EClinicalMedicine 26 (2020) 100479

Contents lists available at ScienceDirect

EClinicalMedicine

journal homepage: https://www.journals.elsevier.com/eclinicalmedicine

Research Paper

Using the COVID-19 to influenza ratio to estimate early pandemic spread in Wuhan, China and Seattle, US

Zhanwei Du[a], Emily Javan[a], Ciara Nugent[a], Benjamin J. Cowling[b], Lauren Ancel Meyers[a,c,*]

[a] The University of Texas at Austin, Austin, TX 78712, United States
[b] The University of Hong Kong, 7 Sassoon Rd, Hong Kong SAR, China
[c] Santa Fe Institute, Santa Fe, NM, United States

ARTICLE INFO

Article History:
Received 3 June 2020
Revised 14 July 2020
Accepted 14 July 2020
Available online 12 August 2020

Keywords:
COVID-19
Pediatric infections
Adult infections
Influenza
Wuhan
Seattle

ABSTRACT

Background: Pandemic SARS-CoV-2 was first reported in Wuhan, China on December 31, 2019. Twenty-one days later, the US identified its first case––a man who had traveled from Wuhan to the state of Washington. Recent studies in the Wuhan and Seattle metropolitan areas retrospectively tested samples taken from patients with COVID-like symptoms. In the Wuhan study, there were 4 SARS-CoV-2 positives and 7 influenza positives out of 26 adults outpatients who sought care for influenza-like-illness at two central hospitals prior to January 12, 2020. The Seattle study reported 25 SARS-CoV-2 positives and 442 influenza positives out of 2353 children and adults who reported acute respiratory illness prior to March 9, 2020. Here, we use these findings to extrapolate the early prevalence of symptomatic COVID-19 in Wuhan and Seattle.
Methods: For each city, we estimate the ratio of COVID-19 to influenza infections from the retrospective testing data and estimate the age-specific prevalence of influenza from surveillance reports during the same time period. Combining these, we approximate the total number of symptomatic COVID-19 infections.
Findings: In Wuhan, there were an estimated 1386 [95% CrI: 420-3793] symptomatic cases over 30 of COVID-19 between December 30, 2019 and January 12, 2020. In Seattle, we estimate that 2268 [95% CrI: 498, 6069] children under 18 and 4367 [95% CrI: 2776, 6526] adults were symptomatically infected between February 24 and March 9, 2020. We also find that the initial pandemic wave in Wuhan likely originated with a single infected case who developed symptoms sometime between October 26 and December 13, 2019; in Seattle, the seeding likely occurred between December 25, 2019 and January 15, 2020.
Interpretation: The spread of COVID-19 in Wuhan and Seattle was far more extensive than initially reported. The virus likely spread for months in Wuhan before the lockdown. Given that COVID-19 appears to be overwhelmingly mild in children, our high estimate for symptomatic pediatric cases in Seattle suggests that there may have been thousands more mild cases at the time.

1. Introduction

On December 31, 2019, a novel coronavirus (SARS-CoV-2) was identified in Wuhan, China. Three weeks later, on January 21st, the US Centers for Disease Control and Prevention (CDC) confirmed the first case of COVID-19 in the US. On January 15th, the man returned from a visit to Wuhan, China to Snohomish County in the Seattle Metropolitan Area of Washington state [1]. To mitigate local transmission and prevent global spread, China imposed a lockdown on Wuhan starting January 23rd. In the first months of the pandemic, confirmed case counts vastly unrepresented the rapid expansion of the pandemic as countries raced to ramp up testing and surveillance capabilities [2–5]. By the time of the Wuhan lockdown, only 571 cases of COVID-19 were reported in mainland China [6], 422 of which were in Wuhan [7]. The Seattle area reported only 245 confirmed COVID-19 cases and 36 COVID-19 deaths by March 9th [8].

Two studies––one in Wuhan [9] and the other in Seattle [10]––re-examined swabs taken from individuals with symptoms of acute respiratory illness during periods where SARS-CoV-2 may have been spreading undetected. Although some of these specimens were previously tested for influenza viruses, none were tested for SARS-CoV-2. The Wuhan study tested 26 throat swabs taken from adults over age 30 who sought outpatient care at one of two central Wuhan hospitals for influenza-like-illness (ILI) between December 30, 2019 and January 12, 2020 [9]. Although no patients were confirmed COVID-19 cases, four retrospectively tested positive for the virus. In addition to

We acknowledge support from NIH grant U01 GM087791 and Tito's Handmade Vodka.
* Corresponding author at: The University of Texas at Austin, Austin, TX 78712, United States.
E-mail address: laurenmeyers@austin.utexas.edu (L.A. Meyers).

https://doi.org/10.1016/j.eclinm.2020.100479

Research in Context section

Evidence before this study

The early pace and extent of the COVID-19 pandemic remains unclear. Given that many countries are still scrambling to provide wide access to coronavirus tests, confirmed case counts underestimate the true prevalence of the virus. Recent studies suggest that SARS-COV-2 may have spread extensively in both Wuhan (China) and Seattle, Washington (US) before the first community-acquired cases were reported in each city.

Added value of this study

We introduce a new method for indirectly gaging the early spread of COVID-19 based on two pieces of information−−the concurrent prevalence of influenza and the ratio of SARS-CoV-2 positive to influenza positive tests among patients with clinical respiratory illness. We apply the method to estimate the dates of emergence and prevalence of COVID-19 in Wuhan prior to the January 23, 2020 lockdown and in Seattle prior to March 9, 2020.

Implications of all the available evidence

Given the epidemiological similarities between influenza and COVID-19, influenza surveillance data can provide a retrospective window into the emergence of COVID-19 in cities around the globe. In both the Wuhan and Seattle metropolitan areas, there were likely thousands of undetected cases of COVID-19 during the first months of transmission. The large discrepancy between confirmed cases and true prevalence of the virus highlights the difficulty of determining infection fatality rates from readily available COVID-19 data.

the four COVID-19 positive samples, seven others tested positive for influenza.

The Seattle study performed RT-PCR tests for SARS-CoV2 and influenza on 2353 mid-nasal swabs collected from 299 children under 18 and 2054 adults who reported symptoms of acute respiratory illness (ARI) between January 1, 2020 and March 9, 2020 [10]. Of these, 442 tested positive for influenza, 25 tested positive for COVID-19, and none tested positive for both viruses.

We note that the two studies have overlapping but not identical case definitions. In Seattle [10], ARI cases had at least two of these symptoms: “feeling feverish, headache, sore throat or itchy/scratchy throat, nausea or vomiting, rhinorrhea, fatigue, myalgia, dyspnea, diarrhea, ear pain or ear discharge, rash, or a new or worsening acute cough alone”. In Wuhan [9], ILI cases included patients reporting fever (with a temperature of at least 100°F/37.8 °C) and a cough or a sore throat without a known cause other than influenza [11].

Our study is premised on the assumption that influenza and SARS-CoV-2 were constrained by similar behavioral and environmental factors in early 2020. The two viruses have overlapping natural histories [12,13] and modes of transmission [13]. Both are respiratory pathogens with a wide spectrum of illness, from asymptomatic to fatal, with severity that depends on age and underlying conditions. They are similarly transmitted from person-to-person through direct contact, droplets and fomites [13−15]. Thus, we expect that once SARS-CoV-2 got a foothold in a city, spreading across multiple communities, its geographic and demographic patterns might mirror those of influenza. In Hong Kong, for example, COVID-19 interventions concurrently reduced the transmission rates (i.e., the daily reproduction number, Rt) of COVID-19 and influenza in early February 2020 [15].

Here, we estimate the early prevalence of symptomatic COVID-19 cases in Wuhan and Seattle based on the ratio of SARS-CoV-2 to influenza test positivity (henceforth, the *covid-to-influenza ratio*) and the local prevalence of influenza in the two cities at the time of the corresponding retrospective study. We derive our estimates of covid-to-influenza positivity directly from the two studies and our estimates of local influenza prevalence from Chinese and US surveillance data.

2. Methods

2.1. Data

2.1.1. COVID-19 and influenza data in Wuhan

To estimate the covid-to-influenza ratio, we used the numbers of COVID-19 positive and influenza positive patients among tested ILI throat swab samples at two hospitals from December 30, 2019 to January 12, 2020 reported by a recent retrospective study [9]. Wuhan has almost 400 hospitals, which collectively have 81,700 beds and 81 million outpatient visits per year [16]. The data we analyzed from ref. [9] were collected from two hospitals that have large and representative catchments: Children’s Hospital of Wuhan (the largest pediatric healthcare center in Wuhan that serves both women and children) [17,18] with 2000 beds and 1.9 million annual outpatient visits and Wuhan No. 1 Hospital [19], with over 3000 beds and 2 million annual outpatient visits. Both serve as sentinel sites in China’s national influenza surveillance system [9]. Together they provide almost 5% of outpatient care in the Wuhan area. The data we analyzed from ref [9]. were collected from two hospitals that have large and representative catchments: Children’s Hospital of Wuhan (the largest pediatric healthcare center in Wuhan, serving both children and adults) [17,18] with 2000 beds and 1.9 million annual outpatient visits and Wuhan No. 1 Hospital [19], with over 3000 beds and 2 million annual outpatient visits. Both serve as sentinel sites in China’s national influenza surveillance system[9]. The SARS-CoV-2 and influenza virus among tested ILI throat swab samples are well kept at −70 °C before the SARS-CoV-2 experiments and detected by real-time PCR with reverse transcription[9].

To estimate the age-stratified numbers of outpatient visits for ILI in Wuhan, we analyzed data from China CDC weekly reports for Wuhan, December 30, 2019-January 12, 2020 [9]. To estimate the age-stratified population sizes of Wuhan’s 13 districts, we obtained data from the Sixth National Census of the People's Republic of China in 2010 [20], and scaled by the growth in overall Wuhan population between 2010 and 2019 reported by Wuhan Statistics Bureau [21] .

2.1.2. COVID-19 and influenza data in Seattle

Our analysis of Seattle is restricted to the portion of the metropolitan area sampled by the Seattle Flu Study in ref. [10]. Specifically, we analyze King county, which contains the city of Seattle, and Snohomish county, where the first US COVID-19 case was identified. Roughly 77% of the 3.5 million metropolitan population reside in the two counties.

To estimate the covid-to-influenza ratio, we used the numbers of COVID-19 positive and influenza positive patients among tested mid-nasal swab samples from participants with symptoms of acute respiratory illness (ARI) in the Seattle Flu pandemic surveillance platform from January 1, 2020 to March 9, 2020 [10]. Our analysis combines viral positivity data from cases with ILI and ARI. We assume that the two populations are the same−individuals with ILI and ARI in Seattle during the study period−and refer to this population as ILI throughout the text and supplement. The ARI case definition in ref. [10] is at least “two of the following: feeling feverish, headache, sore throat or itchy/scratchy throat, nausea or vomiting, rhinorrhea, fatigue, myalgia, dyspnea, diarrhea, ear pain or ear discharge, rash, or a new or worsening acute cough alone”. The CDC’s case definition for ILI is “fever (temperature of 100°F [37.8 °C] or greater) and a cough and/or

a sore throat without a known cause other than influenza"[11]. Thus, the case definitions overlap considerably, but are not identical. The tested mid-nasal swab samples were kept at 4 °C before the influenza and SARS-CoV-2 tests by TaqMan PT-PCR, with an average time from nasal swab collection to receipt at the study laboratory of 2.8 days [10].

We analyzed the age-stratified numbers of outpatient visits for ILI in HHS region 10 between January 1, 2020 and March 9, 2020 available on the CDC's FluView interactive website [22] and the age-stratified population sizes of the 22 Public Use Microdata Areas (PUMA's) in King and Snohomish counties [10,20]. Details are provided in Table 1.

3. Method

Our methods for estimating the prevalence of symptomatic COVID-19 in Wuhan and Seattle are similar, but not identical. We describe our method for Wuhan in this section and our method for Seattle in the Appendix. The key methodological difference is that the retrospective study in Wuhan [9] but not Seattle [10] reported the date of symptom onset for each positive influenza test. For Seattle, we took the extra step of estimating these dates based on the total number of positives and the daily influenza positivity reported by the CDC for HHS Region 10 (Supplementary Figure S1, Tables S1 and S2).

For Wuhan, we assume that the age-specific risks of COVID-19 and influenza infection are identical in all 13 central districts of the city. Therefore, the ratio of COVID-19 to influenza adult outpatients (r) estimated from the subset of outpatients sampled in ref. [9] can be used to estimate the number of COVID-19 infections across all of central Wuhan (Fig. 3).

3.1. Estimating COVID-19 adult infections in Wuhan

To estimate the number of COVID-19 infections we use a binomial distribution, denoted *B(N, p)*, where N is the total population in each district and p is an estimate of the age specific prevalence of symptomatic COVID-19 in the population adjusted by the proportion of individuals in that age group. We chose a binomial distribution as it is the most commonly used distribution to statistically model case counts when the population size and probability are known. We denote by $H_{d,\,\alpha,\,\tau}$ the number of COVID-19 infections in district d and age range α (over 30 years) during the focal fourteen-day period τ, and model it as:

$$H_{d,\alpha,\tau} \mid N_d, \Theta_\tau^\alpha, \Omega_\tau, \Phi_\tau, r \sim B\left(N_d, \frac{\Theta_\tau^\alpha}{\Omega_\tau} \cdot \Phi_\tau \cdot r\right)$$

where N_d is the number of people of all ages in district d; Θ_τ^α is the number of ILI outpatients in age group α over a period of time τ; Ω_τ is the number of all cause outpatients of all ages in Wuhan over a period of time τ; Φ_τ is the percent of influenza tests that are positive in the South Provinces of China during time period τ; the r is the ratio of COVID-19 outpatients to influenza outpatients over age 30. Given the small sample size, we could not reliably estimate COVID-19 prevalence by sex or narrow age brackets.

We take a Bayesian approach using Markov Chain Monte Carlo, where at each iteration we take a draw from the distributions of r and Ω_r, and then use these to draw $H_{d,\,\alpha,\,\tau}$ according to the specified binomial distribution. Since the other parameters are assumed to be known constants, we do not take draws of these parameters; the values of these parameters can be found in Table 1. $H_{d,\,\alpha,\,\tau}$ is then specified by the set of draws, defining a predictive distribution that we use to calculate the mean and credible intervals for the number of COVID-19 infections. We chose a Bayesian approach to allow an intuitive structuring of the model and avoid making assumptions that are not appropriate for our small sample sizes.

To estimate the distribution of r and Ω_r, we first derive r as the following posterior distribution. Let N denote the total number of adults in the sample, and x_c and x_f denote the observed number of adults who tested positive for COVID-19 and influenza, respectively. As before we assume a binomial distribution where

$$x_c \mid p_c, N \sim B(N, p_c) \text{ and } x_f \mid p_f, N \sim B(N, p_f).$$

If we assume uninformative priors on p_c and p_f[23],

$$p_c \sim Beta(1,1) \text{ and } p_f \sim Beta(1,1)$$

then the posterior distributions are known in closed form[23]:

$$p_c \mid x_c, N \sim Beta(1 + x_c, 1 + N - x_c) \sim Beta(5, 23)$$

$$p_f \mid x_f, N \sim Beta(1 + x_f, 1 + N - x_f) \sim Beta(8, 20)$$

We use Markov Chain Monte Carlo to draw from p_c and p_f at each iteration and calculate $r = p_c/p_f$. We combine these draws to obtain the distribution for r. Using this method we estimate that the ratio of COVID-19 to influenza adult hospitalizations across central Wuhan during December 30, 2019 to January 12, 2020 was 0.61 [95% CrI: 0.20−1.64]. We use 10,000 draws and report the medians and 95%

Table 1
Model Parameters and Data Sources. Parameters with an age indicator (a) have separate values for the 30+ age ranges.

Symbol	Description	Values	Sources
$H_{d,\,\alpha,\,\tau}$	Number of COVID-19 outpatients in age group α in district d over time period τ	Estimated	
r	Ratio of ILI outpatients that are COVID-19 positive versus influenza positive (adults over 30)	Age 30+: 0.61 [95% CrI: 0.20−1.64]	Ref. [9]: Of the 26 tested ILI throat swab samples taken from adults over age 30 who sought ILI treatment at two central Wuhan hospitals between December 30, 2019 and January 12, 2020, 7 tested positive for influenza and 4 tested positive for COVID-19. None of the cases tested positive for both viruses.
N_d	Age-stratified population sizes in district d	2010 population scaled by the ratio of the 2019 to the 2010 total Wuhan population	Sixth National Census of the People's Republic of China in 2010 [20] and total population of Wuhan in 2019 (11.08 million) [21]
Ω_τ	Number of outpatient visits (all causes) in Wuhan across all ages over time period τ	42,274 and 38,702 over two weeks, respectively	China CDC weekly reports of outpatient visits in Wuhan, December 30, 2019-January 12, 2020 [9]
Θ_τ^α	Number of ILI outpatients in age group α in Wuhan over time period τ	Age 30+: 61 and 47, for each of the two weeks, respectively	China CDC weekly reports of ILI outpatients in Wuhan, December 30, 2019-January 12, 2020 [9]
Φ_τ	Percent of influenza positive tests	25% and 28.6% for each of the two weeks, respectively	Ref. [9]: 25%, 28.6% adult (30+) influenza positive among 160 ILI throat swab samples, from December 30, 2019 to January 12, 2020.
T_d	Epidemic doubling time	7.3 [95% CrI: 6.3−9.7] days 5.2 [95% CrI: 4.6−6.1] days	Refs. [2] and [24]

credible intervals of the resulting posterior predictive distribution for the number of COVID-19 infections for each district.

3.2. Estimating COVID-19 adult infections prior to the Wuhan lockdown

To project the number of adult infections in Wuhan prior to the closing on January 23, 2020 (H_{cum}), we assume

$$H_{cum} = \sum_{i=t_0}^{L} h_0 \cdot 2^{i/T_d}$$

where T_d is the epidemic doubling time, t_0 is the day of the first adult infection in Wuhan, and L corresponds to January 22, 2020 (the day before the Wuhan lockdown). We use our age- and district-stratified estimates for adult COVID-19 infections for December 30, 2019 to January 12, 2020 to estimate this quantity, under the assumption that the values reflect cumulative incident infections during that fourteen-day period (Fig. 4).

We use Monte Carlo sampling to incorporate the uncertainty in both the epidemic doubling rate in Wuhan during this period [2] and adult infections from December 30, 2019 to January 12, 2020 ($H_{d,\alpha,\tau}$). We take draws from the distribution of $H_{d,\alpha,\tau}$ and T_d (summarized in Table 1) to estimate the time since the first adult infection by

$$\delta = T_d\left(log_2\left(\frac{H_\tau}{\sum_{i=0}^{13} 2^{i/T_d}}\right)\right).$$

That is, the estimated date of the first COVID-19 infection in Wuhan (t_0) is δ days prior to December 30, 2019. We then estimate H_{cum} according to the equation above to project the cumulative COVID-19 adult infections preceding the Wuhan lockdown.

4. Role of funding

This research was made possible, in part, by NIH grant U01 GM087791 and funding from Tito's Handmade Vodka in support of the UT COVID-19 Modeling Consortium. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

5. Results

Estimated symptomatic COVID-19 adult infections in the 13 central districts of Wuhan

Based on the numbers of confirmed SARS-CoV-2 and influenza cases in ref. [9], we estimate that the ratio of symptomatic COVID-19 to influenza infections in Wuhan from December 30, 2019 to January 12, 2020 was 0.61 [95% CrI: 0.20–1.64] for adults over 30. Coupling this ratio with influenza prevalence statistics derived from surveillance data, we estimate there were 1386 [95% CrI: 420-3793] for people over 30 years with symptomatic COVID-19 infections in Wuhan between December 30th and January 12th, ranging from 19 cases [95% CrI: 6-51] in suburban Hannan to 177 cases [95% CrI: 54-485] in central Wuchang (Fig. 1, Table S3). Estimates for the epidemic

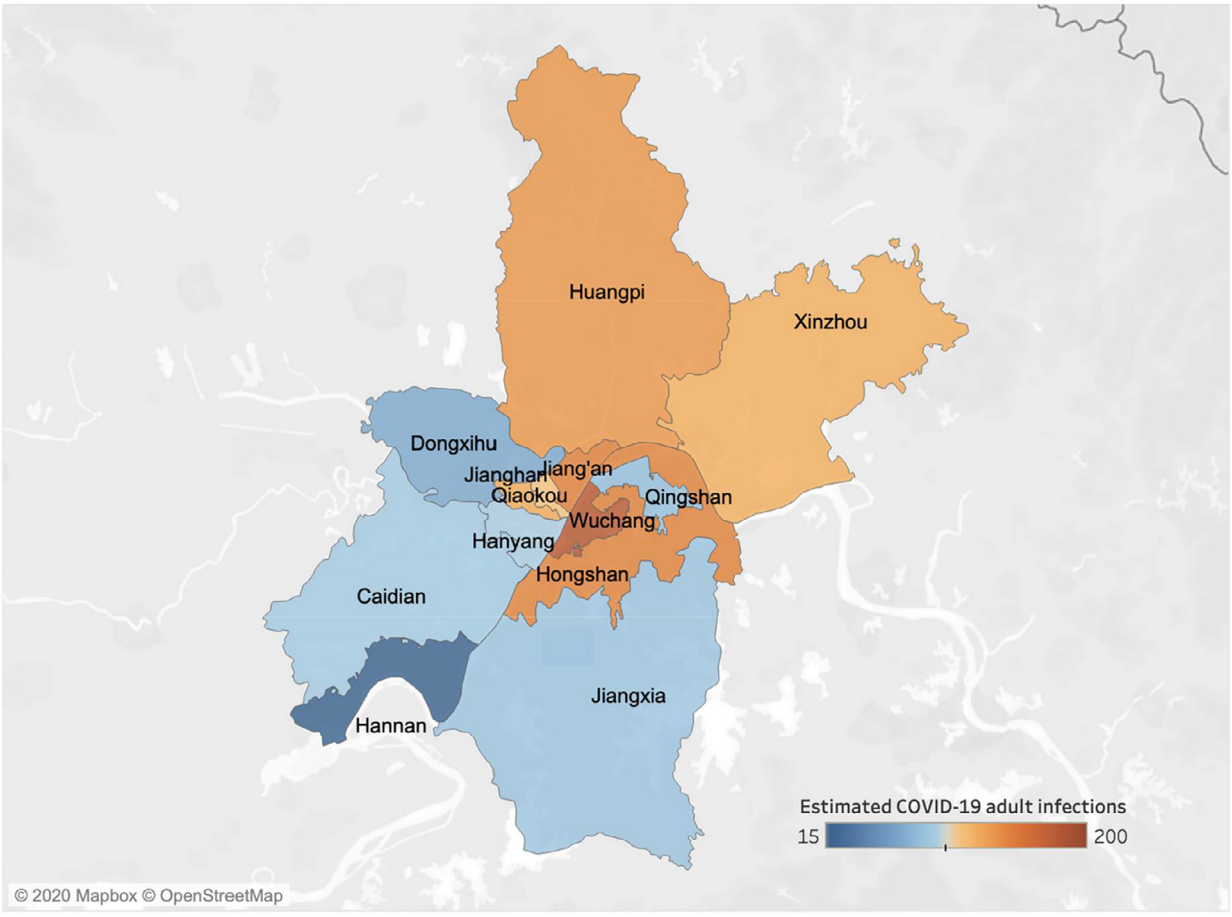

Fig. 1. Estimated symptomatic COVID-19 infections of people over 30 years in the 13 districts of Wuhan from December 30, 2019 to January 12, 2020. A retrospective study identified four ILI cases of COVID-19 from two hospitals in central Wuhan [9]. We estimate that there were a total of 1386 [95% CrI: 420–3793] adult cases of COVID-19 during that 14-day period across the 13 central districts of Wuhan, ranging from 19 cases [95% CrI: 6–51] in suburban Hannan to 177 cases [95% CrI: 54–485] in central Wuchang, as indicated by shading (Table S3).

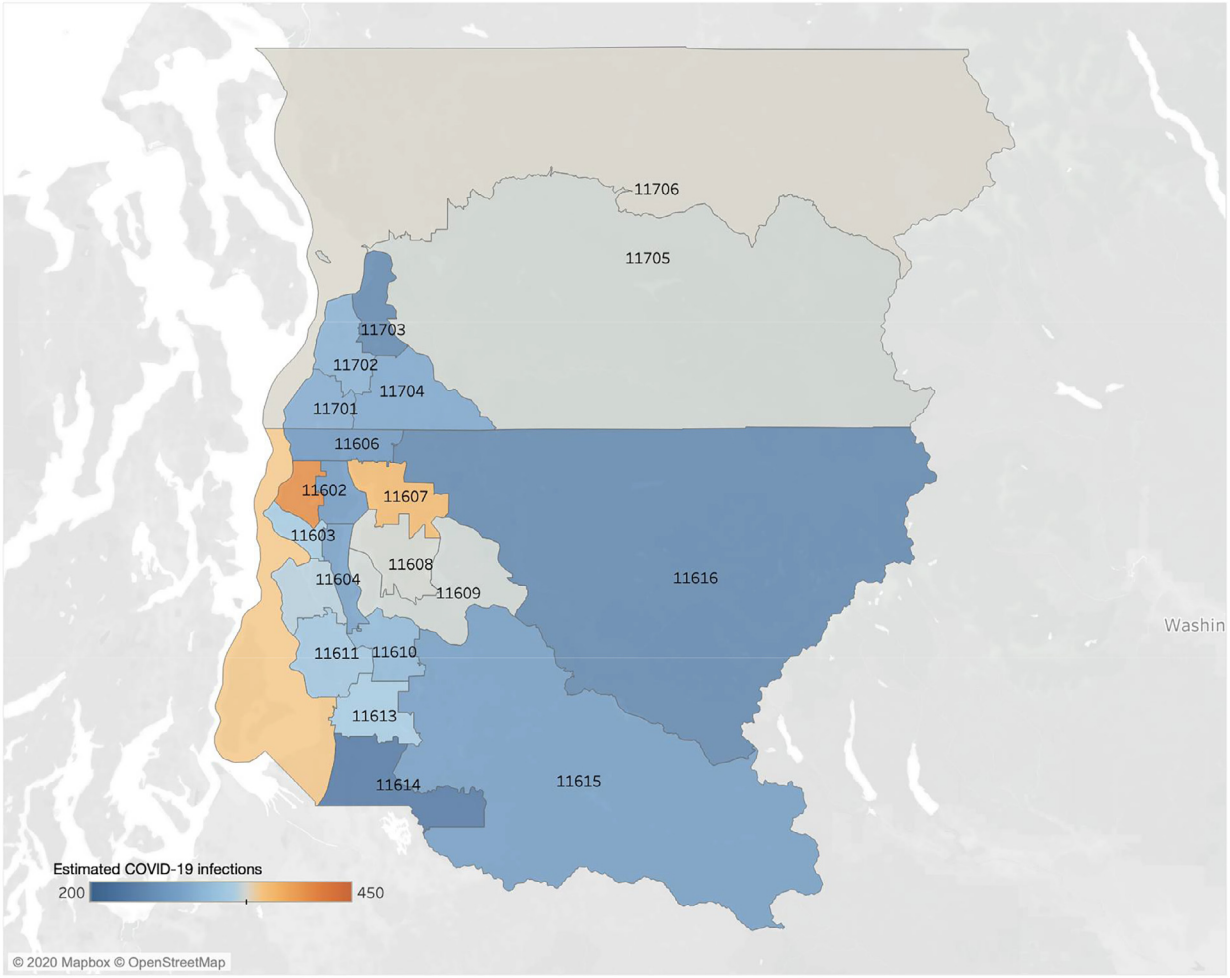

Fig. 2. Estimated symptomatic COVID-19 infections of people over all ages in the 22 studied Public Use Microdata Areas of Seattle from February 24, 2020 to March 9, 2020. A retrospective study identified 2 children and 23 adult ILI cases associated with COVID-19 from the Seattle Flu pandemic surveillance platform [10]. We estimate that there were a total of 6748 [95% CrI: 4133, 11,020] (2268 [95% CrI: 498, 6069] and 4367 [95% CrI: 2776, 6526] for people under and over 18 years, respectively) cases of COVID-19 cases during that 15-day period across the 22 PUMAs of Seattle, ranging from 231 cases [95% CrI:199,265] in PUMA 11,614 to 410 cases [95% CrI:364,459] in PUMA 11,601 (Table S2).

doubling time for COVID-19 in Hubei Province have ranged from 5.2 [24] to 7.31 days [2]. These two values suggest a total of 12,939 [95% CrI: 2728-109,651] or 22,939 [95% CrI: 5034-119,864] symptomatically infected adults over age 30 prior to the January 23rd lockdown, respectively. Both estimates far exceed the 422 documented cases across all age groups [25]. Several studies have estimated that roughly half of infections are asymptomatic [26]. Thus, the number of undetected adult COVID-19 cases at that time may have reached 10,000. We further estimate that the Wuhan epidemic emerged from cases infected around November 17, 2019 [95% CrI: October 26-December 3, 2019] or December 2, 2019 [95% CrI: November 20-December 13, 2019], under the lower or higher reported doubling times, respectively (Fig. 4).

We note that the Wuhan study [9] also tested swabs taken from 54 ILI patients under age 30. Of these, 30 tested positive for influenza and none tested positive for COVID-19. Given that there were likely symptomatic pediatric COVID-19 cases in Wuhan during the study period [27], we do not believe that the true prevalence in this age group was zero. Because estimates close to zero require greater amounts of data to estimate with any certainty, we lack the statistical power to reasonably estimate the COVID-19 to influenza ratio based on the reported zero out of 54 without making additional assumptions. Thus, to avoid potentially problematic assumptions or invalid generalizations, we restricted our analysis to the over 30 age group.

5.1. Estimated symptomatic COVID-19 infections in king and snohomish counties

For the Seattle area, we similarly estimate that the ratio of symptomatic COVID-19 to influenza infections in children under 18 was 0.11 [95% CrI: 0.03–0.33] and in adults was 0.14 [95% CrI: 0.09–0.21] from February 24 and March 9, 2020. Based on this ratio and the concurrent prevalence of influenza in Seattle, we estimate that there were 6748 [95% CrI: 4133, 11,020] symptomatic COVID-19 infections between February 24th and March 9th. The age breakdown is 2268 [95% CrI: 498, 6069] symptomatic cases in children under 18 and 4367 [95% CrI: 2776, 6526] cases in adults. The Seattle Flu Study [10] located the retrospectively detected COVID-19 cases down to the level of Public Use Microdata Areas (PUMA's), which are US Census statistical reporting units. Based simply on the population sizes of the 22 PUMA's in King and Snohomish counties, we estimated that the PUMA-level prevalence of symptomatic COVID-19 cases during this period ranged from 231 cases [95% CrI:199,265] in PUMA 11,614 (Southwest King County) to 410 cases [95% CrI:364,459] in PUMA 11,601 (Northwest Seattle) (Fig. 2).

A prior study estimated COVID-19 had a mean epidemic doubling time in Washington State in January and February 2020 of 6.1 days [90% CrI: 5.1 to 8.2 days] [3]. Under this range of doubling times, we estimate there were a total of 9068 [95% CrI: 8264-10,011] symptomatic COVID-19 infections in Seattle before March 9th. If we assume

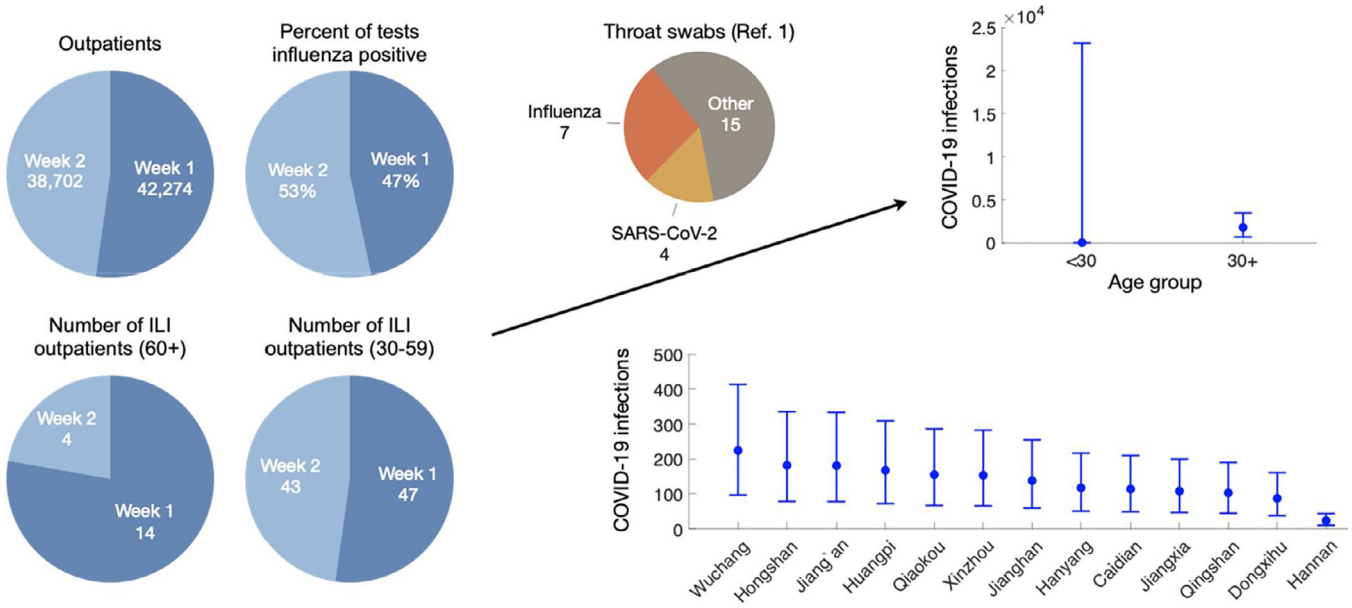

Fig. 3. Estimating adult COVID-19 infections based on the ratio between patients retrospectively testing positive for COVID-19 and influenza in two hospitals in Central Wuhan from December 30, 2019 to January 12, 2020. First we use influenza surveillance data (number of outpatients, percent positive influenza tests, and number of ILI outpatients reported for the Wuhan region by the Chinese CDC) to estimate the proportion of adult outpatients (all cause) testing positive for influenza from December 30, 2019 to January 12, 2020 (left graphs). Second, we estimate the ratio of COVID-19 positive to influenza positive patients among adult outpatients with ILI, based on a recent retrospective study in two Wuhan hospitals (0.61 [95% CrI: 0.20–1.64]) [9]. We then estimate the number of symptomatic COVID-19 infections among adults across Wuhan during this time period based on the proportion of influenza positive outpatients and the ratio of COVID-19 to influenza positive outpatients, using Monte Carlo sampling to incorporate uncertainty in our estimates of both quantities (upper right). Finally, we estimate the age-specific COVID-19 adult infections for the 13 central districts in Wuhan based on the district level population sizes for each age group. Given that the four detected COVID-19 cases lived in central Wuhan in ref. [9], we assumed that risk was uniform across all 13 districts during the 14-day time period. .

50% of infections are asymptomatic [26], then we project there may have been over 15,000 undetected COVID-19 cases at the time. We further estimate that the Seattle epidemic originated with cases that arrived infected around January 6, 2020 [95% CrI: December 25, 2019 - January 15, 2020] (Fig. 4).

6. Discussion

In cities across the Northern Hemisphere, the emergence of the COVID-19 pandemic coincided with the 2019–2020 influenza season [9,10]. Mild COVID-19 and influenza infections have overlapping constellations of symptoms that often fall within the criteria for influenza-like-illness (ILI) and acute respiratory infections (ARI) [28]. Prior to widely available SARS-CoV-2 tests, symptomatic COVID-19 cases who sought care were likely to have been tested for influenza. A few studies have retrieved and retrospectively tested swabs taken from such patients for SARS-CoV-2 and thereby identified early undetected cases of COVID-19 [3,5,9,10]. Given the spatiotemporal overlap and epidemiological similarities between influenza and SARS-CoV-2, we hypothesized that the observed prevalence of influenza might shed light on the unseen early spread of COVID-19. To extrapolate COVID-19 prevalence from influenza surveillance data, we assume that the

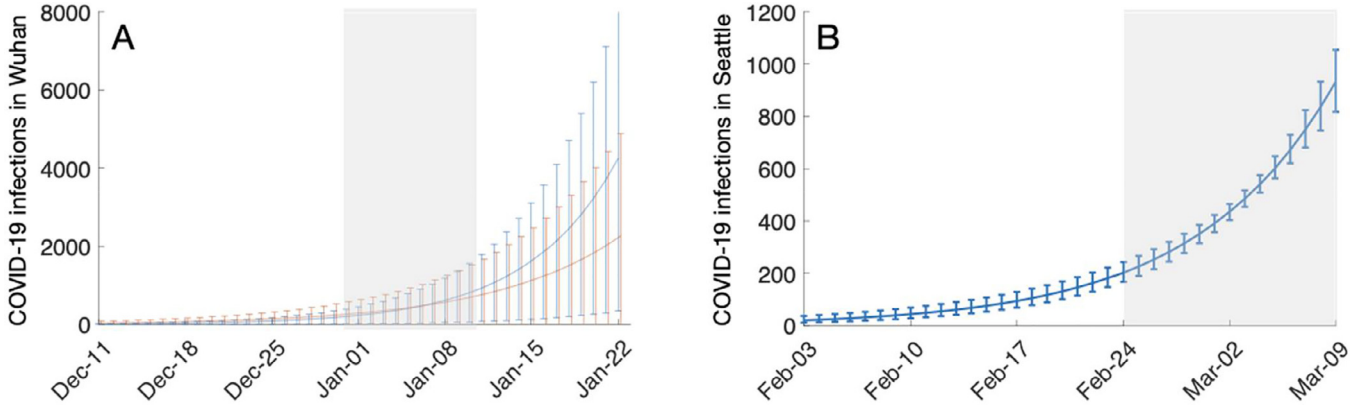

Fig. 4. Estimating the number of symptomatic COVID-19 cases among all age groups in Wuhan prior to January 23, 2020 and all age groups in Seattle prior to March 9, 2020. (A) For Wuhan, we assume an epidemic doubling time of either 7.3 [95% CrI: 6.3–9.7] days (red) or 5.2 [95% CrI: 4.6–6.1] days (blue). We further assume the numbers of COVID-19 infections estimated for the 13 central districts (Table S3) are equal to the sum of the daily number of incident infections from December 30, 2019 to January 12, 2020. Using an exponential model of epidemic growth we estimate that the first COVID-19 infection occurred on (red) November 17, 2019 [95% CrI: October 26-December 3, 2019] or (blue) December 2, 2019 [95% CrI: November 20-December 13, 2019], and then project the daily COVID-19 infections until January 23, 2020. (B) For Seattle, we assume an epidemic doubling time of 6.1 [90% uncertainty interval of 5.1 to 8.2] days [3] and that the numbers of COVID-19 infections estimated across the 22 PUMA's are equal to the sum of the daily number of incident infections from February 24th to March 9th, 2020. Using an exponential model of epidemic growth we estimate the initial pandemic wave in Seattle originated with a single infected case who developed symptoms on January 6, 2020 [95% CrI: December 25, 2019 - January 15, 2020] and then project the daily COVID-19 infections until March 9, 2020. In both graphs, lines and bars indicate the median and 95% CrI estimates, respectively. Gray shading indicates the time period of our initial estimates.

ratio of COVID-19 positive to influenza positive cases detected retrospectively in small samples generally holds for the surrounding metropolitan area.

We analyzed data provided by two studies – one in Wuhan [9] and the other in Seattle [10] – that re-tested swabs taken from ILI and ARI cases in early 2020. The identification of overlooked COVID-19 cases in both cities was not surprising, given the large numbers of cases, hospitalizations and deaths that were detected shortly after these retrospective periods. Nonetheless, the ratios of SARS-CoV-2 to influenza positive swabs were surprisingly high. In Wuhan, there were roughly two symptomatic cases of COVID-19 for every three cases of influenza; in Seattle, there was one pediatric case of symptomatic COVID-19 per every 9 influenza cases, and one per every seven in adults. Given that influenza was circulating widely at the time of these infections, these ratios led us to conclude that there may have been over 5000 undetected cases of symptomatic COVID-19 both in Wuhan prior to January 12th and in Seattle prior to March 9th.

Our results do not imply that health authorities were aware of these undocumented infections, rather that they went unseen during the early and uncertain stages of COVID-19 emergence in the two cities. In Wuhan, other data have suggested similar levels of *unseen* COVID-19 prior to the January 23, 2020 lockdown of the city. For example, we previously estimated that there were 12,400 (95% CrI 3112-58,465) total cases based on extrapolation from the timing and location of the first 19 COVID-19 cases imported from Wuhan to other countries [2]. These numbers are further corroborated by a similarly-derived estimate from Imperial College of 4000 (1000–9700) cases as of January 18, 2020 [29]. Our estimate that the epidemic in Wuhan started in mid to late November of 2019 is consistent with the first known case reporting symptoms starting December 1, 2019 [30].

In Seattle, we estimate that sustained community transmission of SARS-CoV-2 began in early January (Fig. 4), around the time of the first confirmed case [1]. Two recent phylogenetic studies using SARS-CoV-2 genomic data provide conflicting backcasts. The first suggests that a locally-infected case detected on February 24th could be traced back to the initial imported case detected on January 15th [3]; the second calls this claim into question and suggests that the current epidemic originated roughly four weeks later, in early February [5].

6.1. Limitations

Our estimates are based on sparse data and multiple assumptions that have resulted in wide credible intervals and potential biases. For one, we do not explicitly consider the accuracy of the viral tests. For example, the Wuhan study tested oropharyngeal (OP) swabs rather than (NP) nasopharyngeal swabs, which have lower sensitivity [31]. The SARS-CoV-2 RT-PCR tests used have a reported false negative rate of 29% [32] and false positive rate of 0.8% [33]. For influenza, both error rates are under 10% [34]. Under the maximum reported error rates for both viruses, we would expect that ref. [9] may have missed approximately 1.4 SARS-CoV-2 cases and over-diagnosed influenza by 1.5 cases. This would imply an even larger ratio of COVID-19 positive to influenza positive cases and a 41% higher overall prevalence of COVID-19 among adults over 30 in Wuhan during this period than we estimated. Larger samples using NP rather than OP swabs for the SARS-Cov-2 test would allow more precise estimation of the early prevalence of SARS-CoV-2 in cities worldwide.

Both studies leveraged data from existing surveillance systems that are designed to provide reliable and representative data on respiratory virus prevalence. Thus, we made two key assumptions. First, influenza and COVID-19 were widespread and exhibited similar epidemiological patterns throughout the 13 central Wuhan districts and throughout the 19 PUMA's of Seattle, during the study periods. Second, the studies provide representative data for these cities. In Wuhan, if SARS-CoV-2 was only spreading in the 6 districts where it was detected, then our estimate for the prevalence of SARS-CoV-2 would decrease by 51%. Nonetheless, we believe that our methodology and qualitative insights are robust, given that the two Wuhan hospitals serve as sentinels for the Chinese Influenza Surveillance System [9] and the high inter-district mobility within Wuhan [35]. Likewise, the Seattle Flu Study was designed to broadly sample the metropolitan area [36].

Finally, the validity of our estimates hinges on our assumption that influenza and COVID-19 spread similarly during the periods of the two retrospective studies. Both studies tested specimens taken during the heart of the influenza season, when transmission was rampant. The simultaneous global expansion of the COVID-19 pandemic suggests that conditions were equally favorable for the spread of SARS-CoV-2. Moreover, the two studies analyzed specimens collected through surveillance systems in China and the US that were specifically designed to provide reliable estimates of the prevalence of influenza and other similarly-spreading respiratory viruses. That said, influenza is highly seasonal and SARS-CoV-2 may exhibit very different seasonal or non-seasonal transmission dynamics. While we conjecture that our approach was robust for the short period when both viruses were circulating in the focal communities, it may not provide reliable estimates for samples taken over longer periods of time or during the influenza off-season.

7. Conclusions

With these caveats in mind, we conclude that our method provides a way to *roughly triangulate* the unseen emergence of the COVID-19 pandemic in cities around the world during the early months of 2020. Retrospective testing of swabs from ILI and ARI patients stored in laboratories can indicate the local ratio of symptomatic SARS-CoV-2 infections to symptomatic influenza infections. If we know the prevalence of influenza when and where the swabs were taken, then we can extrapolate the concurrent prevalence of COVID-19. This approach can elucidate the past as well as provide sentinel surveillance for novel respiratory viruses that co-circulate with influenza, prior to widely available testing.

The COVID-19 epidemics in Wuhan and Seattle were far more extensive than initially reported and had likely been spreading for several weeks before they became apparent. The large discrepancy between confirmed cases and true prevalence highlights the difficulty of determining infection fatality rates from readily available COVID-19 data.

Declaration of Competing Interest

We declare no competing interests.

Author Contributions

Zhanwei Du, Ciara Nugent and Lauren Ancel Meyers: conceived the study, designed statistical methods, conducted analyses, interpreted results, wrote and revised the manuscript. Benjamin J. Cowling: conceived the study, interpreted results, and revised the manuscript. Emily Javan: collected the demographic and epidemiological data of Census Block Groups and Public Use Microdata Areas in the Seattle area and revised the manuscript.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2020.100479.

References

[1] First travel-related case of 2019 novel coronavirus detected in United States. Center for Disease Control and Prevention; January 2020. p. 21..
[2] Du Z, Wang L, Cauchemez S, Xu X, Wang X, Benjamin JC, et al. Risk for transportation of coronavirus disease from Wuhan to other cities in China. Emerging Infect Dis J 2020;26:1049.
[3] Bedford T., Greninger A.L., Roychoudhury P., Starita L.M., Famulare M., Huang M.-L., et al. Cryptic transmission of SARS-CoV-2 in Washington State. medRxiv2020.
[4] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020;395:689–97.
[5] Worobey M., Pekar J., Larsen B.B., Nelson M.I., Hill V., Joy J.B., et al. The emergence of SARS-CoV-2 in Europe and the US. bioRxiv2020.
[6] Report on Novel Coronavirus−Infected Pneumonia by National Health Commission of the People's Republic of China on January 23, 20202020. http://www.nhc.gov.cn/yjb/s3578/202001/5d19a4f6d3154b9fae328918ed2e3c8a.shtml. (accessed May 30, 2020).
[7] Report on Novel Coronavirus−Infected Pneumonia by Hubei Provincial Health Commission on January 23, 20202020. http://wjw.hubei.gov.cn/fbjd/dtyw/202001/t20200123_2014522.shtml(accessed April 17, 2020).
[8] COVID-19 Case Count Info | Snohomish Health District, WA n.d. https://www.snohd.org/499/COVID-19-Case-Count-Info. (accessed May 23, 2020).
[9] Kong W-H, Li Y, Peng M-W, Kong D-G, Yang X-B, Wang L, et al. SARS-CoV-2 detection in patients with influenza-like illness. NMicrobiol 2020;5:675–8.
[10] Chu HY, Englund JA, Starita LM, Famulare M, Brandstetter E, Nickerson DA, et al. Early detection of Covid-19 through a citywide pandemic surveillance platform. N Engl J Med 2020. doi: 10.1056/NEJMc2008646.
[11] U.S. Influenza Surveillance System: Purpose and Methods | CDC2020. https://www.cdc.gov/flu/weekly/overview.htm. (accessed May 25, 2020).
[12] Cobey S. Modeling infectious disease dynamics. Science 2020;368:713–4.
[13] Q&A: Influenza and COVID-19 - similarities and differences n.d. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-similarities-and-differences-covid-19-and-influenza?gclid=CjwKCAjwiMj2BRBFEiwAYfTbCgmDr3gRpavLHqP5Mfu0scNJ7Y844cF7GDesHoaUVR9l_p7LRzUpKxoC3ZwQAvD_BwE. (accessed May 30, 2020).
[14] Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. N Engl J Med 2020;382:1177–9.
[15] Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, et al. Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study. Lancet Public Health 2020;5:e279–88.
[16] Wuhan Municipal Health Commission. Briefing on the development of health in Wuhan in 2018 n.d. http://wjw.wh.gov.cn/upload/file/20191205/1575536693972018707.pdf. (accessed March 14, 2020).
[17] Children's Hospital of Wuhann.d.http://www.zgwhfe.com/new/info.aspx?cid=70&id=342(accessed May 8, 2020).
[18] Wuhan Children Hospital. Build a Child's Bridge of Life. Student Health News 2020 n.d. http://szb.jkb.com.cn/xsjkb/html/2020-06/10/content_279238.htm (accessed June 23.
[19] Wuhan No.1 Hospitaln.d.http://www.whyyy.com/yygk/single/show/1.aspx. (accessed May 8, 2020).
[20] CNBo S. The sixth national population census of the people's republic of China. China: China National Bureau of Statistics; 2010.
[21] Wuhan Statistics Bureau. Wuhan city2019. http://tjj.wuhan.gov.cn/details.aspx?id=83(accessed March 13, 2020).
[22] Centers for Disease Control and Prevention. National, Regional, and State Level Outpatient Illness and Viral Surveillance2020. https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html. (accessed May 22, 2020).
[23] Berger JO, Yang R. A catalog of noninformative priors. Duke Univ; 1997 ISDS Discussion Paper.
[24] Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan. China Nat Med 2020;26:506–10.
[25] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia. N Engl J Med 2020;382:1199–207.
[26] He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 2020;26:672–5.
[27] Liu W, Zhang Q, Chen J, Xiang R, Song H, Shu S, et al. Detection of Covid-19 in Children in Early January 2020 in Wuhan, China. N Engl J Med 2020;382:1370–1.
[28] Q&A on coronaviruses (COVID-19)n.d.https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses. (accessed May 30, 2020).
[29] Imai N., Dorigatti I., Cori A., Donnelly C., Riley S., Ferguson N.M. Report 2: estimating the potential total number of novel Coronavirus cases in Wuhan City. China (Https://wwwImperialAcUk/media/imperialcollege/medicine/sph/ide/gida-fellowships/2019-nCoV-Outbreak-Report-22-01-2020Pdf 2020.
[30] Baker M. When did the coronavirus arrive in the U.S.? here's a review of the evidence. The New York Times; 2020.
[31] Patel MR, Carroll D, Ussery E, Whitham H, Elkins CA, Noble-Wang J, et al. Performance of oropharyngeal swab testing compared to nasopharyngeal swab testing for diagnosis of COVID-19 -United States, January-February 2020. Clin Infect Dis 2020. doi: 10.1093/cid/ciaa759.
[32] Fang Y., Zhang H., Xie J., Lin M., Ying L., Pang P., et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology2020:200432.
[33] Cohen AN, Kessel B. False positives in reverse transcription PCR testing for SARS-CoV-2. Epidemiology 2020. doi: 10.1101/2020.04.26.20080911.
[34] Clinical Lab Manager Magazine. Influenza diagnostic methods: RT-PCR vs. RIDTsn.d.
[35] Li L, Yang L, Zhu H, Dai R. Explorative analysis of Wuhan intra-urban human mobility using social media check-in data. PLoS One 2015;10:e0135286.
[36] Interactive World Map | University of Washington - Department of Global Health n.d. https://globalhealth.washington.edu/interactive-map/projects/3221/The-Seattle-Flu-Study. (accessed May 27, 2020).