Mathematical Biosciences and Engineering

MBE, 23(8): 2307–2322.
DOI: 10.3934/mbe.2026084
Received: 10 March 2026
Revised: 17 June 2026
Accepted: 07 July 2026
Published: 22 July 2026

https://www.aimspress.com/journal/MBE

Research article

Why accurate risk stratification matters in HIV population modeling

Dobromir Dimitrov[1,2,*], Laura Matrajt[1,2], Xinze Ren[2], Sarah E. Stansfield[1] and Mia Moore[1]

[1] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA 98109, USA
[2] Department of Applied Mathematics, University of Washington, Seattle, WA 98109, USA

* **Correspondence:** Email: ddimitro@fredhutch.org; Tel: +4258948537.

Supplementary

Sexual mixing

1. For a model with 6 risk groups we have assumed the following:
   a) Individuals tend to choose partners with similar risk level and therefore contact matrix $\boldsymbol{C^6}$ is diagonally dominated.

$$C^6 = \begin{pmatrix} 0.80 & 0.10 & 0.05 & 0.05 & 0.00 & 0.00 \\ 0.05 & 0.80 & 0.10 & 0.05 & 0.00 & 0.00 \\ 0.00 & 0.05 & 0.80 & 0.10 & 0.05 & 0.00 \\ 0.00 & 0.00 & 0.05 & 0.80 & 0.10 & 0.05 \\ 0.00 & 0.00 & 0.05 & 0.05 & 0.80 & 0.10 \\ 0.00 & 0.00 & 0.05 & 0.05 & 0.10 & 0.80 \end{pmatrix}$$

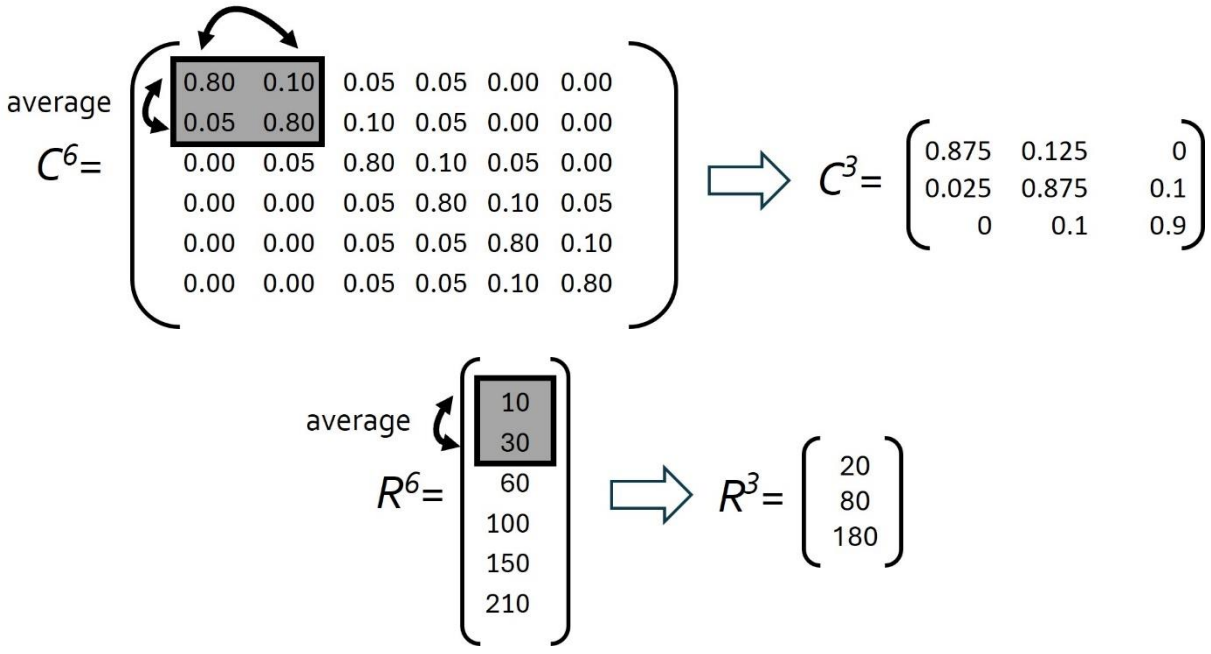
   b) There are significant (up to 20-fold) differences in sexual activity between risk groups. Instead of tracking all types of sexual encounters separately, which would increase model complexity, we opted out for measuring the HIV transmission risk for each group with a single, unifying parameter in vector R^6. One may think of it as the equivalent of the number of unprotected vaginal sex acts individuals from this group have annually. Therefore, the sexual activity of 10 for the lowest risk group, does not mean that they have only sex acts per year but only 10 of those were unprotected. Similarly, individuals in the highest risk group does not need to have

210 acts per year but could have frequent unprotected anal sex which is associated with 10-hold increase in HIV acquisition probability per act.

$$R^6 = \begin{pmatrix} 10 \\ 30 \\ 60 \\ 100 \\ 150 \\ 210 \end{pmatrix}$$

c) Fixed acquisition probability per single act of exposure $\boldsymbol{\beta = 0.002}$

2. To ensure comparability across models with different number of risk groups we reduced the size of the contact matrix $\boldsymbol{C}$ and the sexual activity vector $\boldsymbol{R}$ in the following way (illustrated for transition from 6 to 3 groups):
   a. Reducing contact matrix $\boldsymbol{C}$: To transition from C^6 to C^3 risk groups we aggregate risk groups in pairs by combining (adding) the contacts in the corresponding columns to estimate the total proportion of partners from the aggregated group and then averaging the rows corresponding to the aggregated risk groups
   b. Reducing sexual activity vector $\boldsymbol{R}$: To transition from R^6 to R^3 risk groups we aggregate risk groups in pairs by averaging the activity corresponding to the aggregated risk groups.



Simulation setup

All simulations are initiated with a population of 1000 individuals equally distributed across risk groups and ran for 250 years. All simulations with HIV were initially seeded with 1% people living with HIV (PLHIV). Constant number of individuals (36) join the population each year and it is assumed that they are sexually active for 30 years on average.

Risk-progression scenarios

We analyzed three potential risk progression scenarios:

i) *fixed risk*, in which individuals who join the population (at sexual debut) are equally distributed across risk groups and remain in them for life. Equal departure rates (μ_i = ***1/30*** annually) are assumed for all risk groups.

ii) *age-based risk progression*, assuming that younger age is a strong predictor of higher risk. In this scenario all individuals who join the population start in the highest risk group and gradually progress toward lower risk groups (at rate ***n/30*** annually, where ***n*** is the number of risk groups) with increasing age. Individuals depart (age out) from the population only from the lowest risk group (at rate μ_1 = ***n/30*** annually).

iii) *mixed risk progression*, in which individuals gradually progress toward lower risk groups at half of the rate from the age-based scenario (***n/60*** annually, where ***n*** is the number of risk groups). Different recruitment and departure rates are assumed for all risk groups to ensure the same number of individuals joining the population annually and equal representation of all risk groups in absence of HIV.

Parameter selections for different scenarios

As mentioned in the Methods of the paper, recruitment and departure rates for all scenarios are calculated to satisfy 2 conditions: i) constant annual recruitment and ii) all risk groups being equally represented in the equilibrium population in absence of HIV. In addition to these 2 conditions we wanted to ensure that the parameter sets for models with different numbers of risk groups are reasonably well synchronized (aggregation rule). For instance, given that the 3-group model aggregates risk groups from the 6-group model in pairs, it is reasonable to assume that the number of annual recruits in the 3-group model should match the sum of risk group pairs from the 6-group model.

The model selection was achieved by the following procedure:

1) **Selecting recruitment rates (Λ_i^n)** We selected parameters that ensure that a total of 36 individuals join the population each year. That condition alone, determined the recruitment rates in the fixed and age-based scenario, where new recruits were added to risk groups in equal numbers (Λ_i^n =*36/n*) or all joining the youngest and riskiest group ($\Lambda_n^n = 036,\ \Lambda_i^n = 0, i = 1..n-1$). Parameter selection for the mixed risk scenario was done ad-hoc to guarantee that aggregation rule is satisfied when reducing groups from 6 to 3, 2 and 1. More specially, $\Lambda_i^3 = \Lambda_{2i-1}^6 + \Lambda_{2i}^6,\ \ i = 1..3$ and $\Lambda_i^2 = \Lambda_{3i-2}^6 + \Lambda_{3i-1}^6 + \Lambda_{3i}^6,\ \ i = 1, 2.$

2) **Selecting departure rates (μ_i^n)** We equalize the right-hand side of the equations to zero in absence of HIV and solve for μ_i^n using that all risk groups should be equally represented in the equilibrium distribution ($S_1 = S_2 = \cdots = S_n$).

Additional results

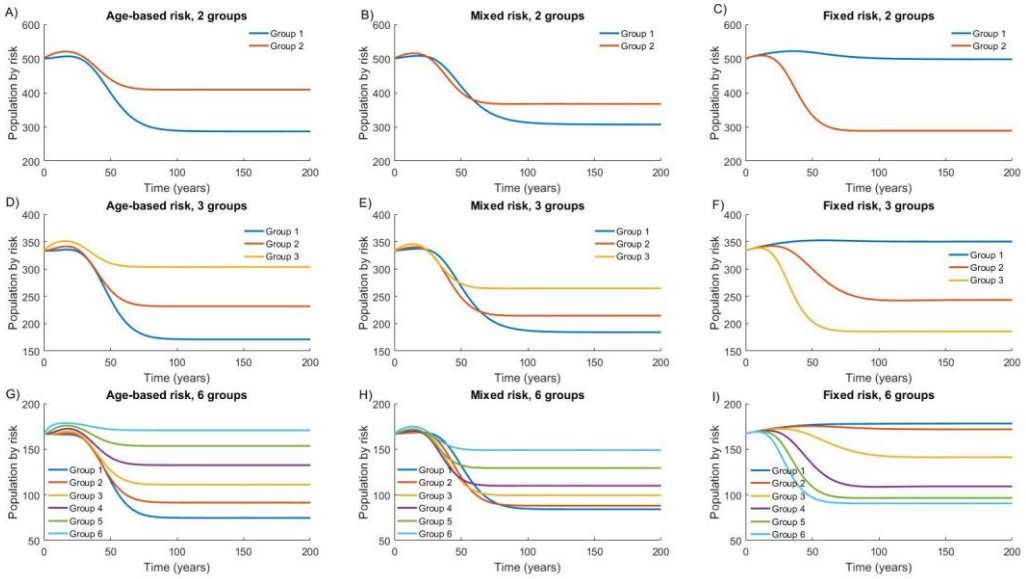


Figure S1. Population distribution by risk for models with different number of risk groups. Groups are numbered in increasing risk order. Note: different y-scale for each panel row.

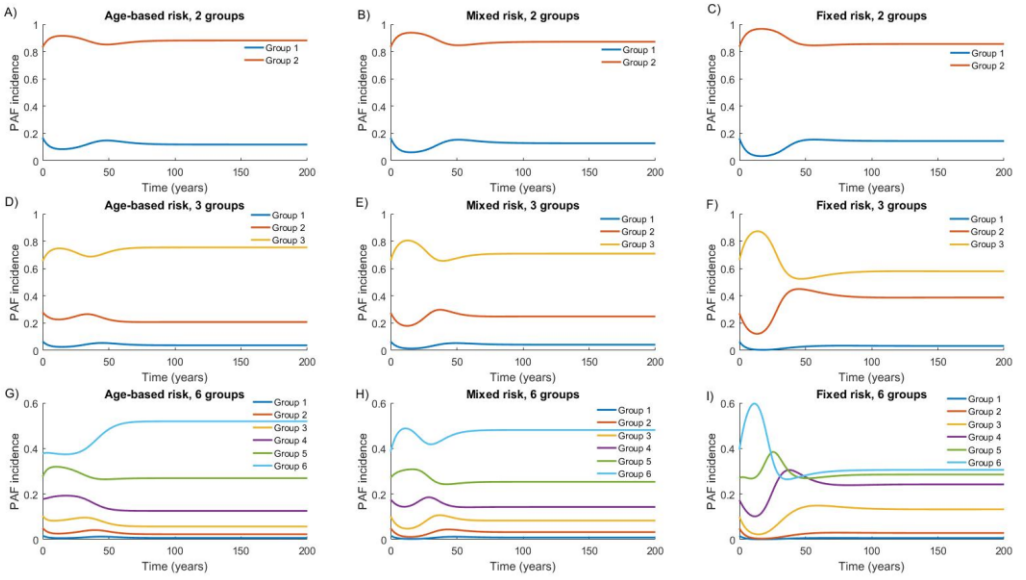


Figure S2. Population attributable fraction of HIV incidence (PAFi) by risk for models with different number of risk groups. PAF is calculated as the proportion of new infections annually which occur among each risk group. Groups are numbered in increasing risk order. Note: different y-scale for each panel row.

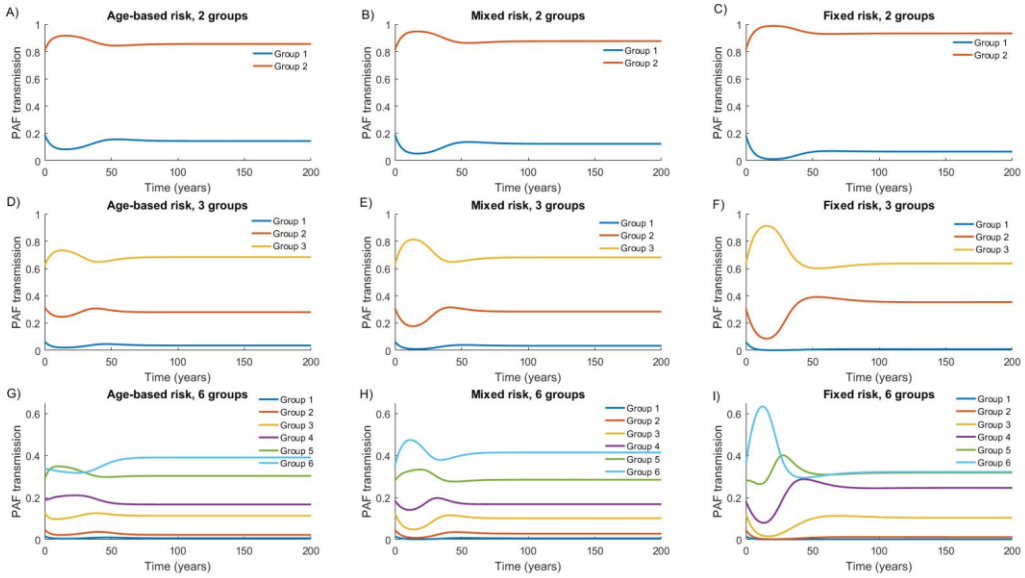


Figure S3. Population attributable fraction of HIV transmission (PAFt) by risk for models with different number of risk groups. PAF is calculated as the proportion of new infections annually which are transmitted by infected individuals from each risk group. Groups are numbered in increasing risk order. Note: different y-scale for each panel row.

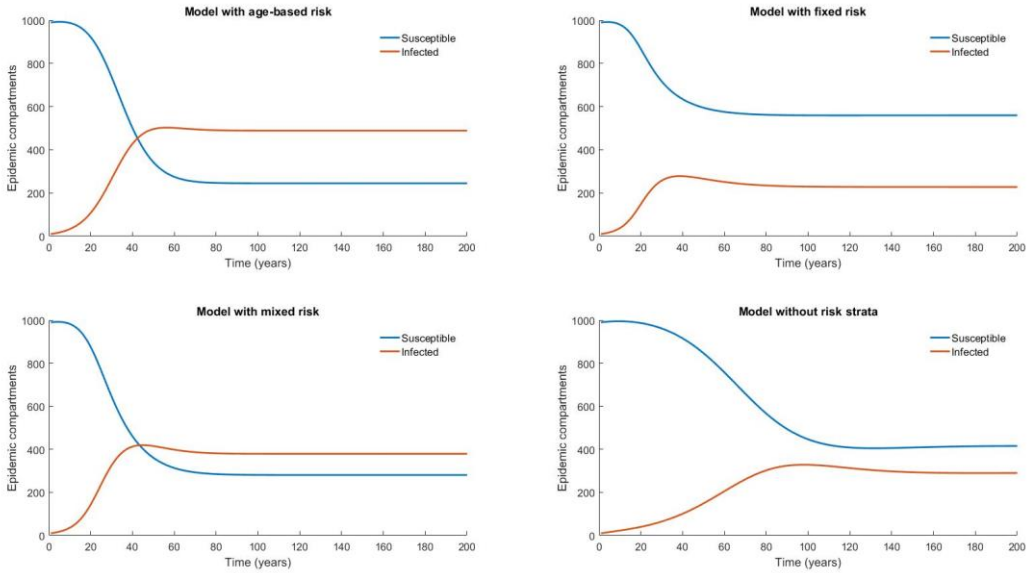


Figure S4. Population distribution by HIV status for models with 6 risk groups and different risk progression mechanisms compared to a model without risk stratification.