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: Aug 6, 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 simulate contact patterns associated with recurrent mobility between patches, such as due to work, school, and other regular travel. We build 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:
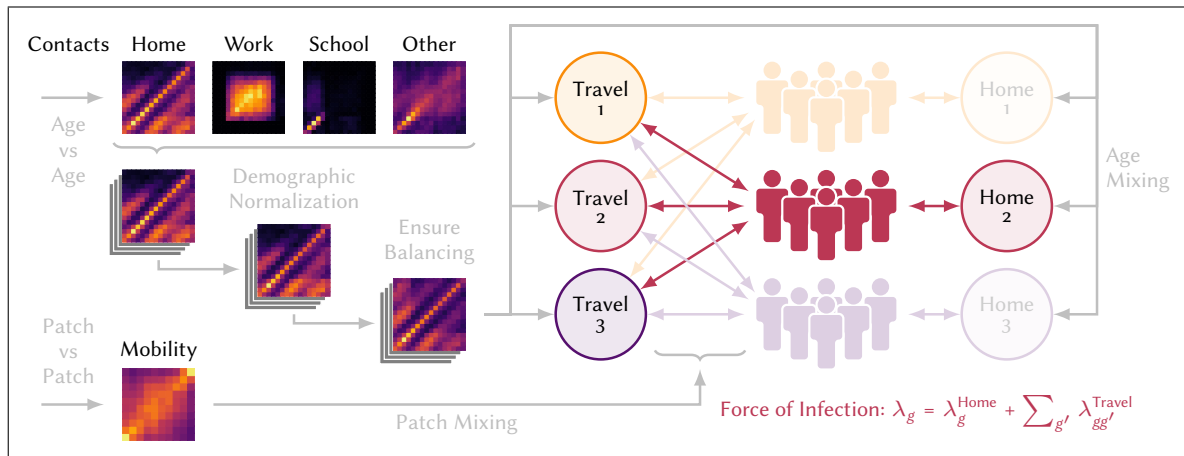
1   Introduction

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. Following foundational work by Balcan and Vespignani [2] and Sattenspiel and Dietz [3], the model incorporates data on short, recurrent mobility patterns to determine contact rates between individuals in different patches and age groups. 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:

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

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 allowing age mixing patterns to respond to local demographic and intervention conditions. 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]:

$$P_a C_a \theta_{aa'} = P_{a'} C_{a'} \theta_{a'a} \qquad (5)$$

   For a model with non-random age mixing and 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, and so the fixed age mixing matrix $\theta_{aa'}$ is applied to all contact types. Then, when mobility restrictions are simulated, and overall average contact rates are reduced to reflect reductions in external versus home contacts, $\theta_{aa'}$ 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, it may be useful to model some or all non-mobile individuals only forming contacts with other individuals from their residence patch, such as reflecting household contacts. As illustrated in Figure A.1, the original approach potentially overestimates the contact network connectivity during reduced mobility versus an approach in which non-mobile populations do not contact visitors. Thus [1] could underestimate the impact of confinement strategies on inter-patch transmission. While it may be desirable to sometimes allow mixing of "non-mobile" individuals with visitors to their residence patch, the original approach does not provide a 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 $C_{gag'a'y}$ provides a representation that is easy to interpret, and allows us to 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 $C_{aa'y}^{u}$ (Figure 4b):[4]

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

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

The next challenge is that $C_{aa'y}^{u}$ 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

---

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

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

In conceptualizing mobility-related mixing, we define two types of contexts in which contacts can be formed:

- **Home pools:** where contacts are formed exclusively with other residents of the same patch (e.g. for household contacts)

- **Travel pools:** where contacts are formed with individuals from any patch who are present in the pool (e.g. for work contacts)

We model one home pool and one travel pool associated with each patch, as illustrated in Figure 1.

---

5 A perfect survey in a closed population would produce a contact matrix $C^u_{aa'y}$ that is already balanced.
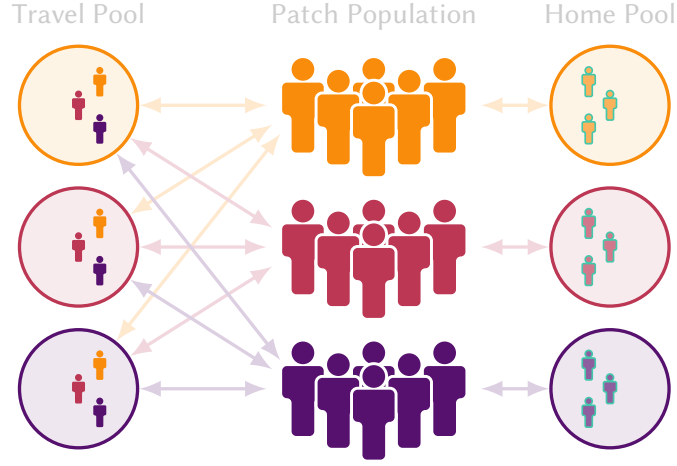
Figure 1: Toy example of "home" vs "travel" mixing pools for a network with 3 patches and 50% individuals mobile. Contacts in the home pool are formed exclusively with other members of the residence patch, whereas contacts in the travel pool may be formed with any visitors to the patch

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

In this conceptualization, only contacts associated with travel pools are influenced by the population mobility matrix $B_{gg'}$, representing the expected proportions of individuals from patch g who visit patch g' per day. For contacts associated with home pools, this matrix is functionally replaced with an identity matrix $\delta_{gg'}$. It is not necessary to assume that all contacts of any particular type are formed in only one type of pool. Rather, we introduce a parameter $h_\gamma \in [0, 1]$ representing the proportion of type γ contacts that are formed in the home pool, and the remainder $(1 - h_\gamma)$ are formed with travel pools. For example, we could have $h_\gamma = 1$ for household contacts, $h_\gamma = 0$ for work contacts, and $h_\gamma = 0.5$ for school contacts. Thus, the expected contacts formed by individuals in patch g are distributed across three situations:

1. **Mobile Away:** individual travelled from patch g to patch g' and formed contacts within travel pool g'
   Proportion of contacts: $(1 - h_\gamma)B_{gg' \, (g \neq g')}$

2. **Mobile at Home:** individual formed contacts within their local travel pool g
   Proportion of contacts: $(1 - h_\gamma)B_{gg' \, (g = g')}$

3. **Non-Mobile at Home:** individual formed contacts within their local home pool g
   Proportion of contacts: $h_\gamma \delta_{gg'}$

The idea of "home pools" is new versus [1], and allows us to address issue 3 by introducing situation 3. Thus in [1], all mixing was implicitly modelled using "travel pools", and individuals described as "non-mobile" reflected situation 2.

In the context of reduced mobility, we do not assume that rows of $B_{gg'}$ sum to 1. The "missing" proportion $1 - \sum_{g'} B_{gg'}$ is then taken to represent non-mobile individuals, who do not form any mobility-related contacts (situations 1 and 2) that day. In § A.3 we discuss some details about generating a mobility matrix $B_{gg'}$ with these properties, based on mobile phone data.

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]

$$P_{gay}^{g^*} = (1 - h_y)B_{gg^*}P_{ga} \tag{11}$$

There is no distinction between situations 1 and 2 in Eq. (11), as both are already reflected in the off-diagonal and diagonal elements of $B_{gg'}$, respectively. 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 $P_{gay}^{g^*}$ and $P_{g'a'y}^{g^*}$ is given by the outer product:

$$X_{gag'a'y}^{g^*\,r} = P_{gay}^{g^*} \otimes P_{g'a'y}^{g^*} \Big/ \sum_{g'a'} P_{g'a'y}^{g^*} \tag{12}$$

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:

$$X_{gag'a'y}^{g^*} = X_{gag'a'y}^{g^*\,r} \, C_{aa'y}^{ub} \Big/ \sum_{a_1'} A_{a'a_1'} \tag{13}$$

since $X_{gag'a'y}^{g^*\,r}$ 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 $X_{gag'a'y}^{g^*\,r}$ and $C_{aa'y}^{ub}$ are proportional to these widths, but the proportionality should only be singular overall. We could have applied this normalization to $C_{aa'y}^{ub}$ in Eq. (10) in the same way as for a, but this would make $C_{aa'y}^{ub}$ harder to interpret, as it would no longer represent the expected numbers of contacts for each age group.

Mixing within home pools (situation 3) can be modelled similar to mixing within travel pools, with one small modification: replacing $(1 - h_y)B_{gg'}$ with $h_y\delta_{gg'}$. Following through Eqs. (11–12), we obtain $X_{gag'a'y}^h$, 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:

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

It may be tempting to simplify the model for home pool contacts by updating the mobility matrix $B_{gg'}$ similar to Eq. (4) from [1], 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. Another implication of our approach is that non-mobile individuals will not form mobility-related contacts. Thus, if $\sum_{g'} B_{gg'}$ is reduced, the total contacts formed by residents of patch g would be reduced proportionately, and changes to mixing patterns by age and patch reflected automatically.

---

[6] If residents of different patches might have relatively different numbers of contacts, a scaling factor could be applied here.
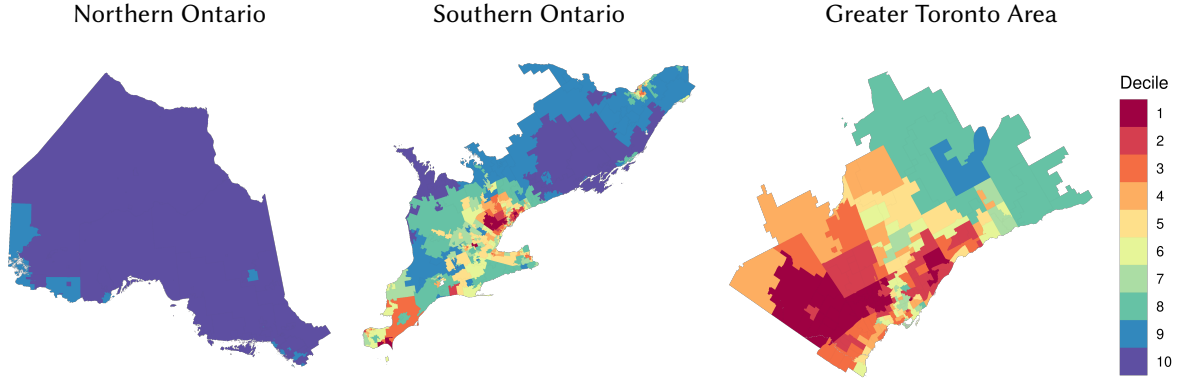
Figure 2: Ontario forward sorting areas (FSAs, $N = 513$), stratified by decile rank in cumulative COVID-19 cases between Dec 2020–May 2021; decile rank was used to group FSAs into 10 patches for transmission modelling.

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 (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.5 plots the daily incidence of COVID-19 cases per patch, and Figure A.6 plots the age distributions of each patch. Age groups were then defined to reflect historical and hypothetical COVID-19 vaccine eligibility in Ontario:

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

---

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

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, only with whom those contacts are formed, although such a belief could 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. Similarly, the average time spent away from home was used to estimate the proportion of devices in each FSA that were mobile within the home FSA. We additionally assumed that individuals not in the dataset (approximately 98%) had 0.9 the odds of being mobile versus the individuals in the dataset, because the number of devices observed each month appeared to vary with mobility restrictions, possibly reflecting an intention to travel preceeding use of geolocation services.

The contribution of each FSA to overall mobility of the patch/decile (group of FSAs) was then aggregated as:

$$B_{gg'} = \sum_{n \in S_g} \sum_{n' \in S_{g'}} B_{nn'} \tag{17}$$

where S_g is the set of FSAs (n) corresponding to patch/decile g. Mobility matrices were estimated for each month in the available dataset, spanning Jan 2020–Dec 2020. 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.

---

[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.
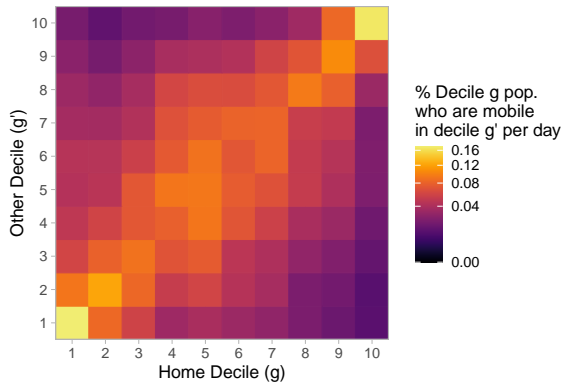
(a) Inter-FSA mobility
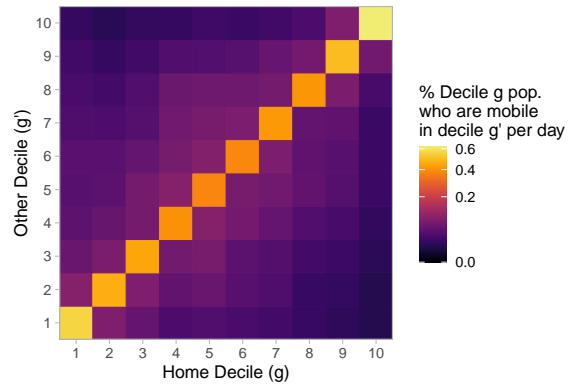
(b) Inter + intra-FSA mobility

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

Derived from mobile device geolocation data; deciles represent groupings of Ontario forward sortation areas (FSAs) by cumulative COVID-19 cases between Dec 2020–May 2021; colour scale is square-root transformed to improve perception of smaller values; reference period: Jan–Feb 2020.

3.2 Results

Figure 4 illustrates the contact matrices $C_{aa'y}$ from Prem, Cook, and Jit [6], before and after the steps of un-weighting 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.6, 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 $C_{aa'y}^{ub}$ before and after bilinear interpolation and aggregation to the target age groups of interest, Eq. (10). The final matrices $C_{a_* a'_* y}^{ub}$ 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 of the above steps, from before Eq. (8) through after Eq. (10). 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:

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

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

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 $B_{gg'}$ (individuals who are mobile within their home 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 degree of mixing between patches using this approach is greater than the mobility matrix alone would suggest, though still less than if completely random mixing was assumed.

---

[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.

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) Original contact matrix from [6], $C_{aa'y}$
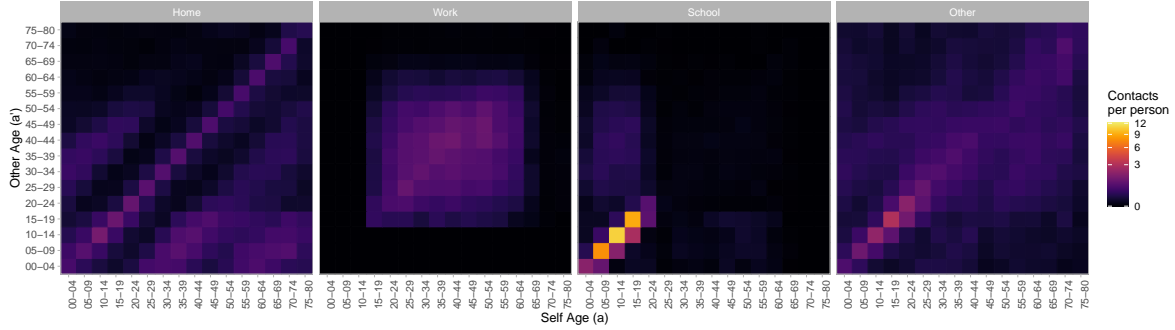
(b) Unweighted contact matrix, $C^u_{aa'y}$; result of Eq. (8)

(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
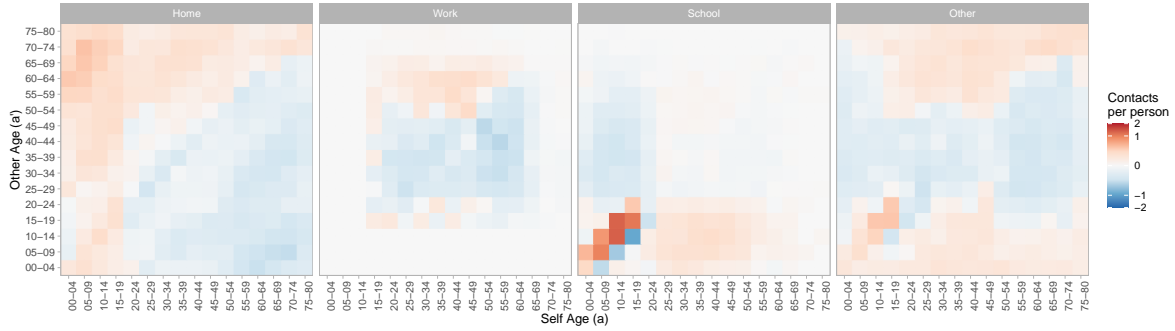
Contact matrices for Canada, derived from [6]; colour scales are square-root transformed to improve perception of smaller values.
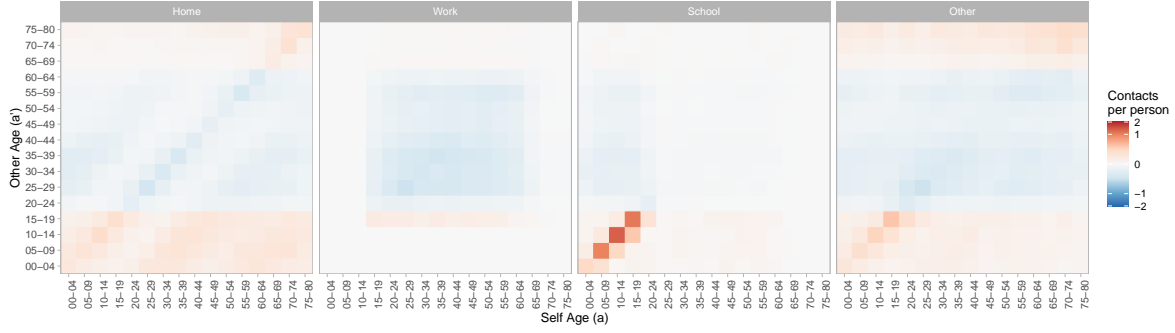
(a) Difference of Figure 4b and Figure 4a: $C^u_{aa'y} - C_{aa'y}$
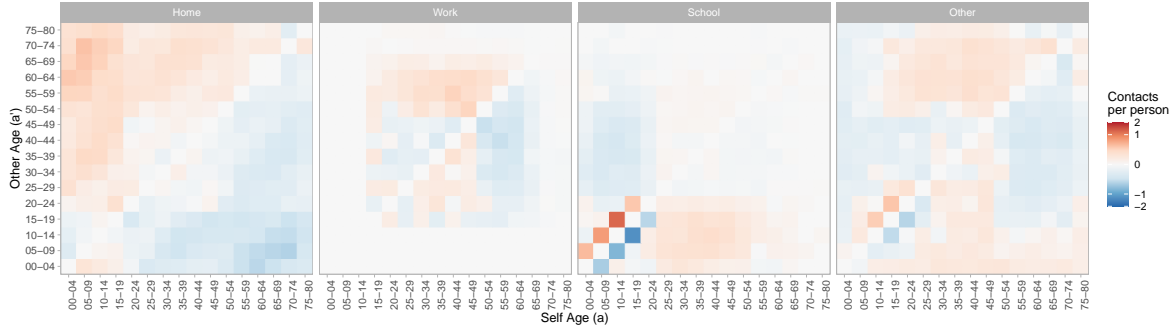
(b) Difference of Figure 4c and Figure 4b: $C^{ub}_{aa'y} - C^u_{aa'y}$

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

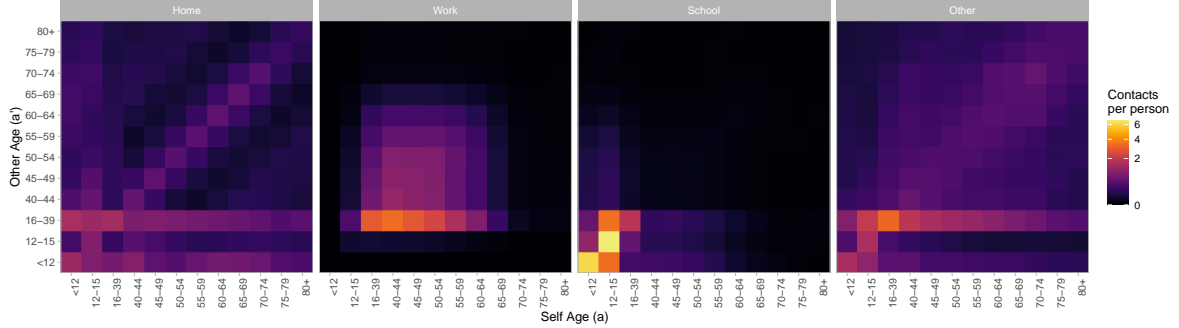Figure 5: Differences between intermediate results shown in Figure 4

Contact matrices for Canada, derived from [6]; 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
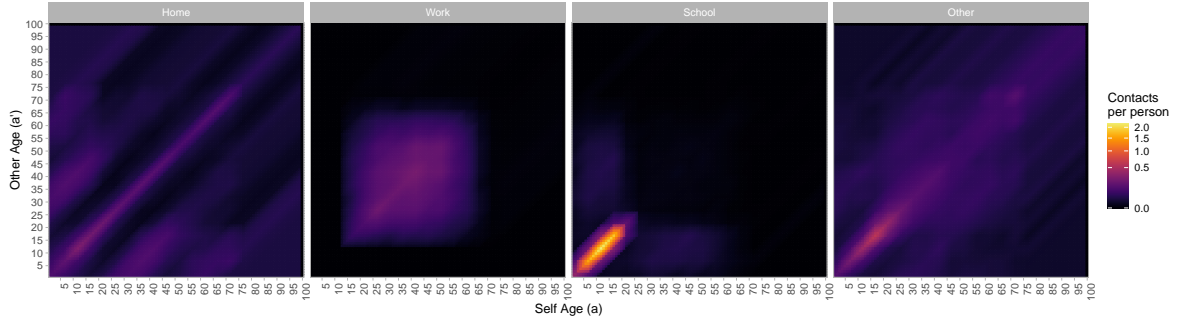
(b) 1-Year age groups $C^{ub}_{a_1 a_1' y}$; result of diagonal edge padding and bilinear interpolation

(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

Contact matrices for Canada, derived from [6]; colour scales are square-root transformed to improve perception of smaller values; the horizontal streaks in (c) corresponding to age groups 0–11 and 16–39 are expected, as more contacts will be formed with larger age groups.
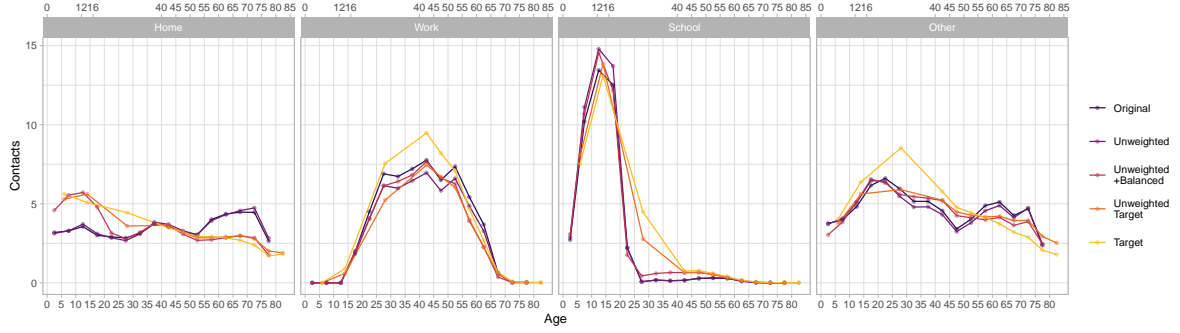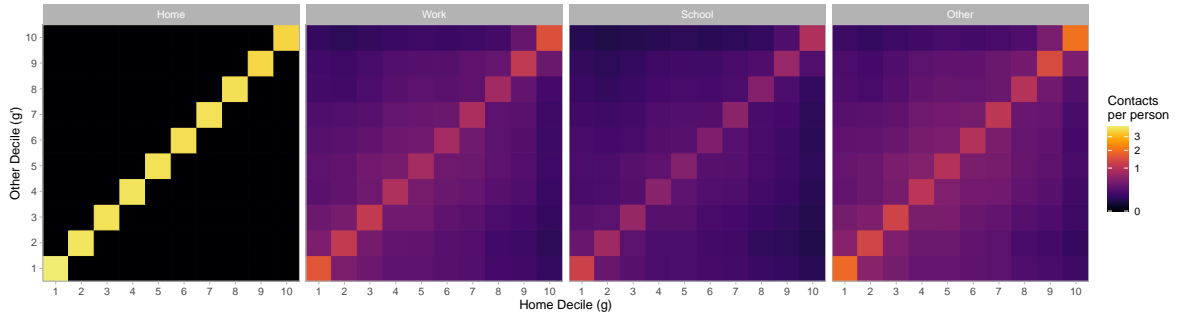
Figure 7: Total contacts per person per day $C_{ay} = \sum_{a'} C_{aa'y}$ for each intermediate step in obtaining $C^{ub}_{a_* a'_* y}$, stratified by contact type.

Contact matrices for Canada, derived from [6]; 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-axes, respectively.

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

(b) Stratified by deciles (patches) and contact type; from Eq. (20)
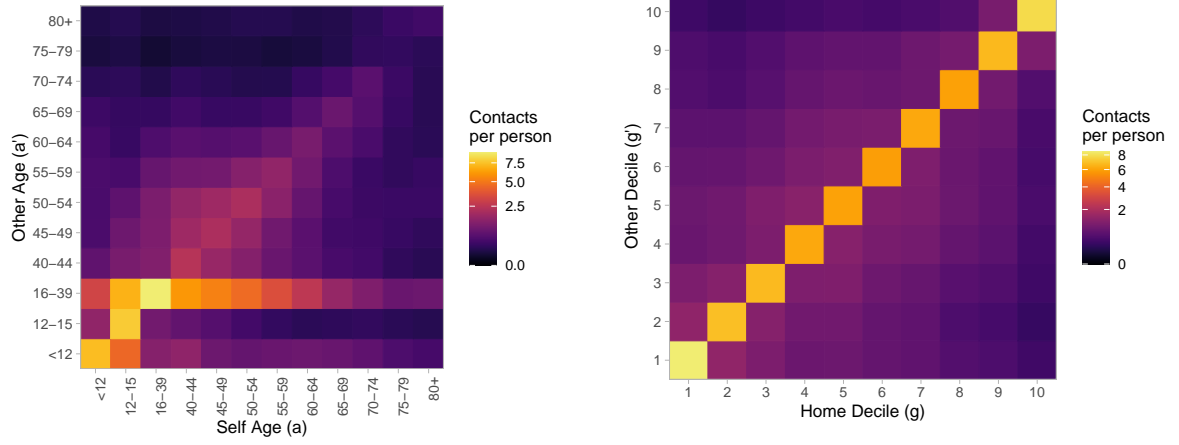
(c) Stratified by age groups, sum across contact types

(d) Stratified by deciles (patches), sum across contact types

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

Contact matrices for Canada, derived from [6]; colour scales are square-root transformed to improve perception of smaller values.

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

Acknowledgements

Contributions

TODO

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: Aug 6, 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
ψ	—	relative time away from home

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 (situation 2 in § 2.2.2), which may not always be desired. The proposed approach to modelling mixing can avoid this assumption if desired through the use of "home pools" (situation 3). 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 has 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, situation 2) the proportions of contact made with other patches increases for all patches, as compared to the proposed approach (Figures A.1a & A.1c, situation 3). The difference is largest in the context of differential mobility by patch (Figures A.1d vs A.1c).
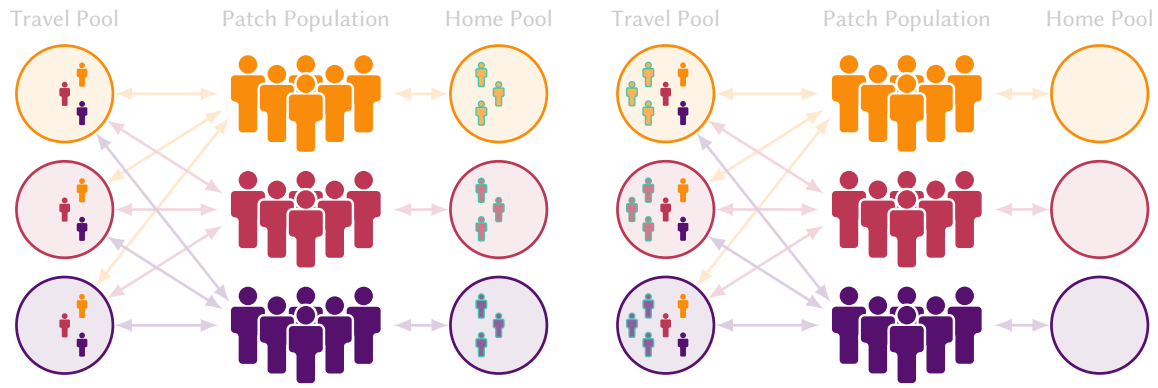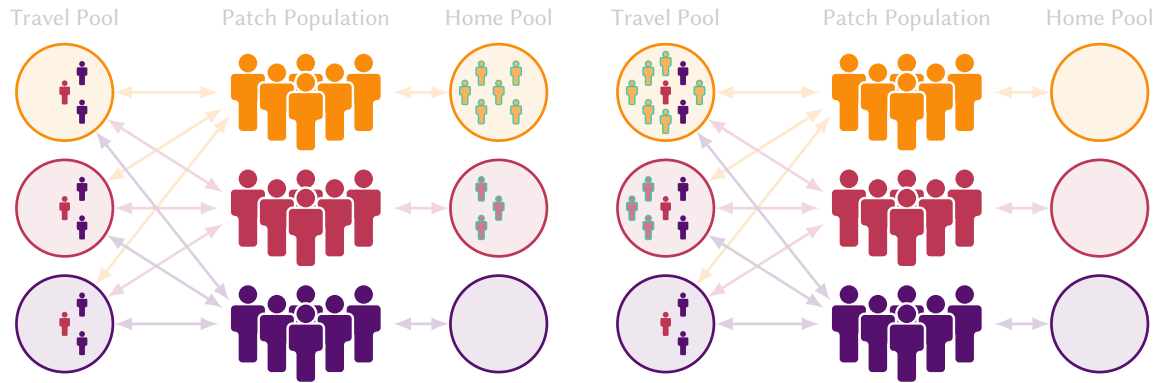

(a) Non-mobile = stay at home, all patches 50% mobile

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, reflecting sitation 3): no contacts with other patches (proposed here); vs "stay within patch" (b,d, reflecting situation 2): 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: Example

Here we describe the methods used to obtain the mobility matrix $B_{gg'}$ for our applied example in Ontario Canada (details in § 3). The mobility matrix represents the expected proportions of individuals residing in decile (patch) g who travel to decile g' each day. We will begin by producing the FSA-level mobility matrix $B_{nn'}$ (513×513), and later aggregate this matrix to obtain $B_{gg'}$ (10×10).

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.

- **Home FSA:** for each unique device, 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.

- **Visited FSA:** FSAs that 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 FSAs, 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.

- **Time away from home:** the proportion of time each device was estimated to be away from the home dwelling (not just home FSA). The home dwelling was defined by the Geohash 6 tile with the greatest daily-average time between 8:00 pm and 5:00 am each month.

These definitions were applied to each calendar month t from Jan 2020–Dec 2020 (12 months) to compute: the mean number of devices with home FSA n per day, denoted H_{nt} or "observed devices"; the mean number of devices with home FSA n that visited FSA n' per day, denoted $V_{nn't}$, or "device visits"; and the mean proportion of time each device spent outside the home per day for each FSA, denoted W_{nt}.

A.3.2   Inter-FSA Mobility

The inter-FSA mobility matrix $B_{nn't}$ ($n \neq n'$) could be defined as the proportion of observed devices that travelled to each FSA ($V_{nn't}/H_{nt}$). 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 (reduced) by pandemic restrictions, and thus $V_{nn't}/H_{nt}$ would overestimate mobility. The

---

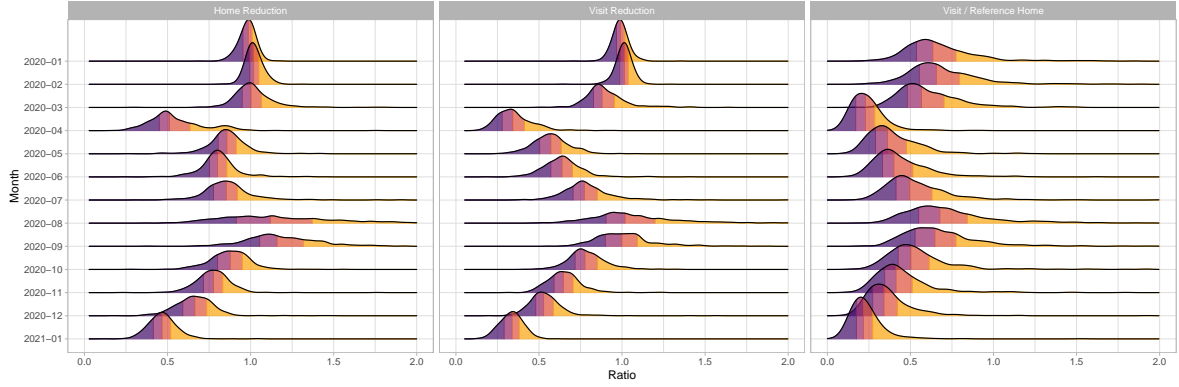[11] https://www150.statcan.gc.ca/n1/en/catalogue/92-179-X

Figure A.2: Ratios of (a) observed devices at home and (b) total device visits during each month versus during the reference period suggesting that the proportion of observed devices travelling each month alone is not reflective of true mobility or else (a) would be 1; (c) adjusted mobility measure: ratio of device visits each month from versus the number of observed devices at home during the reference period

Data from approximately 2% of mobile devices in Ontario, Canada; distributions show the density of ratio values across all 513 FSAs; coloured segments indicate the 4 quantiles of each distributions; reference period: Jan–Feb 2020.

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 chose H_{nt_0} as a more reliable denominator, and thus modelled mobility using: $V_{nn't}/H_{nt_0}$ (Figure A.2c).[12] However, the results of Figure A.2 also suggest that unobserved devices are less mobile than observed devices; thus the unobserved devices would not have the same expected mobility as 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 $D_n - H_{nt_0}$.

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 $D_n - H_{nt_0}$, 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 $\phi_1 \leq 1$ (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 $\phi_2 \leq 1$ (default 0.9) relative to the observed devices. We assume $\phi_1 \leq 1$ and $\phi_2 \leq 1$ because, as noted above, the trends in Figure A.2a suggest that individuals are more likely to be observed via geolocation pings after forming an intention to travel. Thus, the total proportion of individuals in FSA n who travel to FSA n' during month t is modelled as:

$$B_{nn't\,(n \neq n')} = \frac{V_{nn't}}{H_{nt_0}} \Big[\underbrace{1}_{(a)} + \underbrace{\phi_1 \left(D_n - H_{nt_0} \right)}_{(b)} + \underbrace{\phi_2 \left(P_n - D_n \right)}_{(c)} \Big] P_n^{-1} \tag{A.2}$$

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

---

[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.

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

A.3.3 Intra-FSA Mobility

Since device visits to other FSA $V_{nn't}$ do not include "visits" within the home FSA, the proportion of individuals who are mobile (making contacts) within their home FSA is not yet determined. During the reference period, we assume that all individuals who do not travel to other FSAs are still mobile within their home FSA:

$$B_{nn't_0\,(n=n')} = 1 - \sum_{n'} B_{nn't_0\,(n\neq n')} \tag{A.3}$$

During pandemic months, we assume that a proportion of such individuals remain mobile within their home FSA; this proportion is derived from the time away from home W_{nt} as follows. We assume that individuals who are mobile within their home FSA spend $\psi \leq 1$ (default 1) as much time away from home as those who are mobile within other FSAs. Thus, the model for W_{nt} is:

$$W_{nt} = W_n^* \left(\psi B_{nn't\,(n=n')} + \sum_{n'} B_{nn't\,(n\neq n')} \right) \tag{A.4}$$

where W_n^* is the expected proportion of time away from home among travellers from FSA n to other FSAs. During the reference period, all individuals are mobile so $B_{nn't_0\,(n=n')} = 0$ is known, and W_n^* can be calculated by rearranging Eq. (A.4). Then, during the pandemic months, the observed W_{nt} can be used to infer $B_{nn't\,(n=n')}$ by rearranging again:

$$B_{nn't\,(n=n')} = \left(\frac{W_{nt}}{W_n^*} - \sum_{n'} B_{nn't\,(n\neq n')} \right) \psi^{-1} \tag{A.5}$$

Figure A.3 plots the distribution of expected times away from home W_{nt} in hours (a) and relative to the reference period W_{nt_0} (b).

During the pandemic months, $B_{nn't}$ will sum to less than 1 for many FSAs; this is desirable, since we interpret $1 - \sum_{n'} B_{nn't}$ as the expected proportion of individuals in FSA n who did not leave the home each day during month t, and thus would not form any mobility-related contacts. Similarly, if $B_{nn't}$ sums to greater than 1, the net result in the force of infection is that individuals may form more contacts than expected from the reference period, which is also usually acceptable (such as during Aug–Sept 2020).

Finally, as noted in § 3.1, Eq. (17)/(A.6) can be used to aggregate $B_{nn'}$ to obtain $B_{gg'}$ (repeated here for reference):

$$B_{gg'} = \sum_{n \in S_g} \sum_{n' \in S'_g} B_{nn'}^h \tag{A.6}$$

where S_g is the set of FSAs (n) corresponding to decile g. Figure A.4 illustrates the monthly matrices $B_{gg't}$ for the study period with and without the inferred mobility within the home decile. Mobility patterns outlined in Figure A.4 are generally consistent across months, with the overall proportions of travelling individuals proportional to the ratio $V_{nn't}/H_{nt_0}$ depicted in Figure A.2c.

(a) Hours away from home

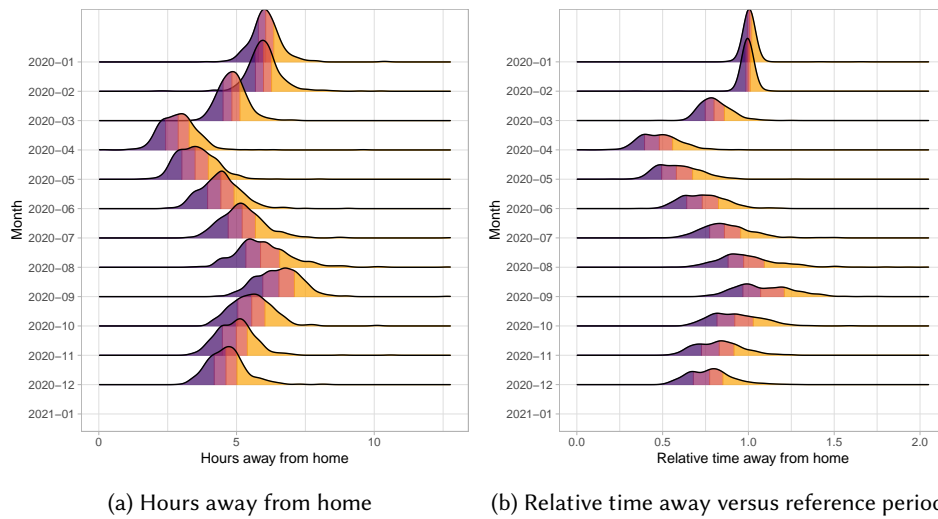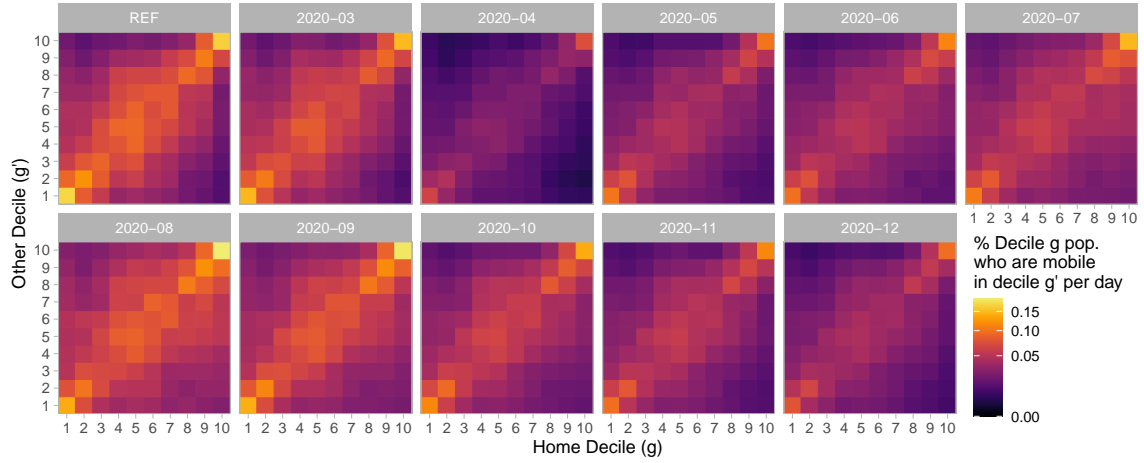(b) Relative time away versus reference period

Figure A.3: Expected time away from home in hours and relative to the reference period

Data from approximately 2% of mobile devices in Ontario, Canada; distributions show the density of ratio values across all 513 FSAs; coloured segments indicate the 4 quantiles of each distributions; reference period: Jan–Feb 2020.

(a) Inter-FSA mobility
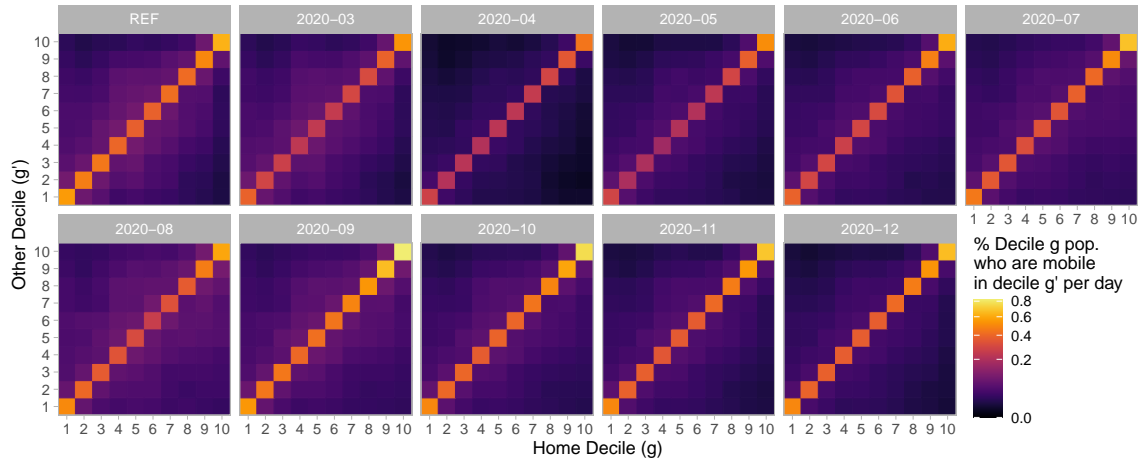
(b) Inter + intra-FSA mobility

Figure A.4: Mobility matrix $B_{gg'}$, representing the expected proportion of individuals in decile (patch) g who are mobile in decile g' per day, stratified by calendar month and the reference period

Derived from mobile device geolocation data; deciles represent groupings of Ontario forward sortation areas (FSAs) by cumulative COVID-19 cases between Dec 2020–May 2021; colour scale is square-root transformed to improve perception of smaller values; reference period: Jan–Feb 2020.

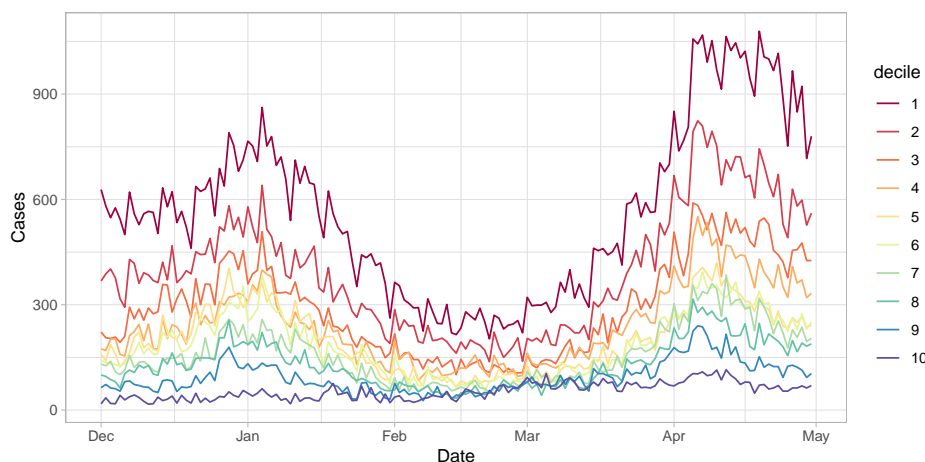## A.4   Additional Data for Ontario Patches

Figure A.5: Time trends in daily COVID-19 cases across the 10 deciles (patches) in Ontario

Deciles represent groupings of Ontario forward sortation areas (FSAs) by cumulative COVID-19 cases between Dec 2020–May 2021.
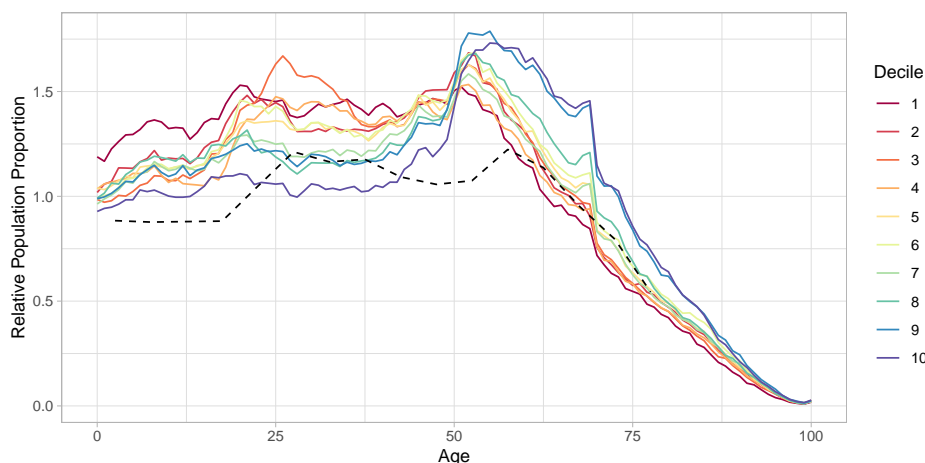

Figure A.6: Population age distributions across the 10 deciles (patches) in Ontario

Deciles represent groupings of Ontario forward sortation areas (FSAs) by cumulative COVID-19 cases between Dec 2020–May 2021; solid coloured lines correspond to deciles; 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.