THE ROLE OF AGE DISTRIBUTION AND FAMILY STRUCTURE ON COVID-19 DYNAMICS: A PRELIMINARY MODELING ASSESSMENT FOR HUBEI AND LOMBARDY

Bryan Wilder[1], Marie Charpignon[2], Jackson A. Killian[1], Han-Ching Ou[1], Aditya Mate[1], Shahin Jabbari[1], Andrew Perrault[1], Angel Desai[3], Milind Tambe[1], Maimuna S. Majumder[4,5]

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
[2]MIT Institute for Data, Systems, and Society, Cambridge, MA, USA
[3]International Society for Infectious Diseases, Brookline, MA, USA
[4]Department of Pediatrics, Harvard Medical School, Boston, MA, USA
[5]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA

March 31, 2020

DISCLAIMER: This work is provisional and will be updated as further information becomes available.

ABSTRACT

Background: The COVID-19 outbreak has already caused significant mortality worldwide. As the epidemic accelerates, understanding the transmission dynamics of COVID-19 is crucial to informing national and regional policies. We develop an individual-level model for SARS-CoV2 transmission which accounts for location-dependent distributions of age and household structure. We apply our model to Hubei, China and Lombardy, Italy to analyze the impact of demographic structure on estimates for key parameters such as the rate of documentation and the reproduction number r_0 for COVID-19 cases. We also assess the effectiveness of potential policies ranging from physical distancing to sheltering in place in Lombardy.

Methods: Our study develops a stochastic, agent-based model for SARS-CoV2 spread. A key feature of the model is the inclusion of population-specific demographic structure, such as the distributions of age, household structure, contact across age groups, and comorbidities. We use prior estimates of these demographic features to instantiate our model for two locations: Hubei, China and Lombardy, Italy. Furthermore, we utilize the data on the number of reported deaths due to COVID-19 in both locations to estimate parameters describing location-specific variation in the transmissibility and fatality of the disease (for reasons beyond demography). The range of the parameters in our model that are consistent with reported data are used to construct plausible ranges for r_0 and the rate of documentation in each location. Finally, we analyze potential policy responses in the context of Lombardy. Our analysis traces out the trade-off between adoption of physical distancing across the entire population and policies that encourage members of a specific age group to shelter at home.

Results: Our estimates for r_0 are comparable to the rest of the literature, with a range of 2.11–2.27 for Hubei and 2.50-3.20 for Lombardy, suggesting higher rates of transmission in the latter. Scenarios where the case fatality rates are higher in Lombardy than Hubei by a factor of 1-5 times appear plausible given the data (even after accounting for differences in age and comorbidity distributions). We estimate the rate at which symptomatic cases are documented to be at 10.3-19.2% in Hubei and 1.2-8% in Lombardy, indicating that the number of undocumented cases may be even higher than has previously been estimated. Evaluation of potential policies suggests that encouraging a single age group to shelter in place is insufficient to control the epidemic by itself, but that targeted "salutary sheltering" by even 50% of a single age group has a substantial impact when combined with adoption of physical distancing by the rest of the population.

1 Introduction

Since December 2019, the ongoing COVID-19 pandemic – caused by the the novel coronavirus, SARS-CoV2 – has resulted in significant morbidity and mortality [1]. As of March 28, 2020, an estimated 664,000 individuals have been infected, with over 30,000 fatalities worldwide [2]. Certain key factors such as existing comorbidities, age, and potentially gender have appeared to play a role in an increased risk of mortality [3]. Epidemiological studies have provided significant insights into the disease to date; however, as the pace of the pandemic continues to accelerate in certain regions of the world, understanding factors related to the transmission dynamics of SARS-CoV2 will be critical to mitigating its spread [4–7]. Moreover, as national and regional governments begin to implement broad-reaching policies in response to rising case counts and stressed healthcare systems, estimating the impact of these policies while accounting for transmission-related factors will be vital.

Prior studies have developed accurate prediction and control models in the setting of other emerging outbreaks including West Nile Virus, Avian influenza and SARS, among others [8–10]. Some previous studies have focused on developing analytical or algorithmic understanding of policies for epidemic control or eradication [11–16], while others have examined techniques for constructing realistic epidemic simulation environments [17–19].

The aim of this study is to employ mathematical modeling to evaluate the impact of age distribution and familial household contacts on transmission using existing data from Hubei, China, and Lombardy, Italy – two regions that have been characterized as epicenters for SARS-CoV2 infection – and describe how the implications of these findings may affect the utility of potential non-pharmaceutical interventions at a country-level.

A key feature of this study is the incorporation of demographic structure, including age distribution, age-stratified contact patterns, household structures, and comorbidity distributions. Existing epidemiological work characterizing the impact of age and household structure (outside the specific context of COVID-19) falls into two categories: simpler, analytically tractable models [13, 20–23] and highly detailed, individual-level simulations, e.g. for national-scale pandemic influenza [24–28]. Such simulations typically contain hundreds of millions of agents and track individual households, communities, and workplaces. However, many do not age-stratify contacts at the level of granularity in our model, instead modeling coarser groups such as school- or work-aged individuals [24, 26, 29], with a few exceptions [27, 28, 30]. Our agent-based model charts a middle course between analytical models and individual-level simulations by focusing on demography and forgoing features such as detailed simulation of the spatial structure of an entire country. This ensures computational tractability, while also avoiding overfitting to the limited data currently available by restricting the number of parameters associated with each agent. While such parsimony limits the granularity of our simulation, it allows for rapid iteration during an emerging epidemic while focusing on the impact of demography on COVID-19 modeling and policy.

2 Methods

This section introduces our agent-based model for COVID-19 spread along with methods used to parameterize the model. Code and data to run the model and reproduce experiments from the study can be found at `https://github.com/bwilder0/COVID19-Demography`.

2.1 Agent-based model description

We develop an agent-based model for COVID-19 spread which accounts for the distributions of age, household types, comorbidities, and contact between different age groups in a given population. The model follows a *susceptible-exposed-infectious-removed (SEIR)* template [11, 14]. Specifically, we simulate a population of n agents (or individuals), each with an age a_i, a set of comorbidities c_i, and a household (a set of other agents). We stratify age into ten-year intervals and incorporate hypertension and diabetes as comorbidities. These comorbidities are common worldwide [31] and have been associated with a higher risk of in-hospital death for COVID-19 patients [3]. However, our model can be expanded to include other comorbidities of interest in the future. The specific procedure we use to sample agents from the joint distribution of age, household structures, and comorbidities may be found in Appendix A.

The simulation tracks two states for each individual: the *infection state* and the *isolation state*. The infection state is divided into *fsusceptible, exposed, infectious, removedg*. *Susceptible* individuals are those who have never been contacted by an infectious individual. *Exposed* individuals are those who have had contact with an infectious individual, though not all exposed individuals become infectious. If an exposed individual contracts the disease, they proceed to the infectious state.[1] *Infectious* is further subdivided into severity levels *fasymptomatic, mild, severe, criticalg*.

[1] Currently, our simulation implementation does not separately track individuals who are exposed but do not become infected, and instead groups them with the susceptible population. This is because we assume that, if exposed again, they will become infected

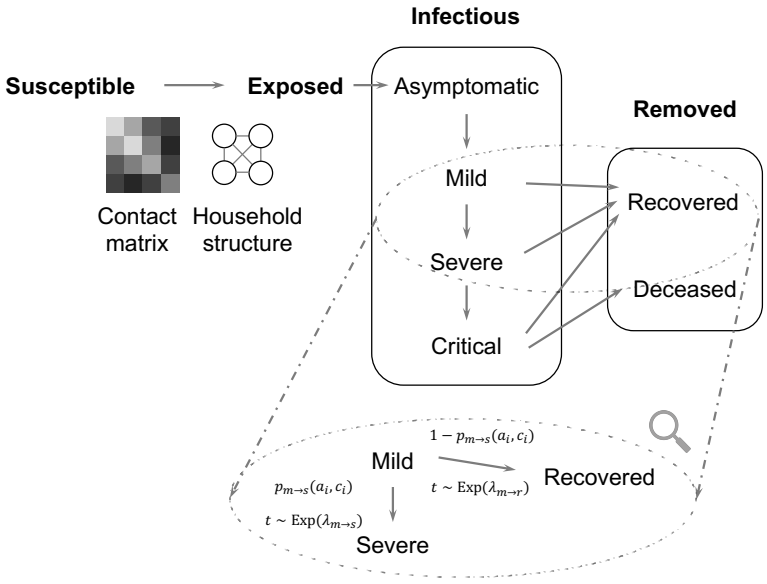

Figure 1: We use a modified SEIR model, where the infectious states are subdivided into levels of disease severity. The transitions are probabilistic and there is a time lag for transitioning between states. For example, the magnified section shows the details of transitions between mild, recovered, and severe states. Each arrow consists of the probability of transition (e.g., $p_{m\,!\,s}(a_i, c_i)$ to progress from mild to severe) as well as the associated time lag for the transition (e.g. the time t to progression from mild to severe is drawn from an exponential distribution with mean $\lambda_{m\,!\,s}$). a_i and c_i denote the age and set of comorbidities of the infected individual i.

We interpret mild severity as symptomatic (but not requiring hospitalization), severe as requiring hospitalization, and critical as eligible for intensive care unit (ICU) care. The *removed* state is further subdivided into {*recovered*, *deceased*}. These states and the transitions between them are summarized in Figure 1. Individuals in all severity levels can transmit the disease, but those in the *asymptomatic* state do so at a rate $\alpha < 1$ times that of symptomatic cases. The decision to incorporate reduced transmission for asymptomatic individuals is based on the fact that, though infection by asymptomatic individuals has been observed in case clusters and in examinations of serial intervals [32–34], available evidence suggests that individuals with no or limited symptoms are less infectious than those with severe symptoms [6]. Currently, our simulation incorporates two levels of infectiousness (before and after the onset of symptoms), but it can be adjusted as better information on how viral shedding increases with severity of illness becomes available. *We acknowledge that our assumptions surrounding transmissibility and disease severity – as derived from existing literature – may serve as a limitation of our model, as many of these factors are evolving over time.*

Each individual has a separate isolation state {*isolated*, *not isolated*}. If isolated, the individual is unable to infect others. We assume that (1) asymptomatic individuals are never isolated, (2) mild individuals become isolated over a mean time of λ_{isolate} days (see Table 1) after the onset of symptoms, and (3) all severe and critical individuals are isolated. However, our simulation framework can easily accommodate different sets of assumptions about isolation (for example, preemptively isolating exposed individuals if they are known to have had contact with an infectious agent).

The disease is transmitted over a contact structure, which is divided into in-household and out-of-household groups. Each agent has a household consisting of a set of other agents (see Appendix A for details on how households are generated using country-specific census information). Individuals infect members of their households at a higher rate than out-of-household agents. We model out-of-household transmission using country-specific estimated contact matrices [35]. These matrices state the mean number of daily contacts an individual of a particular age strata has with individuals from each of the other age strata. *We assume demographics (including age and household distribution) in Hubei and Lombardy are well-approximated by country-level data.*

The model iterates over a series of discrete time steps, each representing a single day, from a starting time t_0 to an end time T. There are two main components to each time step: disease progression and new infections. The progression component is modeled by drawing two random variables for each individual each time they change severity levels (e.g.

with the same probability as an individual who has never been exposed. However, the implementation can be modified to support either differing probabilities of contracting the disease after first exposure or policies that treat exposed and susceptible individuals differently.

Table 1: Model parameters

Parameter	Description	Value and/or source
$p_{m\to s}(a_i, c_i)$	Prob. of progressing from mild to severe given age a_i and comorbidities c_i	Estimated from China CDC and US CDC data (see below)
$p_{s\to c}(a_i, c_i)$	Prob. of progressing from severe to critical given age a_i and comorbidities c_i	As above
$p_{c\to d}(a_i, c_i)$	Prob. of progressing from critical to death given age a_i and comorbidities c_i	As above
p_h	Prob. of infecting each household member each day	Calibrated to match [10]
p_{inf}	Prob. of infecting an outside household contact	Free parameter
$\mu_{e\to m}$	Log-mean time to progress from exposed to mild (mean incubation period)	1.621 [38]
$\sigma^2_{e\to m}$	Log-standard deviation time to progress from exposed to mild	0.418 [38]
$\lambda_{m\to s}$	Mean time to progress from mild to severe	7 days [39]
$\lambda_{s\to c}$	Mean time to progress from severe to critical	7.5 days (using 14.5 days from onset to mechanical ventilation in [3])
$\lambda_{c\to d}$	Mean time to progress from critical to death	4.5 days (subtracting $\lambda_{m\to s}$ and $\lambda_{s\to c}$ from onset-to-death in [3])
λ_{isolate}	Mean time for an individual in the mild state to isolate	4.6 days (time to first medical care [40])
$\lambda_{m\to r}$	Mean time to recovery for an individual in the mild state	14 days [39]
$\lambda_{s\to r}$	Mean time to recovery for an individual in the severe state	$28 - \lambda_{m\to s}$ (midpoint of onset-to-recovery for severe [39])
$\lambda_{c\to r}$	Mean time to recovery for an individual in the critical state	$35 - \lambda_{m\to s} - \lambda_{s\to c}$ (midpoint of [39] onset-to-recovery for critical)
α	Reduction in infectiousness before symptoms	0.55 [6] [2]
M	Contact matrix (for each country)	[35]
t_0	First date with at least 5 infected individuals	Free parameter

on entering the mild state). The first random variable is Bernoulli and indicates whether the individual will recover or progress to the next severity level. The second variable represents the amount of time until progression to the next severity level. We use exponential distributions for almost all time-to-event distributions, a common choice in the absence of specific distributional information [36, 37]. The exception is the incubation time between asymptomatic and mild states, where more specific information is available; here, we use a log-normal distribution (see $\mu_{e\,!\,m}$ and $\sigma^2_{e\,!\,m}$ in Table 1) based on estimates by Lauer et al. [38]. Table 1 summarizes all distributions and their parameters, and Section 2.2 describes how we estimate age- and comorbidity-dependent severity progression.

In the new infections component, individuals in the susceptible state may enter the exposed state. Infected individuals infect each of their household members with probability p_h at each time step. p_h is calibrated so that the total probability of infecting a household member before either isolation or recovery matches the estimated secondary attack rate for household members of COVID-19 patients (i.e., the average fraction of household members infected) [10]. Infected individuals draw outside-of-household contacts from the general population using the country-specific contact matrix. For an infected individual of age group i, we sample w_{ij} Poisson(M_{ij}) contacts for each age group j, where M is the country-specific contact matrix. Poisson distributions are a standard choice for modeling contact distributions [35]. Then, we sample w_{ij} contacts of age j uniformly with replacement, and each contact is infected with the probability p_{inf}, the probability of infection given contact. This probability is an unknown; our experiments test a range of values, and we report which values give results consistent with observations from Lombardy and Hubei.

One key advantage of our modeling framework is its flexibility; namely, we can modify it to test different policies or simulate additional features with greater fidelity. Examples of future work that are easily accommodated by our model include:

Contact-tracing policies: Our simulation tracks the tree of who-infects-whom, easily enabling the simulation of policies that trace the tree with a given probability of successfully identifying each contact (the probability of success potentially varying by household contact or age group).

Health system capacity: Our simulation tracks the number of individuals at each severity level, allowing us to match the number of individuals in the critical state to the number of ICU beds per capita in a given area. We can increase the probability of death for critical individuals once their number exceeds the existing ICU capacity to model the interaction between ICU capacity and mortality.

Age-varying adherence to self-isolation for mildly ill individuals: We can alter the probability that a mildly ill individual will self-isolate based on age group, capturing the hypothesis that younger individuals may disproportionately fail to isolate while mildly symptomatic.

Multiple waves of infection: Our current study only models scenarios where interventions are imposed on a specific date and remain in place thereafter; however, future work could analyze scenarios where interventions are prematurely removed, leading to potential resurgence of the disease.

[2] This setting for α is likely pessimistic in that Li et al.'s estimate for reduction in transmissibility is for undocumented cases, including both asymptomatic cases and those with limited symptoms [6]. Future work should examine the impact of a potentially lower α as better information on transmissibility in the asymptomatic state becomes available.

2.2 Estimating disease progression from age and comorbidities

Many of the parameters for this model are assigned values based on estimates in the literature, shown in Table 1. However, we currently lack a detailed understanding of the joint impact of age and comorbidities on disease progression and mortality. Currently, case fatality rates (CFRs) are available either by age or by individual comorbidity, but not for each specific combination of age and comorbidities. To obtain these estimates, we model the CFR with a logistic regression fit to CFRs from the Chinese Center for Disease Control and Prevention (China CDC) [41]. This model yields $p_{m\,!\,d}(a_i, c_i)$, the country-independent probability that an individual i of age a_i and comborbidity status c_i will die if infected with SARS-CoV-2 (see Appendix B for more details). Corrections for country-specific differences in mortality are discussed in Section 2.3.

The simulation also requires specific values for the probabilities of transitioning between the disease states mild, severe, critical, and death. However, there is currently insufficient information available to infer the probabilities of these individual transitions for each combination of age and comorbidity. We assume that while the absolute values of these probabilities may vary based on age and comorbidity, the *ratios* between them do not exhibit such strong dependency. In particular, we assume that there are coefficients $\gamma_{s\,!\,c}(a_i)$ and $\gamma_{c\,!\,d}$ such that $p_{s\,!\,c}(a_i, c_i) = \gamma_{s\,!\,c}(a_i) p_{m\,!\,s}(a_i, c_i)$ and $p_{c\,!\,d}(a_i) = \gamma_{c\,!\,d} p_{m\,!\,s}(a_i, c_i)$. We allow $\gamma_{s\,!\,c}(a_i)$ to be age-specific while assuming that $\gamma_{c\,!\,d}$ is age-homogeneous because of the information currently available to estimate them; namely, we estimate $\gamma_{s\,!\,c}(a_i)$ based on the relative probabilities of hospitalization and ICU admission by age group in the US [42] and $\gamma_{c\,!\,d}$ based on the probability of death for all critical patients in China [41]. Note that we assume both coefficients to be independent of the comorbodities c_i. Then, we can solve for $p_{m\,!\,s}(a_i, c_i)$ such that

$$p_{m\,!\,s}(a_i, c_i) \quad \gamma_{s\,!\,c}(a_i) p_{m\,!\,s}(a_i, c_i) \quad \gamma_{c\,!\,d} p_{m\,!\,s}(a_i, c_i) = p_{m\,!\,d}(a_i, c_i),$$

and set $p_{s\,!\,c}(a_i, c_i)$ and $p_{c\,!\,d}(a_i, c_i)$ accordingly. Future work can relax the assumptions in this process as more information becomes available about how age and comorbidity impact the progression between disease states.

2.3 Free parameters

There are three main parameters for which values are not precisely estimated in the literature and are varied in-simulation. First is p_{inf}, the probability of infection given contact. This determines the level of transmissibility of the disease. Second is t_0, the start time of the infection, which is not precisely characterized in most locations and has a substantial impact due to rapid doubling times. Third is a parameter d_{mult}, which accounts for differences in the rate of mortality between locations that are *not* captured by demographic factors in the model (e.g., the impact of limited ICU capacity in Lombardy). d_{mult} is a multiplier which is applied to the baseline mortality rate estimated from China CDC data.

3 Results

We use our model to explore the impact of demography on our understanding of existing COVID-19 outbreaks and potential policies that aim to mitigate such outbreaks. Section 3.1 introduces two case studies where we validate the model against confirmed deaths due to COVID-19, explore the range of parameter settings consistent with the data, and illustrate how demography-aware modeling can inform estimates of the rate at which cases are documented. Documentation rates estimated by our model are generally lower than previous estimates, providing an example of how incorporating demography can inform inferences about key unknowns by accounting for how age-dependent patterns of behavior lead to differing disease prevalence across groups. Section 3.2 examines the interplay between different forms of "salutary sheltering" – a term we coin here to describe individuals who shelter in place irrespective of their exposure or infectious state – and physical distancing policies, while simultaneously illustrating how demography impacts the evaluation of these aforementioned policies.

3.1 Validation and inferred parameters

This section instantiates the model for outbreaks in two specific locations: Hubei, China (where the disease originated) and Lombardy, Italy (one of the most heavily-impacted areas thus far, along with Hubei). We show that our model is able to reproduce observed patterns in the number of reported deaths and examine how a range of possible underlying parameters are consistent with the data. Based on this range of parameters (which vary p_{inf}, t_0, and d_{mult}), we propose plausible ranges for r_0 in Hubei and Lombardy, as well as for the rate at which infected individuals are documented.

3.1.1 Hubei, China

We draw a population of individuals from the age, household, and comorbidity distributions for China since more specific information is not available for Hubei (though the fraction of individuals over 65 is within the typical range for

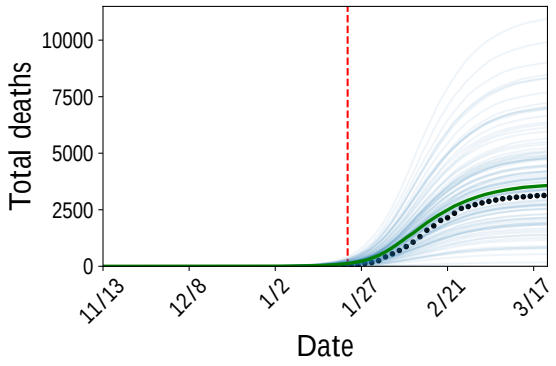

Figure 2: Simulated trajectories of the number of deaths over time in Hubei, China compared to the true number of reported deaths ($p_{\text{inf}} = 0.020$, t_0 = November 15). Light blue lines are individual trajectories, green is the median, and the black dots are the number of reported deaths. The red dashed line represents the January 23 lockdown. The true number of reported deaths is contained within the simulated distribution and lies close to the median (Pearson $r = 0.998$).

many Chinese provinces [43]). We simulate a population of 10 million individuals for computational tractability.[3] All results, here and in the remainder of the paper, use 100 independent runs of the simulation.

We examine times t_0 varying around the November 17 date for the first identified patient in Wuhan [45, 46]. We do not include d_{mult} as a free parameter (that is, $d_{\text{mult}} = 1$) because our baseline mortality rates are estimated on China CDC data drawn mostly from Hubei. We vary p_{inf} in the range $[0.018, 0.023]$ and characterize the resulting patterns in infection spread and fatalities. We simulate through March 21, with a lockdown on January 23. After the lockdown, all contact frequencies are reduced by a factor of 50, reflecting the reported severity of the restrictions imposed [47]. In practice, such constraints correspond to an average of 1 to 2 daily outside-of-household contacts for people aged 15-29 or 30-49, and at most one such contact a day on average for the 50-69 age group [35, 48].

Validating the simulated results is complex because cases are likely substantially underdocumented [44]. We fit to the number of deaths on the simulation end date as a metric for validation, since deaths are believed to be substantially better documented, despite the possibility for both underreporting (e.g., viral pneumonia that is not diagnosed as COVID-19 [49]) and overreporting (e.g., presumed cause of death classified as COVID-19 on death certificates [50] or classification of deaths as caused by COVID-19 independent of possible comorbidities [51]). We assume that all deaths are documented. Future work could relax this assumption by modeling a location-dependent probability for a death to be documented. Figure 2 shows the complete distribution of simulation results for one specific set of well-fitting parameters ($p_{\text{inf}} = 0.020$, t_0 = November 15). We find that the true number of reported deaths is contained within the simulation distribution and lies close to the median. Moreover, the shape of the curves match closely (Pearson $r = 0.998$ between the median of the simulated trajectories and the observations). Note that we only assess goodness of fit using the last entry of the trajectory, so a close match to the rest of the trajectory provides evidence against the possibility of overfitting.

While Figure 2 gives one set of plausible parameters, there are also a range of other parameters that fit the data well. Higher p_{inf} can compensate for a later t_0, allowing a given number of deaths to be reached even with a later start date. Accordingly, Figure 3 shows two outputs of the simulation across the entire parameter range, with heatmap saturation representing the goodness of fit between the simulated and true number of reported deaths. We measure goodness of fit by asking whether the true number of reported deaths is well-contained within the simulated distribution. Specifically, if p is the percentile of the true number of reported deaths in the distribution of the simulation runs, the color of the cell reflects $p(1 \quad p)$. A value of 0 represents that the real number of deaths is either greater or smaller than all simulation runs, and a value of 0.25 (the maximum possible) represents that the real number of deaths is exactly at the 50% percentile of the simulated distribution. Assessing fit via the mean squared error yields a similar parameter order but is more difficult to interpret.

[3] Note that Hubei has a population of approximately 58 million. However, even assuming that the rate of documentation is less than half of current estimates [44], less than 1% of the population of Hubei was infected by the virus by the end of our simulation on March 21. The number of simulated individuals is only relevant when the number of infections begins to saturate the population; otherwise, additional individuals will simply never be encountered in the simulation. 10 million is also comparable to the population of Wuhan, where most cases were concentrated [41].

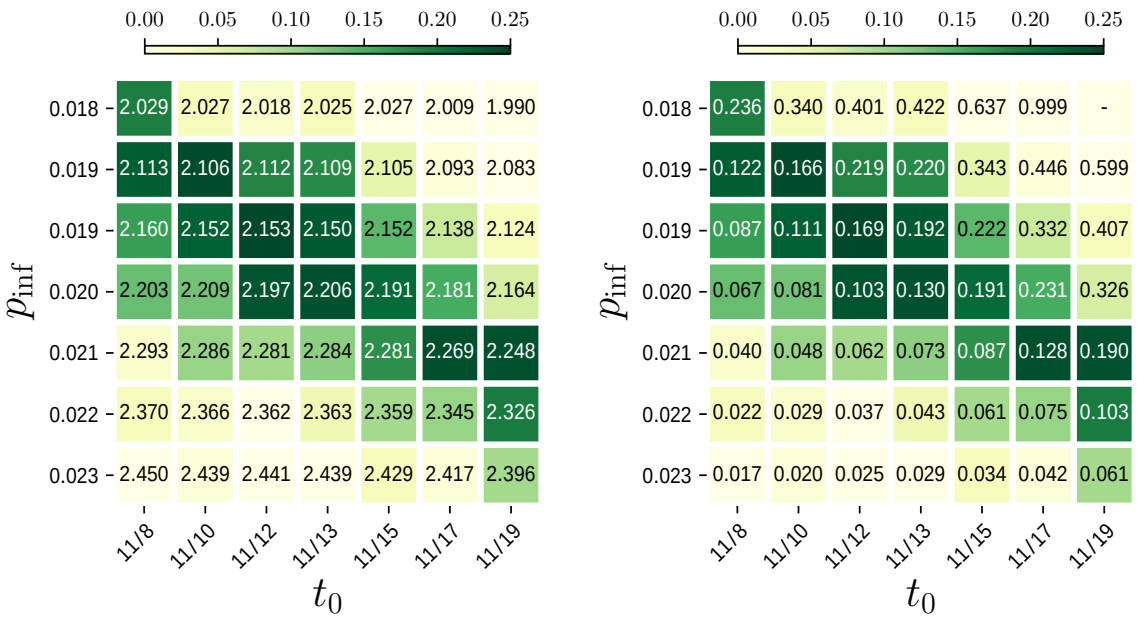

Figure 3: Left: median r_0 as a function of p_{inf} (y-axis) and t_0 (x-axis). Right: median fraction of symptomatic cases documented. Colors indicate the goodness of fit, where darker colors suggest better fit. A large set of parameter settings (the dark green diagonal entries) are consistent with the data, leading to a range of possible values for r_0 and the rate of documentation.

The dark cells along the diagonal have $p(1 \quad p)$ close to 0.25, indicating that the true number of reported deaths is well-contained within the simulated distribution across a range of parameter settings. We examine the implications of this range of possible parameter settings for two quantities of interest: the basic reproduction number r_0 and the documentation rate for symptomatic cases.

The left side of Figure 3 shows the inferred pre-lockdown r_0 given each set of parameters. The values are consistent with existing estimates, which largely fall in the range 2–3 [52]. We calculate a single plausible range for our results, which includes all values for parameter settings where the percentile p of the true number of reported deaths in the simulated distribution satisfies $p(1 \quad p) \quad 0.2$, capturing better-fitting parameterizations. For r_0, the plausible range is 2.11–2.27.

The right side of Figure 3 shows the inferred rate of documentation of symptomatic cases given each set of parameters. We calculate this by dividing the actual number of confirmed cases in Hubei on the simulation end date by the total number of symptomatic infections in the simulation. Each cell reports the median over a set of independent simulations for the corresponding parameters. We find that most scenarios give a somewhat lower rate of documentation than the range 28–33% estimated by Russell et al. [44]. This may be due to differences in documentation in Hubei compared to nationally; however, this would not account for all discrepancies as Hubei experienced most of the cases in China [41]. At least a portion of this lower documentation rate can instead be attributed to our model's inclusion of age-dependent contact patterns and disease severity. Younger individuals may have more total contacts than older individuals [35, 48, 53, 54] because of school or work. As such, younger individuals would be more likely to become infected, while also having a lower risk of fatality [39, 41]. Accordingly, the true CFR among all infected people is expected to be lower than in estimates that assume an age-homogeneous attack rate [44, 55] (where attack rate denotes the fraction of a group which is infected). To demonstrate this point, Figure 4 illustrates simulated attack rates by age for representative parameter settings in both Lombardy and Hubei. As expected, we observe inhomogeneous attack rates with higher attack rates among younger groups. As a result, the plausible range for the documentation rate in our simulation is 10.3-19.2% in Hubei. This illustrates how age-dependent behavioral patterns may impact estimates of important parameters, motivating the inclusion of demographic information in COVID-19 modeling.

3.1.2 Lombardy, Italy

We simulate a population of 10 million individuals (representing the population of Lombardy) drawn from the Italian distribution of age, household structure, and comorbidity status. The full demographic information needed to

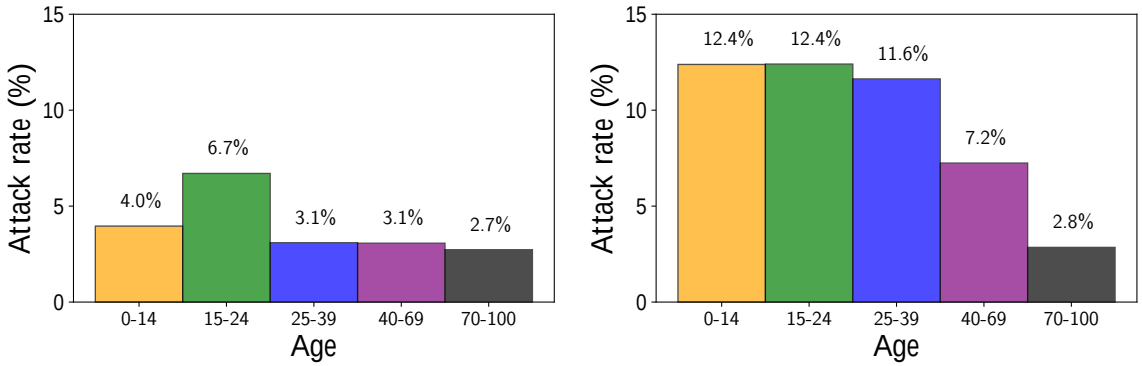

Figure 4: Median simulated attack rate as a function of age. Left: Hubei, China ($p_{\text{inf}} = 0.020$, t_0 = November 15). Right: Lombardy, Italy ($p_{\text{inf}} = 0.029$, t_0 = January 22). Simulated attack rates vary strongly with age and are higher for younger groups.

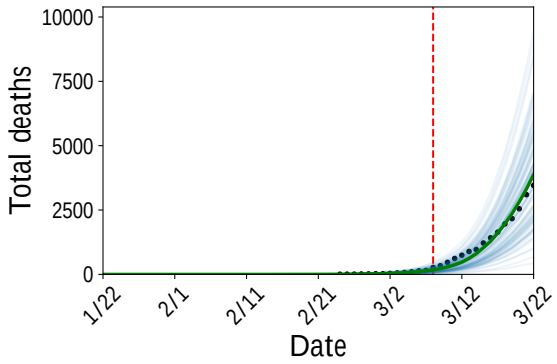

Figure 5: Simulated trajectories in the number of deaths over time in Lombardy, Italy compared to the true number of reported deaths ($p_{\text{inf}} = 0.029$, t_0 = January 22, $d_{\text{mult}} = 4$). Light blue lines are individual trajectories, green is the median, and the black dots give the number of reported deaths. The red dashed line represents the March 8 lockdown in Lombardy, Italy. Once more, the data are close to the median of the simulated distribution (Pearson r = 0.993).

parameterize the simulation was not available for Lombardy specifically, but available information suggests broadly similar characteristics (e.g., the median age in Lombardy is 45 [56], comparable to Italy in general at 46.5 [57]).

Our simulation starts from January 22 and continues till March 22, with a lockdown on March 8. After the lockdown, the number of contacts for all age groups is reduced by a factor of 10, reflecting a less severe lockdown than in Hubei [58]. In practical terms, such a level of physical distancing induces an average of 7 to 8 contacts a day for people aged 15-29 or 30-49, while it corresponds with about 3 to 4 interactions on average for the 50-69 age group [35]. We vary the two parameters p_{inf} and t_0 that we examined for Hubei, but now also consider the impact of d_{mult}. Recall that d_{mult} is a multiplier to the fatality rate across all ages and comorbidities, which capture location-dependent variation in fatalities in excess of differences due to demographic factors.

We measure the goodness of fit as in Section 3.1.1. Figures 6 and 7 show goodness of fit for a wide range of parameters, along with the associated estimates for r_0 and the rate of documentation. Within each figure, each heatmap corresponds to a different value of d_{mult}, the multiplier for fatality rates relative to Hubei. We observe that a wide range of possible scenarios are consistent with the data. Trajectories for one well-fitting set of parameters with $p_{\text{inf}} = 0.029$, t_0 = January 22 are shown in Figure 5. The true number of reported deaths again closely matches the median of the simulated distribution (Pearson $r = 0.993$). January 22nd is a plausible start date because of reports that infected travelers had landed in Milan by January 23 [59] (meaning that later dates are unlikely). However, these parameter settings are by no means the only possibility—it is possible that fatality rates in Lombardy are much higher than in Hubei (e.g., by a factor of 3-5), or that r_0 is significantly higher in Lombardy than in Hubei. Together, the simulations suggest that both factors likely contribute. We find broad support for the hypothesis that Lombardy has thus far experienced a more transmissible and deadlier outbreak in comparison to Hubei [60, 61]. Section 4 discusses possible explanations for this phenomenon.

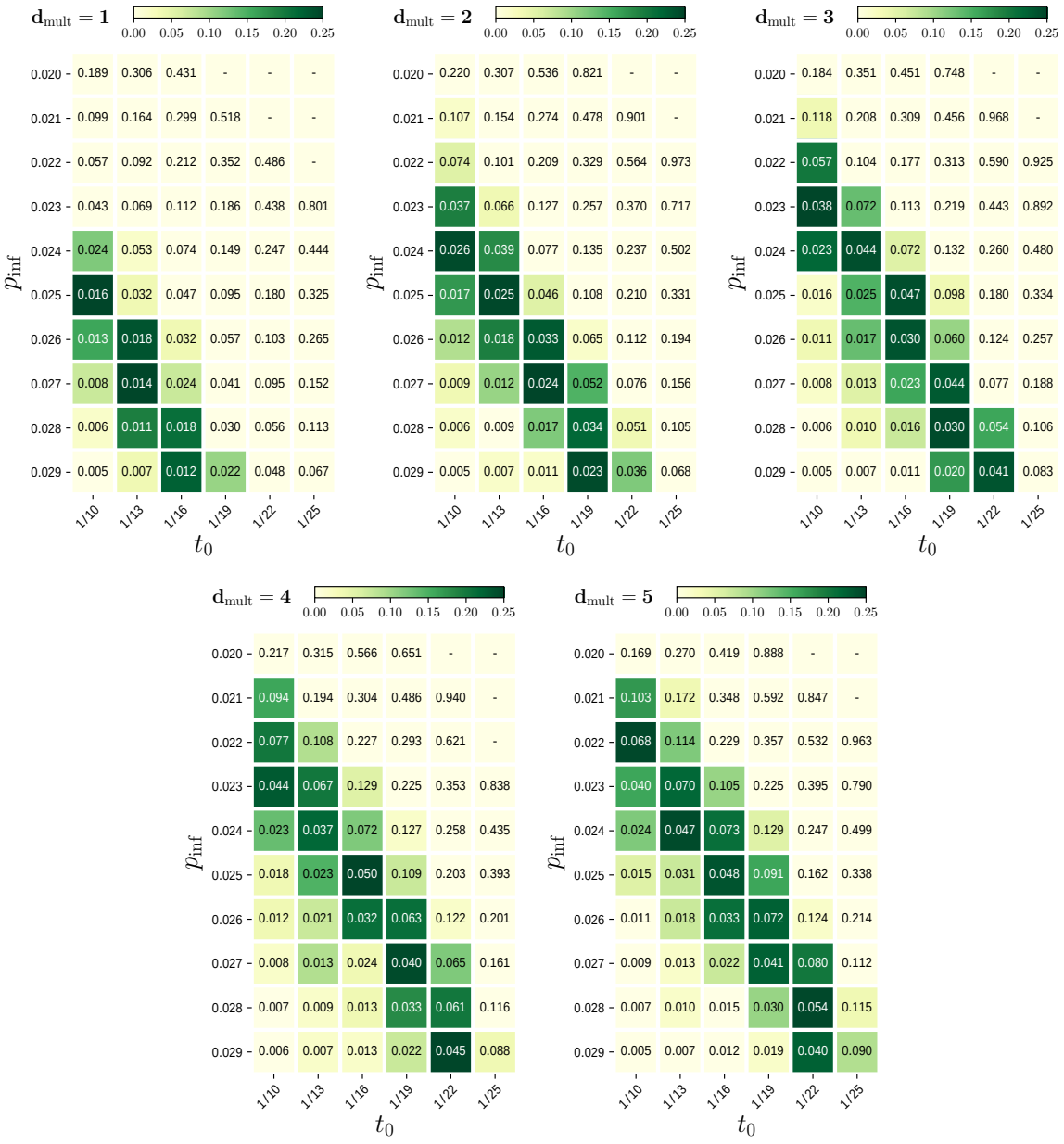

Figure 6: Fraction of symptomatic cases documented in Lombardy as a function of p_{inf} (y axis) and t_0 (x axis). From left to right, top to bottom: a mortality multiplier d_{mult} relative to China of 1, 2, 3, 4, 5. Colors indicate goodness of fit of the parameter settings to the true number of reported deaths on March 21 in Lombardy, with darker cells indicating a better fit. We again find a wide range of possible parameterizations given the observed data, with a plausible range for the documentation rate of 1.17-8.04%.

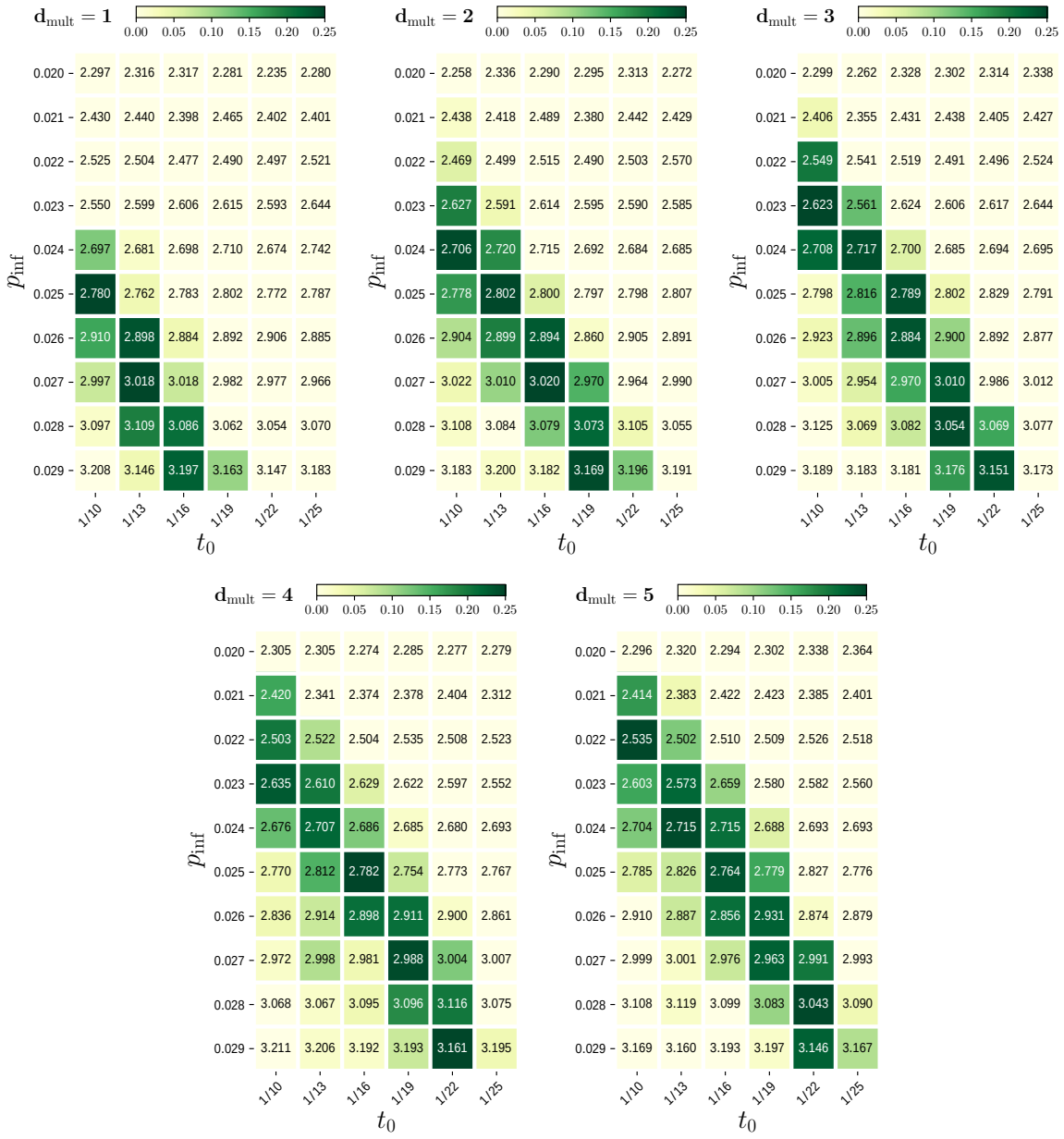

Figure 7: r_0 during the exponential growth (pre-lockdown) phase in Lombardy as a function of p_{inf} (y axis) and t_0 (x axis). From left to right, top to bottom: a mortality multiplier d_{mult} relative to China of 1, 2, 3, 4, 5. Colors indicate goodness of fit of the parameter settings to the true number of reported deaths on March 21 in Lombardy, with darker cells indicating a better fit. The plausible range for r_0 is 2.50-3.20.

Documentation rates appear to be lower under most well-fitting parameter settings than previously reported for Italy overall. Russell et al. [44] reported a 95% confidence interval of 4.6–5.9%, while our model yields a plausible range of 1.17–8.04%. Our plausible ranges broadly agree with Russell et al. [44] that documentation rates are substantially lower in Lombardy than in China. However, our range contains substantially lower values. By accounting for demographic factors, a substantially greater disease prevalence in Lombardy is possible. For r_0, we obtain the plausible range 2.50–3.20, which is disjointed with (and higher than) the corresponding plausible range for Hubei.

3.2 Containment policies in Lombardy: salutary sheltering and physical distancing

Various interventions have been implemented in different countries to slow the spread of SARS-CoV2. These policies range from staggering lockdown of entire cities or even countries, such as the two-month Hubei province lockdown [46] and the ongoing Italian lockdown of Lombardy and 14 neighboring provinces [62], to milder policies such as physical distancing—the practice of reducing physical interactions across the population as a whole. Within these lie a range of possible alternatives. Our model allows us to simulate all these policies, along with many others, because we can implement different forms of daily contact reduction for individual people within specific age groups. For example, a government could encourage some percentage of a given age group to remain sheltered in place (e.g., members of the workforce aged 45 and older who are more likely to be able to continue their professional activity from home [63]), while the rest of the population can continue commuting to their workplace and engaging in in-person social activities. The rationale for age-specific policies is two-fold. First, age-specific policies have already been employed in some countries; in the US for example, the Centers for Disease Control and Prevention (CDC) recommended in early March that people aged 65 years and older shelter in place [64]. Second, along with other factors such as education level and occupation type [65], current telecommuting patterns are intrinsically related to age groups—home workers being more likely to fall into the older age groups as compared to onsite workers [66, 67].

As an initial policy scenario, we investigate to what extent the epidemic can be mitigated by encouraging a single age group to engage in "salutary sheltering" or whether the broader population must also be asked to adopt some form of physical distancing. For example, even those who do not shelter in place could have their work hours staggered to reduce contact, be prescribed specific times to shop for groceries, and so on. Accordingly, we compare two scenarios. In the first scenario, we simulate salutary sheltering for a fraction of a single age cohort while leaving the rest of the population's behavior unaltered. In the second scenario, not only do we simulate salutary sheltering for a fraction of a single age cohort, but we also simulate physical distancing measures among the rest of the population.

In either scenario, we simulate 50% or 100% of a single age cohort as engaged in salutary sheltering. We run these simulations such that (some fraction of) cohorts aged 0–14, 15–29, 30–49, 50–69, and 70 years and older engage in salutary sheltering, to study where targeting such a policy may have the greatest impact, on the total percentage of the population infected, the total (cumulative) number of deaths, or both. We model salutary sheltering as removing all between-household contacts for the "sheltered" individuals, though they are still able to infect and be infected by their household members. We model physical distancing as reducing the expected number of daily contacts between non-sheltered individuals of any two age groups by a factor of two. Here, we study the effect of these transmission-mitigating policies in the Lombardy region of Italy because the epidemic is still ongoing there [62], unlike in Hubei where the situation is currently under control and the end of the lockdown has been officially announced [46]. Throughout this section, we use the same parameter values as in Figure 5 ($p_{\text{inf}} = 0.029, t_0 =$ January 22, $d_{\text{mult}} = 4$), and simulate a scenario where a given policy is implemented on day 46 of the simulation (corresponding to the timing of the actual lockdown in Lombardy on March 8).

Figure 8 shows the first policy scenario, where a portion of a single age group engages in salutary sheltering but the rest of the population does not adopt physical distancing. The left plots (Figures 8(a) and 8(c)) show salutary sheltering of 50% of a single age group, while the right plots (Figures 8(b) and 8(d)) show salutary sheltering of 100% of a single age group. We evaluate these interventions according to two metrics: the percentage of the entire population that is infected (top row) and the total number of deaths (bottom row). Each colored line shows the impact of salutary sheltering of a single given age group. In addition, we consider three baselines policies: the gray dotted curve reflects a baseline scenario with no intervention (no salutary sheltering, no physical distancing)[4] and the blue and pink dotted curves correspond to baselines where 50% and 100% of the entire population would engage in salutary sheltering, respectively.

[4]To put our projections into perspective and assess the credibility of the estimated percentage of infected, we compare the "no intervention" projections to a standard SEIR model [68]. When using the well-fitting set of parameters $p_{\text{inf}} = 0.029$, t_0 = January 22, and d_{mult} = 4, Figure 7 gives an estimated r_0 of 3.16. For this r_0, the standard SEIR model predicts that 95% of the population is eventually infected. Our "no intervention" baseline plateaus at approximately 90% infected, with the difference attributable to the fact that even our baseline scenario contains isolation of individuals in the severely and critically infected states, along with isolation of some mildly infected individuals as determined by the mean time-to-isolation parameter λ_{isolate}. Beyond the standard SEIR model, the fraction infected in "no intervention" baseline is also broadly consistent with other estimates in the literature [69–71].

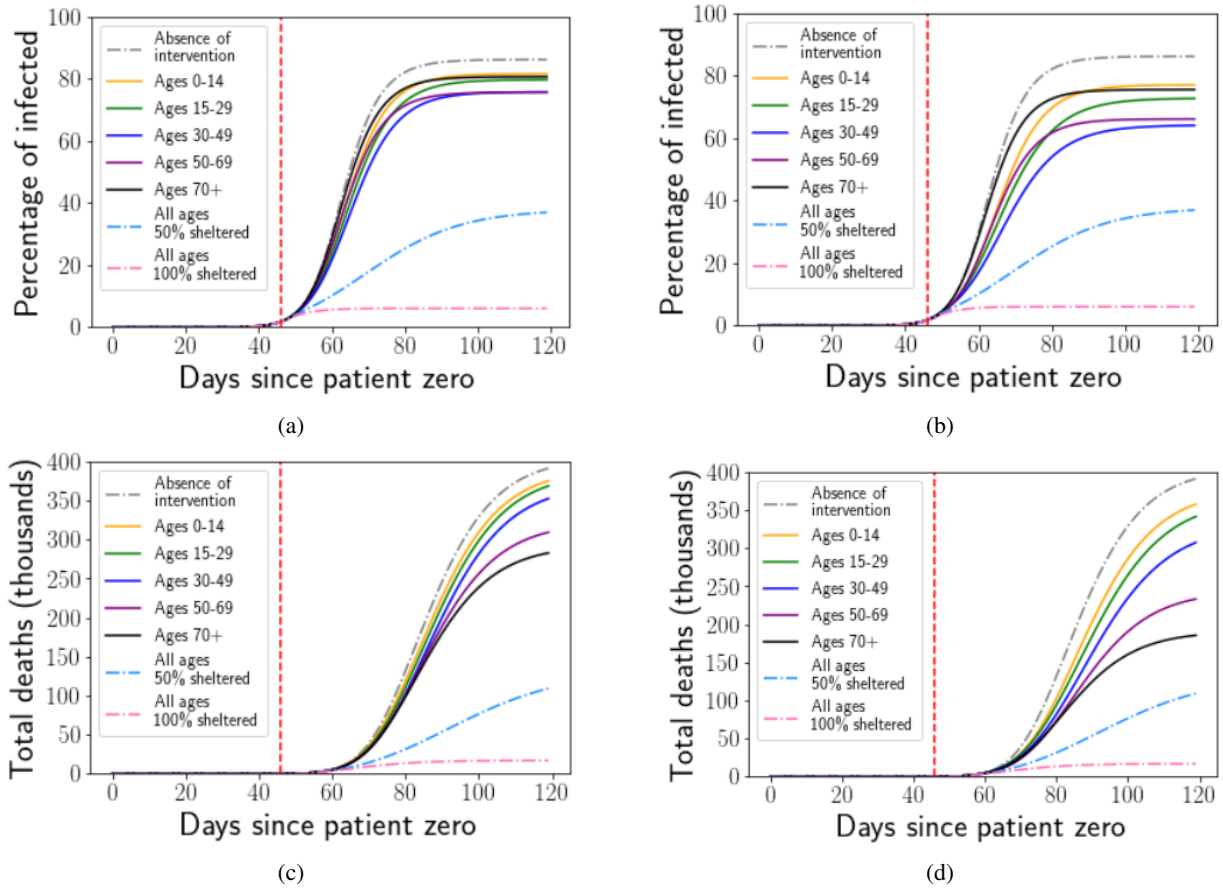

Figure 8: Effect of salutary sheltering for a fraction of each age cohort on the percentage of the population infected (top row, (a) and (b)) and the total number of deaths (bottom row, (c) and (d)), on a four-month horizon after the beginning of the outbreak in Lombardy on January 22 and in the absence of any physical distancing measure for other individuals. Left side, (a) and (c): salutary sheltering 50% of a single given age cohort in place. Right side, (b) and (d): salutary sheltering 100% of a single given age cohort. The red vertical dashed line indicates the start of the intervention, set to match the real March 8 lockdown in Lombardy. Each solid colored curve represents sheltering a specific age cohort in place, while the gray dotted curve represents the "no intervention" scenario (no salutary sheltering, no physical distancing), and the blue and pink dotted curves correspond to baselines where 50% and 100% of the population would be sheltered, respectively.

It is worth noting that while there is no difference on average between 100% of the whole population reducing their physical interactions by a factor of 2 (i.e., physical distancing) and having 50% of the whole population engage in salutary sheltering, the variance in the number of daily contacts–limited to household members–an individual may have in the latter is much lower.

In the case of Lombardy, Italy, we observe that even complete shelter in place of any single age group leaves at least 60% of the population infected (Figure 8b), and nearly 70-80% in the more realistic scenario where only half the group engages in sheltering (Figure 8a). Notably, the only policies that limit the percentage of infected, or the total number of deaths, would require the sheltering of 50% ((a) and (c)) to 100% ((b) and (d)) of the entire population. Except for these two scenarios, illustrated by the blue and pink dotted curves respectively, the total number of deaths as projected by our model in Lombardy, Italy is correspondingly large (above 200 thousand) in the absence of any level of physical distancing and can only be significantly ameliorated by completely sheltering the entire population of 70 years and older – an intervention that would be both unreasonable and unjust, irrespective of age group (Figure 8d).

However, combining partial salutary sheltering of a single age group with physical distancing by the rest of the population has a substantially greater impact, as shown in Figure 9 for Lombardy, Italy. Four months from the beginning of the outbreak, the fraction of the population infected drops to 50% or below (depending on which age group, and what percentage, is sheltered). The total number of deaths projected for this time horizon decreases even more substantially,

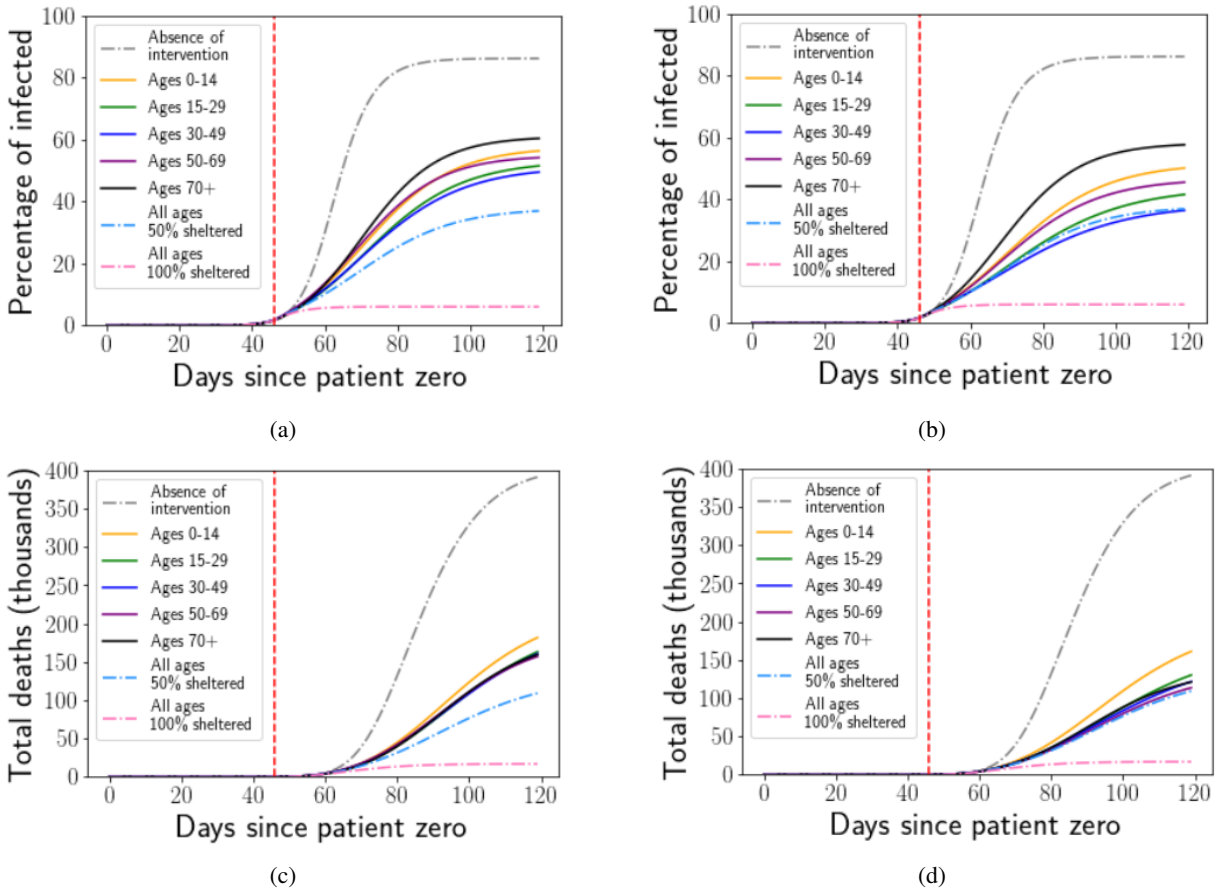

Figure 9: Effect of salutary sheltering for a fraction of each age group on the percentage of the population infected (top row, (a) and (b)) and the total number of deaths (bottom row, (c) and (d)), on a four-month horizon after the beginning of the outbreak in Lombardy on January 22 combined with physical distancing by the rest of the population. Left-hand side, (a) and (c): salutary sheltering of 50% of a single given age group in place. Right-hand side, (b) and (d): salutary sheltering of 100% of a single given age group. The red vertical dashed line indicates the start of the intervention, set to match the real March 8 lockdown in Lombardy. Each solid colored curve represents sheltering a specific age cohort in place, while the gray dotted curve represents the "no intervention" scenario (no salutary sheltering, no physical distancing), and the blue and pink dotted curves correspond to baselines where 50% and 100% of the population would be sheltered, respectively.

to the range of 100–150 thousand (as compared to 275–400 thousand for the previous scenario with 50% sheltering by a single age group and no population-wide physical distancing). In this scenario, where physical distancing is adopted by everyone, there is only a small benefit, particularly with respect to total deaths, to sheltering 100% of an age group (9b and 9d) instead of 50% (9a and 9c). Further, the age group sheltered need not necessarily be the 70+ group; in fact, 50% salutary sheltering of the 30-49 age group in Lombardy, Italy has a somewhat larger impact. This is associated with a higher level of physical interactions on a daily basis among members of this generation [35], especially given that Lombardy is the second most densely populated region in Italy, home to the economic capital Milan. The fact that we observe differences in the effectiveness of sheltering different age groups in Lombardy, Italy thus stems from the interplay of the relative size of each age group [72], daily contact patterns (younger groups have more interactions within and outside their generation [35]), and risk of mortality from the disease (older groups are at higher risk [3, 64]). It is worth noting that the greater impact of the 30-49 age group cannot solely be explained by their prevalence in Lombardy. While this age group accounts for 26% of Italians, they are only the second largest generation after the 50-69 age group that represents 28% of the population [72].

Building upon the case of Lombardy, Italy, our model suggests that hybrid policies, combining targeted salutary sheltering by one part of the population and physical distancing by the rest, could be as effective at limiting the final size of the outbreak as salutary sheltering of an entire sub-population, while preserving our social ties and avoiding

complete disruption of the economy. Future work can use our model to more closely investigate policies which account for these differences and formulate targeted recommendations about levels of salutary sheltering and physical distancing by age group or any other suitable stratification adapted to the country's population and workforce. Such an analysis would be able to draw on additional information related to demography, for example the fraction of a given age group readily prepared to work from home (which has been studied in the labor economics and urban planning literature, for both industrialized [66, 73] and developing countries [74]). To the extent to which such physical distancing measures are compliant with citizen rights and the social acceptance of restricted movement mandated by either national or local governments, corporations, or any other institutions responsible for their practical enforcement, these policies should be adapted to the political system in place, as well as the social characteristics of each country—a capability that our model incorporates.

4 Discussion and future work

In this study, we developed a model of SARS-CoV2 transmission that incorporates household distribution, age, and comorbidities in Hubei, China, and Lombardy, Italy and created population simulations using available demographic information from these two locations. Our findings suggest that population-wide sheltering in place may not be necessary to mitigate the spread of COVID-19; instead, targeted salutary sheltering of specific age groups combined with adherence to physical distancing on the part of the general population may be sufficient to thwart a substantial fraction of cases and deaths. This could be achieved by engaging in activities such as staggered work schedules, increasing spacing in restaurants, and prescribing times to use the gym or grocery store.

From a pragmatic perspective, targeted salutary sheltering may not be realistic for all populations. The feasibility of such a policy relies on those who have access to safe shelter, which does not reflect reality for all individuals. In addition, while this model illustrates that targeted salutary sheltering in combination with widespread physical distancing may decrease transmission of SARS-CoV2, sociopolitical realities may render this recommendation more feasible in certain populations as compared to others. Concerns for personal liberty, discrimination against certain sub-segments of the population, and broad-based societal acceptability may prevent the successful adoption of targeted salutary sheltering in some regions of the world. Allowing salutary sheltering to operate on a voluntary basis that is predicated on the use of a shift system (rather than for indefinite time periods) may address some of these issues.

Existing modeling work of COVID-19 largely focuses on simpler compartmental or branching process models that do not consider demography and thus do not allow for the simulation of creative policies like those aforementioned. Furthermore, while such models have played an important role in estimating key parameters such as r_0 [5, 7] or the rate at which cases are documented [75], as well as in the evaluation of prospective non-pharmaceutical interventions [76–78], they are not able to characterize how differences in demography impact the course of an epidemic in a particular location. Our focus on population-specific demography allows for further refinement of current mortality estimates and is a strength of this study. r_0 estimates in this study are generally comparable to other estimates in the literature to date [52], although our model yields estimates that were higher in Lombardy as compared to Hubei. While the reasons behind this are uncertain, factors including mask-wearing practices in certain regions compared to others [79–81] or differential adoption of behavioral interventions such as hand hygiene [82] may have contributed to these findings. Similarly, our model suggests a higher mortality rate in Lombardy than Hubei for reasons beyond demography. The explanation for this discrepancy is likely multifactorial, including differences in ICU capacity [83], coordination problems in Lombardy specifically to transport patients where there was capacity [83, 84], documented prevalence of antibiotic resistance affecting outcomes in patients affected by secondary bacterial infections [85], and perhaps greater background population exposure to other novel coronaviruses in China [86].

Reporting rates estimated in this study were generally lower than those that have been established via prior studies [44]. This is likely due to our simulations producing infection rates that are not homogeneous across age groups. Heterogeneous infection rates are attributable, at least in part, to that conception that younger people may have more contacts compared to other age groups [35, 48, 53, 54].

Our model has the advantage of being parsimonious while also considering these aforementioned heterogeneities, which is a highly desirable feature given uncertainties in data reporting. Notably, this limits the risk of overfitting when estimating the conditional probability of death.

Nevertheless, this study has several limitations that should be acknowledged. While several comorbidities associated with mortality in COVID-19 were accounted for in this model, the availability of existing data limited the incorporation of all relevant comorbidities. Most notably, chronic pulmonary disease was not included in this study although it has been associated with morbidity and mortality in COVID-19 [41], nor was smoking, despite its prevalence in both China and Italy [87, 88]. Gender-mediated differences were also excluded, which may be important for both behavioral reasons (e.g., adoption of hand-washing [89, 90]) and biological reasons (e.g, the potential protective role of estrogen in

SARS-CoV infections [91]). All of these factors can be incorporated into the model in our future work, particularly as better data becomes available.

Additionally, it is worth noting that we have not yet attempted to model super-spreader events in our existing framework. Such events may have been especially consequential in South Korea [92], and future work could attempt to more closely model the epidemic there via the incorporation of a dispersion parameter into the contact distribution, which has been used in other models [5].

Despite these limitations, this study demonstrates the importance of considering population and household demographics when attempting to better define transmission dynamics for COVID-19. In addition, this model highlights potential policy implications for non-pharmaceutical interventions that account for population-specific demographic features and may provide alternative strategies for national and regional governments moving forward.

References

[1] David Baud, Xiaolong Qi, Karin Nielsen-Saines, Didier Musso, Léo Pomar, and Guillaume Favre. Real estimates of mortality following COVID-19 infection. *The Lancet*, 2020.

[2] Center for Systems Science and Engineering at Johns Hopkins University. Coronavirus COVID-19 Global Cases, 2020. `https://coronavirus.jhu.edu/map.html`, Last Accessed: 2020-03-29.

[3] Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 2020.

[4] Bo Xu, Alomía de Gutiérrez, Sumiko Mekaru, Kara Sewalk, Lauren Goodwin, Alyssa Loskill, Emily Cohn, Yulin Hswen, Sarah Hill, María Mercedes Cobo, Alexander Zarebski, et al. Epidemiological data from the COVID-19 outbreak, real-time case information. *Scientific Data*, 7, 2020.

[5] Julien Riou and Christian Althaus. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance*, 25(4), 2020.

[6] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, 2020.

[7] Adam Kucharski, Timothy Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, and Rosalind Eggo. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 2020.

[8] Vincent Covello, Richard Peters, Joseph Wojtecki, and Richard Hyde. Risk communication, the west nile virus epidemic, and bioterrorism: Responding to the communication challenges posed by the intentional or unintentional release of a pathogen in an urban setting. *Journal of Urban Health*, 78(2):382–391, 2001.

[9] Marc Baguelin, Albert Jan Van Hoek, Mark Jit, Stefan Flasche, Peter White, and W John Edmunds. Vaccination against pandemic influenza A/H1N1v in England: a real-time economic evaluation. *Vaccine*, 28(12):2370–2384, 2010.

[10] Yang Liu, Rosalind Eggo, and Adam Kucharski. Secondary attack rate and superspreading events for SARS-CoV-2. *The Lancet*, 2020.

[11] Frank Ball, Edward Knock, and Philip O'Neill. Stochastic epidemic models featuring contact tracing with delays. *Mathematical Biosciences*, 266:23–35, 2015.

[12] Sudip Saha, Abhijin Adiga, Aditya Prakash, and Anil Kumar Vullikanti. Approximation algorithms for reducing the spectral radius to control epidemic spread. In *Proceedings of SIAM International Conference on Data Mining*, pages 568–576, 2015.

[13] David Tudor. An age-dependent epidemic model with application to measles. *Mathematical Biosciences*, 73(1): 131–147, 1985.

[14] Pauline Van den Driessche, Michael Li, and James Muldowney. Global stability of SEIRS models in epidemiology. *Canadian Applied Mathematics Quarterly*, 7:409–425, 1999.

[15] Fan Chung, Paul Horn, and Alexander Tsiatas. Distributing antidote using pagerank vectors. *Internet Mathematics*, 6(2):237–254, 2009.

[16] Hanghang Tong, Aditya Prakash, Tina Eliassi-Rad, Michalis Faloutsos, and Christos Faloutsos. Gelling, and melting, large graphs by edge manipulation. In *Proceedings of ACM International Conference on Information and Knowledge Management*, pages 245–254, 2012.

[17] Christopher Barrett, Keith Bisset, Stephen G Eubank, Xizhou Feng, and Madhav Marathe. EpiSimdemics: An efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *Proceedings of the ACM/IEEE Conference on Supercomputing*, pages 1–12, 2008.

[18] Stephen Eubank, Hasan Guclu, Anil Kumar, Madhav Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.

[19] Meghendra Singh, Achla Marathe, Madhav Marathe, and Samarth Swarup. Behavior model calibration for epidemic simulations. In *Proceedings of the International Conference on Autonomous Agents and MultiAgent Systems*, pages 1640–1648, 2018.

[20] Frank Ball and Owen Lyne. Stochastic multi-type SIR epidemics among a population partitioned into households. *Advances in Applied Probability*, 33(1):99–123, 2001.

[21] Frank Ball and Owen Lyne. Optimal vaccination policies for stochastic epidemics among a population of households. *Mathematical Biosciences*, 177:333–354, 2002.

[22] Thomas House and Matt Keeling. Deterministic epidemic models with explicit household structure. *Mathematical Biosciences*, 213(1):29–39, 2008.

[23] Bryan Wilder, Sze-Chuan Suen, and Milind Tambe. Preventing infectious disease in dynamic populations under uncertainty. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

[24] Neil Ferguson, Derek Cummings, Christophe Fraser, James Cajka, Philip Cooley, and Donald Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452, 2006.

[25] Steven Riley. Large-scale spatial-transmission models of infectious disease. *Science*, 316(5829):1298–1301, 2007.

[26] Elizabeth Halloran, Neil Ferguson, Stephen Eubank, Ira Longini, Derek Cummings, Bryan Lewis, Shufu Xu, Christophe Fraser, Anil Vullikanti, Timothy Germann, Diane Wagener, et al. Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences*, 105(12):4639–4644, 2008.

[27] Marco Ajelli, Bruno Gonçalves, Duygu Balcan, Vittoria Colizza, Hao Hu, José Ramasco, Stefano Merler, and Alessandro Vespignani. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infectious Diseases*, 10(1):190, 2010.

[28] Marco Ajelli, Stefano Merler, Andrea Pugliese, and Caterina Rizzo. Model predictions and evaluation of possible control strategies for the 2009 A/H1N1v influenza pandemic in Italy. *Epidemiology & Infection*, 139(1):68–79, 2011.

[29] Neil Ferguson, Derek Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sopon Iamsirithaworn, and Donald S Burke. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437(7056):209–214, 2005.

[30] Lorenzo Pellis, Neil Ferguson, and Christophe Fraser. Threshold parameters for a model of epidemic spread among households and workplaces. *Journal of the Royal Society Interface*, 6(40):979–987, 2009.

[31] Gregory Roth, Degu Abate, Kalkidan Hassen Abate, Solomon Abay, Cristiana Abbafati, Nooshin Abbasi, Hedayat Abbastabar, Foad Abd-Allah, Jemal Abdela, Ahmed Abdelalim, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 392(10159):1736–1788, 2018.

[32] Yan Bai, Lingsheng Yao, Tao Wei, Fei Tian, Dong-Yan Jin, Lijuan Chen, and Meiyun Wang. Presumed asymptomatic carrier transmission of COVID-19. *The Journal of the American Medical Association*, 2020.

[33] Camilla Rothe, Mirjam Schunk, Peter Sothmann, Gisela Bretzel, Guenter Froeschl, Claudia Wallrauch, Thorbjörn Zimmer, Verena Thiel, Christian Janke, Wolfgang Guggemos, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *New England Journal of Medicine*, 2020.

[34] Zhanwei Du, Xiaoke Xu, Ye Wu, Lin Wang, Benjamin Cowling, and Lauren Ancel Meyers. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerging Infectious Diseases*, 2020.

[35] Kiesha Prem, Alex Cook, and Mark Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Computational Biology*, 13(9):e1005697, 2017.

[36] Paul Allison. *Survival analysis using SAS: a practical guide*. SAS Institute, 2010.

[37] David Collett. *Modelling survival data in medical research*. CRC Press, 2015.

[38] Stephen Lauer, Kyra Grantz, Qifang Bi, Forrest Jones, Qulu Zheng, Hannah Meredith, Andrew Azman, Nicholas Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 2020.

[39] World Health Organization China. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), 2020. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf, Last Accessed: 2020-03-29.

[40] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy Leung, Eric Lau, Jessica Wong, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 2020.

[41] Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19). *China CDC Weekly*, 2(8):113–122, 2020. http://weekly.chinacdc.cn//article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51, Last Accessed: 2020-03-28.

[42] CDC COVID-19 Response Team. Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12–March 16, 2020, 2020. https://www.cdc.gov/mmwr/volumes/69/wr/mm6912e2.htm, Last Accessed: 2020-03-29.

[43] Xizhe Peng. China's demographic history and future challenges. *Science*, 333(6042):581–587, 2011.

[44] Timothy Russell, Joel Hellewell, Sam Abbott, Christopher Jarvis, Kevin van Zandvoort, Stefan Flasche, Rosalind Eggo, John Edmunds, and Adam Kucharski. Using a delay-adjusted case fatality ratio to estimate under-reporting, 2020. https://cmmid.github.io/topics/covid19/severity/global_cfr_estimates.html, Last Accessed: 2020-03-29.

[45] South China Morning Post. Coronavirus: China's first confirmed COVID-19 case traced back to November 17, 2020. https://www.scmp.com/news/china/society/article/3074991/coronavirus-chinas-first-confirmed-covid-19-case-traced-back, Last Accessed: 2020-03-28.

[46] Wikipedia. Timeline of the 2019–20 coronavirus pandemic in November 2019 – January 2020, 2020. https://en.wikipedia.org/wiki/Timeline_of_the_2019-20_coronavirus_pandemic_in_November_2019_-_January_2020, Last Accessed: 2020-03-28.

[47] New York Times. China tightens Wuhan lockdown in wartime battle with coronavirus, 2020. https://www.nytimes.com/2020/02/06/world/asia/coronavirus-china-wuhan-quarantine.html, Last Accessed: 2020-03-29.

[48] Jonathan Read, Justin Lessler, Steven Riley, Shuying Wang, Li Jiu Tan, Kin On Kwok, Yi Guan, Chao Qiang Jiang, and Derek Cummings. Social mixing patterns in rural and urban areas of Southern China. *Proceedings of the Royal Society B: Biological Sciences*, 281(1785):20140268, 2014.

[49] Harrison Bai, Ben Hsieh, Zeng Xiong, Kasey Halsey, Ji Whae Choi, Thi My Linh Tran, Ian Pan, Lin-Bo Shi, Dong-Cui Wang, Ji Mei, et al. Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. *Radiology*, 2020.

[50] The Telegraph. Why have so many coronavirus patients died in Italy?, 2020. https://www.telegraph.co.uk/global-health/science-and-disease/have-many-coronavirus-patients-died-italy/, Last Accessed: 2020-03-29.

[51] Graziano Onder, Giovanni Rezza, and Silvio Brusaferro. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *The Journal of the American Medical Association*, 2020.

[52] Maimuna Majumder and Kenneth Mandl. Early in the epidemic: Impact of preprints on global discourse of 2019-nCOV transmissibility. *SSRN*, 2020.

[53] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS medicine*, 5(3), 2008.

[54] Sara Del Valle, James Hyman, Herbert Hethcote, and Stephen Eubank. Mixing patterns between age groups in social networks. *Social Networks*, 29(4):539–554, 2007.

[55] Robert Verity, Lucy Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick Walker, Han Fu, et al. Estimates of the severity of COVID-19 disease. *medRxiv*, 2020.

[56] Italian National Institute of Statistics. Median age in Lombardy, 2019. https://www4.istat.it/it/lombardia/dati?qt=gettable&dataset=DCIS_INDDEMOG1&dim=21,0,0, Last Accessed: 2020-03-28.

[57] CIA World Factbook. Field listing: median age, 2020. https://www.cia.gov/library/publications/the-world-factbook/fields/343.html, Last Accessed: 2020-03-28.

[58] Lorenzo Tondo. Italy charges more than 40,000 people with violating lockdown. *The Guardian*, 2020. https://www.theguardian.com/world/2020/mar/18/italy-charges-more-than-40000-people-violating-lockdown-coronavirus, Last Accessed: 2020-03-29.

[59] Fabrizio Carinci. COVID-19: Preparedness, decentralisation, and the hunt for patient zero. *British Medical Journal*, 2020.

[60] Jason Horowitz, Emma Bubola, and Elisabetta Polvedo. Italy, pandemic's new epicenter, has lessons for the world. *The New York Times*, 2020. https://www.nytimes.com/2020/03/21/world/europe/italy-coronavirus-center-lessons.html, Last Accessed: 2020-03-29.

[61] Valentina Di Donato, Sheena McKenzie, and Livia Borghese. Italy's coronavirus death toll passes 10,000. Many are asking why the fatality rate is so high. *CNN*, 2020. https://www.cnn.com/2020/03/28/europe/italy-coronavirus-cases-surpass-china-intl/index.html, Last Accessed: 2020-03-29.

[62] Wikipedia. 2020 coronavirus pandemic in italy, 2020. https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_Italy, Last Accessed: 2020-03-29.

[63] Economics U.S. Department of Commerce and U.S. Census Bureau Statistics Administration. Selected economic characteristics, American community survey, 2018. https://data.census.gov/cedsci/table?q=commute&hidePreview=false&table=DP03&tid=ACSDP5Y2018.DP03&lastDisplayedRow=37, Last Accessed: 2020-03-29.

[64] Centers for Disease Control and Prevention. People who are at higher risk for severe illness, 2020. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-at-higher-risk.html, Last Accessed: 2020-03-29.

[65] Kathy Gardner and Kate Lister. 2017 state of telecommuting in the U.S. employee workforce, 2017. https://www.flexjobs.com/2017-State-of-Telecommuting-US/, Last Accessed: 2020-03-29.

[66] Peter Mateyka, Melanie Rapino, and Liana Christin Landivar. Home-based workers in the United States: 2010, household economic studies, current population reports, 2012. https://www.census.gov/prod/2012pubs/p70-132.pdf, Last Accessed: 2020-03-29.

[67] U.S. Bureau of Labor Statistics. Labor force statistics from the current population survey; employed and unemployed full- and part-time workers by age, sex, race, and hispanic or latino ethnicity (household data, annual averages, item 8), 2019. https://www.bls.gov/cps/cpsaat08.htm, Last Accessed: 2020-03-29.

[68] Joel Miller. A note on the derivation of epidemic final sizes. *Bulletin of Mathematical Biology*, 74:2125–2141, 2012.

[69] Andrea Remuzzi and Giuseppe Remuzzi. COVID-19 and Italy: What next? *The Lancet*, 2020.

[70] Jose Lourenco, Robert Paton, Mahan Ghafari, Moritz Kraemer, Craig Thompson, Peter Simmonds, Paul Klenerman, and Sunetra Gupta. Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic. *medRxiv*, 2020.

[71] The Imperial College COVID-19 Response Team. Report 12: The global impact of COVID-19 and strategies for mitigation and suppression, 2020. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/news--wuhan-coronavirus/, Last Accessed:2020-03-29.

[72] United Nations. World population prospects 2019. 2019. https://population.un.org/wpp/, Last Accessed: 2020-03-29.

[73] Nicholas Bloom, James Liang, John Roberts, and Zhichun Jenny Ying. Does working from home work? evidence from a chinese experiment. *The Quarterly Journal of Economics*, 130(1):165–218, 2014.

[74] Martha Chen and Shalini Sinha. Home-based workers and cities. *Environment and Urbanization*, 28(2):343–358, 2016.

[75] Pablo De Salazar, Rene Niehus, Aimee Taylor, Caroline Buckee, and Marc Lipsitch. Using predicted imports of 2019-nCoV cases to determine locations that may not be identifying all imported cases. *medRxiv*, 2020.

[76] Stephen Kissler, Christine Tedijanto, Marc Lipsitch, and Yonatan Grad. Social distancing strategies for curbing the COVID-19 epidemic. *medRxiv*, 2020.

[77] Joel Hellewell, Sam Abbott, Amy Gimma, Nikos Bosse, Christopher Jarvis, Timothy Russell, James Munday, Adam Kucharski, John Edmunds, Sebastian Funk, and Rosalind Eggo. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 2020.

[78] Neil Ferguson, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Imperial College, London*, 2020.

[79] South China Morning Post. Face masks and coronavirus: how culture affects your decision to wear one, 2020. https://www.scmp.com/news/china/society/article/3075211/face-masks-and-coronavirus-how-culture-affects-your-decision, Last Accessed: 2020-03-29.

[80] Shuo Feng, Chen Shen, Nan Xia, Wei Song, Mengzhen Fan, and Benjamin Cowling. Rational use of face masks in the COVID-19 pandemic. *The Lancet Respiratory Medicine*, 2020.

[81] Time. Why Wearing a Face Mask Is Encouraged in Asia, but Shunned in the U.S., 2020. https://time.com/5799964/coronavirus-face-mask-asia-us/, Last Accessed: 2020-03-29.

[82] Gabriella Di Giuseppe, Rossella Abbate, Luciana Albano, Paolo Marinelli, and Italo Angelillo. A survey of knowledge, attitudes and practices towards avian influenza in an adult population of Italy. *BMC Infectious Diseases*, 8(1):36, 2008.

[83] Giacomo Grasselli, Antonio Pesenti, and Maurizio Cecconi. Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response. *The Journal of the American Medical Association*, 2020.

[84] Benedetta Armocida, Beatrice Formenti, Silvia Ussai, Francesca Palestra, and Eduardo Missoni. The Italian health system and the COVID-19 challenge. *The Lancet Public Health*, 2020.

[85] Alessandro Cassini, Dominique Monnet, Giovanni Mancarella, Michael Borg, José Miguel Cisneros, and Ute Wolff Sönksen. ECDC country visit to Italy to discuss antimicrobial resistance issues, 2017. https://www.ecdc.europa.eu/sites/default/files/documents/AMR-country-visit-Italy.pdf, Last Accessed: 2020-03-29.

[86] Vincent Cheng, Susanna Lau, Patrick Woo, and Kwok Yung Yuen. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clinical Microbiology Reviews*, 20(4):660–694, 2007.

[87] Mark Parascandola and Lin Xiao. Tobacco and the lung cancer epidemic in china. *Translational Lung Cancer Research*, 8:S21–S30, 2019.

[88] Alessandra Lugo, Piergiorgio Zuccaro, Roberta Pacifici, Giuseppe Gorini, Paolo Colombo, Carlo La Vecchia, and Silvano Gallus. Smoking in italy in 2015-2016: Prevalence, trends, roll-your-own cigarettes, and attitudes towards incoming regulations. *Tumori Journal*, 103(4):353–359, 2017.

[89] Maryellen Guinan, Maryanne McGuckin-Guinan, and Alice Sevareid. Who washes hands after using the bathroom? *American Journal of Infection Control*, 25(5):424–425, 1997.

[90] Durell Johnson, Danielle Sholcosky, Karen Gabello, Robert Ragni, and Nicole Ogonosky. Sex differences in public restroom handwashing behavior associated with visual behavior prompts. *Perceptual and Motor Skills*, 97 (3):805–810, 2003.

[91] Rudragouda Channappanavar, Craig Fett, Matthias Mack, Patrick Ten Eyck, David Meyerholz, and Stanley Perlman. Sex-based differences in susceptibility to severe acute respiratory syndrome coronavirus infection. *The Journal of Immunology*, 198(10):4046–4053, 2017.

[92] British Broadcasting Corporation. Coronavirus: South Korea emergency measures as infections increase, 2020. https://www.bbc.com/news/world-asia-51582186, Last Accessed: 2020-03-29.

[93] Zhan Hu and Xizhe Peng. Household changes in contemporary China: An analysis based on the four recent censuses. *The Journal of Chinese Sociology*, 2(1):9, 2015.

[94] Dan He, Xuying Zhang, Zhili Wang, and Yu Jiang. China fertility report, 2006–2016. *China Population and Development Studies*, 2(4):430–439, 2019.

[95] Statista. Household structures in Italy in 2018, 2018. https://www.statista.com/statistics/730604/family-structures-italy/, Last Accessed: 2020-03-28.

[96] Statista. Number of single-person households in Italy from 2012 to 2018, 2018. https://www.statista.com/statistics/728061/number-of-single-person-households-italy/, Last Accessed: 2020-03-28.

[97] Statista. Number of couples with children in Italy from 2012 to 2018, by number of children, 2018. https://www.statista.com/statistics/570106/number-of-couples-with-children-italy/, Last Accessed: 2020-03-28.

[98] Statista. Biennial average number of household members in Italy from 2012 to 2018, 2018. https://www.statista.com/statistics/671945/biennial-average-number-of-families-with-children-italy/, Last Accessed: 2020-03-28.

[99] Statista. Number of single parents in Italy from 2011 to 2018, by number of children, 2018. https://www.statista.com/statistics/570234/number-of-single-parents-with-children-in-italy-by-number-of-children/, Last Accessed: 2020-03-28.

[100] Elisabetta Carrà, Margherita Lanz, and Semira Tagliabue. Transition to adulthood in Italy: An intergenerational perspective. *Journal of Comparative Family Studies*, 45(2):235–248, 2014.

[101] Yu Xu, Limin Wang, Jiang He, Yufang Bi, Mian Li, Tiange Wang, Linhong Wang, Yong Jiang, Meng Dai, Jieli Lu, Min Xu, Yichong Li, Nan Hu, Jianhong Li, Shengquan Mi, Chung-Shiuan Chen, et al. Prevalence and Control of Diabetes in Chinese Adults. *The Journal of the American Medical Association*, 310(9):948–959, 09 2013.

[102] Zengwu Wang, Zuo Chen, Linfeng Zhang, Xin Wang, Guang Hao, Zugui Zhang, Lan Shao, Ye Tian, Ying Dong, Congyi Zheng, Jiali Wang, Manlu Zhu, William Weintraub, and Runlin Gao. Status of hypertension in China. *Circulation*, 137(22):2344–2356, 2018.

[103] Pietro Modesti, Maria Calabrese, Ilaria Marzotti, Hushao Bing, Danilo Malandrino, Maria Boddi, Sergio Castellani, and Dong Zhao. Prevalence, awareness, treatment, and control of hypertension among Chinese first-generation migrants and Italians in Prato, Italy: The CHIP study. *International Journal of Hypertension*, 2017.

[104] Yukako Tatsumi and Takayoshi Ohkubo. Hypertension with diabetes mellitus: significance from an epidemiological perspective for Japanese. *Hypertension Research*, 40(9):795–806, 2017.

A Sampling agents

Our process for sampling agents follows three steps that successively sample households, individual agents within households, and comorbidities for each agent. Because the full joint distributions over all of these quantities are not known, we implement a sampling procedure that respects the marginal distributions of household structure and age, as well as the marginal distribution for the occurrence of comorbidities within each age group.

First, we use information on the distribution of household structures to draw a type of household (e.g., single person, couple, nuclear family, or multigenerational family). Second, we sample the ages of the individual agents according to their role in the household (e.g., parent, child, or grandparent) combined with information about the age distribution of the population and the intergenerational interval. For China, we use household distributions from the 2010 Chinese census [93], intergenerational intervals from He et al. [94], and the age distribution provided by UN population statistics [72]. For Italy, we use demographic statistics from Statista online portal about the following: household structure distribution [95], single-person households [96], couples with children [97] and corresponding family size [98], and single parents with children [99]. Furthermore, we assume that children could stay within the family until the age of 30 and that couples without children were aged 30+, to account for societal patterns reported in familial studies which may have affected household distribution metrics [100]. Third, we sample comorbidities from the corresponding country- and age-specific distributions. For China, we use estimates on age-specific prevalence of diabetes [101] and hypertension [102]. For Italy, we use estimates from the Global Burden of Disease study on diabetes [31] and a recent study of age-stratified hypertension prevalence [103]. We ensure that diabetes and hypertension are appropriately correlated using a single global estimate for the probability of hypertension in individuals with diabetes [104].

B Estimating mortality from age and comorbidities

We require a model of $p_{m\,!\,d}(a_i, c_i)$, however existing data sources only specify $p_{m\,!\,d}(a_i)$ and $p_{m\,!\,d}(c_i)$. To infer the joint distribution, we assume a linear (logistic) interaction between age bracket, diabetes status, and hypertension status. Specifically, we assume

$$p_{m\,!\,d}(a_i, c_i) = \sigma\Big(\beta_{\text{age}}(a_i) + \beta_{\text{diabetes}}\,\mathbb{1}\left[\text{diabetes } \mathit{2}\ c_i\right] + \beta_{\text{hypertension}}\,\mathbb{1}\left[\text{hypertension } \mathit{2}\ c_i\right]\Big),$$

where $\beta_{\text{age}}(a_i)$ has a value for each age bracket (e.g., 20-30, 30-40, etc., 7 in total) and β_{diabetes} and $\beta_{\text{hypertension}}$ are scalars.

The marginal distributions $p_{m\,!\,d}(a_i)$ and $p_{m\,!\,d}(c_i)$ are reported by China CDC [41]. We obtained data from the literature on the prevalence of diabetes and hypertension [102] in China by age [101], as well as a single global estimate of $p(\text{hypertension}|\text{diabetes})$ [104]. We assume that these distributions are the same in COVID-19 patients as in the general population. Given this information, we use gradient descent to find a set of parameters β which minimize the mean squared error in the following marginal consistency constraints:

$$p_{m\,!\,d}(a_i) = \sum_{\text{diabetes,hypertension}} p(\text{diabetes, hypertension}|a_i)p_{m\,!\,d}(a_i, \text{diabetes, hypertension}), \quad \forall a_i,$$

$$p_{m\,!\,d}(\text{diabetes}) = \sum_{a_i} p(a_i|\text{diabetes}) \sum_{\text{hypertension}} p(\text{hypertension}|a_i, \text{diabetes})p_{m\,!\,d}(a_i, \text{diabetes, hypertension}),$$

$$p_{m\,!\,d}(\text{hypertension}) = \sum_{a_i} p(a_i|\text{hypertension}) \sum_{\text{diabetes}} p(\text{diabetes}|a_i, \text{hypertension})p_{m\,!\,d}(a_i, \text{diabetes, hypertension})$$

The set of estimated parameters are

$$\begin{aligned}
\beta_{\text{age}}(18 \quad 30) &= \quad 6.70, \\
\beta_{\text{age}}(30 \quad 40) &= \quad 6.90, \\
\beta_{\text{age}}(40 \quad 50) &= \quad 6.62, \\
\beta_{\text{age}}(50 \quad 60) &= \quad 5.74, \\
\beta_{\text{age}}(60 \quad 70) &= \quad 4.90, \\
\beta_{\text{age}}(70 \quad 80) &= \quad 4.10, \\
\beta_{\text{age}}(80 \quad 100) &= \quad 3.37, \\
\beta_{\text{diabetes}} &= 1.35, \\
\beta_{\text{hypertension}} &= 1.69.
\end{aligned}$$

Over 10 random restarts, the marginal values were always fit to within numerical tolerance by the same set of parameters (less than 0.1% maximum difference in the value of a parameter between runs). This suggests that the model parameters are fully identifiable in this setting. We conclude that the impact of comorbidities is large relative to age; having diabetes or hypertension is estimated to be roughly as impactful as moving from the 50-60 age group to the 70-80 age group. As a baseline, the model estimates that if the entire Chinese population were to contract COVID-19, the CFR would be 1.34%.