Understanding the role of urban design in disease spreading

Noel G. Brizuela [1,2*], Néstor García-Chan [2], Humberto Gutiérrez Pulido [3], Gerardo Chowell [4,5]

1 Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California, USA
2 Department of Physics, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico
3 Department of Mathematics, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico
4 Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, USA
5 Division of International Epidemiology and Population Studies, Fogarty International Center, National Institute of Health, Bethesda, Maryland, USA

* nogutier@ucsd.edu

Abstract

Cities are complex systems whose characteristics impact the health of people who live in them. Nonetheless, urban determinants of health often vary within spatial scales smaller than the resolution of epidemiological datasets. Thus, as cities expand and their inequalities grow, the development of theoretical frameworks that explain health at the neighborhood level is becoming increasingly critical. To this end, we developed a methodology that uses census data to introduce urban geography as a leading-order predictor in the spread of influenza-like pathogens. Here, we demonstrate our framework using neighborhood-level census data for Guadalajara (GDL, Western Mexico). Our simulations were calibrated using weekly hospitalization data from the 2009 A/H1N1 influenza pandemic and show that daily mobility patterns drive neighborhood-level variations in the basic reproduction number R_0, which in turn give rise to robust spatiotemporal patterns in the spread of disease. To generalize our results, we ran simulations in hypothetical cities with the same population, area, schools and businesses as GDL but different land use zoning. Our results demonstrate that the agglomeration of daily activities can largely influence the growth rate, size and timing of urban epidemics. Overall, these findings support the view that cities can be redesigned to limit the geographic scope of influenza-like outbreaks and provide a general mathematical framework to study the mechanisms by which local and remote health consequences result from characteristics of the physical environment.

Introduction

Empirical studies have identified inter-city variations in the timing, intensity and severity of influenza-like outbreaks [1–4]. Aiming to understand the mechanisms through which city characteristics yield such health consequences, epidemiologists have resourced to a variety of methods. Epidemiological data reveal compelling statistical correlations, but do not resolve intra-city variations in health that are driven by lifestyle inequalities at the neighborhood level [2, 4–6]. In contrast, agent-based computational models use massive mobility datasets to recreate the behavior of individuals as they

interact and spread infections [7–12]. Even though this approach allows for household-level analyses, its elevated complexity obscures relationships between model parameters and their health impacts, making it inadequate for the development of general strategies in public health [13]. Furthermore, agent-based simulations are costly and the information necessary to calibrate them is not openly available for most of the world's cities. Thus, despite the fact that epidemics have the potential to be seeded anywhere, these intricate models have been overwhelmingly applied to populations in the developed world. Meanwhile, the fundamental mechanisms that drive health inequalities within metropolitan areas remain elusive.

Urban design determines the densities and relative locations of housing, jobs and services inside a city. Consequently, it influences the transportation choices of the population [14, 15] and thus helps shape interaction networks through which diseases are spread. For example, the agglomeration of jobs and services drives large fractions of a city's population to gather in small fractions of its area, increasing contact rates between the residents of distant neighborhoods [16, 17]. In what follows, we quantify the agglomeration of urban mobility using two Gini coefficients ($0 \leq G_{origins}, G_{destinations} \leq 1$) that use neighborhood-level census data to measure spatial inequalities in the area-density of housing and activities throughout a city. The extreme values $G = 0$ represent homogeneous distributions of the population (trip origins) or its activities (trip destinations); inversely, $G = 1$ indicates that all housing or activities are concentrated within a single location. In epidemiological terms, $G = 1$ produces homogeneous mixing conditions, which allow individuals to interact indiscriminately with all other members of the population. Under this scenario, disease spreading may be modeled using simple ordinary differential equations where the total population $N = S + I + R$ is split into susceptible S, infected I, and recovered R groups. However, reality deviates from these conditions ($G \neq 1$) and successful models must introduce heterogeneous contact networks to link the members of a population [18].

In this article, we introduce urban design into a susceptible-infected-recovered (SIR) epidemiological model and evaluate it as the leading-order predictor for the spatiotemporal evolution of influenza-like outbreaks. Our methodology uses spatially-resolved census data to infer transportation patterns in cities and is meant to bypass the need for large mobility datasets in urban health simulations. In our model, individuals are grouped by age and place of residence, and are subsequently represented through probability density functions that describe their likely travel habits. This approach sacrifices the hyperrealism of agent-based simulations to instead resolve the spatial patterns that arise when a heterogeneous set of metapopulations interact. The Guadalajara Metropolitan Area (GDL) is used as an example, where hospitalization data from the 2009 Influenza A/H1N1 pandemic were used to calibrate disease parameters (S1 Fig). Lastly, we ran simulations in hypothetical cities with the same area, population density, number of schools and businesses as GDL. This allowed us to demonstrate that changes in the spatial distributions of housing, education and economic activities yield large variations in the size and early growth rate of epidemics across cities that would be deemed identical from a large-scale perspective.

Modeling framework

Inferred mobility patterns

High-resolution maps of residential density provide the background distribution of a city's population, introducing a first layer of spatial dependence to the SIR transmission framework. Likewise, the locations and prominence of activity hubs determine common

trip destinations were human interactions occur and diseases are spread. In our gravity model, trips originate at home and the probability density $P(\mathbf{x}, \mathbf{y}')$ that a person residing at $\mathbf{x}$ will visit $\mathbf{y}'$ on a given day is the joint result of two factors: The distance $\Delta r_{xy'}$ between both sites and the overall popularity of $\mathbf{y}'$ as a destination. The effect of distance $\Delta r_{xy'}$ on the likelihood of displacements between two points has been studied empirically, yielding the general form in Equation (1) [19]. Simplifying daily mobility as radial displacements out of a place of residence $\mathbf{x}$, $P(\Delta r_{xy'})$ yields a first estimate of the spatial distribution of individuals as they go through their daily routines. Moreover, parameters $\Delta r_0, b, \kappa$ capture the effects of transportation infrastructure, as the quality of these services largely determines the willingness of individuals to travel long distances [20].

$$P(\Delta r_{xy'}) = (\Delta r_{xy'} + \Delta r_0)^{-b} \exp(-\Delta r_{xy'}/\kappa) \tag{1}$$

Trip Attraction Strength (TAS) further refines our approximation of human mobility patterns in cities. TAS is an estimate of the daily number of visitors driven into a location $\mathbf{y}'$ by education and economic activities [21]. Its value may be calculated using school enrollment data and the number of workers employed at $\mathbf{y}'$ by different economic sectors, allowing different types of establishments to be weighed by the traffic they each induce. Our model incorporates TAS through Equation (2), where parameters Re, Se, In and Pr denote the number of jobs registered at $\mathbf{y}'$ by retail, service, industrial and primary activity organizations respectively [21]. Similarly, St is the number of students enrolled at educational institutions inside the same area. Equation (3) thus defines a probabilistic gravity model that represents mobility as the weighted influence areas of individuals, who try to minimize their displacements but will travel farther given an economic incentive (higher TAS). Here, the normalization factor $c(\mathbf{x})$ allows to adjust for the average number of daily trips made by residents of different neighborhoods. In what follows, we use the empirical values $\kappa = 80$ km, $b = 1.75$ found by González and collaborators [19] but prescribe $\Delta r_0 = 5$. This is a conservative choice, as it facilitates longer trips and thus homogenizes the transportation habits of the city's inhabitants.

$$TAS(\mathbf{y}') = 3.0\,\mathrm{Re} + 2.4\,\mathrm{Se} + 1.9\,\mathrm{In} + 1.3\,(\mathrm{Pr} + \mathrm{St}) \tag{2}$$

$$P_S(\mathbf{x}, \mathbf{y}') = c(\mathbf{x}) \cdot P(\Delta r_{xy'}) \cdot TAS(\mathbf{y}') \tag{3}$$

Before integrating these concepts into an epidemiological transmission model, we must distinguish between the behavior of susceptible and infected groups. We define the mobility of symptomatic groups via Equation (4), where $0 < \alpha(\mathbf{x}) < 1$ is an isolation parameter whose spatial dependence allows the representation of neighborhood-targeted intervention strategies as well as social and economic factors that influence the adaptive behavior of sick individuals. The term $H(\mathbf{x}, \mathbf{y}')$ accounts for the added probability of visits to the nearest hospital, as sick individuals will likely seek diagnosis and treatment. The locations of health care facilities thus determine the last spatial layer of input to our SIR transmission framework. A summary of all spatial components of our framework, as estimated for GDL (see Materials and Methods), is shown in Figure 1.

$$P_I(\mathbf{x}, \mathbf{y}') = \alpha(\mathbf{x}) \cdot P_S(\mathbf{x}, \mathbf{y}') + H(\mathbf{x}, \mathbf{y}') \tag{4}$$

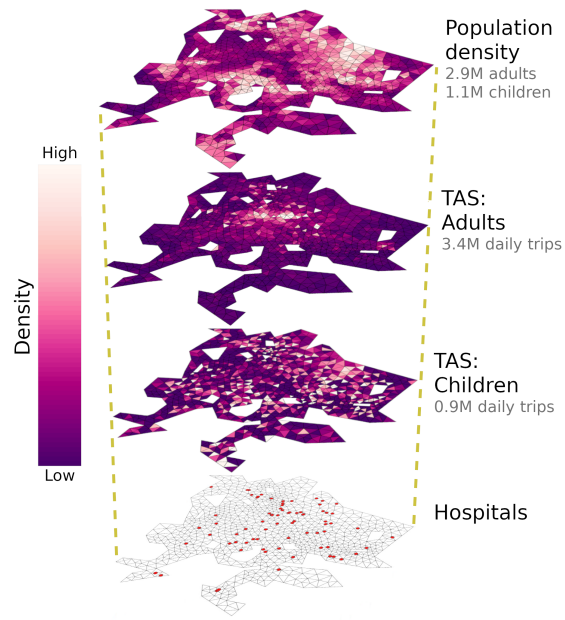

Fig 1. Layers of model input: population density and TAS were calculated for adults and children within a triangular grid representing GDL (Eq. 2, see Materials and Methods). While children are attracted to early-stage schools, the destinations of adult mobility are determined by the locations of retail stores, factories, office buildings, high schools and universities. Given that the spatial distributions of TAS for adults and children differ drastically, individuals are more likely to interact with people within their own age group. Satellite image: Google, Landsat/Copernicus, Maxar Technologies.

Metapopulation model

Let us define three identical domains $\tilde{\Omega} = \Omega' = \Omega \in \mathbb{R}^2$ to represent the urban region under study. Members of the population N are then segregated under an SIR transmission scheme and distributed following population density functions $S(\mathbf{x},t) + I(\mathbf{x},t) + R(\mathbf{x},t) = \rho(\mathbf{x})$ such that $\int_\Omega \rho(\mathbf{x}) d\Omega = N$. This approach links people to their place of residence (henceforth $\mathbf{x} \in \Omega$ for susceptibles and $\tilde{\mathbf{x}} \in \tilde{\Omega}$ for infectives), allowing to formulate a Distributed Contacts SIR model (DC-SIR, Eqs. 5-7 adapted from [22, 23]). Here, $\beta(\mathbf{y}')$ is the probability of contagion given a susceptible-infective interaction that happens at $\mathbf{y}'$ and γ is the recovery rate. The interaction kernel $k(\mathbf{x}, \mathbf{y}', t)$ yields the expected number of interactions between a member of $S(\mathbf{x},t)$ and all infected individuals present at a trip destination $\mathbf{y}' \in \Omega'$ and time t [22, 23]. Thus, $k(\mathbf{x}, \mathbf{y}', t)$ introduces the heterogeneous contact networks through which diseases are spread.

$$\frac{\partial S(\mathbf{x},t)}{\partial t} = -S(\mathbf{x},t) \int_{\Omega'} \beta(\mathbf{y}') \cdot k(\mathbf{x}, \mathbf{y}', t)\, d\Omega' \tag{5}$$

$$\frac{\partial I(\mathbf{x},t)}{\partial t} = -\left(\frac{\partial S(\mathbf{x},t)}{\partial t} + \frac{\partial R(\mathbf{x},t)}{\partial t} \right) \tag{6}$$

$$\frac{\partial R(\mathbf{x},t)}{\partial t} = \gamma I(\mathbf{x},t) \tag{7}$$

To define the interaction kernel $k(\mathbf{x}, \mathbf{y}', t)$, we use origin-destination density functions $P_S(\mathbf{x}, \mathbf{y}', t), P_I(\mathbf{x}, \mathbf{y}', t)$ for members of $S(\mathbf{x},t)$ and $I(\mathbf{x},t)$ respectively. Next,

we compute the total number of infected individuals expected to visit $\mathbf{y}'$ at a given time, which is the integral from all trip origins $\int_{\tilde{\Omega}} P_I(\tilde{\mathbf{x}}, \mathbf{y}', t) \cdot I(\tilde{\mathbf{x}}, t)\, d\tilde{\Omega}$. Taking the product with $P_S(\mathbf{x}, \mathbf{y}', t)$ thus gives an expression for the expected number of SI contacts occurring at $\mathbf{y}'$ for an individual who resides at $\mathbf{x}$

$$k(\mathbf{x}, \mathbf{y}', t) = P_S(\mathbf{x}, \mathbf{y}', t) \int_{\tilde{\Omega}} P_I(\tilde{\mathbf{x}}, \mathbf{y}', t) \cdot I(\tilde{\mathbf{x}}, t)\, d\tilde{\Omega}. \tag{8}$$

A similar procedure yields the reproductive number $R_t(\tilde{\mathbf{x}}, t)$, defined as the number of secondary cases caused by a single infected individual who lives at $\tilde{\mathbf{x}}$ and contracted the disease at time t. Instead of calculating the number of SI interactions for a susceptible in transit, we now use the spatial distribution of susceptibles at a given time. This is given by the integration of susceptible mobility over all trip origins $\int_{\Omega} P_S(\mathbf{x}, \mathbf{y}', t) \cdot S(\mathbf{x}, t)\, d\Omega$. Multiplying this by the origin-destination density for infected mobility $P_I(\tilde{\mathbf{x}}, \mathbf{y}', t)$ then yields the expected number of SI interactions occurring at $\mathbf{y}'$ for a member of $I(\tilde{\mathbf{x}}, t)$

$$q(\tilde{\mathbf{x}}, \mathbf{y}', t) = P_I(\tilde{\mathbf{x}}, \mathbf{y}', t) \int_{\Omega} P_S(\mathbf{x}, \mathbf{y}', t) \cdot S(\mathbf{x}, t)\, d\Omega. \tag{9}$$

Assuming that all variables are slowly-varying over the first generation of disease transmission, we can approximate the basic reproduction number $R_0(\tilde{\mathbf{x}})$ as Equation (10). Notice that $k(\mathbf{x}, \mathbf{y}', t)$ and $q(\tilde{\mathbf{x}}, \mathbf{y}', t)$ represent the vulnerability and disease spreading capacity of individuals during an infectious outbreak. Moreover, their spatial dependence implies that a person's role during an epidemic is a function of their place of residence but determined by characteristics of the locations they visit.

$$R_0(\tilde{\mathbf{x}}) \sim 1/\gamma \int_{\Omega'} \beta(\mathbf{y}') \cdot q(\tilde{\mathbf{x}}, \mathbf{y}', t = 0)\, d\Omega' \tag{10}$$

Results

Our modeling framework was tested using data for GDL, where census and economic data [24–26] were processed to derive the daily mobility patterns of children (age ≤ 15) and adults (age > 15, see Materials and Methods). Lorenz curves in Figure 2 represent the different degrees of agglomeration driven by housing ($G_{origins} = 0.32$) and daily activities ($G_{destinations} = 0.56$). As is true for large cities [16, 17], inferred trips were largely directed towards a few, hyper-affluent areas. In our model, these areas receive 25% of all daily trips but occupy less than 5% of the metropolis; meanwhile, housing places 25% of the population throughout 13.5% of the city's most densely populated regions. The condition $G_{destinations} > G_{origins}$ requires that some neighborhoods have a net loss of occupants during workdays, as their inhabitants leave and agglomerate around major activity hubs. This is true for 60% of all neighborhoods in GDL, which then act as net sources of mobility whose role is to supply more affluent destinations with visitors.

Values of $\beta(\mathbf{y}')$ were calibrated using city-wide hospitalization data from the 2009 A/H1N1 influenza pandemic (S1 Fig). In our model, this parameter yields a linear relationship for the probability of falling ill given the mean area density of infectives that one encounters throughout the day. We assumed a homogeneous transmission rate $\beta = \beta(\mathbf{y}')$ and obtained that the daily risk of contagion increased by $1.1 \pm 0.1\%$ for every 1000 infective adults per km^2 added to one's surroundings. Similarly, the addition of 1000 infective children per km^2 raised the probability of falling ill on a given day by $3.2 \pm 0.3\%$. With β fixed everywhere, spatial patterns in the evolution of epidemics

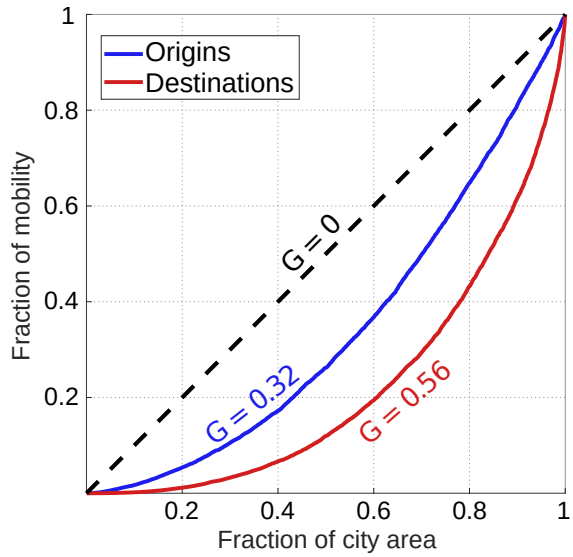

Fig 2. Lorenz curves show the cumulative distributions of housing (trip origins, blue) and daily activities (trip destinations, red) throughout portions of the city area. Deviations from homogeneous urban design ($G = 0$) arise due to the existence of activity hubs where people agglomerate, interact, and thus increase contact rates. At the limit value $G = 1$, all of the city's housing or daily activities are concentrated in a single location, allowing all members of the population to interact with each other (homogeneous mixing).

result entirely from the number and age of people who visit each one of the city's neighborhoods (Eqs. 8, 9). Hence, spatial variations in our simulation results and in the basic reproductive number $R_0(\tilde{\mathbf{x}})$ are a direct consequence of the city's transportation network (Eq. 10), which is by definition a product of urban design and land-use patterns (Eqs. 1 - 3).

Spatial variations in the basic reproductive number $R_0(\tilde{\mathbf{x}})$ (Fig. 3) had a profound impact on the spatiotemporal evolution of outbreaks in our model. Firstly, the age and place of residence of patient zero (initial conditions) significantly influenced the early rate of epidemic growth. In fact, the R_0 of patient zero can delay the peak of an epidemic by as much as 9 days (Fig. 4.a,b). Secondly, people with the highest $R_0(\tilde{\mathbf{x}})$ also had the highest probability of falling ill. In particular, the increased contact rates of downtown residents allowed them to spread pathogens 50% more efficiently than their suburban counterparts (Fig. 3), but also put them at a higher risk of contagion. Regardless of initial conditions, this particular situation gave rise to a spatial pattern in which influenza spread as waves of disease that emanate out of the city center (Fig. 4.c, S2 Fig, S1 Video).

To better appreciate the structure of wave-like patterns and inequalities in the spread of disease, we define the relative incidence $\Upsilon(\tilde{\mathbf{x}}, t)$ (Eq. 11) and use simulation results to inspect its statistical dependence on the distance $\Delta r_{\tilde{x}x^*}$ between neighborhoods $\tilde{x}$ and the city's largest retail hub $\mathbf{x}^*$, where TAS peaks. For all initial conditions, the incidence of influenza was highest in the city centre and decreased towards the suburbs (Fig. 4.c). Time dependence represented in S2 Fig suggests a relationship of the form $\Upsilon(\tilde{\mathbf{x}}, t) = \Delta r_{\tilde{x}x^*}^{-d(t)}$. Here, $d(t)$ is a non-negative function that decreases with time, as $\Upsilon(\tilde{\mathbf{x}}, t)$ approaches unity everywhere towards the end of outbreaks.

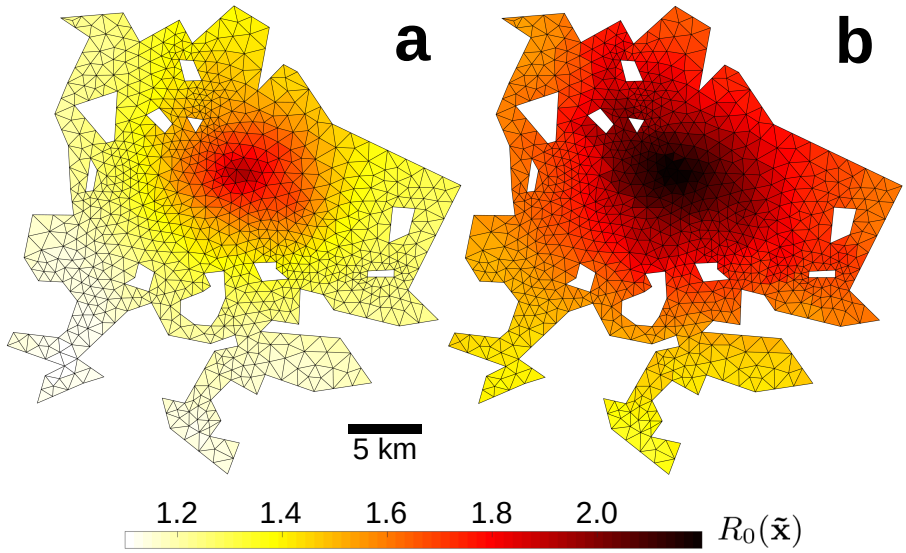

Fig 3. The basic reproductive number R_0 (Eq. 10) varies at the neighborhood scale, showing that adults (**a**) and children (**b**) living near the city center play a disproportionate role in disease spreading. Because a substantial fraction of the city's entire population visits downtown on any given day, people living there have the highest contact rates and can thus spread diseases to the residents of distant neighborhoods. Population-averaged values of R_0 were 1.27 for adults and 1.57 for children.

$$\Upsilon(\tilde{\mathbf{x}}, t) = \frac{I(\tilde{\mathbf{x}}, t)}{\rho(\tilde{\mathbf{x}})} \left(\frac{I(\mathbf{x}^*, t)}{\rho(\mathbf{x}^*)} \right)^{-1} \tag{11}$$

It is clear from Figures 3 and 4 that mobility hotspots play an important role in setting the growth rate and evolution of epidemics in our model. To gain more insight, we relocated businesses, schools and housing across the GDL numerical grid and ran simulations in the resulting geographies. This allowed us to compare outbreaks across cities with unique transportation networks but the same demographics, area, number of schools, businesses and daily trips as GDL. In one set of experiments we modified the Gini coefficients of mobility by relocating businesses, schools and housing across areas with low and high TAS or population density. As a result, we modulated the agglomeration of urban mobility but preserved the spatial structure of GDL. In another experiment, we redistributed housing, schools and businesses at random to evaluate whether the effects of agglomeration can be generalized to all cities despite their spatial structure. All simulations used the same values of β (previously calibrated for GDL) and an infective period $1/\gamma$ of 4 days. Results are summarized in Figure 5.

Our simulations suggest that, for all cities, the basic reproduction number R_0 varies as a function of $G_{destinations}$ and is virtually independent of $G_{origins}$ (Fig. 5.a,b). This relationship is nonlinear for adults and linear for children, mainly because $G_{destinations}$ is set by commercial and economic superclusters that primarily attract adults. Furthermore, our results suggest that housing and activities in metropolitan areas with $G_{destinations} < 0.58$ can be redistributed so that the mean basic reproduction number of adults remains under the threshold value $R_0 = 1$. Namely, in this illustrative case where β is fixed everywhere and transmission dynamics are solely driven by urban design, cities can be redesigned to render their populations incapable of sustaining epidemics.

At the end of an outbreak, the attack rate z measures the fraction of the population

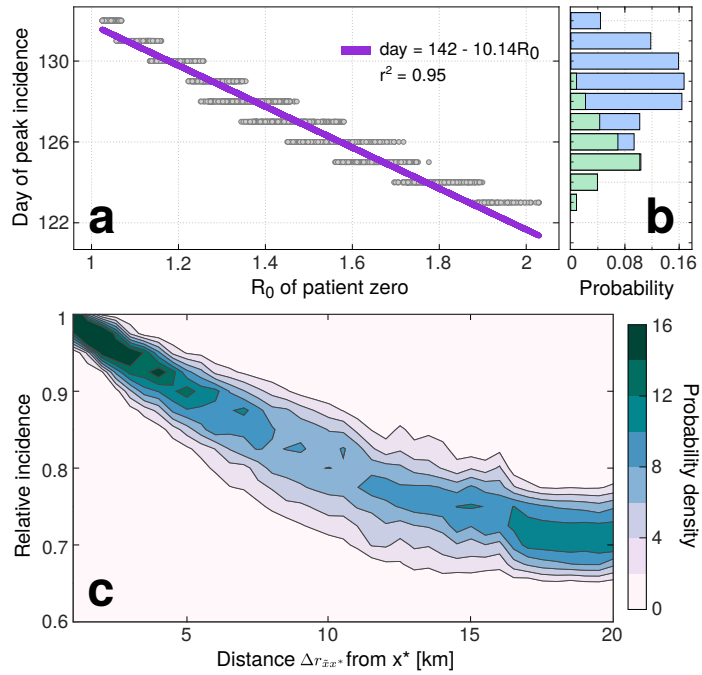

Fig 4. Over 3160 simulations with different initial conditions, panel **a** shows the time of peak incidence (when $\int_{\tilde{\Omega}} I(\tilde{x}, t) d\Omega$ reaches a maximum) plotted against the R_0 of patient zero. A histogram in panel **b** shows the probability that an outbreak peaks on a given day and the separate contributions of children (green) and adults (blue) to this variability. Color in panel **c** shows the expected probability distribution of relative incidence $\Upsilon(\tilde{\mathbf{x}}, t)$ between $t = 120$ and 140 days given the distance $\Delta r_{\tilde{x}x^*}$ between a location $\tilde{\mathbf{x}}$ and the city's largest retail hub $\mathbf{x}^*$.

that contracted the disease. Under homogeneous mixing conditions ($G_{destinations} = G_{origins} = 1$), R_0 and z are linked by Equation (12) [27]. At $R_0 = 1.05$, this relationship predicts $z_{theory} = 0.094$; however, heterogeneity introduced by urban design in our simulations yields attack rates in the range $0.001 < z_{sim} < 0.003$. In other words, the assumption of homogeneous mixing in cities can lead models to overestimate epidemic size by as much as two orders of magnitude for low values of R_0 (Fig. 5.c). This disagreement is partly caused by neighborhood-level variations in $R_0(\tilde{\mathbf{x}})$, which can produce a mean $R_0 > 1$ even when a portion of the population has $R_0(\tilde{\mathbf{x}}) < 1$ and thus cannot drive epidemic growth. At higher values of R_0, the homogeneous mixing solution (Eq. 12) continues to overestimate z, which further varies with the city's spatial structure. Differences between the GDL (where the spatial structure of GDL is preserved) and Random experiments (Fig. 5) suggest that, at equal R_0, realistic urban layouts may act to decrease the size of epidemics when compared to cases where housing and activities are distributed randomly. However, a lower R_0 can be achieved with a given $G_{destinations}$ when cities are designed at random.

$$1 - z = e^{-R_0 z} \tag{12}$$

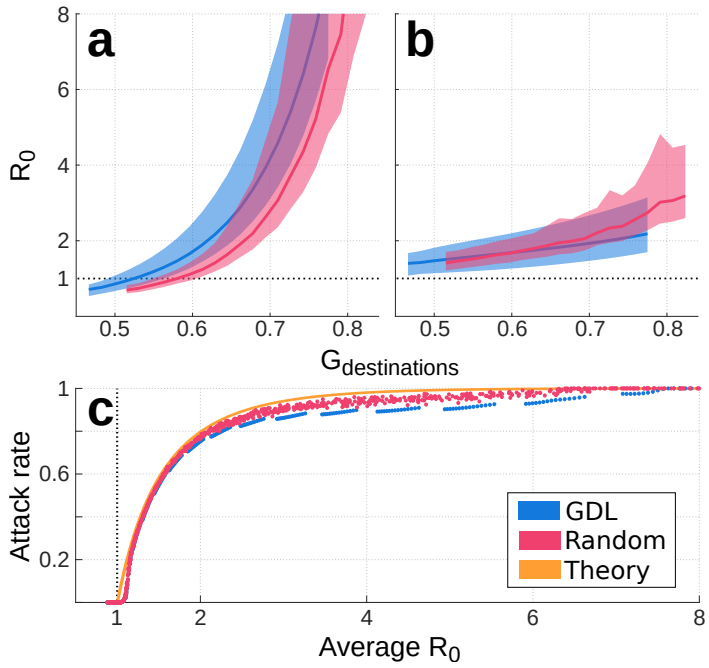

Fig 5. The population-averaged basic reproduction number R_0 is plotted as thick lines for adults (**a**) and children (**b**). Color shading shows the range of values $R_0(\tilde{\mathbf{x}})$ calculated with equation (10) in cities with the spatial structure of GDL (blue) and those with randomized urban design (pink). The relationship between population-averaged R_0 and attack rate z is shown in panel **c** for cities with realistic (blue) and random (pink) spatial structures. The theoretical relation for the attack rate under homogeneous mixing conditions (Eq. 12) is shown in orange. Disagreements between these curves show the impacts of urban design on the growth rate and size of influenza-like outbreaks for cities with the same area, population and economic activities.

Discussion

Our parameterization of contact heterogeneity in metropolitan areas produces smaller epidemics than homogeneous mixing models (Fig. 5) and predicts that, depending on where they live, people play unequal roles in disease spreading (Fig. 3). These inequalities relate the identity of patient zero to the timing of influenza-like outbreaks (Fig. 4.a,b) and can give rise to spatiotemporal patterns in the spread of disease (Fig. 4.c, S2 Fig, S1 Video). In fact, these processes mirror those of large scale transportation networks, whose heterogeneity leads to spatial inequalities in the size, timing and growth rate of epidemics [1, 28, 29]. As similar insight is used for the development of optimal intervention strategies in global-scale infectious outbreaks [30], public health officials may use results like those presented in Figure 3 to device neighborhood-specific campaigns that target those who are most vulnerable or generate the most secondary cases. For example, the efficiency of vaccination campaigns is likely to vary whether they target the inhabitants of the city center, its visitors, or people living in suburban areas 3. Similarly, the need for treatment and diagnosis is expected to differ across neighborhoods (Fig. 4.c), suggesting that the distribution of medical personnel and equipment throughout a city can be accordingly optimized.

Although our figures present the health impacts of urban geography as a function of place of residence, inequalities result primarily from the remote influence of people in

other neighborhoods and the characteristics of places where interactions occur (Eqs. 5-10). The inclusion of non-residental processes is essential to the accurate representation of environmental health impacts [31, 32] and is achieved here through interaction kernels $k(\mathbf{x}, \mathbf{y}', t)$ and $q(\tilde{\mathbf{x}}, \mathbf{y}', t)$. While we used a constant disease parameter, the representation of local and remote environmental conditions can be further refined by assigning full spatial dependence to transmission parameters so that $\beta = \beta(\mathbf{x}, \mathbf{y}', \tilde{\mathbf{x}}, t)$. Generally speaking, our analytical framework is designed to map geographic information onto social space, establishing connections that are weighed by a risk of contagion β. This assigns realistic network characteristics to a spatially-explicit disease transmission model, which simplifies the interpretation of simulation results and may enable additional theoretical and statistical analyses [33, 34].

Our results highlight the extent to which the assumption of homogeneous mixing in metropolitan areas can bias model results. Most notably, epidemic size in simulations that use the same transmission parameter β throughout hypothetical, seemingly identical cities, covers the entire range $0 < z \leq 1$ (Fig. 5.c). This suggests that macroscopic quantities such as population size and density hold little to no dynamical significance in the evolution of infectious outbreaks. Instead, the spatial organization of human mobility (quantified here using the Gini coefficient $G_{destinations}$, Fig. 2) seems to control epidemic growth (Figs. 5.a,b) and is statistically correlated with population size [4, 12]. Dalziel and collaborators explored this notion [12] by comparing agent-based model results across 48 Canadian cities, but were not able to isolate the effects of spatial heterogeneity from all other information embedded in massive mobility datasets that were used as model input.

Empirical relationships between $G_{destinations}$, R_0 and the attack rate z shown in Fig. 5 suggest that, through its influence on transportation networks, urban design can be optimized to render cities incapable of seeding large outbreaks. The existence of invasion thresholds under which populations cannot sustain epidemic growth is a general characteristic of metapopulation models [35]. While many authors have conjectured about the possibility of designing cities to minimize disease prevalence [4, 12, 36, 37], we present the first dynamical argument to link these ideas and the invasion threshold proposition of Colizza and Vespignani [35]. More specifically, our results suggest that, by scattering activity hubs throughout a city (instead of clustering them in the city center as in GDL), city planners can segregate subsets of the population and potentially inhibit the rapid transmission of pathogens across neighborhoods (Fig. 5). Consistent with known statistical relations between the transmission potential of influenza and the spatial organization of human behavior [4], our framework explains one of the mechanisms behind health inequality across the world's cities and their neighborhoods.

Agent-based models are the standard method to run realistic simulations of disease spreading in cities [7, 8, 10–12]. However, these simulations are costly and require input from massive mobility data that are not openly available for most of the world's cities, making them inaccessible to many research groups and public health organizations. Likewise, large-scale epidemiological models often assume homogeneous mixing conditions in cities and neglect all processes that occur at the neighborhood level [1, 4, 9, 38]. When compared to agent-based model output, this simplification can lead to overestimate the size of epidemics and miscalculate their timing, and can thus introduce considerable bias [9]. Although far simpler, our parameterization of small-scale heterogeneity leads to the same conclusions in such a comparison (Figs. 4, 5) . Thus, we believe that the mathematical framework presented here is an adequate alternative to introduce neighborhood-level processes in large scale epidemiological simulations.

Because it uses standard census and economic data to infer daily mobility patterns

(see Materials and Methods), our method enables researchers to investigate transmission dynamics in virtually any city. Unfortunately, mobility parameters used here (Eqs. 1-3) were not calibrated with real origin-destination data (which exist for GDL during the 2009 A/H1N1 pandemic [11] but are not openly available) and may thus be inaccurate. Nonetheless, our analyses rely on the patterns that arise from fundamental processes driven by the agglomeration of economic activities and highlight small-scale inequalities that are not resolved by most epidemiological datasets (Fig. 3). In reality, spatial inequalities in health result from the complex interactions of social, environmental and biological processes among which urban design is only one. Nonetheless, the development of theoretical frameworks that seek to understand these complex interactions is crucial to help design new policies and research that aim to improve health in cities [36, 37, 39].

Materials and methods

Inclusion of census data

We solved equations (5)-(7) in finite differences by mapping all their variables onto a 1580-element triangular mesh representing GDL ([40], Fig. 1). Neighborhood-level data from the 2010 census [25] were used to calculate the total adult (age > 15) and infantile (age ≤ 15) populations at each grid element. $TAS(\mathbf{y}')$ was estimated for adults and children from two publicly available datasets: The 2015 National Statistical Directory of Economical Units (DENUE) lists all registered employers in Mexico along with their sector, number of workers and location [24]. Similarly, the National System for School Information (SNIE) locates all of the country's schools and universities and lists their enrollment at each educational stage [26]. DENUE employment data were combined with SNIE enrollment at the high school and university levels to calculate TAS for the adult population via equation 2. On the other hand, TAS for infantile mobility was inferred using SNIE enrollment data for educational stages up to the middle school level, roughly corresponding with the age threshold between our age groups. Next, origin-destination matrices were obtained for all grid elements using equations (3) and (4). Resulting maps of population density and TAS for adults and children are shown in Figure 1.

Mobility parameters were established a priori as $\Delta r_0 = 5$ km, $b = 1.75$, $\kappa = 80$ km in equation 1. Sensibility analyses showed that variations in these values primarily impact $G_{destinations}$, whose consequences are shown in Figure 5, but do not change the fundamental phenomena described in this study. Similarly, $c(\mathbf{x})$ in equation 3 was chosen to yield an average of 1.18 and 0.85 daily trips per person for adults and children respectively. The mobility of infected individuals was computed via equation 4 with $\alpha = 0.8$ and an 8% probability of visiting the hospital once in an infective period.

Numerical solutions

Time-varying solutions to equations (5)-(7) were obtained using a forward finite-difference scheme that considered piece-wise constant functions defined over the triangular mesh shown in Figures 1 and 3. Continuous input variables were transformed to be constant at all grid elements, which are henceforth noted as $\Omega_j \subset \Omega$, and whose area is $A_j = \int_{\Omega_j} d\Omega$. For instance, the number of susceptibles in Ω_j at time t_k, was calculated as $S_j(t_k) = \int_{\Omega_j} S(\mathbf{x}, t)\, d\Omega$. Origin-destination matrices were defined using the corresponding origin-destination density functions (Eq. 3) as

$$P_{S:jm} = \frac{1}{A_j} \int_{\Omega_j} \int_{\Omega'_m} P_S(\mathbf{x}, \mathbf{y}')\, d\Omega'\, d\Omega. \tag{13}$$

Similarly, $\beta(\mathbf{y}')$ can be mapped onto the grid as

$$\beta_m = \frac{1}{A_m^2} \int_{\Omega'_m} \beta(\mathbf{y}') d\Omega'. \tag{14}$$

With all variables defined within the numerical grid, integrals over $\tilde{\Omega}$ and Ω' in equation (5) were replaced by sums over infective origins $\tilde{\Omega}_h$ and trip destinations Ω'_m. Thus, the temporal evolution in the number of susceptibles at Ω_j over one timestep Δt was computed as

$$S_j(t_{k+1}) = S_j(t_k) \left[1 - \Delta t \sum_m \beta_m P_{S:jm} \sum_h P_{I:hm} I_h(t_k) \right]. \tag{15}$$

Supporting information

S1 Fig. Comparison of data and model output. The city-wide number of weekly hospitalizations in GDL during the 2009 AH1N1 influenza pandemic (blue) is compared to simulation results under various initial conditions (gray). Because our model does not feature interventions nor a time-dependent behavioral response to outbreak intensity, values of β were calibrated using only 9 weeks of data at the onset of the third and largest wave in the AH1N1 epidemic (pink shading).

S2 Fig. Radial dependence of relative incidence. Expected evolution of relative incidence at different times since the onset of 1580 outbreak simulations, each run under a different initial condition. At any time, the proportion of people who are sick is maximum near the city center where TAS (Eq. 2) peaks, and decreases radially towards the suburbs. As epidemics die out and the pool of susceptibles available to spread infection is depleted, relative incidence approaches unity everywhere, weakening the health contrasts that exist between residents of the downtown and suburban areas.

S1 Video. Spatiotemporal evolution of an outbreak. Model output shows the time-dependent fraction of people who are sick at each location in our numerical grid. The outbreak shown in this video was seeded by adding 15 infective adults to a grid element in the suburbs.

Acknowledgments

The authors wish to acknowledge Haley McInnis, Thomas Gorin and Tarik Benmarhnia for their valuable feedback and suggestions.

References

1. Viboud C, Bjørnstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. Synchrony, waves, and spatial hierarchies in the spread of influenza. science. 2006;312(5772):447–451.

2. Chowell G, Bettencourt LM, Johnson N, Alonso WJ, Viboud C. The 1918–1919 influenza pandemic in England and Wales: spatial patterns in transmissibility and mortality impact. Proceedings of the Royal Society of London B: Biological Sciences. 2008;275(1634):501–509.

3. Yang W, Lipsitch M, Shaman J. Inference of seasonal and pandemic influenza transmission dynamics. Proceedings of the National Academy of Sciences. 2015;112(9):2723–2728.

4. Dalziel BD, Kissler S, Gog JR, Viboud C, Bjørnstad ON, Metcalf CJE, et al. Urbanization and humidity shape the intensity of influenza epidemics in US cities. Science. 2018;362(6410):75–79.

5. World Health Organization. Centre for Health Development, World Health Organization. Hidden Cities: unmasking and overcoming health inequities in urban settings. World Health Organization; 2010.

6. Rocha LE, Thorson AE, Lambiotte R. The non-linear health consequences of living in larger cities. Journal of Urban Health. 2015;92(5):785–799.

7. Eubank S, Guclu H, Kumar VA, Marathe MV, et al. Modelling disease outbreaks in realistic urban social networks. Nature. 2004;429(6988):180.

8. Lee BY, Bedford VL, Roberts MS, Carley KM. Virtual epidemic in a virtual city: simulating the spread of influenza in a US metropolitan area. Translational Research. 2008;151(6):275–287.

9. Ajelli M, Gonçalves B, Balcan D, Colizza V, Hu H, Ramasco JJ, et al. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. BMC infectious diseases. 2010;10(1):190.

10. Cooley P, Brown S, Cajka J, Chasteen B, Ganapathi L, Grefenstette J, et al. The role of subway travel in an influenza epidemic: a New York City simulation. Journal of Urban Health. 2011;88(5):982.

11. Frias-Martinez E, Williamson G, Frias-Martinez V. An agent-based model of epidemic spread using human mobility and social network information. In: Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. IEEE; 2011. p. 57–64.

12. Dalziel BD, Pourbohloul B, Ellner SP. Human mobility patterns predict divergent epidemic dynamics among cities. Proceedings of the Royal Society of London B: Biological Sciences. 2013;280(1766):20130763.

13. Arino J, Brauer F, van den Driessche P, Watmough J, Wu J. Simple models for containment of a pandemic. Journal of the Royal Society Interface. 2006;3(8):453–457.

14. Hanson S, Schwab M. Accessibility and intraurban travel. Environment and Planning A. 1987;19(6):735–748.

15. Frank LD, Pivo G. Impacts of mixed use and density on utilization of three modes of travel: single-occupant vehicle, transit, and walking. Transportation research record. 1994;1466:44–52.

16. Chowell G, Hyman JM, Eubank S, Castillo-Chavez C. Scaling laws for the movement of people between locations in a large city. Physical Review E. 2003;68(6):066102.

17. Dannemann T, Sotomayor-Gómez B, Samaniego H. The time geography of segregation during working hours. Royal Society open science. 2018;5(10):180749.

18. Chowell G, Sattenspiel L, Bansal S, Viboud C. Mathematical models to characterize early epidemic growth: a review. Physics of life reviews. 2016;18:66–97.

19. Gonzalez MC, Hidalgo CA, Barabasi AL. Understanding individual human mobility patterns. Nature. 2008;453(7196):779.

20. Suárez M, Murata M, Delgado Campos J. Why do the poor travel less? Urban structure, commuting and economic informality in Mexico City. Urban studies. 2016;53(12):2548–2566.

21. Jakimavičius M, Burinskiene M. A GIS and multi-criteria-based analysis and ranking of transportation zones of Vilnius city. Technological and Economic Development of Economy. 2009;15(1):39–48.

22. Medlock J, Kot M. Spreading disease: integro-differential equations old and new. Mathematical Biosciences. 2003;184(2):201–222.

23. Lee C, Hoopes M, Diehl J, Gilliland W, Huxel G, Leaver E, et al. Non-local concepts and models in biology. Journal of theoretical biology. 2001;210(2):201–219.

24. Instituto Nacional de Estadística y Geografía. Directorio Estadístico Nacional de Unidades Económicas; 2015.

25. Instituto Nacional de Estadística y Geografía. Población por colonia; 2010.

26. Secretaría de Educación Pública. Sistema Nacional de Información de Escuelas; 2017.

27. Diekmann O, Heesterbeek H, Britton T. Mathematical tools for understanding infectious disease dynamics. vol. 7. Princeton University Press; 2012.

28. Colizza V, Barrat A, Barthélemy M, Vespignani A. The role of the airline transportation network in the prediction and predictability of global epidemics. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(7):2015–2020.

29. Merler S, Ajelli M. The role of population heterogeneity and human mobility in the spread of pandemic influenza. Proceedings of the Royal Society B: Biological Sciences. 2009;277(1681):557–565.

30. Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS. Strategies for mitigating an influenza pandemic. Nature. 2006;442(7101):448.

31. Inagami S, Cohen DA, Finch BK. Non-residential neighborhood exposures suppress neighborhood effects on self-rated health. Social science & medicine. 2007;65(8):1779–1791.

32. Chaix B, Duncan D, Vallée J, Vernez-Moudon A, Benmarhnia T, Kestens Y. The "residential" effect fallacy in neighborhood and health studies. Epidemiology. 2017;28(6):789–797.

33. Pellis L, Ball F, Bansal S, Eames K, House T, Isham V, et al. Eight challenges for network epidemic models. Epidemics. 2015;10:58–62.

34. Roberts M, Andreasen V, Lloyd A, Pellis L. Nine challenges for deterministic epidemic models. Epidemics. 2015;10:49–53.

35. Colizza V, Vespignani A. Invasion threshold in heterogeneous metapopulation networks. Physical review letters. 2007;99(14):148701.

36. Rydin Y, Bleahu A, Davies M, Dávila JD, Friel S, De Grandis G, et al. Shaping cities for health: complexity and the planning of urban environments in the 21st century. Lancet. 2012;379(9831):2079.

37. Grant M, Brown C, Caiaffa WT, Capon A, Corburn J, Coutts C, et al. Cities and health: an evolving global conversation. Cities & Health. 2017;1(1):1–9. doi:10.1080/23748834.2017.1316025.

38. Balcan D, Gonçalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A. Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. Journal of computational science. 2010;1(3):132–145.

39. Cummins S, Curtis S, Diez-Roux AV, Macintyre S. Understanding and representing 'place'in health research: a relational approach. Social science & medicine. 2007;65(9):1825–1838.

40. Geuzaine C, Remacle JF. Gmsh: A 3-D finite element mesh generator with built-in pre-and post-processing facilities. International journal for numerical methods in engineering. 2009;79(11):1309–1331.