How do Urban Factors Control the Severity of COVID-19?

Jacob Roxon, Marie-Sophie Dumont, Eric Vilain, Roland J.M. Pellenq

EpiDaPo Lab - CNRS / George Washington University Children’s National Medical Center, Children’s Research Institute, 111 Michigan Ave. NW, Washington, DC 20010, USA

*Roland J.M. Pellenq.

Email: roland.pellenq@cnrs.fr

Abstract

Human health in urban environment has emerged as a primary focus of sustainable development during the time of global pandemic caused by a severe acute respiratory syndrome due to the SARS-CoV-2 coronavirus, COVID-19. It has reshaped the world with the way our communities interact, people work, commute, and spend their leisure time. While different mitigation solutions for controlling COVID-19 virus transmission have already been established, global models that would explain and predict the impact of urban environments on the case fatality ratio CFR of COVID-19 (defined as the number of deaths divided by the number of cases over a time window) are missing. Here, with readily available data from public sources, we study the CFR of the coronavirus for 118 locations (city zip-codes, city boroughs, and cities) worldwide to identify the links between the CFR and outdoor, indoor and personal urban factors. We show that a probabilistic model, optimized on the sample of 20 districts from 4 major US cities, provides an accurate predictive tool for the CFR of COVID-19 regardless of the geographical location. When adjusted for the population, our model can be used to evaluate risk and severity of the disease at multi-geospatial scales worldwide ranging from zip-codes and neighborhoods to cities and countries for different waves of the pandemic. Our results suggest that although disease screening and vaccination policies to containment and lockdowns remain critical in controlling the spread of airborne diseases, urban factors such as population density, humidity, or order of buildings, should all be taken into consideration when identifying resources and planning targeted responses to mitigate the impact and severity of the viruses transmitted through air. We advocate the study of urban factors as a path towards facilitating timely deployment of targeted countermeasures and confinement strategies where sharing of personal information and availability of tests may be restricted or limited.

Introduction

Towards the end of January 2020, the World Health Organization (WHO) declared a Public Health Emergency of International Concern, which within less than 6 weeks, on March 11th 2020 was described as a global pandemic of coronavirus (COVID-19) disease. The most severe impact of the COVID-19 has been observed in urban areas (1, 2), which in most countries are home to over 70% of population. And while the implementation of pandemic response measures, such as vaccinations or social distancing and face covering restrictions, have provided a path towards controlling the epidemic (3, 4), they may not be sufficient to end the disease, which is likely to become endemic (5) resulting in infections and mortalities, severity of which, seen through the prism of the Case Fatality Ratio, CFR, (i.e. the ratio of the number of deaths to the number of confirmed cases) changes from country to country (Fig. 1).
Figure 1. The severity of the COVID-19 throughout the world showing cumulative COVID-19 CFR (Table S1) at the country scale between 01/2020 and 01/2022 with a geo-location of 30 cities (Table S2) used in this study.

SARS-CoV-2 virus stability is affected by physical, chemical, and biological factors (6). On inanimate surfaces, the most important are surface type, relative humidity (7), temperature (8), moisture content of the suspending medium (mucus). Although among humans, there are different possible transmission routes of COVID-19, the focal point of scientists has been airborne propagation of the SARS-CoV-2 (9), which involves projection of virions in aerosolized droplets of mucus from an infected individual, inhaled by sane individual. Due to air drag and gravity, virus-laden droplets with diameter, \( d \geq 5 \, \mu m \) emitted in a cough or sneeze, have a generally low projection distance of less than few meters (10, 11). Smaller virus-laden particles (\( d < 5 \, \mu m \)), can remain in air for hours and consequently be dispersed through airborne route both indoors and outdoors. In addition, the propagation of COVID-19 has been linked to particulate matter (PM) pollution levels showing a positive correlation between long-term exposure to high concentration of PM particles and the number of COVID-19 deaths (12). At a nanoscale, the viral load and virulence through the SARS-CoV-2 spike proteins and ACE2 human cells receptors fusion mechanisms, correlate directly with infectivity, disease phenotype, morbidity, and mortality, which should be accounted for correlating conventional aerosol metrics, such as PM2.5 or PM10 concentrations to COVID-19 cases, or deaths, and provide a direct explanation of the variability of the COVID-19 severity through the Case Fatality Ratio (13). CFR has also been linked to the viremia (14, 15), a measure of the quantity of virions emitted by infected patients in the form of aerosolized droplets that can circulate both indoors and outdoors. Although, CFR is considered a crucial quantity to predict emergency and health care infrastructure deployment to manage the spread of COVID-19, understanding of mechanisms that in an urban setting contribute to the COVID-19 severity and the airborne propagation of SARS-CoV-2 remains limited, subsequently preventing accurate determination of COVID-19 CFR at the scale of neighborhoods, cities or even countries (16–22). Here, we propose a probabilistic model to predict CFR based on 3 distinctive Urban Factors (\( U_i \)): Personal (\( P_i \)), Indoor (\( I_i \)), and Outdoor (\( O_i \)). We find that while the approach works best for city-wide predications (\( N = 30 \) cities worldwide), we can also successfully utilize it at other scales, ranging from zip codes or city districts (Fig. S1) to county or even country levels.
Results

The main component of the CFR model relies on the data from the first year of the pandemic for 30 cities worldwide (Fig. 1, Table S1) at various geospatial scales. Since at the very beginning of the pandemic there were major challenges in reporting values of infected people, $C19_c$, we chose a minimal time window of six months, which is sufficient to study the first wave of the COVID-19 virus and investigate CFR correlations between $C19_c$ and the number of deaths, $C19_d$ in the form of:

$$C19_d = a \times C19_c^k$$

(Eq. 1)

where, $a$ is an algebraic proportionality constant and $k$ is an exponent factor, which for COVID-19 has been $\equiv 1$ (Fig. S2). We began evaluating $a = CFR = C19_d/C19_c$ at a local scale for city wards and districts (Fig. S1, Table S2-S3) in 4 US cities: New-York NY (Figs. 2a-b), Los Angeles CA, Chicago IL and Seattle WA, during the first COVID-19 wave ($\lambda_1$ factor) from March 2020. Although, similar data for the wards of Washington D.C. is available and has been used throughout this study, we chose not to use it in the model optimization process due to low (<100,000) population of D.C. wards but rather keep this information for validation purpose at the wards scale. While for most regions available data (Table S2) extends beyond the first 6-12 months of the pandemic and captures multiple waves, herein defined $\lambda_1$ captures the earliest and most severe time frame of the pandemic. Upon investigating time series of CFRs, we find that during $\lambda_1$ they resemble sigmoid functions. To eliminate time component, we use average $CFR(\lambda_1)$ obtained directly from upper plateau range of CFR distribution (Fig. 2a). Although, one can use the steepest slope (Fig. 2b) or the maximum value (Fig. 2a) of CFR distribution to quantify its severity cities districts and cities ($N = 50$, Figs. 2c-d), $CFR(\lambda_1)$ provides the most stable averaging approach that works with inconstancies and fluctuations in raw COVID-19 data. With such defined $CFR(\lambda_1)$ we proceeded to define Personal ($P_i$), Outdoor ($O_i$) and Indoor ($I_i$) Urban Factors, $U_i$s (Fig. 3a), where $i$ represents a geographical location, which to formulate the model is limited to wards or districts of cities.

Figure 2. COVID-19 CFR methodology showing (a) $CFR$ time series with examples for calculating $CFR(\lambda_1)$ and $CFR_m(\lambda_1)$ and (b) $CFR_s(\lambda_1)$ using Eq.1 for the first wave of the pandemic in New York City, NY. Linear correlation between (c) $CFR(\lambda_1)$ vs. $CFR_s(\lambda_1)$ and (d) $CFR(\lambda_1)$ vs. $CFR_m(\lambda_1)$ for city districts ($N = 20$, Table S4) and cities ($N = 30$, Table S5).
To identify differences in the COVID-19 spread in each location as captured with CFR, we adopted city texture methodology (see Materials and Methods, SI), among other variable used in the determination of Urban Factors (Tables. S4-S8). However, for the calibration of weight parameters in equations 3-5 only city district ($N = 20$) from Table S3 have been used to allow a sufficient sample of uncorrelated data from various geographical regions to be used for validation of the model ($N = 118$ for cities, zip codes and boroughs, including 9 outliers).

To correct measured CFR values, the dimensionless weight factors derived from $CFR(\lambda_1)$ for 20 city districts from 4 major US cities were adjusted using a theoretical model $a_i$, defined as the $CFR(\lambda_1)$ for region $i$ during time window $\lambda_1$ (first pandemic: March-August 2020, Fig. 4a):

$$a_i = \frac{a_0}{[1 - P_i][1 - I_i][1 - O_i]}$$

(Eq. 2)

where, $a_0$ is a constant indicative of intrinsic $CFR(\lambda_1)$ (corrected for the urban factors under the form of a product of 3 independent probabilities) during the initial, first wave of the pandemic with no, or minimal, city-wide social distancing and vaccine protocols. Thus, $a_0$ becomes $CFR(\lambda_1)$ that corresponds to natural human response to COVID-19 virus corrected for all type of Urban Factors, a correction derived using independent and uncorrelated $U_i$'s (Fig. 3b): Personal, Indoor and Outdoor Factors, where each $[1 - U_i]$ component reflects the correction of measured $CFR(\lambda_1)$.

![Urban Factor Distribution](image)

**Figure. 3.** (a) Urban Factors with intrinsic $CFR$ values for different waves of the pandemic and model weight parameters optimized for the city districts (SI, Table S4) with (b) their probability distributions.

From the distribution of factors, it can be concluded that age index, captured by $P_i$, has much lower impact on $CFR(\lambda_1)$ than $O_i$ or $I_i$, where their mean values closer to 1 have greater influence over $CFR(\lambda_1)$. With a very low absolute residual sum between predicted and measured $CFR$, which is < 15% of $\sum_{i=1}^{20} CFR_i(\lambda_1)$ (SI) built upon 20 city districts, the $CFR(\lambda_1)$ model results in $a_0 = 0.0089$, which at the global scale is lower than previously measured intrinsic $a_0(\lambda_1) = 0.014$ (22) for single geographical region. While real-time predictability of $a_0$ can be improved by considering a day-to-day and local variability of relative humidity, here it is assumed to be an external constraint affecting the whole urban area using annual average values. To adequately capture risk and severity of $CFR(\lambda_1)$ at the local scale of cities, for zip-code, wards, districts and boroughs values we multiple measured and predicted case fatality ratio values by the fraction of the population of a specific city (Fig. 4b). For cities taken as a whole, this ratio remains unity.
Fig. 4. Urban Factors model predictions showing (a) comparison between measured and predicted $CFR(\lambda_1)$ from Eq.2 for 20 city districts for 4 major US cities (Fig. S1, Table S4) between 03/2020-08/2020 used for optimizing urban factor weight parameters and intrinsic $CFR$. Linear fitting with slope coefficient of unity provides $R^2 = 0.65$ and $RMSE = 0.012$. (b) Comparison between measured and predicted population adjusted $CFR_{pop}(\lambda_i)$ for 118 location worldwide (Tables S4-S9) during different waves of the pandemic (excluded 9 outliers for cities in Table S5 for the correlation calculations). Linear fitting with slope coefficient of unity provides $R^2 = 0.95$ and $RMSE = 0.31$ in the loglog scale. Predicted $CFR$ and $CFR_{pop}$ values used the same urban factor weight parameters in both figures.

Discussion

The ability of our model to extend beyond city districts scale was then demonstrated at the global city scale with no adjustment of urban factor weight and $a_0$ parameters for 30 cities and 88 smaller scale locations (i.e. zip-codes, wards) worldwide. Regardless of the scale we find that very accurate population adjusted $CFR_{pop}(\lambda_i)$ predictions ($R^2 = 0.95$) can be made for any place in the world, provided that no extreme countermeasures were adopted at the beginning of the pandemic. Furthermore, the Urban Factors model from Eq.2 can be applied to make prediction over different time scales of the pandemic. With additional cumulative $C19_c$ and $C19_d$ data between the start of the pandemic and third wave (March 2020 - June 2021) and forth wave (March 2020 - January 2022) respectively, and keeping all $w$'s parameters constant (Fig. 2), we could solely adjust $a_0(\lambda_i)$ to $a_0(\lambda_3) = 0.002$ and $a_0(\lambda_4) = 0.001$ respectively while maintaining acceptable accuracy of predictions. The sharp decline in $a_0(\lambda_i)$ values over time reflects the immense effect of the vaccine campaign on the severity of the pandemic and the transition to virus variants being less harmful in terms of severe cases (i.e. deaths). The variability in $a_0(\lambda_i)$ is expected as over the course of the entire pandemic, with yearly average $HR$ values, city layouts and census data exposed to minimal changes, $w$ weights parameters should not vary to a noticeable extent. It is also important to note that the model significantly overestimated $CFR(\lambda_1)$ for 9 cities (Fig. 4a, Table S5) where measured $CFR(\lambda_1)$ were very close to $a_0(\lambda_1)$. Upon further investigation, these cities were found to be either isolated territories (such as Maldives or Malta, which based on size and population are treated as cities) or quick to adopt social distancing and/or lockdown measures (as in San Francisco CA, Berlin in Germany, or Singapore) during the first wave, $\lambda_1$, of the COVID-19 pandemic. We can conclude therefore that measured $CFR$ as predicted by the coronavirus severity equation (Eq.2)
reflects the absence or inefficiency of sanitary measures taken by local and/or national authorities to manage the first wave of COVID-19.

Multi-spatial scale application of the model with no adjustment to its weight factors allows us to study the impact of specific variables that make up the three distinctive urban factors. Relative humidity takes 90.5% of \( O_i \), while \( \phi \) (characterizing city texture, see SI) is responsible for the remaining 9.5%. High percentage value of \( H_p \) supports the airborne nature of pandemic transmission underlying the role of evaporation and thermodynamic conditions of the surrounding atmosphere (7, 9). The effect of city texture parameter \( \phi \) can be linked to persistence of active outdoor forms of SARS-CoV-2 previously related to urbanization and particulate atmospheric pollution (13, 23). This also suggests that ordered districts in a given city (i.e. Brooklyn in NYC) are more likely to retain laden droplets than less ordered ones (i.e. Bronx in NYC). Indoor Factor, defined by the household size and density of units have similar contribution, 54% and 46%, respectively, which can be explained by the close-range contaminations and circulation of the virus not just inside units, but also in between them due air fluxes and recirculation (24, 25) and its presence on common surfaces and spaces such as elevators, stair cases, lobbies (8, 26). Since the higher density of inhabitants per housing unit results in higher probability of CFR, this work reveals an aspect of social discriminations linked to wealth and living standards. Finally, the Personal Factor is derived solely from population older than 65 years old, which suggests that the current healthcare systems in place across the 30 cities we considered, were able to respond to patient’s needs regardless of their income.

Reduction in global \( a_0 \) by a factor 4 (from 0.0089 to 0.0022) during the first three waves followed by a further reduction by 2.8 between the third and fourth waves (from 0.0022 to 0.0008) reflects the joint positive impact of social distancing, hygiene measures and above all vaccination combined with the virus mutations to less harmful variants (i.e. Omicron and subsequent more recent versions of the virus). Each \( U_i \) changes predicted CFR according to expected norms that is an increase in relative humidity, density of urban dwellers or the proportion of population above 65 years of age, all increase CFR values. It is demonstrated that the approach remains operative at the global city scale making model fit predictions with \( R^2 = 0.95 \) and a low RMSE = 0.31 population adjusted CFR\(_{pop}\). It is interesting to note that averaging measured CFRe, across all US and UK cities considered in this work, gives comparable to cumulative at country scale values for CFR(\( \lambda_3 \)): 2.0% vs 1.8% (US) and 2.7% vs. 2.7% (UK) and CFR(\( \lambda_4 \)): 1.1% vs 1.2% (US) and 0.9% vs. 1.0% (UK) reflecting the fact that the COVID-19 pandemic is mainly an urban phenomenon.

In the field of epidemiology, it is often considered that the CFR has limited time validity (usually a few months) and is spatially restricted to a country level. This work demonstrates, however, that the CFR carries a multiscale length and time scale stability and thus not only can be used to infer disease severity in both real time and during the entire pandemic. It can also be averaged by location from the zip code to city scales worldwide. Therefore, predicting the severity of COVID-19 using CFR provides new dynamic means for evaluating cities in real-time with changing urban factors to equip decision makers and public authorities with quantifiable data to optimize and target their seasonal response to airborne infectious diseases, such as COVID-19 at any geo-spatial scale.

**Materials and Methods**

For all input data, we resort to using publicly available building footprints, weather, and census data (Tables S1-S2). We define the first \( U_i \), Personal Factor as:

\[
P_i = w_A \frac{N_{A,i}}{N_{0,i}} + w_I \frac{N_{I,i}}{N_{0,i}} \quad \{0 \leq w_A, w_I \leq 1 \quad w_A + w_I = 1\}
\]

(Eq. 3)
where \( w_A \) is a weight factor for inhabitants of age, \((A, i) \geq 65\) years old, \( N_{A,i} \) is the total number of people \( A, i \geq 65 \), \( w_I \) is a weight factor for an annual income, \( I, i \leq \$50,000 \), \( N_{I,i} \) is the total number of people with \( I, i \leq \$50,000 \), and \( N_{0,i} \) is the total number of people living in the region \( i \). The second \( U_i \), Indoor Factor is defined as:

\[
U_i = w_{\rho_b} [1 - \rho_b^{-1}] + w_{\rho_r} [1 - \rho_r^{-1}] \quad \{0 \leq w_{\rho_b}, w_{\rho_r} \leq 1\, w_{\rho_b} + w_{\rho_r} = 1\}
\]

(Eq. 4)

where \( w_{\rho_b} \) is a weight factor for building density, \( \rho_{b,i} \) (number of housing units per building), and \( w_{\rho_r} \) is a weight factor for resident density, \( \rho_{r,i} \) (household size, which is the number of residents per housing unit) in region \( i \). The final \( U_i \), Outdoor Factor is calculated as:

\[
O_i = w_{\rho} \rho_i + w_{\varphi} \varphi_i + w_{H_R} H_{R,i} \quad \{0 \leq w_{\rho}, w_{\varphi}, w_{H_R} \leq 1\, w_{\rho} + w_{\varphi} + w_{H_R} = 1\}
\]

(Eq. 5)

where \( w_{\rho} \) is a weight factor for the planar density of building footprints, \( \rho_i \) (ratio between total area of buildings and area of region \( i \)), \( w_{\varphi} \) is a weight factor for the order parameter of buildings, \( \varphi_i \), and \( w_{H_R} \) is a weight factor for the relative humidity \( H_{R,i} \), which is assumed constant at the city-scale, except for New York, Los Angeles and Chicago, where city district-level \( H_{R,i} \) was available. The buildup surface area and the local order parameter, \( \varphi_i \) reflects the local urban texture of the first shell of neighboring building with respect to a central one and is used to quantify 2-D angular order of a local urban texture obtained from the 2-D building-building pair correlation function, \( g(r) \) (\((27), \text{ SI}\)), which is defined as:

\[
g(r) = \frac{1}{N} \sum_{i=1}^{N} \frac{n_i(r + dr) - n_i(r)}{\rho_{\text{city}} 2 \pi r \times dr}
\]

(Eq. 6)

where \( n_i(r) \) denotes the number of buildings within the radial distance \( r \) from building \( i \), and \( dr \) is distance increment, which for \( g(r) \) calculations we chose to be 5% of the average building size, \( L \). From the distribution of \( g(r) \) we can directly obtain coordination number, \( C_n \) (i.e. average number of neighboring buildings) using its integral in the form:

\[
C_n = 2 \pi \rho_{\text{city}} \int_0^{r_{\text{min}}} g(r) dr
\]

(Eq. 7)

The approach for selecting \( C_n \) should be evaluated based on its application, which depending on desired accuracy may lead to different results. Here, the application is defining average local configuration of buildings as captured by \( g(r) \) thus leading to our use of Eq. 7, where \( r_{\text{min}} \) is the first local minimum in the \( g(r) \) distribution following the main peak (i.e. one with the maximum \( g(r) \) value). Due to variability in local city texture between zip codes or wards within a given city, integral
of the first peak may not always lead to $C_n > 1$, which is required to be able to obtain angular order parameter with a minimum configuration of 3 buildings, or 2 neighbors. If such situation exists, we proceed with the next local minimum in the distribution of $g(r)$ until $C_n > 1$ has been obtained. With such defined $C_n$ we proceed with calculations of the second city texture value, order parameter. Exact $r_{\text{min}}$ values used to define the first peak are presented in Tables S1-S3, with $g(r)$ functions visualized in figs. S2-S4. Applied at the city scale, we use the absolute values of $C_n$, to find $\phi$, which characterizes the average angular distortion of buildings compared to a perfect angular local order of a city at fixed $m = C_n$ within the first shell distance determined from the integral of $g(r)$ (Eq. 7):

$$\phi = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{N_a(j)} \left( \sum_{k=1}^{N_a(j)} \exp (i C_n \vartheta_k) \right)$$

(Eq. 8)

where, $N$ is the number of buildings.

The model parameters have been optimized by minimizing the error between predicted and measured $CFR(\lambda_1)$ values using standard statistical methods. Although, it is common to apply sum of square errors when quantifying errors in regression analyses, here we adopted a sum of absolute errors, $\varepsilon$ in the form of:

$$\varepsilon = \sum_{i=1}^{n=20} |CFR_i(\lambda_1) - a_i(\lambda_1)|$$

(Eq. 9)

where $i$ is the city district ID from Table S4. Since $CFR$ values are fractions, differences between any predicted and measured values lead to small fractions and thus squaring the difference would lead to even smaller values thereby introducing bias to any optimization approach that is trying to minimize the sum of errors. As such the absolute sum $\varepsilon$ provides a more stable optimization parameter. We minimized $\varepsilon$ using a non-linear generalized reduced gradient (GRD) method with constraint convergence of 0.00001, forward derivate population size of 1000. With such defined parameters, we obtained optimized urban factors and intrinsic $CFR$, $a_0(\lambda_1)$ (Fig. 3).

**Acknowledgments**

We acknowledge CNRS-Innovation for the SURe project funding (www.sure-tech.fr) through the pre-maturation and RISE programs.

**References**


