COVID-19 Healthcare Demand Projections: Houston-The Woodlands-Sugar Land MSA, Texas

Kelly Pierce, Ethan Ho, Xutong Wang, Remy Pasco, Zhanwei Du, Spencer Fox, Greg Zynda, Jawon Song, Lauren Ancel Meyers

COVID-19 Healthcare Demand Projections: Houston-The Woodlands-Sugar Land MSA, Texas

The University of Texas COVID-19 Modeling Consortium
Contributors: Kelly Pierce, Ethan Ho, Xutong Wang, Remy Pasco, Zhanwei Du,
Spencer Fox, Greg Zynda, Jawon Song, Lauren Ancel Meyers
Contact: utpandemics@austin.utexas.edu

Overview

To support healthcare planning, we analyzed the Houston-The Woodlands-Sugar Land MSA module of our *US COVID-19 Pandemic Model* to project the number of cases, healthcare requirements and deaths under different scenarios. Note that the results presented herein are based on multiple assumptions about the transmission rate and age-specific severity of COVID-19. There is still much we do not understand about the transmission dynamics of this virus, including the extent of asymptomatic infection and transmission. These results do not represent the full range of uncertainty. Rather, they are meant to serve as plausible scenarios for gauging the likely impacts of control measures in the Houston-The Woodlands-Sugar Land MSA.

We have updated our model inputs based on the daily number of COVID-19 hospitalizations in Houston-The Woodlands-Sugar Land between April 2, 2020 and April 20, 2020, provided by the Southeast Texas Regional Advisory Council (SETRAC). The projections assume that schools were closed on March 19, 2020 (start of state mandated school closures) and extensive social distancing began on March 24, 2020 with Houston and Harris County's Stay Home Work Safe order [1]. The data suggest that recent social distancing has reduced transmission by anywhere between 80% and 100% relative to the period prior to March 19th. We make projections for five different scenarios. The first three–80%, 95% and 100% reductions in transmission–fall within this range of current estimates; the other two–0% and 50% reductions in transmission–provide more pessimistic projections that could occur with extreme relaxation of social distancing measures. For each of the scenarios, the graphs project COVID-19 cases, hospitalizations, patients requiring ICU care, patients requiring ventilation and deaths.

We are posting these results prior to peer review to provide intuition for both policy makers and the public regarding both the immediate threat of COVID-19 and the extent to which continued social distancing, transmission-reducing precautions such as keeping physical distance and wearing cloth face coverings, can mitigate that threat.

COVID-19 projections for Houston-The Woodlands-Sugar Land MSA

We updated the Houston-The Woodlands-Sugar Land MSA module of our US COVID-19 Pandemic Model to simulate COVID-19 epidemics under various assumptions about the efficacy of social distancing measures.

Hospitalization data from the Houston-The Woodlands-Sugar Land MSA suggest that the ongoing COVID-19 epidemic was likely seeded in early February. The simulations thus assume the epidemic began with a single infected adult on February 10, 2020 and project transmission for 175 days through mid-August based on the following assumptions:

- Starting deterministic condition: February 10, 2020 with 1 infected adult

- Starting stochastic condition: at least 10 symptomatic cases total. For the Houston-The Woodlands-Sugar Land model this threshold is met on February 23, 2020.

- Time course of interventions

    - February 10 - March 18, 2020: No interventions

    - March 19 - August 16, 2020: Schools closed

    - March 24 - August 16, 2020: social distancing intervention reduces transmission (beyond school closures) by an additional 50%, 80%, 95% and 100%

- β = 0.035 (based on fitting our model to daily COVID-19 hospitalizations in Austin Round-Rock MSA for March 13-April 24, 2020). This corresponds to an epidemic doubling time of 2.9 days prior to March 19, 2020.

- Average incubation period (assuming 12.1% of transmission happens pre-symptomatically): 6.9 days [2]

- Proportion of cases asymptomatic (assumed 46% as infectious as symptomatic cases): 17.9% [3]

Table 1 and Figures 1-5 summarize results of COVID-19 simulations for the Houston-The Woodlands-Sugar Land MSA. Each stochastic simulation began on February 23, 2020 with approximately 10 infectious COVID-19 cases and ended on August 16, 2020. The model structure and parameters, including age-specific hospitalization and fatality rates, are described in the Appendices below.

Table 1. Estimated cumulative COVID-19 cases, hospitalizations, ICU cases, cases requiring mechanical ventilatory therapy, and deaths for the Houston-The Woodlands-Sugar Land MSA from February 23, 2020 to August 16 2020. The values are medians (with full range in parentheses) across 300 stochastic simulations based on the parameters given in Appendix 1.

| Contact reduction	Outcomes (Cumulative February 23 - August 16, 2020)				
	Cases	**Hospitalizations**	**ICU beds**	**Ventilators**	**Deaths**
School closure only	5,734,903 (5,732,110-5,737,977)	312,253 (311,283-313,606)	37,425 (37,313-37,605)	17,822 (17,768-17,907)	38,786 (38,136-39,288)
50%	5,492,249 (5,363,306-5,560,703)	285,971 (271,713-294,474)	34,355 (3,2646-35,373)	16,360 (15,546-16,844)	33,393 (30,684-35,202)
80%	1,175,502 (551,894-2,426,190)	45,247 (20,543-103,154)	5,467 (2,478-12,450)	2,604 (1,180-5,929)	4,154 (1,856-10,125)
95%	76,918 (37,818-196,765)	3,554 (1,759-9,054)	430 (214-1,096)	205 (102-522)	355 (168-874)
100%	50,099 (18,193-135,650)	2,299 (859-6,291)	278 (104-759)	133 (50-362)	225 (92-615)
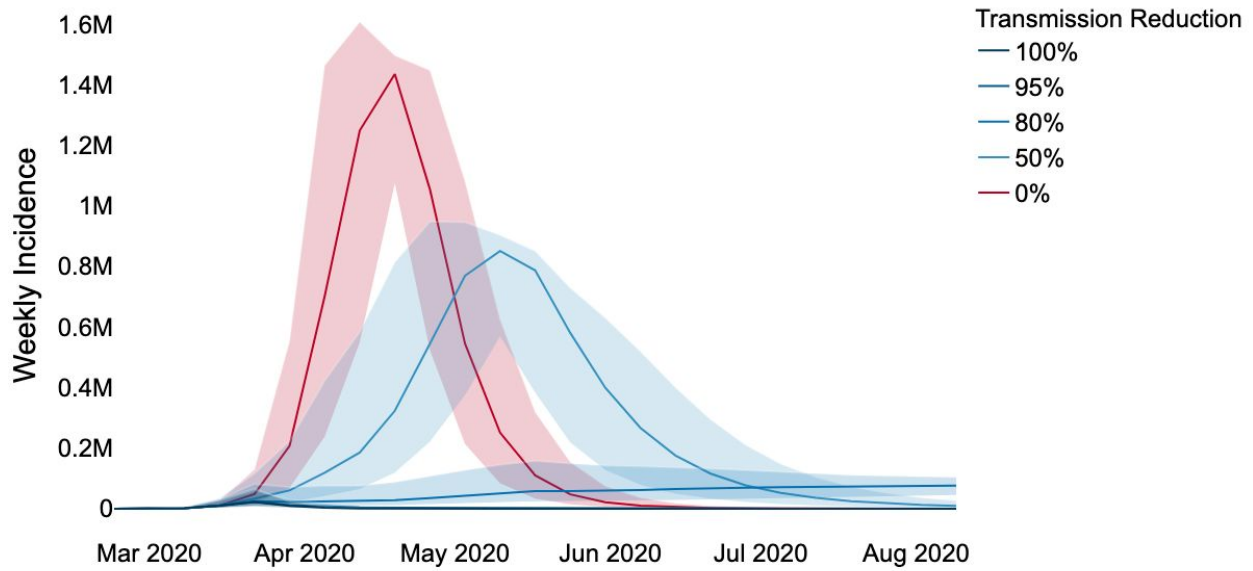
Figure 1. Projected COVID-19 cases in the Houston-The Woodlands-Sugar Land MSA from February 23, 2020 to August 16, 2020 coupled with different degrees of social distancing starting March 24, 2020. The red line projects COVID-19 transmission assuming that there was no reduction in transmission from social distancing beyond a modest effect of school closures. The blue lines show increasing levels of social distancing efficacy, from light to dark: 50%, 80%, 95%, and 100% reductions in transmission. Lines and shading indicate the minimum, median and maximum values across 300 stochastic simulations.
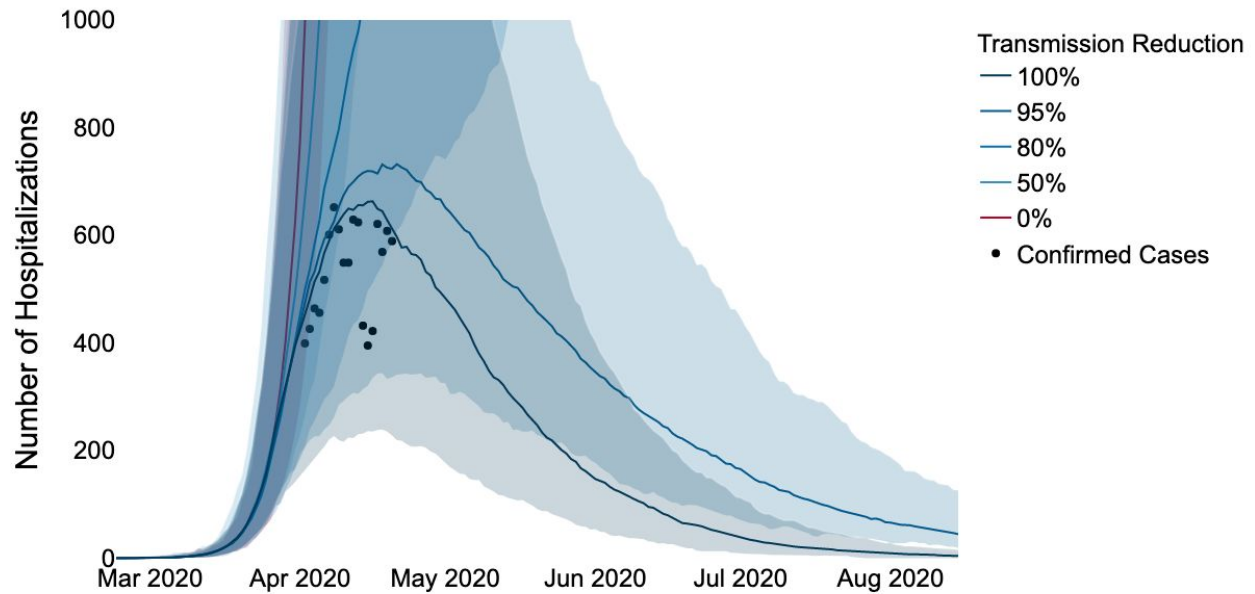
Figure 2. Projected daily COVID-19 hospitalizations in Houston-The Woodlands-Sugar Land MSA from February 23, 2020 to August 16, 2020 coupled with different degrees of social distancing starting March 24, 2020. The two graphs are identical except that the top graph only shows hospitalizations up to 1000 per day and includes the reported total daily COVID-19 hospitalizations across all reporting Houston-The Woodlands-Sugar Land MSA hospitals (black points). In both graphs, the red lines project COVID-19 hospitalizations assuming that there was no reduction in transmission from social distancing beyond a modest effect of school closures. The blue lines show increasing levels of social distancing efficacy, from light to dark: 50%, 80%, 95%, and 100% reductions in transmission. Lines and shading indicate the minimum, median and maximum values across 300 stochastic simulations.
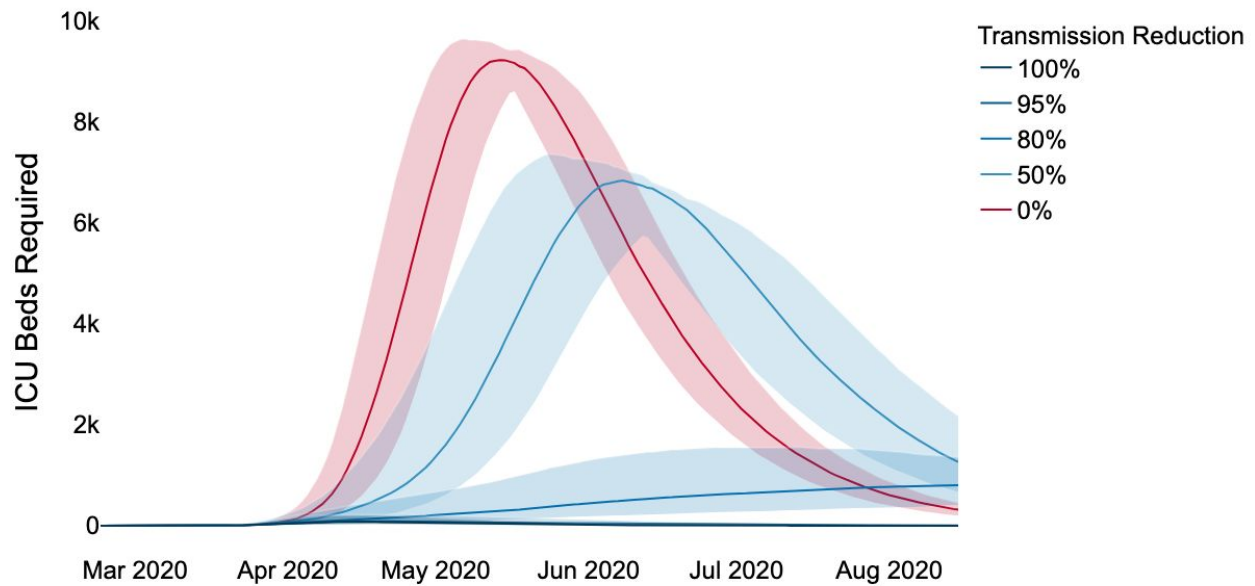
Figure 3. Projected COVID-19 cases requiring ICU treatment in Houston-The Woodlands-Sugar Land MSA from February 23, 2020 to August 16, 2020 coupled with different degrees of social distancing starting March 24, 2020. The red line projects COVID-19 ICU patients assuming that there was no reduction in transmission from social distancing beyond a modest effect of school closures. The blue lines show increasing levels of social distancing efficacy, from light to dark: 50%, 80%, 95%, and 100% reductions in transmission. Lines and shading indicate the minimum, median and maximum values across 300 stochastic simulations.
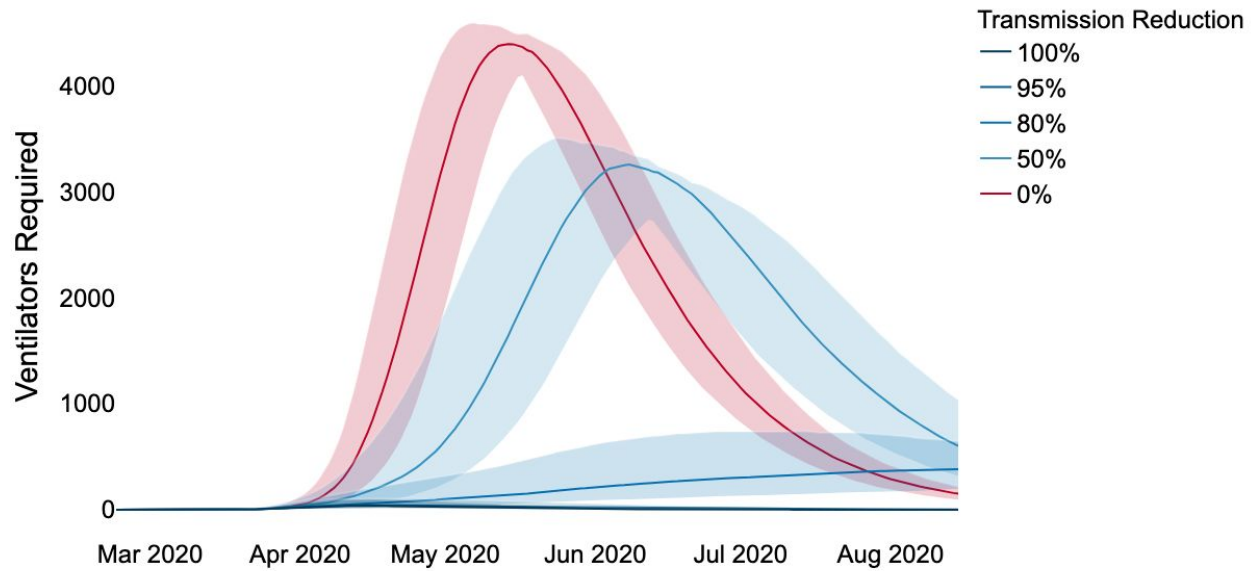
Figure 4. Projected COVID-19 cases requiring mechanical ventilation in Houston-The Woodlands-Sugar Land MSA from February 23, 2020 to August 16, 2020 coupled with different degrees of social distancing starting March 24, 2020. The red line projects ventilated COVID-19 patients assuming that there was no reduction in transmission from social distancing beyond a modest effect of school closures. The blue lines show increasing levels of social distancing efficacy, from light to dark: 50%, 80%, 95%, and 100% reductions in transmission. Lines and shading indicate the minimum, median and maximum values across 300 stochastic simulations.
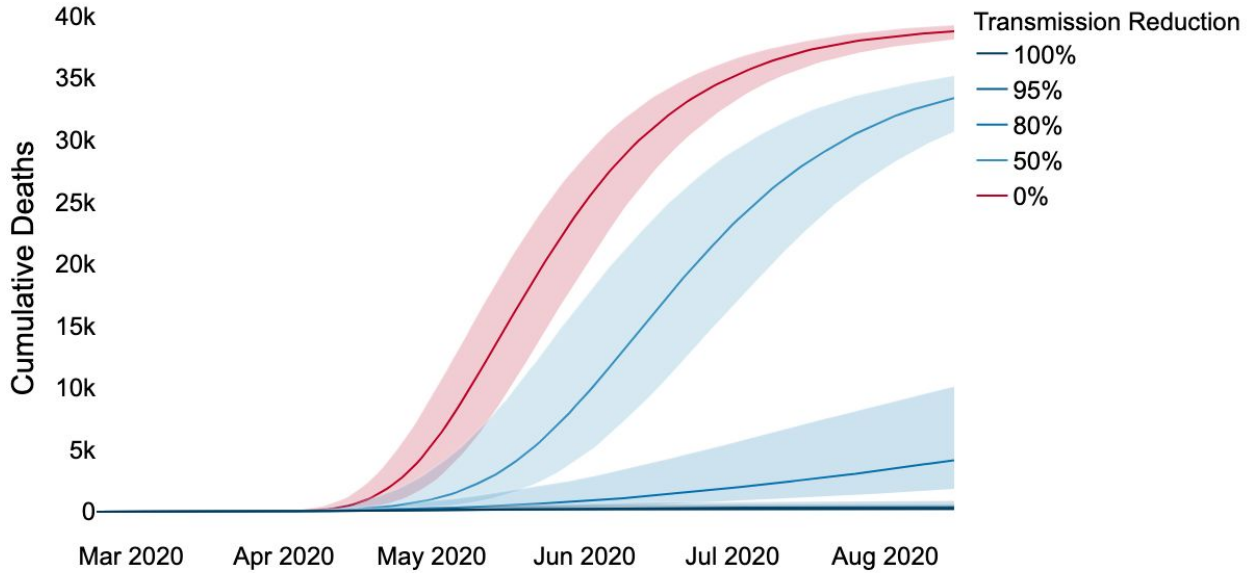
Figure 5. Projected cumulative COVID-19 deaths in Houston-The Woodlands-Sugar Land MSA from February 23, 2020 to August 16, 2020 coupled with different degrees of social distancing starting March 24, 2020. The red line projects cumulative COVID-19 deaths assuming that there was no reduction in transmission from social distancing beyond a modest effect of school closures. The blue lines show increasing levels of social distancing efficacy, from light to dark: 50%, 80%, 95%, and 100% reductions in transmission. Lines and shading indicate the minimum, median and maximum values across 300 stochastic simulations.

Appendix

COVID-19 Epidemic Model Structure and Parameters

The model structure is diagrammed in Figure A1 and described in the equations below. For each age and risk group, we build a separate set of compartments to model the transitions between the states: susceptible (S), exposed (E), symptomatic infectious (I^Y), asymptomatic infectious (I^A), symptomatic infectious that are hospitalized (I^H), recovered (R), and deceased (D). The symbols S, E, I^Y, I^A, I^H, R, and D denote the number of people in that state in the given age/risk group and the total size of the age/risk group is $N = S + E + I^Y + I^A + I^H + R + D$. The model for individuals in age group a and risk group r is given by:

$$\frac{dS_{a,r}}{dt} = -\sum_{i \in A}\sum_{j \in K}(I_{i,j}^Y \omega^Y + I_{i,j}^A \omega^A + E_{i,j}\omega^E)\beta\phi_{a,i}/N_i$$

$$\frac{dE_{a,r}}{dt} = \sum_{i \in A}\sum_{j \in K}(I_{i,j}^Y \omega^Y + I_{i,j}^A \omega^A + E_{i,j}\omega^E)\beta\phi_{a,i}/N_i - \sigma E_{a,r}$$

$$\frac{dI_{a,r}^A}{dt} = (1 - \tau)\sigma E_{a,r} - \gamma^A I_{a,r}^A$$

$$\frac{dI_{a,r}^Y}{dt} = \tau\sigma E_{a,r} - (1 - \pi)\gamma^Y I_{a,r}^Y - \pi\eta I_{a,r}^Y$$

$$\frac{dI_{a,r}^H}{dt} = \pi\eta I_{a,r}^Y - (1 - \nu)\gamma^H I_{a,r}^H - \nu\mu I_{a,r}^H$$

$$\frac{dR_{a,r}}{dt} = \gamma^A I_{a,r}^A + (1 - \pi)\gamma^Y I_{a,r}^Y + (1 - \nu)\gamma^H I_{a,r}^H$$

$$\frac{dD_{a,r}}{dt} = \nu\mu I_{a,r}^H$$

where A and K are all possible age and risk groups, ^{A}, ^{Y}, ^{H} are relative infectiousness of the I^A, I^Y, E compartments, respectively, β is transmission rate, $_{a,i}$ is the mixing rate between age group a, $i \in A$, ^{A}, ^{Y}, ^{H} are the recovery rates for the I^A, I^Y, I^H compartments, respectively, σ is the exposed rate, τ is the symptomatic ratio, π is the proportion of symptomatic individuals requiring hospitalization, η is rate at which hospitalized cases enter the hospital following symptom onset, ν is mortality rate for hospitalized cases, and μ is rate at which terminal patients die.

We simulate the model using a hybrid approach combining a deterministic initial phase (up to 10 cumulative cases) followed by a stochastic phase implemented as follows. Transitions between compartments are governed using the τ-leap method [4,5] with key parameters given in Table S1. Assuming that the events at each time-step are independent and do not impact the underlying transition rates, the numbers of each type of event should follow Poisson distributions with means equal to the rate parameters. We thus simulate the model according to the following equations:

$$S_{a,r}(t+1) - S_{a,r}(t) = -P_1$$
$$E_{a,r}(t+1) - E_{a,r}(t) = P_1 - P_2$$
$$I_{a,r}^A(t+1) - I_{a,r}^A(t) = (1 - \tau)P_2 - P_3$$
$$I_{a,r}^Y(t+1) - I_{a,r}^Y(t) = \tau P_2 - P_4 - P_5$$
$$I_{a,r}^H(t+1) - I_{a,r}^H(t) = P_5 - P_6 - P_7$$
$$R_{a,r}(t+1) - R_{a,r}(t) = P_3 + P_4 + P_6$$
$$D_{a,r}(t+1) - D_{a,r}(t) = P_7 \, ,$$

with

$$P_1 \sim \text{Pois}(S_{a,r}(t)F_{a,r}(t))$$
$$P_2 \sim \text{Pois}(\sigma E_{a,r}(t))$$

$$P_3 \sim \text{Pois}(\gamma^A I_{a,r}^A(t))$$

$$P_4 \sim \text{Pois}((1-\pi)\gamma^Y I_{a,r}^Y(t))$$

$$P_5 \sim \text{Pois}(\pi\eta I_{a,r}^Y(t))$$

$$P_6 \sim \text{Pois}((1-\nu)\gamma^H I_a^H)$$

$$P_7 \sim \text{Pois}(\nu\mu I_{a,r}^H(t))$$

and where $F_{a,r}$ denotes the force of infection for individuals in age group a and risk group r and is given by:

$$F_{a,r}(t) = \sum_{i \in A} \sum_{j \in K} (I_{i,r}^Y(t)\omega^Y + I_{i,r}^A(t)\omega^A + E_{i,j}(t)\omega^E)\beta_{a,i}\phi_{a,i}/N_i$$
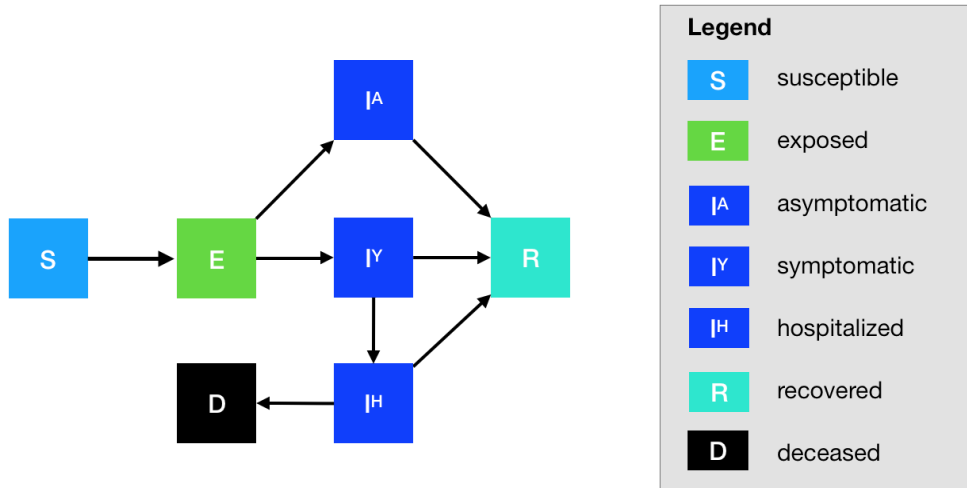

Figure A1. Compartmental model of COVID-19 transmission in a US city. Each subgroup (defined by age and risk) is modeled with a separate set of compartments. Upon infection, susceptible individuals (S) progress to exposed (E) and then to either symptomatic infectious (I^Y) or asymptomatic infectious (I^A). All asymptomatic cases eventually progress to a recovered class where they remain protected from future infection (R); symptomatic cases are either hospitalized (I^H) or recover. Mortality (D) varies by age group and risk group and is assumed to be preceded by hospitalization.

Estimating the effect of the Stay Home-Work Safe order

We estimated the transmission rate of COVID-19 in the Houston-The Woodlands-Sugar Land MSA before and after the March 24th Stay at Home order using least-squares fitting, which compares the predicted and observed numbers of daily hospitalizations (i.e., heads in beds) for the Houston-The Woodlands-Sugar Land MSA. We assume that: (i) the epidemic starts with an initial transmission rate of $\beta = 0.035$, (ii) the transmission rate decreases when school closures are enacted on March 19, 2020 (by an amount determined by our pre-set contact matrices), (iii) the transmission rate decreases further by an amount d on March 24th following the *Stay Home-Work Safe* order.

We estimate d using a nonlinear least squares fitting procedure in the SciPy/Python package across a range of possible start dates [6]. For a given start date, we run a deterministic simulation of our model assuming a central value for d. Using a trust region method, the algorithm finds value of d that minimize the sum of squared daily differences between the simulated ($\hat{H}_t$) and actual (H_t) daily hospitalizations from April 2, 2020 through April 20, 2020:

$$S(d) = \sum_t (H_t - \hat{H}_t)^2$$

for the assumed start date as ... We then select the start date that produces the lowest normalized mean square deviation and fix that start date for subsequent simulations. For the Houston-The Woodlands-Sugar Land MSA, the best-fit start date is February 10, 2020.

We calculated 95% confidence intervals for the social distancing parameter d indirectly by running 500 stochastic simulations for each of the following possible values of d': 0.0, 0.05, ..., 0.95, 1.0. For each value of d', we conducted the following analysis to determine if d' lies inside the 95% confidence interval for d.

- For all simulations, we calculate the day-to-day difference in hospitalizations (i.e., heads in beds) during the period following the *Stay Home-Work Safe* order: $\hat{z}_t = \hat{H}_t - \hat{H}_{t-1}$. We do the same for the actual data: $z_t = H_t - H_{t-1}$.

- We compute the 95% prediction interval for $\hat{z}_t$ across all 500 stochastic simulations for d' for each day t.

- We then conduct a test of the null hypothesis $H_0 : d' = d$. Under this null hypothesis, we would expect roughly 95% of the observed data (z_t) to fall within the 95% prediction band for $\hat{z}_t$ that we constructed from our simulations. By analyzing the day-to-day difference in hospitalizations rather than daily hospitalizations, we can assume that the data are independent from one day to the next. Then the expected number of observed values contained in the 95% prediction band is given by the binomial expression:
$$N_{\text{contained}} \sim B\left(N_{\text{points}}, 0.95\right)$$
where $N_{\text{contained}}$ is the number of data points contained within the 95% prediction band and N_{points} is the total number of data points (i.e., days).

- If the binomial probability of $N_{\text{contained}}$ is less than 0.05, we reject the null hypothesis $H_0 : d' = d$

To construct a 95% confidence interval for d we take the minimum and maximum d' for which we did not reject the null hypothesis $H_0 : d' = d$.

Table A1. Initial conditions, school closures and social distancing policies

Variable	Settings
Initial day of simulation	February 10, 2020
Initial infection number in locations	1 symptomatic case in 18-49y age group
School closure	March 19, 2020 - August 16, 2020
Social distancing reduction in contacts	Five scenarios: [0.0, 0.5, 0.8, 0.95, 1.0]

Age-specific and day-specific contact rates	Home, work, other and school matrices provided in Tables S4.1-S4.4
	
	● From February 10, 2020 to March 18, 2020
	   ○ Weekday = home + school + work + other
	   ○ Weekend = home + other
	   ○ Weekday holiday = home + other
	● From March 19, 2020 to March 23, 2020
	   ○ Weekday = home + work + other
	   ○ Weekend = home + other
	   ○ Weekday holiday = home + other
	● From March 24, 2020
	   ○ Weekday = (1-α)*(home + work + other)
	   ○ Weekend = (1-α)*(home + other)
	   ○ Weekday holiday = (1-α)*(home + other)

Table A2. Model parameters[a]

Parameters	Value	Source
β: transmission rate	0.035	Fitted to daily COVID-19 hospitalizations in Austin-Round Rock MSA
γ^A: recovery rate on asymptomatic compartment	Equal to γ^Y	
γ^Y: recovery rate on symptomatic non-treated compartment	$\frac{1}{\gamma^Y} \sim \mathrm{Triangular}(21.2, 22.6, 24.4)$	Verity et al. [7]
τ: symptomatic proportion (%)	82.1	Mizumoto et al. [3]

σ : exposed rate	$\frac{1}{\sigma} \sim \text{Triangular}(5.6, 7, 8.2)$	Lauer et al. [2]
P: proportion of pre-symptomatic transmission (%)	12.6	Du et al. [8]
ω^E : relative infectiousness of infectious individuals in compartment E	$\omega^E = \frac{(\frac{YHR}{\eta} + \frac{1-YHR}{\gamma^Y})\omega^Y \sigma P}{1 - P}$	
ω^A : relative infectiousness of infectious individuals in compartment I^A	0.4653	Set to mean of ω^E
IFR: infected fatality ratio, age specific (%)	Low risk: [0.0009, 0.0022, 0.0339, 0.2520, 0.6440] High risk: [0.0092, 0.0218, 0.3388, 2.5197, 6.4402]	Age adjusted from Verity et al. [7]
YFR: symptomatic fatality ratio, age specific (%)	Low risk: [0.0011165, 0.0027 , 0.0412, 0.3069, 0.7844] High risk: [0.0112, 0.0265, 0.4126, 3.0690, 7.8443]	$YFR = \frac{IFR}{1 - \tau}$
h : high-risk proportion, age specific (%)	[8.2825, 14.1121, 16.5298, 32.9912, 47.0568]	Estimated using 2015-2016 Behavioral Risk Factor Surveillance System (BRFSS) data with multilevel regression and poststratification using CDC's list of conditions that may increase the risk of serious complications from influenza [9–11]

[a]Values given as five-element vectors are age-stratified with values corresponding to 0-4, 5-17, 18-49, 50-64, 65+ year age groups, respectively.

Table A3 Hospitalization parameters

Parameters	Value	Source
γ^H: recovery rate in hospitalized compartment	1/14	14 day-average from admission to discharge (UT Austin Dell Med)
YHR: symptomatic case hospitalization rate (%)	Low risk: [0.0279, 0.0215, 1.3215, 2.8563, 3.3873] High risk: [0.2791, 0.2146, 13.2154, 28.5634, 33.8733]	Age adjusted from Verity et al. [7]
π: rate of symptomatic individuals go to hospital, age-specific	$$\pi = \frac{\gamma^Y * YHR}{\eta + (\gamma^Y - \eta)YHR}$$	
η: rate from symptom onset to hospitalized	0.1695	5.9 day average from symptom onset to hospital admission Tindale et al.[12]
μ: rate from hospitalized to death	1/14	14 day-average from admission to death (UT Austin Dell Med)
HFR: hospitalized fatality ratio, age specific (%)	[4, 12.365, 3.122, 10.745, 23.158]	$$HFR = \frac{IFR}{YHR(1 - \tau)}$$
ν: death rate on hospitalized individuals, age specific	[0.0390, 0.1208, 0.0304, 0.1049, 0.2269]	$$\nu = \frac{\gamma^H HFR}{\mu + (\gamma^H - \mu)HFR}$$
ICU: proportion hospitalized people in ICU	[0.15, 0.20, 0.15, 0.20, 0.15]	CDC planning scenarios (based on US seasonal flu data)
Vent: proportion of individuals in ICU needing ventilation	0.67	Assumption
d_{ICU}: duration of stay in ICU	10 days	Assumption, set equal to duration of ventilation
d_V: duration of ventilation	10 days	Assumption

Table A4.1 Home contact matrix. Daily number contacts by age group at home.

	0-4y	5-17y	18-49y	50-64y	65y+
0-4y	0.5	0.9	2.0	0.1	0.0
5-17y	0.2	1.7	1.9	0.2	0.0
18-49y	0.2	0.9	1.7	0.2	0.0
50-64y	0.2	0.7	1.2	1.0	0.1
65y+	0.1	0.7	1.0	0.3	0.6

Table A4.2 School contact matrix. Daily number contacts by age group at school.

	0-4y	5-17y	18-49y	50-64y	65y+
0-4y	1.0	0.5	0.4	0.1	0.0
5-17y	0.2	3.7	0.9	0.1	0.0
18-49y	0.0	0.7	0.8	0.0	0.0
50-64y	0.1	0.8	0.5	0.1	0.0
65y+	0.0	0.0	0.1	0.0	0.0

Table A4.3 Work contact matrix. Daily number contacts by age group at work.

	0-4y	5-17y	18-49y	50-64y	65y+
0-4y	0.0	0.0	0.0	0.0	0.0
5-17y	0.0	0.1	0.4	0.0	0.0
18-49y	0.0	0.2	4.5	0.8	0.0
50-64y	0.0	0.1	2.8	0.9	0.0
65y+	0.0	0.0	0.1	0.0	0.0

Table A4.4 Others contact matrix. Daily number contacts by age group at other locations.

	0-4y	5-17y	18-49y	50-64y	65y+
0-4y	0.7	0.7	1.8	0.6	0.3
5-17y	0.2	2.6	2.1	0.4	0.2
18-49y	0.1	0.7	3.3	0.6	0.2
50-64y	0.1	0.3	2.2	1.1	0.4
65y+	0.0	0.2	1.3	0.8	0.6

Estimation of age-stratified proportion of population at high-risk for COVID-10 complications

We estimate age-specific proportions of the population at high risk of complications from COVID-19 based on data for the Houston-The Woodlands-Sugar Land MSA from the CDC's 500 cities project (Figure A2) [13]. We assume that high risk conditions for COVID-19 are the same as those specified for influenza by the CDC [9]. The CDC's 500 cities project provides city-specific estimates of prevalence for several of these conditions among adults [14]. The estimates were obtained from the 2015-2016 Behavioral Risk Factor Surveillance System (BRFSS) data using a small-area estimation methodology called multi-level regression and poststratification [10,11]. It links geocoded health surveys to high spatial resolution population demographic and socioeconomic data [11].

Estimating high-risk proportions for adults. To estimate the proportion of adults at high risk for complications, we use the CDC's 500 cities data, as well as data on the prevalence of HIV/AIDS, obesity and pregnancy among adults (Table A6).

The CDC 500 cities dataset includes the prevalence of each condition on its own, rather than the prevalence of multiple conditions (e.g., dyads or triads). Thus, we use separate co-morbidity estimates to determine overlap. Reference about chronic conditions [15] gives US estimates for the proportion of the adult population with 0, 1 or 2+ chronic conditions, per age group. Using this and the 500 cities data we can estimate the proportion of the population p_{HR} in each age group in each city with at least one chronic condition listed in the CDC 500 cities data (Table A6) putting them at high-risk for flu complications.

HIV: We use the data from table 20a in CDC HIV surveillance report [16] to estimate the population in each risk group living with HIV in the US (last column, 2015 data). Assuming independence between HIV and other chronic conditions, we increase the proportion of the population at high-risk for influenza to account for individuals with HIV but no other underlying conditions.

Morbid obesity: A BMI over 40kg/m^2 indicates morbid obesity, and is considered high risk for influenza. The 500 Cities Project reports the prevalence of obese people in each city with BMI over 30kg/m^2 (not necessarily morbid obesity). We use the data from table 1 in Sturm and Hattori [17] to estimate the proportion of people with BMI>30 that actually have BMI>40 (across the US); we then apply this to the 500 Cities obesity data to estimate the proportion of people who are morbidly obese in each city. Table 1 of Morgan et al. [18] suggests that 51.2% of morbidly obese adults have at least one other high risk chronic condition, and update our high-risk population estimates accordingly to account for overlap.

Pregnancy: We separately estimate the number of pregnant women in each age group and each city, following the methodology in CDC reproductive health report [19]. We assume independence between any of the high-risk factors and pregnancy, and further assume that half the population are women.

Estimating high-risk proportions for children. Since the 500 Cities Project only reports data for adults 18 years and older, we take a different approach to estimating the proportion of children at high risk for severe influenza. The two most prevalent risk factors for children are asthma and obesity; we also account for childhood diabetes, HIV and cancer.
From Miller et al. [20], we obtain national estimates of chronic conditions in children. For asthma, we assume that variation among cities will be similar for children and adults. Thus, we use the relative prevalences of asthma in adults to scale our estimates for children in each city. The prevalence of HIV and cancer in children are taken from CDC HIV surveillance report [16] and cancer research report [21], respectively.

We first estimate the proportion of children having either asthma, diabetes, cancer or HIV (assuming no overlap in these conditions). We estimate city-level morbid obesity in children using the estimated morbid obesity in adults multiplied by a national constant ratio for each age group estimated from Hales et al. [22], this ratio represents the prevalence in morbid obesity in children given the one observed in adults. From Morgan et al. [18], we estimate that 25% of morbidly obese children have another high-risk condition and adjust our final estimates accordingly.

Resulting estimates. We compare our estimates for the Houston-The Woodlands-Sugar Land MSA to published national-level estimates [23] of the proportion of each age group with underlying high risk conditions (Table A7). We estimate a higher proportion of individuals at high-risk for complications for COVID-19 compared to the national average in almost every age group, with the biggest difference observed in 10-19y and 35-64y groups.
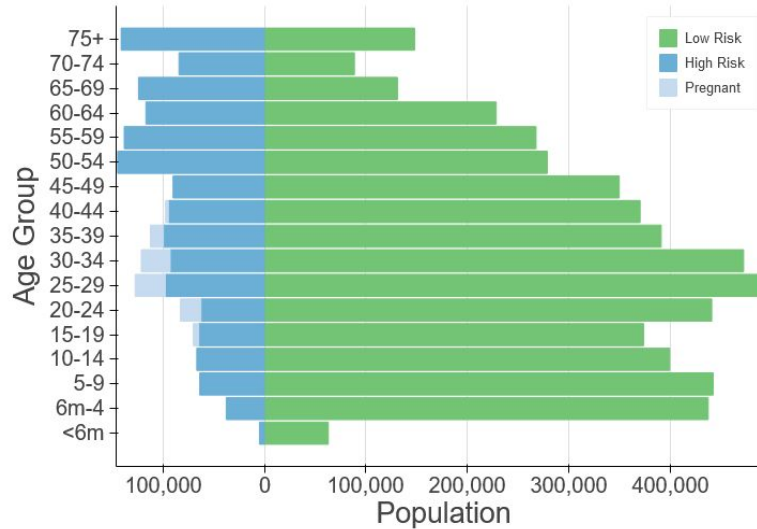
Figure A2. Demographic and risk composition of the Houston-The Woodlands-Sugar Land MSA. Bars indicate age-specific population sizes, separated by low risk, high risk, and pregnant. High risk is defined as individuals with cancer, chronic kidney disease, COPD, heart disease, stroke, asthma, diabetes, HIV/AIDS, and morbid obesity, as estimated from the CDC 500 Cities Project [13], reported HIV prevalence [16] and reported morbid obesity prevalence [17,18], corrected for multiple conditions. The population of pregnant women is derived using the CDC's method combining fertility, abortion and fetal loss rates [24–26].

Table A6. High-risk conditions for influenza and data sources for prevalence estimation

Condition	Data source
Cancer (except skin), chronic kidney disease, COPD, coronary heart disease, stroke, asthma, diabetes	CDC 500 cities [13]
HIV/AIDS	CDC HIV Surveillance report [16]
Obesity	CDC 500 cities [13], Sturm and Hattori [17], Morgan et al. [18]
Pregnancy	National Vital Statistics Reports [24] and abortion data [25]

Table A7. Comparison between published national estimates and Houston-The Woodlands-Sugar Land MSA estimates of the percent of the population at high-risk of influenza/COVID-19 complications.

Age Group	National estimates [22]	Houston (excluding pregnancy)	Pregnant women (proportion of age group)
0 to 6 months	NA	7.1	-
6 months to 4 years	6.8	7.9	-
5 to 9 years	11.7	12.6	-
10 to 14 years	11.7	14.3	-
15 to 19 years	11.8	14.6	2.5
20 to 24 years	12.4	11.9	7.9
25 to 34 years	15.7	15.7	9.8
35 to 39 years	15.7	19.8	5.2
40 to 44 years	15.7	20.2	1.2
45 to 49 years	15.7	20.5	-
50 to 54 years	30.6	34.1	-
55 to 60 years	30.6	34.0	-
60 to 64 years	30.6	33.8	-
65 to 69 years	47.0	48.6	-
70 to 74 years	47.0	48.7	-
75 years and older	47.0	48.8	-

References

1. Ready Harris > Stay Home. [cited 27 Apr 2020]. Available: https://www.readyharris.org/Stay-Home

2. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med. 2020. doi:10.7326/M20-0504

3. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the Asymptomatic Proportion of 2019 Novel Coronavirus onboard the Princess Cruises Ship, 2020. Infectious Diseases (except HIV/AIDS). medRxiv; 2020. doi:10.1101/2020.02.20.20025866

4. Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton University Press; 2011.

5. Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. J Chem Phys. 2001;115: 1716–1733.

6. minimize(method='trust-constr') — SciPy v1.4.1 Reference Guide. [cited 19 Apr 2020]. Available: https://docs.scipy.org/doc/scipy/reference/optimize.minimize-trustconstr.html

7. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of COVID-19 disease. Epidemiology. medRxiv; 2020. doi:10.1101/2020.03.09.20033357

8. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. The serial interval of COVID-19 from publicly reported confirmed cases. Epidemiology. medRxiv; 2020. doi:10.1101/2020.02.19.20025452

9. CDC. People at High Risk of Flu. In: Centers for Disease Control and Prevention [Internet]. 1 Nov 2019 [cited 26 Mar 2020]. Available: https://www.cdc.gov/flu/highrisk/index.htm

10. CDC - BRFSS. 5 Nov 2019 [cited 26 Mar 2020]. Available: https://www.cdc.gov/brfss/index.html

11. Zhang X, Holt JB, Lu H, Wheaton AG, Ford ES, Greenlund KJ, et al. Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. Am J Epidemiol. 2014;179: 1025–1033.

12. Tindale L, Coombe M, Stockdale JE, Garlock E, Lau WYV, Saraswat M, et al. Transmission interval estimates suggest pre-symptomatic spread of COVID-19. Epidemiology. medRxiv; 2020. doi:10.1101/2020.03.03.20029983

13. 500 Cities Project: Local data for better health | Home page | CDC. 5 Dec 2019 [cited 19 Mar 2020]. Available: https://www.cdc.gov/500cities/index.htm

14. Health Outcomes | 500 Cities. 25 Apr 2019 [cited 28 Mar 2020]. Available: https://www.cdc.gov/500cities/definitions/health-outcomes.htm

15. Part One: Who Lives with Chronic Conditions. In: Pew Research Center: Internet, Science & Tech [Internet]. 26 Nov 2013 [cited 23 Nov 2019]. Available: https://www.pewresearch.org/internet/2013/11/26/part-one-who-lives-with-chronic-conditions/

16. for Disease Control C, Prevention, Others. HIV surveillance report. 2016; 28. URL: http://www cdc gov/hiv/library/reports/hiv-surveil-lance html Published November. 2017.

17. Sturm R, Hattori A. Morbid obesity rates continue to rise rapidly in the United States. Int J Obes . 2013;37: 889–891.

18. Morgan OW, Bramley A, Fowlkes A, Freedman DS, Taylor TH, Gargiullo P, et al. Morbid obesity as a risk factor for hospitalization and death due to 2009 pandemic influenza A(H1N1) disease. PLoS One. 2010;5: e9694.

19. Estimating the Number of Pregnant Women in a Geographic Area from CDC Division of Reproductive Health. Available: https://www.cdc.gov/reproductivehealth/emergency/pdfs/PregnacyEstimatoBrochure508.pdf

20. Miller GF, Coffield E, Leroy Z, Wallin R. Prevalence and Costs of Five Chronic Conditions in Children. J Sch Nurs. 2016;32: 357–364.

21. Cancer Facts & Figures 2014. [cited 30 Mar 2020]. Available: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2014.html

22. Hales CM, Fryar CD, Carroll MD, Freedman DS, Ogden CL. Trends in Obesity and Severe Obesity Prevalence in US Youth and Adults by Sex and Age, 2007-2008 to 2015-2016. JAMA. 2018;319: 1723–1725.

23. Zimmerman RK, Lauderdale DS, Tan SM, Wagener DK. Prevalence of high-risk indications for influenza vaccine varies by age, race, and income. Vaccine. 2010;28: 6470–6477.

24. Martin JA, Hamilton BE, Osterman MJK, Driscoll AK, Drake P. Births: Final Data for 2017. Natl Vital Stat Rep. 2018;67: 1–50.

25. Jatlaoui TC, Boutot ME, Mandel MG, Whiteman MK, Ti A, Petersen E, et al. Abortion Surveillance - United States, 2015. MMWR Surveill Summ. 2018;67: 1–45.

26. Ventura SJ, Curtin SC, Abma JC, Henshaw SK. Estimated pregnancy rates and rates of pregnancy outcomes for the United States, 1990-2008. Natl Vital Stat Rep. 2012;60: 1–21.