 15-59, 0.36 (0.29-0.44) [41], for age group 2. We expanded the lower and upper limits by ± 20% to account for the uncertainty due to inconsistencies of age range and calendar year between MPES estimates and the model requirements. For age group 2, both upper and lower limits were further reduced by 25% according to the fact that the proportion of undiagnosed infections was likely to be lower than the population average.

Similarly to prevalence of HIV, there was a higher undiagnosed proportion in the high-activity than in the low-activity group as suggested by GMSHS data. Consequently, we accounted for the effects of sexual activity level using multiplicative term that were assumed to range between 1 and 2 (Table S13). During the fitting process (Section 4), the proportions of undiagnosed HIV in low-activity men of both age groups were first sampled based on uniform priors. After that we drew the multiplicative term for effects of sexual activity level and calculated the undiagnosed proportion for all classes.

Next, we derived the proportion of ART-treated to all diagnosed MSM by dividing the reported number of MSM who have ever been on ART obtained from the SOPHID data by the estimated number of MSM living with diagnosed HIV in 2001 [40, 41]. We obtained an estimate of 75% of diagnosed MSM who have ever been treated with ART. For the remaining 25% who have never been on ART, we split them by the five CD4 stages based on the data from the CD4 cell counts database of patients seen for

HIV care in the UK. At that time, the number of MSM living with diagnosed HIV during PHI stage should be minimal. Hence we allocated only 1% of all diagnosed MSM to the stage of PHI.

The distribution of undiagnosed MSM by CD4 stages was adopted from a recent study on HIV incidence in MSM in England and Wales [48]. The lower and upper limits were arbitrarily assumed according to the adopted estimates (Table S13). The proportion of MSM in the undiagnosed PHI stage was assumed to be very low (2%). Due to a considerable uncertainty in this parameter, it was included to the model fitting and sampled from the Dirichlet distribution (Section 4).

Combing all the above parameters (summarised in Table S13) resulted in the initial total number of MSM fully stratified by age group, sexual activity, CD4 disease stage, HIV diagnosis, and ART status. We further stratified the modelled population into current and past MSM. NATSAL estimates of the proportion of past MSM were approximately 7% and 28% of all MSM in age group 1 and 2, respectively. Presanis et al [40] also provided estimates of the proportion of past MSM among undiagnosed and diagnosed HIV prevalence (Table S13). We combined all this information to construct the initial numbers of past MSM by age group and disease stage. Sex between men in past MSM was assumed absent and therefore stratification by sexual activity was no longer applied. We captured the ratio between past and current MSM by age group and disease stage from the initial MSM size. This ratio was used for balancing the numbers of current and past MSM in the model.

Finally, among current MSM, a group of individuals who reported as currently in a relationship with a man based on GYM data was assigned to the ‘single’ group while the remaining men belonged to the ‘pair’ group (Table S13). All calculations for the initial numbers of MSM were carried out in the model fitting before simulations.

4. Model fitting

We fitted our model using Monte Carlo filtering method [1] to match the HIV epidemic in MSM in the UK during 2001–2009. We began by conducting preliminary one-way sensitivity analyses on all model parameters to see the effects on the model estimates of HIV prevalence. All parameter values were varied by $\pm 50\%$ and the corresponding estimates of the overall HIV prevalence at the end of 2009 were compared to that of the baseline simulation. Any parameters that caused the 2009 prevalence to change greater than 10% compared to the baseline parameters were included in the fitting. Table S14 summarises results from the preliminary sensitivity analysis. We manually added the one-off sexual partnership formation rates, the proportion of susceptibles who have UAI with a diagnosed HIV positive repeat sexual partner, and three key model initial conditions: the HIV prevalence, the proportion of undiagnosed prevalence, and the distribution of undiagnosed prevalence by disease stages at the beginning of 2000. The probability distributions of all fitted parameters are summarised in Table S14 and the ranges can be found in Section 3. The other parameters not involved in the fitting were held constant at their baseline values.

We then sampled 20,000 different combinations of the fitted parameters using Latin Hypercube sampling [49] and ran 20,000 model simulations from 2000 to 2009. Using Monte Carlo filtering, model outputs from each parameter set were compared against the 2001–2009 estimates of the reported HIV prevalence, both overall and undiagnosed [23, 40, 50]. Any parameter sets that were unable to provide estimates within ranges of all fitting data simultaneously were filtered out. The interval estimates of HIV prevalence were taken directly from the reported ranges of MSM

aged 15-59, the upper limits of which were then increased by 15% to account for the difference in the age range.

The filtering process reduced the number of parameter sets from 20,000 to 1,093. The model fit to the HIV prevalence estimates are shown in Figure S2. The filtered parameter sets also provided an adequate fit to the diagnosed prevalence data judging from the resulting median estimates (Figure S2). Our estimated number of MSM aged 15-64 living with HIV in the UK ranged from around 24,000 in 2001 to 38,000 in 2009 with a rather constant number of 9,000 men who were unaware of their infection. These 1,093 parameter sets were later used for simulating the HIV epidemic in MSM in the UK from 2001 to 2020 and evaluating the contributions to HIV transmission of various biological and behavioural factors.

5. Model validation

The model and its parameters values were validated collectively by comparing the output estimates derived from 1,093 parameter sets to the time-series surveillance data and reported estimates that were not used for fitting the model, which were the annual number of new HIV infections, the number of new HIV diagnoses, and the number of ART-treated MSM. Figure S3 shows a comparison of new HIV infections. Our model yields a total of 21,677 (15,505-29,596) new infections during 2001–2009 with the annual numbers lie within the reported ranges [48] across the entire time period (Figure S3a). The upward trend of new infections during the early periods followed by a slight decline and levelling off can be observed from both our and reported estimates. There were no reported estimates stratified by age groups. However, the smaller number of new infections of the older group (Figure S3c) compared to the younger group (Figure S3b) were also consistent with the 2010 data from the national monitoring system of recent HIV infections which suggesting much lesser proportions of recent infection in newly-diagnosed MSM aged 35 and over (7-23%) in contrast to those of aged less than 35 (24-41%) [19, 41].

We then compared in Figure S4 our median estimates of individuals in the ART stage at the end of the year during 2001–2009 with the observed numbers of MSM on ART obtained from the PHE SOPHID database. The upward trend in the numbers of ART-treated MSM was observed during the last decade (from around 9,200 in 2001 to 21,300 in 2009) and our estimates follow the trend consistently (from 10,952 in 2001 to 23,865 in 2009). Note that our estimates were somewhat higher than the observed data, particularly in the early phase, probably because we assumed that once an

individual is treated with ART he remains in the ART stage until being removed from the model, while in the actual situation a proportion of MSM may stop the treatment due to a number of reasons, e.g. poor adherence.

The numbers of new diagnoses stratified by age groups are shown in Figure S5. The 2001–2009 new diagnoses data in UK MSM aged 15-64 were obtained from PHE New diagnoses databases. Since the selected data matched exactly with the demographics of our modelled populations, we validated our HIV transmission model more quantitatively by calculating a simple goodness-of-fit indicator from percentage differences between observed data and model estimates for each available data points, $1-\frac{|O_t - E_t|}{O_t} \times 100\%$, where O_t and E_t are, respectively, the observed data and model estimates at time *t*. The combined value is an average of all individual data points' average goodness-of-fit. A larger value reflected a better consistency with the data. For the younger group, the goodness-of-fit value of the new diagnoses during 2001–2009 was 93.9% when comparing using the median estimates. For MSM aged 35-64, the goodness-of-fit value was 88.6%. Overall, the predicted trends of new HIV diagnoses in this research were well consistent with the national surveillance data with the goodness-of-fit value of 91.2%.

The CD4-specifc new diagnoses data and our estimates are shown in Figure S6. Due to uncertainty associated with missing values of the reported data, we skipped the calculation of goodness-of-fit and visually inspected the consistency. We found that our median estimates resembled quite closely both trend and magnitude of the new diagnoses data in all CD4 stages and both age groups. At the end of 2009, our

estimation of the numbers of newly diagnosed MSM were 757, 571, 400, and 504 for CD4≥500, 350-499, 200-349, and <200 cells/μL, respectively, which matched closely the 805, 535, 426, and 440 cases of the reported data.

In addition, our model predicted a total of 39,295 HIV-positive MSM aged 15-64 at the end of 2010 which was in line with the 2010 estimates of 40,100 (35,300-46,700) MSM aged 15-59 living with HIV in the UK [19]. At the same period of time, we also approximated that 21% of HIV-positive MSM were unaware of their infection compared to the reported estimate of 26% (16-36%) [19].

6. Supplementary results

6.1 Epidemic projections

MSM living with HIV

The projections for the number of MSM living with HIV in the UK stratified by diagnosis status, age group, and sexual activity level are shown in Figure 2 in the main article. The model estimated that there would be 43,682 (35,833-56,387) MSM living with HIV by the end of 2013 corresponding to a prevalence rate of 6.3% (5.1-8.1%). By 2020, the estimated HIV prevalence rose to 52,268 (38,064-81,006) equivalent to 7.3% (5.3-11.3%). During this period (2013–2020), the overall HIV prevalence increased at an average annual rate of 2.6% with the majority of men being diagnosed, particularly those aged 35-64. The numbers of undiagnosed individuals stabilised over time in all age groups. For both sexual activity groups, the overall HIV prevalence at the end of 2013 are similar, but the numbers of undiagnosed men in the high-activity group is roughly twice the number in the low-activity group (5,584 vs. 2,311). The number of diagnosed HIV infections is similar in both activity groups (low: 17,951, high: 17,376).

New HIV infections

According to the median estimates of new HIV infections in Figure 2 in the main article, our model indicated that there was an increase in the overall HIV incidence during 2001–2005 before a slight decline during 2005–2009 and levelling off at around 2,400 cases per year (incidence rate of 0.35%) after that, however this is

associated with large uncertainty. The 2001–2020 cumulative numbers of new infections were estimated to be 48,148 (29,254-83,713). The rates of new infections in the younger group were about twice as many as those in the older group (0.49% vs. 0.25% in 2013), while unsurprisingly, the incidence in the high-activity men was substantially higher than in the low-activity men (0.64% vs. 0.14% in 2013).

6.2 Population attributable fractions

The population attributable fraction (PAF) stratified by age groups and sexual activity levels are shown in Figure S7. The details of analysis and calculation of PAF are provided in the main text.

The PAFs for all MSM are presented in Figure S7a. Figure S7b and Figure S7c show that eliminating HIV spread from MSM aged 15-34, while being highly beneficial within the age group (average PAF of 75%), would prevent a almost half of new infections in MSM aged 35-64 (average PAF of 46%) which is considerably high compared to only 61% in the case where there were no transmission from within the older group itself. Moreover, the contributions remained stable over the 20-year period which, taken together, indicated that interventions targeting young MSM could prove effective in tackling HIV infection. Other factors were shown to uniformly affect both age groups.

The PAF calculated from the incidence rates stratified by sexual activity levels are shown in Figure S7d and Figure S7e. For both groups, the largest PAF is still associated with transmission from men at non-PHI stages while the contributions of

men at PHI are among the lowest. The smallest PAF for the low-activity group is 5% on average for one-off sexual partnership due to low rates of acquiring new partners (≤1 per year) of men in this group (Figure S7d). The treated men seem to contribute considerably low (average PAF of 14%) to HIV infections in high-activity group as a result of small proportions of men who still performed high activity after HIV diagnosis (Figure S7e). This also explains why the low-activity PAF of HIV diagnosis and treatment changed more rapidly over time than that of high-activity men. The analyses also indicate that HIV-positive men in high-activity group accounted for more infections in low-activity populations than do the low-activity men (Figure S7d and Figure S7e), which is similar to how young MSM contributed to new infections in both age groups (Figure S7b and Figure S7c).

Moreover, the repeat sexual partnerships of high-activity MSM make a small contribution to new infections in low-activity men (Figure S7d), whereas the repeat sexual partnerships of low-activity MSM make almost no contribution to HIV infection in high-activity men (Figure S7e), demonstrating the minor impact of HIV transmission between the two sexual activity groups.

6.3 Sensitivity analysis

The sensitivity of model predictions was assessed using the regression trees technique [51]. We selected total new HIV infections during 2001–2020 as the response variable and all filtered parameters except the model initial conditions as the predictor variables (Table S14). We grew the tree and pruned it by selecting the tree size that minimised the sum of squared errors [51].

The sensitivity analysis regression tree a total of 7 splits, 8 terminal nodes, and 4 model parameters (Figure S8). The most influential model parameter for the new infections during 2001–2020 based on the number of appearances and its level in the tree was the per-URAI act transmission probability in non-PHI stages. The incidence estimates were also highly sensitive to the per-UIAI act non-PHI transmission probability, condom efficacy, and the frequency of sex acts with a repeat sexual partner.

The uncertainty of the estimated infections (43,560-64,980) indicated the importance of these factors in the epidemic of HIV in UK MSM. The number of new infection over 2001–2020 could be as high as 64,980 if the mean infectivity of an URAI and UIAI act at non-PHI stages was greater than 2.0% and 1.0%, respectively. Conversely, if several conditions represented by the route on the far left of the tree (Figure S8) were met simultaneously, the incidence of HIV could have been as low as 43,560, corresponding to 13% reduction from the mean estimates of 50,110 cases in the baseline scenario (root node).

Supplementary tables

Table S1: Model parameters

Parameter	Definition
$N_{j,k,h}$	Total number of MSM in age group *j*, sexual activity group *k*, and HIV stage *h*
$X_{j,k,h}$	Total number of single MSM in age group *j*, sexual activity group *k*, and HIV stage *h*
$Y_{j,k,h}$	Number of current single MSM in age group *j*, sexual activity group *k*, and HIV stage *h*
$Z_{j,k,h}$	Number of past single MSM in age group *j*, sexual activity group *k*, and HIV stage *h*
$P_{j,k,h}^{m,n,r}$	Number of pairs between an individual of age group *j*, sexual activity group *k*, and HIV stage *h* and an individual of age group *m*, sexual activity group *m*, and HIV stage *r*
$\nu_{j,k}$	Influx of new susceptibles in age group *j* (*j*=1), and sexual activity group *k*
$\eta_{j,k}$	Number of current MSM in age group *j*, and sexual activity group *k* (*k*=1) moving to past MSM
α_j	Ageing rate from an individual in age group *j* to next age group or out of the model
$\mu_{j,h}$	Mortality rate of an individual in age group *j* and HIV stage *h*
γ_h	Disease progression rate of an individual in disease stage *h* (*h*=2,...,11) to the next stage
$\phi_{j,k,h}$	HIV diagnosis rate of an individual in age group *j*, sexual activity group *k*, and undiagnosed stage *h* (*h*=2,4,6,8,10)
s_k	The proportion of MSM in sexual activity group *k* who switch to another sexual activity group after being diagnosed with HIV
τ_h	ART initiating rate of an individual in diagnosed stage *h* (*h*=5,7,9,11)
$\rho_{j,k}^{rep}$	Repeat sexual partnership formation rate of an individual in age group *j*, and sexual activity group *k*
$\sigma_{j,k}$	Repeat sexual partnership dissolution rate of an individual in age group *j*, and sexual activity group *k*
$\psi_{j,k,h}^{rep,m,n,r}$	Chance of an individual in age group *j*, sexual activity group *k*, and HIV stage *h* to acquire a repeat sexual partner of age group *m*, sexual activity group *n*, and HIV stage *r*
$\psi_{j,k}^{one,m,n,r}$	Chance of a susceptible in age group *j*, and sexual activity group *k* to acquire a one-off sexual partner of age group *m*, sexual activity group *n*, and HIV stage *r*

Parameter	Definition
$\lambda_{Y,j,k}^{one}$	Force of infection in a relationship between a one-off sexual partner and a single susceptible individual of age group *j*, and sexual activity group *k*
$\lambda_{P,j,k}^{one}$	Force of infection in a relationship between a one-off sexual partner and a paired susceptible individual of age group *j*, and sexual activity group *k*
$\lambda_{j,k}^{rep,m,n,r}$	Force of infection in a relationship between a repeat sexual partner of age group *m*, sexual activity group *n* , and disease stage *h* (*h*=2,...,1) and a susceptible individual of age group *j*, and sexual activity group *k*

Table S2: Number of survey respondents included for data analysis

MSM	NATSAL [a]		GMSHS		GYM	
	n	%	n	%	n	%
Current MSM						
Aged 16-34	101.5	65.4	9,936	57.5	1,894	44.2
Aged 35-64	53.7	34.6	7,347	42.5	2,393	55.8
Total	155.2	100.0	17,283	100.0	4,287	100.0
Past MSM						
Aged 16-34	7.1	25.7	-	-	-	-
Aged 35-64	20.5	74.3	-	-	-	-
Total	27.6	100.0	-	-	-	-
All MSM						
Aged 16-34	108.6	59.4	9,936	57.5	1,894	44.2
Aged 35-64	74.2	40.6	7,347	42.5	2,393	55.8
Total	182.8	100.0	17,283	100.0	4,287	100.0

[a] With weights for core sample and the ethnic minority boost sample

Table S3: Weight adjustment for GMSHS and GYM data according to four variables: age of survey respondents, number of male sexual partners in the last year, number of UAI male sexual partners in the last year, and time of last HIV test

Variable	Reported								Adjusted (all weights simultaneously) [a]			
	NATSAL		GMSHS			GYM			GMSHS		GYM	
	n [b]	%	n	%	Weight	n	%	Weight	n	%	n	%
Age of respondents [c]												
16-24	29.4	18.9	892	16.6	1.138	68	5.2	3.645	1,322.9	23.4	318.7	23.8
25-34	72.1	46.5	2,763	51.5	0.902	652	49.8	0.933	2,626.8	46.5	596.1	44.5
35-44	53.7	34.6	1,711	31.9	1.086	590	45.0	0.769	1,702.4	30.1	425.3	31.7
median		32		32			35			32		32
Number of male sexual partners in the last year [d]												
Aged 16-34 [e]												
0	17.9	17.9	63	1.8	9.857	-	-	-	693.8	18.4	-	-
1	32.4	32.4	494	14.3	2.271	-	-	-	1,354.3	36.0	-	-
2	7.4	7.4	249	7.2	1.024	-	-	-	267.0	7.1	-	-
3	12.4	12.4	237	6.8	1.813	-	-	-	489.2	13.0	-	-
4	11.0	11.0	207	6.0	1.841	-	-	-	366.0	9.7	-	-
5-9	4.3	4.3	542	15.7	0.273	-	-	-	144.8	3.8	-	-
10-14	6.2	6.2	451	13.0	0.474	-	-	-	195.9	5.2	-	-
15-19	1.2	1.2	190	5.5	0.215	-	-	-	39.6	1.1	-	-
20-29	2.1	2.1	350	10.1	0.212	-	-	-	63.8	1.7	-	-
30-49	0.5	0.5	271	7.8	0.065	-	-	-	16.1	0.4	-	-
50+	4.6	4.6	408	11.8	0.389	-	-	-	134.6	3.6	-	-
mean		7.30		21.61			-			6.38		-
Aged 35-64												
0	9.7	18.1	84	3.8	4.714	-	-	-	353.3	15.7	-	-
1	19.8	37.0	320	14.6	2.532	-	-	-	921.0	41.0	-	-
2	5.8	10.8	161	7.4	1.468	-	-	-	241.4	10.7	-	-
3	2.3	4.3	139	6.3	0.678	-	-	-	89.6	4.0	-	-
4	2.6	4.9	106	4.8	1.008	-	-	-	113.4	5.0	-	-
5-9	3.8	7.1	282	12.9	0.552	-	-	-	158.4	7.0	-	-
10-14	5.6	10.4	284	13.0	0.802	-	-	-	221.1	9.8	-	-
15-19	1.1	2.0	92	4.2	0.475	-	-	-	38.6	1.7	-	-

Variable	Reported								Adjusted (all weights simultaneously) [a]			
	NATSAL		GMSHS			GYM			GMSHS		GYM	
	n [b]	%	n	%	Weight	n	%	Weight	n	%	n	%
20-29	1.8	3.4	254	11.6	0.289	-	-	-	69.9	3.1	-	-
30-49	0.3	0.6	164	7.5	0.086	-	-	-	13.6	0.6	-	-
50+	0.8	1.5	304	13.9	0.105	-	-	-	28.5	1.3	-	-
mean		4.48		23.14			-			4.87		-
Number of UAI male sexual partners in the last year [f]												
0	30.6	34.8	2,914	53.3	0.653	514	45.9	0.758	2,112.5	41.1	449.9	40.0
1	35.9	40.9	1,506	27.5	1.485	335	29.9	1.366	2,463.8	48.0	421.4	37.5
2	10.2	11.6	464	8.5	1.365	105	9.4	1.234	305.8	6.0	137.6	12.2
3	6.8	7.7	167	3.1	2.524	54	4.8	1.597	177.3	3.5	77.9	6.9
4+	4.4	5.0	418	7.6	0.653	111	9.9	0.503	78.4	1.5	37.2	3.3
mean		1.52		1.90			2.13			1.18		1.34
Time of last HIV test												
Aged 16-34												
In the last year	12.5	12.7	1,184	34.7	0.365	217	31.9	0.397	352.5	9.4	108.5	12.6
More than a year ago	34.5	34.9	1,147	33.6	1.039	290	42.6	0.819	1116.6	29.7	269.8	31.3
Never had an HIV test	51.8	52.4	1,080	31.7	1.654	173	25.4	2.059	2290.7	60.9	482.7	56.1
Aged 35-64												
In the last year	8.5	18.9	529	25.0	0.755	172	24.9	0.758	339.4	15.7	107.1	19.7
More than a year ago	11.1	24.8	948	44.8	0.554	355	51.4	0.483	518.2	23.9	136.6	25.1
Never had an HIV test	25.2	56.3	640	30.2	1.863	164	23.7	2.373	1308.5	60.4	300.8	55.2
Total weights [g]												
Aged 16-34												
Minimum	-	-	-	-	0.014	-	-	0.187	-	-	-	-
Maximum	-	-	-	-	18.558	-	-	11.984	-	-	-	-
Aged 35-64												
Minimum	-	-	-	-	0.031	-	-	0.187	-	-	-	-
Maximum	-	-	-	-	9.537	-	-	3.243	-	-	-	-

Abbreviation:
UAI: Unprotected anal intercourse
Proportions of some variables shown here may not sum to one due to rounding off decimals.
NATSAL data of the year 2000
GMSHS data of the year 1999, 2000, and 2001

GYM data of the year 2000 and 2001

[a] Adjustment with all weights simultaneously results in similar but non-identical proportions to NATSAL.
[b] The reported numbers may not be an integer due to the original NATSAL weighting.
[c] Age range in NATSAL is 16-44. Therefore, we can adjust the age of GMSHS and GYM respondents only of the same range. The remaining ages were left as reported.
[d] The number of male sexual partners in the last year is not available in GYM.
[e] The reported numbers and proportions of NATSAL for this variable are shown identical due to rounding off decimals.
[f] The variable was not categorised by age group due to the small number of NATSAL cases.
[g] Total weights are the product of all independent weights of the above four variables. The minimum and maximum rows represent the minimum and maximum total weight for the corresponding age groups, respectively. The total weights were then used for adjusting the original GMSHS and GYM data.

Table S4: Estimated 5-year age distribution of MSM aged 15-64 of year 2000

Age group	Proportion (%) [a]
15-19	3.13
20-24	7.57
25-29	10.45
30-34	17.08
35-39	15.63
40-44	10.48
45-49	9.64
50-54	10.31
55-59	8.38
60-64	7.33
Total	100.00

[a] Estimated from NATSAL and ONS data

Table S5: Data used for deriving MSM population size in the UK of year 2000

Description	London	Elsewhere in England	Outside England	Entire UK
Adjusted proportion of MSM to male population (%) [a,b]				
Aged 15-34	6.60	2.67	2.78	-
Aged 35-64	8.55	3.26	0.94	-
Mid-year estimates size of male population size [b,c]				
Aged 15-34	1,196,000	5,436,300	1,266,400	-
Aged 35-64	1,263,000	8,087,000	1,824,800	-
Proportion of current MSM (%) [d,e]				
Aged 15-34	-	-	-	93.38
Aged 35-64	-	-	-	72.42
Proportion of past MSM (%) [d,e]				
Aged 15-34	-	-	-	6.62
Aged 35-64	-	-	-	27.58

[a] Data from NATSAL and PHE

[b] Data for the entire UK are not shown because they would result in total number of MSM population that are different from the current calculation which based on the three regions.

[c] Data from ONS

[d] Data from NATSAL

[e] We were unable to stratify the proportions by regions because of the small number of cases.

Table S6: Demographic parameters

Parameter	Symbol	Value		Source
		Aged 15-34	Aged 35-64	
MSM size of year 2000				
Current MSM	-	242,678	285,122	NATSAL, ONS
Past MSM	-	16,822	103,878	NATSAL, ONS
All MSM	-	259,500	389,000	NATSAL, ONS
Proportion of past MSM	-	0.0662	0.2758	NATSAL
Ageing rate per year	a	0.0523	0.0261	NATSAL, ONS
Growth rate per year	g	0.0043	0.0056	NATSAL, ONS
Mortality rate per year				
HIV-negative MSM	μ	0.00094	0.00449	NATSAL, ONS
HIV-positive MSM [a]	μ	0.00843	0.00899	Assumed [b]

[a] Applied only to HIV-positive MSM on antiretroviral treatment
[b] Assumed based on findings from ref [22]

Table S7: Number of MSM stratified by sexual activity level

MSM	Aged 16-34	%	Aged 35-64	%	Total	%
NATSAL						
Low	50.3	50.3	29.4	55.1	79.7	52.0
High	49.7	49.7	24.0	44.9	73.7	48.0
Total	100.0	100.0	53.4	100.0	153.4	100.0
GMSHS						
Low	4943.9	54.1	3692.1	56.8	8636	55.2
High	4195.4	45.9	2809.4	43.2	7004.8	44.8
Total	9139.3	100.0	6501.5	100.0	15640.8	100.0
GYM						
Low	608.7	29.9	607.0	35.9	1215.7	32.6
High	1425.7	70.1	1084.6	64.1	2510.3	67.4
Total	2034.4	100.0	1691.6	100.0	3726.0	100.0

Table S8: Data used for calculating HIV transmission probability of MSM who have never been treated with antiretroviral treatment

Description	Viral load (copies/mL)				
	0- 399	400-999	1,000-9,999	10,000-49,999	50,000+
Median viral load (copies/mL) [a]	49	1,690	6,234	22,948	102,077
Proportion of men by viral load in each CD4 stage [a]					
CD4 ≥ 500	0.1051	0.1700	0.1763	0.3219	0.2267
CD4 350-499	0.0700	0.1133	0.1487	0.3552	0.3129
CD4 200-349	0.0830	0.0840	0.1168	0.3341	0.3821
CD4 < 200	0.1085	0.0814	0.0492	0.2136	0.5475
HIV transmission probability (%) [b]					
URAI	0.135 (0.019-0.241)	0.535 (0.076-0.956)	0.889 (0.127-1.588)	1.477 (0.211-2.637)	2.640 (0.377-4.714)
UIAI	0.060 (0.007-0.162)	0.237 (0.027-0.642)	0.394 (0.044-1.067)	0.654 (0.074-1.772)	1.169 (0.132-3.168)
UROI	0.004 (0.001-0.016)	0.015 (0.004-0.065)	0.025 (0.006-0.108)	0.042 (0.011-0.179)	0.075 (0.019-0.321)

The values in parentheses represent the lower and upper limits of the corresponding parameters that were included in the model fitting.

Abbreviations:
URAI: Unprotected receptive anal intercourse
UIAI: Unprotected insertive anal intercourse
UROI: Unprotected receptive oral intercourse

[a] Estimated from SOPHID 2005–2009 data
[b] Not the final transmission probability

Table S9: Condom efficacy in preventing HIV transmission

Parameter	Symbol	Value	Source
Condom efficacy per sexual act	ε_{cdm}	0.80 (0.75-0.95)	Assumed [a]

The values in parentheses represent the lower and upper limits of the corresponding parameters that were included in the model fitting.

[a] Assumed based on findings from ref [37-39]

Table S10: HIV diagnosis parameters

Parameter	Symbol	Value
Initial HIV diagnosis rates per year [a]	ϕ	0.102-0.191
Ratio of diagnosis rate of high-activity to low-activity MSM [b]	-	1-3
Increase in initial diagnosis rates per a CD4 stage [c]	-	0.05-0.15
Changes per year in HIV diagnosis rate during 2000–2004 [d]	$\Delta\phi$	0.001-0.0339
Proportion of MSM who change sexual activity group after HIV diagnosis		
From low-activity to high-activity group	s_1	0.07
From high-activity to low-activity group	s_2	0.66

We assumed identical interval estimates of HIV diagnosis rate for MSM of different sexual activity level, age group, and disease stage. The interval estimates in this table will be used for constructing uniform priors used for estimating the diagnosis rates via model fitting.

[a] At the beginning of year 2000; for MSM at CD4 350-499 cells/μL
[b] Derived from NATSAL, GMSHS, and GYM data
[c] An absolute increase of the corresponding rates
[d] The lower bound was assumed 0.001 and the upper bound was derived from simulations based on uniform distribution.

Table S11: Rate of initiating HIV treatment

Parameter	Symbol	Value
Rate before 1st January 2009 (per year)		
CD4 ≥ 500	τ_5	0.002
CD4 350-499	τ_7	0.02
CD4 200-349	τ_9	1.80
CD4 < 200	τ_{11}	3.80
Rate from 1st January 2009 onwards (per year) [a]		
CD4 ≥ 500	τ_5	0.02
CD4 350-499	τ_7	1.80
CD4 200-349	τ_9	3.80
CD4 < 200	τ_{11}	3.80

HIV treatment during PHI stage was omitted from the model.

[a] Revised due to an increased ART initiation threshold [8]

Table S12: Duration of HIV stages

Parameter	Symbol	Value (months)
PHI to CD4 ≥ 500	$1/\gamma_2, 1/\gamma_3$	2.90 (1.24-6.00)
CD4 ≥ 500 to 350-499	$1/\gamma_4, 1/\gamma_5$	67.27
CD4 350-499 to 200-349	$1/\gamma_6, 1/\gamma_7$	23.92
CD4 200-349 to < 200 [a]	$1/\gamma_8, 1/\gamma_9$	20.12

The values in parentheses represent the lower and upper limits of the corresponding parameters that were included in the model fitting. An inverse of the above stage durations is the disease progression rates used in the model.

Abbreviation:
PHI: Primary HIV infection

[a] In our model, individuals remained in the stage of CD4<200 cells/μL until they are either diagnosed, treated with ART, or removed from the model.

Table S13: Data used for constructing initial numbers of MSM by subgroups

Description	Value
HIV prevalence [a]	
Baseline estimates [b]	0.025-0.045
Multiplicative factor for high-activity group	1-2
Proportion of undiagnosed HIV to all HIV-positive MSM	
Undiagnosed proportion for age group 1	0.26-0.57
Undiagnosed proportion for age group 2	0.18-0.40
Multiplicative factor for high-activity group	1-2
Proportion distribution of diagnosed MSM by disease stage [c]	
PHI	0.01
CD4 ≥ 500	0.42
CD4 350-499	0.27
CD4 200-349	0.21
CD4 < 200	0.09
Proportion distribution of undiagnosed MSM by disease stage [c]	
PHI	0.02 (0.01-0.04)
CD4 ≥ 500	0.50 (0.40-0.60)
CD4 350-499	0.28 (0.18-0.38)
CD4 200-349	0.15 (0.05-0.25)
CD4 < 200	0.05 (0.02-0.08)
Proportion of ART-treated MSM to all diagnosed MSM	0.75
Proportion of past MSM to all MSM by HIV status	
Susceptibles	
Aged 15-34	0.066
Aged 35-64	0.276
Undiagnosed HIV	0.028
Diagnosed HIV	0.025
Proportion of single MSM to all current MSM by class [e]	
Aged 15-34	
Low-activity	0.463
High-activity	0.530

Description	Value
Aged 35-64	
Low-activity	0.334
High-activity	0.493

The values in parentheses represent the lower and upper limits of the corresponding parameters that were included in the model fitting.

Abbreviations:
ART: Antiretroviral treatment
PHI: Primary HIV infection

[a] At the beginning of year 2000; assumed for MSM at CD4 350-499 cells/μL
[b] Assumed for low-activity MSM in model fitting.
[c] The sum of proportions equals one.
[d] Single MSM are men who are currently not in a repeat sexual partnership with a man.

Table S14: Model parameters included in the model fitting

Parameter	Effects on prevalence (%) [a]	Distribution
UIAI transmission probability	18.31	Beta
URAI transmission probability	47.40	Beta
Condom efficacy	35.63	Uniform
Rate of repeat sexual partnership formation	19.21	Beta (low-activity) and gamma (high-activity)
Rate of one-off sexual partnership formation	-	Beta (low-activity) and gamma (high-activity)
Proportion of repeat sexual partnerships that involve anal sex to all repeat sexual partnerships	50.53	Beta
Proportion of susceptibles who have UAI with a perceived HIV negative repeat sexual partner	25.10	Uniform
Proportion of susceptibles who have UAI with a diagnosed HIV positive repeat sexual partner	-	Uniform
Frequency of sexual acts with a repeat sexual partner	59.71	Uniform
HIV diagnosis rate	17.06	Uniform
Duration of PHI stage	31.37	Gamma
Initial HIV prevalence at the beginning of 2000	-	Uniform
Initial proportion of undiagnosed HIV to all HIV-positive MSM at the beginning of 2000	-	Uniform
Initial proportion distribution of undiagnosed MSM by disease stage	-	Dirichlet

Abbreviations:
URAI: Unprotected receptive anal intercourse
UIAI: Unprotected insertive anal intercourse
UAI: Unprotected anal intercourse
PHI: Primary HIV infection

[a] The results from preliminary one-way sensitivity analysis prior to the model fitting that shows the maximum percentage changes in the model estimates of HIV prevalence at the end of 2009 due to ± 50% changes of the fitted parameters.

Supplementary figures

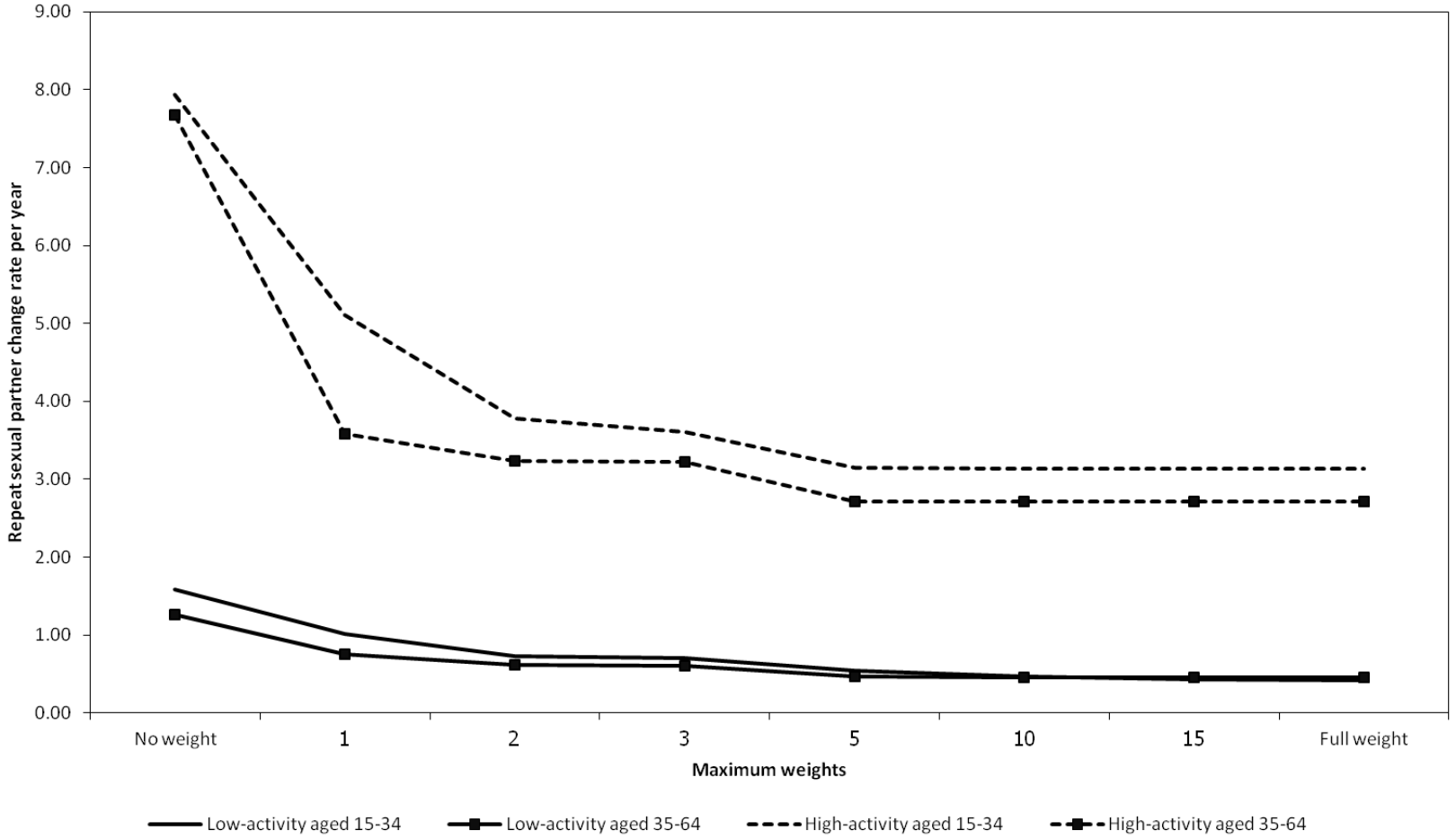


Figure S1: Comparison of repeat sexual partnership formation rates using different adjustment weights

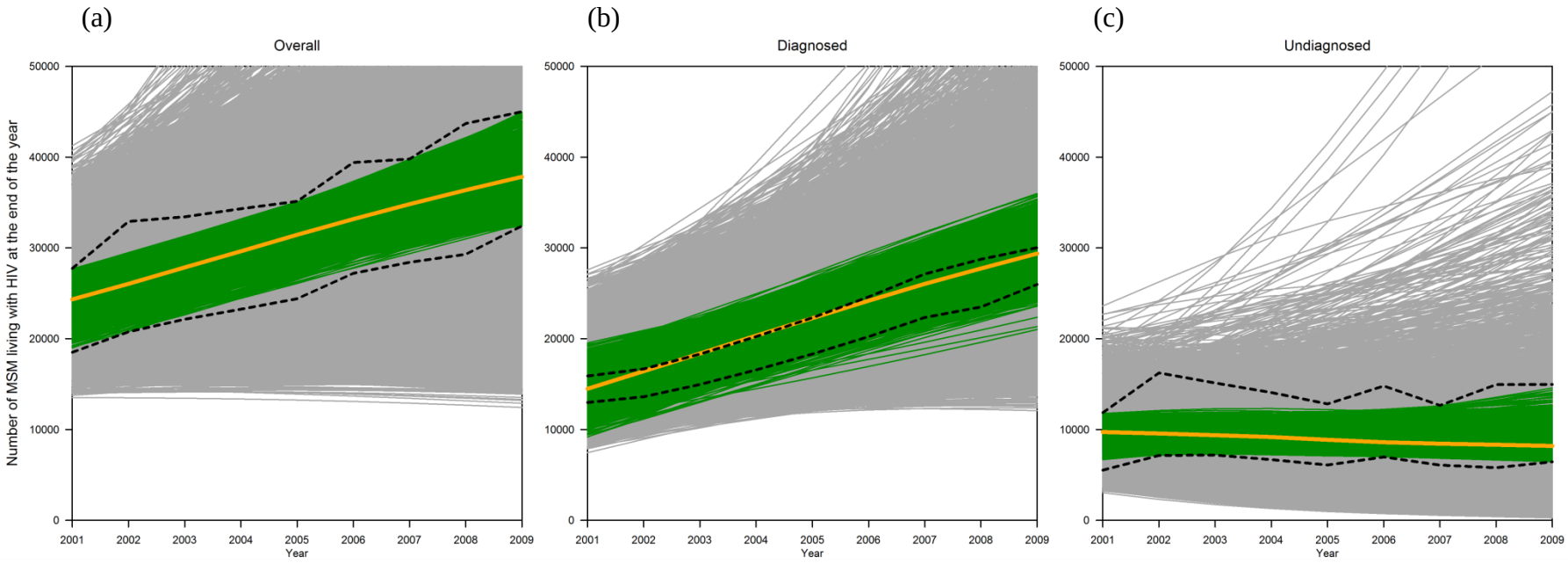


Figure S2: The model fitting of the number of MSM aged 15-64 living with HIV in the UK

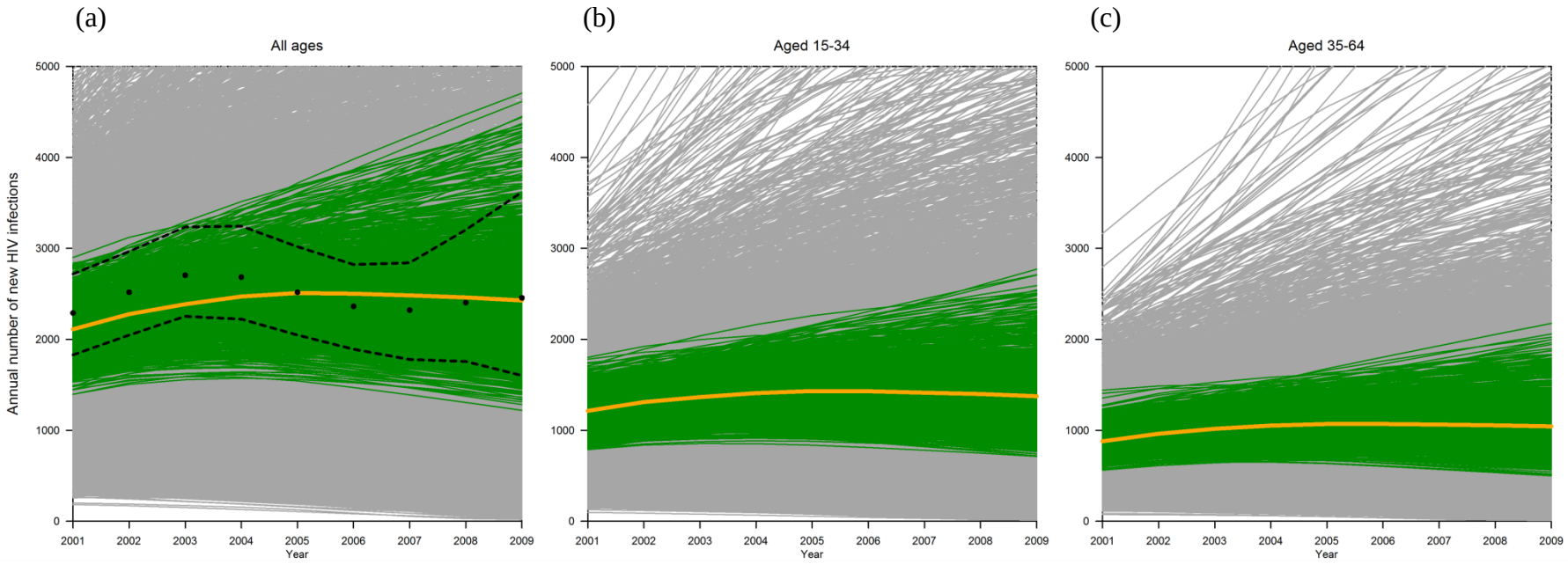


Figure S3: Comparison of annual numbers of new HIV infections between model and reported estimates for model validation

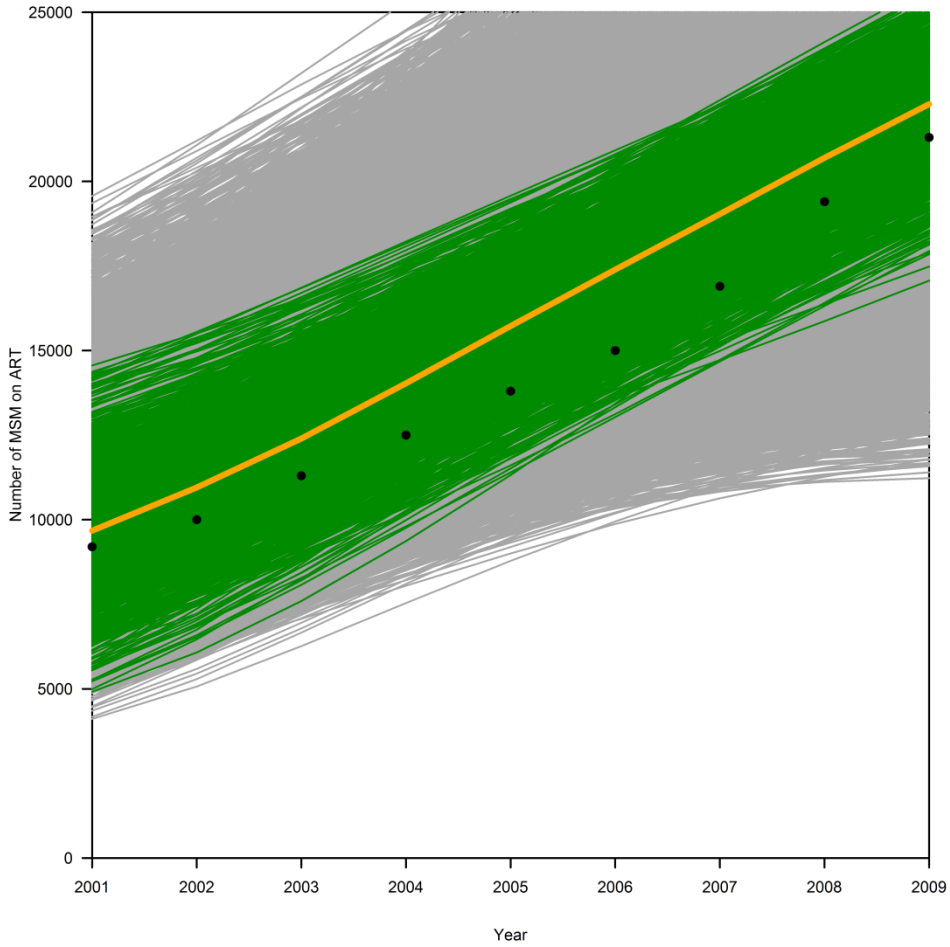


Figure S4: Comparison of the number of MSM on ART between model estimates and reported data for model validation

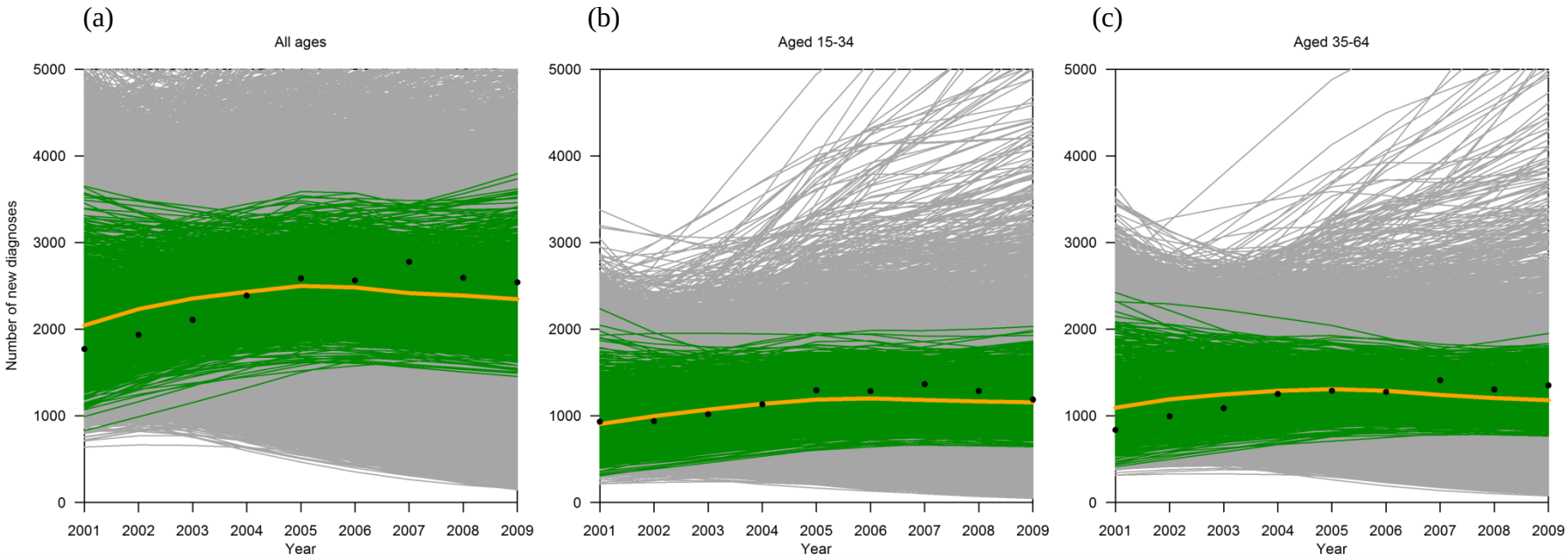


Figure S5: Comparison of the number of new diagnoses between model estimates and reported data for model validation

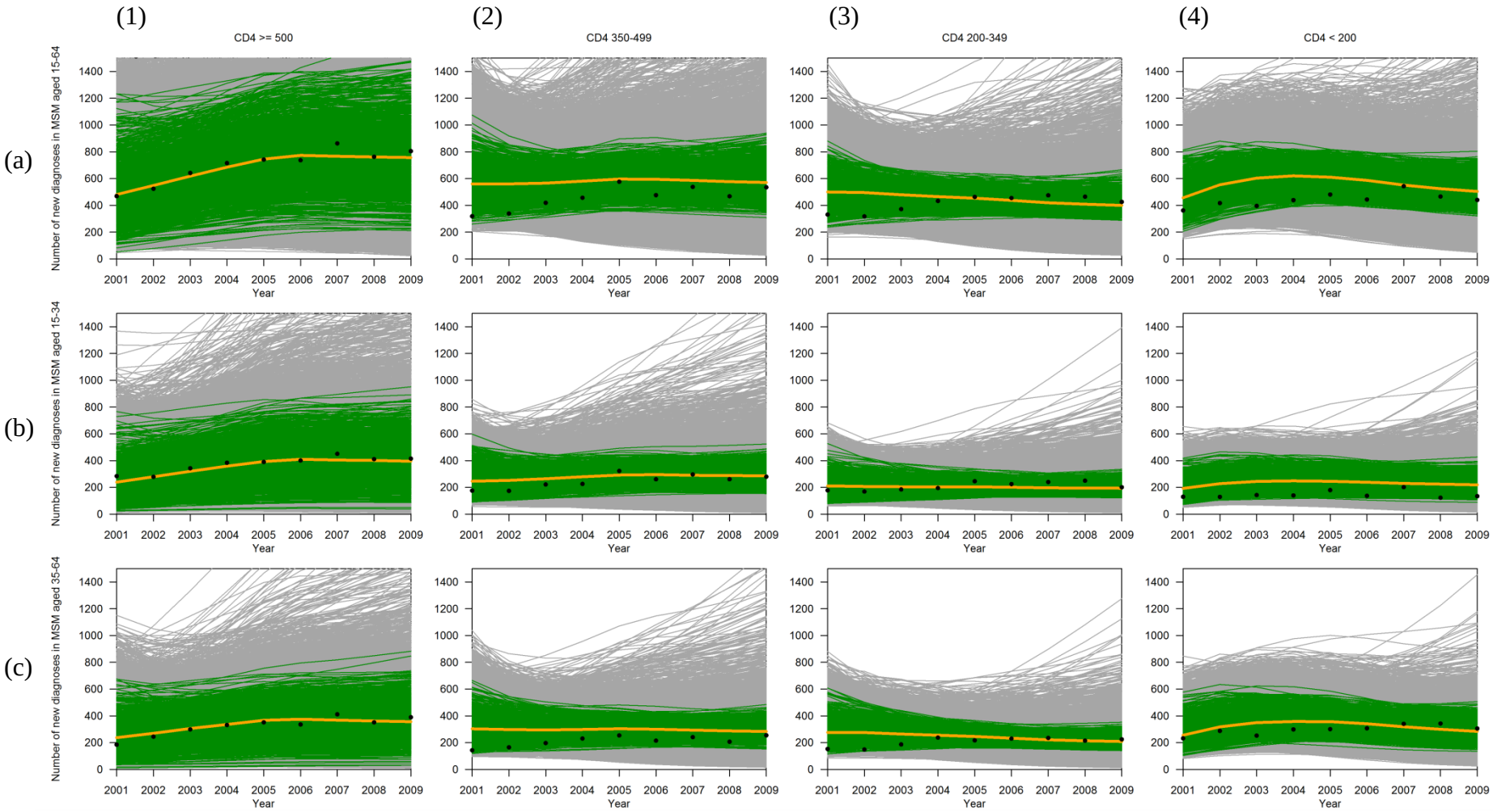
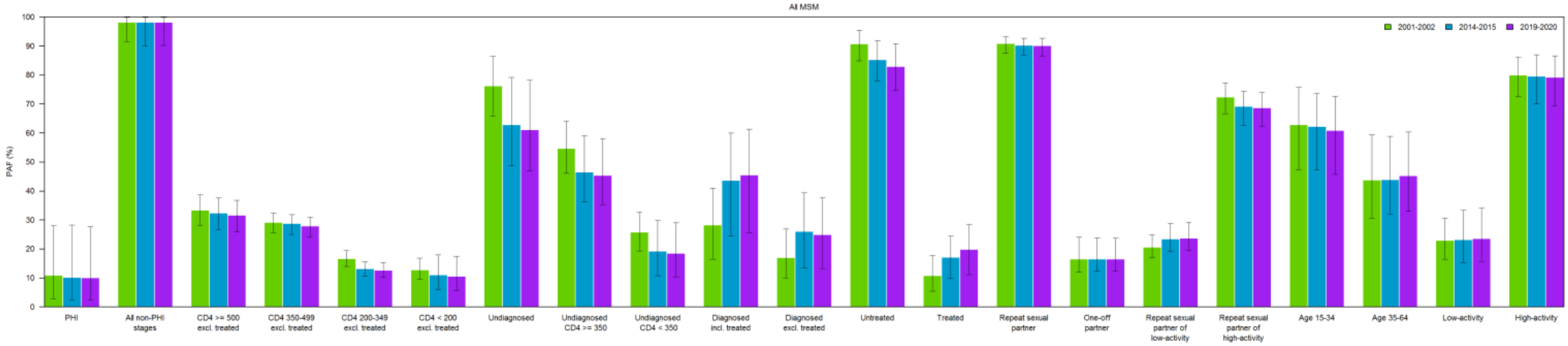
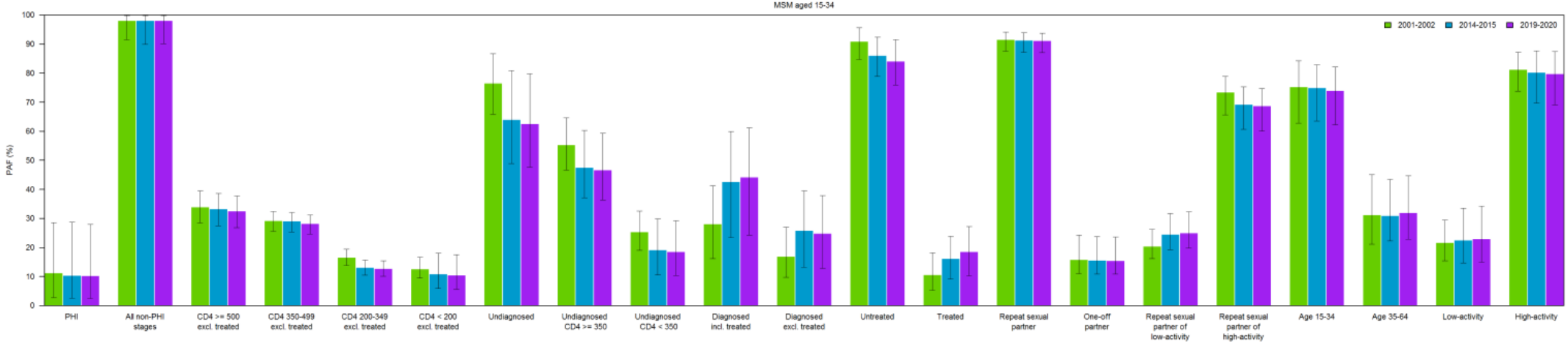


Figure S6: Comparison of the number of new diagnoses stratified by CD4 stages between model estimates and reported data for model validation

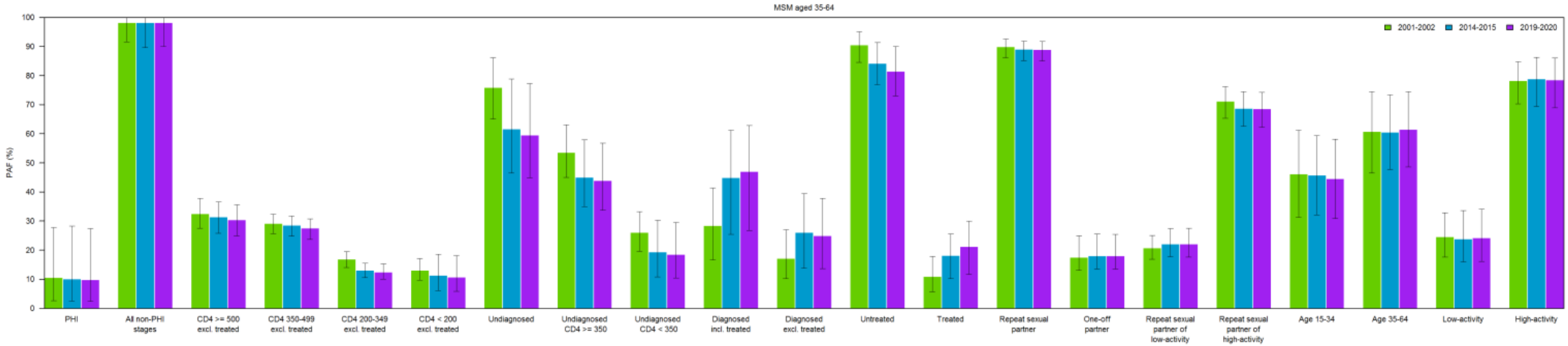
(a) All MSM



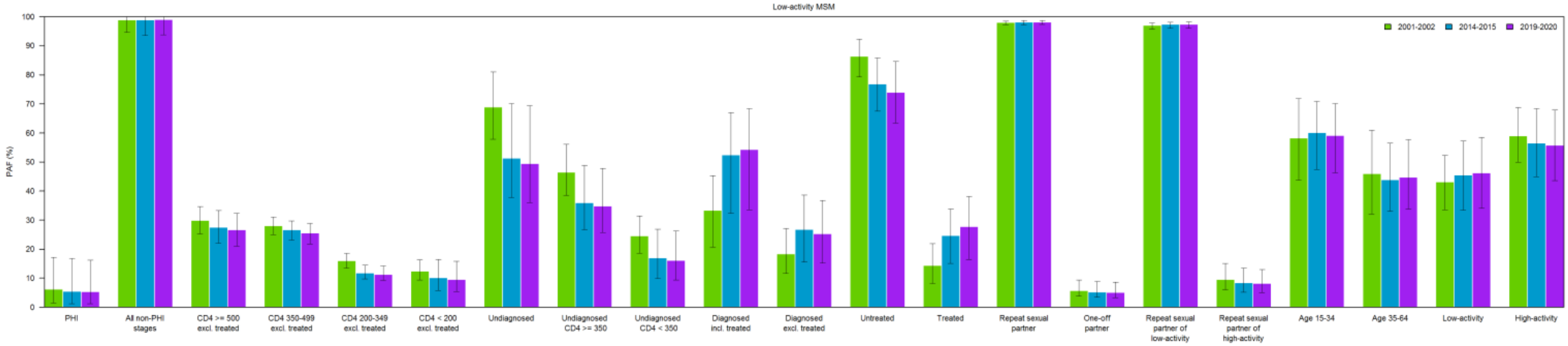
(b) MSM aged 15-34



(c) MSM aged 35-64



(d) Low-activity MSM



(e) High-activity MSM

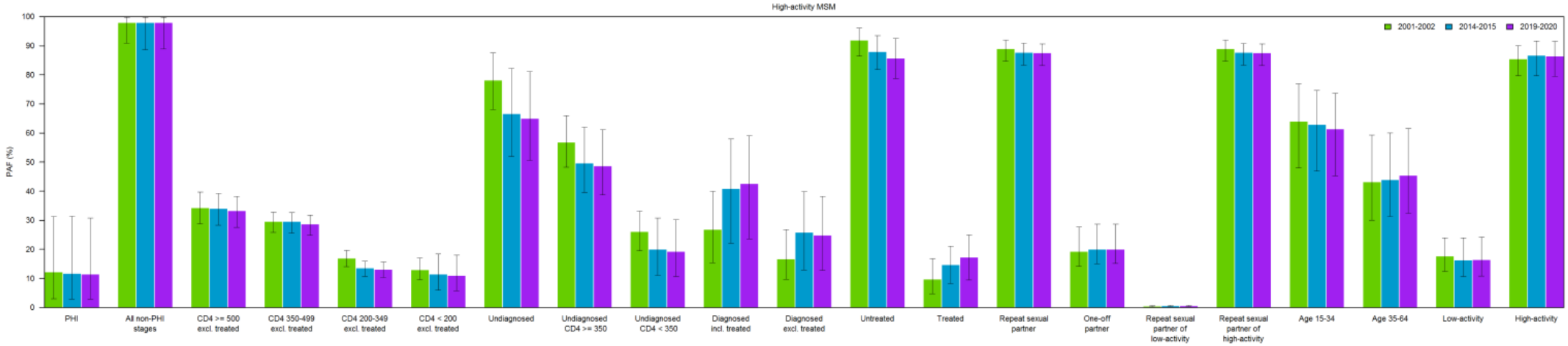


Figure S7: Population attributable fractions (PAFs) of various factors for HIV incidence among MSM in the UK stratified by age groups and sexual activity levels

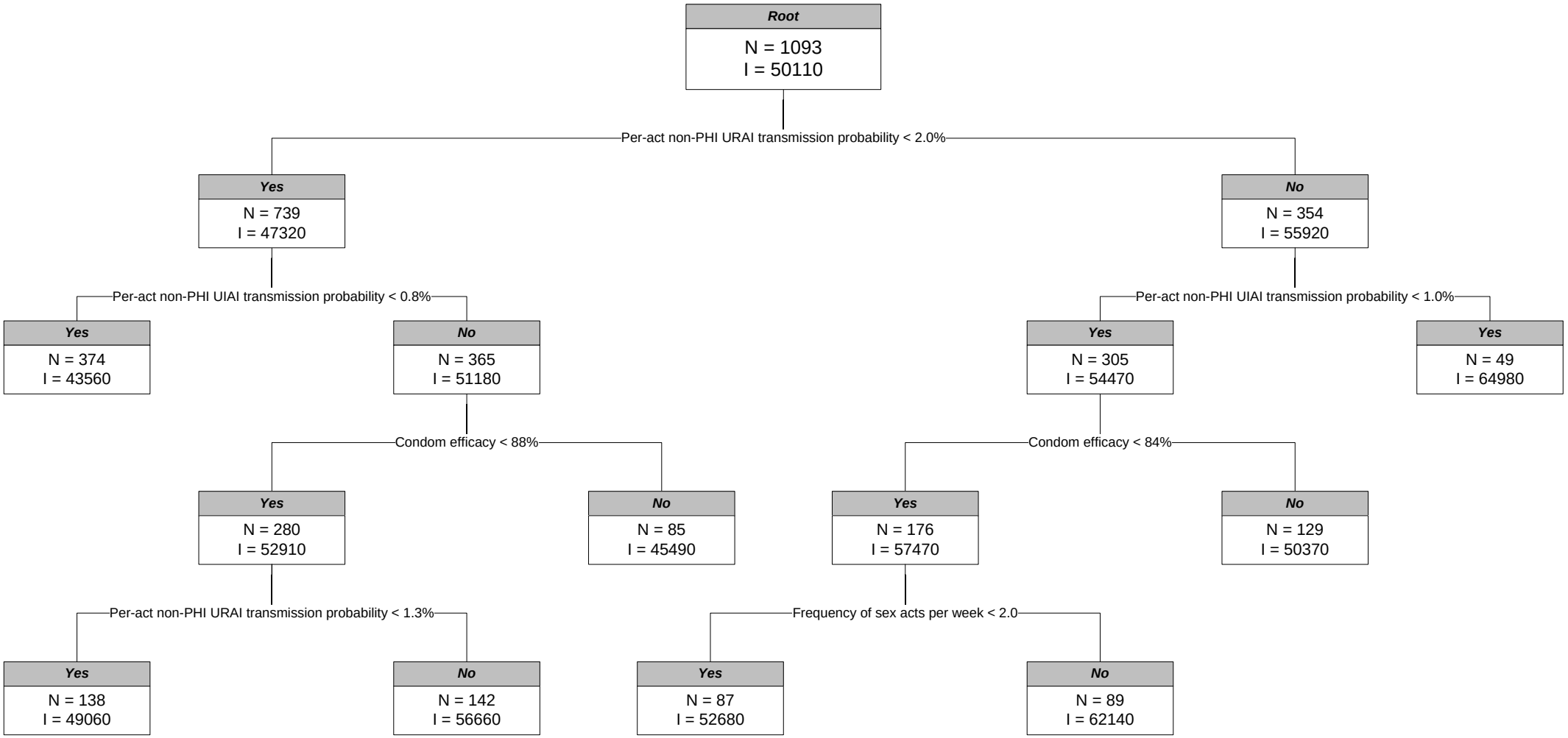


Figure S8: Model regression tree for sensitivity analysis

Supplementary figure legends

Figure S1

Each line illustrates changes in repeat sexual partnership formation rates per year of the corresponding classes for different adjustment weights ranging from no weight at all to full weight. The differences become hardly noticeable for the weights of five or above. The full weights were used in this study.

Figure S2

(a) Overall HIV prevalence, (b) Diagnosed HIV prevalence (not used for model fitting), (c) Undiagnosed HIV prevalence. The dashed lines showed ranges of the fitting data. Note that the upper limits of reported estimates were increased by 15% to compensate for difference of age range between the reported estimates and this study (15-59 vs. 15-64, respectively). The grey lines represent model estimates from all 20,000 sampled parameter sets. The green lines represent model estimates from 1,093 parameter sets that yielded both the overall and undiagnosed HIV prevalence within the range of the 2001–2009 fitting data, simultaneously. The yellow lines indicated median of the green-line estimates.

Figure S3

(a) Overall new HIV infections, (b) new HIV infections in MSM aged 15-34, (c) new HIV infections in MSM aged 35-64. The dots and dashed lines indicated the reported estimates and 95% credible intervals of annual numbers of new HIV infections in

MSM in England and Wales. The explanations of all other items in this figure are provided in the legends of Figure S2.

Figure S4

ART: antiretroviral treatment. The dots indicated the reported numbers of UK MSM on ART from 2001–2009. The explanations of all other items in this figure are provided in the legends of Figure S2.

Figure S5

(a) Overall new HIV diagnoses, (b) new HIV diagnoses in MSM aged 15-34, (c) new HIV diagnoses in MSM aged 35-64. The dots indicated the reported new HIV diagnoses in UK MSM from 2001–2009. The explanations of all other items in this figure are provided in the legends of Figure S2.

Figure S6

Rows (a), (b), and (c): Overall new HIV diagnoses, new HIV diagnoses in MSM aged 15-34, and new HIV diagnoses in MSM aged 35-64, respectively. Columns (1), (2), (3), and (4): New HIV diagnoses at CD4 $\geq$500, 350-499, 200-349, and <200 cells/μL, respectively. The dots indicated the reported new HIV diagnoses in UK MSM from 2001–2009. The explanations of all other items in this figure are provided in the legends of Figure S2.

Figure S7

PAF: Population attributable fractions; PHI: Primary HIV infection. The figure shows the impacts of removing the corresponding factors from all MSM on the incidence of HIV in (a) all MSM, (b) MSM aged 15-34, (c) MSM aged 35-64, (d) low-activity MSM, and (e) high-activity MSM. The green, blue, and purple bars represent 2001–2002, 2014–2015, and 2019–2020 PAF respectively. The error bars show the ranges between 2.5^{th} and 97.5^{th} percentiles of the estimated PAF.

Figure S8

The response variable was the total new HIV infections during 2001–2020. The predictor variables were all filtered parameters. The per-act non-PHI URAI and UIAI HIV transmission probability, condom efficacy, and the frequency of sex acts with a repeat sexual partner were included in the optimal tree. The number of simulations denoted by "N" and the total number of new infections denoted by "I" are shown inside each node. The root node contains 1,093 parameter sets and yields a total of 50,110 new infections. The splitting conditions are located on the routes between nodes. The nodes on the left of routes represent a case when the splitting condition is met, and vice versa for the right nodes. The tree expands until the terminal nodes of each branch are reached where the new infections varied from 43,560 to 64,980.

References

1. Rose KA, Smith EP, Gardner RH, Brenkert AL, Bartell SM. **Parameter Sensitivities, Monte-Carlo Filtering, and Model Forecasting under Uncertainty**. *Journal of Forecasting* 1991,**10**:117-133.
2. Punyacharoensin N, Edmunds WJ, De Angelis D, White RG. **Mathematical models for the study of HIV spread and control amongst men who have sex with men**. *Eur J Epidemiol* 2011,**26**:695-709.
3. Johnson AM, Mercer CH, Erens B, Copas AJ, McManus S, Wellings K, *et al.* **Sexual behaviour in Britain: partnerships, practices, and HIV risk behaviours**. *Lancet* 2001,**358**:1835–1842.
4. Dietz K. **On the transmission dynamics of HIV**. *Math Biosci* 1988,**90**:397–414.
5. Xiridou M, Geskus R, De Wit J, Coutinho R, Kretzschmar M. **The contribution of steady and casual partnerships to the incidence of HIV infection among homosexual men in Amsterdam**. *AIDS* 2003,**17**:1029–1038.
6. Xiridou M, Geskus R, de Wit J, Coutinho R, Kretzschmar M. **Primary HIV infection as source of HIV transmission within steady and casual partnerships among homosexual men**. *AIDS* 2004,**18**:1311–1320.
7. Powers KA, Ghani AC, Miller WC, Hoffman IF, Pettifor AE, Kamanga G, *et al.* **The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: a modelling study**. *Lancet* 2011,**378**:256-268.
8. Gazzard BG, Anderson J, Babiker A, Boffito M, Brook G, Brough G, *et al.* **British HIV Association Guidelines for the treatment of HIV-1-infected adults with antiretroviral therapy 2008**. *HIV Med* 2008,**9**:563–608.
9. Goodreau SM, Golden MR. **Biological and demographic causes of high HIV and sexually transmitted disease prevalence in men who have sex with men**. *Sex Transm Infect* 2007,**83**:458–462.
10. Cassels S, Menza TW, Goodreau SM, Golden MR. **HIV serosorting as a harm reduction strategy: evidence from Seattle, Washington**. *AIDS* 2009,**23**:2497-2506.
11. Johnson AM, Wadworth J, Wellings K, Field J. ***Sexual attitudes and lifestyles***. Oxford: Blackwell Scientific Press; 1994.
12. NatCen Social Research. **The National Survey of Sexual Attitudes and Lifestyles**. In; 2012.
13. Dodds JP, Mercey DE, Parry JV, Johnson AM. **Increasing risk behaviour and high levels of undiagnosed HIV infection in a community sample of homosexual men**. *Sex Transm Infect* 2004,**80**:236-240.
14. Elford J, Bolding G, Davis M, Sherr L, Hart G. **Trends in sexual behaviour among London homosexual men 1998-2003: implications for HIV prevention and sexual health promotion**. *Sex Transm Infect* 2004,**80**:451–454.
15. Health Protection Agency. **United Kingdom; New HIV Diagnoses data to end December 2011. Tables No.2:2011.** In. London: Health Protection Agency; 2011.

16. The UK Collaborative Group for HIV and STI Surveillance. **A Complex Picture. HIV and other Sexually Transmitted Infections**. In. London: Health Protection Agency, Centre for Infections; 2006.
17. Reidy WJ, Goodreau SM. **The role of commercial sex venues in the HIV epidemic among men who have sex with men**. *Epidemiology* 2010,**21**:349-359.
18. Office for National Statistics. **Population Estimates**. In; 2012.
19. Health Protection Agency. **Sexually transmitted infections and men who have sex with men in the UK: 2011 report**. In. London: Health Protection Agency; 2011.
20. Office for National Statistics. **Population Projections**. In; 2012.
21. Office for National Statistics. **Life Tables**. In; 2012.
22. May M, Gompels M, Delpech V, Porter K, Post F, Johnson M, *et al.* **Impact of late diagnosis and treatment on life expectancy in people with HIV-1: UK Collaborative HIV Cohort (UK CHIC) Study**. *BMJ* 2011,**343**:d6016.
23. Health Protection Agency. **Sexually transmitted infections and men who have sex with men in the UK: 2008 report**. In. London: Health Protection Agency; 2008.
24. Vittinghoff E, Douglas J, Judson F, McKirnan D, MacQueen K, Buchbinder SP. **Per-contact risk of human immunodeficiency virus transmission between male sexual partners**. *Am J Epidemiol* 1999,**150**:306–311.
25. Marks G, Crepaz N, Senterfitt JW, Janssen RS. **Meta-analysis of high-risk sexual behavior in persons aware and unaware they are infected with HIV in the United States: implications for HIV prevention programs**. *J Acquir Immune Defic Syndr* 2005,**39**:446-453.
26. Jin F, Crawford J, Prestage GP, Zablotska I, Imrie J, Kippax SC, *et al.* **Unprotected anal intercourse, risk reduction behaviours, and subsequent HIV infection in a cohort of homosexual men**. *AIDS* 2009,**23**:243-252.
27. Crepaz N, Marks G, Liau A, Mullins MM, Aupont LW, Marshall KJ, *et al.* **Prevalence of unprotected anal intercourse among HIV-diagnosed MSM in the United States: a meta-analysis**. *AIDS* 2009,**23**:1617–1629.
28. Macdonald N, Elam G, Hickson F, Imrie J, McGarrigle CA, Fenton KA, *et al.* **Factors associated with HIV seroconversion in gay men in England at the start of the 21st century**. *Sex Transm Infect* 2008,**84**:8-13.
29. Ford K, Sohn W, Lepkowski J. **American adolescents: sexual mixing patterns, bridge partners, and concurrency**. *Sex Transm Dis* 2002,**29**:13–19.
30. Baggaley RF, White RG, Boily M-C. **HIV transmission risk through anal intercourse: systematic review, meta-analysis and implications for HIV prevention**. *Int J Epidemiol* 2010:dyq057.
31. Jin F, Jansson J, Law M, Prestage GP, Zablotska I, Imrie JC, *et al.* **Per-contact probability of HIV transmission in homosexual men in Sydney in the era of HAART**. *AIDS* 2010,**24**:907–913.
32. Smith RJ, Blower SM. **Could disease-modifying HIV vaccines cause population-level perversity?** *The Lancet Infectious Diseases* 2004,**4**:636–639.
33. Boily MC, Baggaley RF, Wang L, Masse B, White RG, Hayes RJ, *et al.* **Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies**. *The Lancet Infectious Diseases* 2009,**9**:118–129.

34. Doerner R, McKeown E, Nelson S, Anderson J, Low N, Elford J. **Circumcision and HIV Infection among Men Who Have Sex with Men in Britain: The Insertive Sexual Role**. *Arch Sex Behav* 2013.
35. Bonell C, Weatherburn P, Hickson F. **Sexually transmitted infection as a risk factor for homosexual HIV transmission: a systematic review of epidemiological studies**. *Int J STD AIDS* 2000,**11**:697-700.
36. Jin F, Prestage GP, Imrie J, Kippax SC, Donovan B, Templeton DJ, *et al.* **Anal sexually transmitted infections and risk of HIV infection in homosexual men**. *J Acquir Immune Defic Syndr* 2010,**53**:144-149.
37. Pinkerton SD, Abramson PR. **Effectiveness of condoms in preventing HIV transmission**. *Soc Sci Med* 1997,**44**:1303-1312.
38. Davis KR, Weller SC. **The effectiveness of condoms in reducing heterosexual transmission of HIV**. *Fam Plann Perspect* 1999,**31**:272-279.
39. Weller S, Davis K. **Condom effectiveness in reducing heterosexual HIV transmission**. *Cochrane Database Syst Rev* 2002:CD003255.
40. Presanis AM, Gill ON, Chadborn TR, Hill C, Hope V, Logan L, *et al.* **Insights into the rise in HIV infections, 2001 to 2008: a Bayesian synthesis of prevalence evidence**. *AIDS* 2010,**24**:2849-2858.
41. Presanis AM, De Angelis D, Goubar A, Gill ON, Ades AE. **Bayesian evidence synthesis for a transmission dynamic model for HIV among men who have sex with men**. *Biostatistics* 2011,**12**:666-681.
42. Health Protection Agency. **HIV in the United Kingdom: 2011 Report**. In. London: Health Protection Agency; 2011.
43. Fox J, White PJ, Macdonald N, Weber J, McClure M, Fidler S, *et al.* **Reductions in HIV transmission risk behaviour following diagnosis of primary HIV infection: a cohort of high-risk men who have sex with men**. *HIV Med* 2009,**10**:432-438.
44. Bansi L, Sabin C, Delpech V, Hill T, Fisher M, Walsh J, *et al.* **Trends over calendar time in antiretroviral treatment success and failure in HIV clinic populations**. *HIV Med* 2010,**11**:432-438.
45. Birrell PJ, Presanis AM, De Angelis D. **Multi-state models of HIV progression in homosexual men: an application to the CASCADE collaboration. Technical report.** In. Cambridge: MRC Biostatistics Unit; 2012.
46. Hollingsworth TD, Anderson RM, Fraser C. **HIV-1 transmission, by stage of infection**. *J Infect Dis* 2008,**198**:687-693.
47. Unlinked Anonymous Surveys Steering Group. **Prevalence of HIV and hepatitis infections in the United Kingdom 2000**. In. London: Department of Health; 2001.
48. Birrell PJ, Gill ON, Delpech VC, Brown AE, Desai S, Chadborn TR, *et al.* **HIV incidence in men who have sex with men in England and Wales 2001-10: a nationwide population study**. *Lancet Infect Dis* 2013,**13**:313-318.
49. Blower SM, Dowlatabadi H. **Sensitivity and Uncertainty Analysis of Complex Models of Disease Transmission: An HIV Model, as an Example**. *International Statistical Review* 1994,**62**:229–243.
50. Health Protection Agency. **HIV in the United Kingdom: 2010 Report**. In. London: Health Protection Agency; 2010.
51. Hastie T, Tibshirani R, Friedman J. ***The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition***. 2 ed. New York: Springer Sceince+Business Media; 2009.