Title: Adaptive data-driven age and patch mixing in contact networks with recurrent mobility

Authors: Jesse Knight[1,2,a,*], Huiting Ma[1,b], Amir Ghasemi[3], Mackenzie Hamilton[1,d], Kevin Brown[4,5,e], and Sharmistha Mishra[1,2,5,6,f]

[1]MAP Centre for Urban Health Solutions, Unity Health Toronto
[2]Institute of Medical Science, University of Toronto
[3]Communications Research Centre Canada, Ottawa
[4]Public Health Ontario, Canada
[5]Dalla Lana School of Public Health, University of Toronto
[6]Division of Infectious Diseases, Department of Medicine, University of Toronto
[a]ORCID: 0000-0002-0455-5455
[b]ORCID: 0000-0003-1910-5614
[d]ORCID: 0000-0003-2423-3629
[e]ORCID: 0000-0002-1483-2188
[f]ORCID: 0000-0001-8492-5470
[*]Corresponding Author: Jesse Knight. jesse.knight@mail.utoronto.ca

Keywords: COVID-19, Mathematical modelling, Heterogeneous mixing, Population mobility

Journal: MethodsX (preprint)

Submission: Direct

Subject Area: Bioinformatics: Epidemic modelling

Method: A. Arenas et al. "Modeling the Spatiotemporal Epidemic Spreading of COVID-19 and the Impact of Mobility and Social Distancing Interventions". In: *Physical Review X* 10.4 (2020), p. 041055

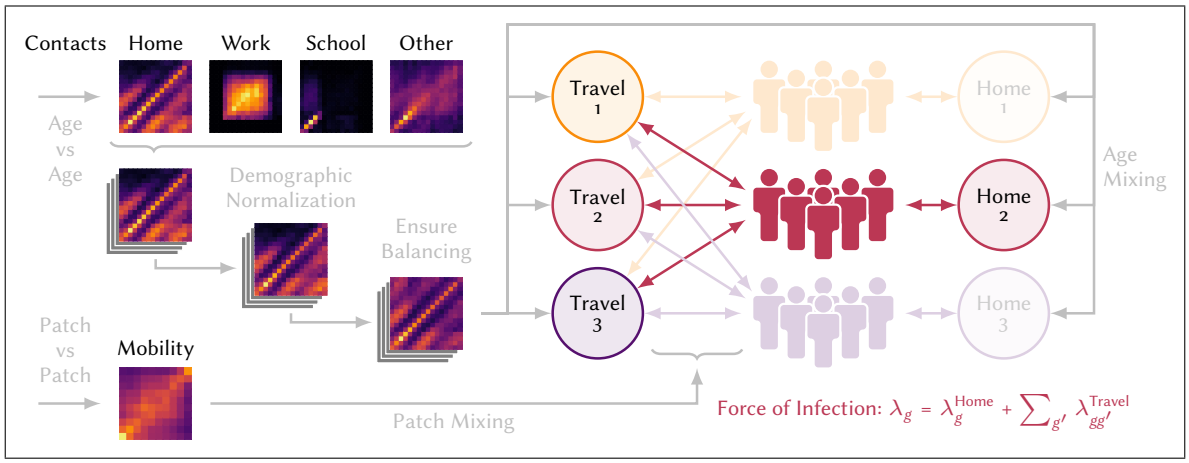
Resource: https://github.com/mishra-lab/covid-on-model

Date: July 25, 2021

Abstract:

Infectious disease transmission models often stratify populations by age and geographic patches. Contact patterns between age groups and patches are key parameters in such models. Arenas et al. (2020) develop an approach to simulating contact patterns associated with recurrent mobility between patches, such as due to work, school, and other regular travel. We built upon this approach to address several potential weaknesses. In our approach, the distribution of contacts by age is responsive to underlying age distribution of the mixing pool. Different age distributions by contact type are also maintained in our model, such that changes to the numbers of different types of contacts are appropriately reflected in changes to the overall age mixing patterns. Finally, we introduce and distinguish between two mixing pools associated with each patch, with possible implications for the overall connectivity of the population: the home pool, in which contacts can only be formed with other individuals residing in the same patch; and the travel pool, in which contacts can be formed with some residents and any other visitor to the patch. We describe in detail the steps required to implement our approach, and present results of an example application.

Graphical Abstract:



1 Introduction

Contact patterns are key determinants of epidemic dynamics. Arenas et al. [1] develop a patch-based model of COVID-19 transmission applied to Spain, in which the modelled population is stratified by geographic patches and three age groups. The model incorporates data on short, recurrent mobility patterns to determine contact rates between individuals in different patches and age groups, following foundational work by Balcan and Vespignani [2] and Sattenspiel and Dietz [3]. We build upon this contact model to incorporate age mixing patterns for different contact types which are responsive to the age distributions of mixing populations, as proposed by Arregui et al. [4]. We also explore some practical challenges in parameterizing such models.

2 Method

Consider a population stratified by N_g patches and N_a age groups.[1] Let P_{ga} be the number of people in patch g and age group a. Let y denote N_y different types of contacts. Let $B_{gg'}$ be the proportion of population P_g who travel to g' each day, or the "mobility matrix".[2]

2.1 Original Approach

Arenas et al. [1] model the force of infection (incidence) experienced by population P_{ga} as:

$$\lambda_{ga}(t) = (1 - \rho_a)\,\Phi_{ga}(t) + \rho_a \sum_{g^*} B_{gg^*a}\Phi_{g^*a}(t) \tag{1}$$

where: $\Phi_{ga}(t)$ is the probability of acquiring infection while in patch g; and $\rho_a \in [0, 1]$ is an age-specific overall mobility factor. Thus, $\lambda_{ga}(t)$ is the sum of infection probabilities from the residence patch g, and from visited patches $g^* \neq g$. The probability $\Phi_{g^*a}(t)$ is modelled using the chained binomial for multiple exposures [5]:

$$\Phi_{g^*a}(t) = 1 - \prod_{a'}\prod_{g'}\prod_{i'}(1 - \beta_{i'})^{f_{g^*} C_a \theta_{aa'} \Omega_{g^* g' i' a'}} \tag{2}$$

where: β_i is the per-contact transmission probability associated with infectious state i; f_{g^*} is a density factor associated with patch g^*; C_a is the expected number of contacts made per person per day in age group a; $\theta_{aa'}$ is the age distribution of those contacts, derived from [6] for Spain, such that $\sum_{a'} \theta_{aa'} = 1$; and $\Omega_{g^* g' i' a'}$ is the proportion of individuals present in patch g^* who reside in patch g' and who are in infectious state i', for each age group a'. This proportion $\Omega_{g^* g' i' a'}$ is defined as:

$$\Omega_{g^* g' a' i'} = \frac{P_{g' i' a'} M_{g' g^* a'}}{\sum_{g' i'} P_{g' i' a'} M_{g' g^* a'}} \tag{3}$$

[1] We use different notation than Arenas et al. [1]; a comparison is given in Table A.1.

[2] Arenas et al. [1] consider different mobility patterns by age: $B_{gg'a}$. For simplicity, we consider $B_{gg'}$ unstratified by age, but age stratification could be added to our approach.

where $M_{gg'a}$ is a convenience simplification of the mobility matrix:

$$M_{gg'a} = (1 - \rho_a)\,\delta_{gg'} + \rho_a B_{gg'a} \tag{4}$$

This model of infection captures important mixing patterns related to recurrent mobility that are relevant to epidemic modelling on relatively small spatial and time scales. However, the model could be improved by separating different contact types throughout the force of infection equation, and by not fixing the age distribution of contacts. Three specific issues with the original approach are as follows.

1. **Contact balancing:** The contact balancing principle states that the total number of contacts formed by group a with group a' should equal the number formed by group a' with group a [4]:

$$P_a C_a \theta_{aa'} = P_{a'} C_{a'} \theta_{a'a} \tag{5}$$

   For a model with non-random age mixing but random mixing by patches, Eq. (5) could be satisfied by a single fixed age mixing matrix $\theta_{aa'}$, i.e. for the population overall. However, in the context of patch-based mixing reflecting recurrent mobility, Eq. (5) should be satisfied in each mixing context (patch). If different patches have different age distributions, then it would not be possible to satisfy Eq. (5) with a single fixed age mixing matrix $\theta_{aa'}$. The implications of violating Eq. (5) depend on the context. For example, if $\theta_{aa'}$ overestimates the proportion of contacts formed with a highly infectious age group, then the modelled incidence would be higher than it should be, etc.

2. **Age mixing by contact type:** A related issue is that the expected contact rates by age group C_a reflect the summation of different types in [1], and so the fixed age mixing matrix $\theta_{aa'}$ is applied to all contact types. Then, when restrictions are simulated, overall average contact rates C_a are reduced by an amount reflecting the proportion of work versus home contacts (and various related assumptions), but $\theta_{aa'}$ is not updated. As illustrated by the POLYMOD study [7], age mixing patterns vary by contact type. Thus, differential reductions in each contact type would affect age mixing patterns. For example, if reduced contacts between working-aged adults were not reflected in $\theta_{aa'}$, then the relative contribution of this age group to transmission could be overestimated.

3. **Modelling contact & mobility reductions:** The term $(1-\rho_a)\,\Phi_{ga}(t)$ in Eq. (1) represents transmission to non-mobile individuals in patch g. The associated definitions in Eqs. (2–4) consider transmission to these individuals from visitors to patch g. Such definitions therefore imply that non-mobile individuals still form contacts with visitors to their residence patch. However, when simulating confinement measures in [1], mobility reductions are always paired with proportional reductions in non-home contacts, suggesting that non-mobile individuals form only home contacts. It seems incongruous to model the formation of home contacts among non-mobile residents of patch g with mobile visitors to patch g. As illustrated in Figure A.1, this approach potentially overestimates the contact network connectivity during reduced mobility versus an approach in which non-mobile populations do not contact visitors, and thus could underestimate the impact of confinement strategies. While it may be sometimes desirable to allow mixing of "non-mobile" individuals with visitors to their residence patch, the original approach does not provide parameterization to prevent this from happening.

We therefore develop a refinement of the original approach, with the aim of addressing these three issues.

2.2 Proposed Approach

In the proposed approach, the contributions of different contact types to the force of infection are added to the binomial function for multiple exposures:

$$\lambda_{ga}(t) = 1 - \prod_{y}\prod_{g'}\prod_{a'}\prod_{i'}(1-\beta_{i'})^{C_{gag'a'y}\Omega_{g'a'i'}} \tag{6}$$

where: $C_{gag'a'y}$ is the expected number of type y contacts formed per person per day among individuals in population P_{ga} with those in population $P_{g'a'}$; and $\Omega_{g'a'i'}$ is the proportion of individuals in residing in patch g' and age group a' who are in infectious state i':

$$\Omega_{g'a'i'} = \frac{P_{g'a'i'}}{P_{g'a'}} \tag{7}$$

For each type of contact, $C_{gag'a'y}$ is defined to reflect both age-related and mobility-related mixing factors, as described in the following subsections. To support these descriptions, we will refer to Figures from an example application, although the details of the application and the Figures are given in § 3. Collecting the full network of contacts in the matrix $C_{gag'a'y}$ is parsimonious, and allows us to easily compute various properties, like the margins in a, a' or g, g', and whether contact balancing is satisfied per Eq. (5). Additionally, separating contact types allows the incorporation of different probabilities of transmission per contact type $\beta_{i'y}$, if desired.

2.2.1 Age Mixing

Prem, Cook, and Jit [6] project contact patterns by 5-year age groups from the POLYMOD study [7] onto 152 countries, considering various demographic data. These contact matrices represent $C_{aa'y}$: the expected number of type y contacts formed per day among individuals in age group a with those in age group a'. Four types of contact are considered: "home", "work", "school", and "others"[3] (Figure 4a). We aim to incorporate these contact numbers and patterns into $C_{gag'a'y}$.

The first challenge is that these contact matrices $C_{aa'y}$ are inherently weighted by the underlying population age distribution—the proportion of expected contacts with age group a' is proportional to the size of age group a'. To overcome this challenge and apply these patterns to new population age structures, Arregui et al. [4] suggest to divide by the population age distribution to obtain an "unweighted" matrix $C^u_{aa'y}$ (Figure 4b):[4]

$$C^u_{aa'y} = C_{aa'y}\frac{\bar{P}}{P_{a'}} \tag{8}$$

where $\bar{P}$ is the mean of $P_{a'}$.

[3] The "others" contact type in [6] is itself derived from the combination of "leisure", "transport", and "other" contact types from [7], while the "home", "work", and "school" types are the same between the studies.

[4] The matrix $C^u_{aa'y}$ could also be interpreted as the expected contact matrix for a population with a rectangular demographic pyramid [4].

The next challenge is that $C^u_{aa'y}$ may not satisfy the contact balancing principle, Eq. (5), due to potential differences in age distribution of the POLYMOD survey respondents versus that of the contacts the respondents reported. In order to ensure that $C_{gag'a'y}$ will satisfy the balancing principle, $C^u_{aa'y}$ must satisfy the principle. A simple solution is to average $C^u_{aa'y}$ with its transpose to obtain the "balanced" matrix $C^{ub}_{aa'y}$ (Figure 4C):

$$C^{ub}_{aa'y} = \frac{1}{2}\left[C^u_{aa'y} + C^u_{aa'y}{}^T\right] \tag{9}$$

This operation may change the margin C_{ay}, representing the total type y contacts formed by individuals in age group a. However, such changes are reasonable if understood as a correction for sampling bias.[5]

A final challenge in applying the contact matrices from [6] is that the 5-year age groups may not align with the age groups of interest. Overcoming this challenge is not theoretically required to obtain $C_{gag'a'y}$, but we describe a solution here in case it is useful for modelling applications. We begin by upsampling the contact matrix from 5-year age groups a_5 to 1-year age groups a_1 using bilinear interpolation, based on the midpoints of each age group, and scaled by a factor of $1/5$. To avoid edge effects associated with many interpolation implementations, we first pad the matrix by replicating the edges diagonally. If the desired age groups extend beyond the maximum age group of 80 available in [1], diagonal padding can also be used to approximate the trends in the additional age groups. Then, given the age groups of interest a_* (which may have irregular widths), we aggregate $C^{ub}_{a_1a'_1y}$ to obtain $C^{ub}_{a_*a'_*y}$ using matrix multiplication with indicator matrix A:

$$C^{ub}_{a_*a'_*y} = \frac{A_{a_*a_1}}{\sum_{a_1} A_{a_*a_1}} C^{ub}_{a_1a'_1y} A^T_{a'_*a'_1}, \qquad A_{a_*a_1} = \begin{cases} 1, & a_1 \in a_* \\ 0, & a_1 \notin a_* \end{cases} \tag{10}$$

The right-hand A^T term *sums* the total number of contacts formed with the 1-year "other" age groups a'_1 corresponding to a'_*. The left-hand A term *averages* the total number of contacts formed from the 1-year "self" age groups a_1 corresponding to a_*. The average weights each 1-year age group a_1 equally, although other weights could be incorporated through the nonzero values of A. Another interpretation of the normalization sum is the widths of the age groups a_*.

The resulting matrix $C^{ub}_{a_*a'_*y}$ represents the expected contacts among age groups a_* when mixing with a population having equal proportion in all age groups a'_* (regardless of their width). Thus, $C^{ub}_{a_*a'_*y}$ can later be multiplied by the population age distribution of interest—reversing Equation (8)—to obtain the expected number of contacts when mixing with that population. This approach then addresses issues 1 and 2 described in § 2.1.

2.2.2 Mobility-Related Mixing

For mobility-related mixing, we use the same mobility matrix $B_{gg'}$ as [1], representing the expected proportions of individuals in patch g who travel to patch g' each day. We do not assume that rows of $B_{gg'}$ sum to 1; however, we do assume that the diagonal ($g = g'$) represents the proportion of individuals who are mobile within their own patch. Thus $1 - \sum_{g'} B_{gg'}$ represents the proportion of individuals in patch g' who are not mobile. In § A.3 we discuss some details about generating such a matrix $B_{gg'}$.

[5] A perfect survey in a closed population would produce a contact matrix $C^u_{aa'y}$ that is already balanced.

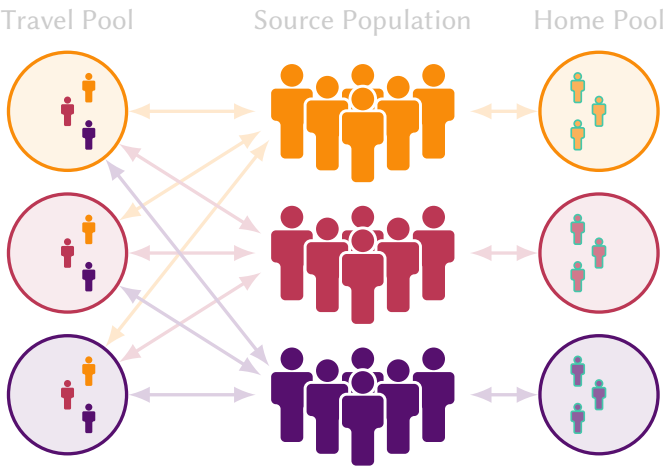


Figure 1: Toy example of "home" vs "travel" mixing pools for a network with 3 patches and 50% individuals mobile. Contacts in the home pool are formed exclusively with other members of the residence patch, whereas contacts in the travel pool may be formed with any visitors to the patch

Non-mobile populations are indicated with faded colour and green outline

The distinction between "not mobile" and "mobile within the residence patch" is new versus the original approach [1]. As described in § 2.1, in the original approach, individuals with reduced mobility were assumed to form fewer contacts, but continue mixing with visitors to their residence patch. Our proposed distinction allows us to separate two "mixing pools" where contacts might be formed: "home pools", where contacts are formed exclusively with other residents of the same patch g (e.g. household contacts); and "travel pools", where contacts are formed with any mobile individuals visiting patch g, which may include individuals who are mobile within their residence patch and travellers from other patches (e.g. work contacts). An illustration of these pools is given in Figure 1.

We assume that individuals who are not mobile will only form contacts within the home pool, but that the number of contacts formed per person within the home pool is fixed and not affected by the proportion of individuals who are not mobile. It is not necessary to assume that all contacts of any particular type are formed with only one type of pool; rather we introduce a parameter $h_y \in [0, 1]$ representing the proportion of type y contacts that are formed with the home pool. For example, we could have $h_y = 1$ for household contacts, $h_y = 0$ for work contacts, and perhaps $h_y = 0.5$ for school contacts.

To calculate $C_{gag'a'y}$ using these assumptions, we begin by considering the travel pool in patch g^*. The effective number of individuals from population P_{ga} who are present in the pool is given by:[6]

$$P^{g^*}_{gay} = h_y P_{ga} B_{gg^*} \tag{11}$$

If we assume that mixing by residence patch g within the pool is random, we need only consider age mixing within the pool. Under completely random mixing, the total number of contacts formed between $P^{g^*}_{gay}$ and

[6] If residents of different patches might have relatively different numbers of contacts, a scaling factor could be applied here.

$P^{g^*}_{g'a'y}$ is given by the outer product:

$$X^{g^*r}_{gag'a'y} = P^{g^*}_{gay} \otimes P^{g^*}_{g'a'y} \Big/ \sum_{g'a'} P^{g^*}_{g'a'y} \tag{12}$$

where the first term represents the absolute population size of "self", and the second term represents proportions of their contacts among "other" strata. Then, the patterns of age mixing can be applied via multiplication:

$$X^{g^*}_{gag'a'y} = X^{g^*r}_{gag'a'y} \; C^{ub}_{aa'y} \Big/ \sum_{a'_1} A_{a'a'_1} \tag{13}$$

since $X^{g^*r}_{gag'a'y}$ is proportional to the population age distribution of "others", and will therefore act to reverse Eq. (8) as planned. The term $\sum_{a'_1} A_{a'a'_1}$ is from Eq. (10), representing the widths of the age groups a'. It is necessary to divide by the widths of age groups a' since both $X^{g^*r}_{gag'a'y}$ and $C^{ub}_{aa'y}$ are proportional to these widths, but the proportionality should only be singular overall. We could have applied this normalization to $C^{ub}_{aa'y}$ in Eq. (10) in the same way as for a, but this would make $C^{ub}_{aa'y}$ harder to interpret, as it would no longer represent the expected numbers of contacts for each age group.

Mixing within home pools can be modelled similar to mixing within travel pools, with one small modification: replacing $B_{gg'}$ with the identity matrix $\delta_{gg'}$. Following through Eqs. (11–12), we obtain $X^{h}_{gag'a'y}$, representing the total contacts formed within home pools. Then, the total type y contacts formed between populations P_{ga} and $P_{g'a'}$ across all relevant mixing pools is given by the sum:

$$X_{gag'a'y} = X^{h}_{gag'a'y} + \sum_{g^*} X^{g^*}_{gag'a'y} \tag{14}$$

It may be tempting to simplify the model for home pool contacts by updating the mobility matrix $B_{gg'}$ similar to Eq. (4), with $h_y = (1 - \rho_a)$. However, such an approach does not produce the same result as Eq. (14), and indeed underpins issue 3 described in § 2.1 regarding mixing of non-mobile individuals with mobile visitors to their patch. On the other hand, if the interpretation of "non-mobile" is intended to allow mixing with visitors, then $B_{gg'}$ can still be adjusted per Eq. (4) to simulate this behaviour.

Finally, the number of type y contacts formed *per person* in population P_{ga} with population $P_{g'a'}$ can be obtained by dividing $X_{gag'a'y}$ by the population size:

$$C_{gag'a'y} = \frac{X_{gag'a'y}}{P_{ga}} \tag{15}$$

3 Example

We applied the proposed methodology for generating a mixing matrix $C_{gag'a'y}$, which reflects patterns of age mixing, recurrent mobility between patches, and different contact types, to the population of Ontario, Canada, in the context of COVID-19 transmission modelling. Ten patches were defined based on groupings of the 513 forward sortation areas (FSAs)[7] in Ontario. The FSA groupings reflect deciles of cumulative COVID-19 cases between Dec 2020–May 2021. Thus each patch represents approximately 10% of the Ontario population

[7] Each FSA is the first 3 characters of the postal code.

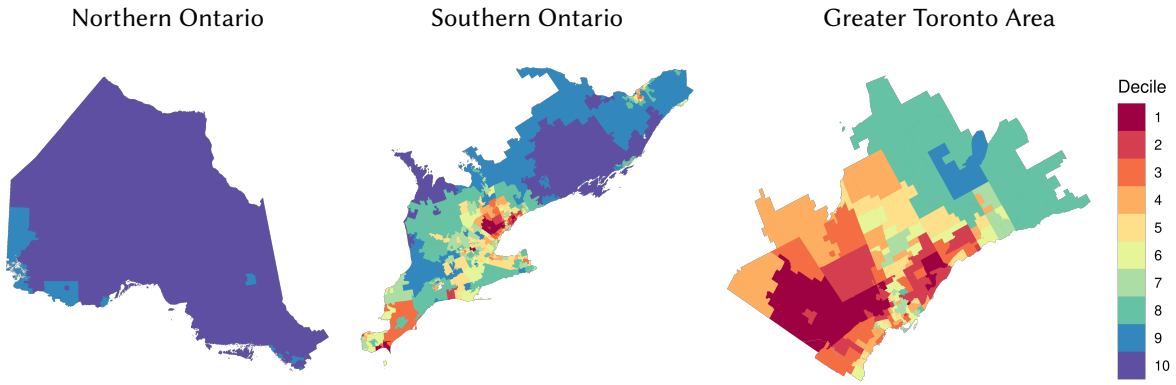


Figure 2: Ontario forward sorting areas (FSAs, $N = 513$), stratified by decile rank in cumulative COVID-19 cases between Dec 2020–May 2021; decile rank was used to group FSAs into 10 patches for transmission modelling.

(37–68 FSAs), but not contiguous geographic regions. Such definitions were used to support prioritization of COVID-19 vaccines to "hot spot" neighbourhoods in Ontario [8]. Figure 2 illustrates the locations of the FSAs and their decile rank, which is synonymous with their patch index. Figure A.4 plots the trends in daily COVID-19 case trends in each patch, and Figure A.5 plots the age distributions of each patch. Age groups were then defined to reflect historical and hypothetical COVID-19 vaccine eligibility in Ontario:

$$a_* = \{\text{0-11, 12-15, 16-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+}\} \tag{16}$$

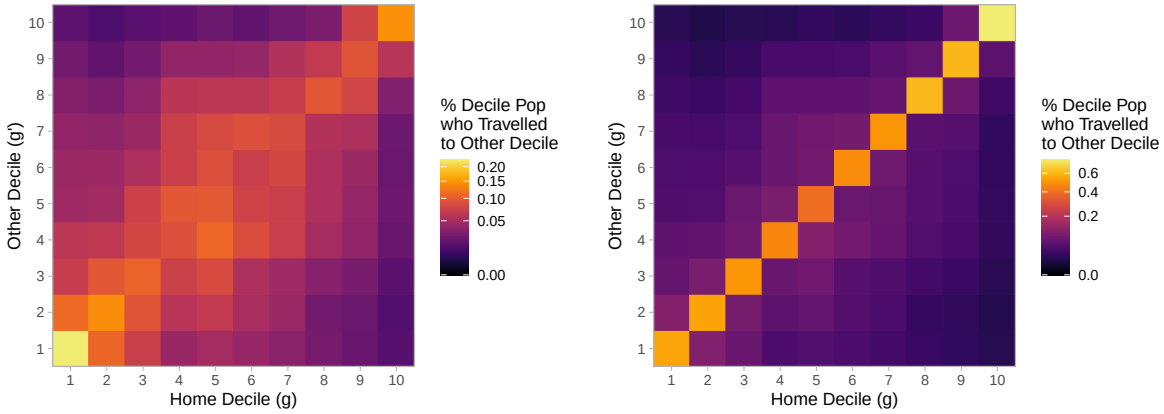
3.1 Data

Ontario population sizes by age and FSA P_{ga} were obtained from the 2016 Canadian Census via Statistics Canada[8] and aggregated from 1-year age groups (a_1) into 5-year (a_5) and target (a_*) age groups as needed. We obtained the final output contact matrices $C_{aa'y}$ from [6], for each of the "home", "work", "school", and "others" contact types, as well as the population size of each 5-year age group used in [6].[9] We assumed that residence patch did not influence the numbers of contacts formed per person, although such a belief could easily be incorporated in the model, perhaps in Eq. (13).

The mobility matrix $B_{gg'}$ between patches was derived using private data on geolocation service usage among a sample of approximately 2% of mobile devices in Ontario; Appendix A.3 details the specific methods and assumptions used. To summarize: Each device was assigned a home FSA (n) based on the most common location during overnight hours for each calendar month. Visits to other FSAs (n') by each device were used to estimate the proportion of devices in each FSA that are expected to travel to each other FSA each day, averaged monthly. This proportion was adjusted for potential changes to the denominator associated with

[8] https://www150.statcan.gc.ca/n1/en/catalogue/98-400-X2016008

[9] The rows for Canada in `contacts_home.rdata`, `contacts_work.rdata`, `contacts_school.rdata`, and `contacts_others.rdata` from `/generate_synthetic_matrices/output/syntheticcontactmatrices2020/` and `poptotal.rdata` from `/generate_synthetic_matrices/input/pop/` within https://github.com/kieshaprem/synthetic-contact-matrices/tree/67c824b.



(a) Directly observed mobility: no data for mobility within the residence FSA

(b) Complete modelled mobility: including assumed mobility within the residence FSA

Figure 3: Mobility matrix $B_{gg'}$, representing the proportion of individuals in decile (patch) g who are expected to travel to decile g' each day

Time period: Jan–Feb 2020 (average). Colour scales are square-root transformed to improve perception of smaller values.

intention to travel. We also assumed that individuals not in the dataset (approximately 98%) had 0.9 the odds of being mobile versus the individuals in the dataset, and that a proportion of individuals who were not observed to travel to any other FSA were still mobile within their residence FSA. Then, the contribution of each FSA to overall mobility of the patch/decile (group of FSAs) was aggregated as follows:

$$B_{gg'} = \sum_{n \in S_g} \sum_{n' \in S'_g} B_{nn'} \tag{17}$$

where S_g is the set of FSA n corresponding to patch/decile g. Mobility matrices were estimated for each month in the available dataset, spanning Jan 2020–Jan 2021. A reference period reflecting pre-pandemic conditions was defined as Jan–Feb 2020; unless otherwise specified, all subsequent results use the average mobility patterns during that period (Figure 3). We did not model any differences in mobility by age group, although such differences could be included in the model by adding a relative rate in Eq. (11).

Finally, we specified the proportions of each contact type assumed to be formed with the home pool:

$$h_y = \{\text{home} : 1, \quad \text{work} : 0, \quad \text{school} : 0, \quad \text{others} : 0\} \tag{18}$$

The parameters P_{ga}, $C_{aa'y}$, $B_{gg'}$, and h_y represent the necessary inputs to our approach for calculating $C_{gag'a'y}$. The following § 3.2 walks step-wise through the approach and presents all major intermediate results.

3.2 Results

Figure 4 illustrates the contact matrices $C_{aa'y}$ from Prem, Cook, and Jit [6], before and after the steps of unweighting by population age distributions, Eq. (8), and ensuring contact balancing, Eq. (9). Figure 5 illustrates the differences in contact matrices between each step. These differences can be explained as follows. The Canadian age distribution used by Prem, Cook, and Jit [6] (Figure A.5, black dashed line), is below the mean for the youngest and oldest age groups; thus inverting the weighting by this age distribution increases the contacts expected with these age groups (Figure 5a). By contrast, Figure 5b is purely symmetric (and opposite about the central diagonal), reflecting differences from the symmetric mean matrix.

Figure 6 illustrates the unweighted and balanced contact matrices $C^{ub}_{aa'y}$ before and after bilinear interpolation and aggregation to the target age groups of interest, Eq. (10). The final matrices $C^{ub}_{a_* a'_* y}$ include dominant horizontal streaks corresponding to larger age groups. These streaks are expected, as more contacts are expected to form with larger "other" age groups. Vertical streaks do not appear, as each column represents the expected contacts for each person in the "self" age group, not the total contacts formed by that age group.[10]

Figure 7 plots the total expected contacts per person per day, $C_{ay} = \sum_{a'} C_{aa'y}$, before and after each step. Overall, patterns remained roughly consistent across transformations, although some details among the large 16–39 age group are lost due to substantial averaging.

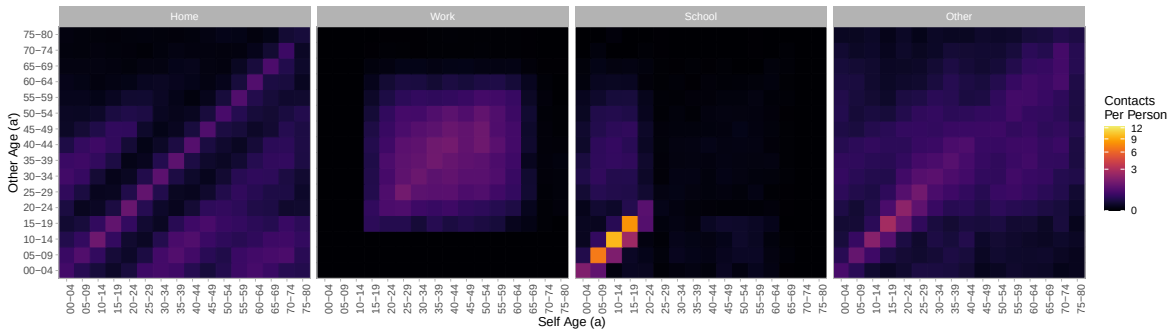
Finally, Figure 8 illustrates the margins (sum over "other" strata and population-weighted average over "self" strata) of the complete mixing matrices $C_{gag'a'y}$, in terms of age groups a & a' (Figure 8a), and patches/deciles g & g' (Figure 8b). Such margins are computed as follows:

$$C_{aa'y} = \sum_{g} P_{ga} \sum_{g'} C_{gag'a'y} \Bigg/ \sum_{g} P_{ga} \tag{19}$$

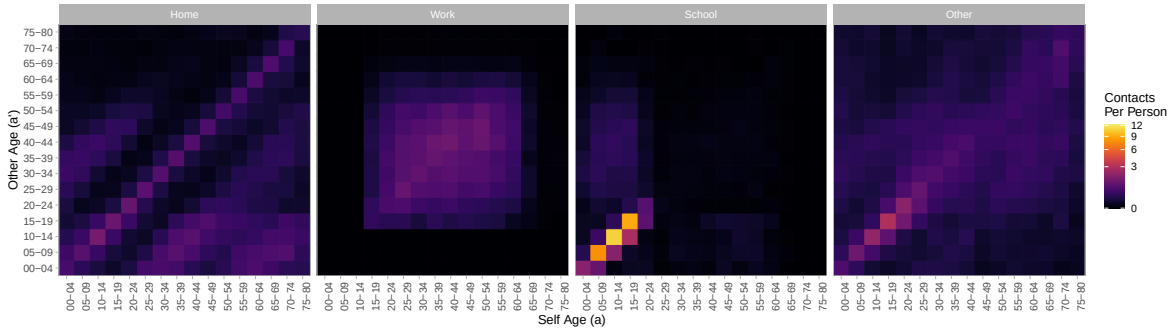
$$C_{gg'y} = \sum_{a} P_{ga} \sum_{a'} C_{gag'a'y} \Bigg/ \sum_{a} P_{ga} \tag{20}$$

The equivalent matrices for total number of contacts per person of all types ($C_{aa'}$ and $C_{gg'}$) are also given in Figures 8c and 8d, respectively. The marginal matrices $C_{aa'y}$ are identical to the input age mixing matrices from Eq. (10), which could be used as an implementation check. Since $h_y = 1$ for "home" contacts, $C_{gg'y}$ is an identity matrix. The equivalent matrices for "work", "school", and "others" contact types also feature a strong diagonal, due to a strong diagonal in the source mobility matrix $B_{gg'}$ (which includes assumptions about the mobility of some individuals within their own FSA). However, the off-diagonal elements are less clustered towards the central diagonal versus the mobility matrix $B_{gg'}$ (Figure 3); mobile individuals from patches g and g' may form contacts not only when either travels to the others' patch, but also when they both travel to a third patch. Thus, the connectivity of the network is greater than the mobility matrix alone would suggest, but still less than random mixing.

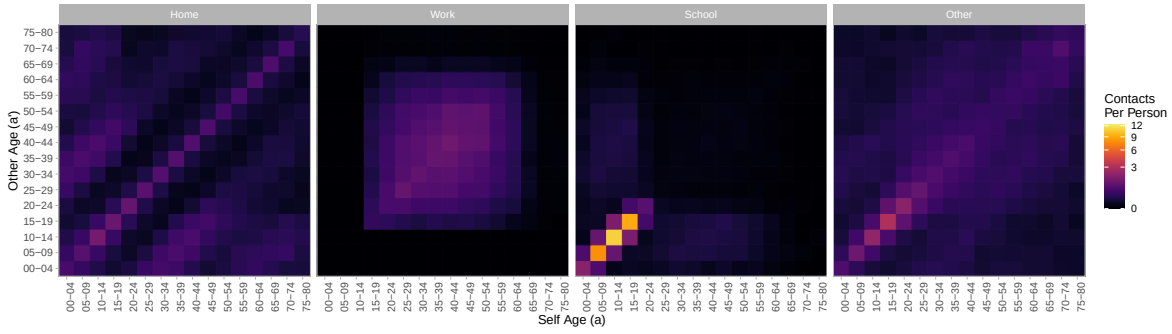
[10] The corresponding matrix of absolute contact $X_{aa'y} = C_{aa'y} P_{a'}$ will be symmetric, and thus will include symmetric horizontal and vertical streaks.



(a) Original contact matrix from [6], $C_{aa'y}$



(b) Unweighted contact matrix, $C^{u}_{aa'y}$; result of Eq. (8)

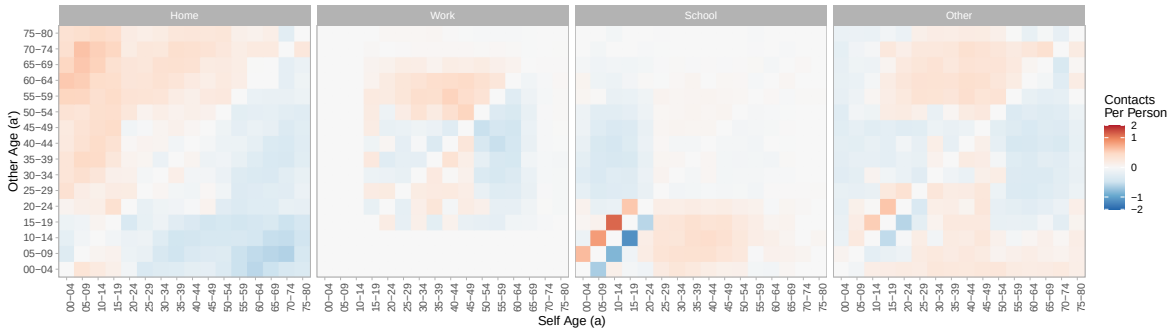


(c) Balanced contact matrix, $C^{ub}_{aa'y}$; result of Eq. (9)

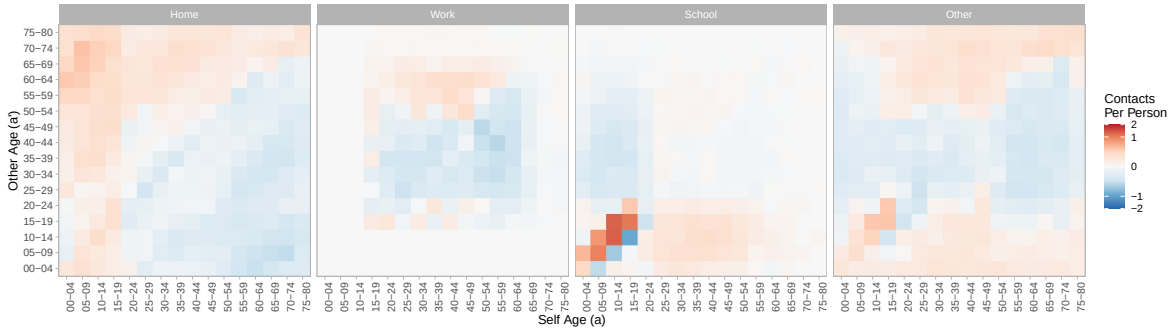
Figure 4: Intermediate results in obtaining unweighted and balanced age contact matrices $C^{ub}_{aa'y}$ (expected number of type y contacts per person per day in each age group a, with those other age groups a') from population-weighted matrices $C_{aa'y}$ from [6] which may not satisfy contact balancing

Colour scales are square-root transformed to improve perception of smaller values.

(a) Difference of Figure 4b and Figure 4a: $C^{u}_{aa'y} - C_{aa'y}$



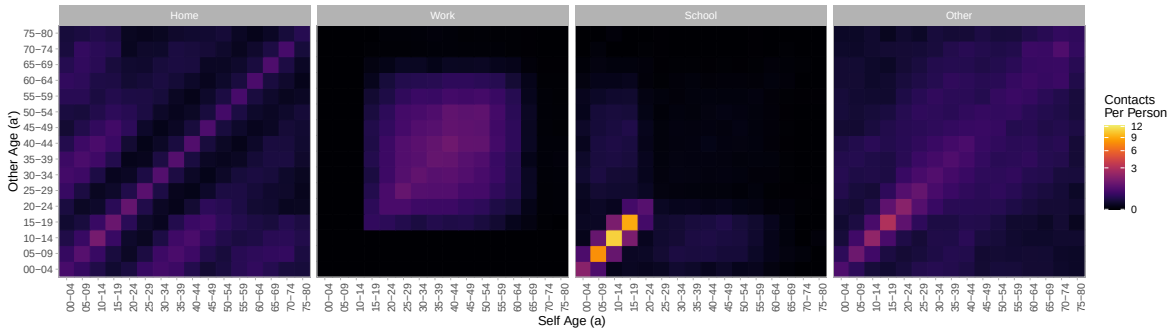
(b) Difference of Figure 4c and Figure 4b: $C^{ub}_{aa'y} - C^{u}_{aa'y}$



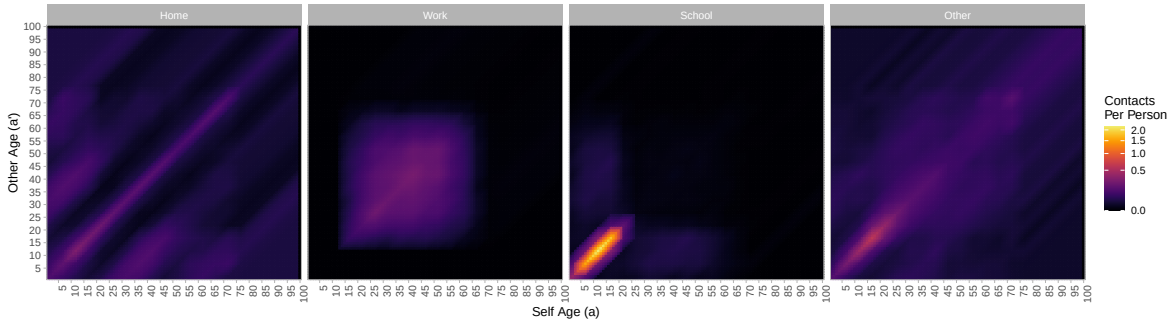
(c) Difference of Figure 4c and Figure 4a: $C^{ub}_{aa'y} - C_{aa'y}$

Figure 5: Differences between intermediate results shown in Figure 4

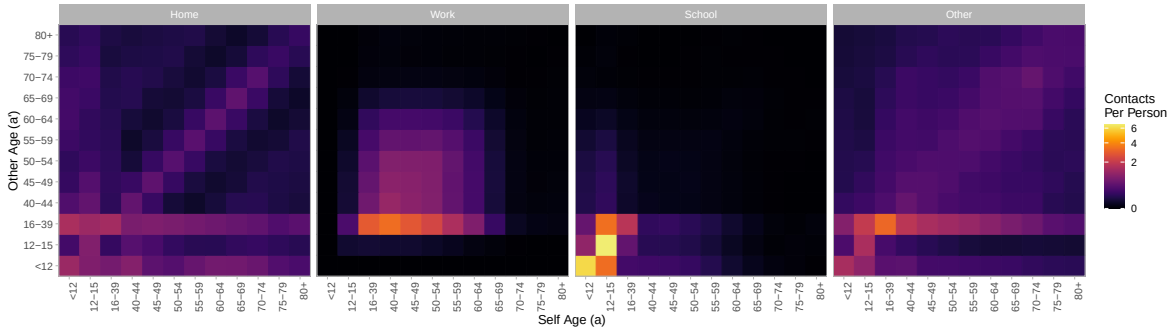
Colour scales are square-root transformed to improve perception of smaller values.



(a) Original age groups $C^{ub}_{a_5 a'_5 y}$; result of Eq. (9); identical to Figure 4c



(b) 1-Year age groups $C^{ub}_{a_1 a'_1 y}$; result of diagonal edge padding and bilinear interpolation



(c) Target age groups $C^{ub}_{a_* a'_* y}$; result of Eq. (10)

Figure 6: Intermediate results in obtaining age-restratified contact matrices $C^{ub}_{a_* a'_* y}$ (expected number of type y contacts per person per day in each age group a_*, with those from age groups a'_*) from matrices $C^{ub}_{a_5 a'_5 y}$ with 5-year age stratifications a_5

Colour scale is square-root transformed to improve perception of smaller values. The vertical streaks in (Figure 6c) corresponding to age groups 0–11 and 16–39 are expected, as more contacts will be formed with larger age groups.

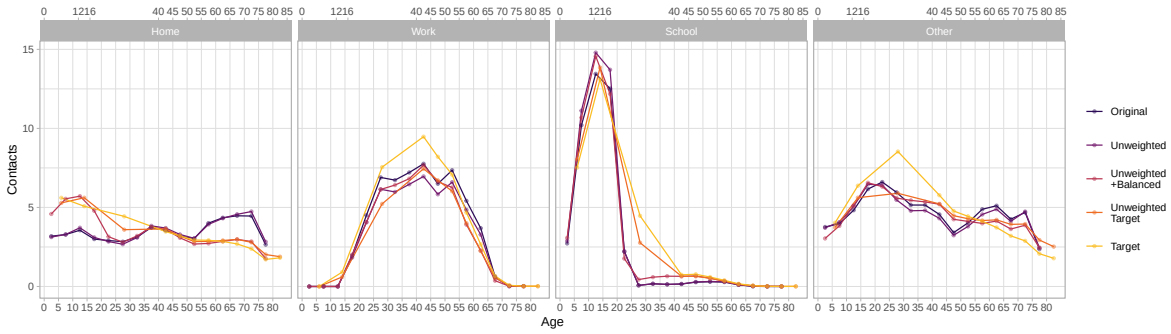


Figure 7: Total contacts per person per day $C_{ay} = \sum_{a'} C_{aa'y}$ for each intermediate step in obtaining $C^{ub}_{a_* a'_* y}$, stratified by contact type.

Modelled contacts for each age group are plotted at the midpoint of the age group. The cut points for the original age groups a_5 from [6] and the target age groups a_* in our application are indicated on the bottom and top x-axis, respectively.

4 Conclusion

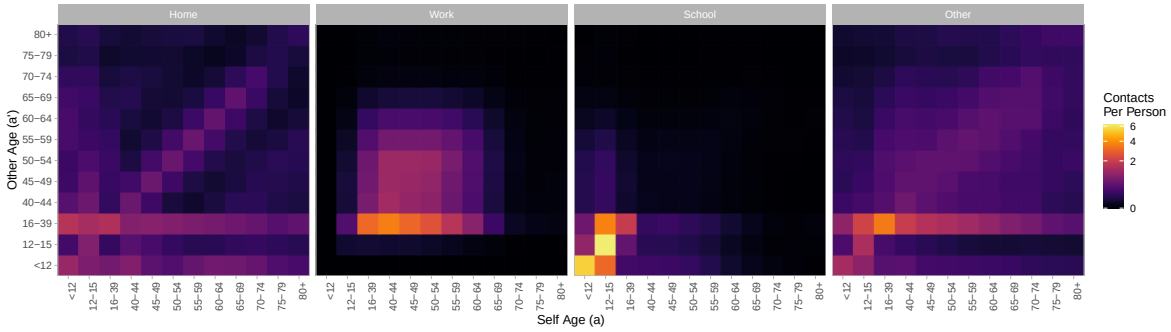
Arenas et al. [1] develop an approach to modelling contact patterns associated with recurrent mobility, which is relevant to dynamic models of infectious disease transmission, such as COVID-19. The original approach simulates contact patterns between age groups and geographic patches connected by recurrently mobile individuals, and considers changes to mixing between patches due to reduced mobility among some individuals. However, the method could be improved to: ensure contact balancing between age groups; model changes to age contact patterns in response to reduced mobility; and allow complete isolation of non-mobile individuals from mobility-related contacts. We propose a modification of the approach that achieves these improvements.

The first key change in the proposed approach draws on [4] to combine preferential patterns of age mixing with the age distribution of the mixing population, such that the actual number of contacts formed reflects both elements. This change is incorporated into each separate mixing "pool" where contacts are formed, and ensures that the number of contacts simulated from age group a to age group a' will equal those from age group a' to age group a.

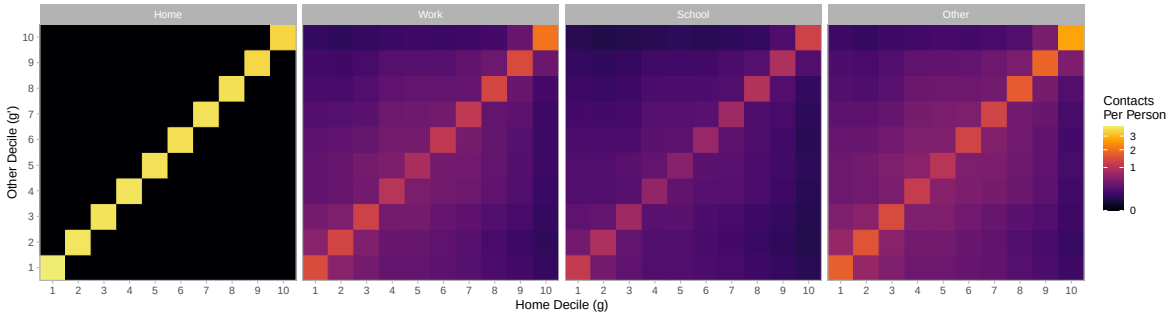
The second key change in the proposed approach is to maintain separate mixing patterns for each type of contact, only aggregating the contribution of different contact types to overall transmission within the force of infection equation. With this change, the age mixing patterns associated with any contact type are not influenced by changes to the numbers or mixing patterns of any other contact type. This change also supports differential probability of transmission by contact type.

The final key change in the proposed approach is to introduce two separate mixing pools where contacts can form: within "home" pools, contacts can only be formed with other residents of the same patch; within "travel" pools, contacts cab be formed with other residents of the same patch who are mobile within their residence patch, or with any mobile visitors to the patch. Home pools therefore allow true isolation of some individuals from mobility-related contacts, with implications for overall network connectivity.

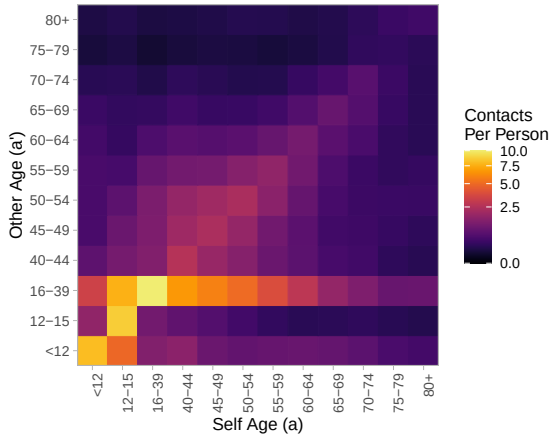
In developing and applying the proposed approach to an example context, we present the methodological details and results of each intermediate step, so that they may be reproduced or built upon in future work.



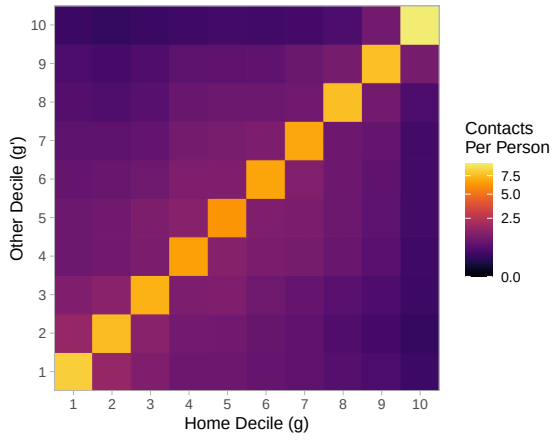
(a) Stratified by age groups and contact type; from Eq. (19)



(b) Stratified by patches/deciles and contact type; from Eq. (20)



(c) Stratified by age groups (all types)



(d) Stratified by patches/deciles (all types)

Figure 8: Expected contacts per person per day, stratified by age, patch/decile, and contact type, computed as the margins of the overall contact matrices $C_{gag'a'y}$

References

[1] A. Arenas, W. Cota, J. Gómez-Gardeñes, et al. "Modeling the Spatiotemporal Epidemic Spreading of COVID-19 and the Impact of Mobility and Social Distancing Interventions". In: *Physical Review X* 10.4 (2020), p. 041055.

[2] D. Balcan and A. Vespignani. "Phase transitions in contagion processes mediated by recurrent mobility patterns". In: *Nature Physics* 7.7 (2011), pp. 581–586.

[3] L. Sattenspiel and K. Dietz. "A structured epidemic model incorporating geographic mobility among regions". In: *Mathematical Biosciences* 128.1-2 (1995), pp. 71–91.

[4] S. Arregui, A. Aleta, J. Sanz, et al. "Projecting social contact matrices to different demographic structures". In: *PLoS Computational Biology* 14.12 (2018), e1006638.

[5] E. H. Kaplan. "Modeling HIV infectivity: Must sex acts be counted?" In: *Journal of Acquired Immune Deficiency Syndromes* 3.1 (1990), pp. 55–61.

[6] K. Prem, A. R. Cook, and M. Jit. "Projecting social contact matrices in 152 countries using contact surveys and demographic data". In: *PLoS Computational Biology* 13.9 (2017), e1005697.

[7] J. Mossong, N. Hens, M. Jit, et al. "Social contacts and mixing patterns relevant to the spread of infectious diseases". In: *PLoS Medicine* 5.3 (2008), pp. 0381–0391.

[8] S. Mishra, N. M. Stall, H. Ma, et al. "A Vaccination Strategy for Ontario COVID-19 Hotspots and Essential Workers". In: *Science Briefs of the Ontario COVID-19 Science Advisory Table* 2.26 (2021).

Funding

The study was supported by: the Natural Sciences and Engineering Research Council of Canada (NSERC CGS-D); Ontario Early Researcher Award No. ER17-13-043; and the 2020 COVID-19 Centred Research Award from the St Michael's Hospital Foundation Research Innovation Council.

Acknowledgements

We thank: Kristy Yiu (Unity Health Toronto) for research coordination support; Gary Moloney for support with geographic data processing.

Contributions

TODO

APPENDIX

Title: Adaptive data-driven age and patch mixing in contact networks with recurrent mobility

Authors: Jesse Knight[1,2,a,*], Huiting Ma[1,b], Amir Ghasemi[3], Mackenzie Hamilton[1,d], Kevin Brown[4,5,e], and Sharmistha Mishra[1,2,5,6,f]

[1]MAP Centre for Urban Health Solutions, Unity Health Toronto
[2]Institute of Medical Science, University of Toronto
[3]Communications Research Centre Canada, Ottawa
[4]Public Health Ontario, Canada
[5]Dalla Lana School of Public Health, University of Toronto
[6]Division of Infectious Diseases, Department of Medicine, University of Toronto
[a]ORCID: 0000-0002-0455-5455
[b]ORCID: 0000-0003-1910-5614
[d]ORCID: 0000-0003-2423-3629
[e]ORCID: 0000-0002-1483-2188
[f]ORCID: 0000-0001-8492-5470
[*]Corresponding Author: Jesse Knight. jesse.knight@mail.utoronto.ca

Resource: https://github.com/mishra-lab/covid-on-model

Date: July 25, 2021

A Supporting Materials

A.1 Notation

Table A.1 summarizes our notation, and any corresponding parameters from [1].

Table A.1: Notation

Symbol in this	Symbol in [1]	Definition
g	i	home patch
g'	j	other patch
g^*	—	visited patch
a	g	self age group
a'	h	other age group
y	—	contact type
n	—	home FSA
n'	—	visited FSA
i	m	infection state
P	N	population size
B	R	mobility matrix
—	M	convenience mobility matrix
λ	Π	force of infection
ρ	p	mobility factor
h	—	home pool contact proportion
C	k	contacts per person
X	—	total absolute contacts
θ	C	contacts age distribution
ϕ	—	odds of mobility if unobserved

A.2 Interpretation of "Non-mobile"

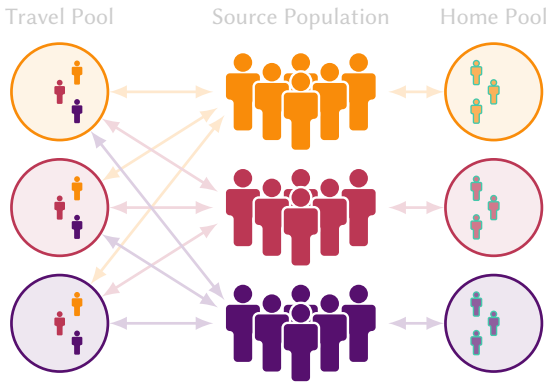
As described in § 2.1, Arenas et al. [1] model the degree of mobility of age group a using a parameter ρ_a (our notation), incorporated into a convenience matrix $M_{gg'}$ per Eq. (4) / (A.1), repeated here for reference:

$$M_{gg'a} = (1 - \rho_a)\,\delta_{gg'} + \rho_a B_{gg'a} \tag{A.1}$$

The implicit assumption of this approach is that non-mobile individuals may form contacts with visitors to their patch, which may not be as intended. The proposed approach to modelling mixing can avoid this assumption if desired. Figure A.1 explores the potential influence of this assumption on network connectivity, as measured by the expected proportion of contacts formed with other patches, in a toy example. The example with 3 patches, each having equal population size and random mobility ($B_{gg'} = 1/N_{g'}$). Age is not considered.

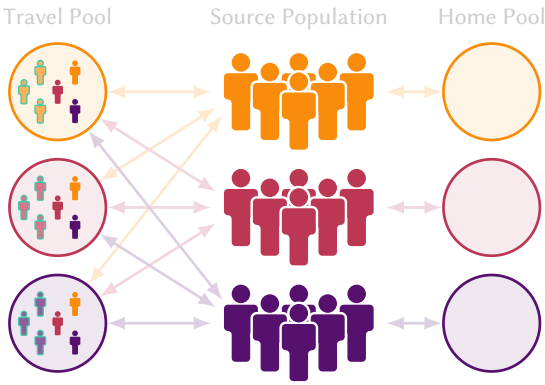
By simulating mixing between non-mobile individuals and mobile visitors (Figures A.1b & A.1d) the proportions of contact made with other patches increases for all patches, as compared to the proposed approach (Fig-

ures A.1a & A.1c). The difference is largest in the context of differential mobility by patch (Figures A.1d vs A.1c).



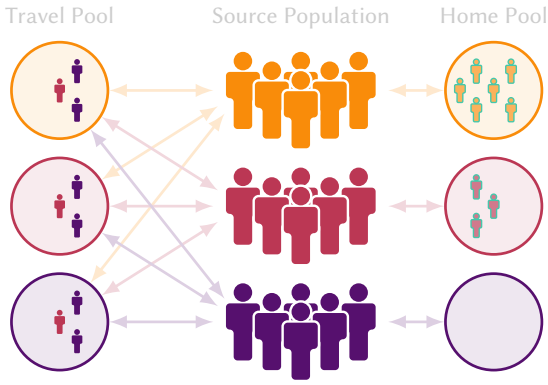
(a) Non-mobile = stay at home, all patches 50% mobile

Contacts formed with other patches:
33% if non-mobile have equal contact rates vs mobile
44% if non-mobile have 1/2 contact rates vs mobile



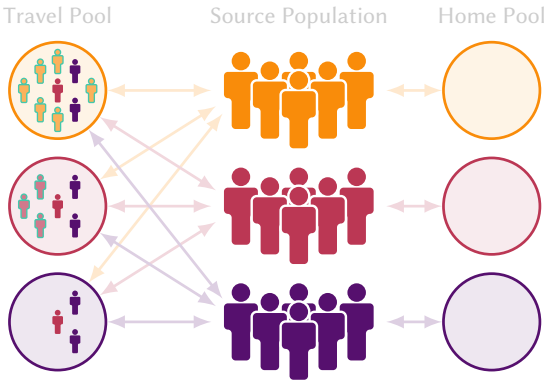
(b) Non-mobile = stay in patch, all patches 50% mobile

Contacts formed with other patches:
50% if non-mobile have equal contact rates vs mobile
59% if non-mobile have 1/2 contact rates vs mobile



(c) Non-mobile = stay at home, patches 0, 50, 100 % mobile

Contacts formed with other patches:
0, 33, 33 % if non-mobile have equal contact rates vs mobile
0, 44, 33 % if non-mobile have 1/2 contact rates vs mobile



(d) Non-mobile = stay in patch, patches 0, 50, 100 % mobile

Contacts formed with other patches:
33, 48, 59 % if non-mobile have equal contact rates vs mobile
50, 58, 52 % if non-mobile have 1/2 contact rates vs mobile

Figure A.1: Mixing implications for two assumptions about the behaviour of non-mobile populations: "stay at home" (a,c): no contacts with other patches (proposed here); vs "stay within patch" (b,d): no travel, but may form contacts with travellers from other patches (as in [1]).

Non-mobile populations are indicated with faded colour and green outline; patch-specific values ordered from top (yellow) to bottom (purple). Additional assumptions: large population size; equal population sizes and contact rates by patch; random mobility (equal probability of visiting any patch if mobile).

A.3 Deriving the Mobility Matrix

Here we describe the methods used to obtain the mobility matrix $B_{gg'}$, representing the expected proportions of individuals residing in decile/patch g, who travel to decile/patch g' each day, based on mobile device data. We will begin by producing the FSA-level mobility matrix $B_{nn'}$.

A.3.1 Data

The raw data represent logs of: timestamp, geolocation (latitude / longitude), and unique device ID. A log entry ("ping") is generated when an app on the device requests the current geolocation from the service provider. Geolocation is logged using Geohash level 6 (1.2 km $\times$ 609.4 m). The midpoint of each Geohash is then compared to boundary files representing all 513 Ontario FSAs[11] to determine which FSA the ping is attributed to. The following definitions were then used to determine device mobility.

The "home FSA" of each unique device is defined as the FSA with the greatest daily-average time between 8:00 pm–5:00 am during each calendar month. Devices for which it was not possible to determine the home FSA were excluded from all further analysis.

The "visited FSAs" represent FSAs a device travelled to within a 24-hour period, as defined by at least 2 consecutive pings spanning at least 2 hours within the FSA. By this definition, it was possible for a device to visit multiple FSA, or no FSA during a given day. Repeated visits to the same FSA by the same device on the same calendar date are treated as a single visit.

These definitions were applied to each calendar month t from Jan 2020–Jan 2021 (13 months) to compute: the mean number of devices with home FSA n per day, denoted H_{nt} or "observed devices"; and the mean number of devices with home FSA n that visited FSA n' per day, denoted $V_{nn't}$, or "device visits".

A.3.2 Bias Correction

Elements of the mobility matrix $B_{gg't}$ could be defined as the proportion of observed devices that travelled to each FSA ($V_{nn'}/H_{nt}$). However, the following analysis of the data suggested that such an approach may bias estimates of mobility. A reference period t_0 was defined as Jan–Feb 2020, to reflect pre-pandemic conditions. The expected values of H_{nt} and $V_{nt} = \sum_{n'} V_{nn't}$ for each FSA during this period were then computed and compared to the expected values during each subsequent month. Figure A.2 plots the distribution of ratios H_{nt}/H_{nt_0} (a) and V_{nt}/V_{nt_0} ((b) across all 513 FSAs, for each month. These ratios illustrate that both V_{nt} and H_{nt} were influenced by pandemic restrictions, and thus $V_{nn't}/H_{nt}$ would overestimate mobility. The trend in H_{nt} might be because apps accessing geolocation services are only opened after the user intends to travel, such as map-related apps.

To address this bias, we modelled mobility with the denominator as the observed devices during the reference period: $V_{nn't}/H_{nt_0}$ (Figure A.2c).[12] However, the results of Figure A.2 suggest that unobserved devices are less mobile than observed devices; thus the unobserved devices would not have the same expected mobility as

[11] https://www150.statcan.gc.ca/n1/en/catalogue/92-179-X

[12] One limitation of this approach is that the reference period only includes winter months, which may not be representative of mobility year-round. Indeed, all three ratios in Figure A.2 have substantial tails above 1 during Aug–Sept 2020, suggesting that mobility was higher during that period versus the pre-pandemic reference period.

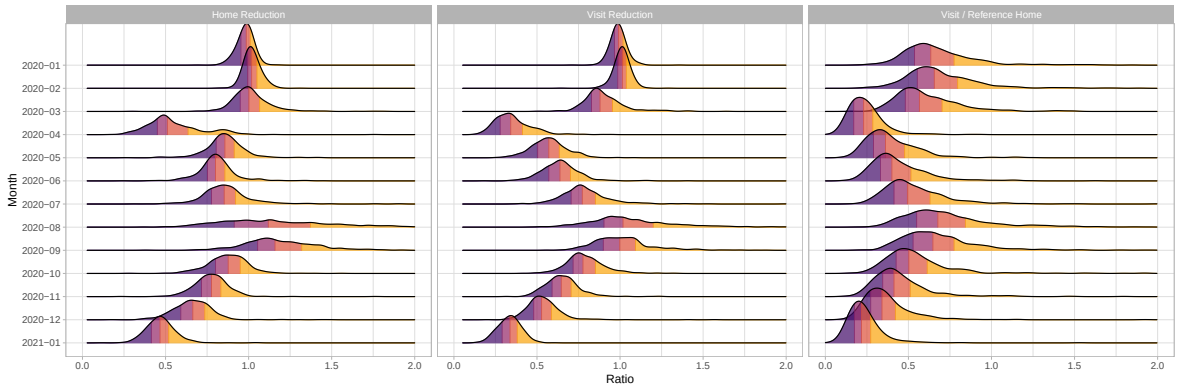


Figure A.2: Ratio of observed devices at home (a) and device visits (b) during each month versus during the reference period (Jan–Feb 2020), suggesting that the proportion of observed devices travelling each month alone is not reflective of true mobility; (c) adjusted mobility measure: ratio of device visits each month versus the number of observed devices at home during the reference period

Distributions show the density of ratio values across all 513 FSAs; coloured segments indicate the 4 quantiles of each distributions.

the observed devices. To account for this observation, we first estimated the total number of devices per FSA, denoted D_n, using Statistics Canada data on smart-phone ownership by age group.[13] Based on these data, we estimate that H_{nt} represents 2.2 [IQR: 1.9, 2.6] % of devices pre-pandemic, and 1.8 [IQR: 1.4, 2.3] % after from March 2020–Jan 2021. The number of unobserved devices in FSA n before the pandemic is therefore $D_n - H_{nt_o}$.

During the pandemic, additional devices go unobserved (H_{nt} decreases) but we assume that the newly unobserved devices do not travel. Therefore, the number of devices that may travel unobserved is a proportion of $D_n - H_{nt_o}$, not a proportion of $D_n - H_{nt}$. Since we have no data on these individuals, we assume that they are mobile at a rate $\phi_1 < 1$ (default 0.9) relative to the observed devices. Similarly, for individuals who do not own devices ($P_n - D_n$) we assume that they are mobile at a rate $\phi_2 < 1$ (default 0.9) relative to the observed devices. Thus, the total proportion of individuals in FSA n who travel to FSA n' during month t is modelled as:

$$B_{nn't} = \frac{V_{nn't}}{H_{no}} \Big[\underbrace{1}_{\text{(a)}} + \underbrace{\phi_1 \left(D_n - H_{no}\right)}_{\text{(b)}} + \underbrace{\phi_2 \left(P_n - D_n\right)}_{\text{(c)}} \Big] \tag{A.2}$$

where (a) represents the observed devices, (b) represents unobserved devices, and (c) represents individuals without devices.

The current data do not provide information on mobility within the home FSA, although the aggregated $B_{gg't}$ includes values along the diagonal due to travel to other FSA within the same decile. We assume that, during the pre-pandemic period, all individuals who do not travel to other FSA are still mobile within their home FSA. During the pandemic months, we assume that individuals who are mobile within their home FSA have reduced mobility proportional to the reduced mobility of individuals who are mobile outside their FSA, as measured by $B_{nt} = \sum_{n'} B_{nn't}$. Thus $B^h_{nn't}$ that considers mobility within the home decile is defined as:

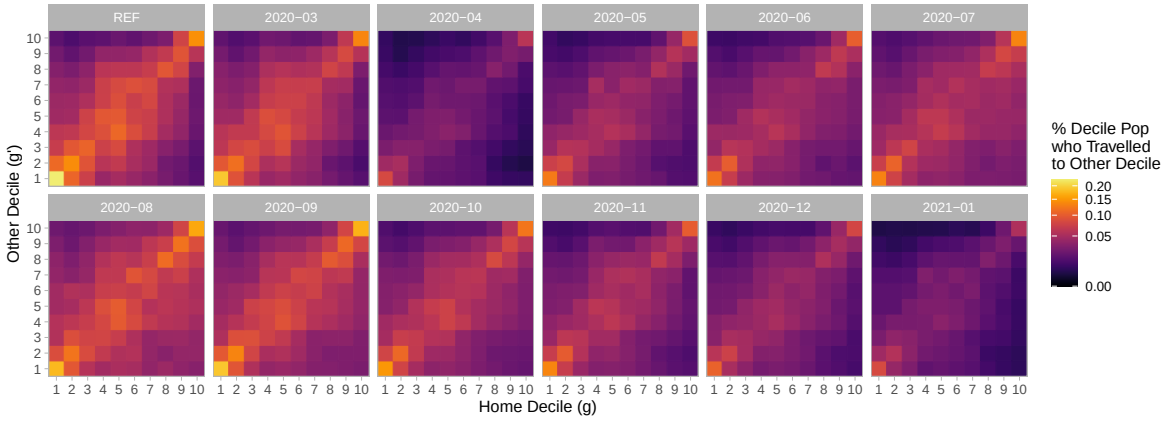
$$B^h_{gg't} = B_{gg't} + \frac{B_{gt}}{B_{gt_o}} \operatorname{diag}\left(1 - B_{gt_o}\right) \tag{A.3}$$

[13] https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2210011501

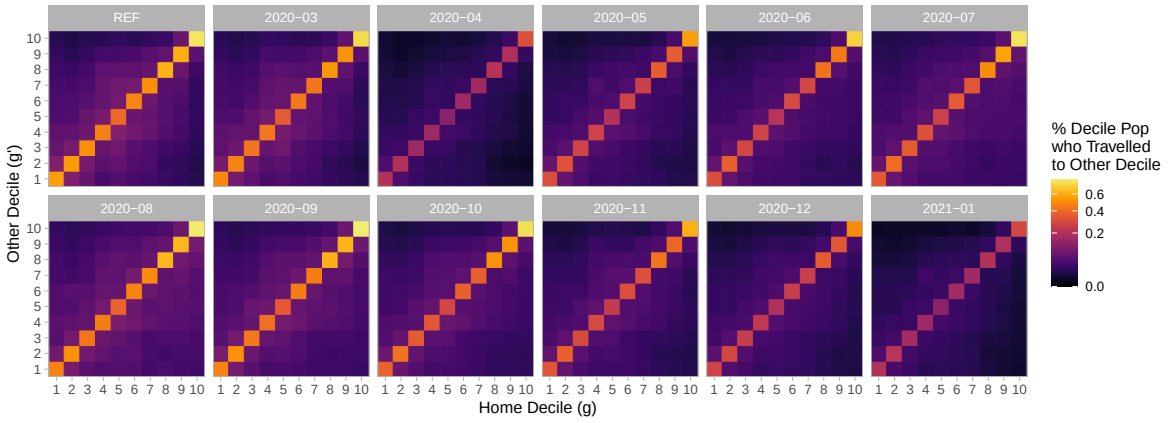
with the constraint that $B_{gt}/B_{gt_0} \leq 1$. Finally, as noted in § 3.1, Eq. (17) / (A.4) can be used to aggregate $B_{nn'}$ to obtain $B_{gg'}$ (repeated here for reference):

$$B_{gg'} = \sum_{n \in S_g} \sum_{n' \in S'_g} B^h_{nn'} \tag{A.4}$$

where S_g is the set of FSA n corresponding to patch g. Figure A.3 illustrates the matrices $B_{gg't}$ for the reference period, and each subsequent month, with and without the assumed mobility within the home FSA / decile. Mobility patterns are overall consistent across months, with the overall proportions of travelling individuals proportional to the ratio depicted in Figure A.2c. Due to Eq. (A.3), $B_{gg't}$ during the reference period will have exactly $\sum_{g'} B_{gg't_0} = 1$. However, for pandemic months, $\sum_{g'} B_{gg't} \leq 1$. Thus, when incorporating this mobility matrix into Eq. (11), the total number of contacts formed by population P_{ga} will be implicitly reduced by the proportional reduction in mobility.



(a) Directly observed mobility: no data for mobility within the residence decile



(b) Complete modelled mobility: including assumed mobility within the residence decile

Figure A.3: Mobility matrices $B_{gg't}$, representing the expected proportions of individuals in patch/decile g who travelled to patch/decile g' each day, stratified by calendar month plus the reference period (Jan–Feb 2020)

Colour scale is square-root transformed to improve perception of smaller values.

A.4 Additional Data for Ontario Patches

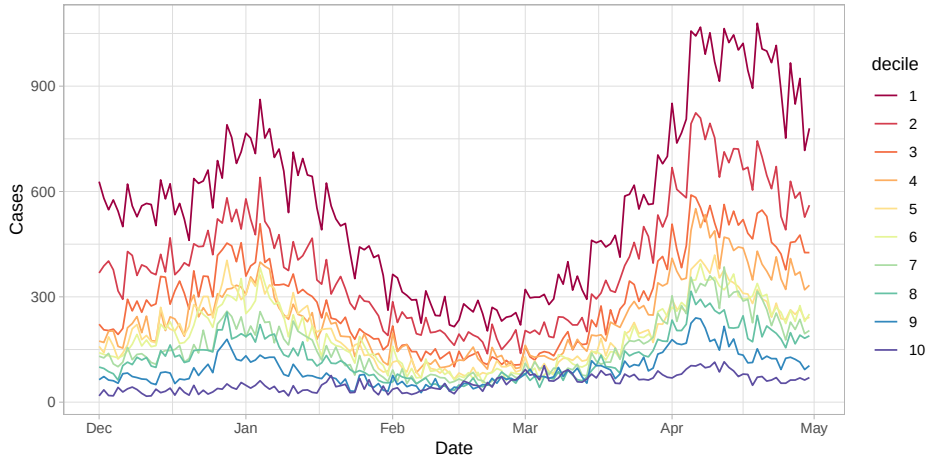


Figure A.4: Time trends in daily COVID-19 cases across the 10 modelled patches in Ontario

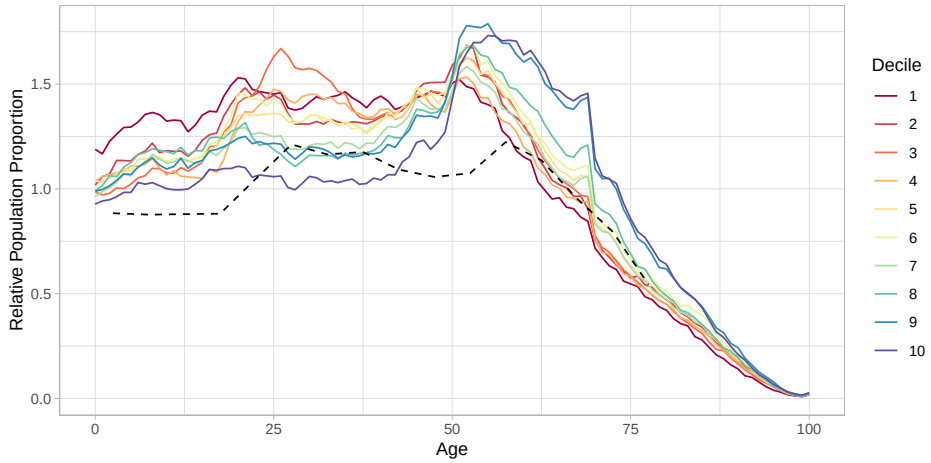


Figure A.5: Population age distributions across the 10 modelled patches in Ontario

Solid coloured lines correspond to the 10 patches defined by deciles of cumulative COVID-19 cases between Dec 2020–May 2021 in Ontario. Dashed black line corresponds to the Canadian age distribution used by Prem, Cook, and Jit [2].

References

[1] A. Arenas, W. Cota, J. Gómez-Gardeñes, et al. "Modeling the Spatiotemporal Epidemic Spreading of COVID-19 and the Impact of Mobility and Social Distancing Interventions". In: *Physical Review X* 10.4 (2020), p. 041055.

[2] K. Prem, A. R. Cook, and M. Jit. "Projecting social contact matrices in 152 countries using contact surveys and demographic data". In: *PLoS Computational Biology* 13.9 (2017), e1005697.