Supplementary Information: Modeling the effect of transient populations on epidemics in Washington DC

NIDHI PARIKH, MINA YOUSSEF, SAMARTH SWARUP, STEPHEN EUBANK

Networks Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech, USA

This paper provides the supplementary information for the article "Modeling the effect of transient populations on epidemics in Washington DC" submitted to *Scientific Reports*. Section 1 describes the detailed process for generating transient population followed by supporting tables for simulation results in section 2. Details of the ordinary differential equation (ODE) model are explained in Section 3.

1 Synthetic Transient Population

We generated an augmented synthetic population for the Washington DC Metro Area, which combines a previously generated resident population (the "base population" consisting of 4.13 million people) with a transient population consisting of tourists and business travelers. Since details about generating the base synthetic population are not novel to the present work and are described elsewhere [1], we only describe in detail the methodology for generating the synthetic transient population.

1.1 Data Available

Demographic data about transients were obtained from Destination DC. We also used data from the the Smithsonian Institution about daily numbers of visits to various Smithsonian museums. Finally, we used data from Dun & Bradstreet to identify places that tourists visit, based on Standard Industrial Classification (SIC) codes. The data sets used for generating the synthetic population are listed in table S1.

The methodology for generating the transient population broadly follows that for generating the base population. We first use demographic data to represent transient individuals and transient parties (groups). Each transient party is placed in a hotel which serves as their home for the period of the visit. Each transient individual is then assigned activities to perform during the day like staying in the hotel, visiting museums and other tourist destinations (or work activities, for business travelers), going to restaurants, and various night life activities. Each activity is represented by the type of activity, the time each activity begins and ends, and the location for the activity. A location is chosen for each activity based on the type of activity using Dun & Bradstreet data.

Used for	data source
Base US population	American Community Survey
	National Center for Education Stat.
	National Household Travel Survey
	Navteq
	Dun & Bradstreet
Transient population	Destination DC
(additional)	Smithsonian visit counts

Table 1: Datasets used for population generation.

1.2 Tourist Population

1.2.1 Generating Synthetic Tourists

The goal here is to combine various demographic distributions and represent synthetic tourist parties and individuals with demographics drawn from these distributions. According to data from Destination DC, about 50000 visitors visit Washington DC every day, 55% of these are leisure travelers and the rest are business travelers. They also provide distributions of age, household income, party size, marital status and if the household has children. Please note that these data are given only for adult, overnight leisure travelers. Also age and marital status are individual level demographics while household income and party size are household (or party) level demographics, and hence they need to be treated differently. These distributions are not independent of each other within a party i.e, a married couple is more likely to travel together and hence we cannot sample independently from the given distributions.

High Level View

Our approach is simple: we assume a small set of rules about party structure and then do sampling without replacement from the given demographic distributions (since we know the total number of individuals to be generated) in combination with these rules to generate the tourist population.

We start by generating first party member (called householder) by sampling age, marital status, income and party size independently from the corresponding distributions and then generate other party members in relation to the householder. For example, if a party member is married then with a certain probability his/her spouse will also be part of the party and the age difference between them is assumed to be within a certain range (± 5 years in this case). All party members should have same household/party level demographics, household income and party size. Whenever an individual is assigned a demographic, the probabilities for selecting various categories of that demographic is adjusted to model sampling without replacement. We also assign all individuals some other demographic variables as assigned in the synthetic base population, e.g., sex (at random and in accordance with marital status) and employee status record (*esr*) and occupation code (*socp*) (by finding an individual with the closest income from the synthetic population of Washington DC metro area and assigning corresponding *esr* and *socp* codes).

For the present study most of these demographic details are irrelevant because disease parameters are not chosen to vary with demographic. However, that could be done in future work, and the synthetic populations can also be used for other studies where the demographic details are important.

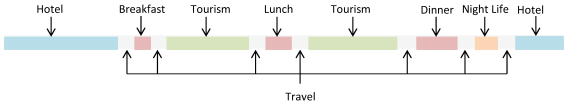
1.2.2 Assigning hotels

We identified hotels and lodging locations within I-495 loop area in Washington DC from Dun & Bradstreet (*D&B*) data. *D&B* is a commercial data set that gives information about business locations like longitude-latitude of buildings, number of employees (relative numbers), type of business going on there etc.

Each tourist party is assumed to stay at a hotel, which serves as their home location for the duration of the visit. Taking into consideration that tourists prefer to stay near downtown and each hotel has a capacity proportional to number of employees there, a hotel location (i) is chosen from the available pool with probability proportional to $num\_employees(i) \times e^{\delta \times distance\_from\_white\_house}$.

1.2.3 Assigning Activities

Since we could not find any data about activity sequences for tourists, we assumed a template for it, as illustrated in figure S1. We assume that all individuals in a party travel together and hence have the same activity sequence and go to the same locations. However within a location (building) they may go to different sublocations (rooms). Each party's activity sequence contains information about the type of activity and the start time and duration. Location and sublocation are decided later.



Activity	Start time	Duration
Hotel	12 am and after dinner/night life activity	Until breakfast
Breakfast	between 7 am to 10 am	30 mins to 1 hour
Tourism	after breakfast, lunch or other tourism activity	30 mins to 30 hours
Lunch	after noon	1 hour to 1.5 hours
Dinner	after 6 pm	1 hour to 2 hours
Night Life	after dinner	1 hour to 3 hours
Travel	between two activities	0 mins to 1 hour

Figure 1: Activity template for tourists.

Each party starts the day with a hotel activity. It is followed by breakfast which could happen at the same hotel (with 60% probability) or at some other location (with 40% probability). Each tourism activity shown in figure S1 is divided into one or more tourism activities with some travel time between them. Each party goes for lunch after $12:00$ pm which is again followed by one or more tourism activities and then dinner. After dinner, with 50% probability they go back to the hotel directly and stay for the rest of the day. Otherwise they go for a night life activity and then back to their hotel. Each pair of activities is separated by travel time of 0 mins to one hour.

1.2.4 Locating Activities

We identified locations for tourism, eating, and night life activities from *D&B* data. Tourism activity locations include places like museums, art galleries, planetarium, historical societies, and botanical and zoological gardens. Eating activity locations include various restaurants and night life activity locations include bars and pubs, night clubs, and movie theatres.

Assuming that most of the transients to Washington DC visit museums which are around the National Mall and plan their trip around that area, we choose locations for all activities based on the distance from the National Mall. Each location has a capacity (again assumed to be proportional to the number of employees at that location according to *D&B*). Considering both of these factors, a location (i) is chosen from the available locations for a given activity type with probability proportional to $num\_employees(i) \times e^{\delta \times distance\_from\_national\_mall}$.

The Smithsonian Institution provides data about daily visit counts at various museums. To match the number of visits in our synthetic population at museum locations with these counts, we adjusted weights (number of employees) from *D&B* data for these museums. However, the number of transients is not sufficient to account for all the visits to some museum locations. For example, the National Air and Space Museum has about 80000 visits per day. So we adjusted the activities of some individuals in the base population and routed them to these locations to match the visit counts exactly. This also creates mixing between the transient and the base population, which is an important factor in the spread of disease.

1.2.5 Sublocation Modeling

An activity location typically corresponds to a building and sublocations correspond to rooms in the building. Sublocation modeling involves deciding which room a person visits and hence with whom he comes into contact. All individuals present at the same sublocation at the same time are assumed to be in contact with each other. All individuals in a party are assumed to meet each other at the hotel and hence are assigned same sublocation.

For other locations, we follow the assumption made in the creation of the base population [1], that sublocations have a capacity of 25 people, and that each person, upon arriving at a location is assigned to a sublocation where he remains for the duration of his activity at that location. In reality, people would come into contact with more than 25 people at major tourist venues like the National Air and Space Museum and the National Museum of Natural History. Also, inside museums, they do not stay at the same location during the entire period of their visit. They keep moving from one exhibition to another. We therefore create a simple stochastic process modeling movement between sublocations at for the four biggest tourism locations - the National Air and Space Museum (NASM), the National Museum of Natural History (NMNH), the National Museum of American History (NMAH) and the National Art Gallery (NAG). For these four locations, we decided the number of sublocations by looking at their floor plans. While modeling visits to these locations, a person's visit is divided into the interval of 5 to 15 minutes and a person keeps moving to different sublocations (chosen at random) within the location.

1.3 Business Travelers

The process used is similar to the synthetic tourist population generation process.

1.3.1 Generating Synthetic Business Travelers

We could not find any demographic data for business travelers. The only information available is that about 45% of the transients are business travelers. We followed the same procedure as for generating tourists but with some assumptions. Each business traveler is assumed to be by himself and hence party size is assumed to be 1. Age is assumed to be between 18 to 70 years. Marital status is chosen from unmarried, married, and divorced/widowed with equal probability. The household income distribution is assumed to be Gaussian with peak and standard deviation equal to the maximum and average household income in Washington DC metro area respectively.

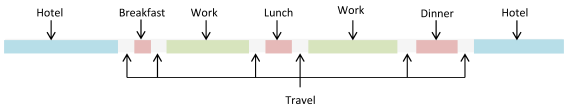
We assigned other demographic variables i.e., sex, *socp*, *esr* following the same process as for tourist population.

1.3.2 Assigning Hotels

Business travelers are assigned hotels the exact same way as tourists.

1.3.3 Assigning Activities

Here also, since we could not find any data about the activity sequences for business travelers, we assumed a template for it. The activity sequence created contains information about the type of activity and the start time and duration. Location and sublocation choice are described in the next subsections. The template for activities is as shown in figure S2.



Activity	Start time	Duration
Hotel	12 am and after dinner activity	Until breakfast
Breakfast	between 7 am to 9 am	30 mins to 1 hour
Work	after breakfast and lunch	until lunch or dinner
Lunch	between 11:30 am to 12:30 pm	1 hour to 1.5 hours
Dinner	between 5 pm to 6 pm	1 hour to 2 hours
Travel	between two activities	0 mins to 1 hour

Figure 2: Activity template for business travelers.

Each business traveler starts the day at a hotel and hence with hotel activity. It is followed by breakfast which could happen at the same hotel (with 60% probability)or at some other location (with 40% probability). After breakfast, he leaves for work and stays there until lunch. After lunch, he goes back to work and stays there until dinner. After dinner, he goes back to the hotel and stays for the rest of the day. Here also, each pair of activities is separated by a travel time of 0 mins to one hour.

1.3.4 Locating Activities

The process used to assign activity locations is quite similar to that of tourists. Here we identified locations for work from the D&B data (these are also used as work locations for the base synthetic population) and eating (same as for tourists). Activity locations are assigned the same way as for tourists.

1.3.5 Sublocation Modeling

For sublocation modeling, we follow the same process as used for the base synthetic population.

2 Detailed simulation results

In this section we present results from simulation including statistical analysis of outcomes.

Table 2: The fraction of infections (residents + transients) over 120 days at four major tourist locations (average over 50 iterations): the National Air and Space Museum (NASM), the National Museum of Natural History (NMNH), the National Museum of American History (NMAH), and the National Gallery of Art (NGA).

Museum	No Intervention	Museums closed (5 days)	Museums closed (14 days)	Healthy behavior 80%	Healthy behavior 60%	Healthy behavior 40%	Healthy behavior 20%
Residents only							
NASM	0.033413	0.028599	0.027349	0.030576	0.029372	0.053294	0.000972
NMNH	0.028815	0.023072	0.020470	0.025957	0.023839	0.021263	0.000681
NMAH	0.017859	0.014188	0.011787	0.015265	0.012453	0.003830	0.000299
NGA	0.003242	0.002632	0.001025	0.001616	0.000692	0.000196	0.000034
Residents and transients							
NASM	0.067039	0.060212	0.055637	0.050779	0.038752	0.024470	0.002433
NMNH	0.056805	0.049966	0.044430	0.042688	0.031852	0.015943	0.001745
NMAH	0.031314	0.026539	0.022244	0.024012	0.016521	0.006404	0.000762
NGA	0.007305	0.005076	0.002608	0.004030	0.001656	0.000459	0.000091

2.1 Statistical Analysis

We compare various scenarios (residents only, residents + transients, and two intervention strategies, closing museums (four most-visited locations) and practice of helathy behavior (at these museums with the compliance rate of 50%), with 50 simulations for each case) in terms of the day when disease peaks, the fraction of residents infected at peak and the fraction of residents infected cumulatively over the course of simulation.

For comparing various scenarios, we first visualize data (i.e., the day of peak for each scenario) as a scatter plot and remove outliers before performing statistical tests. We perform following set of comparisons:

- **Evaluating the effect of transients:** To see if having transients in the city makes any difference to disease dynamics of the city, we compare residents only and residents + transients scenarios (without any intervention). We use independent samples t-test for comparison.
- **Evaluating the effect of interventions in the presence of transients:** For residents + transients population, we compare various intervention strategies to no intervention scenario (for residents + transients) to see if these interventions make any difference. We use Tukey's HSD test (with $\alpha = 0.05$) for comparison. Tukey's HSD test assumes data to be normally distributed and homogeneity of variances. As we have enough number of samples (even after removing outliers), we can assume data to be normally distributed using central limit theorem. However not all groups have equal variance (as seen in scatter plots). Hence, we choose maximal set of scenarios which satisfy the test of homogeneity of variances and compare those using Tukey's HSD test. For the rest of the intervention scenarios, we do pairwise comparison with no intervention scenario using Welch two sample t-test. Doing multiple t-tests in this fashion can lead to type I error (rejecting null hypothesis when it is actually true) but all the p-values that we obtain from t-tests are very small ($< 1.333e-08$). Hence, we are fairly confident that there is a significant difference whenever t-test rejects the null hypothesis.
- **Evaluating the effect of interventions in the absence of transients:** For residents only population also, we compare various intervention strategies to no intervention scenario (for residents only) to see if we get similar

results in the absence of transients as well. We use the same methodology and tests as used for the residents + transients population.

2.1.1 Comparing the day of peak

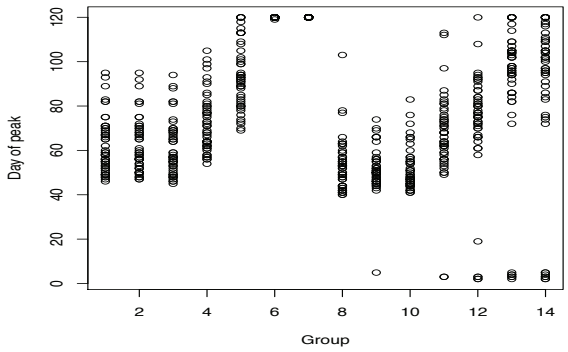


Figure 3: Scatter plots showing the day of peak verses group where groups are defined as follows: 1 - No interventions (residents only), 2 - Museums closed for 5 days (residents only), 3 - Museums closed for 14 days (residents only), 4 - Healthy behavior 80% (residents only), 5 - Healthy behavior 60% (residents only), 6 - Healthy behavior 40% (residents only), 7 - Healthy behavior 20% (residents only), 8 - No interventions (residents + transients), 9 - Museums closed for 5 days (residents + transients), 10 - Museums closed for 14 days (residents + transients), 11 - Healthy behavior 80% (residents + transients), 12 - Healthy behavior 60% (residents + transients), 13 - Healthy behavior 40% (residents + transients), 14 - Healthy behavior 20% (residents + transients). We remove outliers from each group before performing statistical tests.

Evaluating the effect of transients

Table 3: Independent sample t-test ($\alpha = 0.05$) comparing the day of peak for residents only and residents + transients scenarios (without any interventions). Significance level (0.019) for the Levene's test for equality of variance is less than $\alpha = 0.05$, which suggests that the variances of these two scenarios are not equal. As the significance level (0.000) for t-test (in the line for "equal variance not assumed") is less than $\alpha = 0.05$, we can conclude that disease peaks significantly earlier when the transients are considered.

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig.(2 tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	5.722	0.019	4.646	97	0.000	10.116	2177	5.795	14.437
Equal variances not assumed			4.661	89.509	0.000	10.116	2.170	5.804	14.428

Evaluating the effect of interventions in the presence of transients

Table 4: Levene test of homogeneity of variances (α=0.05) comparing the day of peak for following scenarios, for residents + transient population: no intervention, close museums (5 days), close museums (14 days), healthy behavior 80%, and healthy behavior 60%. The significance value (0.065) is greater than α (0.05). Hence, the variances are equal for all scenarios and we can proceed towards ANOVA. *Note:* Variances for "Healthy behavior 40%" and "healthy beahvior 20%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

Levene statistic	df1	df2	Sig.
2.249	4	232	0.065

Table 5: Analysis of variance (ANOVA) (α=0.05) comparing the day of peak, to see if any of the following scenarios for residents + transient population differ: no intervention, close museums (5 days), close museums (14 days), healthy behavior 80%, and healthy behavior 60%. The significance value (0.000) is less than α (0.05). Hence, at least one scenario differs from others and we can proceed towards Tukey's HSD test to see which scenarios differ. *Note:* Variances for "Healthy behavior 40%" and "healthy beahvior 20%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	25292.020	4	6323.005	71.004	0.000
Within Groups	20659.913	232	89.051		
Total	45951.932	236			

Table 6: Tukey's HSD test ($\alpha = 0.05$) comparing the day of peak for following scenarios, for residents + transients population: no intervention, close museums (5 days), close museums (14 days), healthy behavior 80%, and healthy behavior 60%. Scenarios in the same group (e.g. "No intervention", "Close museum (5 days)", and "close Museums (14 days)" are in group 1) are statistically similar to each other and scenarios in different groups (e.g. "No intervention" is in group 1 and "Healthy behavior 80%" is in group 2) are statistically different from each other and hence one is significantly better than the other. *Note:* Variances for "Healthy behavior 40%" and "healthy beahvior 20%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

Group	N	Subsets for α=0.05		
		1	2	3
No intervention	49	52.20		
Close museums (5 days)	49	52.24		
Close museums (14 days)	50	51.68		
Healthy behavior 80%	46		67.22	
Healthy behavior 60%	43			77.74
Sig.		0.998	1.000	1.000

Table 7: Welch t-tests ($\alpha = 0.05$, it assumes inequality of variances) comparing the day of peak for following intervention scenarios to the "no intervention" scenario (with mean = 52.2041), for residents + transients population: healthy behavior 40% and healthy behavior 20%. In all cases, p-values (2.2e-16) are less than α (0.05). So these scenarios differ significantly from the "no intervention" scenario. Doing multiple t-tests in this fashion can lead to type I error (rejecting null hypothesis when it is actually true) but all p-values obtained here are very small. So we are reasonably confident that these interventions delay peak significantly as compared to the "no intervention" scenario.

Scenario	t	df	p-value	mean	95% Confidence Interval	
					Lower	Upper
Healthy behavior 40%	-21.2452	78.08	2.2e-16	100.11364	-52.39900	-43.42011
Healthy behavior 20%	-20.0582	69.712	2.2e-16	102.32558	-55.10557	-45.13743

Evaluating the effect of interventions in the absence of transients

Table 8: Levene test of homogeneity of variances (α=0.05) comparing the day of peak for following scenarios, for resident population: no intervention, close museums (5 days), close museums (14 days), healthy behavior 80%, and healthy behavior 60%. The significance value (0.694) is greater than α (0.05). Hence, the variances are equal for all scenarios and we can proceed towards ANOVA. *Note:* Variances for "Healthy behavior 40%" and "healthy beahvior 20%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

Levene statistic	df1	df2	Sig.
0.557	4	245	0.694

Table 9: Analysis of variance (ANOVA) (α=0.05) comparing the day of peak to see if any of the following scenarios for resident population differ: no intervention, close museums (5 days), close museums (14 days), healthy behavior 80%, and healthy behavior 60%. The significance value (0.000) is less than α (0.05). Hence, at least one scenario differs from others and we can proceed towards Tukey's HSD test to see which scenarios differ. *Note:* Variances for "Healthy behavior 40%" and "healthy beahvior 20%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	38861.560	4	9715.390	58.069	0.000
Within Groups	40990.040	245	167.306		
Total	79851.600	249			

Table 10: Tukey's HSD test ($\alpha = 0.05$) comparing the day of peak for following scenarios, for resident population: no intervention, close museums (5 days), close museums (14 days), healthy behavior 80%, and healthy behavior 60%. Scenarios in the same group (e.g. "No intervention", "Close museum (5 days)", and "close Museums (14 days)" are in group 1) are statistically similar to each other and scenarios in different groups (e.g. "No intervention" is in group 1 and "Healthy behavior 80%" is in group 2) are statistically different from each other and hence one is significantly better than the other. *Note:* Variances for "Healthy behavior 40%" and "healthy beahvior 20%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

Group	N	Subsets for α=0.05		
		1	2	3
No intervention	49	62.32		
Close museums (5 days)	49	62.28		
Close museums (14 days)	50	61.22		
Healthy behavior 80%	46		72.36	
Healthy behavior 60%	43			94.02
Sig.		0.993	1.000	1.000

Table 11: Welch t-tests ($\alpha = 0.05$, it assumes inequality of variances) comparing the day of peak for following intervention scenarios to the "no intervention" scenario (with mean = 62.32), for resident population: healthy behavior 40% and healthy behavior 20%. In all cases, p-values (2.2e-16) are less than α (0.05) and hence these scenarios differ significantly from the "no intervention" scenario. Doing multiple t-tests in this fashion can lead to type I error (rejecting null hypothesis when it is actually true) but all p-values obtained here are very small. So we are reasonably confident that these interventions delay peak significantly as compared to the "no intervention" scenario.

Scenario	t	df	p-value	mean	95% Confidence Interval	
					Lower	Upper
Healthy behavior 40%	-32.9557	49.013	2.2e-16	119.98	-61.17597	-54.14403
Healthy behavior 20%	-32.9693	49	2.2e-16	120.00	-61.19576	-54.16424

2.1.2 Comparing the fraction of residents infected at peak

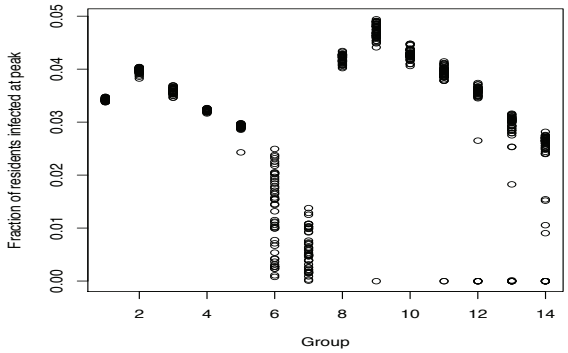


Figure 4: Scatter plots showing the fraction of residents infected at peak verses group where groups are defined as follows: 1 - No interventions (residents only), 2 - Museums closed for 5 days (residents only), 3 - Museums closed for 14 days (residents only), 4 - Healthy behavior 80% (residents only), 5 - Healthy behavior 60% (residents only), 6 - Healthy behavior 40% (residents only), 7 - Healthy behavior 20% (residents only), 8 - No interventions (residents + transients), 9 - Museums closed for 5 days (residents + transients), 10 - Museums closed for 14 days (residents + transients), 11 - Healthy behavior 80% (residents + transients), 12 - Healthy behavior 60% (residents + transients), 13 - Healthy behavior 40% (residents + transients), 14 - Healthy behavior 20% (residents + transients). We remove outliers from each group before performing statistical tests.

Evaluating the effect of transients

Table 12: Independent sample t-test ($\alpha = 0.05$) comparing the fraction of resident infections at peak for residents only and residents + transients scenarios (without any interventions). Significance level (0.000) for the Levene's test for equality of variance is less than $\alpha = 0.05$, which suggests that the variances of the two scenarios are not equal. As the significance level (0.000) for t-test (in the line for "equal variance not assumed") is less than $\alpha = 0.05$, we can conclude that there are significantly more number of residents infected at peaks when the transients are considered.

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig.(2 tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	52.505	0.000	-68.897	98	0.000	-0.0077	0.0001	-0.0080	-0.0075
Equal variances not assumed			-68.897	59.649	0.000	-0.0077	0.0001	-0.0080	-0.0075

Table 13: Levene test of homogeneity of variances (α=0.05) comparing the the fraction of residents infected at peak for following scenarios, for residents + transient population: no intervention, close museums (14 days), healthy behavior 80%, healthy behavior 40%, and healthy behavior 20%. The significance value (0.302) is greater than α (0.05). Hence, the variances are equal for all scenarios and we can proceed towards ANOVA. *Note:* Variances for "close museums (5 days)" and "healthy beahvior 60%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

Levene statistic	df1	df2	Sig.
1.222	4	222	0.302

Table 14: Analysis of variance (ANOVA) (α=0.05) comparing the fraction of residents infected at peak to see if any of the following scenarios for residents + transient population differ: no intervention, close museums (14 days), healthy behavior 80%, healthy behavior 40%, and healthy behavior 20%. The significance value (0.000) is less than α (0.05). Hence, at least one scenario differs from others and we can proceed towards Tukey's HSD test to see which scenarios differ. *Note:* Variances for "close museums (5 days)" and "healthy beahvior 60%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.010	4	0.002	2787.491	0.000
Within Groups	0.000	222	0.000		
Total	0.010	226			

Table 15: Tukey's HSD test ($\alpha = 0.05$) comparing the fraction of residents infected at peak for following scenarios, for residents + transients population: no intervention, close museums (14 days), healthy behavior 80%, healthy behavior 40%, and healthy behavior 20%. Scenarios in the same group are statistically similar to each other and cases in different groups (e.g. "No intervention" is in group 4 and "Healthy behavior 80%" is in group 3) are statistically different from each other. So one is significantly better than the other. *Note:* Variances for "close museums (5 days)" and "healthy beahvior 60%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

Group	N	Subsets for α=0.05				
		1	2	3	4	5
Healthy behavior 20%	39	0.02625249354				
Healthy behavior 40%	40		0.03002136255			
Healthy behavior 80%	48			0.03945132879		
No intervention	50				0.04201985456	
Close museums (14 days)	50					0.04285123094
Sig.		1.000	1.000	1.000	1.000	1.000

Table 16: Welch t-tests ($\alpha = 0.05$, it assumes inequality of variances) comparing the fraction of residents infected at peak for following intervention scenarios to the "no intervention" scenario (with mean = 0.04201985), for residents + transients population: Close museums (5 days) and healthy behavior 60%. In all cases, p-values (2.2e-16) are less than α (0.05). Hence, these scenarios differ significantly from the "no intervention" scenario. Doing multiple t-tests in this fashion can lead to type I error (rejecting null hypothesis when it is actually true) but all p-values obtained here are very small. So we are reasonably confident that these interventions reduces the fraction of residents infected at peak significantly as compared to the "no intervention" scenario.

Scenario	t	df	p-value	mean	95% Confidence Interval	
					Lower	Upper
Close museums (5 days)	-24.3452	76.697	2.2e-16	0.04725868	-0.005667348	-0.004810302
Healthy behavior 60%	41.9429	92	2.2e-16	0.03590205	0.005828117	0.006407501

Evaluating the effect of interventions in the absence of transients

Table 17: Levene test of homogeneity of variances (α=0.05) comparing the the fraction of residents infected at peak for following scenarios, for resident population: no intervention, healthy behavior 80%, and healthy behavior 60%. The significance value (0.391) is greater than α (0.05). Hence, the variances are equal for all scenarios and we can proceed towards ANOVA. *Note:* Variances for "close museums (5 days)", close museums (14 days)", "healthy behavior 40%", and healthy behavior 20%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

Levene statistic	df1	df2	Sig.
0.945	2	146	0.391

Table 18: Analysis of variance (ANOVA) (α=0.05) comparing the fraction of residents infected at peak to see if any of the following scenarios for resident population differ: no intervention, healthy behavior 80%, and healthy behavior 60%. The significance value (0.000) is less than α (0.05). Hence, at least one scenario differs from others and we can proceed towards Tukey's HSD test to see which scenarios differ. *Note:* Variances for "close museums (5 days)", close museums (14 days)", "healthy behavior 40%", and healthy behavior 20%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.001	2	0.000	5047.264	0.000
Within Groups	0.000	146	0.000		
Total	0.001	148			

Table 19: Tukey's HSD test ($\alpha = 0.05$) comparing the fraction of residents infected at peak for following scenarios, for resident population: no intervention, healthy behavior 80%, and healthy behavior 60%. Scenarios in the same group are statistically similar to each other and cases in different groups (i.e., "no intervention" is in group 1 and "healthy beahvior 80%" is in group 2)are statistically different from each other. Hence, one is significantly better than the other. *Note:* Variances for "close museums (5 days)", close museums (14 days)", "healthy behavior 40%", and healthy behavior 20%" scenarios differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

Group	N	Subsets for α=0.05		
		1	2	3
Healthy behavior 60%	49	0.02927613465		
Healthy behavior 80%	50		0.03220738820	
No intervention	50			0.03424248026
Sig.		1.000	1.000	1.000

Table 20: Welch t-tests ($\alpha = 0.05$, it assumes inequality of variances) comparing the fraction of residents infected at peak for following intervention scenarios to the "no intervention" scenario (with mean = 0.03424248), for resident population: Close museums (5 days), close museums (14 days), healthy behavior 40% and healthy behavior 20%. In all cases, p-values (2.2e-16) are less than α (0.05). Hence, these scenarios differ significantly from the "no intervention" scenario. Doing multiple t-tests in this fashion can lead to type I error (rejecting null hypothesis when it is actually true) but all p-values obtained here are very small. So we are reasonably confident that these interventions reduces the fraction of residents infected at peak significantly as compared to the "no intervention" scenario.

Scenario	t	df	p-value	mean	95% Confidence Interval	
					Lower	Upper
Close museums (5 days)	-76.0382	77.12	2.2e-16	0.03975947	-0.005661459	-0.005372514
Close museums (14 days)	-19.7753	68.955	2.2e-16	0.03591980	-0.001846534	-0.001508112
Healthy behavior 40%	20.613	49.117	2.2e-16	0.01302156	0.01915220	0.02328964
Healthy behavior 20%	55.1045	49.469	2.2e-16	0.005872231	0.02733588	0.02940462

2.1.3 Comparing the fraction of residents infected cumulatively

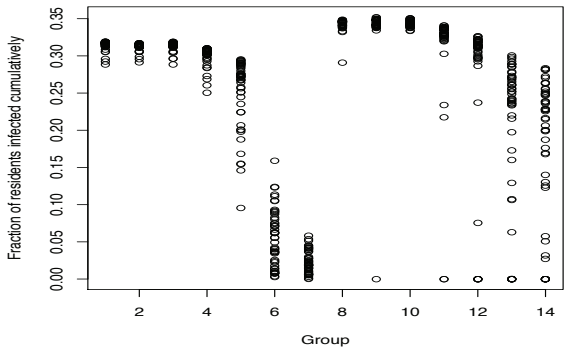


Figure 5: Scatter plots showing the fraction of residents infected cumulatively, and the day of peak verses group where groups are defined as follows: 1 - No interventions (residents only), 2 - Museums closed for 5 days (residents only), 3 - Museums closed for 14 days (residents only), 4 - Healthy behavior 80% (residents only), 5 - Healthy behavior 60% (residents only), 6 - Healthy behavior 40% (residents only), 7 - Healthy behavior 20% (residents only), 8 - No interventions (residents + transients), 9 - Museums closed for 5 days (residents + transients), 10 - Museums closed for 14 days (residents + transients), 11 - Healthy behavior 80% (residents + transients), 12 - Healthy behavior 60% (residents + transients), 13 - Healthy behavior 40% (residents + transients), 14 - Healthy behavior 20% (residents + transients). We remove outliers from each group before performing statistical tests.

Evaluating the effect of transients

Table 21: Independent sample t-test ($\alpha = 0.05$) comparing the fraction of resident infections cumulatively over the course of simulation for residents only and residents + transients scenarios (without any interventions). Significance level (0.228) for the Levene's test for equality of variance is greater than $\alpha = 0.05$, which suggests that the variances of the two scenarios are equal. As the significance level (0.000) for t-test (in the line for equal variance assumed case) is less than $\alpha = 0.05$, we can conclude that there are significantly more number of residents infected cumulatively over the course of simulation when the transients are considered.

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig.(2 tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	1.474	0.228	27.933	97	0.000	0.0294	0.0011	-0.0315	-0.0273
Equal variances not assumed			28.073	79.194	0.000	0.0294	0.0010	-0.0315	-0.0273

Table 22: Levene test of homogeneity of variances (α=0.05) comparing the the fraction of residents infected cumulatively over the course of simulation for following scenarios, for residents + transient population: no intervention, close museums (5 days), and close museums (14 days). The significance value (0.177) is greater than α (0.05). Hence, the variances are equal for all scenarios and we can proceed towards ANOVA. *Note:* Variances for all "healthy beahvior" scenarios (with efficacy 80%, 60%, 40%, and 20%") differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

Levene statistic	df1	df2	Sig.
1.753	2	145	0.177

Table 23: Analysis of variance (ANOVA) (α=0.05) comparing the fraction of residents infected cumulatively over the course of simulation to see if any of the following scenarios for residents + transient population differ: no intervention, close museums (5 days), and close museums (14 days). The significance value (0.445) is greater than α (0.05). Hence, there is not a significant difference between these scenarios. *Note:* Variances for all "healthy beahvior" scenarios (with efficacy 80%, 60%, 40%, and 20%") differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.000	2	0.000	0.815	0.445
Within Groups	0.002	145	0.000		
Total	0.002	147			

Table 24: Welch t-tests ($\alpha = 0.05$, it assumes inequality of variances) comparing the fraction of residents infected cumulatively over the course of simulation for following intervention scenarios to the "no intervention" scenario (with mean = 0.3437282), for residents + transients population: healthy behavior 80%, healthy behavior 60%, healthy behavior 40%, and healthy behavior 20%. In all cases, p-valuse (2.2e-16) are less than α (0.05). Hence, these scenarios differ significantly from the "no intervention" scenario. Doing multiple t-tests in this fashion can lead to type I error (rejecting null hypothesis when it is actually true) but all p-values obtained here are very small. So we are reasonably confident that these interventions reduces the fraction of residents infected cumulatively over the course of simulation significantly as compared to the "no intervention" scenario.

Scenario	t	df	p-value	mean	95% Confidence Interval	
					Lower	Upper
Healthy behavior 80%	12.668	76.323	2.2e-16	0.3313268	0.01045175	0.01435098
Healthy behavior 60%	19.6952	52.856	2.2e-16	0.3128457	0.02773718	0.03402767
Healthy behavior 40%	13.2146	41.421	2.2e-16	0.2452898	0.08339898	0.11347778
Healthy behavior 20%	16.535	38.418	2.2e-16	0.2247314	0.1044331	0.1335604

Table 25: Levene test of homogeneity of variances (α=0.05) comparing the the fraction of residents infected cumulatively over the course of simulation for following scenarios, for resident population: no intervention, close museums (5 days), and close museums (14 days). The significance value (0.660) is greater than α (0.05). Hence, the variances are equal for all scenarios and we can proceed towards ANOVA. *Note:* Variances for all "healthy beahvior" scenarios (with efficacy 80%, 60%, 40%, and 20%") differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

Levene statistic	df1	df2	Sig.
0.417	2	147	0.660

Table 26: Analysis of variance (ANOVA) (α=0.05) comparing the fraction of residents infected cumulatively over the course of simulation to see if any of the following scenarios for resident population differ: no intervention, close museums (5 days), and close museums (14 days). The significance value (0.331) is greater than α (0.05). Hence, there is not a significant difference between these scenarios. *Note:* Variances for all "healthy beahvior" scenarios (with efficacy 80%, 60%, 40%, and 20%") differ from other scenarios. So we compare them with "no intervention" scenario using Welch t-test later.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	0.000	2	0.000	1.177	0.311
Within Groups	0.005	147	0.000		
Total	0.005	149			

Table 27: Welch t-tests ($\alpha = 0.05$, it assumes inequality of variances) comparing the fraction of residents infected cumulatively over the course of simulation for following intervention scenarios to the "no intervention" scenario (with mean = 0.3143110), for resident population: healthy behavior 80%, healthy behavior 60%, healthy behavior 40%, and healthy behavior 20%. In all cases, p-values (¡1.333e-08) are less than α (0.05). Hence, these scenarios differ significantly from the "no intervention" scenario. Doing multiple t-tests in this fashion can lead to type I error (rejecting null hypothesis when it is actually true) but all p-values obtained here are very small. So we are reasonably confident that these interventions reduces the fraction of residents infected cumulatively over the course of simulation significantly as compared to the "no intervention" scenario.

Scenario	t	df	p-value	mean	95% Confidence Interval	
					Lower	Upper
Healthy behavior 80%	6.4221	70.841	1.333e-08	0.3010247	0.009160976	0.017411638
Healthy behavior 60%	9.6737	50.227	4.627e-13	0.2563238	0.04594861	0.07002577
Healthy behavior 40%	50.4469	51.021	2.2e-16	0.05365425	0.2502838	0.2710297
Healthy behavior 20%	124.5587	65.684	2.2e-16	0.02328112	0.2863645	0.2956952

3 Ordinary Differential Equation (ODE) Model

This section describes analysis of each case (effect of transients and intervention scenarios like closing four big museums and promoting healthy behavior at these museums) using ordinary differential equation (ODE) model.

3.1 Effect of transients

This section analyses the effect of transient population on epidemics. Table S28 lists the subscripts used for referring to various sub-populations in the following analysis.

Table 28: Definitions of populations and subpopulations

Population	Definition
r	Resident population
t	Transient population
Subpopulation	Definition
rr	Residents who only meet residents
rt	Residents who meet residents and transients
tt	Transients who only meet transients
tr	Transients who meet transients and residents
Subpopulation	Definition (based on activity at museums)
rr^-	Residents who only meet residents and they do not go to museums
rt^-	Residents who meet residents and transients and they do not go to museums
tr^-	Transients who meet transients and residents and they do not go to museums
rr^+	Residents who only meet residents and they go to museums
rt^+	Residents who meet residents and transients and they go to museums
tr^+	Transients who meet transients and residents and they go to museums

3.1.1 Homogeneously-mixing SEIR model

Two populations are considered in this paper. Resident population represents individuals who live in a given location/city without any birth and death process. Consequently, resident population does not change over time. Transient population represents individuals who stay for a short time in the same location/city as resident population. In particular, a transient individual arrives at the same location/city of the resident population, stays for a certain number of days and leaves the location/city. Every transient arrives to the residency city is assumed to be susceptible and stays on the average for 5 days. Therefore, there exist birth and death processes for the transient population with rate $\rho = 0.2$. Individuals from each population are in contact with individuals from the same population as well as with the other population. To clarify, a resident individual is in contact with resident individuals and with transient individuals. The same is true for transient individuals.

We now assume that a hypothetical Influenza-Like-Illness ILI described by Susceptible/Exposed/Infected/Recovered SEIR compartmental model spreads in the resident population. Due to the existence of contact mixing between resident and transient populations, the spread of ILI can reach the transient population. Let S_r, E_r, I_r, and R_r represent the fraction of susceptbile, exposed, infected/infectious, and recovered resident individuals, respectively. In the same fashion, let S_t, E_t, I_t, and R_t represent the fraction of susceptbile, exposed, infected/infectious, and recovered transient individuals, respectively. We assume that there is an initial small fraction of infected residents. Every infected resident tries to transmit the infection to susceptible residents and transients with infection rates $\beta_{r\to r}$

and $\beta_{t\to r}$, respectively. The infection rate is a function of the contact rate, probability of infection transmission, susceptibility and infectivity as follows [4]:

$$\beta_{i\to j} = \alpha_i \xi_j C_{ij}(1 - e^{T_{ij}p}) \tag{1}$$

where α_i and ξ_i are the susceptibility and infectivity for populations i and j, respectively. $C_{i,j}$ is the contact rate per individual in population i with individuals in population j, $T_{i,j}$ is the average duration per contact between populations i and j and p is the transmissibility value, which is set to 4×10^{-5} transmission per minute. For simplicity, we assume that all individuals have the same level of susceptibility and infectivity, i.e. α_i =1 and ξ_j = 1. The fraction of susceptible individuals who receive the infection become exposed for $\frac{1}{\gamma}$ time units. Fraction of exposed individuals become infected and infectious for $\frac{1}{\mu}$ time units, during which they infect susceptible individuals. After being infected and infectious for $\frac{1}{\mu}$ time units, infected individuals recover without any further infection. Mathematically, the spread of ILI in the two populations is represented using the following system of ordinary differential equations:

$$\frac{dS_r}{dt} = -S_r(\beta_{r\to r} I_r + \beta_{r\to t} I_t) \tag{2a}$$
$$\frac{dE_r}{dt} = S_r(\beta_{r\to r} I_r + \beta_{r\to t} I_t) - \gamma E_r \tag{2b}$$
$$\frac{dI_r}{dt} = \gamma E_r - \mu I_r \tag{2c}$$
$$\frac{dR_r}{dt} = \mu I_r \tag{2d}$$
$$\frac{dS_t}{dt} = -S_t(\beta_{t\to r} I_r + \beta_{t\to t} I_t) + \rho(1 - S_t) \tag{2e}$$
$$\frac{dE_t}{dt} = S_t(\beta_{t\to r} I_r + \beta_{t\to t} I_t) - E_t(\gamma + \rho) \tag{2f}$$
$$\frac{dI_t}{dt} = \gamma E_t - I_t(\mu + \rho) \tag{2g}$$
$$\frac{dR_t}{dt} = \mu I_t - \rho R_t. \tag{2h}$$

Depending on the degree of mixing within the resident population, within transient population and between resident and transient population, the spread can reach a non-negligible fraction of the populations.

Basic Reproductive Number R_o:

The basic reproductive number R_o is defined as the average number of secondary infection caused by a single infected case in a fully susceptible population. The reproductive numbers for the resident population and the transient population can be easily found to be $R_o^r = \frac{\beta_{r\to r}}{\mu_r}$ and $R_o^t = \frac{\gamma\beta_{t\to t}}{(\rho+\gamma)(\rho+\mu)}$, respectively. These reproductive numbers do not reflect the actual reproductive number of the whole system R_o, but they only represent the reproductive number of their populations when the two populations are studied independently. To compute the overall reproductive number, we apply the next generation method [5, 2]. Let F be a matrix with entries f_{ij} representing the rate of appearance of new infection case in state i. In addition, let $V = V^- - V^+$, where V^- is the transfer rate matrix of individuals out of a given state and V^+ is the transfer rate matrix of individuals into a given state. We are only concern about the infection states. Therefore the matrices F and V have 4×4 dimension representing the states E_r, E_t, I_r, I_t. Using the system of differential equations (2a-2h), the matrices F and V are as follows:

$$F = \begin{bmatrix} 0 & 0 & \beta_{r\to r} & \beta_{r\to t} \\ 0 & 0 & \beta_{t\to r} & \beta_{t\to t} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} \gamma & 0 & 0 & 0 \\ 0 & \gamma+\rho & 0 & 0 \\ -\gamma & 0 & \mu & 0 \\ 0 & -\gamma & 0 & \mu+\rho \end{bmatrix}$$

The matrix V is non-singular. The reproductive number R_o is the maximum eigenvalue of the matrix FV^{-1}

$$FV^{-1} = \begin{bmatrix} \frac{\beta_{r\to r}}{\mu} & \frac{\beta_{r\to t}\gamma}{(\gamma+\rho)(\mu+\rho)} & \frac{\beta_{r\to r}}{\mu} & \frac{\beta_{r\to t}}{(\mu+\rho)} \\ \frac{\beta_{t\to r}}{\mu} & \frac{\beta_{t\to t}\gamma}{(\gamma+\rho)(\mu+\rho)} & \frac{\beta_{t\to r}}{\mu} & \frac{\beta_{t\to t}}{(\mu+\rho)} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The reproductive number R_o is found to be as follows:

$$R_o = \frac{\beta_{r\to r}}{2\mu} + \frac{\beta_{t\to t}\gamma}{2(\gamma+\rho)(\mu+\rho)} + \frac{1}{2}[(\frac{\beta_{r\to r}}{\mu})^2 - 2\frac{\beta_{r\to r}}{\mu}\frac{\beta_{t\to t}\gamma}{(\gamma+\rho)(\mu+\rho)} + (\frac{\beta_{t\to t}\gamma}{(\gamma+\rho)(\mu+\rho)})^2 + 4\frac{\beta_{r\to t}\beta_{t\to r}\gamma}{\mu(\gamma+\rho)(\mu+\rho)}]^{\frac{1}{2}} \tag{3}$$

After rearrangement, R_o is found to be:

$$R_o = \frac{R_o^r + R_o^t}{2} + \frac{1}{2}[(R_o^r - R_0^t)^2 + 4\frac{\beta_{t\to r}\beta_{r\to t}\gamma}{\mu(\gamma+\rho)(\mu+\rho)}]^{\frac{1}{2}} \tag{4}$$

which is a function of the individual reproductive numbers for both the resident population and the transient population. The term $\frac{\beta_{t\to r}\beta_{r\to t}\gamma}{\mu(\gamma+\rho)(\mu+\rho)}$ is the competing reproductive number, which represents the average number of secondary infected cases in a susceptible population caused by an infected individual from the other population. To clarify, it represents the average number of secondary infected transients caused by a single infected resident in a fully susceptible transient population and vice versa. Below, we discuss the competing reproductive number R_o^c.

Competing Reproductive Number R_o^c:

Assume that there are contacts between transient and resident population, while there is no contact among individuals in each population. The sysytem of differential equations represents the competing behavior can be obtained by letting $\beta_{r\to r} = \beta_{t\to t} = 0$ in (2). Based on the susceptible, exposed and recovered compartments for each population, the Jacobian matrix for the disease-free equilibrium is as follows:

$$J = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -\beta_{r\to t} \\ 0 & -\gamma & 0 & 0 & 0 & \beta_{r\to t} \\ 0 & \gamma & -\mu & 0 & 0 & 0 \\ 0 & 0 & -\beta_{t\to r} & -\rho & 0 & 0 \\ 0 & 0 & \beta_{t\to r} & 0 & -\gamma-\rho & \beta_{t\to t} \\ 0 & 0 & 0 & 0 & \gamma & -\mu-\rho \end{bmatrix}. \tag{5}$$

There are six eigenvalues for the Jacobian matrix, which are as follows:

$$0 \tag{6a}$$

$$-\rho \tag{6b}$$

$$\frac{-\mu-\gamma-\rho}{2} \pm \frac{1}{2}\Big(\mu^2 - 2\gamma\mu + \gamma^2 + \rho^2 \pm 2((\mu\rho)^2 - 2\mu\gamma\rho^2 + (\gamma\rho)^2 + 4\beta_{r\to t}\beta_{t\to r}\gamma^2)^{\frac{1}{2}}\Big)^{\frac{1}{2}}. \tag{6c}$$

The last four eigenvalues in (6c) represent the stability condition that each eigenvalue is less than 0. After rearrangement, the four eigenvalues lead to the same competing reproductive number as follows:

$$R_o^c = \frac{\beta_{r\to t}\beta_{t\to r}\gamma}{\mu(\mu+\rho)(\gamma+\rho)} < 1. \tag{7}$$

The reproductive number R_o^c can also be obtained by letting R_o^r and R_o^t equal 0 in Eq. 4. Therefore, if $R_o^c < 1$, there is no secondary infection case in the susceptible population that happens due to a single infected individual who has contacts with individuals from the former susceptible population. If $R_o^c > 1$, the epidemic invades the fully susceptible population due to the existence of an infected individual in the other population.

Stability Analysis:

To study the stability analysis, we first formulate the Jacobian matrix for the system of differential equations (2a-2h) based 6 states, namely $S_r, E_r, I_r, S_t, E_t, I_t$ as follows:

$$J = \begin{bmatrix} -\beta_{r\to r}I_r - \beta_{r\to t}I_t & 0 & -\beta_{r\to r}S_r & 0 & 0 & -\beta_{r\to t}S_r \\ \beta_{r\to r}I_r + \beta_{r\to t}I_t & -\gamma & \beta_{r\to r}S_r & 0 & 0 & \beta_{r\to t}S_r \\ 0 & \gamma & -\mu & 0 & 0 & 0 \\ 0 & 0 & -\beta_{t\to r}S_t & -\beta_{t\to t}I_t - \beta_{t\to r}I_r - \rho & 0 & -\beta_{t\to t}S_t \\ 0 & 0 & \beta_{t\to r}S_t & \beta_{t\to t}I_t + \beta_{t\to r}I_r & -\gamma-\rho & \beta_{t\to t}S_t \\ 0 & 0 & 0 & 0 & \gamma & -\mu-\rho \end{bmatrix}. \tag{8}$$

The system has three equilibrium points $S_r, E_r, I_r, R_r, S_t, E_t, I_t, R_t$ representing the diease-free point $P1 : (1, 0, 0, 0, 1, 0, 0, 0)$, non-endemic disease point $P2 : (S_r^*, 0, 0, R_r^*, 1, 0, 0, 0)$, and the transient endemic point $P3 : (0, 0, 0, 1, S_t^*, E_t^*, I_t^*, R_t^*)$, respectively. Below, we address each equilibrium point in detail.

Disease-free equilibrium $P1$ point

The point $P1$ represents the disease-free equilibrium point that the initial infected cases die out without causing any new infection cases. Therefore, both resident and transient populations are susceptible. The reproductive number of the system is less than one, $R_o < 1$. The following theorem addresses the stability of the system at point $P1$.

Theorem 1 *Consider the disease model for the resident and transient populations being in equilibrium point $P1$. If $R_o < 1$ the equilibrium point $P1$ is locally asymptotically stable, while if $R_o > 1$ the equilibrium point $P1$ is unstable.*

Proof. Consider the following matrix properties: 1) F is nonnegative matrix, 2) the eigenvalues of V are the diagonal elements, 3) V is non singular M matrix where all eigenvalues are positive, and 4) V has Z sign pattern property since all elements $v_{ij} \leq 0 \, \forall \, i \neq j$. Let $J_1 = F - V$ be the matrix representing the system of differential equation describing the infection states E_r, E_t, I_r, I_t. Using the properties of F and V matrices, $-J_1 = V - F$ has the Z sign pattern. In addition, the matrix $-J_1 V^{-1} = I - FV^{-1}$ has the Z sign pattern because FV^{-1} is nonnegative matrix.
Let the matrix $I - FV^{-1}$ be non singular M matrix. It follows that maximum eigenvalue of FV^{-1} is less than 1. Since both $-J_1$ and $-J_1 V^{-1}$ have Z sign pattern and V^{-1} is a lower triangular with positive eigenvalues (V^{-1} is nonsingular M matrix), the matrix $-J_1$ is non singular M matrix [3, 5]. Therefore, the maximum eigenvalue of J_1 is less than 0 if and only if $-J_1 V^{-1}$ is nonsingular M matrix if and only if the leading eigenvalue of FV^{-1} is $R_o < 1$.
Now, let the matrix $I - FV^{-1}$ be singular M matrix with 0 leading eigenvalue, which implies $-J_1$ is singular M matrix with leading eigenvalue equals 0 if and only if the leading eigenvalue of FV^{-1} $R_o = 1$. It also follows that the leading eigenvalue of J_1 is greater than 0 if and only if $R_o > 1$. ■

Non-endemic disease equilibrium $P2$ and transient endemic equilibrium $P3$ points

The second equilibrium point $P2$ represents the disease invasion in both populations. At equilibrium, resident population are divided into susceptible and recovered states. Although the disease eventually reach the transient population whenever $R_o > 1$, transient population becomes suspetible at equilibrium. To clarify, there is a birth and death processes in transient population disease model. In such process, the transient individuals in each disease state are removed from the population and replaced with susceptible transients at rate ρ. Therefore, at equilibrium, there are only susceptible transient populations. The process is conditioned by the reproductive number for transient population R_o^t as shown in the following theorem.

Theorem 2 *Consider the disease model for the resident and transient populations being at equilibrium point $P2$ with $S_r^* = 0$ and $R_r^* = 1$. The point $P2$ is unstable if both $R_o > 1$ and $R_o^t > 1$, and the transient disease model has endemic equilibrium point $P3$.*

Proof. Let $S_r^* = 0$ and $R_r^* = 1$ and the point $P2$ becomes $P2 : (0, 0, 0, 1, 1, 0, 0, 0)$. Substitute $P2$ in the Jacobian matrix 8 considering six variables $(S_r, E_r, I_r, S_t, E_t, I_t)$=$(0, 0, 0, 1, 0, 0, 0)$, the matrix becomes as follows:

$$J(P2) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\gamma & 0 & 0 & 0 & 0 \\ 0 & \gamma & -\mu & 0 & 0 & 0 \\ 0 & 0 & -\beta_{t\to r} & -\rho & 0 & -\beta_{t\to t} \\ 0 & 0 & \beta_{t\to r} & 0 & -\gamma-\rho & \beta_{t\to t} \\ 0 & 0 & 0 & 0 & \gamma & -\mu-\rho \end{bmatrix}. \tag{9}$$

The matrix $J(P2)$ is structured such that its overall eigenvalues are the eigenvalues of the diagonal blocks as shown in 9. Thus, the eigenvalues are 0, $-\gamma$, $-\mu$, $-\rho$, and $-\frac{\gamma+\mu+2\rho}{2} \pm \frac{1}{2}\sqrt{(\gamma-\mu)^2 + 4\gamma\beta_{t\to t}}$. The disease model is asymptotically stable at the point $P2$, if and only if all the eigenvalues of $J(P2)$ are negative. Note that the zero eigenvalue is due to the existence of a raw and a column with zero entries. Therefore, the stability condition is as follows:

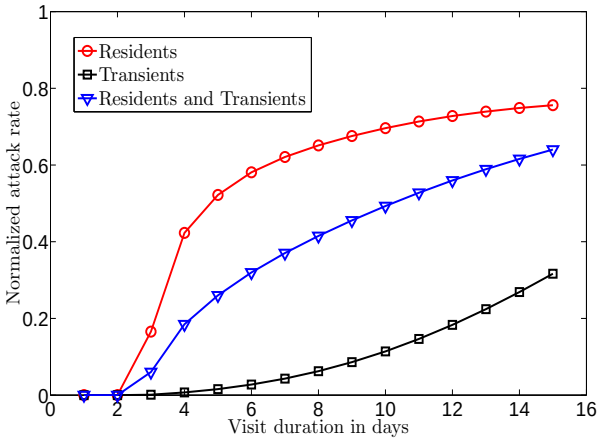


Figure 6: Evaluation of the nomralized attack rates of residents, transients and both residents and transients for a range of visit duration from 1 day until 15 days.

$$-\frac{\gamma+\mu+2\rho}{2}+\frac{1}{2}\sqrt{(\gamma-\mu)^2+4\gamma\beta_{t\to t}}<0. \tag{10}$$

After rearrangement, the stability condition becomes as follows:

$$\frac{\gamma\beta_{t\to t}}{(\rho+\gamma)(\rho+\mu)}<1. \tag{11}$$

The left-hand-side of the inequality is the reproductive number of the disease transmission within the transient population R_o^t. ■

The fraction of transient individuals in each state at endemic equilibrium $P3$ are as follows:

$$S_t^* = \frac{1}{R_o^t} \tag{12}$$

$$E_t^* = \frac{\rho}{(\gamma+\rho)}(1-\frac{1}{R_o^t}) \tag{13}$$

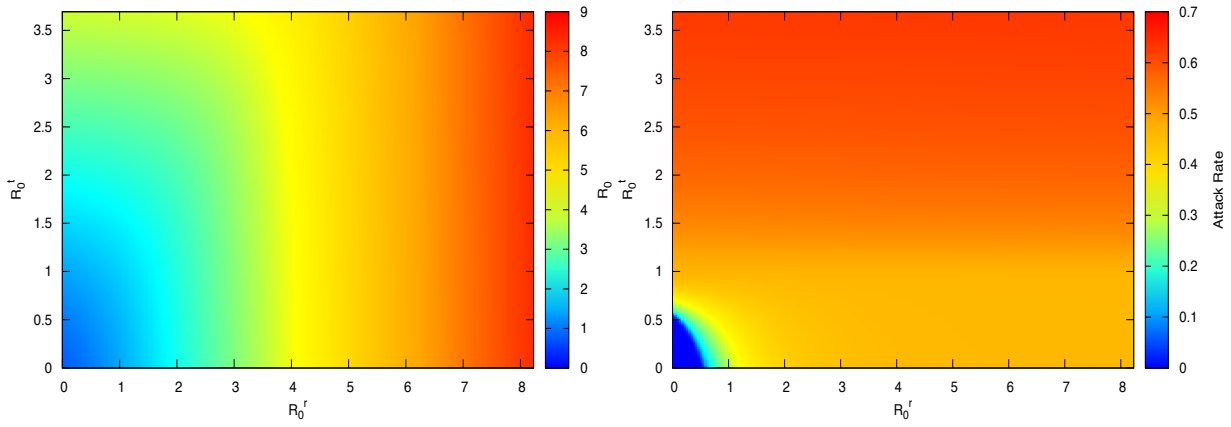
$$I_t^* = \frac{\rho R_o^t}{\beta_{t\to t}}(1-\frac{1}{R_o^t}) \tag{14}$$

$$R_t^* = \frac{\mu R_o^t}{\beta_{t\to t}}(1-\frac{1}{R_o^t}) \tag{15}$$

Sensitivity of the overall reproductive number and the attack rate with respect to the visit duration and the individual reproductive numbers

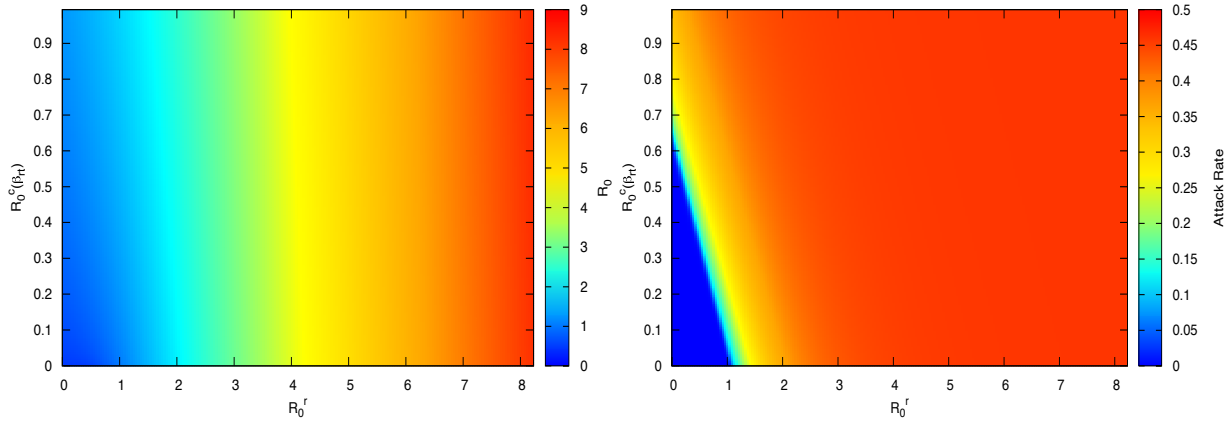
We study the effect of visit duration of transients on the attack rate as shown in Figure S6. A large change in the resident attack rate and the overall attack rate take place when the visit duration increases from 2 days to 4 days. The transient attack rate increases nonlinearly as a convex function with the visit duration showing that the transient attack rate is less sensitive to the increase of visit duration than the resident attack rate.

We also study the sensitivity of the overall reproductive number R_o and the total attack rate with respect to the individual reproductive numbers R_o^r, R_o^t and R_o^c. Because the competing reproductive number R_o^c is a function of two



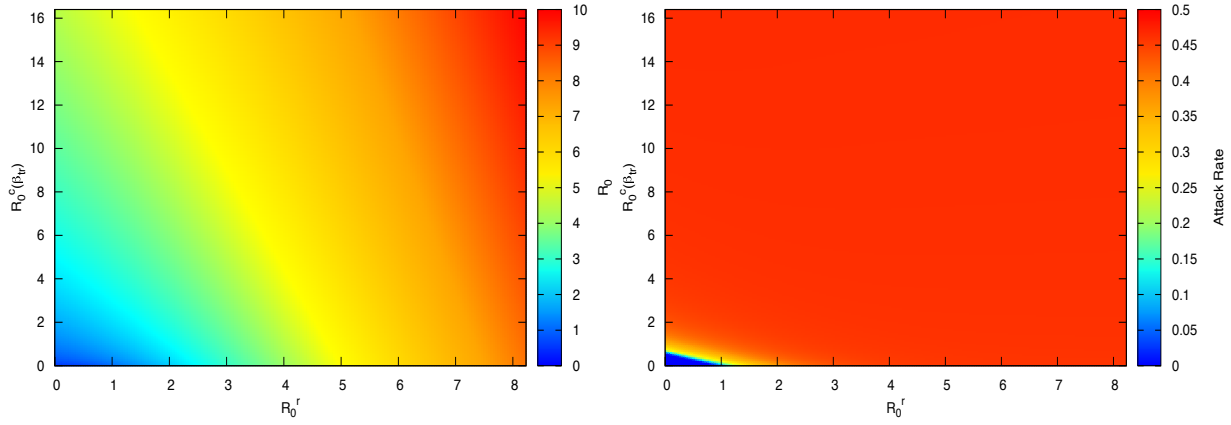
(a) Evaluation of R_o as a function of R_o^r and R_o^t

(b) Evaluation of attack rate as a function of R_o^r and R_o^t



(c) Evaluation of R_o as a function of R_o^r and $R_o^c(\beta_{r\to t})$

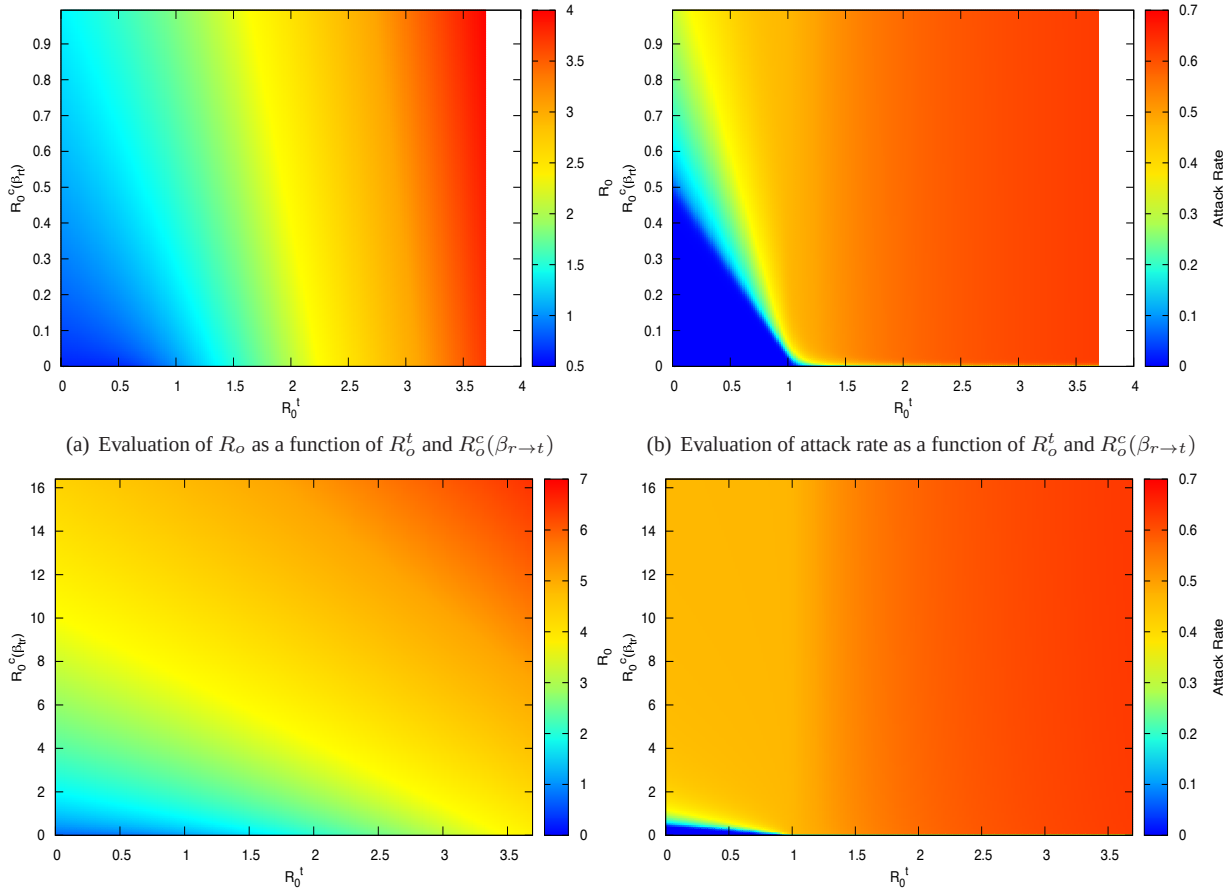
(d) Evaluation of attack rate as a function of R_o^r and $R_o^c(\beta_{r\to t})$



(e) Evaluation of R_o as a function of R_o^r and $R_o^c(\beta_{t\to r})$

(f) Evaluation of attack rate as a function of R_o^r and $R_o^c(\beta_{t\to r})$

Figure 7: Evaluation of the reproductive number R_o and the attack rate as a function of resident reproductive number R_o^r, and R_o^t, $R_o^c(\beta_{r\to t})$, and $R_o^c(\beta_{t\to r})$.



(a) Evaluation of R_o as a function of R_o^t and $R_o^c(\beta_{r\to t})$

(b) Evaluation of attack rate as a function of R_o^t and $R_o^c(\beta_{r\to t})$

(c) Evaluation of R_o as a function of R_o^t and $R_o^c(\beta_{t\to r})$

(d) Evaluation of attack rate as a function of R_o^t and $R_o^c(\beta_{t\to r})$

Figure 8: Evaluation of the reproductive number R_o and the attack rate as a function of transient reproductive number R_o^t, and $R_o^c(\beta_{r\to t})$, and $R_o^c(\beta_{t\to r})$.

infection transmission rates $\beta_{r\to t}$ and $\beta_{t\to r}$, there are four different infection rates to be considered. We evaluate the overall reproductive number and the attack rate as a function of two infection rates, while fixing the other two infection rates at their original values. As shown in Figures S7(a) and S7(b), the reproductive number and the attack rate are evaluated when the infection rates $\beta_{r\to t}$ and $\beta_{t\to r}$ are fixed and hence R_o^c is fixed, while the infection rates within resident population $\beta_{r\to r}$ and within transient population $\beta_{t\to t}$ are varied. Consequently, their reproductive numbers R_o^r and R_o^t are varied. The figures show that even the reproductive numbers R_o^r and R_o^t are less than 1, the overall reproductive number can be greater than 1 and the epidemic spreads in the two populations. This observation complies with the non-endemic disease equilibrium point $P2$ where there is no endemic equilibrium for the transient population. The endemic equilibrium point $P3$ is observed for $R_o^t > 1$ where the attack rate becomes high 0.45. Also the two figures show that the transient reproductive number changes slower than the resident reproductive number when their infection rates are changed similarly. Figures S7(c) and S7(d) show the evaluation of the overall reproductive and the attack rate when the resident infection rate varies and so the resident reproductive number, and the infection rate between residents and transients varies and so the competing reproductive number, while the transient infection rate and the infection rate between transients and residents are fixed. The reproductive number $R_o^c(\beta_{r\to t})$ changes slower than R_o^r. Also, for very small values of $R_o^c(\beta_{r\to t})$, the epidemic spreads between the two populations leading to a large attack rate. Figures S7(e) and S7(f) show the evaluation of the reproductive number and the attack rate when $\beta_{r\to r}$ and $\beta_{t\to r}$ vary and so the reproductive numbers R_o^r and $R_o^c(\beta_{t\to r})$. The competing reproductive number vary faster than the resident reproductive number showing that any small change in $R_o^c(\beta_{t\to r})$ leads to learge attack rate. If we assume that $R_o > 1$, then Figures S7(d) and S7(f) show that the system can easily become epidemic free ($R_o < 1$) if the infection rate $\beta_{r\to t}$ is slightly reduced by reducing the contact rates and/or the duration per contact of residents with transients. On the other hand, a slight reduction in the infection rate from transients to residents may not reduce R_o to be less than 1.

3.1.2 Effective residents and transients populations

The assumption that every resident has a certain number of contacts with transients is vague and misleading. Actually, not every resident has contacts with transients. To shade light on this argument, we decompose the resident and transient populations to four subpopulations. The first subpopulation represents residents who only have contacts with other residents. The second subpopulation represents residents who have contacts with other residents and transients. The third subpopulation represents transients who have contacts with other transients and residents. The last subpopulation represents transients who only have contacts with other transients. We denote the subpopulations as rr, rt, tr, and tt, respectively. The four subpopulations are shown in Figure S9, where the ellipse shape represents a subpopulation and the arrow highlights the contacts between subpopulations. Using our synthetic social network, we extract the total number of contacts and the total contact duration between individuals who belong to two subpopulations. The average number of contact of an individual in subpopulation rr with other individuals in the same subpopulation equals the total number of contacts among individuals in subpopulation rr divided by number of individuals in subpopulation rr. The average number of contacts of an individual in subpopulation rr with other individuals in subpopulation rt equals the total number of contacts between the two subpopulations divided by number of individuals in subpopulation rr. The contact rates of every individual with individuals in the four subpopulations and the duration per contact are reported in Table S29 for the non intervention scenario (left). Obviously, there is no contact between residents in subpopulation rr and transients in subpopulations tr and tt. The same is true for transients in subpopulation tt and residents in subpopulations rr and rt. In addition, we notice that every transient has contacts with other transients and residents because there is no contact rate between transients in subpopulation tt with the four subpopulations.

To study the spread of infectious diseases among the subpopulations, we develop and ODE system that accounts for each state for every subpopulation as shown in Eq.(16a-16d,16e-16h, 16m-16p,16i-16l). We denote the fraction of susceptible, exposed, infected and recovered individuals in subpopulation rr as S_{rr}, E_{rr}, I_{rr} and R_{rr}, respectively. The fraction of individuals in each state in each subpopulation is denoted in the same way. In addition, there are ten different infection rates between subpopulations as shown in Figure S9. Each infection rate is denoted as $\beta_{wx\to yz}$ where wx represents individuals in subpopulation wx and have contacts with individuals in subpopulation yz. For example, the infection rate of subpopulation rr due to contacts in infected individuals in subpopulation rt is denoted as $\beta_{rr\to rt}$.

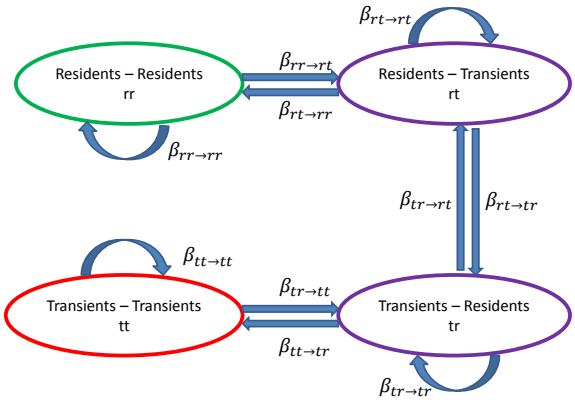


Figure 9: Contact pattern among four subpopulations. In general, $\beta_{ab\to cd}$ represents the infection transmission rate due to the contact between subpopulation ab to subpopulation cd. The infection rates $\beta_{rr\to rr}$, $\beta_{rr\to rt}$, $\beta_{rt\to rr}$, $\beta_{rt\to rt}$, $\beta_{rt\to tr}$, $\beta_{tr\to rt}$ and $\beta_{tr\to tr}$ have positive values, while the infection rates $\beta_{tt\to tt}$, $\beta_{tt\to tr}$ and $\beta_{tr\to tt}$ equal 0 because all transients have contacts with transients and residents resulting in the population tt equals 0.

$$\frac{dS_{rr}}{dt} = -S_{rr}(\beta_{rr\to rr}I_{rr} + \beta_{rr\to rt}I_{rt}) \tag{16a}$$
$$\frac{dE_{rr}}{dt} = S_{rr}(\beta_{rr\to rr}I_{rr} + \beta_{rr\to rt}I_{rt}) - \gamma E_{rr} \tag{16b}$$
$$\frac{dI_{rr}}{dt} = \gamma E_{rr} - \mu I_{rr} \tag{16c}$$
$$\frac{dR_{rr}}{dt} = \mu I_{rr} \tag{16d}$$
$$\frac{dS_{rt}}{dt} = -S_{rt}(\beta_{rt\to rt}I_{rt} + \beta_{rt\to rr}I_{rr} + \beta_{rt\to tr}I_{tr}) \tag{16e}$$
$$\frac{dE_{rt}}{dt} = S_{rt}(\beta_{rt\to rt}I_{rt} + \beta_{rt\to rr}I_{rr} + \beta_{rt\to tr}I_{tr}) - \gamma E_{rt} \tag{16f}$$
$$\frac{dI_{rt}}{dt} = \gamma E_{rt} - \mu I_{rt} \tag{16g}$$
$$\frac{dR_{rt}}{dt} = \mu I_{rt} \tag{16h}$$
$$\frac{dS_{tt}}{dt} = -S_{tt}(\beta_{tt\to tt}I_{tt} + \beta_{tt\to tr}I_{tr}) + \rho(1 - S_{tt}) \tag{16i}$$
$$\frac{dE_{tt}}{dt} = S_{tt}(\beta_{tt\to tt}I_{tt} + \beta_{tt\to tr}I_{tr}) - E_{tt}(\gamma + \rho) \tag{16j}$$
$$\frac{dI_{tt}}{dt} = \gamma E_{tt} - I_{tt}(\mu + \rho) \tag{16k}$$
$$\frac{dR_{tt}}{dt} = \mu I_{tt} - \rho R_{tt} \tag{16l}$$
$$\frac{dS_{tr}}{dt} = -S_{tr}(\beta_{tr\to tr}I_{tr} + \beta_{tr\to rt}I_{rt} + \beta_{tr\to tt}I_{tt}) + \rho(1 - S_{tr}) \tag{16m}$$
$$\frac{dE_{tr}}{dt} = S_{tr}(\beta_{tr\to tr}I_{tr} + \beta_{tr\to rt}I_{rt} + \beta_{tr\to tt}I_{tt}) - E_{tr}(\gamma + \rho) \tag{16n}$$
$$\frac{dI_{tr}}{dt} = \gamma E_{tr} - I_{tr}(\mu + \rho) \tag{16o}$$
$$\frac{dR_{tr}}{dt} = \mu I_{tr} - \rho R_{tr}. \tag{16p}$$

Using the next generation method, we obtain the F, V and FV^{-1} matrices as follows:

$$F = \begin{bmatrix} 0 & 0 & 0 & \beta_{rr\to rr} & \beta_{rr\to rt} & 0 \\ 0 & 0 & 0 & \beta_{rt\to rr} & \beta_{rt\to rt} & \beta_{rt\to tr} \\ 0 & 0 & 0 & 0 & \beta_{tr\to rt} & \beta_{tr\to tr} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{17}$$

$$V = \begin{bmatrix} \gamma & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma+\rho & 0 & 0 & 0 \\ -\gamma & 0 & 0 & \mu & 0 & 0 \\ 0 & -\gamma & 0 & 0 & \mu & 0 \\ 0 & 0 & -\gamma & 0 & 0 & \rho+\mu \end{bmatrix}. \tag{18}$$

The matrix FV^{-1}

$$FV^{-1} = \begin{bmatrix} \frac{\beta_{rr\to rr}}{\mu} & \frac{\beta_{rr\to rt}}{\mu} & 0 & \frac{\beta_{rr\to rr}}{\mu} & \frac{\beta_{rr\to rt}}{\mu} & 0 \\ \frac{\beta_{rt\to rr}}{\mu} & \frac{\beta_{rt\to rt}}{\mu} & \frac{\beta_{rt\to tr}\gamma}{(\gamma+\rho)(\mu+\rho)} & \frac{\beta_{rt\to rr}}{\mu} & \frac{\beta_{rt\to rt}}{\mu} & \frac{\beta_{rt\to tr}}{(mu+rho)} \\ 0 & \frac{\beta_{tr\to rt}}{\mu} & \frac{\beta_{tr\to tr}\gamma}{(\gamma+\rho)(\mu+\rho)} & 0 & \frac{\beta_{tr\to rt}}{\mu} & \frac{\beta_{tr\to tr}}{\mu+\rho} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{19}$$

The reproductive number R_o is the largest eigenvalue of the matrix FV^{-1} as follows:

$$Ro = \frac{R_{rr\leftrightarrow rr} + R_{rt\leftrightarrow rt} + R_{tr\leftrightarrow tr}}{3} + \frac{Y}{X} + X \tag{20}$$

where

$$X = \left(\left(\frac{R_{rr\leftrightarrow rr}+R_{rt\leftrightarrow rt}+R_{tr\leftrightarrow tr}}{3} \right)^3 + \right. \tag{21}$$

$$\left(\left(\frac{R_{rr\leftrightarrow rr}R_{rt\leftrightarrow tr}}{2} - \left(\frac{R_{rr\leftrightarrow rr}+R_{rt\leftrightarrow rt}+R_{tr\leftrightarrow tr}}{3} \right)^3 + \frac{R_{rr\leftrightarrow rt}R_{tr\leftrightarrow tr}}{2} - \frac{R_{rr\leftrightarrow rr}R_{rt\leftrightarrow rt}R_{tr\leftrightarrow tr}}{2} + Z \right)^2 - Y^3 \right)^{\frac{1}{2}} -$$

$$\left. \frac{R_{rr\leftrightarrow rr}R_{rt\leftrightarrow tr}}{2} - \frac{R_{rr\leftrightarrow rt}R_{tr\leftrightarrow tr}}{2} + \frac{R_{rr\leftrightarrow rr}R_{rt\leftrightarrow rt}R_{tr\leftrightarrow tr}}{2} - Z \right)^{\frac{1}{3}}$$

$$Y = \left(\frac{R_{rr\leftrightarrow rr} + R_{rt\leftrightarrow rt} + R_{tr\leftrightarrow tr}}{3} \right)^2 - \frac{\left(R_{rr\leftrightarrow rr}R_{rt\leftrightarrow rt} + R_{rr\leftrightarrow rr}R_{tr\leftrightarrow tr} + R_{rt\leftrightarrow rt}R_{tr\leftrightarrow tr} \right)}{3} + \frac{\left(R_{rr\leftrightarrow rt} + R_{rt\leftrightarrow tr} \right)}{3}$$

$$Z = \frac{\left(R_{rr\leftrightarrow rr} + R_{rt\leftrightarrow rt} + R_{tr\leftrightarrow tr} \right)}{2} \frac{\left(R_{rr\leftrightarrow rr}R_{rt\leftrightarrow rt} - R_{rt\leftrightarrow tr} - R_{rr\leftrightarrow rt} + R_{rr\leftrightarrow rr}R_{tr\leftrightarrow tr} + R_{rt\leftrightarrow rt}R_{tr\leftrightarrow tr} \right)}{3}$$

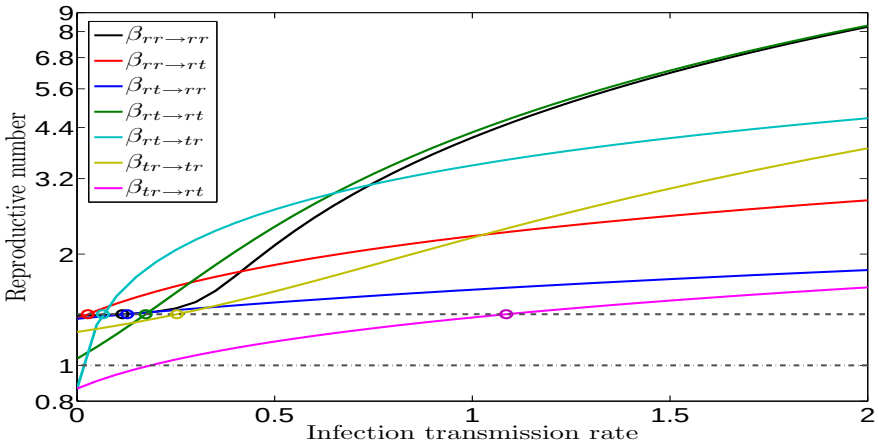


Figure 10: Evaluation of reproductive number as a function of the infection transmission rates. The circles represent the estimated infection transmission rate values based on the synthetic social network. The thin dash line represents the value of reproductive number $Ro = 1.375$ (Eq. 20), while the thick dash line represents reproductive number $Ro = 1$ below which the epidemic dies out. For every infection transmission rate, we sweep the transmission rate value between 0 and 2 and we evaluate the reproductive number R_o using Eq. 20.

$$
\begin{aligned}
R_{rr\leftrightarrow rr} &= \frac{\beta_{rr\to rr}}{\mu} \\
R_{rr\leftrightarrow rt} &= \frac{\beta_{rr\to rt}\beta_{rt\to rr}}{\mu^2} \\
R_{rt\leftrightarrow rt} &= \frac{\beta_{rt\to rt}}{\mu} \\
R_{rt\leftrightarrow tr} &= \frac{\beta_{rt\to tr}\beta_{tr\to rt}\gamma}{\mu(\gamma+\rho)(\mu+\rho)} \\
R_{tr\leftrightarrow tr} &= \frac{\beta_{tr\to tr}\gamma}{(\gamma+\rho)(\mu+\rho)}
\end{aligned}
\tag{22}
$$

where $R_{rr\leftrightarrow rr}$ is the reproductive number for the resident subpopulation that only have contacts with residents rr, $R_{rr\leftrightarrow rt}$ is the reproductive number between resident subpopulation that only have contacts with residents rr and resident subpopulation that have contacts with both residents and transients rt, $R_{rt\leftrightarrow rt}$ is the reproductive number of the resident subpopulation that have contacts with both residents and transients rt, $R_{rt\leftrightarrow tr}$ is the reproductive number between resident and transients subpopulations that have contacts with both residents and transients (rt and tr) and $R_{tr\leftrightarrow tr}$ is the reproductive number of transient subpopulation that have contacts with both transients and residents tr.

3.2 Interventions

3.2.1 Closing museums

For closing museums intervention, in Table S29 we compare the average contact rates and duration per contact among the four subpopulations (rr, rt, tt and tr) for the non intervention scenario and closing museums scenario.

Table S30 shows the infection transmission rates among the four subpopulations given non intervention scenario and closing museums intervention. Figure 11 shows the attack rate for the two cases: 1) museums are opened representing the non-intervention scenario and 2) closed museums representing the intervention scenario. The figure shows that the closing-museums intervention does not decrease the final number of infected individuals for both residents and transients. This conclusion is in agreement with the conclusion drawn from the agent-based model.

Table 29: The number of contacts per day (upper table) and the duration per contact in hours (lower table) among four subpopulations without any intervention (left) and for closed museums (right)

Subpopulation	Without intervention				Closed museums			
	rr	rt	tt	tr	rr	rt	tt	tr
rr	23.6	8.5	0	0	23.3	8.3	0	0
rt	39.2	413.1	0	242	38.5	52.7	0	23.5
tt	0	0	0	0	0	0	0	0
tr	0	4010.9	0	719.1	0	388.3	0	71
rr	2.04	1.35	0	0	2.05	1.36	0	0
rt	1.35	0.17	0	0.11	1.34	0.17	0	0.11
tt	0	0	0	0	0	0	0	0
tr	0	0.11	0	0.15	0	0.85	0	1.14

Table 30: Infection rates for non intervention scenario and closing museums intervention

Subpopulation	Without intervention				Closing museums			
	rr	rt	tt	tr	rr	rt	tt	tr
rr	0.1150	0.0274	0	0	0.1147	0.0270	0	0
rt	0.1268	0.1737	0	0.0655	0.1245	0.0222	0	0.0064
tt	0	0	0	0	0	0	0	0
tr	0	1.0857	0	0.2526	0	0.7910	0	0.1944

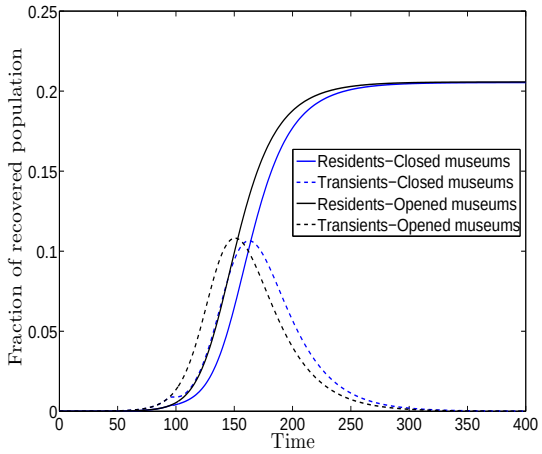


Figure 11: Fraction of removed residents and transients after contracting the infection when the museums are opened (non intervention) and the museums are closed (social distancing intervention).

Symbol	Definition	Count
rr^+	Residents who meet only residents and who go to museums	0
rr^-	Residents who meet residents and who do not go to museums	3400201
rt^+	Residents who meet residents and transients and who go to museums	190124
rt^-	Residents who meet residents and transients and who do not go to museums	544236
tr^+	Transients who meet residents and transients and who go to museums	24450
tr^-	Transients who meet residents and transients and who do not go to museums	19859

Table 31: Definitions for healthy behavior subpopulations

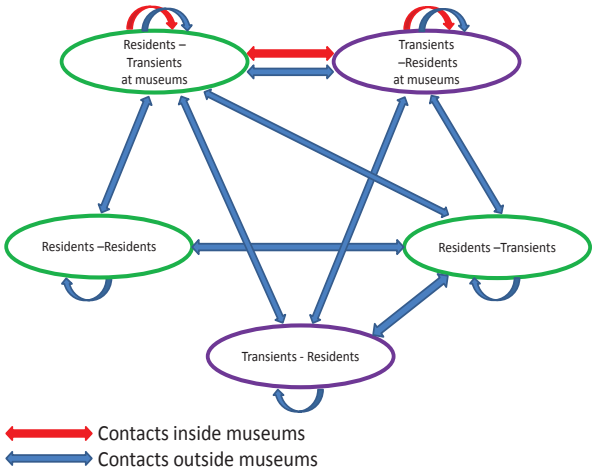


Figure 12: Resident and transient populations are divided based on museum visit for healthy behavior intervention. The $Residents - Transients\ at\ museums$ subpopulation represents residents who visit museums and they meet residents and transients. Similarly, the $Transients - Residents\ at\ museums$ subpopulation represents transients who visit museums and they meet transients and residents. These two subpopulations are denoted as rt^+ and tr^+ and they have contacts inside the museums (red) and outside the museums (blue). The other three subpopulations (rr^-, rt^- and tr^-) represent subpopulations of individuals who do not visit museums.

3.2.2 Healthy behavior

We can also see museums where a lot of mixing happens as a places where we can promote healthy behavior and hence reduce the number of infections that happen within museums. Hence, we evaluate a scenario where people are encouraged to practice healthy behavior (like using hand sanitizer or covering cough) at the four big museums. As data about how much infectivity and susceptibility are reduced by application of healthy behavior is unavailable, we did a series of experiments assuming that the healthy behavior reduces infectivity and susceptibility to 20%, 40%, 60%, and 80% of the original values (effective only inside the four museums). We assume that 50% of the people going to these engage in healthy behavior.

As healthy behavior intervention is assumed to be effective only inside the museums, we further divide each of the supopulation used for the "close museum" case into that going to museums and not going to museums. Subpopulations obtained are listed with the number of people in each category in Table S31 and resulting contact pattern is shown in Figure S12 and Table S32. The infection transmission rates outside museums, inside museums among individuals who do not comply with the intervention and inside museums among individuals who comply with the intervention are shown in Tables S33, S34 and S35, respectively.

$$\frac{\mathbf{dS}}{\mathbf{dt}} = [diag(-S)]([\beta]\mathbf{I}) + \rho(\mathbf{U} - \mathbf{S}) \tag{23}$$

$$\frac{\mathbf{dE}}{\mathbf{dt}} = [diag(S)]([\beta]\mathbf{I}) - \gamma\mathbf{E} - \rho\mathbf{E} \tag{24}$$

$$\frac{\mathbf{dI}}{\mathbf{dt}} = \gamma\mathbf{E} - \rho\mathbf{I} - \mu\mathbf{I} \tag{25}$$

$$\frac{\mathbf{dR}}{\mathbf{dt}} = \mu\mathbf{I} - \rho\mathbf{R} \tag{26}$$

where $\frac{\mathbf{dS}}{\mathbf{dt}} = [\frac{dS_{rr^-}}{dt}\ \frac{dS_{rt^-}}{dt}\ \frac{dS_{tr^-}}{dt}\ \frac{dS_{rt^+}}{dt}\ \frac{dS_{tr^+}}{dt}]^T$, $\mathbf{I} = [I_{rr^-}\ I_{rt^-}\ \dots I_{tr^+}]^T$, ρ is a diagonal matrix with diagonal elements $\{0\ 0\ \rho\ 0\ \rho\}$, $\mathbf{U} = [1\ 1\ \dots 1]^T$ and $[\beta]$ represents the transmission rate matrix among the subpopulations as follows

$$\begin{array}{ccccccc}
\beta_{rr^-\to rr^-} & \beta_{rr^-\to rt^-} & 0 & \beta_{rr^-\to rt^+} & 0 \\
\beta_{rt^-\to rr^-} & \beta_{rt^-\to rt^-} & \beta_{rt^-\to tr^-} & \beta_{rt^-\to rt^+} & \beta_{rt^-\to tr^+} \\
0 & \beta_{tr^-\to rt^-} & \beta_{tr^-\to tr^-} & \beta_{tr^-\to rt^+} & \beta_{tr^-\to tr^+} \\
\beta_{rt^+\to rr^-} & \beta_{rt^+\to rt^-} & \beta_{rt^+\to tr^-} & \frac{1}{2}(\beta_{rt^+\to rt^+} + \beta_{rt^+\to rt^{+\prime}})_m + \beta_{rt^+\to rt^+}|_{Om} & \frac{1}{2}(\beta_{rt^+\to tr^+} + \beta_{rt^+\to tr^{+\prime}})_m + \beta_{rt^+\to tr^+}|_{Om} \\
0 & \beta_{tr^+\to rt^-} & \beta_{tr^+\to tr^-} & \frac{1}{2}(\beta_{tr^+\to rt^+} + \beta_{tr^+\to rt^{+\prime}})_m + \beta_{tr^+\to rt^+}|_{Om} & \frac{1}{2}(\beta_{tr^+\to tr^+} + \beta_{tr^+\to tr^{+\prime}})_m + \beta_{tr^+\to tr^+}|_{Om}
\end{array} \tag{27}$$

where β'_m is the reduced infection rate due to engaging in healthy behavior inside museums, β_m is the infection rate inside museums without the practice of healthy behavior and $\beta|_{Om}$ is the infection rate outside museums.

Table 32: The number of contacts per day (upper table) and the duration per contact in hours (lower table) among the subpopulations outside the museums (left) and inside the museums (right) for healthy behavior case

Subpopulation	Outside museums					Inside museums	
	rr^-	rt^+	rt^-	tr^+	tr^-	rt^+	tr^+
rr^-	23.5519	2.7365	5.7353	0	0		
rt^+	48.9407	1.6189	9.4496	0.1510	0.3170	1547.2	922.611
rt^-	35.8319	3.3011	13.1099	1.8263	2.2491		
tr^+	0	1.1740	40.6526	22.4098	14.2144	7174.3	1260
tr^-	0	3.0345	61.6374	17.5004	8.1431		
rr^-	2.0409	1.0828	1.4771	0	0		
rt^+	1.0828	0.9739	1.2873	1.4621	1.0619	0.1008	0.0984
rt^-	1.4771	1.2873	2.8793	0.7583	1.5581		
tr^+	0	1.4621	0.7583	1.9681	0.8888	0.0984	0.1008
tr^-	0	1.0619	1.5581	0.8888	1.2636		

References

[1] C. Barrett, R. Beckman, K. Berkbigler, K. Bisset, B. Bush, K. Campbell, S. Eubank, K. Henson, J. Hurford, D. Kubicek, M. Marathe, P. Romero, J. Smith, L. Smith, P. Speckman, P. Stretz, G. Thayer, E. Eeckhout, and

Table 33: Infection rates outside museums among subpopulations with individuals who go and do not go to museums

	rr^-	rt^+	rt^-	tr^+	tr^-
rr^-	0.1151	0.0071	0.0203	0	0
rt^+	0.1270	0.0038	0.0292	0.0005	0.0008
rt^-	0.1268	0.0102	0.0903	0.0033	0.0084
tr^+	0	0.0041	0.0739	0.1056	0.0303
tr^-	0	0.0077	0.2301	0.0373	0.0247

Table 34: Infection rates inside museums among subpopulations for half of individuals who go to museums and do not comply with healthy behavior intervention (compliance rate 50%)

	rt^+	tr^+
rt^+	0.3773	0.2165
tr^+	1.7294	0.3037

M. D. Williams. TRANSIMS: Transportation analysis and simulation system. Technical Report LA-UR-00-1725, Los Alamos National Laboratory, 2001.

[2] J. Heffernan, R. Smith, and L. Wahl. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2:281 – 293, 2005.

[3] R. Horn and C. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge UK, 1991.

[4] S. D. Valle, J. M. Hyman, H. W. Hethcote, and S. G. Eubank. Mixing patterns between age groups using social networks. *Social Networks*, 29:539 – 554, 2007.

[5] P. van den Driessche and J. Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Journal of Mathematical Biosciences*, 2:29 – 48, 2002.

Table 35: Infection rates inside museums among subpopulations for half of individuals who go to museums and comply with healthy behavior intervention (compliance rate 50%)

	Efficacy 80%		Efficacy 60%		Efficacy 40%		Efficacy 20%	
Subpopulation	rt^+	tr^+	rt^+	tr^+	rt^+	tr^+	rt^+	tr^+
rt^+	0.3019	0.1732	0.2264	0.1299	0.1509	0.0866	0.0755	0.0433
tr^+	1.3835	0.2430	1.0377	0.1823	0.6918	0.1215	0.3459	0.0608