

# HIV transmission model to investigate the impact of targeted interventions in the Dutch Caribbean

*G. Rozhnova*

In these notes we will specify the HIV metapopulation model we are studying and outline the methods we use to study it. Our aim in the first three sections is to give a non-technical introduction to the methods we use; technical descriptions already exist in several published papers. We will explain the formalism in detail for the usual (one-country) model in such a way that the generalisation to two, or  $n$ , countries is straightforward.

As discussed in project meetings the countries are

- (i) Aruba, Curaçao, Sint Maarten, Bonaire, St. Eustatius and Saba (6 islands of interest);
- (ii) United States, Netherlands (two developed countries);
- (iii) Colombia, Venezuela (two developing countries);
- (iv) Suriname;
- (v) A grouping of other Caribbean islands (Antigua and Barbuda, Barbados, Dominica, Dominican Republic, Grenada, Guyana, Haiti, Honduras, Jamaica etc).

These notes describe

1. The one-country stochastic SIR model;
2. The two-country stochastic SIR model;
3. The  $n$ -country stochastic SIR model;
4. The  $n$ -country stochastic HIV model.

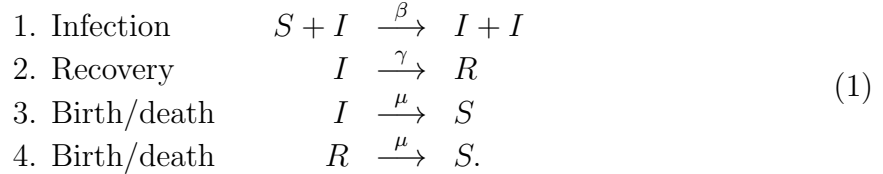
## 1 The One-Country Stochastic SIR model

### 1.1 Formulation of the individual-based model and the master equation

We begin with the simple SIR model — or in the context of our project, the one-country SIR model. We define it in terms of “reactions” between the  $N$  constituents of the system, which in this case are  $N$  individuals which are either of type  $S$  (susceptible), type  $I$  (infected) or type  $R$  (recovered). These individuals are born and die at the same rate,

and in addition these events are linked, so that the total population  $N$  remains constant. Thus instead of having distinct birth and death events, only combined birth/death events occur. Since all newly-born individuals are susceptible, these events correspond to an infected individual being replaced by a susceptible individual or a recovered individual being replaced by a susceptible individual. A susceptible individual dying results in no change in the state of the system, since they are simply replaced by another susceptible individual.

The reactions that define the one-country SIR model are then:



Here  $\beta, \gamma$  and  $\mu$  are respectively the rate of infection, the rate of recovery and the rate of birth/death. The reactions will be labelled by  $\alpha$  ( $= 1, 2, 3, 4$ ).

Frequently, birth and death processes are assumed to happen at the same rate, but remain as distinct events. This still results in fluctuations in the total population size, even if the average population size remains constant. By linking these events at the stochastic level, the population size remains constant at any system size, so that we can still eliminate the variable relating to recovered individuals:  $R = N - I - S$ . This means that there are only two variables which define the state of the system: the number of susceptible individuals at a given time which we will denote by  $S$  and the number of infected individuals at a given time which we will denote by  $I$ . The use of the same symbol for an individual of a particular type and the number of individuals of that type should cause no confusion. The general state will then be denoted by  $\mathbf{n} = (S, I)$ .

The four processes (1) in the SIR model which cause transitions to a new state have the following transition rates:

$$\begin{array}{lll}
1. \text{ Infection} & T_1(S - 1, I + 1 | S, I) & = \beta SI / N \\
2. \text{ Recovery} & T_2(S, I - 1 | S, I) & = \gamma I \\
3. \text{ Birth/death} & T_3(S + 1, I - 1 | S, I) & = \mu I \\
4. \text{ Birth/death} & T_4(S + 1, I | S, I) & = \mu (N - S - I).
\end{array} \tag{2}$$

We use the convention whereby the initial state is on the right and the final state is on the left, so that  $T(\mathbf{n} | \mathbf{n}')$  represents the transition rate from  $\mathbf{n}'$  to  $\mathbf{n}$ .

The transition rates (2) define the individual-based SIR model, and when substituted in the master equation, give an equation for the probability,  $P_{\mathbf{n}}(t)$ , of finding the system in the state  $\mathbf{n}$  at time  $t$ . Since the master equation can only be solved for very simple linear systems, we will simulate the model dynamics using the Gillespie's algorithm or its extensions or modifications.

## 2 The Two-Country Stochastic SIR model

Having discussed in detail the formalism and general procedure in the one-country case in a way that naturally generalises, the description of the two-country case can be more

concise. What is novel is the nature of the interactions between the two countries, which is what we discuss first of all.

## 2.1 Form of the interaction between the two countries

The interaction between the two countries is only reflected in the first type of reaction in Eq. (1); the other three remain essentially unchanged since they involve only one individual, and so only one country.

The number of individuals in the three classes born in country  $j$  ( $j = 1, 2$ ) will be denoted by  $S_j$ ,  $I_j$  and  $R_j$  respectively. We again assume that births and deaths are coupled at the individual level, so that when an individual dies another (susceptible) individual is born. Therefore the number of recovered individuals is not an independent variable:  $R_j = N_j - S_j - I_j$ , where  $j = 1, 2$ .

To model HIV transmission in the Dutch Caribbean we will not focus on specific individuals and we will not be concerned with the precise movements of migrants between the countries — the frequency or duration of their migration ‘events’ — but only the fraction of the population which is migrated from its home country at any given time. It is precisely this fraction (generally fairly stable across a number of years) that we estimated from the migration data in the second year of the project. For migrants from country  $j$  to country  $k$  we will denote this fraction by  $f_{kj}$ . It follows that the number of individuals in country 1 is  $M_1 = (1 - f_{21})N_1 + f_{12}N_2$  and in country 2 is  $M_2 = (1 - f_{12})N_2 + f_{21}N_1$ .

We further denote  $N_j = (1 - f_{kj})N_j + f_{kj}N_j = S_j + I_j + R_j = S_{jj} + I_{jj} + R_{jj} + S_{kj} + I_{kj} + R_{kj}$ , where the first three terms denote the residents in country  $j$  and the last three terms denote the migrants from country  $j$  ( $j = 1, 2$ ) to country  $k$  ( $k = 1, 2$ ). Therefore, we have  $(1 - f_{kj})N_j = S_{jj} + I_{jj} + R_{jj}$  and  $f_{kj}N_j = S_{kj} + I_{kj} + R_{kj}$ ;  $S_j = S_{jj} + S_{kj}$ ,  $I_j = I_{jj} + I_{kj}$ , and  $R_j = R_{jj} + R_{kj}$ .

We will assume that the birth/death rate and the recovery rate are the same in both countries, but that the infection rates are country dependent:  $\beta_1$  in country 1 and  $\beta_2$  in country 2. When considering the HIV model instead of the SIR model, as we are in this section, the HIV infection rates can be country-dependent, but HIV disease progression parameters are the same in both countries.

There are four different types of infection events.

- (i) Infective residents of one country (say  $j$ ) can infect susceptible residents of the same country. The rate for this to occur is  $\beta_j S_{jj} I_{jj}/M_j$ .
- (ii) Infective migrants from country  $k$  to country  $j$  can infect susceptible residents in their home country  $j$ . The rate for this to occur is  $\beta_j S_{jj} I_{jk}/M_j$ .
- (iii) Infective residents in their home country (say  $k$ ) can infect susceptible migrants from the other country ( $j$ ). The rate for this to occur is  $\beta_k S_{kj} I_{kk}/M_k$ .
- (iv) Infective migrants can infect susceptible migrants away from their home country  $j$ . The rate for this to occur is  $\beta_k S_{kj} I_{kj}/M_k$ .

Adding these rates together we obtain the total transition rate for infection of  $S_j$  individuals as

$$\begin{aligned} & \beta_j S_{jj} \frac{I_{jj}}{M_j} + \beta_j S_{jj} \frac{I_{jk}}{M_j} + \beta_k S_{kj} \frac{I_{kk}}{M_k} + \beta_k S_{kj} \frac{I_{kj}}{M_k} = \\ & \beta_j S_{jj} \frac{(I_{jj} + I_{jk})}{M_j} + \beta_k S_{kj} \frac{(I_{kk} + I_{kj})}{M_k} \end{aligned} \quad (3)$$

Equation (3) gives us the generalisation of the infection rate to the two-country case. There are now 16 processes rather than the four processes (1) of the one-country SIR model which cause transitions from one state to another. They are

1. Infection of  $S_{11}$   $T_1(S_{11} - 1, I_{11} + 1 | S_{11}, I_{11}) = \beta_1 S_{11} \frac{(I_{11} + I_{12})}{M_1}$
2. Infection of  $S_{21}$   $T_2(S_{21} - 1, I_{21} + 1 | S_{21}, I_{21}) = \beta_2 S_{21} \frac{(I_{22} + I_{21})}{M_2}$
3. Infection of  $S_{22}$   $T_3(S_{22} - 1, I_{22} + 1 | S_{22}, I_{22}) = \beta_2 S_{22} \frac{(I_{22} + I_{21})}{M_2}$
4. Infection of  $S_{12}$   $T_4(S_{12} - 1, I_{12} + 1 | S_{12}, I_{12}) = \beta_1 S_{12} \frac{(I_{11} + I_{12})}{M_1}$
5. Recovery of  $I_{11}$   $T_5(S_{11}, I_{11} - 1 | S_{11}, I_{11}) = \gamma I_{11}$
6. Recovery of  $I_{21}$   $T_6(S_{21}, I_{21} - 1 | S_{21}, I_{21}) = \gamma I_{21}$
7. Recovery of  $I_{22}$   $T_7(S_{22}, I_{22} - 1 | S_{22}, I_{22}) = \gamma I_{22}$
8. Recovery of  $I_{12}$   $T_8(S_{12}, I_{12} - 1 | S_{12}, I_{12}) = \gamma I_{12}$
9. Birth/death in country 1  $T_9(S_{11} + 1, I_{11} - 1 | S_{11}, I_{11}) = \mu I_{11}$
10. Birth/death in country 1  $T_{10}(S_{21} + 1, I_{21} - 1 | S_{21}, I_{21}) = \mu I_{21}$
11. Birth/death in country 2  $T_{11}(S_{22} + 1, I_{22} - 1 | S_{22}, I_{22}) = \mu I_{22}$
12. Birth/death in country 2  $T_{12}(S_{12} + 1, I_{12} - 1 | S_{12}, I_{12}) = \mu I_{12}$
13. Birth/death in country 1  $T_{13}(S_{11} + 1, I_{11} | S_{11}, I_{11}) = \mu[(1 - f_{21})N_1 - S_{11} - I_{11}]$
14. Birth/death in country 1  $T_{14}(S_{21} + 1, I_{21} | S_{21}, I_{21}) = \mu[f_{21}N_1 - S_{21} - I_{21}]$
15. Birth/death in country 2  $T_{15}(S_{22} + 1, I_{22} | S_{22}, I_{22}) = \mu[(1 - f_{12})N_2 - S_{22} - I_{22}]$
16. Birth/death in country 2  $T_{16}(S_{12} + 1, I_{12} | S_{12}, I_{12}) = \mu[f_{12}N_2 - S_{12} - I_{12}].$

(4)

Note that we have not listed all the state variables as arguments of the transition rates  $T_\alpha(\dots|\dots)$  — only those which are most relevant to the reaction under consideration.

Introducing  $\mathbf{n} = (S_{11}, S_{21}, S_{22}, S_{12}, I_{11}, I_{21}, I_{22}, I_{12})$ , we may now set up the master equation in the same way as for the one-country case and simulate the dynamics of the model using e.g. Gillespie's algorithm. Note that the total number of independent variables is # of countries  $\times$  # of disease states (e.g. infected, susceptible)  $\times$  # groups of individuals (i.e. residents, migrants).

## 2.2 The Extended Two-Country Stochastic SIR model

So far we have assumed that the infection rates are country dependent:  $\beta_1$  in country 1 and  $\beta_2$  in country 2. In the context of an HIV model it means that HIV infection rates are country-dependent. However, even within one country the rates of infection can be different between different groups of individuals (i.e. residents, migrants). We could take this into account by introducing infection rate  $\beta_{ij,ij}$  ( $i, j = 1, 2$ ), where the first pair of indices denotes the susceptible group and the second pair of indices denotes the infectives group. For example,  $\beta_{11,11}$  is the rate of infection of  $S_{11}$  (susceptible residents in country 1) by  $I_{11}$  (infective residents in country 1);  $\beta_{12,11}$  is the rate of infection of  $S_{12}$  (susceptible migrants from country 2 to country 1) by  $I_{11}$  (infective residents in country 1). In total, there are 8 different infection rates for the two-country SIR model.

The transition rates for infection then become

$$\begin{aligned}
1'. \text{ Infection of } S_{11} \quad T_1(S_{11} - 1, I_{11} + 1 | S_{11}, I_{11}) &= S_{11} \frac{(\beta_{11,11} I_{11} + \beta_{11,12} I_{12})}{M_1} \\
2'. \text{ Infection of } S_{21} \quad T_2(S_{21} - 1, I_{21} + 1 | S_{21}, I_{21}) &= S_{21} \frac{(\beta_{21,22} I_{22} + \beta_{21,21} I_{21})}{M_2} \\
3'. \text{ Infection of } S_{22} \quad T_3(S_{22} - 1, I_{22} + 1 | S_{22}, I_{22}) &= S_{22} \frac{(\beta_{22,22} I_{22} + \beta_{22,21} I_{21})}{M_2} \\
4'. \text{ Infection of } S_{12} \quad T_4(S_{12} - 1, I_{12} + 1 | S_{12}, I_{12}) &= S_{12} \frac{(\beta_{12,11} I_{11} + \beta_{12,12} I_{12})}{M_1}.
\end{aligned} \tag{5}$$

Why is this extension useful? If, for some reason, we believe that there is not much transmission between e.g. infective migrants from country 2 to country 1,  $I_{12}$ , and susceptible residents in country 1,  $S_{11}$ , we can easily work with this extended model by setting  $\beta_{11,12} = 0$ . In the simpler model this was not possible, because there was only one infection rate in country 1,  $\beta_1$ .

## 2.3 The Two-Country Stochastic SIR model for main and satellite countries

We would like to model HIV transmission in the 6 main countries (Aruba, Curaçao, Sint Maarten, Bonaire, St. Eustatius and Saba). The rest of the countries are satellite

countries (United States, Netherlands, Colombia, Venezuela, Suriname, and a grouping of other Caribbean islands), linked to the main countries by migration. We do not aim to model HIV transmission in satellite countries. For the two-country model two situations are possible

- (a) both countries are main countries;
- (b) country 1 is a main country and country 2 is a satellite country.

In case of (a), no simplification is possible. In case of (b), further simplification is possible because we do not intend to model HIV transmission in a satellite country. This means that we can set the rate of infection of  $S_{22}$  (susceptible residents in the satellite country),  $T_3$ , to zero, and drop  $S_{22}$  variable. In the same way, we are probably not interested in the rate of infection of  $S_{21}$  (susceptible migrants from the main country to the satellite country),  $T_2$ , because those infections occur in the satellite country. Therefore, it would seem that we only want to keep track of  $S_{11}$  (susceptible residents in the main country) and  $S_{12}$  (susceptible migrants from the satellite country to the main country). Both these types are part of the effective population size of the main country, where the impact of different interventions to prevent HIV spread will be investigated. The relevant transition rates (of infection) are

$$\begin{aligned}
 1'. \text{ Infection of } S_{11} \quad T_1(S_{11} - 1, I_{11} + 1 | S_{11}, I_{11}) &= S_{11} \frac{(\beta_{11,11}I_{11} + \beta_{11,12}I_{12})}{M_1} \\
 4'. \text{ Infection of } S_{12} \quad T_4(S_{12} - 1, I_{12} + 1 | S_{12}, I_{12}) &= S_{12} \frac{(\beta_{12,11}I_{11} + \beta_{12,12}I_{12})}{M_1}.
 \end{aligned} \tag{6}$$

Let us recall that  $M_1$  and  $M_2$  denote the number of individuals in the main and satellite countries, respectively,  $M_1 = (1 - f_{21})N_1 + f_{12}N_2$  and  $M_2 = (1 - f_{12})N_2 + f_{21}N_1$ . We can set  $f_{21} = 0$ , assuming that there are no migrants from the main country to the satellite country.  $M_1$  and  $M_2$  then become  $M_1 = N_1 + f_{12}N_2$  and  $M_2 = (1 - f_{12})N_2$ , so that  $M_1 + M_2 = N_1 + N_2$ . This should be simplified further to take into account the data on HIV prevalence in the satellite country.

### 3 The $n$ -Country Stochastic SIR model

We will be relatively brief in this section, and only outline the results, since the formalism and general procedure is as for the two-country case. The main difference is that there is a fifth type of infection event — in addition to those mentioned for two countries. This is due to the fact that infective migrants from country  $k$  can infect susceptible migrants from country  $j$  in country  $\ell$ , where  $j, k$  and  $\ell$  are all different. This is only possible when there are three or more countries.

It is not clear whether this type of interaction is necessary. In simple terms, it means that infective migrants from Bonaire could infect susceptible migrants from Aruba in Curaçao. While in theory this interaction exists, in reality its contribution to HIV transmission could be small. If this is true, the rates for this type of interactions could be set to zero.

### 3.1 Form of the interaction between the $n$ countries

The number of individuals in the three classes born in country  $j$  are denoted by  $S_j$ ,  $I_j$  and  $R_j = N_j - S_j - I_j$  as before, where now  $j = 1, \dots, n$ . We will also introduce the notation

$$f_j = \sum_{k \neq j} f_{kj}, \quad (7)$$

so that the number of individuals in country  $j$  may be written as

$$\begin{aligned} M_j &= \left[1 - \sum_{k \neq j} f_{kj}\right] N_j + \sum_{k \neq j} f_{jk} N_k \\ &= (1 - f_j) N_j + \sum_{k \neq j} f_{jk} N_k. \end{aligned} \quad (8)$$

As for the two country case, we will assume that the birth/death rate and the recovery rate are the same in both countries, but that the infection rate for country  $j$  is  $\beta_j$ . Later, we can extend this model as before.

As mentioned above, there are five different types of infection events:

- (i) Infective residents of one country (say  $j$ ) can infect susceptible residents of the same country. The rate for this to occur is  $\beta_j S_{jj} I_{jj}/M_j$ .
- (ii) Infective migrants from country  $k$  to country  $j$  can infect susceptible residents in their home country  $j$ . The rate for this to occur is  $\beta_j S_{jj} I_{jk}/M_j$ .
- (iii) Infective residents in their home country (say  $k$ ) can infect susceptible migrants from the other country ( $j$ ). The rate for this to occur is  $\beta_k S_{kj} I_{kk}/M_k$ .
- (iv) Infective migrants from country  $j$  infect susceptible migrants from  $j$  in country  $k$  ( $k \neq j$ ). The rate for this to occur is  $\beta_k S_{kj} I_{kj}/M_k$ .
- (v) Infective migrants from country  $k$  can infect susceptible migrants from country  $j$  in country  $\ell$  ( $\ell \neq j, k$ ). The rate for this to occur is  $\beta_\ell S_{\ell j} I_{\ell k}/M_\ell$ .

Adding these rates together we obtain the total transition rate for infection of  $S_j$  individuals as

$$\begin{aligned} &\beta_j S_{jj} \frac{I_{jj}}{M_j} + \sum_{k \neq j} \beta_j S_{jj} \frac{I_{jk}}{M_j} + \sum_{k \neq j} \beta_k S_{kj} \frac{I_{kk}}{M_k} + \sum_{k \neq j} \beta_k S_{kj} \frac{I_{kj}}{M_k} + \sum_{\ell \neq j, k} \sum_{k \neq j} \beta_\ell S_{\ell j} \frac{I_{\ell k}}{M_\ell} = \\ &\beta_j S_{jj} \frac{\left(I_{jj} + \sum_{k \neq j} I_{jk}\right)}{M_j} + \sum_{k \neq j} \beta_k S_{kj} \frac{(I_{kk} + I_{kj})}{M_k} + \sum_{\ell \neq j, k} \sum_{k \neq j} \beta_\ell S_{\ell j} \frac{I_{\ell k}}{M_\ell}. \end{aligned} \quad (9)$$

In the  $n$ -countries case we have

$$\begin{aligned}
N_j &= \left(1 - \sum_{k \neq j} f_{kj}\right) N_j + \sum_{k \neq j} f_{kj} N_j \\
&= (1 - f_j) N_j + f_j N_j \\
&= S_j + I_j + R_j \\
&= (S_{jj} + I_{jj} + R_{jj}) + \sum_{k \neq j} (S_{kj} + I_{kj} + R_{kj}), \tag{10}
\end{aligned}$$

where the first three terms denote the residents in country  $j$  and the last three terms denote the migrants from country  $j$  to country  $k \neq j$  ( $j, k = 1, 2, \dots, n$ ). We have

$$\begin{aligned}
(1 - f_j) N_j &= \left(1 - \sum_{k \neq j} f_{kj}\right) N_j = S_{jj} + I_{jj} + R_{jj}, \\
f_j N_j &= \sum_{k \neq j} f_{kj} N_j = \sum_{k \neq j} (S_{kj} + I_{kj} + R_{kj}), \\
S_j &= S_{jj} + \sum_{k \neq j} S_{kj}, \quad I_j = I_{jj} + \sum_{k \neq j} I_{kj}, \quad R_j = R_{jj} + \sum_{k \neq j} R_{kj}. \tag{11}
\end{aligned}$$

Term by term we set

$$f_{kj} N_j = S_{kj} + I_{kj} + R_{kj}, \quad k \neq j. \tag{12}$$

## 4 The n-Country Stochastic HIV model

In this model, the countries of interest are divided into main countries and satellite countries. In each country, we stratify the population into men who have sex with men (MSM) and heterosexual individuals (two submodels, running in parallel). Each sub-model would involve infection interactions described in the previous sections. The SIR-type of model is extended to more realistic HIV model with three HIV stages (primary, chronic, AIDS), treatment uptake and drop-out in each stage, and pre-exposure prophylaxis uptake by HIV negative persons, who in the model could be both MSM and heterosexual individuals. The schematic of this model was published in Rozhnova et al AIDS 2018.

## 5 Modeling the movement among the main countries

### 5.1 Migration

We first consider migration in a population distributed into  $n$  main countries (islands). By migration we mean the change of place of residence (literally, home address) of a person from one country to another country.



Let  $N_{ii}(t)$  be the number of persons born in country  $i$  ( $i = 1, \dots, n$ ) who are resident in their country of birth at time  $t$ , and let  $N_{ij}(t)$  be the number of persons born in country  $i$  who migrated to country  $j$  ( $j \neq i, j = 1, \dots, n$ ) at time  $t$ . That is, in the notation  $N_{ij}(t)$ , the first index  $i$  denotes the country of birth of a person, and the second index  $j$  denotes the country of residence of a person. If the country of birth  $i$  is different from the country of residence  $j$ , we say that a person is a migrant from country  $i$  in country  $j$ . Let us denote  $m_{ij}$  a rate of migration from country  $j$  to country  $i$ . By definition,  $m_{ii} = 0$ .

The migration of individuals among countries is described by the equations

$$\begin{aligned}\frac{dN_{ii}}{dt} &= \sum_{j=1, j \neq i}^n m_{ij} N_{ij} - \sum_{j=1, j \neq i}^n m_{ji} N_{ii}, \\ \frac{dN_{ij}}{dt} &= m_{ji} N_{ii} - m_{ij} N_{ij}, \quad i \neq j.\end{aligned}\tag{13}$$

Let  $N_i$  be the total number of persons born in country  $i$  at time  $t$

$$N_i = N_{ii}(t) + \sum_{j=1, j \neq i}^n N_{ij}(t).\tag{14}$$

It is easy to check that this number is constant ( $N_i(t) \equiv N_i$ ) because

$$\frac{dN_{ii}}{dt} + \sum_{j=1, j \neq i}^n \frac{dN_{ij}}{dt} = 0.\tag{15}$$

At equilibrium we have

$$\begin{aligned}0 &= \sum_{j=1, j \neq i}^n m_{ij} N_{ij}^* - \sum_{j=1, j \neq i}^n m_{ji} N_{ii}^*, \\ 0 &= m_{ji} N_{ii}^* - m_{ij} N_{ij}^*, \quad i \neq j,\end{aligned}\tag{16}$$

where  $N_{ii}^*$  and  $N_{ij}^*$  denote the steady state values.

The steady state solution of the model is

$$\begin{aligned}N_{ii}^* &= N_i \left( 1 + \sum_{j=1, j \neq i}^n \frac{m_{ji}}{m_{ij}} \right)^{-1}, \\ N_{ij}^* &= N_i \frac{m_{ji}}{m_{ij}} \left( 1 + \sum_{j=1, j \neq i}^n \frac{m_{ji}}{m_{ij}} \right)^{-1}.\end{aligned}\tag{17}$$

For two countries, we have

$$\begin{aligned}N_{11}^* &= \frac{N_1 m_{12}}{(m_{12} + m_{21})}, & N_{22}^* &= \frac{N_2 m_{21}}{(m_{12} + m_{21})}, \\ N_{12}^* &= \frac{N_1 m_{21}}{(m_{12} + m_{21})}, & N_{21}^* &= \frac{N_2 m_{12}}{(m_{12} + m_{21})}.\end{aligned}\tag{18}$$

For three countries, we have

$$\begin{aligned}
N_{11}^* &= \frac{N_1 m_{12} m_{13}}{(m_{12} m_{13} + m_{21} + m_{31})}, \\
N_{12}^* &= \frac{N_1 m_{21} m_{13}}{(m_{12} m_{13} + m_{21} + m_{31})}, \\
N_{13}^* &= \frac{N_1 m_{31} m_{12}}{(m_{12} m_{13} + m_{21} + m_{31})}, \text{ etc.}
\end{aligned} \tag{19}$$

**NB!** According to Arino's work there is a unique positive equilibrium for this system, that is asymptotically stable. From data we see that  $N_{ii}$  and  $N_{ij}$  measured at 5-year time intervals are fairly stable. Therefore, we will take the latest measurement and assume that it corresponds to  $N_{ii}^*$  and  $N_{ij}^*$ . We will then have to solve Eq. (16) to obtain  $m_{ij}$  ( $n \times n$  matrix with 0's on the diagonal).

### 5.1.1 Example for migration between two countries

$$\begin{aligned}
\text{Equation for } N_{11}^* : \quad 0 &= m_{12} N_{12}^* - m_{21} N_{11}^*, \\
\text{Equation for } N_{12}^* : \quad 0 &= m_{21} N_{11}^* - m_{12} N_{12}^*, \\
\text{Equation for } N_{22}^* : \quad 0 &= m_{21} N_{21}^* - m_{12} N_{22}^*, \\
\text{Equation for } N_{21}^* : \quad 0 &= m_{12} N_{22}^* - m_{21} N_{21}^*.
\end{aligned} \tag{20}$$

We have 4 equations for 2 unknowns ( $m_{12}$  and  $m_{21}$ ). However, the first and the second equation are the same. Likewise, the third and the fourth equation are the same. This is not surprising because equations for  $dN_{ii}/dt$  and for  $\sum_{j=1, j \neq i}^n dN_{ij}/dt$  are not linearly independent. This is equivalent to saying that the total number of people born in country  $i$  is constant:  $N_i = N_{ii}(t) + \sum_{j=1, j \neq i}^n N_{ij}(t)$ . Thus,  $N_i$  is just a model parameter. Therefore, we are left with two equations and 2 unknowns ( $m_{12}$  and  $m_{21}$ )

$$\begin{aligned}
\text{Equation for } N_{11}^* : \quad 0 &= m_{12} N_{12}^* - m_{21} N_{11}^*, \\
\text{Equation for } N_{22}^* : \quad 0 &= m_{21} N_{21}^* - m_{12} N_{22}^*.
\end{aligned} \tag{21}$$

It looks like the only solution of this model is zero migration rates ( $m_{12}=0$  and  $m_{21}=0$ ), which is rather strange. However, the correct interpretation is different. The migration rates can be nonzero but they must satisfy simultaneously the following relationships

$$\begin{aligned}
\frac{m_{21}}{m_{12}} &= \frac{N_{12}^*}{N_{11}^*}, \\
\frac{m_{21}}{m_{12}} &= \frac{N_{22}^*}{N_{21}^*}.
\end{aligned} \tag{22}$$

It means that to have nonzero migration rates there are constraints on population sizes in different groups. This is not very realistic, especially given that we want to use data estimates for  $(N_{11}^*, N_{12}^*, N_{22}^*$  and  $N_{21}^*)$  to obtain the migration rates ( $m_{12}$  and  $m_{21}$ ).

Currently, the migration rate  $m_{ij}$  depends only on the country of residence of a person prior to (country  $j$ ) and right after (country  $i$ ) migrating. For example,

$$\begin{aligned} N_{11}^* &\xrightarrow{m_{21}} N_{12}^*, \\ N_{21}^* &\xrightarrow{m_{21}} N_{22}^*. \end{aligned} \quad (23)$$

The solution we propose is to make the migration rate also depend on the country of birth of a person. For example, if the countries we consider are Curaçao (country 1) and the Netherlands (country 2), then we allow the (forward) migration rate of people who were born and live in Curaçao to the Netherlands ( $N_{cc}^* \rightarrow N_{cn}^*$ ) to be different from (the backward) migration rate of people to the Netherlands who were born in the Netherlands but currently live in Curaçao ( $N_{nc}^* \rightarrow N_{nn}^*$ ). At first sight, this seems to be quite a reasonable assumption. For example, people from low-income countries might have a much higher (forward) migration rate than the (backward) migration rate of people to high-income countries. Overall, there is no reason why the artificial constraints of the original model should exist in the real world.

Using a subscript in the migration rate to denote a person's birth country, we obtain

$$\begin{aligned} N_{11}^* &\xrightarrow{m_{21}^1} N_{12}^*, \\ N_{21}^* &\xrightarrow{m_{21}^2} N_{22}^*. \end{aligned} \quad (24)$$

The steady state equations of the model, Eq. (21) now become

$$\begin{aligned} \text{Equation for } N_{11}^* : \quad 0 &= m_{12}^1 N_{12}^* - m_{21}^1 N_{11}^*, \\ \text{Equation for } N_{22}^* : \quad 0 &= m_{21}^2 N_{21}^* - m_{12}^2 N_{22}^*. \end{aligned} \quad (25)$$

We end up with a system of 2 equations and 4 unknowns. If we know the steady state values for population sizes in different groups, from each equation we can determine the ratio of incoming to outgoing rates (but not each rate individually).

**NB!** Suggestions? Suggestions on notation?

### 5.1.2 Example for migration between three countries

Just as a check, we consider an example with three countries. The independent steady state equations are

$$\begin{aligned} \text{Equation for } N_{12}^* : \quad 0 &= m_{21}^1 N_{11}^* - m_{12}^1 N_{12}^*, \\ \text{Equation for } N_{13}^* : \quad 0 &= m_{31}^1 N_{11}^* - m_{13}^1 N_{13}^*, \\ \text{Equation for } N_{21}^* : \quad 0 &= m_{12}^2 N_{22}^* - m_{21}^2 N_{21}^*, \\ \text{Equation for } N_{23}^* : \quad 0 &= m_{32}^2 N_{22}^* - m_{23}^2 N_{23}^*, \\ \text{Equation for } N_{31}^* : \quad 0 &= m_{13}^3 N_{33}^* - m_{31}^3 N_{31}^*, \\ \text{Equation for } N_{32}^* : \quad 0 &= m_{23}^3 N_{33}^* - m_{32}^3 N_{32}^*, \end{aligned}$$

We have 6 equations and 12 unknowns, 6 ratios for the incoming and outgoing migration rates can be determined. In a general model with  $n$  countries, we will have  $n(n-1)$  equations (number of off-diagonal elements in an  $n \times n$  square matrix) and the double of unknowns,  $n(n-1)$  ratios can be estimated.

### 5.1.3 Updated equations

In case when migration rate also depends on the country of birth of a person, the general equations for the migration of individuals among countries are updated as follows

$$\begin{aligned}\frac{dN_{ii}}{dt} &= \sum_{j=1, j \neq i}^n m_{ij}^i N_{ij} - \sum_{j=1, j \neq i}^n m_{ji}^i N_{ii}, \\ \frac{dN_{ij}}{dt} &= m_{ji}^i N_{ii} - m_{ij}^i N_{ij}, \quad i \neq j.\end{aligned}\tag{26}$$

### 5.1.4 Birth and death

We can extend our model to include demographic processes such as birth and death. Birth is assumed to occur in the  $N_{ii}$  compartment at rate  $\mu_i$ . In other words, the birth place and residence of newly born individuals coincide. Natural death occurs in all  $N_{ij}$  compartments at the rate  $\mu_j$ , i.e. the death rate depends on the country of residence of a person.

$$\begin{aligned}\frac{dN_{ii}}{dt} &= \sum_{j=1, j \neq i}^n m_{ij}^i N_{ij} - \sum_{j=1, j \neq i}^n m_{ji}^i N_{ii} + \beta_i \left( N_{ii} + \sum_{j=1, j \neq i}^n N_{ji} \right) - \mu_i N_{ii}, \\ \frac{dN_{ij}}{dt} &= m_{ji}^i N_{ii} - m_{ij}^i N_{ij} - \mu_j N_{ij}, \quad i \neq j.\end{aligned}\tag{27}$$

**NB!** Our previous conclusions regarding the dynamics of the system are still applicable? If we have data to estimate  $\mu$ , we can solve the respective steady state equations to obtain  $m_{ij}$ ?

For HIV transmission I would like to model, birth and death rates have the meaning of rates of recruitment into and leaving the sexually active population. It is reasonable to assume that the duration of sexual life is the same in all countries as people tend to start and stop being sexually active at more or less the same age. We also assume that people start being sexually active and stop being sexually active in their country of residence.

**NB!** 1) What if birth rate is different from death rate, and differs by country? This is a more general scenario. Can we still use the method to estimate  $m_{ij}$ ?

### 5.1.5 Birth and death for two countries

$$\begin{aligned}\text{Equation for } N_{11}^* : \quad 0 &= m_{12}^1 N_{12}^* - m_{21}^1 N_{11}^* + \beta_1 (N_{11}^* + N_{21}^*) - \mu_1 N_{11}^*, \\ \text{Equation for } N_{12}^* : \quad 0 &= m_{21}^1 N_{11}^* - m_{12}^1 N_{12}^* - \mu_2 N_{12}^*, \\ \text{Equation for } N_{22}^* : \quad 0 &= m_{21}^2 N_{21}^* - m_{12}^2 N_{22}^* + \beta_2 (N_{22}^* + N_{12}^*) - \mu_2 N_{22}^*, \\ \text{Equation for } N_{21}^* : \quad 0 &= m_{12}^2 N_{22}^* - m_{21}^2 N_{21}^* - \mu_1 N_{21}^*.\end{aligned}\tag{28}$$

Summing  $N_{11}^*$  and  $N_{21}^*$  we obtain the population resident in country 1. Similarly, the population resident in country 2 is given by  $N_{22}^*$  and  $N_{12}^*$ . At the steady state we have

$$\begin{aligned} \text{Equation for } (N_{11}^* + N_{21}^*) : \quad 0 &= m_{12}^1 N_{12}^* - m_{21}^1 N_{11}^* + \beta_1 (N_{11}^* + N_{21}^*) - \mu_1 N_{11}^* + \\ &+ m_{12}^2 N_{22}^* - m_{21}^2 N_{21}^* - \mu_1 N_{21}^*, \\ \text{Equation for } (N_{22}^* + N_{12}^*) : \quad 0 &= m_{21}^2 N_{21}^* - m_{12}^2 N_{22}^* + \beta_2 (N_{22}^* + N_{12}^*) - \mu_2 N_{22}^* + \\ &+ m_{21}^1 N_{11}^* - m_{12}^1 N_{12}^* - \mu_2 N_{12}^*. \end{aligned}$$

The terms with  $m$ 's cancel out. Further simplifying we get

$$\begin{aligned} \text{Equation for } (N_{11}^* + N_{21}^*) : \quad 0 &= \beta_1 (N_{11}^* + N_{21}^*) - \mu_1 (N_{11}^* + N_{21}^*), \\ \text{Equation for } (N_{22}^* + N_{12}^*) : \quad 0 &= \beta_2 (N_{22}^* + N_{12}^*) - \mu_2 (N_{22}^* + N_{12}^*). \end{aligned}$$

This means what for a model that includes both migration and birth and death we must have  $\beta_1 = \mu_1$ ,  $\beta_2 = \mu_2$  etc.

## 5.2 Short-term visits including tourism

In addition to migration, we know that short-term visits could aid in the transmission of HIV on the Caribbean islands. The yearly number of persons visiting the islands is much larger than the population of the islands themselves. To account for short-term visits including tourism, the model needs to be extended further. By a short-term visit we mean the temporary change of location of a person from the country of one's residence to another country. Example: Ganna was born in Ukraine, and is currently resident in the Netherlands. However, Ganna will be temporary away from the Netherlands next December to attend the Epidemics conference in the USA.

We extend our notation to take into account the current location of a person as follows. In  $N_{ijk}(t)$ , the first index  $i$  denotes the country of birth of a person, the second index  $j$  denotes the country of residence of a person, and the third index  $k$  denotes the current location (country) of a person. We assume that immediately after migration the country of residence and the location of a person coincide, i.e. the second and the third index are equal. Therefore, Eq. (26) is modified as follows

$$\begin{aligned} \frac{dN_{iii}}{dt} &= \sum_{j=1, j \neq i}^n m_{ij} N_{ijj} - \sum_{j=1, j \neq i}^n m_{ji} N_{iii}, \\ \frac{dN_{ijj}}{dt} &= m_{ji} N_{iii} - m_{ij} N_{ijj}, \quad i \neq j. \end{aligned} \tag{29}$$

Let us denote  $p_{ij}$  a rate of short-term movement from country  $j$  to country  $i$ . By definition,  $p_{ii} = 0$ . The equations with this type of movement become

$$\begin{aligned}
\frac{dN_{iii}}{dt} &= \sum_{j=1, j \neq i}^n m_{ij} N_{ijj} - \sum_{j=1, j \neq i}^n m_{ji} N_{iii} + \sum_{j=1, j \neq i}^n p_{ij} N_{iij} - \sum_{j=1, j \neq i}^n p_{ji} N_{iii}, \\
\frac{dN_{iij}}{dt} &= p_{ji} N_{iii} - p_{ij} N_{iij} \quad i \neq j \\
\frac{dN_{ijj}}{dt} &= m_{ji} N_{iii} - m_{ij} N_{ijj} + \sum_{k=1, k \neq j}^n p_{jk} N_{ijk} - \sum_{k=1, k \neq j}^n p_{kj} N_{ijj}, \quad i \neq j \\
\frac{dN_{ijk}}{dt} &= p_{kj} N_{ijj} - p_{jk} N_{ijk}, \quad i \neq j, k \neq j.
\end{aligned} \tag{30}$$

The sum of Eq. (30) is zero, so the total population size of all islands does not change. Note that we assume that a person leaving on a short-term trip from country  $i$  to country  $j$  must return back to its country of residence  $i$  prior to leaving on another short-term trip to country  $k$ . We think this is justified because the islands are geographically isolated communities. The same assumption was made regarding migration: a person had to return to its country of birth prior becoming a migrant again.

**NB!** Is it possible to estimate the rates of short-term movement similarly to the migration rates? The short-term movement data refer to the yearly number of visitors to country  $i$  from country  $j$  (cumulative number not the instantaneous number as we had for migration data). Also, it is quite diverse data set, some of it refers to arrivals of non-resident tourists, some to arrivals of non-resident visitors at national borders, by country of residence (some part) or by nationality (some part). Some assumptions must be made here, for example, that the country of residence and nationality are the same and refer to the second index in our notation.

### 5.3 Some useful quantities

- $N$  is the total population size of all countries (constant)

$$N = \sum_{i=1}^n N_i \tag{31}$$

- $N_i$  is the total number of persons born in country  $i$  at time  $t$  (constant)

$$N_i = \sum_{k=1}^n N_{iik}(t) + \sum_{k=1}^n \sum_{j=1, j \neq i}^n N_{ijk}(t) \tag{32}$$

- $R_j$  is the total number of persons resident in country  $j$  at time  $t$

$$R_j(t) = \sum_{k=1}^n \sum_{i=1}^n N_{ijk}(t) \tag{33}$$

- $N_{ii}(t)$  is the number of persons born in country  $i$  ( $i = 1, \dots, n$ ) who are resident in their country of birth at time  $t$

$$N_{ii}(t) = \sum_{k=1}^n N_{iik}(t) \quad (34)$$

- $N_{ij}(t)$  is the number of persons born in country  $i$  who migrated to country  $j$  ( $j \neq i$ ,  $j = 1, \dots, n$ ) at time  $t$

$$N_{ij}(t) = \sum_{k=1}^n N_{ijk}(t) \quad (35)$$

- $V_k(t)$  is the number of short-term visitors in country  $k = 1, \dots, n$

$$V_k(t) = \sum_{i=1}^n \sum_{j=1, j \neq k}^n N_{ijk}(t) \quad (36)$$

- $V_{jk}(t)$  is the number of short-term visitors in country  $k = 1, \dots, n$  who are resident in country  $j \neq k$

$$V_{jk}(t) = \sum_{i=1}^n N_{ijk}(t), \quad j \neq k \quad (37)$$

## 5.4 Summary of available data

1. Demographics data: Total population at mid-year by country, 1990-2017 (every 5 years)
2. Migration data: International migrant stock at mid-year by country, 1990-2017 (every 5 years)
  - International migrant: any person who changes his or her country of usual residence (UN DESA, 1998)
  - International migrant stocks: the total number of international migrants present in a given country at a particular point in time (UN SD, 2017: 9)
  - The data used to produce the estimates refer to the foreign-born population
3. Non-resident tourists or non-resident visitors data (average quality)
  - Arrivals by air of non-resident tourists or non-resident visitors at national borders, by country of residence or by nationality, 2013-2017 (yearly)
  - Outbound tourism - trips abroad by resident visitors to countries of destination (basis: arrivals in destination countries), 1995-2017 (yearly)

## 6 Alternative model formulation

Several constraints are desirable in the model:

- The stratification of the population by ethnicity into migrants and natives in the main countries (Curaçao, Aruba, St Maarten, BES) should not change (apart from decreasing due to the HIV-related mortality).
- The main countries (Curaçao, Aruba, St Maarten, BES) should be linked among themselves so that HIV-related interventions in one country can potentially affect HIV transmission in another country.

As the basis for an alternative model we consider the study by Xiridou et al published in *AIDS* 2010, 24:2081-2088. The population in each of the main countries (Curaçao, Aruba, St Maarten, BES) is stratified by ethnicity into natives, migrants from the main countries (Curaçao, Aruba, St Maarten, BES) and migrants from other locations (Asia, Other Caribbean, Europe, Latin America, Northern America, Oceania). Let  $N_{me}(t)$  denote the number of individuals of ethnicity  $e$  in the main country  $m$  at time  $t$ . The index  $m = 1, 2, 3, 4$  goes through the main countries (Curaçao, Aruba, St Maarten, BES) and the index  $e = 1, 2, \dots, 10$  goes through the main countries and all other locations (Curaçao, Aruba, St Maarten, BES, Asia, Other Caribbean, Europe, Latin America, Northern America, Oceania). The total population size of the main country,  $N_m(t)$ , is obtained

by summing over all ethnic groups living in the country as follows  $N_m(t) \equiv \sum_{e=1}^{e=10} N_{me}(t)$ .

For example,  $N_{11}(t)$  is the number of Curaçaoans (Curaçao natives) in Curaçao;  $N_{22}(t)$  is the number of Aruba natives (Arubans) in Aruba;  $N_{36}(t)$  is the number of migrants from other Caribbean countries in St Maarten. Note that the only available data are total population size of the main countries and the stratification of the main countries by ethnicity based on the number of people who are born abroad. To stress that these quantities ( $N_m^0$  and  $N_{me}^0$ , respectively) are constant, we have added the superscript 0 (the initial condition at  $t = 0$ ).

To keep the population stratification constant in the absence of infection, we assume that the per capita rates of entering the sexually active population and leaving it due to causes unrelated to HIV are equal and given by the average duration of the sexual lifespan. Denoting this rate  $\mu$ , we can write the equation for  $N_{me}(t)$  as follows

$$\frac{dN_{me}(t)}{dt} = \mu N_{me}^0 - \mu N_{me}(t). \quad (38)$$

In total, we would have  $4 \times 10$  equations for  $N_{me}(t)$ .

In the simplest case of an SIR model,  $N_{me}(t) = S_{me}(t) + I_{me}(t) + R_{me}(t)$ . We can further take into account that natives always enter the population as susceptible. However, other ethnic groups can enter the population in infected state with the probability equal to the HIV prevalence in their country of origin,  $P_e(t)$ . The equations for the number of susceptible,  $S_{me}(t)$ , and infected,  $I_{me}(t)$ , individuals read



$$\begin{aligned}\frac{dS_{me}(t)}{dt} &= [1 - P_e(t) (1 - \delta_{me})] \mu N_{me}^0 - \mu S_{me}(t) - \text{infection term}, \\ \frac{dI_{me}(t)}{dt} &= P_e(t) (1 - \delta_{me}) \mu N_{me}^0 - \mu I_{me}(t) + \text{infection term} - \text{recovery term},\end{aligned}\quad (39)$$

where  $\delta_{me} = 1$  if  $m = e$  and 0 otherwise. Here  $P_e(t) \equiv P_{ee}(t) = I_{ee}(t)/N_{ee}(t)$  is the HIV prevalence of the ethnic group  $e$  in their country of origin. If the country of origin is the main country (Curaçao, Aruba, St Maarten, BES), i.e.  $e \in m$  then  $P_e(t)$  will be given by the model equations. In case of effective measures against HIV such as the increased treatment or PrEP uptake in the country of origin, the influx of infectious individuals in other main islands would decrease. For non-main countries (Asia, Other Caribbean, Europe, Latin America, Northern America, Oceania), the prevalence will be taken as a model parameter and will be fixed in the analyses ( $P_e = \text{const}$ ).

$$\begin{aligned}\frac{dS_{me}(t)}{dt} &= \left[ 1 - \frac{I_{ee}(t)}{N_{ee}(t)} (1 - \delta_{me}) \right] \mu N_{me}^0 - \mu S_{me}(t) - \text{infection term}, \\ \frac{dI_{me}(t)}{dt} &= \frac{I_{ee}(t)}{N_{ee}(t)} (1 - \delta_{me}) \mu N_{me}^0 - \mu I_{me}(t) + \text{infection term} - \text{recovery term}.\end{aligned}\quad (40)$$

The infection term consists of two parts. The first term corresponds to infection of  $S_{me}$  individuals in the main country  $m$  due to sexual partners from all other ethnic groups  $e$ . The second term corresponds to infection of  $S_{me}$  individuals due to sexual partners while they visit their country of origin  $e$ . Note there are multiple ways to write down the second term (e.g. by using information on frequency of trips back to the country of origin and the percentage of people who have partners back in their country of origin etc.). I use the simple form such as

$$\text{infection term} = S_{me}(t) [J_{me}(t) + J'_{me}(t)], \quad (41)$$

where  $J_{me}(t)$  and  $J'_{me}(t)$  are the forces of infection per unit of time for the ethnic group  $e$  living in country  $m$  due to sexual partners within the country and back in their country of origin  $e$ , respectively. Note that we assume that  $S_{me}$  individuals do not have partners in countries other than  $m$  and  $e$ . Native individuals from the main countries  $m$  are assumed to have sexual partners only within their own country (with all other ethnic groups) but not in other countries. In a very simplified way, we could write

$$\text{infection term} = S_{me}(t) \epsilon c_{me} \left[ p_{me} \sum_{e'=1}^{e'=10} M_{me,me'}(t) \frac{I_{me'}(t)}{N_{me'}(t)} + (1 - p_{me}) \frac{I_{ee}(t)}{N_{ee}(t)} \right]. \quad (42)$$

Here  $\epsilon$  is the probability of transmission per partner,  $c_{me}$  is the partner change rate for ethnic group  $e$  in country  $m$ .  $p_{me}$  is the proportion of sexual partners for the ethnic group  $e$  in the country  $m$  with individuals from the same country;  $(1 - p_{me})$  then denotes the proportion of sexual partners this ethnic group has in their country of origin  $e$ . By

definition, for native individuals from the main countries  $m$  we have  $p_{mm} = 1$  by definition.  $M_{me,me'}(t)$  denotes the mixing of the ethnic group  $e$  with ethnic group  $e'$  in the country  $m$ :

$$M_{me,me'}(t) = \omega \frac{c_{me'} N_{me'}(t)}{\sum_{e''=1} c_{me''} N_{me''}(t)} + (1 - \omega) \delta_{ee'}, \quad (43)$$

where  $\omega \in [0, 1]$  is a mixing parameter.

## 7 A slight reformulation of the model

We start with the migration equations proposed in Section 5. As above, we denote by  $N_{ij}(t)$  the number of individuals born in country  $i$  residing at country  $j$  at time  $t$ , and denote by  $m_{ij}$  the per-capita rate of individuals migrating from country  $j$  to country  $i$  per unit of time (and define  $m_{ii} = 0$  for all  $i$ ). The equations read

$$\begin{aligned}\frac{dN_{ii}}{dt} &= \sum_{j=1, j \neq i}^n m_{ij}N_{ij} - \sum_{j=1, j \neq i}^n m_{ji}N_{ii}, \\ \frac{dN_{ij}}{dt} &= m_{ji}N_{ii} - m_{ij}N_{ij}, \quad i \neq j.\end{aligned}$$

We propose the following modification of the second equation above, to include migration for individuals from country  $i$  between two countries  $j$  and  $k$ , possibly different from their origin country:

$$\frac{dN_{ii}}{dt} = \sum_{j=1}^n m_{ij}N_{ij} - \sum_{j=1}^n m_{ji}N_{ii}, \quad (44)$$

$$\frac{dN_{ij}}{dt} = \sum_{k=1}^n m_{jk}N_{ik} - \sum_{k=1}^n m_{kj}N_{ij}, \quad i \neq j. \quad (45)$$

Note that this new set of equations can be unified in a single structure (which we should also be able to extend to include the births and deaths of individuals - see below):

$$\frac{dN_{ij}}{dt} = \sum_{k=1}^n m_{jk}N_{ik} - \left( \sum_{k=1}^n m_{kj}N_{ij} \right). \quad (46)$$

That is, for each nationality  $i$ , we want to look at how its population residing at each country  $j$  changes over time (in particular, at country  $j = i$  as well). In addition to the natural dynamics of the population, this variation includes two migration-related processes: i) individuals of nationality  $i$  arrive from other countries  $k$  (at a rate  $m_{jk}$ ), and ii) individuals of nationality  $i$  leave to other countries  $k$  (at a rate  $m_{kj}$ ). Note that we are implicitly assuming that the rate of migration from country  $j$  to country  $i$  is the same for individuals from all nationalities (probably unrealistic - more on this below).

Let us focus on Eq. (46). UN provides data on the population from every nationality at each country in the world, every 5 years. The strategy is now to infer the migration rates in the system from these discrete “snapshots” of the population. That is, we want to solve Eq. (46) **for**  $m_{ij}$ , rather than for the  $N_{ij}$  (which is the data we have at hand). To do so, we approximate both sides of the equation by their finite difference expression

$$N_{ij}(t+1) - N_{ij}(t) = \sum_{k=1}^n m_{jk}N_{ik}(t) - \left( \sum_{k=1}^n m_{kj}N_{ij}(t) \right). \quad (47)$$

In principle, we could also use the backward finite difference  $N_{ij}(t) - N_{ij}(t-1)$  or the average  $N_{ij}(t+1) - N_{ij}(t-1)$  in the left-hand side above. Note that we are also implicitly incorporating the natural life cycle of the population in the yearly variation of the

total residents of individuals from each nationality. This could be corrected by explicitly including the births and deaths of individuals from nationality  $j$  at country  $i$  in the left-hand side above; we can also ignore these and assume that they balance each other for the moment.

We consider 5 years as the unit of time, and denote  $n_{ij} = N_{ij}(t+1) - N_{ij}(t)$ ; these coefficients can be computed straightforwardly from the UN data. Note that we are making a second assumption here: that the migration rates do not change over time. There is a total of  $N^2 - N$  unknowns in this system: the migration rates  $m_{ij}$  for  $i, j = 1 \dots N$  (recall  $m_{ii} = 0$  for all  $N$ ). Moreover, exactly  $N$  equations are redundant in the system, as adding the equations for  $N_{ij}(t)$  over all  $j$  gives a null condition (ie the total population of people of the same nationality is not changed by where this population is located), and the remaining equations are linearly independent. That provides a system with a single solution, that can be found from the coefficients  $n_{ij}$  with standard techniques for systems of linear equations. Written in terms of the coefficients  $n_{ij}$ , system (47) takes the form

$$n_{ij} = \sum_k N_{ik} m_{jk} - N_{ij} \sum_k m_{kj} \quad (48)$$

As an example, let us consider the case of  $n = 3$  countries. The system is the following:

$$\begin{array}{rrrr} -N_{11}(m_{21} + m_{31}) & +N_{12}m_{12} & +N_{13}m_{13} & = n_{11} \\ N_{11}m_{21} & -N_{12}(m_{12} + m_{32}) & +N_{13}m_{23} & = n_{12} \\ N_{11}m_{31} & +N_{12}m_{32} & -N_{13}(m_{13} + m_{23}) & = n_{13} \\ -N_{21}(m_{21} + m_{31}) & +N_{22}m_{12} & +N_{23}m_{13} & = n_{21} \\ N_{21}m_{21} & -N_{22}(m_{12} + m_{32}) & +N_{23}m_{23} & = n_{22} \\ N_{21}m_{31} & +N_{22}m_{32} & -N_{23}(m_{13} + m_{23}) & = n_{23} \\ -N_{31}(m_{21} + m_{31}) & +N_{32}m_{12} & +N_{33}m_{13} & = n_{31} \\ N_{31}m_{21} & -N_{32}(m_{12} + m_{32}) & +N_{33}m_{23} & = n_{32} \\ N_{31}m_{31} & +N_{32}m_{32} & -N_{33}(m_{13} + m_{23}) & = n_{33} \end{array} \quad (49)$$

Adding the first three equations, the second three equations, and the third three equations, we obtain the conditions

$$0 = n_{11} + n_{12} + n_{13}, \quad (50)$$

$$0 = n_{21} + n_{22} + n_{23}, \quad (51)$$

$$0 = n_{31} + n_{32} + n_{33}, \quad (52)$$

$$(53)$$

which amount to assuming that the global births and deaths of individuals of nationality  $i$  cancel out, for all nationalities. These births and deaths could also be explicitly included in the right-hand sides of the system to obtain corrected migration rates that take into account these dynamics, replacing the  $n_{ij}$  coefficients in the right hand sides by  $n'_{ij} = n_{ij} + b_{ij} - d_{ij}$ , where  $b_{ij}$  and  $d_{ij}$  are the births and deaths of individuals of nationality  $i$  at country  $j$ . In any case, if we want our system to be consistent, we will need to use input data  $n_{ij}$  so that these conditions are met. We will assume that these conditions are satisfied for the moment and show in the next section how to deal with this from the data.

If we label the migration rates  $m_{21}, m_{31}, m_{12}, m_{13}, m_{23}, m_{32}$  as the unknowns of the system (in this order), the augmented matrix of coefficients of the system reads

$$\left( \begin{array}{cccccc|c} -N_{11} & -N_{11} & N_{12} & N_{13} & 0 & 0 & n_{11} \\ N_{11} & 0 & -N_{12} & 0 & N_{13} & -N_{12} & n_{12} \\ 0 & N_{11} & 0 & -N_{13} & -N_{13} & N_{12} & n_{13} \\ -N_{21} & -N_{21} & N_{22} & N_{23} & 0 & 0 & n_{21} \\ N_{21} & 0 & -N_{22} & 0 & N_{23} & -N_{22} & n_{22} \\ 0 & N_{21} & 0 & -N_{23} & -N_{23} & N_{22} & n_{23} \\ -N_{31} & -N_{31} & N_{32} & N_{33} & 0 & 0 & n_{31} \\ N_{31} & 0 & -N_{32} & 0 & N_{33} & -N_{32} & n_{32} \\ 0 & N_{31} & 0 & -N_{33} & -N_{33} & N_{32} & n_{33} \end{array} \right). \quad (54)$$

Removing the third, sixth and ninth equations, we obtain

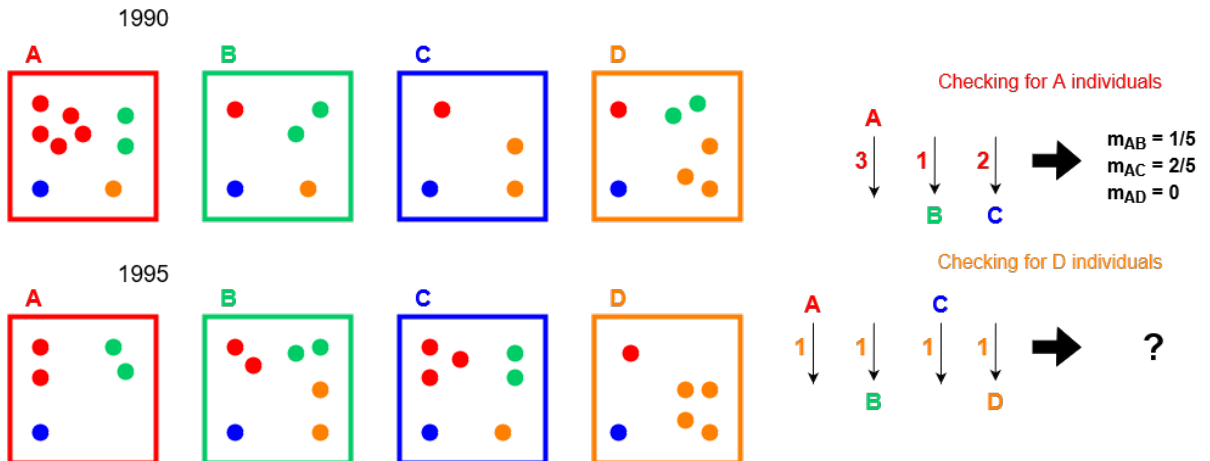
$$\left( \begin{array}{cccccc|c} -N_{11} & -N_{11} & N_{12} & N_{13} & 0 & 0 & n_{11} \\ N_{11} & 0 & -N_{12} & 0 & N_{13} & -N_{12} & n_{12} \\ -N_{21} & -N_{21} & N_{22} & N_{23} & 0 & 0 & n_{21} \\ N_{21} & 0 & -N_{22} & 0 & N_{23} & -N_{22} & n_{22} \\ -N_{31} & -N_{31} & N_{32} & N_{33} & 0 & 0 & n_{31} \\ N_{31} & 0 & -N_{32} & 0 & N_{33} & -N_{32} & n_{32} \end{array} \right). \quad (55)$$

This is a full-rank  $6 \times 6$  matrix, and thus the corresponding system has a unique solution.

Another possibility would be to assume that the migration rates  $m_{ij}$  measure the per-capita rate of individuals born in  $i$  moving to country  $j$  per unit of time (instead of the rate of individuals residing in country  $i$  moving to country  $j$ ), and to assume that this is independent of their current country of residence. This may also be unrealistic, of course (more on this below). The system would now read

$$\frac{dN_{ij}}{dt} = m_{ji} \sum_{k=1}^n N_{ik} - \left( \sum_{k=1}^n m_{ki} \right) N_{ij}. \quad (56)$$

One can see that this system shares the properties highlighted for the first one: there are  $N^2 - N$  unknowns and  $N^2$  equations that reduce to  $N^2 - N$  linearly independent ones. Thus, another set of migration rates  $m_{ij}$  can be found from this system.



## 8 Computing migration rates from migrant population data

We group migrant population data according to the following groupings:

- Locations of interest: Aruba, BES, Curaçao.
- Interacting regions: Latin America (comprising Central and Southern America), Rest of Caribbean (comprising the Caribbean countries except the locations of interest), Netherlands, and rest of the world.

We collected data concerning these groupings from the following sources:

- UN population data: Estimates for the total population at each country and region in the world were obtained from the UN population division data portal.
- UN migrant stock data: figures on the number of individuals from each nationality living at each country and region if the world were extracted from the “Origin and destination” table. These data were aggregated according to the groupings above.

The total populations of migrants were scaled down to approximate the population aged 15-64, using the age distributions of the migrant population at each country and region in the world from the “Age and sex” table.

The total international migrant stock at each country and region in the world was extracted from the “Sex and destination” table as well, and subtracted from the UN population data (after grouping the figures accordingly), to obtain the number of national individuals at each country and region in the world.

- CBS public repository: the foreign population at BES was not available at the UN data table, and was obtained from the national Dutch statistics institute (CBS).

After processing and joining these sources of data, we arrive at a dataset for analysis that includes the total population from each location of interest and interacting region in the others. This reference dataset is file `mig_pop.csv`.

### 8.1 A simple example: country-specific rates for ABC

We start by considering only the migratory flows between the three locations of interest (A, B and C). After processing the data, we arrive at the following figures:

We encounter a key matter here (and the reason why we obtained negative migration rates in our first computation). A central assumption for the validity of our reasoning is that only  $N^2 - N$  equations in the system are linearly independent; this condition is achieved by imposing

$$\sum_j n_{ij} = 0 \quad (57)$$

for all  $i$ . That is, we assume that births and deaths of individuals of each nationality cancel out. For the case of 3 nodes, this allows us to reduce system (54) to system (55),

dest	orig	y2010	y2015	y2020	v2020	v2015
C	C	124750	131961	132078	117	7211
C	A	1573	1708	1992	284	135
C	B	2149	2334	2722	388	185
A	C	2335	2455	2745	290	120
A	A	66013	68143	52992	-15151	2130
A	B	283	297	331	34	14
B	C	3216	3540	3909	369	324
B	A	524	625	675	50	101
B	B	8268	8870	9254	384	602

Table 1: Number of individuals from orig at dest for years 2010, 2015 and 2020, and their 5-year variation. File `mig_pop_ABC.xlsx`

which has a unique solution. Real data does not verify this, as can be easily checked by adding the corresponding rows and columns in the table above. While we can derive the raw variation of the total number of individuals of each nationality between years, data on the births and deaths of these are not available by country.

What I did for the moment was to assume that the raw change of of the total number of individuals of each nationality was compensated at their origin node. That is: I computed the total variation of the number of individuals from Curaçao from year 2015 to year 2020 over the three regions as a whole, and manually changed the number of individuals from Curaçao at Curaçao in year 2020 so that this total variation was 0. While this change allows us to proceed, and it may not cause a significant distortion in the real data (e.g. total variation for BES = 806, for Curaçao = 776), in some cases this may be actually an important modification of the data (e.g. total variation for Aruba = -14817). This resulted in the numbers in table 2, that do verify the desired assumptions.

dest	orig	y2010	y2015	y2020	v2020	v2015
C	C	124750	131961	131302	-659	7211
C	A	1573	1708	1992	284	135
C	B	2149	2334	2722	388	185
A	C	2335	2455	2745	290	120
A	A	66013	68143	67809	-334	2130
A	B	283	297	331	34	14
B	C	3216	3540	3909	369	324
B	A	524	625	675	50	101
B	B	8268	8870	8448	-422	602

Table 2: Number of individuals from orig at dest for years 2010, 2015 and 2020, and their yearly variation, after adjusting the global variations by nationality to compensate births and deaths (only done for 2020 in this table). File `mig_pop_ABC_adj.xlsx`.

Solving system (55) with this data provides the migration rates in table 3, taking in all cases expected values, and checked to work (i.e. taking 2015 populations as input and

assuming these migration rates leads to 2020 populations). Excel versions of these files can be found in the root folder.

rate	value	notes
$m_{BA}$	0.001077975	from Aruba to BES
$m_{CA}$	0.003909974	from Aruba to Curaçao
$m_{AB}$	0.003421627	from BES to Aruba
$m_{AC}$	0.002198626	from Curaçao to Aruba
$m_{BC}$	0.004082292	from Curaçao to BES
$m_{CB}$	0.045264758	from BES to Curaçao

Table 3: Migration rates for the locations of interest, derived under the assumptions explained in the text, after adjusting for 2020 total variations by nationality. File `mig_rates_2020_adjusted.xlsx`.

If we perform this adjustment for both timespans under consideration (2010 to 2015 and 2015 to 2020) subsequently (ie, taking the adjusted 2015 populations as base for the 2020 adjustment), we obtain the populations in table 4. This accounted for smaller corrections than before, resulting in 2366 and -12451 individuals from Aruba subtracted from Aruba’s population in 2015 and 2020 respectively, 801 and 1607 individuals from BES subtracted from BES’s population, and 7655 and 8431 individuals from Curaçao subtracted from Curaçao’s population.

dest	orig	y2010	y2015	y2020	v2020	v2015
C	C	124750	124306	123647	-659	-444
A	C	2335	2455	2745	290	120
B	C	3216	3540	3909	369	324
C	A	1573	1708	1992	284	135
A	A	66013	65777	65443	-334	-236
B	A	524	625	675	50	101
C	B	2149	2334	2722	388	185
A	B	283	297	331	34	14
B	B	8268	8069	7647	-422	-199

Table 4: Number of individuals from orig at dest for years 2010, 2015 and 2020, and their yearly variation, after adjusting the global variations by nationality to compensate births and deaths (done for 2015 and 2020 in this table). File `mig_pop_ABC_adj_2.xlsx`.

Following the same reasoning as above, we obtain the migration rates shown in table 5. They should be interpreted as before: taking the 2010 populations as initial conditions leads to the 2015 populations after one time unit if the 2010 to 2015 migrations rates are considered. Analogously, taking the 2015 populations as initial conditions leads to the 2020 populations after one time unit if the 2015 to 2020 migrations rates are considered. Note that the time unit here is 5 years (the timespan between observations). For shorter times, the migration rates would need to be scaled down (e.g. divided by 5 for yearly rates).



rate	2010 to 2015	2015 to 2020	notes
$m_{BA}$	0.001651642	0.001153624	from Aruba to BES
$m_{CA}$	0.001959365	0.004020058	from Aruba to Curaçao
$m_{AB}$	0.001559737	0.003730443	from BES to Aruba
$m_{AC}$	0.000989303	0.002328895	from Curaçao to Aruba
$m_{BC}$	0.003209725	0.004473128	from Curaçao to BES
$m_{CB}$	0.023399759	0.049904817	from BES to Curaçao

Table 5: Migration rates for the locations of interest, derived under the assumptions explained in the text, after adjusting for 2015 and 2020 total variations by nationality. File `mig_rates_2015_2020_adjusted.xlsx`.

## 8.2 A simple example II: nationality-specific rates for ABC

For the case of three nationalities, equation (56) reads

$$\begin{aligned}
n_{11} &= m_{11}(N_{12} + N_{13}) - m_{21}N_{11} - m_{31}N_{11} \\
n_{12} &= -m_{11}N_{12} + m_{21}(N_{11} + N_{13}) - m_{31}N_{12} \\
n_{13} &= -m_{11}N_{13} - m_{21}N_{13} + m_{31}(N_{11} + N_{12}) \\
n_{21} &= m_{12}(N_{22} + N_{23}) - m_{22}N_{21} - m_{32}N_{21} \\
n_{22} &= -m_{12}N_{22} + m_{22}(N_{21} + N_{23}) - m_{32}N_{22} \\
n_{23} &= -m_{12}N_{23} - m_{22}N_{23} + m_{32}(N_{21} + N_{22}) \\
n_{31} &= m_{13}(N_{32} + N_{33}) - m_{23}N_{31} - m_{33}N_{31} \\
n_{32} &= -m_{13}N_{32} + m_{23}(N_{31} + N_{33}) - m_{33}N_{32} \\
n_{33} &= -m_{13}N_{33} - m_{23}N_{33} + m_{33}(N_{31} + N_{32})
\end{aligned}$$

where, now,  $m_{ij}$  denotes the rate of migration of individuals from nationality  $j$  to country  $i$  (independently of their current country of residence). Note that in this case,  $m_{ii}$  is not necessarily zero.

We see that the system consists of three independent subsystems: those concerning the movement of individuals of each of the nationalities (first, second and third groups of three equations). We also see that each of these blocks contains a linearly dependent equation, as the sum of the three equations in the block leads to the empty condition  $0 = 0$ . This shows the phenomenon we observed during previous discussions: if we observe 20 more Arubans at BES in 2020 than in 2015, is it because 20 Arubans moved to BES, or because 30 Arubans moved to BES and 10 moved from BES to other locations?

This introduces a degree of freedom at each of the blocks. In terms of migration rates, this means that rather than solving the system for some values of  $m_{11}, m_{12}, \dots$ , we will be able to find particular solutions for these rates, that can be perturbed by the addition of a parameter (one for each nationality), that needs to be fitted to data. That is, the solution is now

$$\begin{pmatrix} m_{11} \\ m_{21} \\ m_{31} \end{pmatrix} + \lambda_1 v_1, \quad \begin{pmatrix} m_{12} \\ m_{22} \\ m_{32} \end{pmatrix} + \lambda_2 v_2, \quad \begin{pmatrix} m_{13} \\ m_{23} \\ m_{33} \end{pmatrix} + \lambda_3 v_3, \quad (58)$$

where  $\lambda_1, \lambda_2, \lambda_3$  control the raw volume of movement for individuals of nationalities 1, 2 and 3, and the vectors  $v_1, v_2, v_3$  span the null space of each block of equations. Any choice of values for the  $\lambda_i$  will lead to the same final populations in year 2020; their particular values will determine the amount of individuals that moved during the considered period of time.

Table 6 shows the values of these parameters found from the population data for the timespan under study, after performing the same adjustment as in previous section (that is, assuming that the total population of people from A, B and C remained constant from 2010 to 2020). Note that the migration rates to C are 0 in all cases, and the corresponding coordinate of the vector spanning the nullspace is 1; this is arbitrary, and just a consequence of the freedom in the choice of rates and R’s linear system solver default behaviour. Equally valid solutions can be found substituting the values in the table in equation 58.

rate	2010 to 2015 solution	2010 to 2015 nullspace	2015 to 2020 solution	2010 to 2015 nullspace
$m_{AA}$	-0.086645887	41.96630642	-0.165484646	38.51112412
$m_{BA}$	0.000822619	0.333121424	-0.0007917	0.365925059
$m_{CA}$	0	1	0	1
$m_{AB}$	-0.000968457	0.131689158	-0.001436706	0.127249357
$m_{BB}$	-0.085118094	3.84737087	-0.164801512	3.457155099
$m_{CB}$	0	1	0	1
$m_{AC}$	0.000984724	0.018717435	0.0023255	0.01974965
$m_{BC}$	0.002574394	0.025779559	0.002975933	0.02847811
$m_{CC}$	0	1	0	1

Table 6: Nationality-specific migration rates and null spaces for the locations of interest, derived under the assumptions explained in the text, after adjusting for 2015 and 2020 total variations by nationality. The columns labelled with “nullspace” contain the vectors  $(v_1, v_2, v_3)$  (in this order), as denoted in Eq. (58). File `mig_rates_nat_2015_2020_adjusted.xlsx`.

## 9 Arbitrary number of locations of interest

Let us now show how to systematically treat the general case of  $n$  locations of interest under study. We start from the system

$$n_{ij} = \sum_k N_{ik} m_{jk} - N_{ij} \sum_k m_{kj} \quad (59)$$

We label the  $n^2$  unknowns under consideration (the migration rates  $m_{ij}$ ) in the order that would result if we were to read them from their matrix arrangement, top-to-bottom and

left-to-right. That is, we write

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} & \dots & m_{1n} \\ m_{21} & m_{22} & m_{23} & \dots & m_{2n} \\ m_{31} & m_{32} & m_{33} & \dots & m_{3n} \\ & & & \dots & \end{pmatrix}, \quad (60)$$

and thus label

$$x_1 = m_{11}, \quad (61)$$

$$x_2 = m_{21}, \quad (62)$$

$$\vdots \quad \vdots \quad (63)$$

$$x_n = m_{n1}, \quad (64)$$

$$x_{n+1} = m_{12}, \quad (65)$$

$$x_{n+2} = m_{22}, \quad (66)$$

$$x_{n+3} = m_{32}, \quad (67)$$

$$\vdots \quad \vdots \quad (68)$$

$$(69)$$

That is, the migration rate  $m_{ab}$  is the  $(n(b-1)+a)$ -th variable of the system (we will later drop the null rates  $m_{ii}$ ). This allows us build the matrix of coefficients for the system for any number of locations of interest (see the script `all_locations/migratory_rates_all_locations.R`).

To compute the migration rates, we follow an analogous procedure to the case of ABC:

1. We obtain population data  $N_{ij}$  for 2010, 2015 and 2020 from UN data.
2. We assume that the total number of individuals from country  $i$  remains constant over the years. We manually adjust the data to force this assumption, changing the number of individuals from country  $i$  at country  $i$  to meet this requirement. This results in the populations in `all_locations/mig_pop_adj.csv`.
3. We use this data to compute the migration rates per 5-years during the period 2015-2020, based on 2015 population data and the variation in numbers from 2015 to 2020. This results in the rates in `all_locations/mig_rates_adj_2020.xlsx` for 2020. Note that these are the country-specific version of the rates.

## 10 SIR model with migration for ABC only

In this section, I will try to outline a simple SIR-type epidemiological model with migration. We start with the migration model given by Eq. (46) that describes the time-dependent dynamics of the number of individuals of nationality  $i$  in the country  $j$ ,  $N_{ij}(t)$  (including the case  $i = j$ ):

$$\frac{dN_{ij}}{dt} = \sum_{k=1}^n m_{jk} N_{ik} - \sum_{k=1}^n m_{kj} N_{ij}.$$

In this equation,  $m_{ij}$  is the per-capita rate of migration from country  $j$  to country  $i$  ( $m_{ii} = 0$  for all  $i$ ) that are estimated in Table 8.1.

In the simplest case, we stratify the population by disease status into susceptible, infectious, and recovered individuals,  $N_{ij}(t) = S_{ij}(t) + I_{ij}(t) + R_{ij}(t)$ . Migration applies to all types of individuals. Individuals become sexually active as susceptible and leave the population in any disease state. For simplicity, we assume that the rates of recruitment into and leaving the sexually active population are equal.

$$\frac{dS_{ij}}{dt} = -S_{ij}\beta_i \frac{1}{N_j} \sum_{k=1}^n I_{kj} + \mu N_{ij} - \mu S_{ij} + \sum_{k=1}^n m_{jk} S_{ik} - \sum_{k=1}^n m_{kj} S_{ij}, \quad (70)$$

$$\frac{dI_{ij}}{dt} = S_{ij}\beta_i \frac{1}{N_j} \sum_{k=1}^n I_{kj} - \nu I_{ij} - \mu I_{ij} + \sum_{k=1}^n m_{jk} I_{ik} - \sum_{k=1}^n m_{kj} I_{ij}, \quad (71)$$

$$\frac{dR_{ij}}{dt} = \nu I_{ij} - \mu R_{ij} + \sum_{k=1}^n m_{jk} R_{ik} - \sum_{k=1}^n m_{kj} R_{ij}. \quad (72)$$

Here,  $\nu$  and  $\mu$  are rates of recovery and becoming sexually active. The transmission rate for nationality  $i$  is  $\beta_i$ . Naturally, substituting in the above equations,  $N_{ij}(t) = S_{ij}(t) + I_{ij}(t) + R_{ij}(t)$ , and summing all equations leads to Eq. (46). Note that to close the above system of equations, we substitute the number of individuals residing in the country  $j$  as follows

$$N_j(t) = \sum_{i=1}^n N_{ij}(t) = \sum_{i=1}^n [S_{ij}(t) + I_{ij}(t) + R_{ij}(t)].$$

The comparison of the SIR transmission model with and without migration is done by setting  $m_{i,j} = 0$  for all  $i$  and  $j$ .

## 11 HIV model with migration for ABC only

We formulate a minimalistic model that describes HIV transmission in the population with migration. Individuals can be in the following disease stages: susceptible ( $S$ ), infected but untreated ( $I$ ), infected on antiretroviral treatment ( $T$ ), and deceased from AIDS without or after treatment ( $A$ ). The HIV model equations are

$$\frac{dS_{ij}}{dt} = -S_{ij}\beta_i \frac{1}{N_j} \sum_{k=1}^n (I_{kj} + \epsilon T_{kj}) + \mu N_{ij} - \mu S_{ij} + \sum_{k=1}^n m_{jk} S_{ik} - \sum_{k=1}^n m_{kj} S_{ij}, \quad (73)$$

$$\frac{dI_{ij}}{dt} = S_{ij}\beta_i \frac{1}{N_j} \sum_{k=1}^n (I_{kj} + \epsilon T_{kj}) - (\nu_1 + \tau + \mu) I_{ij} + \phi T_{ij} + \sum_{k=1}^n m_{jk} I_{ik} - \sum_{k=1}^n m_{kj} I_{ij}, \quad (74)$$

$$\frac{dT_{ij}}{dt} = \tau I_{ij} - (\nu_2 + \phi + \mu) T_{ij} + \sum_{k=1}^n m_{jk} T_{ik} - \sum_{k=1}^n m_{kj} T_{ij}, \quad (75)$$

$$\frac{dA_{ij}}{dt} = \nu_1 I_{ij} + \nu_2 T_{ij}. \quad (76)$$

Here  $N_j(t) = \sum_{i=1}^n N_{ij}(t) = \sum_{i=1}^n [S_{ij}(t) + I_{ij}(t) + T_{ij}(t)]$  because  $A_{ij}(t)$  denotes the number of individuals deceased from HIV in the AIDS stage. We will use only the first three equations excluding the equations for  $A_{ij}(t)$ . The model parameters are  $\beta_i$  - transmission rate for the nationality  $i$ ,  $\epsilon$  - infectivity of treated individuals relative to untreated,  $\mu$  - rate of recruitment to and leaving sexually active population,  $\nu_1$  and  $\nu_2$  - rates of HIV-related mortality for untreated and treated individuals,  $\tau$  - treatment uptake rate, and  $\phi$  - treatment dropout rate.

## 12 Model assumptions: HIV model with migration

### Migration model assumptions:

- Locations of interest (3 locations, ABC): Aruba, BES (grouped), Curacao.
- Interacting regions (4 regions): Latin America (comprising Central and Southern America), rest of the Caribbean region (comprising the Caribbean countries except the locations of interest), The Netherlands, and the rest of the world.
- For technical reasons, we assume that the global population for each of the locations of interest and interacting regions did not change during 2010-2020 (e.g., the total number of Arubans in the world is the same during 2010-2020). This is forced by manually adjusting the national population at each location of interest and interacting region, without changing the number of migrants.
- In each of Aruba, BES, Curacao, the resident population is stratified into 7 groups (3+4) by place of birth: those born in Aruba, BES, Curacao, Latin America, rest of Caribbean region, The Netherlands, and the rest of the world.
- For the migration model, we used the most recent migrant population data from the UNAIDS.
- Individuals in all disease states (living without HIV, living with undiagnosed HIV, living with diagnosed HIV and receive ART) can migrate to/and from ABC and the locations of interest.

- We assume that the rate at which individuals migrate from one location of interest/interacting region to another depends only on the location, and is independent of the individual's nationality (e.g., Arubans and Dutch living in Aruba migrate to the Netherlands at the same rate).
- Immediately upon migration individuals do not change their disease status. This means that individuals who live with undiagnosed HIV continue to be undiagnosed immediately upon migration (but can be diagnosed later on). This also means that individuals who live with diagnosed HIV and receive ART continue to be receiving ART immediately upon migration (but can drop off ART later on). NB! Here, "individuals" means people born in 7 (3+4) regions and residing in Aruba, BES, Curacao. There is no distinction in these migration rules by country of birth or residence of migrants.
- We know that some proportion of migrants on ABC are illegal, and we need estimates of the illegal migration (and possibly, where it comes from because it is not homogeneous across countries). The UNAIDS population division data contain the numbers of refugees and asylum seekers on ABC. Can I assume these numbers are equal to the number of illegal migrants? If not, which proportion of migrants is illegal per island (ABC)? This is very important for accessing the impact of offering ART to illegal migrants.

#### **Transmission model assumptions:**

- Individuals are stratified by disease states into living without HIV, living with undiagnosed HIV, living with diagnosed HIV and receive ART (virally suppressed). We assume that individuals can drop off ART and then move back to the group of people with undiagnosed HIV.
- HIV dynamics is modelled on ABC, but not in interacting regions.
- Average duration of sexual life (from sexual debut until the cessation of sexual activity): 45 years. This duration is assumed to be the same for migrants and locally born on ABC. This value is currently based on MSM population in the UK. A. Johnson, J. Wadsworth, K. Wellings, and J. Field, *Sexual attitudes and lifestyles*. Oxford: Blackwell Scientific Publications, 1994.
- Infectivity of diagnosed treated individuals relative to untreated undiagnosed individuals: 0.01 (so 1%). The infectivity is assumed the same for migrants and locally born on ABC. This captures that individuals need some time to get virally suppressed and a small proportion of individuals on treatment may develop resistant strain and transmit further until put back on the second-line regimen. The value is based on R. M. Granich, C. F. Gilks, C. Dye, K. M. De Cock, and B. G. Williams, "Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model," *Lancet*, vol. 373, pp. 48–57, 2009.

- Percentage of people dropping off ART annually (and thus moving from diagnosed treated group to undiagnosed untreated group): 5% (pure assumption). Assumed to be the same for migrants and locally born on ABC.
- Sexual behavior of individuals on each of ABC is the same. Sexual behavior is also the same for locally born and migrants on ABC. This is pure assumption.
- Sexual mixing on ABC: individuals mix homogeneously on ABC, i.e., probability to form a sexual partnership with a migrant or a local person depends only on the proportions of migrants or locals on ABC.
- Percentage of people starting ART annually (and thus moving from undiagnosed untreated group to diagnosed treated group): unknown. To know this, we would need to know how many people live with HIV on the islands overall. Assumed to be the same for migrants and locally born on ABC. How should we fix this?
- Average time between getting infected with HIV and death (related to HIV) if untreated: 11.08 years. Assumed to be the same for migrants and locally born. This value is taken from Collaborative Group on AIDS Incubation and HIV Survival including the CASCADE EU Concerted Action. Concerted Action on SeroConversion to AIDS and Death in Europe. “Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis,” *Lancet*, vol. 355, pp. 1131–1137, 2000; P. J. Birrell, A. M. Presanis, D. De Angelis, and The CASCADE Collaboration, “Multi-state models of HIV progression in homosexual men: an application to the CASCADE collaboration,” tech. rep., MRC Biostatistics Unit, 2012.
- Average time between getting diagnosed and treated and death (related to HIV, e.g., cardiovascular diseases, etc.) if treated: unknown. Assumed to be the same for migrants and locally born. This value depends on how late into their HIV infection people get diagnosed and start treatment. For MSM, I used 67.4 years.

## 13 HIV model with migration for ABC and interacting regions

We note that our intention is to model HIV dynamics on ABC and not on interacting regions. In total, we have 3 locations of interest and 7 ( $n-3$ ) interacting regions. We extend the HIV model equations from the previous section as follows. As before, the first index  $i$  denotes the country of birth and the second index  $j$  denotes the country of residence.

For  $i = 1, \dots, n$  and  $j = 1, 2, 3$  we have transmission and migration dynamics:

$$\begin{aligned}
\frac{dS_{ij}}{dt} &= -S_{ij}\beta_i \frac{1}{N_j} \sum_{k=1}^n (I_{kj} + \epsilon T_{kj}) + \mu N_{ij} - \mu S_{ij} + \\
&\quad + \sum_{k=1}^3 m_{jk} S_{ik} + \sum_{k=4}^n m_{jk} N_{ik} (1 - P_k) - \sum_{k=1}^n m_{kj} S_{ij}, \\
\frac{dI_{ij}}{dt} &= S_{ij}\beta_i \frac{1}{N_j} \sum_{k=1}^n (I_{kj} + \epsilon T_{kj}) - (\nu_1 + \tau + \mu) I_{ij} + \phi T_{ij} + \\
&\quad + \sum_{k=1}^3 m_{jk} I_{ik} + \sum_{k=4}^7 m_{jk} N_{ik} P_k (1 - f_k) - \sum_{k=1}^n m_{kj} I_{ij}, \tag{77}
\end{aligned}$$

$$\begin{aligned}
\frac{dT_{ij}}{dt} &= \tau I_{ij} - (\nu_2 + \phi + \mu) T_{ij} + \\
&\quad + \sum_{k=1}^3 m_{jk} T_{ik} + \sum_{k=4}^7 m_{jk} N_{ik} P_k f_k - \sum_{k=1}^n m_{kj} T_{ij}, \tag{78}
\end{aligned}$$

$$\frac{dA_{ij}}{dt} = \nu_1 I_{ij} + \nu_2 T_{ij}, \tag{79}$$

where as before  $N_j(t) = \sum_{i=1}^n N_{ij}(t) = \sum_{i=1}^n [S_{ij}(t) + I_{ij}(t) + T_{ij}(t)]$ .

For  $i = 1, \dots, n$  and  $j = 4, \dots, n$  we have only migration dynamics:

$$\frac{dN_{ij}}{dt} = \sum_{k=1}^n m_{jk} N_{ik} - \sum_{k=1}^n m_{kj} N_{ij}.$$

Assumptions involved in this model formulation:

- Proportion  $(1-P_k)$  of individuals residing in the interacting regions,  $N_{ik}$ , where  $i = 1, \dots, n$  and  $k = 4, \dots, n$  migrate as susceptible, and, respectively,  $P_k$  proportion migrate either as infected untreated or as infected treated.
- $P_k$  is, thus, HIV prevalence in the interacting region  $k = 4, \dots, n$  which, for simplicity, is assumed to be independent of the country of birth (first index  $i$ ) and constant in time.
- $f_k$  is the proportion of infected individuals on treatment in the interaction region  $k = 4, \dots, n$ . This proportion is assumed to be independent of the country of birth and constant with time too.
- We need estimates of  $P_k$  (true prevalence = proportion of the total population living with HIV) and  $f_k$  (proportion of all those living with HIV on treatment and suppressed) for the four interacting regions [Latin America (comprising Central and Southern America), rest of the Caribbean region (comprising the Caribbean countries except the locations of interest), The Netherlands, and the rest of the world].



- Importantly, we assume that there is an influx of infectious migrants into ABC due to migration but that individuals born on ABC who migrate to the interacting regions don't get infected there. Should we change this? In a simplified way?

## 14 Model extension for illegal migration

We know that some proportion of migrants are illegal, and we need estimates of the illegal migration (and possibly, where it comes from because it is not homogeneous across countries). For the moment, I propose to use the numbers of refugees and asylum seekers from the UNAIDS population division data to calculate the proportion of migrants who are illegal on ABC (I don't really know whether this assumption is OK, will check). What is known about illegal migrants is that some of them get tested and get diagnosed with HIV but these individuals are not offered subsequent treatment because they don't have health insurance. The question in which stakeholders are interested and which we could address is whether offering treatment to these individuals could reduce the number of HIV infections on ABC.

In the model, I assume that illegal migrants get tested and diagnosed at the same rate  $\tau$  as legal migrants or individuals born on ABC. However, only legal migrants or individuals born on ABC are offered treatment. Therefore,  $I_{ij}$  individuals become  $T_{ij}$  ( $i = 1, \dots, n$ ,  $j = 1, 2, 3$ ) with rate  $\tau[1 - g_j(1 - \delta_{ij})]$ , where  $\delta_{ij} = 1$  for  $i = j$ ,  $j = 1, 2, 3$  and  $\delta_{ij} = 0$  for  $i \neq j$ ,  $j = 1, 2, 3$ . Parameter  $g_j$  ( $j = 1, 2, 3$ ) can be interpreted as the fraction of migrants who are illegal on ABC, which we assume to be independent of the country of birth of migrants. The respective equations are modified as follows:

For  $i = 1, \dots, n$  and  $j = 1, 2, 3$  we have transmission and migration dynamics:

$$\begin{aligned} \frac{dS_{ij}}{dt} = & -S_{ij}\beta_i \frac{1}{N_j} \sum_{k=1}^n (I_{kj} + \epsilon T_{kj}) + \mu N_{ij} - \mu S_{ij} + \\ & + \sum_{k=1}^3 m_{jk} S_{ik} + \sum_{k=4}^n m_{jk} N_{ik} (1 - P_k) - \sum_{k=1}^n m_{kj} S_{ij}, \end{aligned} \quad (80)$$

$$\begin{aligned} \frac{dI_{ij}}{dt} = & S_{ij}\beta_i \frac{1}{N_j} \sum_{k=1}^n (I_{kj} + \epsilon T_{kj}) - (\nu_1 + \tau[1 - g_j(1 - \delta_{ij})] + \mu) I_{ij} + \phi T_{ij} + \\ & + \sum_{k=1}^3 m_{jk} I_{ik} + \sum_{k=4}^n m_{jk} N_{ik} P_k (1 - f_k) - \sum_{k=1}^n m_{kj} I_{ij}, \end{aligned} \quad (81)$$

$$\begin{aligned} \frac{dT_{ij}}{dt} = & \tau[1 - g_j(1 - \delta_{ij})] I_{ij} - (\nu_2 + \phi + \mu) T_{ij} + \\ & + \sum_{k=1}^3 m_{jk} T_{ik} + \sum_{k=4}^n m_{jk} N_{ik} P_k f_k - \sum_{k=1}^n m_{kj} T_{ij}, \end{aligned} \quad (82)$$

$$\frac{dA_{ij}}{dt} = \nu_1 I_{ij} + \nu_2 T_{ij}, \quad (83)$$

where as before  $N_j(t) = \sum_{i=1}^n N_{ij}(t) = \sum_{i=1}^n [S_{ij}(t) + I_{ij}(t) + T_{ij}(t)]$ .

For  $i = 1, \dots, n$  and  $j = 4, \dots, n$  we have only migration dynamics:

$$\frac{dN_{ij}}{dt} = \sum_{k=1}^n m_{jk} N_{ik} - \sum_{k=1}^n m_{kj} N_{ij}.$$

Summary:

- $g_j$  ( $j = 1, 2, 3$ ) is the proportion of migrants who are illegal on ABC.
- $g_j$  will be taken from the UNAIDS data (i.e., the proportion of migrant stock on ABC who are refugees and asylum seekers for the most recent year).

## 15 Model extension for mixing of migrant and local populations

We extend the model to incorporate heterogeneity in transmission between “locals” and migrants in the locations of interest. We assume that partner change rates,  $c_i$ , defined as the rate of acquisition of sexual partners per unit of time, depend on the nationality  $i$ ,  $i = 1, \dots, n$  (first index). We further define the  $n \times n$  mixing matrix  $\mathbf{M}(t) = [M_{ik}(t)]_{i,k \in \{1, \dots, n\}}$ , where the elements  $M_{ik}(t)$  denote sexual mixing of individuals of the nationality  $i$  and individuals of the nationality  $k$ . To be as general as possible, we allow mixing to vary between fully proportionate to fully assortative by introducing the mixing parameter  $\omega \in [0, 1]$  as follows

$$M_{ik}(t) = \omega \frac{c_k N_{kj}(t)}{\sum_{k'=1}^n c_{k'} N_{k'j}(t)} + (1 - \omega) \delta_{ik}, \quad (84)$$

where  $i, k, k' = 1, \dots, n$ ,  $j = 1, 2, 3$  and  $\delta_{ik} = 1$  for  $i = k$ ,  $i, k = 1, \dots, n$  and  $\delta_{ik} = 0$  otherwise.

In the general case, Eq. (84) means that a proportion  $(1 - \omega)$  of the partnerships of individuals of nationality  $i$  are formed with individuals of the same nationality, whereas the remaining proportion  $\omega$  of the partnerships is formed with each nationality  $k$  ( $k = 1, \dots, n$ ) proportionally to the number of partnerships offered by those nationalities in the location of interest  $j$  ( $j = 1, 2, 3$ ). Therefore, when  $\omega = 0$ , mixing is fully assortative

$$M_{ik}(t) = \delta_{ik}. \quad (85)$$

When  $\omega = 1$ , mixing is fully proportionate

$$M_{ik}(t) = \frac{c_k N_{kj}(t)}{\sum_{k'=1}^n c_{k'} N_{k'j}(t)}. \quad (86)$$

We write the infection term in the equation for susceptibles as follows

$$\frac{dS_{ij}(t)}{dt} = -S_{ij}(t) \lambda c_i \sum_{k=1}^n M_{ik}(t) \left( \frac{I_{kj}(t)}{N_{kj}(t)} + \epsilon \frac{T_{kj}(t)}{N_{kj}(t)} \right) + \dots, \quad (87)$$

where  $M_{ik}(t)$  is given by Eq. (84),  $\lambda$  is probability of transmission per partnership, and  $c_i$  is the partner change rate of nationality  $i$ .

We further consider two special cases of fully proportionate and fully assortative mixing. Note that, for simplicity, and in the absence of detailed data on sexual behavior, we assume that partner change rates  $c_i \equiv c$  are the same for all nationalities  $i$  ( $i = 1, \dots, n$ ).

**Proportionate mixing.** We substitute  $\omega = 1$  and  $c_i \equiv c$  ( $i = 1, \dots, n$ ). The mixing matrix becomes

$$M_{ik}(t) = \frac{N_{kj}(t)}{\sum_{k'=1}^n N_{k'j}(t)} = \frac{N_{kj}(t)}{N_j(t)}. \quad (88)$$

The equation for the susceptibles becomes

$$\frac{dS_{ij}(t)}{dt} = -S_{ij}(t)\lambda c_i \sum_{k=1}^n \left( \frac{I_{kj}(t)}{N_j(t)} + \epsilon \frac{T_{kj}(t)}{N_j(t)} \right) + \dots \quad (89)$$

Note that with substituting  $\lambda c_i \equiv \beta_i$  we arrive at Eq. (80).

**Assortative mixing.** We substitute  $\omega = 0$  and  $c_i \equiv c$  ( $i = 1, \dots, n$ ). The mixing matrix becomes

$$M_{ik} = \delta_{ik}. \quad (90)$$

The equation for the susceptibles becomes

$$\frac{dS_{ij}(t)}{dt} = -S_{ij}(t)\lambda c_i \left( \frac{I_{ij}(t)}{N_{ij}(t)} + \epsilon \frac{T_{ij}(t)}{N_{ij}(t)} \right) + \dots \quad (91)$$

**General mixing.** Below are given the equations for the general mixing that I implement in the Mathematica notebook.

For  $i = 1, \dots, n$  and  $j = 1, 2, 3$  we have transmission and migration dynamics:

$$\begin{aligned} \frac{dS_{ij}}{dt} = & -S_{ij}\lambda c_i \sum_{k=1}^n M_{ik}(t) \left( \frac{I_{kj}}{N_{kj}} + \epsilon \frac{T_{kj}}{N_{kj}} \right) + \mu N_{ij}^0 - \mu S_{ij} + \\ & + \sum_{k=1}^3 m_{jk} S_{ik} + \sum_{k=4}^n m_{jk} N_{ik} (1 - P_k) - \sum_{k=1}^n m_{kj} S_{ij}, \end{aligned} \quad (92)$$

$$\begin{aligned} \frac{dI_{ij}}{dt} = & S_{ij}\lambda c_i \sum_{k=1}^n M_{ik}(t) \left( \frac{I_{kj}}{N_{kj}} + \epsilon \frac{T_{kj}}{N_{kj}} \right) - (\nu_1 + \tau[1 - g_j(1 - \delta_{ij})] + \mu) I_{ij} + \phi T_{ij} + \\ & + \sum_{k=1}^3 m_{jk} I_{ik} + \sum_{k=4}^7 m_{jk} N_{ik} P_k (1 - f_k) - \sum_{k=1}^n m_{kj} I_{ij}, \end{aligned} \quad (93)$$

$$\begin{aligned} \frac{dT_{ij}}{dt} = & \tau[1 - g_j(1 - \delta_{ij})] I_{ij} - (\nu_2 + \phi + \mu) T_{ij} + \\ & + \sum_{k=1}^3 m_{jk} T_{ik} + \sum_{k=4}^7 m_{jk} N_{ik} P_k f_k - \sum_{k=1}^n m_{kj} T_{ij}, \end{aligned} \quad (94)$$

$$\frac{dA_{ij}}{dt} = \nu_1 I_{ij} + \nu_2 T_{ij}, \quad (95)$$

where as before  $N_j(t) = \sum_{i=1}^n N_{ij}(t) = \sum_{i=1}^n [S_{ij}(t) + I_{ij}(t) + T_{ij}(t)]$ .

For  $i = 1, \dots, n$  and  $j = 4, \dots, n$  we have only migration dynamics:

$$\frac{dN_{ij}}{dt} = \sum_{k=1}^n m_{jk} N_{ik} - \sum_{k=1}^n m_{kj} N_{ij}.$$

Note, the recruitment to the susceptible population occurs with a constant rate  $\mu N_{ij}^0$ , where  $N_{ij}^0 \equiv N_{ij}(t = 0)$ .

## 16 Input data

Three more sources of data are required for the model:

1. **Estimates for HIV prevalence in the interacting regions.** We downloaded estimates of HIV prevalence for all the countries in the world from Global Health Data Exchange (GHDx), a global health data catalog. We then combined the prevalences of the countries in each interacting region to obtain an approximation of the prevalence at the whole region, as follows:

$$prev(\text{region } i) = \sum_k \frac{\text{population at country } k}{\text{total population at region } i} prev(\text{country } k),$$

where *prev* denotes prevalence, and the sum runs over all the countries in region *i*. The resulting data for 2019, the latest year available in GHDx (expressed as number of cases/inhabitant), can be found in `all_locations/HIV_prevalences_aggregated.xlsx`, together with the total population at each interacting region. 95% confidence intervals are obtained by performing the same computation as in Eq.(1) on the confidence intervals provided by GHDx (note that this procedure is not statistically correct).

2. **Proportion of undocumented migrants.** We extracted from the UN migrant stock data tables the estimated proportion of migrants that are refugees or asylum seekers at the locations of interest. This data was available for Aruba (year 2020) and Curaçao (years 2010, 2015, 2020), as seen in `all_locations/proportion_undocumented_migrants_ABC.xlsx`.
3. **Population aged 15-64.** We processed UN world's population data to obtain the proportion of the population aged 15-64 at the locations of interest and the interacting regions, resulting in the data in `all_locations/pop_ages.xlsx`.