Supplementary Appendix

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

Zhanwei Du[1], Emily Javan[1], Ciara Nugent[1], Benjamin J. Cowling[2], and Lauren Ancel Meyers[1,3*]

1. The University of Texas at Austin, Austin, Texas 78712, The United States of America
2. The University of Hong Kong, 7 Sassoon Rd, Hong Kong SAR, China
3. Santa Fe Institute, Santa Fe, New Mexico, The United States of America

[*] Correspondence: Lauren Ancel Meyers (laurenmeyers@austin.utexas.edu)

Table S1. Model Parameters and Data Sources. Parameters with an age indicator (a) have separate values for the 0–17 and 18+ age ranges.

Symbol	Description	Values	Sources
$H^a_{d,\tau}$	Number of symptomatic COVID-19 cases in age group a in location d at time τ	Estimated at the Public Use Microdata Areas (PUMA) level per day for 0-17y and over 18y age groups.	Estimated
r^a	Ratio of ARI patients that are COVID-19 positive versus influenza positive in age group a	Between February 24 and March 9, 2020: Age 0-17y: 0.11 [95% CrI: 0.03–0.33] Over 18y: 0.14 [95% CrI: 0.09–0.21]	Ref. [1] tested 2353 mid-nasal swab samples from patients with acute respiratory illness (ARI), January 1-March 9, 2020. Of those, 442 and 25 tested positive for influenza and COVID-19, respectively (none were double positive).
N_d	Population size of location d	Public Use Microdata Area (PUMA) population sizes in Seattle metropolitan area estimated for 2014-2018	2018 American Community Survey 5-Year Data, released in 2019 [2]
Ω_τ	Number of outpatient visits (all causes and ages) at time τ	MMWR week 1 to week 11, 2020 for Region 10: [71870, 73781, 70333, 72732, 73329, 72597, 73331, 73118,	CDC weekly reports for HHS Region 10 (Alaska, Idaho, Oregon, and Washington), January 1-March 9, 2020[3]

Appendix: Symptomatic COVID-19 cases in Seattle

		72963, 72812, 68179]	
Θ_τ^a	Number of ILI outpatient visits in age group a at time τ	MMWR week 1 to week 11, 2020 for Region 10: Age 0-17y: [1645, 1125, 1037, 1198, 1169, 1088, 1111, 1172, 1164, 1389, 1419] Over 18y: [2305, 1855, 1436, 1546, 1581, 1375, 1420, 1429, 1650, 2436, 2592]	CDC weekly reports of ILI in HHS Region 10 (Alaska, Idaho, Oregon, and Washington), January 1-March 9, 2020[3]
Φ_τ	Percent of influenza positive tests at time τ	MMWR week 1 to week 11, 2020 for Region 10: [23.275, 20.9372, 18.8126, 21.2625, 19.097, 16.3656, 16.3466, 17.2397, 18.3453, 17.5661, 10.8454]	CDC weekly reports of influenza positive percents in HHS Region 10 (Alaska, Idaho, Oregon, and Washington), January 1-March 9, 2020[3]
ι_τ	Number of influenza positive in ref. [1] at time τ	442	Ref. [1] reports that a total of 442 swabs tested positive for influenza from January 1-March 9, 2020, but does not report influenza positivity by day or week.
x_c^a	Number of symptomatic COVID-19 cases in age group a at time τ	Between February 24 and March 9, 2020: Age 0-17y: 2 Over 18y: 23	Ref. [1] tested 2 and 23 tested COVID-19 positive for two age groups, respectively.
x_f^a	Number of influenza positive in age group a between February 24 and March 9, 2020	Age 0-17y: 23 Over 18y: 170	Estimated
$N_{a,\tau}$	Number of total tests of age group a in ref. [1] at time τ	Extracted from ref.[1] for 0-17y and over 18y age groups.	Data extracted from ref.[1] using a web plot digitizer[4], , January 1-March 9, 2020.
T_d	Epidemic doubling time	6.1 [90% uncertainty interval of 5.1 to 8.2] days	Ref. [5]

Appendix: Symptomatic COVID-19 cases in Seattle

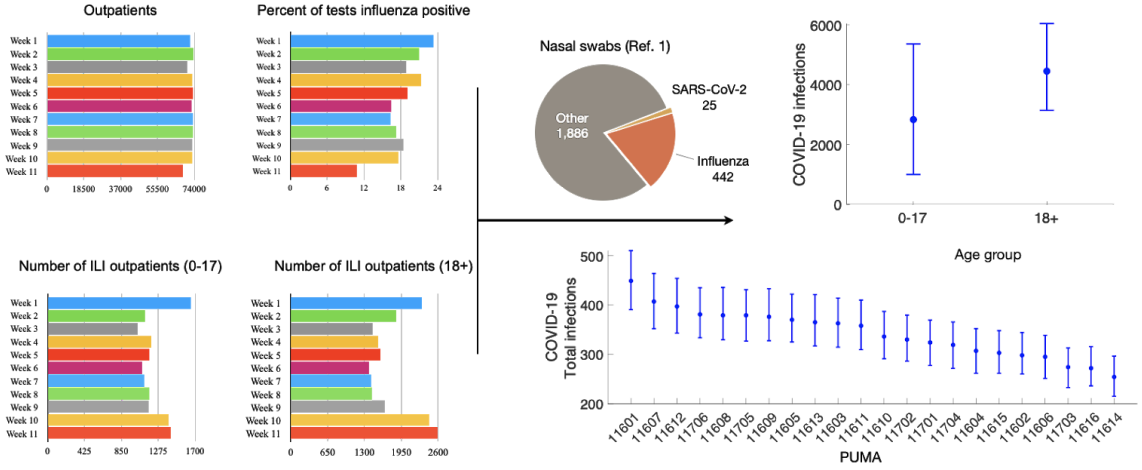

Figure S1. Estimating the number of symptomatic pediatric and adult COVID-19 infections based on the ratio between samples retrospectively testing positive for COVID-19 and influenza in Seattle from January 1st to March 9th, 2020 [6]**.** First, we analyze influenza surveillance data from the CDC FluView platform at the level of HHS region 10 [3], including the number of outpatients, percent positive influenza tests, and number of ILI outpatients. We combined these to estimate the number of outpatients (all cause) testing positive for influenza from January 1, 2020 to March 9, 2020 (left graphs). Second, we estimate the ratio of COVID-19 positive to influenza positive patients among pediatric and adult patients with ARI, based on a recent retrospective study in Seattle [6]. The ARI case definition in ref. [1] 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", in contrast with 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"[7]. Thus, the case definitions overlap considerably, but are not identical. We estimate ratios of 0.11 [95% CrI: 0.03–0.33] and 0.14 [95% CrI: 0.09–0.21] for children under 18 and adults, respectively, between February 24 and March 9, 2020. We then estimate the number of symptomatic COVID-19 infections among pediatric and adult patients in Seattle during this time period based on the number of influenza positive patients and the ratio of COVID-19 to influenza positive patients, using Monte Carlo sampling to incorporate uncertainty in our estimates of both quantities (upper right). Finally, we estimate the age-specific COVID-19 infections for the 22 PUMA's of King and Snohomish counties in the Seattle Metropolitan Area based on their age-stratified population sizes, ordered from highest to lowest expected COVID-19 prevalence.

Appendix: Symptomatic COVID-19 cases in Seattle

Method

We estimate the ratio of COVID-19 to influenza patients (r) from ref. [6], the age-specific prevalence of influenza during the corresponding time period based on CDC surveillance data [3], and then combine the two to estimate the PUMA level prevalence of COVID-19 (Figure S1).

Estimating the number of influenza positive samples during the period of undetected COVID-19 transmission

Ref. [1] provides the total number of influenza positive swabs from January 1, 2020-March 9, 2020, but does not break down the results by date or age group. We use regional influenza surveillance data to estimate that breakdown, under the assumption that weekly and age-specific positivity in the ref. [1] sample mirrored that observed for HHS Region 10 as a whole. HHS Region 10 encompasses Alaska, Idaho, Oregon, and Washington.

In particular, the study reported 442 influenza positive patients for January 1, 2020-March 9, 2020 [1]. We apply the following method to estimate the subset of those patients who were influenza positive between February 24, 2020 and March 9, 2020, corresponding to the dates of the SARS-CoV-2 positive cases identified in ref. [1]. In short, we assume that the daily influenza positivity in the sample mirrored the overall influenza positivity reported by CDC FluView for HHS region 10 [3].

For a given day τ, we assume that the number of influenza positive cases in the sample for age group a is simply the number of tests conducted in that age group on that day $N_{a,\tau}$ times the regional influenza positivity ι_τ on that day, scaled to ensure that the total number of cases across the entire time period totalled 442.

$$x_{a,\tau} = N_{a,\tau}\iota_\tau \left(\frac{442}{\sum_{j=\text{Jan 1}}^{\text{Mar 9}} (N_{\text{ped},j} + N_{\text{adult},j})\iota_j} \right).$$

The CDC's FluView provides the number of tests performed and number of tests positive for influenza on a weekly basis. We assume that the daily positivity is equal to that of the corresponding week. The denominator is the *expected* number of influenza positive tests in the study based on the total number of tests performed daily and the weekly proportions of positive tests reported by the regional surveillance system. Remarkably it is equal to 433, which is almost identical to the actual 442 cases reported by the study [1].

For the purposes of estimating the ratio of SARS-CoV-2 to influenza positive cases, we aggregate these estimates into the total influenza positive pediatric ($x_f^{\text{ped}} = 23$) and adult ($x_f^{\text{adult}} = 170$) cases between February 24th and March 9th.

Estimating COVID-19 infections in Seattle

Let $H_{d,\tau}^a$ denote the number of COVID-19 infections in PUMA *d* (for each of the 22 PUMA's in King and Snohomish counties) and age group *a* (i.e., children under 18y or adults), during the focal fifteen-day period τ. We first estimate the ratio (r) of COVID-19 cases to influenza cases among pediatric and adult cases from the

retrospective sample [6] using a Bayesian approach (described below), and then use this to predict the patients in each PUMA and age range by assuming:

$$H^a_{d,\tau}|N_d, \lambda, \Theta^a_\tau, \Omega_\tau, \Phi_\tau, r \sim B(N_d, \frac{\lambda\Theta^a_\tau}{\Omega_\tau}\Phi_\tau r)$$

where N_d is the number of people in PUMA d; Θ^a_τ is the number of ILI outpatients in age group a over a period of time τ; Ω_τ is the number of all cause outpatients of all ages over a period of time τ; Φ_τ is the percent of influenza tests that are positive in the HHS Region 10 during time period τ; r is the ratio of COVID-19 to influenza cases. *B(N, p)* denotes the standard Binomial distribution. We assume that *H* is distributed binomially where *N* is the total population and *p* is the estimate of the prevalence of symptomatic COVID-19 in the population. In other words, we multiply the number of influenza positive cases by the ratio of ARI patients that have COVID-19 versus influenza, both estimated for Seattle from February 24 to March 9, 2020.

We take a Bayesian approach to derive r as the following posterior distribution. Let N^a denote the total number of cases in the sample in the sample for age group a (pediatric under 18y or adult). Let x^a_c denote the observed number SARS-CoV-12 positive cases in age group a between February 24th and March 9th and recall that x^a_f is the estimated number of influenza positive cases during this time period. Then

$$x^a_c|p^a_c, N^a \sim B(N^a, p^a_c) \text{ and } x^a_f|p^a_f, N^a \sim B(N^a, p^a_f).$$

If we assume uninformative priors on p^a_c and p^a_f,

$$p^a_c \sim Beta(1,1) \text{ and } p^a_f \sim Beta(1,1)$$

then the posterior distributions are known in closed form[8]:

$$p^{0\text{–}17}_c|x^{0\text{–}17}_c, N^{0\text{–}17} \sim Beta(1+x^{0\text{–}17}_c, 1+N^{0\text{–}17}-x^{0\text{–}17}_c) \sim Beta(3,125)$$

$$p^{0\text{–}17}_f|x^{0\text{–}17}_f, N^{0\text{–}17} \sim Beta(1+x^{0\text{–}17}_f, 1+N^{0\text{–}17}-x^{0\text{–}17}_f) \sim Beta(24,104)$$

and

$$p^{18+}_c|x^{18+}_c, N^{18+} \sim Beta(1+x^{18+}_c, 1+N^{18+}-x^{18+}_c) \sim Beta(24,923)$$

$$p^{18+}_f|x^{18+}_f, N^{18+} \sim Beta(1+x^{18+}_f, 1+N^{18+}-x^{18+}_f \sim Beta(171,776)$$

We use MCMC to take draws from p^a_c and p^a_f, and then calculate $r^a = \frac{p^a_c}{p^a_f}$ to obtain the distribution for r^a. We thereby estimate that the ratios of COVID-19 to influenza during February 24, 2020 to March 9, 2020 were 0.11 [95% CrI: 0.03–0.33] and 0.14 [95% CrI: 0.09–0.21] for children 0–17y and adults over 18y, respectively.

Appendix: Symptomatic COVID-19 cases in Seattle

Using 1,000 draws from the distribution of r^a, we estimate $H^a_{d,\tau}$ for each PUMA and age group, and point estimates for all other parameters given in Table S1. We report the means and 95% credible intervals of the resulting posterior predictive distribution for each PUMA.

Estimating COVID-19 infections prior to March 9, 2020

To backcast the number of infections in Seattle prior to March 9, 2020 (H_{cum}), we assume

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

where T_d is the epidemic doubling time, t_0 is the day of the first infection in Seattle, and L corresponds to March 9, 2020. We use our age- and PUMA-stratified estimates for adult COVID-19 infections for February 24, 2020 to March 9, 2020 to estimate this quantity, under the assumption that the values reflect cumulative incident infections during that fourteen-day period (Figure 4).

We use Monte Carlo sampling to incorporate the uncertainty in both the epidemic doubling rate in Seattle during this period [9] and total infections from February 24, 2020 to March 9, 2020 ($H^a_{d,\tau}$). We take draws from the distribution of $H_\tau = \sum_d \sum_a H^a_{d,\tau}$ and T_d (summarized in Table S1) to estimate the time since the first infection by

$$\delta = T_d(log_2(\frac{H_\tau}{\sum_{i=0}^{14} 2^{i/T_d}})) .$$

That is, the estimated date of the first COVID-19 infection in Seattle (t_0) is δ days prior to February 24, 2020. We then estimate H_{cum} according to the equation above to project the cumulative COVID-19 infections preceding the Seattle lockdown.

Appendix: Symptomatic COVID-19 cases in Seattle

Table S2. Estimated COVID-19 infections in the 22 PUMAs of Seattle, from February 24, 2020 to March 9, 2020. Values are medians and 95% bounds across 1,000 Monte Carlo samples.

PUMA	Population	Infections		
		all	0-17	18+
11601	175,213	410 [95% CrI:364,459]	150 [95% CrI:121,179]	261 [95% CrI:227,295]
11602	116,083	272 [95% CrI:236,311]	99 [95% CrI:76,125]	173 [95% CrI:146,201]
11603	141,776	332 [95% CrI:292,375]	121 [95% CrI:96,153]	210 [95% CrI:184,243]
11604	119,177	280 [95% CrI:244,318]	101 [95% CrI:78,128]	178 [95% CrI:152,205]
11605	144,312	340 [95% CrI:298,381]	124 [95% CrI:99,151]	216 [95% CrI:184,246]
11606	114,608	269 [95% CrI:234,306]	98 [95% CrI:75,121]	171 [95% CrI:146,200]
11607	158,258	372 [95% CrI:329,417]	135 [95% CrI:108,164]	237 [95% CrI:208,269]
11608	147,938	348 [95% CrI:308,389]	126 [95% CrI:100,157]	221 [95% CrI:192,255]
11609	146,632	346 [95% CrI:304,390]	125 [95% CrI:99,156]	220 [95% CrI:189,252]
11610	131,325	308 [95% CrI:273,349]	111 [95% CrI:88,141]	196 [95% CrI:167,224]
11611	139,603	327 [95% CrI:289,368]	118 [95% CrI:94,145]	209 [95% CrI:178,241]
11612	154,565	362 [95% CrI:318,407]	131 [95% CrI:104,162]	231 [95% CrI:199,265]
11613	142,658	335 [95% CrI:297,379]	121 [95% CrI:98,149]	212 [95% CrI:183,245]
11614	98,399	231 [95% CrI:199,265]	84 [95% CrI:63,108]	147 [95% CrI:123,173]
11615	118,069	278 [95% CrI:243,316]	100 [95% CrI:77,126]	177 [95% CrI:149,206]
11616	105,922	249 [95% CrI:217,287]	89 [95% CrI:69,114]	158 [95% CrI:134,188]
11701	126,114	296 [95% CrI:261,335]	108 [95% CrI:85,135]	188 [95% CrI:159,216]
11702	128,584	302 [95% CrI:265,343]	109 [95% CrI:86,137]	192 [95% CrI:164,221]
11703	106,551	249 [95% CrI:214,288]	91 [95% CrI:69,118]	159 [95% CrI:133,185]
11704	124,149	291 [95% CrI:256,332]	106 [95% CrI:80,134]	186 [95% CrI:161,212]
11705	147,548	347 [95% CrI:304,389]	126 [95% CrI:99,155]	221 [95% CrI:189,251]
11706	148,447	350 [95% CrI:308,390]	127 [95% CrI:101,156]	223 [95% CrI:191,251]
Seattle	2,935,931	6748 [95% CrI: 4133, 11020]	2268 [95% CrI: 498, 6069]	4367 [95% CrI: 2776, 6526]

Appendix: Symptomatic COVID-19 cases in Seattle

Table S3. Estimated COVID-19 adult infections in the 13 central districts of Wuhan, from December 30, 2019 to January 12, 2020. Values are medians and 95% bounds across 1,000 Monte Carlo samples.

District	Population	Infections
		≥30
Hongshan*	1,574,315	143 (43-391)
Wuchang*	1,375,062	177 (54-485)
Jiangan*	1,014,584	144 (44-393)
Huangpi	990,782	132 (40-362)
Xinzhou	961,138	121 (37-331)
Qiaokou*	939,515	123 (37-335)
Jiangxia	894,731	85 (26-234)
Jianghan*	776,487	109 (33-298)
Caidian	700,950	90 (27-246)
Hanyang*	661,434	93 (28-254)
Qingshan	549,903	81 (25-223)
Dongxihu	511,906	69 (21-189)
Hannan	130,192	19 (6-51)
Wuhan	11,081,000	1386 (420-3793)

* These 7 districts are located in central Wuhan; the other six are suburban.

References

[1] 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. https://doi.org/10.1056/NEJMc2008646.

[2] US Census Bureau. American Community Survey 5-Year Data (2009-2018) n.d.

[3] Centers for Disease Control and Prevention. National, Regional, and State Level Outpatient Illness and Viral Surveillance 2020. https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html (accessed May 22, 2020).

[4] Rohatgi A. WebPlotDigitizer- Extract data from plots, images, and maps 2010.

[5] Bedford T, Greninger AL, Roychoudhury P, Starita LM, Famulare M, Huang M-L, et al. Cryptic transmission of SARS-CoV-2 in Washington State. medRxiv 2020.

[6] 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. Nat Microbiol 2020. https://doi.org/10.1038/s41564-020-0713-1.

[7] U.S. Influenza Surveillance System: Purpose and Methods | CDC 2020. https://www.cdc.gov/flu/weekly/overview.htm (accessed May 25, 2020).

[8] Berger JO, Yang R. A catalog of noninformative priors. ISDS Discussion Paper, Duke Univ; 1997.

[9] Zhanwei Du, Lin Wang, Simon Cauchemez, Xiaoke Xu, Xianwen Wang, Benjamin J. Cowling, et al. Risk for Transportation of 2019 Novel Coronavirus Disease from Wuhan to Other Cities in China. Emerging Infectious Disease Journal 2020;26. https://doi.org/10.3201/eid2605.200146.