



ELSEVIER

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

Method Article

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



Jesse Knight^{a,b,*}, Huiting Ma^a, Amir Ghasemi^c, Mackenzie Hamilton^a,
Kevin Brown^{d,e}, Sharmistha Mishra^{a,b,e,f}

^a MAP Centre for Urban Health Solutions, Unity Health Toronto, Canada

^b Institute of Medical Science, University of Toronto, Canada

^c Communications Research Centre Canada, Ottawa, Canada

^d Public Health Ontario, Canada

^e Dalla Lana School of Public Health, University of Toronto, Canada

^f Division of Infectious Diseases, Department of Medicine, University of Toronto, Canada

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. Using their approach, mixing between patches is greater than mobility data alone would suggest, because individuals from patches A and B can form contacts if they meet in patch C. We build upon their approach to address three potential gaps that remain, outlined in the bullets below. We describe the steps required to implement our approach in detail, and present step-wise results of an example application to generate contact matrices for SARS-CoV-2 transmission modelling in Ontario, Canada. We also provide methods for deriving the mobility matrix based on GPS mobility data (appendix).

- Our approach includes a distribution of contacts by age that is responsive to the underlying age distributions of the mixing populations.
- Our approach maintains different age mixing patterns by contact type, such that changes to the numbers of different types of contacts are appropriately reflected in changes to overall age mixing patterns.
- Our approach distinguishes 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 of, and any other visitors to the patch.

© 2021 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

* Corresponding author.

E-mail address: jesse.knight@mail.utoronto.ca (J. Knight).

ARTICLE INFO

Method name: Contact patterns in patch-based models with recurrent mobility

Keywords: COVID-19, transmission modelling, heterogeneous mixing, contact patterns, age groups, population mobility

Article history: Received 1 October 2021; Accepted 21 December 2021; Available online 28 December 2021

Specifications Table

Subject Area:	Mathematics
Specific Subject Area:	Epidemic Modelling
Method Name:	Contact patterns in patch-based models with recurrent mobility
Reference of Original Method:	Arenas et al. [1]
Resource Availability:	github.com/mishra-lab/age-patch-mobility-mixing

Introduction

Contact patterns are key determinants of epidemic dynamics because they define who can be infected, by whom, and how quickly [2]. Arenas et al. [1] develop a patch-based model of SARS-CoV-2 transmission applied to Spain, in which the modelled population is stratified by geographic patches and three age groups. Following foundational work by [3–5], 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 improved age mixing patterns, which are stratified by different contact types and are responsive to the age distributions of mixing populations, as proposed by [6]. We also explore some practical considerations in parameterizing such models.

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 (e.g. household, workplace, etc.). Let $B_{gg'}$ be the proportion of population P_g who travel to g' each day, or the “mobility matrix”.²

Original Approach

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

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

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 [7]:

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

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 [8] for Spain, such that

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.

$\sum_{a'} \theta_{aa'} = 1$; 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:

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

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 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 [6]:

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

For a model with non-random age mixing and random (proportionate) 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). Specifically, if different patches have different age distributions, or different rates of per-person contact formation due to household size, employment, etc., 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 relative differences in demographics and/or contact rates by patch and/or age group. For example, if a given patch skews younger than average in age, and most contacts are formed with other members of the same patch, then fixed average $\theta_{aa'}$ would underestimate the number of younger contacts among residents of this patch, and overestimate the number of older contacts.

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 of contacts, and so the fixed age mixing matrix $\theta_{aa'}$ is applied to all contact types. As such, changes to the numbers of each type of contact are not paired with changes the overall mixing patterns. As illustrated by the POLYMOD study [2], age mixing patterns vary by contact type, such as highly age-assortative mixing in schools. Thus, differential reductions in each contact type would affect overall age mixing patterns. For example, if reductions in school-related contacts due to school closures were not reflected in $\theta_{aa'}$, then the relative contribution of children to overall transmission could be overestimated during the period of school closures.
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 non-mobile 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 as only forming contacts with other individuals from their residence patch. That is, scenarios may exist wherein a fraction of the population only has household contacts, as could be the case with public health measures such as lockdowns. As illustrated in Figure A.1, the original approach may overestimate inter-patch connectivity during periods of reduced mobility (via lockdowns) versus an approach in which some or all non-mobile populations are limited to contacts with others from their residence patch and not with visitors. Thus, the original approach [1] could underestimate the impact of confinement lockdown strategies on inter-patch transmission reduction.

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

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.

Age Mixing

Prem et al. [9] project contact patterns by 5-year age groups from the POLYMOD study [2] onto 177 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 the 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. [6] 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 sampling and/or reporting error in the POLYMOD survey. To ensure that the overall mixing matrix $C_{gag'a'y}$ will satisfy the balancing principle, the input age mixing matrix $C_{aa'y}^u$ must satisfy the principle. A simple solution is to average $C_{aa'y}^u$ with its transpose to obtain the “balanced” matrix $C_{aa'y}^{ub}$ (Figure 4b):

$$C_{aa'y}^{ub} = \frac{1}{2} [C_{aa'y}^u + C_{aa'y}^{uT}] \tag{9}$$

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

³ The “others” contact type in [9] is itself derived from the combination of “leisure”, “transport”, and “other” contact types from [2], 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 [6].

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

A final challenge in applying the contact matrices from [9] 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 $C_{gg'a'y}$, 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 [9], 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 $C_{a_1a'_1y}^{ub}$ to obtain $C_{a_*a'_*y}^{ub}$ using matrix multiplication with indicator matrix A :

$$C_{a_*a'_*y}^{ub} = \frac{A_{a_*a_1}}{\sum_{a_1} A_{a_*a_1}} C_{a_1a'_1y}^{ub} A_{a'_*a'_1}^T, \quad A_{a_*a_1} = \begin{cases} 1, & a_1 \in a_* \\ 0, & a_1 \notin a_* \end{cases} \quad (10)$$

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 $C_{a_*a'_*y}^{ub}$ 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, $C_{a_*a'_*y}^{ub}$ can later be multiplied by the population age distribution of interest –reversing Eq. (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.

Mobility-Related Mixing

In conceptualizing mobility-related mixing, we define two types of contexts in which contacts can be formed, similar to “residences” and “common” sites in [10]:

- **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.

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 $h_y \in [0, 1]$ representing the proportion of type y contacts that are formed in the home pool, and the remainder $(1 - h_y)$ are formed with travel pools. For example, we could have $h_y = 1$ for household contacts, $h_y = 0$ for work contacts, and $h_y = 0.5$ 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: $(1 - h_y)B_{gg'} (g \neq g')$
2. **Mobile at Home:** individual formed contacts within their local travel pool g
Proportion of contacts: $(1 - h_y)B_{gg'} (g = g')$
3. **Non-Mobile at Home:** individual formed contacts within their local home pool g
Proportion of contacts: $h_y\delta_{gg'}$

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.

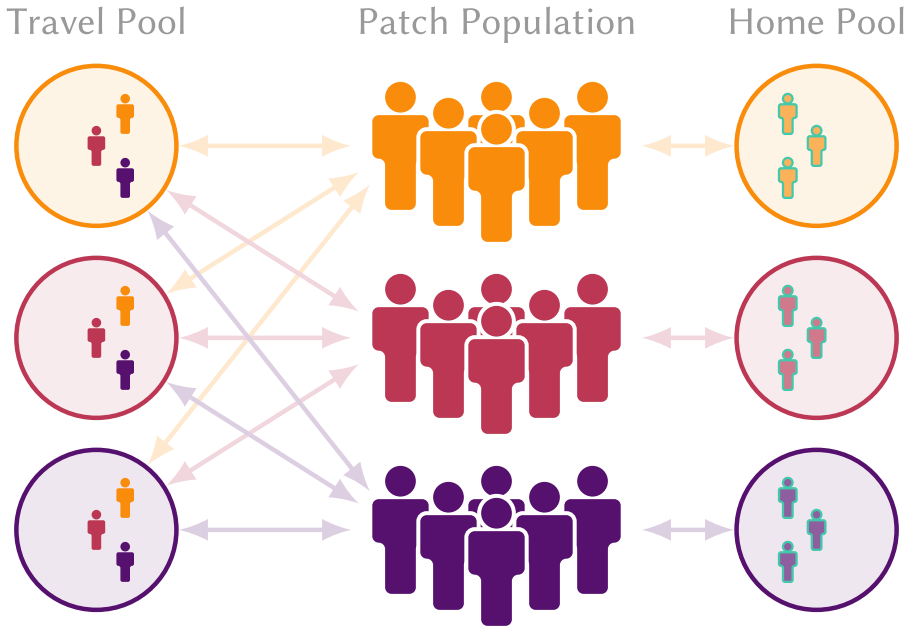


Fig. 1. Toy example of “home” vs “travel” mixing pools for a network with 3 patches and 50% individuals mobile. Contacts in the home pool are formed exclusively with other members of the residence patch, whereas contacts in the travel pool may be formed with any visitors to the patch. Non-mobile populations are indicated with faded colour and green outline

In 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 and with 1 contact per person, the total number of contacts formed between $P_{gay}^{g^*}$ and $P_{g'a'y}^{g^*}$ is given by the outer product:

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

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 numbers and patterns of contacts by age can be applied via multiplication:

$$X_{gag'a'y}^{g^*} = X_{gag'a'y}^{g^*r} C_{aa'y}^{ub} / \sum_{a'_1} A_{a'a'_1} \tag{13}$$

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

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 $\sum_{a'_1} A_{a'_1}$ 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.

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:

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

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 (14 million) 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, excluding cases among residents of long-term care homes, between 15 January 2020 and 28 March 2021 [11]. Thus each patch represents approximately 10% of the Ontario population (37–68 FSAs), but not contiguous geographic regions. Such definitions were used to support allocation and prioritization of COVID-19 vaccines to “hot spot” neighbourhoods in Ontario [12,13]. Figure 2 illustrates the locations of the FSAs and their decile rank, which is synonymous with their patch index. Figure A.10 plots the daily incidence of COVID-19 cases per patch, and Figure A.11 plots the age distributions of each patch. Age groups were then defined to reflect historical and hypothetical COVID-19 vaccine eligibility in Ontario:

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

Data

Ontario population sizes by age and FSA P_{ga} were obtained from the 2016 Canadian Census via Statistics Canada⁸ and aggregated from 1-year age groups (a_1) into 5-year (a_5) and target (a_*) age

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

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

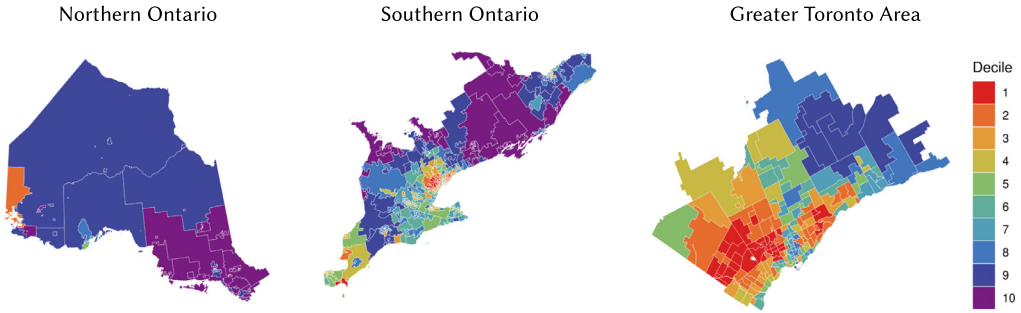


Fig. 2. Ontario forward sorting areas (FSAs, $N = 513$), stratified by decile rank in cumulative covid-19 cases between 15 Jan 2020–28 Mar 2021; decile rank was used to group FSAs into 10 patches for transmission modelling.

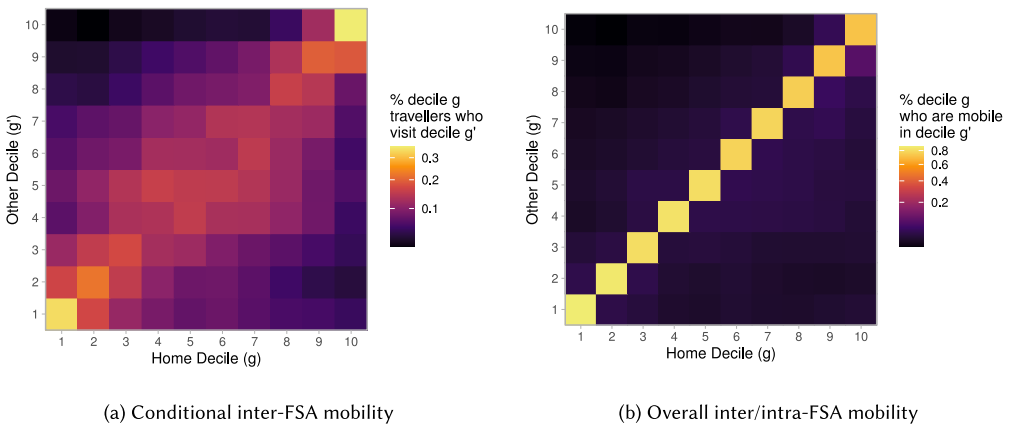


Fig. 3. Mobility matrix $B_{gg'}$, representing the expected proportion of individuals in decile (patch) g who are mobile in decile g' per day. Derived from mobile device geolocation data; deciles represent groupings of Ontario forward sorting areas (FSAs) by cumulative covid-19 cases between 15 Jan 2020–28 Mar 2021; colour scale is square-root transformed to improve perception of smaller values; reference period: Jan–Feb 2020.

groups as needed. We obtained the final output contact matrices $C_{aa'y}$ for Canada from [9], 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 [9].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 [14] during January–December 2020. Appendix A.3 details the specific methods and assumptions used; to summarize: Each device was assigned an approximate home location ($152.9 \text{ m} \times 152.4 \text{ m}$) based on the most common location during overnight hours for each calendar month. This location was then used to determine the home FSA (n). The proportion of time spent outside the home location each day, stratified by inside vs outside the home FSA, was also used to estimate the relative proportions of intra- vs inter-FSA mobility. Finally, the total numbers of visits to other FSAs (n') by all devices were used to estimate

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/6e0eebc>.

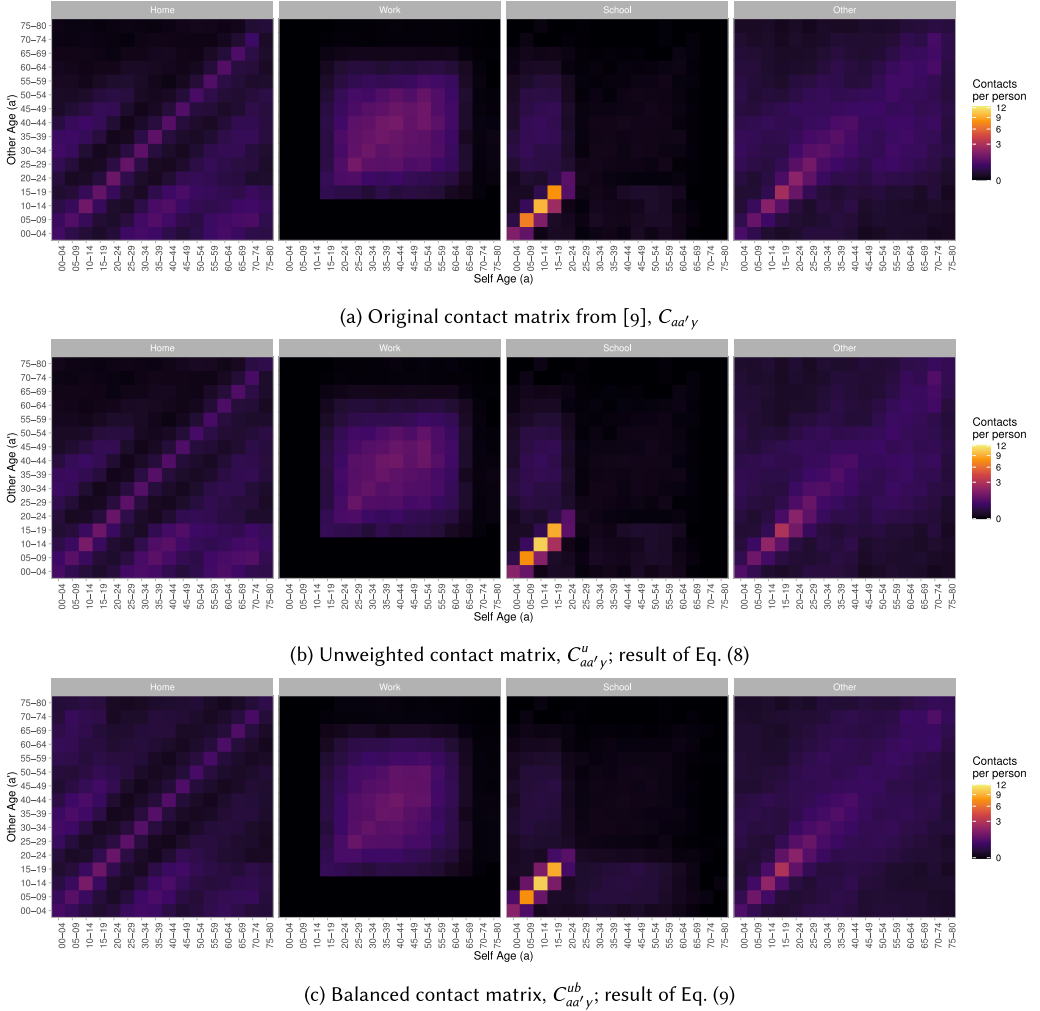


Fig. 4. Intermediate results in obtaining unweighted and balanced age contact matrices $C_{aa^y}^{ub}$ (expected number of type y contacts per person per day in each age group a , with those other age groups a') from population-weighted matrices C_{aa^y} from [9] which may not satisfy contact balancing. Contact matrices for Canada, derived from [9]; colour scales are square-root transformed to improve perception of smaller values.

the conditional probability of travelling from FSA n to FSA n' , given that an individual will travel outside the home FSA n .

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 (Jan–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

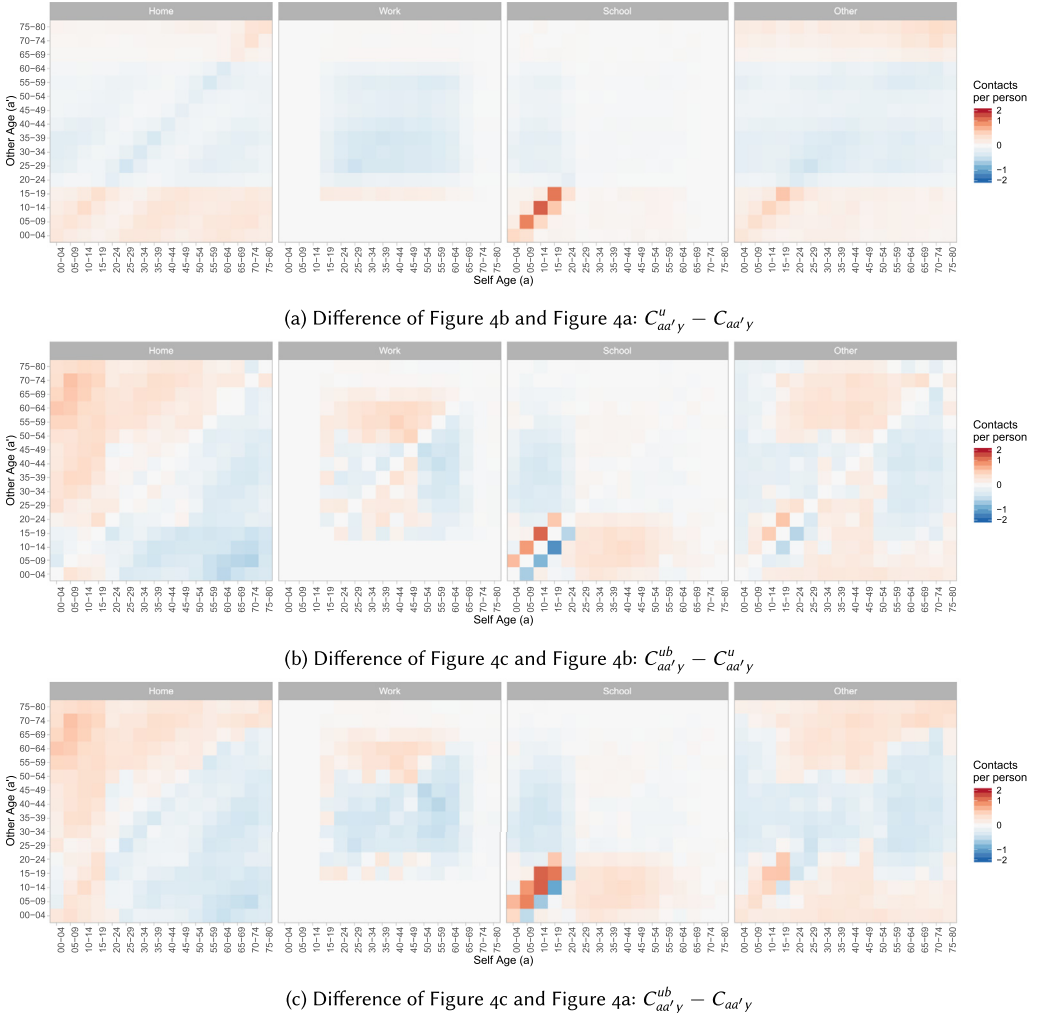


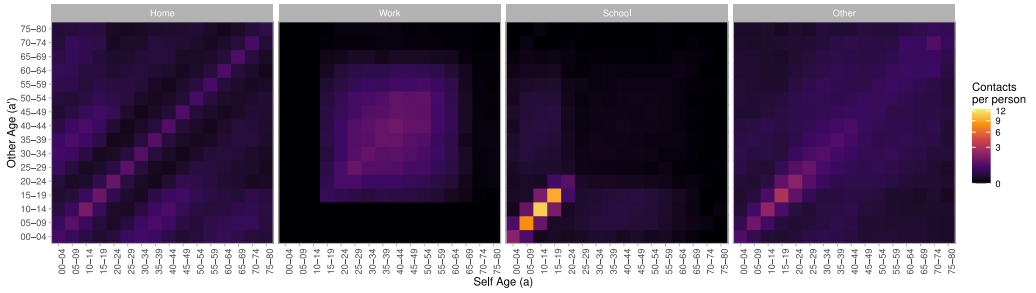
Fig. 5. Differences between intermediate results shown in Figure 4. Contact matrices for Canada, derived from [9]; colour scales are square-root transformed to improve perception of smaller values.

by age group, although such differences could be included in the model by adding a relative rate in Eq. (11).

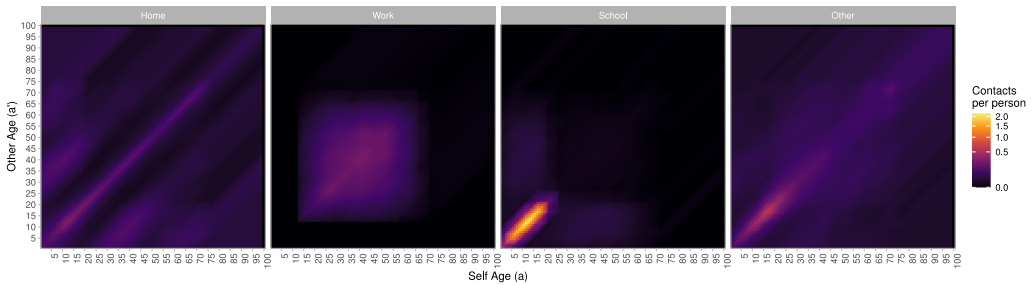
Finally, we specified the proportions of each contact type assumed to be formed with the home pool:

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

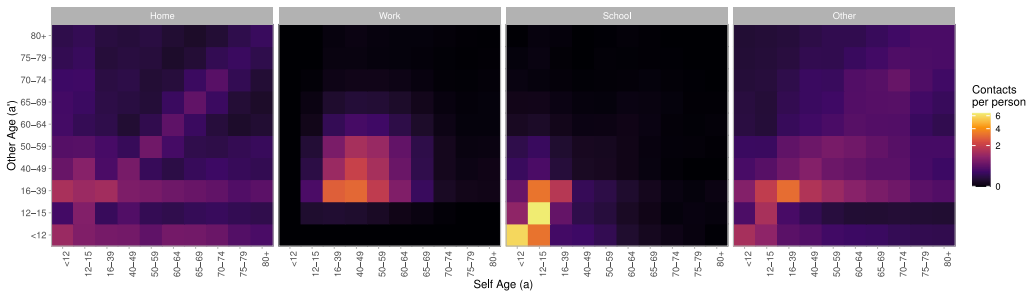
The parameters P_{ga} , C_{aa^y} , B_{gg^y} , and h_y represent the necessary inputs to our approach for calculating $C_{gag^a}^y$. The following § 3.2 walks step-wise through the approach and presents all major intermediate results.



(a) Original age groups $C_{a_5 a_5}^{ub}$; result of Eq. (9); identical to Figure 4c



(b) 1-Year age groups $C_{a_1 a_1}^{ub}$; result of diagonal edge padding and bilinear interpolation



(c) Target age groups $C_{a_w a_w}^{ub}$; result of Eq. (10)

Fig. 6. Intermediate results in obtaining age-restratified contact matrices C_{aa}^{ub} (expected number of type y contacts per person per day in each age group a_* , with those from age groups a'_*) from matrices $C_{a_5 a_5}^{ub}$ with 5-year age stratifications a_5 . Contact matrices for Canada, derived from [9]; 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.

Results

Figure 4 illustrates the contact matrices C_{aa}^{ub} from Prem et al. [9], 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 et al. [9] (Figure A.11 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.

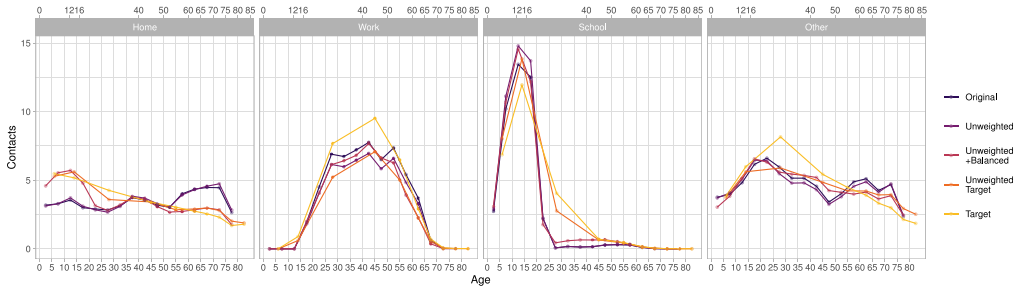


Fig. 7. Total contacts per person per day $C_{ay} = \sum_{a'} C_{aa'y}$ for each intermediate step in obtaining $C_{a,a',y}^{ub}$, stratified by contact type. Contact matrices for Canada, derived from [9]; modelled contacts for each age group are plotted at the midpoint of the age group; the cut points for the original age groups a_5 from [9] and the target age groups a_* in our application are indicated on the bottom and top x-axes, respectively.

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.

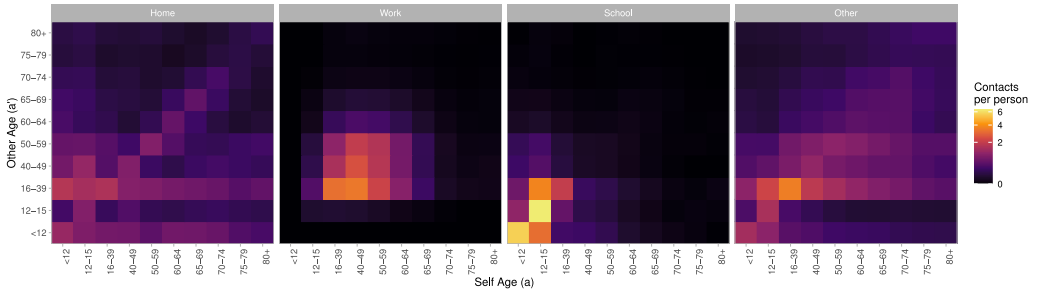
Finally, Figure 8 illustrates the margins (sum over “other” strata and population-weighted average over “self” strata) of the complete mixing matrices $C_{gag'a'y}$, in terms of age groups a & a' (Figure 8a), and patches/deciles g & g' (Figure 8b). Such margins are computed as follows:

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

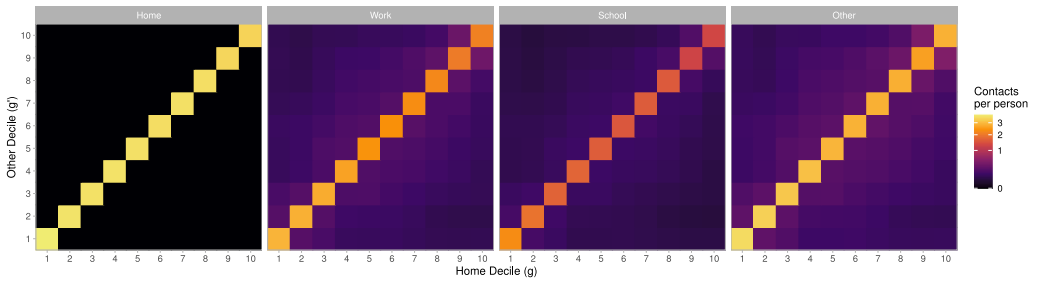
$$C_{gg'y} = \sum_a P_{ga} \sum_{a'} C_{gag'a'y} / \sum_a P_{ga} \tag{20}$$

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 less than if completely random mixing was assumed.

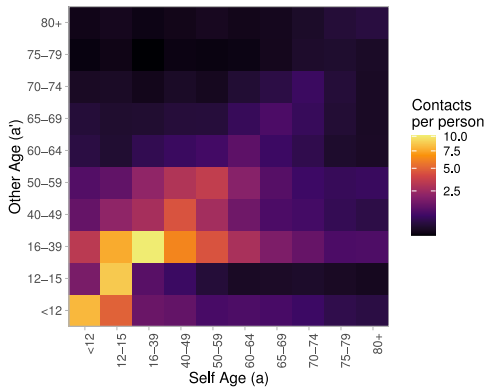
¹⁰ The corresponding matrix of absolute contact $X_{aa'y} = C_{aa'y}P_{a'}$ will be symmetric, and thus will include symmetric horizontal and vertical streaks.



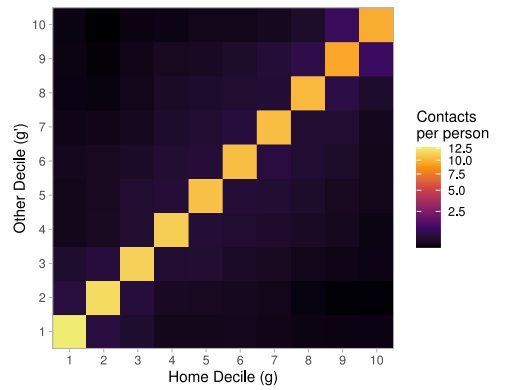
(a) Stratified by age groups and contact type; from Eq. (19)



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



(c) Stratified by age groups, sum across contact types



(d) Stratified by patches, sum across contact types

Fig. 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 [9]; colour scales are square-root transformed to improve perception of smaller values.

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 SARS-CoV-2. 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. In this paper, we proposed approaches to improve upon the approach 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.

The first key change in the proposed approach draws on [6] 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 . This revised approach can also ensure contacts balance if population sizes change over time, such as in the case of large differential mortality by age group or patch; or if the numbers of contacts formed per-person differ by patch, such as if individuals in some patches have higher numbers of work contacts.

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 can 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.

Funding

The study was supported by: the Natural Sciences and Engineering Research Council of Canada (NSERC CGS-D); the Canadian Institutes of Health Research (VR5-172683); and the 2020 COVID-19 Centred Research Award from the St Michael's Hospital Foundation Research Innovation Council.

Contributions

JK conceived of and developed the approach, reviewed the literature, conducted the mixing and mobility analyses, and wrote the code and manuscript. AG conducted and led data analyses and developed the mobility metrics. HM provided critical input into the assumptions and bias adjustments with the mobility data. MH, KB, and SM contributed critical insights to the assumptions, data, approach, and interpretation. All authors were involved in interpretation and manuscript revisions.

Acknowledgements

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

Reported COVID-19 cases were obtained from the Public Health Ontario (PHO) integrated Public Health Information System (iPHIS), via the Ontario COVID-19 Modelling Consensus Table, and with approval from the University of Toronto Health Sciences REB (protocol No. 39253).

Appendix

The Appendix for this article can be found, in the online version, at doi:[10.1016/j.mex.2021.101614](https://doi.org/10.1016/j.mex.2021.101614)

References

- [1] A. Arenas, W. Cota, J. Gómez-Gardeñes, S. Gómez, C. Granell, J.T. Matamalas, D. Soriano-Paños, B. Steinegger, Modeling the Spatiotemporal Epidemic Spreading of COVID-19 and the Impact of Mobility and Social Distancing Interventions, *Physical Review X* 10 (4) (2020) 041055, doi:[10.1103/PhysRevX.10.041055](https://doi.org/10.1103/PhysRevX.10.041055).
- [2] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G.S. Tomba, J. Wallinga, J. Hejine, M. Sadkowska-Todys, M. Rosinska, W.J. Edmunds, Social contacts and mixing patterns relevant to the spread of infectious diseases, *PLoS Medicine* 5 (3) (2008) 0381–0391, doi:[10.1371/journal.pmed.0050074](https://doi.org/10.1371/journal.pmed.0050074).
- [3] L. Sattenspiel, K. Dietz, A structured epidemic model incorporating geographic mobility among regions, *Mathematical Biosciences* 128 (1-2) (1995) 71–91, doi:[10.1016/0025-5564\(94\)00068-B](https://doi.org/10.1016/0025-5564(94)00068-B).
- [4] D. Balcan, A. Vespignani, Phase transitions in contagion processes mediated by recurrent mobility patterns, *Nature Physics* 7 (7) (2011) 581–586, doi:[10.1038/nphys1944](https://doi.org/10.1038/nphys1944).
- [5] J. Gómez-Gardeñes, D. Soriano-Paños, A. Arenas, Critical regimes driven by recurrent mobility patterns of reaction-diffusion processes in networks, *Nature Physics* 14 (4) (2018) 391–395, doi:[10.1038/s41567-017-0022-7](https://doi.org/10.1038/s41567-017-0022-7).
- [6] S. Arregui, A. Aleta, J. Sanz, Y. Moreno, Projecting social contact matrices to different demographic structures, *PLoS Computational Biology* 14 (12) (2018) e1006638, doi:[10.1371/journal.pcbi.1006638](https://doi.org/10.1371/journal.pcbi.1006638).
- [7] E.H. Kaplan, Modeling HIV infectivity: Must sex acts be counted? *Journal of Acquired Immune Deficiency Syndromes* 3 (1990) 55–61. <https://pubmed.ncbi.nlm.nih.gov/2293643/>
- [8] K. Prem, A.R. Cook, M. Jit, Projecting social contact matrices in 152 countries using contact surveys and demographic data, *PLoS Computational Biology* 13 (9) (2017) e1005697, doi:[10.1371/journal.pcbi.1005697](https://doi.org/10.1371/journal.pcbi.1005697).
- [9] K. Prem, K. van Zandvoort, P. Klepac, R.M. Eggo, N.G. Davies, A.R. Cook, M. Jit, Projecting contact matrices in 177 geographical regions: An update and comparison with empirical data for the COVID-19 era, *PLoS Computational Biology* 17 (7) (2021), doi:[10.1371/journal.pcbi.1009098](https://doi.org/10.1371/journal.pcbi.1009098).
- [10] C. Granell, P.J. Mucha, Epidemic spreading in localized environments with recurrent mobility patterns, *Physical Review E* 97 (5) (2018) 052302, doi:[10.1103/PhysRevE.97.052302](https://doi.org/10.1103/PhysRevE.97.052302).
- [11] ICES, COVID-19 Dashboard, Applied Health Research Questions, volume 2021-0950-080-000, 2020. <https://www.ices.on.ca/DAS/AHRQ/COVID-19-Dashboard>
- [12] S. Mishra, N.M. Stall, H. Ma, A. Odutayo, J.C. Kwong, U. Allen, K.A. Brown, I.I. Bogoch, A. Erman, T. Huynh, S. Ikura, A. Maltsev, A. McGeer, G. Moloney, A.M. Morris, M. Schull, A. Siddiqi, J. Smylie, T. Watts, K. Yiu, B. Sander, P. Jüni, A Vaccination Strategy for Ontario COVID-19 Hotspots and Essential Workers, *Science Briefs of the Ontario COVID-19 Science Advisory Table 2* (26) (2021), doi:[10.47326/ocsat.2021.02.26.1.0](https://doi.org/10.47326/ocsat.2021.02.26.1.0).
- [13] K.A. Brown, N.M. Stall, E. Joh, U. Allen, I.I. Bogoch, S.A. Buchan, N. Daneman, G.A. Evans, D.N. Fisman, J.L. Gibson, J. Hopkins, T. van Ingen, A. Maltsev, A. McGeer, S. Mishra, F. Razak, B. Sander, B. Schwartz, K. Schwartz, A. Siddiqi, J. Smylie, P. Jüni, A Strategy for the Mass Distribution of COVID-19 Vaccines in Ontario Based on Age and Neighbourhood, *Science Briefs of the Ontario COVID-19 Science Advisory Table 2* (10) (2021), doi:[10.47326/ocsat.2021.02.10.1.0](https://doi.org/10.47326/ocsat.2021.02.10.1.0).
- [14] A. Ghasemi, N. Daneman, I. Berry, S.A. Buchan, J.-P. Soucy, S. Sturrock, K.A. Brown, Impact of a nighttime curfew on overnight mobility, *medRxiv* (2021), doi:[10.1101/2021.04.04.21254906](https://doi.org/10.1101/2021.04.04.21254906).