Visualizing the invisible: The effect of asymptomatic transmission on the outbreak dynamics of COVID-19

Mathias Peirlinck[1], Kevin Linka[1,2], Francisco Sahli Costabal[3,4],
Eran Bendavid[5], Jay Bhattacharya[5], John P.A. Ioannidis[5,6], Ellen Kuhl[1]

[1] *Department of Mechanical Engineering, Stanford University, Stanford, California, United States*
[2] *Institute of Continuum and Materials Mechanics, Hamburg University of Technology, Hamburg, Germany*
[3] *Department of Mechanical and Metallurgical Engineering, Pontificia Universidad Catolica de Chile, Chile*
[4] *Institute for Biological and Medical Engineering, Pontificia Universidad Catolica de Chile, Santiago, Chile*
[5] *Department of Medicine, Stanford University, Stanford, California, United States*
[6] *Department of Epidemiology & Population Health, Stanford University, Stanford, California, United States*

Understanding the outbreak dynamics of the COVID-19 pandemic has important implications for successful containment and mitigation strategies. Recent studies suggest that the population prevalence of SARS-CoV-2 antibodies, a proxy for the number of asymptomatic cases, could be an order of magnitude larger than expected from the number of reported symptomatic cases. Knowing the precise prevalence and contagiousness of asymptomatic transmission is critical to estimate the overall dimension and pandemic potential of COVID-19. However, at this stage, the effect of the asymptomatic population, its size, and its outbreak dynamics remain largely unknown. Here we use reported symptomatic case data in conjunction with antibody seroprevalence studies, a mathematical epidemiology model, and a Bayesian framework to infer the epidemiological characteristics of COVID-19. Our model learns, in real time, the time-varying contact rate of the outbreak, and projects the temporal evolution and credible intervals of the effective reproduction number and the symptomatic, asymptomatic, and recovered popu-

lations. Our study reveals that the outbreak dynamics of COVID-19 are sensitive to three parameters: the effective reproduction number, the ratio between the symptomatic and asymptomatic populations, and the infectious periods of both groups. For three distinct locations, Santa Clara County (CA, USA), New York City (NY, USA), and Heinsberg (NRW, Germany), our model estimates the fraction of the population that has been infected and recovered by May 13, 2020 to 6.2% (95% CI: 3.3%-9.0%), 22.7% (95% CI: 15.7%-29.8%), and 20.5% (95% CI: 17.0%-24.3%). Our method traces the initial outbreak date in Santa Clara County back to January 20, 2020 (95% CI: January 16, 2020 - January 24, 2020). Our results could significantly change our understanding and management of the COVID-19 pandemic: A large asymptomatic population will make isolation, containment, and tracing of individual cases challenging. Instead, if needed, managing community transmission through increasing population awareness, promoting physical distancing, and encouraging behavioral changes could become more relevant.

Introduction

Since its outbreak in December 2019, the COVID-19 pandemic has rapidly swept across the globe and is now affecting 188 countries with more than 5 million cases reported worldwide [8]. In the early stages of a pandemic, doctors, researchers, and political decision makers mainly focus on symptomatic individuals that come for testing and ad-

dress those who require the most urgent medical attention [11]. In the more advanced stages, the interest shifts towards mildly symptomatic and asymptomatic individuals who– by definition–are difficult to trace and likely to retain normal social and travel patterns [27]. In this manuscript, we collectively use the term "asymptomatic" for individuals who have mild symptoms that are not directly associated with COVID-19 or display no symptoms at all. Recent antibody seroprevalence studies suggests that the number of asymptomatic COVID-19 cases outnumbers the symptomatic cases by an order of magnitude or more [3,4,7,10,13,41–43,46,47,49,52]. Estimating the prevalence and contagiousness of these asymptomatic cases is critical since it will change our understanding of the overall dimension and the pandemic potential of COVID-19 [12]. Yet, at this stage, the effect of the asymptomatic population, its size, and its outbreak dynamics remain largely unknown.

The first evidence of asymptomatic individuals in a family cluster of three was reported in late January, where one individual was mildly symptomatic and two remained asymptomatic, with normal lymphocyte counts and chest computer tomography images, but positive quantitative reverse transcription polymerase chain reaction tests [32]. As of today, more than 50 studies have reported an asymptomatic population, twelve of them with a sample size of at least 500 [22], with a median undercount of 20 across all studies, suggesting that only one in twenty COVID-19 cases is noticed and reported. These studies are based on polymerase chain reaction or antibody seroprevalence tests in different subgroups of the population, at different locations, at different points in time [3,5,46]. To no surprise, the reported undercount varies hugely, ranging from 5 to 627 with maximum

values in Oise, France [13] and in Kobe, Japan [7]. Most of these studies are currently only available on preprint servers, but an increasing number is now passing peer review, including a study of 1402 individuals in Wuhan City with an undercount of 22.1 [52], a study of 400 health care workers in London with an undercount of 35.0 [50], a community spreading study of 131 patients with influenza-like symptoms in Los Angeles with an undercount of 100.0 in [45], and a seroprevalence study in Los Angeles county with an undercount of 43.5 [44]. The reported trend across all studies is strikingly consistent: A much larger number of individuals displays antibody prevalence than we would expect from the reported symptomatic case numbers. Knowing the exact dimension of the asymptomatic population is critical for two reasons: first, to truly estimate the severity of the outbreak, e.g., hospitalization or mortality rates [12], and second, to reliably predict the success of surveillance and control strategies, e.g., contact tracing or vaccination [14].

While there is a pressing need to better understand the prevalence of asymptomatic transmission, it is also becoming increasingly clear that it will likely take a long time until we can, with full confidence, deliver reliable measurements of this asymptomatic group. In the meantime, mathematical modeling can provide valuable insight into the tentative outbreak dynamics and outbreak control of COVID-19 for varying asymptomatic scenarios [27]. Many classical epidemiology models base their predictions on compartment models in which individuals pass through different stages as they experience the disease [23]. A popular model to simulate the outbreak dynamics of COVID-19 is the SEIR model [11], which is made up of four compartments for the susceptible, exposed, infectious, and recovered

populations [2]. Here, to explicitly account for the asymptomatic population, we introduce an SEIIR model, which further divides the infectious population into symptomatic and asymptomatic groups. Similar models have recently been used to study the general role of asymptomatic carriers in disease transmission [33] and to illustrate how asymptomatic individuals have facilitated the rapid spread of COVID-19 throughout China [27], South Korea [48], and Italy [15]. While it is tempting–and easily possible–to introduce many more sub-populations into the model, for example a pre-symptomatic, hospitalized, or mortality group [36], here, we focus on the simplest possible model that allows us to explore the role of the asymptomatic population throughout the COVID-19 pandemic. To systematically probe different scenarios, we combine this deterministic SEIIR model with a dynamic effective reproduction number and adopt machine learning and uncertainty quantification techniques to learn the reproduction number, in real time, and quantify uncertainties in the symptomatic-to-asymptomatic ratio, and the initial exposed and infectious populations [29]. We show that this not only allows us to visualize the dynamics and uncertainties of the dynamic contact rate, the effective reproduction number, and the symptomatic, asymptomatic, and recovered populations, but also to estimate the initial date of the outbreak.

Results

Outbreak dynamics of COVID-19 in Santa Clara County. Figure 1 illustrates the outbreak dynamics of COVID-19 in Santa Clara County. The first day, March 2, 2020, is the day on which the number of confirmed cases exceeded 19 cases, 0.001% of the popu-

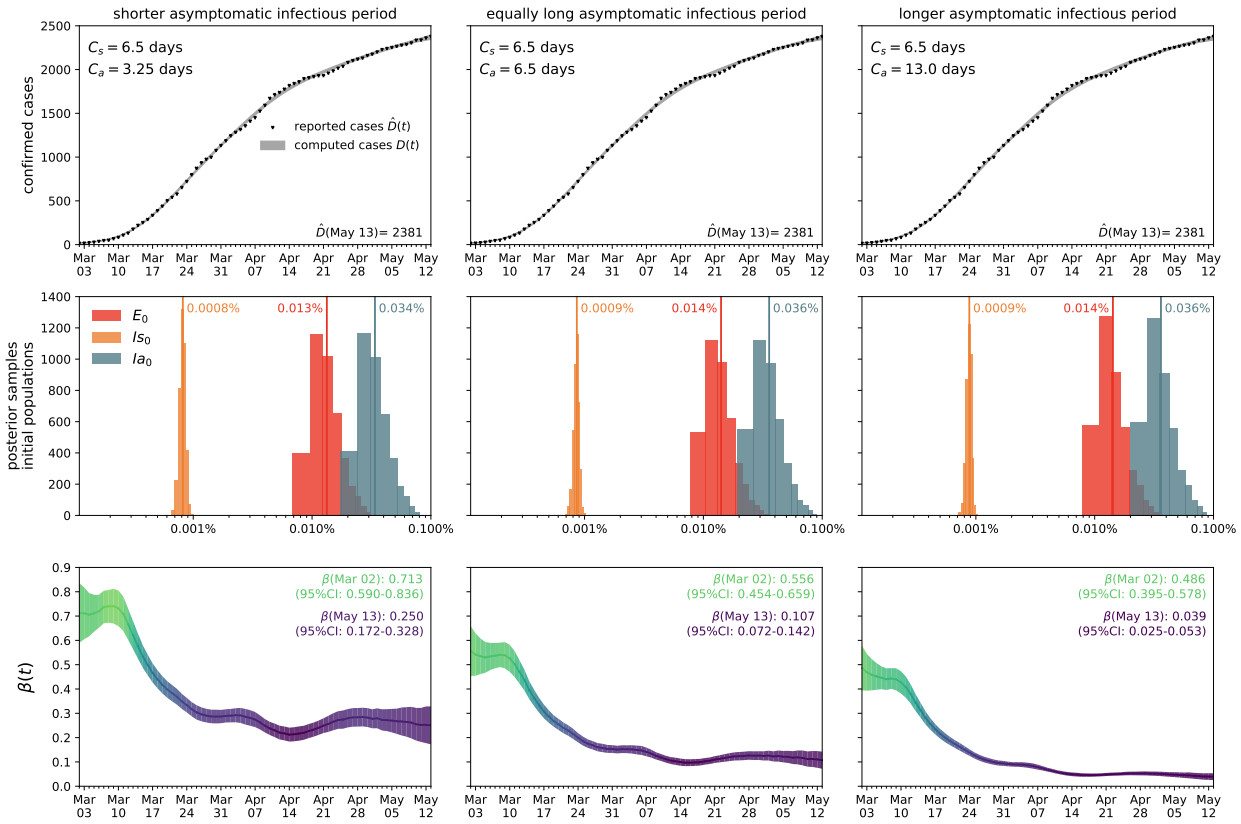

Figure 1: **Outbreak dynamics of COVID-19 in Santa Clara County.** The simulation learns the time-varying contact rate $\beta(t)$ for fixed latent and symptomatic infectious periods $A = 2.5$ days and $C_{\rm s} = 6.5$ days, and for three asymptomatic infectious periods $C_{\rm a} =$ 3.25 days, 6.5 days, and 13.0 days (from left to right). Computed and reported confirmed cases in Santa Clara County, $D(t) = I_{\rm s}(t) + R_{\rm s}(t)$ and $\hat{D}(t)$ (top), initial exposed and infectious populations, E_0, $I_{\rm s0}$, and $I_{\rm a0}$ (middle), and dynamic contact rate, $\beta(t)$ (bottom). The gray and green-blue regions highlight the 95% credible intervals on the confirmed cases $D(t)$ (top) and the contact rate $\beta(t)$ (bottom) based on the reported cases $\hat{D}(t)$, while taking into account uncertainties on the fraction of the symptomatic infectious population $\nu_{\rm s} = I_{\rm s}/I$, and the initial exposed and infectious populations E_0, $I_{\rm s0}$, and $I_{\rm a0}$.

lation. The black dots highlight the reported cases $\hat{D}(t)$ from this day forward. Based on these data points, we learn the posterior distributions of our SEIIR model parameters for fixed latent and symptomatic infectious periods $A = 2.5$ days and $C_{\rm s} = 6.5$ days, and for three asymptomatic infectious periods, $C_{\rm a} = 3.25$, 6.5, and 13.0 days, from left to right. The gray and green-blue regions highlights the 95% credible intervals on the confirmed cases $D(t)$, top row, and the contact rate $\beta(t)$, bottom row, based on the reported cases $\hat{D}(t)$, while taking into account uncertainties on the fraction of the symptomatic infectious population $\nu_{\rm s} = I_{\rm s}/I$, and the initial exposed and infectious populations E_0, $I_{\rm s0}$, and $I_{\rm a0}$. The red, orange, and gray histograms display the learnt initial exposed and infectious populations, E_0, $I_{\rm s0}$, and $I_{\rm a0}$, for the three different asymptomatic infectious periods, middle row. The graphs confirm that our dynamic SEIIR epidemiology model is capable of correctly capturing the gradual flattening of the curve of confirmed cases in agreement with the decrease in new cases reported in Santa Clara County, top row. The consistent downward trend of the contact rate $\beta(t)$ quantifies the efficiency of public health interventions. The different magnitudes in the contact rate highlight the effect of the three different asymptomatic infectious periods $C_{\rm a}$: For larger asymptomatic infectious periods $C_{\rm a}$, from left to right, to explain the same number of confirmed cases $D(t) = I_{\rm s}(t) + R_{\rm s}(t)$, the contact rate $\beta(t)$ has to decrease. On March 2, 2020, when the detected population amounted to 0.001% in Santa Clara County, the mean contact rate $\beta(t)$ was 0.713 (95% CI: 0.590 - 0.836) for an infectious period of $C_{\rm a}$=3.25 days, 0.556 (95% CI: 0.454 - 0.659) for $C_{\rm a}$= 6.5 days, and 0.486 (95% CI: 0.395 - 0.578) for $C_{\rm a}$=13.0 days. By March 17, 2020,

the day Santa Clara County announced the first county-wide shelter-in-place order in the entire United States, these mean contact rates $\beta(t)$ were 0.466 (95% CI: 0.427 - 0.506) for an infectious period of C_{a}=3.25 days, 0.308 (95% CI: 0.280 - 0.336) for C_{a}= 6.5 days, and 0.235 (95% CI: 0.214 - 0.256) for C_{a}=13.0 days.

Effect of asymptomatic transmission of COVID-19 in Santa Clara County. Figure 2 visualizes the effect of asymptomatic transmission in Santa Clara County. The simulation learns the time-varying contact rate $\beta(t)$, and with it the time-varying effective reproduction number $R(t)$, top row, for three asymptomatic infectious periods $C_{\mathrm{a}} = 3.25$ days, 6.5 days, and 13.0 days, from left to right. The effective reproduction number $R(t)$ follows a similar downward trend as the contact rate $\beta(t)$. For larger asymptomatic infectious periods C_{a}, from left to right, since $R(t) = C_{\mathrm{s}}\,\beta(t)/[\,\nu_{\mathrm{s}} + \nu_{\mathrm{a}}\,C_{\mathrm{s}}/C_{\mathrm{a}}\,]$, as C_{a} increases, $C_{\mathrm{s}}/C_{\mathrm{a}}$ decreases, and $R(t)$ increases. Since $R(t)$ represents the number of new infections from a single case, a decrease below $R(t) < 1$ implies that a single infectious individual infects less than one new individual, which indicates that the outbreak decays. The dashed vertical lines indicate the time window of $R(t) = 1$ during which one infectious individual, either symptomatic or asymptomatic, infects on average one other individual. For an asymptomatic infectious period of C_{a}=3.25 days, this critical transition occurred from March 24 until April 6, 2020 and lasted 13 days. For C_{a}=6.5 days, this occurred from March 27 until April 7, 2020 and lasted 11 days. For C_{a}=13.0 days, this occurred from April 4 until April 8, 2020 and lasted 4 days. This confirms our intuition that, the larger the asymptomatic infectious period C_{a}, for example because asymptomatic individuals will not

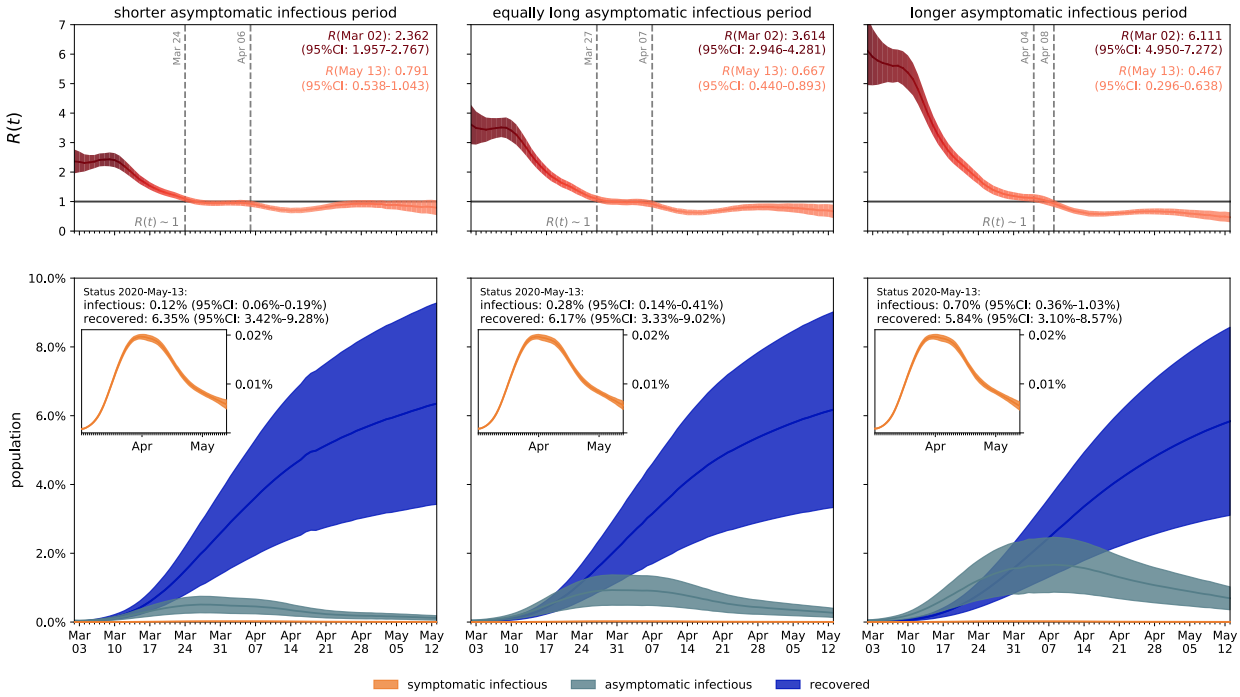

Figure 2: **Effect of asymptomatic transmission of COVID-19 in Santa Clara County.** The simulation learns the time-varying contact rate $\beta(t)$, and with it the time-varying effective reproduction number $R(t)$, for fixed latent and symptomatic infectious periods $A = 2.5$ days and $C_\mathrm{s} = 6.5$ days, and for three asymptomatic infectious periods $C_\mathrm{a} = 3.25$ days, 6.5 days, and 13.0 days (from left to right). The downward trend of the effective reproduction number $R(t)$ reflects the efficiency of public health interventions (top row). The dashed vertical lines mark the critical time period during which the effective reproductive reproduction number fluctuates around $R(t) = 1$. The simulation predicts the symptomatic infectious, asymptomatic infectious, and recovered populations I_s, I_a, and R (bottom row). The colored regions highlight the 95% credible interval for uncertainties in the number of confirmed cases D, the fraction of the symptomatic infectious population $\nu_\mathrm{s} = I_\mathrm{s}/I$, the initial exposed population E_0 and the initial infectious populations $I_{\mathrm{s}0}$ and $I_{\mathrm{a}0}$.

isolate as strictly as symptomatic individuals, the higher the effective reproduction number $R(t)$, and the more difficult it will be to control $R(t)$ by public health interventions. For each of the three cases, the symptomatic infectious, asymptomatic infectious, and recovered population, are shown in the bottom row. For larger asymptomatic infectious periods C_{a}, from left to right, the total infectious population I increases and its maximum occurs later in time. Specifically, the maximum infectious population since March 2, 2020 amounts to 0.53% (95% CI: 0.29%-0.77%) on March 27, 2020 for $C_{\mathrm{a}} = 3.25$ days, 0.95% (95% CI: 0.51%-1.39%) on March 30, 2020 for $C_{\mathrm{a}} = 6.5$ days, and 1.69% (95% CI: 0.89%-2.48%) on April 7, 2020 for $C_{\mathrm{a}} = 13.0$ days. For larger asymptomatic infectious periods C_{a}, from left to right, the recovered population R decreases. Specifically, on May 13, 2020, the recovered population R amounts to 6.35% (95% CI: 3.42%-9.28%) for an infectious period of C_{a} =3.25 days, 6.17% (95% CI: 3.33%-9.02%) for C_{a} =6.5 days, and 5.84% (95% CI: 3.10%-8.57%) for C_{a} =13.0 days. Similarly, and important when considering different exit strategies, the total infectious population, $I = I_{\mathrm{s}} + I_{\mathrm{a}}$, on May 13, 2020 is estimated to 0.12% (95% CI: 0.06%-0.19%) for C_{a} =3.25 days, 0.28% (95% CI: 0.14%-0.41%) for C_{a} =6.5 days, and 0.70% (95% CI: 0.36%-1.03%) for C_{a} =13.0 days.

Outbreak dynamics of COVID-19 in Santa Clara County, New York City, and Heinsberg. Figures 3 and 4 illustrate the outbreak dynamics of COVID-19 in three different locations that reported COVID-19 antibody prevalence in a representative sample of the population: Santa Clara County (CA, USA), New York City (NY, USA) and Heinsberg (NRW, Germany). To compare the different outbreak dynamics, we first learn the asymp-

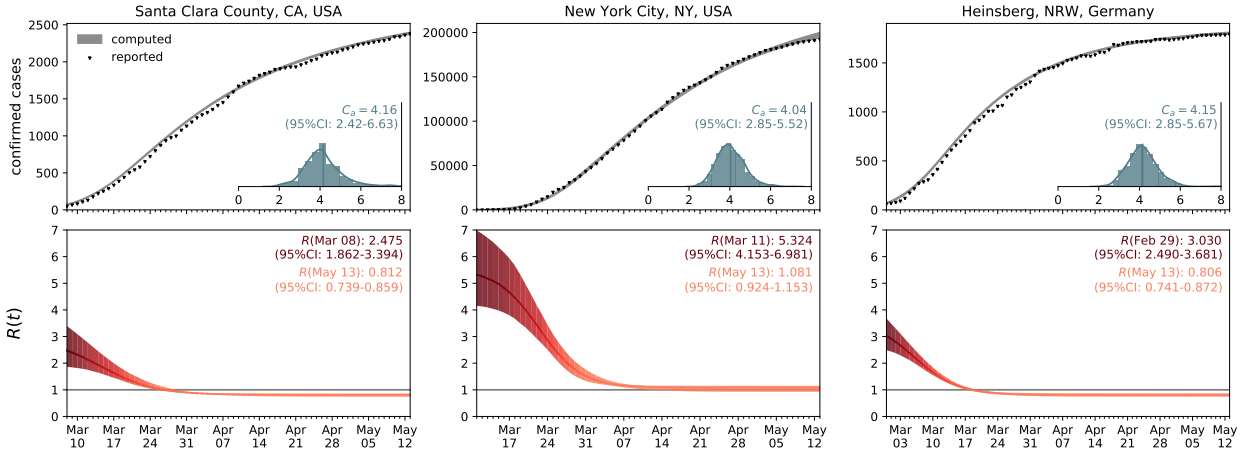

Figure 3: **Outbreak dynamics of COVID-19 in Santa Clara County, New York City, and Heinsberg.** Dynamic effective reproduction number $R(t)$ and reported and simulated detected cases $\hat{D}(t)$ and $D(t)$ at three different locations where antibody prevalence studies were performed. The simulation learns the asymptomatic infectious period C_{a}, the time-varying contact rate $\beta(t)$, and with it the time-varying effective reproduction number $R(t)$, for fixed latent and symptomatic infectious periods $A = 2.5$ days and $C_{\mathrm{s}} = 6.5$ days. The colored regions highlight the 95% credible interval for the effective reproductive number $R(t)$ (top) and for the detected cases $D(t)$ compared to the reported cases $\hat{D}(t)$ (bottom). This asymptomatic infectious periods are C_{a} = 4.16 (95% CI: 2.42-6.63) days for Santa Clara County, C_{a} = 4.04 (95% CI: 2.85-5.52) days for New York City, and C_{a} = 4.15 (95% CI: 2.85-5.67) days for Heinsberg.

tomatic infectious period C_{a} for a simplified contact rate β, and then learn the dynamic contact rate $\beta(t)$ for fixed C_{a}. Figure 3 illustrates the learnt asymptomatic infectious periods C_{a}, the dynamic effective reproduction number $R(t)$, and the reported and simulated cases $\hat{D}(t)$ and $D(t)$ in all three locations. Here, to keep the parameter space manageable, we approximate the contact rate, $\beta(t) = \beta_0 - \frac{1}{2}[\,1 + \tanh\left([\,t - t^*\,]/T\right)][\,\beta_0 - \beta_{\mathrm{t}}\,]$, by a hyperbolic tangent function in terms of the initial and current contact rates β_0 and β_t, the adaptation time t^*, and the transition time T [29]. To account for variability between the three locations while simultaneously taking advantage of the entire data, we adopt a hierarchical model to learn the asymptomatic infectious period C_{a} [17]. For each location, we draw C_{a} from normal distributions using weakly informative priors as $\mu_{C_{\mathrm{a}}} \sim \mathcal{N}_i(6.5, 2)$ and $\sigma_{C_a} \sim$ half-normal(2). This results in asymptomatic infectious periods of C_{a} = 4.16 (95% CI: 2.42-6.63) days for Santa Clara County, C_{a} = 4.04 (95% CI: 2.85-5.52) days for New York City, and C_{a} = 4.15 (95% CI: 2.85-5.67) days for Heinsberg.

Figure 4 illustrates the learnt effective reproduction number $R(t)$, and the symptomatic and asymptomatic infectious populations I_{s} and I_{a}, and the recovered population R for all three locations. Here, we assume fixed latent and infectious periods of $A = 2.5$ days, $C_{\mathrm{s}} = 6.5$ days, and $C_{\mathrm{a}} = 6.5$ days. For all three locations, the calculated metrics display similar trends, although their absolute numbers and percentage values are different. The downward trend of the dynamic effective reproductive number evolution $R(t)$ quantifies how fast each location managed to control the spreading of COVID-19. The dashed vertical lines indicate the critical time window during which the effective reproduction num-

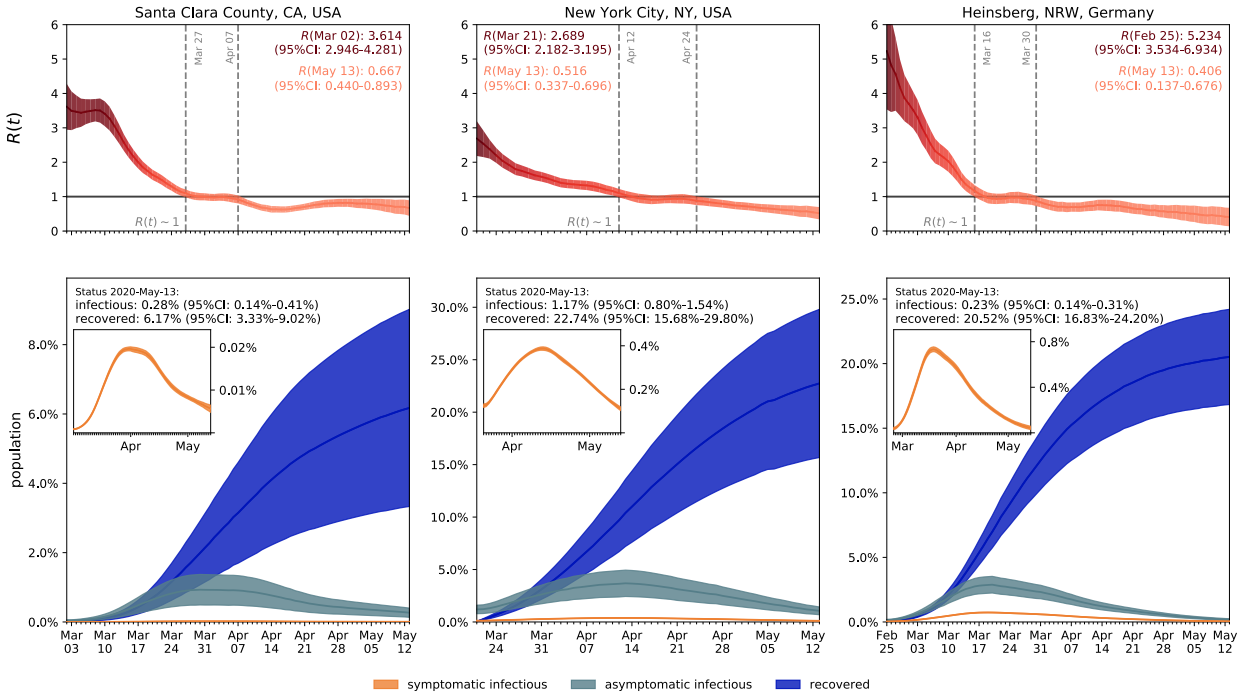

Figure 4: **Outbreak dynamics of COVID-19 in Santa Clara County, New York City, and Heinsberg.** Dynamic effective reproduction number R(t) and symptomatic, asymptomatic, and recovered populations at three different locations where antibody prevalence studies were performed. The simulation learns the time-varying contact rate $\beta(t)$, and with it the time-varying effective reproduction number $R(t)$, to predict the symptomatic infectious, asymptomatic infectious, and recovered populations $I_{\rm s}$, $I_{\rm a}$, and R, for fixed latent and infectious periods A = 2.5 days, $C_{\rm s}$ = 6.5 days, and $C_{\rm a}$ = 6.5 days. The dashed vertical lines mark the critical time period during which the effective reproduction number fluctuates around $R(t) = 1$ (top). The colored regions highlight the 95% credible interval for the effective reproductive number $R(t)$ (top), the symptomatic and asymptomatic populations $I_{\rm s}$ and $I_{\rm a}$, and the recovered population R (bottom), for uncertainties in the number of confirmed cases D, the fraction of the symptomatic infectious population $\nu_{\rm s} = I_{\rm s}/I$, the initial exposed population E_0, and the initial infectious populations $I_{\rm s0}$ and $I_{\rm a0}$.

ber fluctuates around $R(t) = 1$. For Santa Clara County, this critical transition occurred from March 27 until April 7, 2020 and lasted 11 days. For New York City, this occurred from April 12 until April 24, 2020 and lasted 13 days. For Heinsberg, this occurred from March 16 until March 30, 2020 and lasted 15 days. Based on our simulations, the maximum infectious population size amounted to 0.95% (95% CI: 0.51%-1.39%) on March 30, 2020 in Santa Clara County, to 4.07% (95% CI: 2.81%-5.33%) on April 13, 2020 in New York City, and to 3.62% (95% CI: 2.97%-4.27%) on March 20, 2020 in Heinsberg. On May 13, 2020, the estimated recovered population in Santa Clara County, New York City, and Heinsberg reached 6.17% (95% CI: 3.33%-9.02%), 22.74% (95% CI: 15.68%-29.80%), 20.52% (95% CI: 16.83%-24.20%). When using an asymptomatic infectious period of $C_{\mathrm{a}} = 4.1$ days, the best fit value in Figure 3, the maximum infectious population would be slightly smaller and the recovered population would be slightly larger than estimated here for $C_{\mathrm{a}} = 6.5$ days.

Forecasting the COVID-19 dynamics in Santa Clara County. Figure 5 shows a forecast of the COVID-19 dynamics in Santa Clara County for an increase of the contact rate $\beta(t)$ by 10% after May 13, 2020. Based on the inferred posterior distribution of the localized SEIIR model parameters, Figure 5 predicts the effective reproduction number $R(t)$, the symptomatic and asymptomatic infectious populations I_{s} and I_{a}, the recovered population R, and the fraction of the population that will be hospitalized and will require intensive care beds. In addition to the different sizes of infectious and recovered populations throughout the past, up to the dashed line, the three columns show how different asymp-

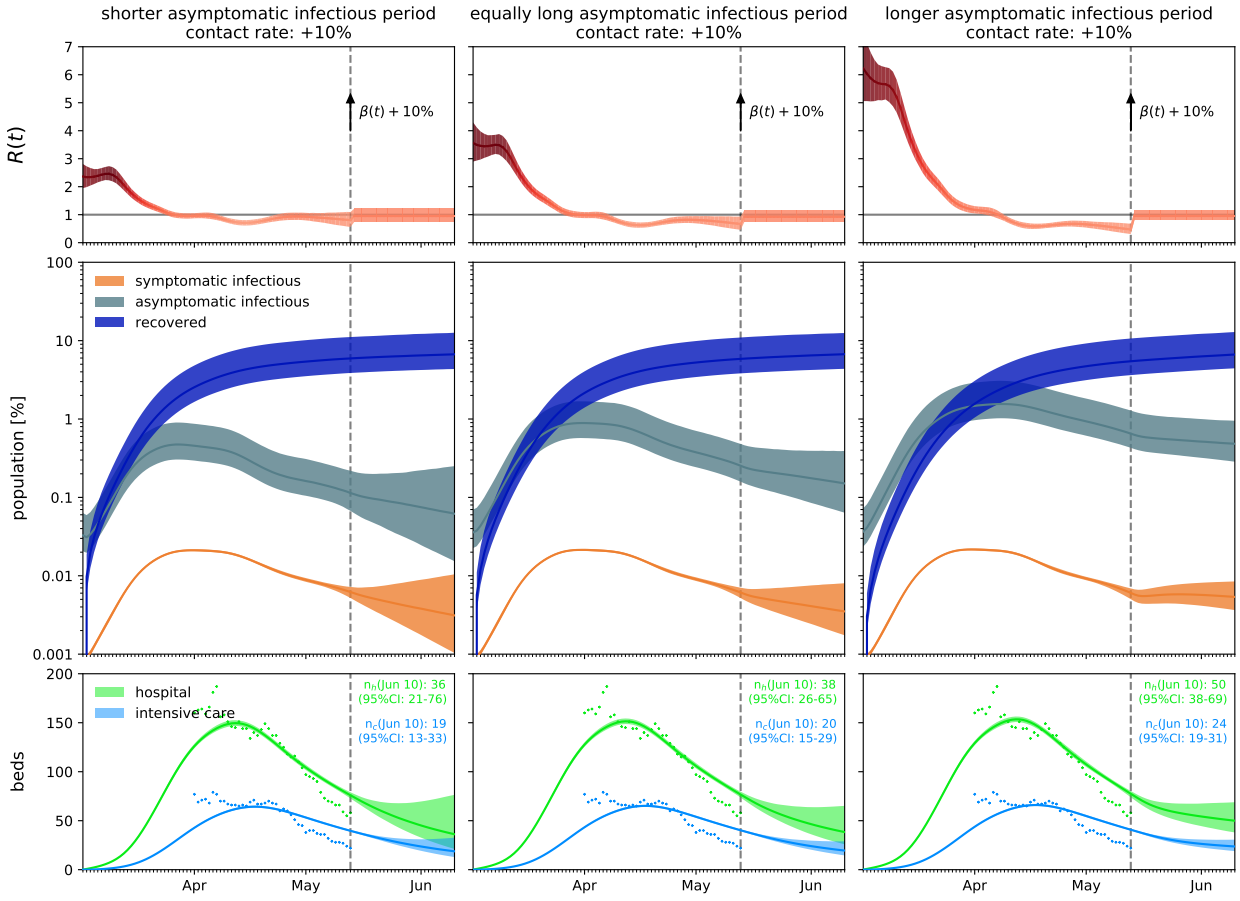

Figure 5: **Forecasting the COVID-19 dynamics in Santa Clara County.** Prediction of the effective reproduction number $R(t)$, the symptomatic and asymptomatic infectious populations I_{s} and I_{a}, the recovered population R, the number of hospitalizations, and the required intensive care beds, from top to bottom, for fixed latent and symptomatic infectious periods $A = 2.5$ days and $C_{\mathrm{s}} = 6.5$ days, and for three asymptomatic infectious periods $C_{\mathrm{a}} = 3.25$ days, 6.5 days, and 13.0 days (from left to right). As a forecasting scenario, we assume a 10% contact increase after May 13, 2020.

tomatic infectious periods $C_{\rm a}$ affect the projected evolution of the infectious populations and the need for hospital and intensive care beds. We identify the hospitalization fraction based on data until May 13, 2020, to $\nu_{\rm h} = 35.24\%$ and the intensive care unit fractions to $\nu_{\rm c} = 32.38\%$. Our projections show with 95% confidence that a 10% increase of the contact rate $\beta(t)$, compared to today's contact rate, would lead to an effective reproduction number $R(t)$ slightly below one. Values of $R(t) > 1$ would re-initiate the exponential growth dynamics of the COVID-19 outbreak. Four weeks into the future, for asymptomatic infectious periods of $C_{\rm a} = 3.25$, 6.5, and 13.0 days, a 10% increase in the contact rate would lead to the need of 36 (95%CI: 21-76), 38 (95%CI: 26-65), and 50 (95%CI: 38-69) hospital beds, of which 19 (95%CI: 13-32), 20 (95%CI: 15-29), and 24 (95%CI: 19-31) would be in the intensive care unit, compared to 888 acute hospital beds, 186 intensive care unit beds, and 1231 surge beds currently available in Santa Clara County [40].

Estimating the outbreak date. Figure 6 shows the estimated outbreak date of COVID-19 in Santa Clara County. For fixed latent and symptomatic infectious periods $A = 2.5$ days and $C_{\rm s} = 6.5$ days, and for three asymptomatic infectious periods $C_{\rm a} = 3.25$ days, 6.5 days, and 13 days, the graphs highlight the estimated date of the first COVID-19 case in the county. Based on the reported case data from March 2, 2020 onward, and taking into account uncertainty on the fraction of the symptomatic infectious population $\nu_{\rm s}$, on the initial exposed population E_0, and on the initial symptomatic and asymptomatic infectious populations $I_{\rm s0}$ and $I_{\rm a0}$, we systematically backtracked the date of the first undetected infectious individual. The three graphs show that, the longer the asymptomatic infectious

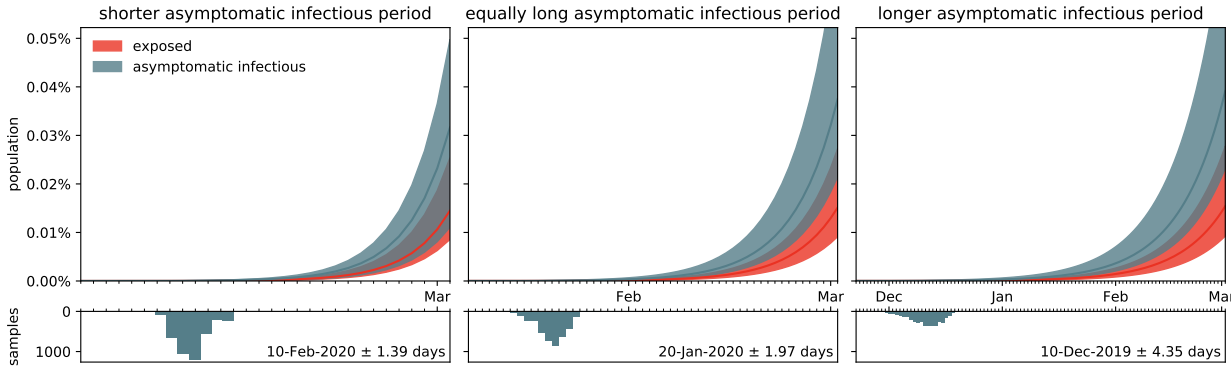

Figure 6: **Estimating the outbreak date of COVID-19 in Santa Clara County varying asymptomatic infectious periods C_{a}.** Estimated date of the first COVID-19 case in Santa Clara County for fixed latent and symptomatic infectious periods $A = 2.5$ days and $C_{\mathrm{s}} = 6.5$ days, and for three asymptomatic infectious periods $C_{\mathrm{a}} = 3.25$ days, 6.5 days, and 13 days (from left to right). The colored regions highlight the 95% credible interval for the exposed and asymptomatic infectious populations E_0 and I_{a} estimated based on the reported cases $\hat{D}(t)$ from March 2, 2020 onward and taking into account uncertainties on the fraction of the symptomatic infectious population $\nu_{\mathrm{s}} = I_{\mathrm{s}}/I$, and the exposed and asymptomatic infectious populations E_0 and $I_{\mathrm{a}0}$ on March 2, 2020 (top row). The histograms show the distribution of the most probable origin dates to February 10, 2020 (95% CI: February 9, 2020 - February 14, 2020) for an asymptomatic infectious period $C_{\mathrm{a}} = 3.25$ days, to January 20, 2020 (95% CI: January 16, 2020 - January 24, 2020) for $C_{\mathrm{a}} = 6.5$ days and to December 10, 2020 (95% CI: December 1, 2020 - December 18, 2020) for $C_{\mathrm{a}} = 13$ days (bottom row).

period C_{a}, the further the predicted first undetected case would date back in time. Our results suggest that the first case of COVID-19 in Santa Clara County dates back to February 10, 2020 (95% CI: February 9, 2020 - February 14, 2020) for an asymptomatic infectious periods $C_{\mathrm{a}} = 3.25$ days, to January 20, 2020 (95% CI: January 16, 2020 - January 24, 2020) for $C_{\mathrm{a}} = 6.5$ days and to December 10, 2020 (95% CI: December 1, 2020 - December 18, 2020) for $C_{\mathrm{a}} = 13$ days. For an asymptomatic infectious period of $C_{\mathrm{a}} = 4.1$ days, the best fit value in Figure 3, the first case would date back to late January.

Discussion

A key question in understanding the outbreak dynamics of COVID-19 is the dimension of the asymptomatic population and its role in disease transmission. Throughout the past three months, dozens of studies have been initiated to quantify the fraction of the general population that displays antibody prevalence but did not report symptoms of COVID-19. Here we assume that this subgroup of the population has been infected with the novel coronavirus, but has remained asymptomatic, or only displayed mild symptoms that were not directly reported in the context of COVID-19. We collectively map this subgroup into an asymptomatic population and additively decompose the total infectious population, $I = I_{\mathrm{s}} + I_{\mathrm{a}}$, into a symptomatic group I_{s} and an asymptomatic group I_{a}. We parameterize this decomposition in terms of a single scalar valued parameter, the symptomatic fraction ν_{s}, such that $I_{\mathrm{s}} = \nu_{\mathrm{s}}\, I$ and $I_{\mathrm{a}} = \nu_{\mathrm{a}}\, I = [\, 1 - \nu_{\mathrm{s}}\,]\, I$. Within this paradigm, we can conceptually distinguish two scenarios: the special case for which both subgroups display identical

contact rates β, latent periods A, and infectious periods C, and the general case for which these transition dynamics are different.

For comparable dynamics, the size of the asymptomatic population does not affect overall outbreak dynamics. For the special case in which both subgroups display identical contact rates β, latent rates α, and infectious rates γ [51], our study shows that the overall outbreak dynamics can be represented by the classical SEIR model [20] using equations (7). Importantly, however, since the reported case data only reflect the symptomatic infectious and recovered groups I_{s} and R_{s}, the true infectious and recovered populations $I = I_{\mathrm{s}}/\nu_{\mathrm{s}}$ and $R = R_{\mathrm{s}}/\nu_{\mathrm{s}}$ could be about an order of magnitude larger than the SEIR model predictions. From an individual's perspective, a smaller symptomatic group ν_{s}, or equivalently, a larger asymptomatic group $\nu_{\mathrm{a}} = [1 - \nu_{\mathrm{s}}]$, could have a personal effect on the likelihood of being unknowingly exposed to the virus, especially for high-risk populations: A larger asymptomatic fraction ν_{a} would translate into an increased risk of community transmisson and would complicate outbreak control [12]. From a health care perspective, however, the special case with comparable transition dynamics would not pose a threat to the health care system since the overall outbreak dynamics would remain unchanged, independent of the fraction ν_{a} of the asymptomatic population: A larger asymptomatic fraction would simply imply that a larger fraction of the population has already been exposed to the virus–without experiencing significant symptoms–and that the true hospitalization and mortality rates would be much lower than the reported rates [22].

For different dynamics, the overall outbreak dynamics depend on both size and in-

fectiousness of the asymptomatic group. For the general case in which the transition rates for the symptomatic and asymptomatic groups are different, the overall outbreak dynamics of COVID-19 become more unpredictable, since little is known about the dynamics of the asymptomatic population [33]. To study the effects of different dynamics between the symptomatic and asymptomatic groups, we decided to collectively represent a lower infectivity of the asymptomatic population through a smaller infectious period $C_{\mathrm{a}} < C_{\mathrm{s}}$ and a lack of early isolation of the asymptomatic population through a larger infectious period $C_{\mathrm{s}} < C_{\mathrm{a}}$, while, for simplicity, keeping the latent period A and contact rate β similar across both groups [26]. Our study shows that the overall reproduction number, $R(t) = [C_{\mathrm{a}}\, C_{\mathrm{s}}]/[\nu_{\mathrm{s}}\, C_{\mathrm{a}} + \nu_{\mathrm{a}}\, C_{\mathrm{s}}]\ \beta(t)$, and with it the outbreak dynamics, depend critically on the fractions of the symptomatic and asymptomatic populations ν_{s} and ν_{a} and on the ratio of the two infectious periods C_{s} and C_{a}. To illustrate these effects, throughout this study, we consistently report the results for three different scenarios where the asymptomatic group is half as infectious, $C_{\mathrm{a}} = 0.5\, C_{\mathrm{s}}$, equally infectious, $C_{\mathrm{a}} = C_{\mathrm{s}}$, and twice as infectious $C_{\mathrm{a}} = 2.0\, C_{\mathrm{s}}$ as the symptomatic group. The second case, the middle column in Figures 1, 2, 5, and 6, corresponds to the special case with comparable dynamics and similar parameters. Our learnt asymptomatic infectious periods of $C_{\mathrm{a}} = 4.1$ days in Figure 3 suggest that C_{a} is consistently smaller than the symptomatic infectious period of $C_{\mathrm{s}} = 6.5$ days and that the asymptomatic population is about two third as infectious as the symptomatic population.

Dynamic contact rates are a metric for the efficiency of public health interven-

tions. Classical SEIR epidemiology models with static parameters are well suited to model outbreak dynamics under unconstrained conditions and predict how the susceptible, exposed, infectious, and recovered populations converge freely toward the endemic equilibrium [20]. However, they cannot capture changes in disease dynamics and fail to converge towards a temporary equilibrium before the entire population has become sufficiently immune to prevent further spreading [36]. To address this limitation, we introduce a time-dependent contact rate $\beta(t)$, which we learn dynamically from the reported case data. Figure 1 demonstrates that our approach can successfully identify a dynamic contact rate that not only decreases monotonically, but is also capable of reproducing local contact fluctuations. With this dynamic contact rate, our model can capture the characteristic S-shaped COVID-19 case curve that plateaus before a large fraction of the population has been affected by the disease, resembling a Gompertz function. Previous studies have inferred discrete date points at which the contact rates vary [6] or used sliding windows over the amount of novel reported infections [34] to motivate dynamic contact rates. As such, our framework provides a model-based method for statistical inference of virus transmissibility: It naturally learns the most probable contact rate from the changing time evolution of new confirmed cases and concomitantly quantifies the uncertainty on that estimation.

The dynamics of the asymptomatic population affect the effective reproduction number. Our analysis in equations (5) and (6) and our simulations in Figure 2 illustrate how asymptomatic transmission affects the effective reproduction number, and with it the outbreak dynamics of COVID-19. Our results show that, the larger the infec-

tious period C_{a} of the asymptomatic group, the larger the effective reproduction number, $R(t) = C_{\mathrm{s}}\,\beta(t)/[\,\nu_{\mathrm{s}} + \nu_{\mathrm{a}}\,C_{\mathrm{s}}/C_{\mathrm{a}}\,]$, and the later the drop of $R(t)$ below the critical value of one. A recent study analyzed the dynamics of the asymptomatic population in three consecutive windows of two weeks during the early outbreak in China [27]. The study found relatively constant latency and infectious periods A and C, similar to our assumption, and a decrease in the contact rate $\beta = 1.12, 0.52, 0.35$ days^{-1} and in the effective reproduction number $R(t) = 2.38, 1.34, 0.98$, which is consistent with our results. However, rather than assuming constant outbreak parameters within pre-defined time windows, our study learns the effective reproduction number dynamically, in real time, from the available data. Figures 2 and 4 demonstrate that we can successfully learn the critical time window of $R(t) = 1$, which, in Santa Clara County, starts as early as March 24 for C_{a} =3.25 days, on March 27 for $C_{\mathrm{a}} =$ 6.5 days, and on April 4 for $C_{\mathrm{a}} =$ 13.0 days. Our findings are consistent with the observation that the basic reproduction number will be over-estimated if the asymptomatic group has a shorter generation interval, and underestimated if it has a longer generation interval than the symptomatic group [33]. Naturally, these differences are less pronounced under current conditions where the effective reproduction number is low and the entire population has been sheltering in place for ten weeks. It will be interesting to see if the effects of asymptomatic transmission become more visible as we gradually relax the current constraints and allow all individuals to move around and interact with others more freely. Seasonality, effects of different temperature and humidity, and other unknown factors may also influence the extent of transmission.

Estimates of the infectious asymptomatic population may vary, but general trends are similar. Throughout the past months, an increasing number of researchers around the globe have started to characterize the size of the asymptomatic population to better understand the outbreak dynamics of COVID-19 [22]. Two major challenges drive the interest in these studies: estimating the severity of the outbreak, e.g., hospitalization and mortality rates [12], and predicting the success of surveillance and control efforts, e.g., contact tracing or vaccination [14]. This is especially challenging now–in almost complete lockdown–when the differences in transmission dynamics between the symptomatic and asymptomatic populations are small and difficult to quantify. However, as Figure 2 suggests, these transmission dynamics can have a significant effect on the size of the asymptomatic population: For infectious periods of $C_{\mathrm{a}} = 0.5$, 1.0, and 2.0 C_{s}, the maximum infectious population varies from 0.53% to 0.95% and 1.69%. Interestingly, not only the sum of the infectious and recovered populations, but also the uncertainty of their prediction, remain relatively insensitive to variations in the infectious period. To explore whether this is a universal trend, we perform the same analysis for three different locations, Santa Clara County [3], New York City [5], and Heinsberg [46]. The fraction of the symptomatic population in these three locations is ν_{s} = 2.44%, 10.15%, and 20.76% and falls right within the range of reported symptomatic versus estimated total cases [3,4,7,10,13,41–43,46,47,49,52]. Of the three locations we analyzed here, Santa Clara County tested IgG and IgM, New York City tested IgG, and Heinsberg tested IgG and IgA. While we did include reported uncertainty on the seroprevalence data, seroprevalence would likely have been higher if

all three locations had tested for all three antibodies. Despite these differences, the effective reproduction numbers $R(t)$ and the infectious and recovered populations I_{s}, I_{a}, and R in Figure 4 display remarkably similar trends: In all three locations, the effective reproduction number $R(t)$ drops rapidly to values below one within a window of about three weeks. However, the maximum infectious population, a value that is closely monitored by hospitals and health care systems, varies significantly between 0.95% in Santa Clara County, 4.07% in New York City, and 3.62% in Heinsberg. An effect that we do not explicitly address is that immune response not only results COVID-19 antibodies (humoral response), but also from innate and cellular immunity [18]. While it is difficult to measure the effects of the unreported asymptomatic group directly, and discriminate it precisely from innate and cellular immunity, mathematical models can provide valuable insight into how this population modulates the outbreak dynamics and the potential of successful outbreak control [27].

Simulations provide a window into the outbreak date. Santa Clara County was home to the first individual who died with COVID-19 in the United States. Although this happened as early as February 6, the case remained unnoticed until April 22 [1]. The unexpected new finding suggests that the new coronavirus was circulating in the Bay Area as early as January. The estimated uncertainty on the exposed, symptomatic infectious, and asymptomatic infectious populations of our model allows us to estimate the initial outbreak date. Figure 6 indicates that this initial outbreak data is sensitive to the asymptomatic transmission and moves towards later points in time for increasing asymptomatic

infectious periods: The outbreak dates back to February 10 for $C_{\mathrm{a}} = 0.5\, C_{\mathrm{s}}$, around January 20 for $C_{\mathrm{a}} = 1.0\, C_{\mathrm{s}}$, and around December 10 for $C_{\mathrm{a}} = 2.0\, C_{\mathrm{s}}$. These back-calculated dates not only confirm the undetected community spreading of COVID-19 before the first death of an individual with no history of travel on February 6, but also suggest that an asymptomatic contact period close to $C_{\mathrm{a}} = C_{\mathrm{s}}$ would be more realistic to describe these early events. These back-calculated early outbreak dates are in line with our intuition that COVID-19 is often present in a population long before the first official case is reported. Interestingly, our analysis comes to this conclusion purely based on a local serology antibody study [3] and the number of reported cases [40].

Limitations. Our approach naturally builds in and learns several levels of uncertainties. By design, this allows us to estimate sensitivities and credible intervals for a number of important model parameters and discover important features and trends. Nevertheless, it has a few limitations, some of them by design, some simply limited by the current availability of data: First, our current SEIIR model assumes a similar contact rate $\beta(t)$ for symptomatic and asymptomatic individuals. While we can easily adjust this in the model by defining individual symptomatic and asymptomatic rates $\beta_{\mathrm{s}}(t)$ and $\beta_{\mathrm{a}}(t)$, we currently do not have data about the temporal evolution of the hidden asymptomatic infectious population $I_{\mathrm{a}}(t)$ and longitudinal large population antibody studies would be needed to appropriately calibrate $\beta_{\mathrm{a}}(t)$. Second, the ratio between the symptomatic and asymptomatic populations $\nu_{\mathrm{s}} : \nu_{\mathrm{a}}$ can vary over time, especially, as we have shown, if both groups display notably different dynamics, in our model represented through C_{s} and C_{a}. Since this

can have serious effects on the overall reproduction number $R(t)$, and with it on required outbreak control strategies, it seems critical to perform more tests and learn the dynamics of the fractions $\nu_{\mathrm{s}}(t)$ and $\nu_{\mathrm{a}}(t)$ of both groups. Third, and this is not only true for our specific model, but for COVID-19 forecasts in general, all predictions can be sensitive to the amount of testing in time. As such, they crucially rely on testing policies and testing capacities. We expect to see a significant increase in the symptomatic-to-asymptomatic, or rather detected-to-undetected, ratio as we move towards systematically testing larger fractions of the population and more and more people who have no symptoms at all. The intensity of testing increased significantly in all locations during our simulation period. For example, in Santa Clara County, testing was extremely limited until early April, increased substantially in the first three weeks of April, and even more after. Including limited testing and more undocumented cases during the early outbreak would shift the case distribution towards earlier days, result in lower C_{a} values, and predict an even earlier outbreak date. Fourth, while we have included uncertainty in the seroprevalence data, the three locations we analyzed here tested different types of antibodies and had different sampling procedures. Seroprevalence could have been higher if all three locations had tested for the same three antibodies and data may differ depending on biases introduced by the sampling procedure. Finally, our current model does not explicitly account for innate and cellular immunity. If the fraction of the population with innate and cellular immunity is substantially high, we would anticipate a smaller susceptible population and a larger and earlier protective immunity overall. These, and other limitations related to the availability

of information, can be easily addressed and embedded in our model and will naturally receive more clarification as studies and data become available in the coming months.

Conclusions.

The rapid and devastating development of the COVID-19 pandemic has raised many open questions about its outbreak dynamics and unsuccessful outbreak control. From an outbreak management standpoint–in the absence of effective vaccination and treatment–the two most successful strategies in controlling an infectious disease are isolating symptomatic infectious individuals and tracing and quarantining their contacts. Both critically rely on a rapid identification of infections, typically through clinical symptoms. Recent antibody prevalence studies could explain why these strategies have largely failed in containing the COVID-19 pandemic: Increasing evidence suggests that the number of unreported asymptomatic cases could outnumber the reported symptomatic cases by an order of magnitude or more. Mathematical modeling, in conjunction with reported symptomatic case data, antibody seroprevalence studies, and machine learning allows us to infer, in real time, the epidemiology characteristics of COVID-19. We can now visualize the invisible asymptomatic population, estimate its role in disease transmission, and quantify the confidence in these predictions. A better understanding of asymptomatic transmission will help us evaluate strategies to manage the impact of COVID-19 on both our economy and our health care system.

Methods

Epidemiology modeling. We model the epidemiology of COVID-19 using an SEIIR model with five compartments, the susceptible, exposed, symptomatic infectious, asymptomatic infectious, and recovered populations. Figure 7 illustrates our SEIIR model, which is governed by a set of five ordinary differential equations,

$$\begin{aligned}
\dot{S} &= -S\,[\,\beta_{\rm s}\,I_{\rm s}+\beta_{\rm a}\,I_{\rm a}\,] \\
\dot{E} &= +S\,[\,\beta_{\rm s}\,I_{\rm s}+\beta_{\rm a}\,I_{\rm a}\,] - \alpha\,E \\
\dot{I}_{\rm s} &= +\nu_{\rm s}\,\alpha\,E - \gamma_{\rm s}\,I_{\rm s} \\
\dot{I}_{\rm a} &= +\nu_{\rm a}\,\alpha\,E - \gamma_{\rm a}\,I_{\rm a} \\
\dot{R} &= +\gamma_{\rm s}\,I_{\rm s}+\gamma_{\rm a}\,I_{\rm a}\,,
\end{aligned} \tag{1}$$

where the fractions of all five populations add up to one, $S+E+I_{\rm s}+I_{\rm a}+R=1$. We assume that both the symptomatic group $I_{\rm s}$ and the asymptomatic group $I_{\rm a}$ can generate new infections. We introduces these two groups as fractions $\nu_{\rm s}$ and $\nu_{\rm a}$ of the total infectious group I,

$$I=I_{\rm s}+I_{\rm a} \quad \text{with} \quad I_{\rm s}=\nu_{\rm s}\,I \quad \text{and} \quad I_{\rm a}=\nu_{\rm a}\,I \quad \text{where} \quad 0\le\nu_{\rm s},\nu_{\rm a}\le 1 \quad \text{and} \quad \nu_{\rm s}+\nu_{\rm a}=1\,. \tag{2}$$

We postulate that the two infectious groups $I_{\rm s}$ and $I_{\rm a}$ have the same latent period $A=1/\alpha$, but can have individual contact periods $B_{\rm s}=1/\beta_{\rm s}$ and $B_{\rm a}=1/\beta_{\rm a}$ to mimic their different community spreading, and individual infectious periods $C_{\rm s}=1/\gamma_{\rm s}$ and $C_{\rm a}=1/\gamma_{\rm a}$ to mimic their different likelihood of isolation. From the infectious fractions (2), we can derive the overall contact and infectious rates β and γ from their individual symptomatic

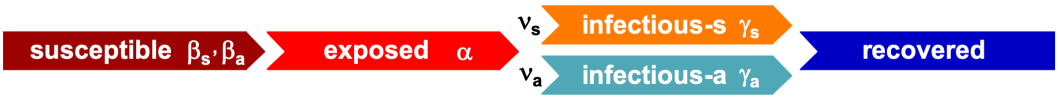

Figure 7: **SEIIR epidemiology model.** The SEIIR model contains five compartments for the susceptible, exposed, symptomatic infectious, asymptomatic infectious, and recovered populations. The transition rates between the compartments, β, α, and γ are inverses of the contact period $B = 1/\beta$, the latent period $A = 1/\alpha$, and the infectious period $C = 1/\gamma$. The symptomatic and asymptomatic groups have the same latent period A, but they can have individual contact periods $B_\mathrm{s} = 1/\beta_\mathrm{s}$ and $B_\mathrm{a} = 1/\beta_\mathrm{a}$ and individual infectious periods $C_\mathrm{s} = 1/\gamma_\mathrm{s}$ and $C_\mathrm{a} = 1/\gamma_\mathrm{a}$. The fractions of the symptomatic and asymptomatic subgroups of the infectious population are ν_s and ν_a. We assume that the infection either goes through the symptomatic or the asymptomatic path, but not both for one individual.

and asymptomatic counterparts, β_s, β_a, γ_s, and γ_a,

$$\beta = \nu_\mathrm{s}\,\beta_\mathrm{s} + \nu_\mathrm{a}\,\beta_\mathrm{a} \quad \text{and} \quad \gamma = \nu_\mathrm{s}\,\gamma_\mathrm{s} + \nu_\mathrm{a}\,\gamma_\mathrm{a}\,. \tag{3}$$

Similarly, we can express the overall contact and infectious periods B and C in terms of their symptomatic and asymptomatic counterparts, B_s, B_a, C_s, and C_a,

$$B = \frac{B_\mathrm{a}\,B_\mathrm{s}}{\nu_\mathrm{s}\,B_\mathrm{a} + \nu_\mathrm{a}\,B_\mathrm{s}} \quad \text{and} \quad C = \frac{C_\mathrm{a}\,C_\mathrm{s}}{\nu_\mathrm{s}\,C_\mathrm{a} + \nu_\mathrm{a}\,C_\mathrm{s}}\,. \tag{4}$$

Naturally, the different dynamics for the symptomatic and asymptomatic groups also affect the basic reproduction number R_0, the number of new infections caused by a single one

individual in an otherwise uninfected, susceptible population,

$$R_0 = \frac{C}{B} = \frac{C_\mathrm{a}\, C_\mathrm{s}}{B_\mathrm{a}\, B_\mathrm{s}} \frac{\nu_\mathrm{s}\, B_\mathrm{a} + \nu_\mathrm{a}\, B_\mathrm{s}}{\nu_\mathrm{s}\, C_\mathrm{a} + \nu_\mathrm{a}\, C_\mathrm{s}} = \frac{\nu_\mathrm{s}\, \beta_\mathrm{s} + \nu_\mathrm{a}\, \beta_\mathrm{a}}{\nu_\mathrm{s}\, \gamma_\mathrm{s} + \nu_\mathrm{a}\, \gamma_\mathrm{a}} = \frac{\beta}{\gamma}\,. \tag{5}$$

For a large asymptomatic group $\nu_\mathrm{a} \to 1$, the basic reproduction number approaches the ratio between the infectious and contact periods of the asymptomatic population, $R_0 \to C_\mathrm{a}/B_\mathrm{a}$, which could be significantly larger than the basic reproduction number for the symptomatic group, $R_0 = C_\mathrm{s}/B_\mathrm{s}$, that we generally see reported in the literature. To characterize the effect of changes in social behavior and other interventions that may affect contact, we assume that the contact rate $\beta(t)$ can vary as a function of time [29], but is the same for the symptomatic and asymptomatic groups,

$$\beta = \beta_\mathrm{s} = \beta_\mathrm{a} = \beta(t) \quad \text{such that} \quad R(t) = \frac{C}{B(t)} = \frac{C_\mathrm{a}\, C_\mathrm{s}}{[\nu_\mathrm{s}\, C_\mathrm{a} + \nu_\mathrm{a}\, C_\mathrm{s}]\, B(t)} = \frac{\beta(t)}{\nu_\mathrm{s}\, \gamma_\mathrm{s} + \nu_\mathrm{a}\, \gamma_\mathrm{a}} = \frac{\beta(t)}{\gamma}\,. \tag{6}$$

This introduces a time-varying effective reproduction number $R(t)$, which is an important real time characteristic of the current outbreak dynamics. For the special case when the dynamics of the symptomatic and asymptomatic groups are similar, i.e., $\beta_\mathrm{s} = \beta_\mathrm{a} = \beta$ and $\gamma_\mathrm{s} = \gamma_\mathrm{a} = \gamma$, we can translate the SEIIR model (1) into the classical SEIR model (7) with four compartments, the susceptible, exposed, infectious, and recovered populations [28]. For this special case, we can back-calculate the symptomatic and asymptomatic groups from equation (7.3) as $I_\mathrm{s} = \nu_\mathrm{s}\, I$ and $I_\mathrm{a} = \nu_\mathrm{a}\, I$. Figure 8 illustrates the SEIR model, which

Figure 8: **SEIR epidemiology model.** The SEIR model contains four compartments for the susceptible, exposed, infectious, and recovered populations. The transition rates between the compartments, β, α, and γ are inverses of the contact period $B = 1/\beta$, the latent period $A = 1/\alpha$, and the infectious period $C = 1/\gamma$. If the transition rates are similar for the symptomatic and asymptomatic groups, the SEIIR model simplifies to the SEIR model with $I_\mathrm{s} = \nu_\mathrm{s} I$ and $I_\mathrm{a} = \nu_\mathrm{a} I$.

is governed by a set of four ordinary differential equations [20],

$$\begin{aligned} \dot{S} &= -\beta\, SI \\ \dot{E} &= +\beta\, SI - \alpha\, E \\ \dot{I} &= \qquad + \alpha\, E - \gamma\, I \\ \dot{R} &= \qquad\qquad + \gamma\, I\,. \end{aligned} \tag{7}$$

We draw the daily number of confirmed cases for Santa Clara County, CA, USA [40], New York City, NY, USA [31], and Heinsberg, NRW, Germany [19], and scale the number of reported cases by the total population N to obtain the relative detected population $\hat{D}(t)$. The day on which the relative detected population passed the pandemic outbreak threshold, $\hat{D}(t) > 0.001\%$, marks day 0 and the beginning of our simulation. From this day on, we calculate the simulated detected population, $D(t) = I_\mathrm{s}(t) + R_\mathrm{s}(t)$ with $\dot{R}_\mathrm{s}(t) = \gamma_\mathrm{s}\, I_\mathrm{s}(t)$ and compare it against the relative detected population $\hat{D}(t)$.

Uncertainty quantification. Our SEIIR model uses the following set of parameters, $\boldsymbol{\vartheta} = \{\, A, C_\mathrm{s}, C_\mathrm{a}, \nu_\mathrm{s},\, \beta(t),\, E_0,\, I_{\mathrm{s}0},\, I_{\mathrm{a}0},\, \sigma\,\}$. To reduce the set of unknowns, we fix the la-

tency period $A = 2.5$ days and and the symptomatic infectious period $C_\mathrm{s} = 6.5$ days [25,26,39]. Since the asymptomatic infectious period C_a is unreported, we study three cases with $C_\mathrm{a} : C_\mathrm{s} =$ 0.5, 1.0, and 2.0, resulting in infectious periods of $C_\mathrm{a} =$ 3.25, 6.5, and 13.0 days. We assume that the symptomatic fraction of the infectious group, $\nu_\mathrm{s} = I_\mathrm{s}/I$, is normally distributed. For Santa Clara County, on April 3, 2020, there were 956 detected confirmed cases [40], and an antibody serology study estimated the number of total cases to 25,000 to 91,000 (95% CI) [3] resulting in $\nu_\mathrm{s} =$ 2.43% (95% CI: 1.05%-3.82%). For Heinsberg, between March 31 and April 6, 2020, the detected confirmed cases made up 3.10% of the population [19], and a local antibody prevalence study estimated the fraction of the total cases to 12.30% to 19.00% (95% CI) [46] resulting in $\nu_\mathrm{s} =$ 20.76% (95% CI: 16.32%-25.20%). For New York City, on May 2, 2020, the detected confirmed cases made up 2.02% of the population [31], and a local large-scale antibody testing survey estimated the fraction of the total cases to 19.90% [5] resulting in $\nu_\mathrm{s} =$ 10.15% (95% CI: 6.37%-13.93%), where we assume the 95% confidence width as the weighted average of the relative confidence widths for Santa Clara County and Heinsberg since the New York City credible interval was unreported. We estimate the remaining parameters including the time-varying contact rate $\beta(t)$, the initial exposed population E_0, and the initial symptomatic and asymptomatic infectious populations I_s0 and I_a0 using Bayesian inference. For the time-varying contact rate $\beta(t)$, we set a log-Gaussian random walk prior, which we construct with weakly informative priors with a drift μ_RW and a daily step width σ_RW. For the initial symptomatic infectious population I_s0, we set a weakly informative log-

normal prior distribution with a mean equal to the amount of detected confirmed cases on day 0 and a standard deviation equal to one. We express the initial asymptomatic infectious population as $I_{\mathrm{a0}} = [\,1-\nu_{\mathrm{s}}\,]/\nu_{\mathrm{s}}\,I_{\mathrm{s0}}$, and approximate the initial exposed population as $E_0 = [\,A/C_{\mathrm{s}}\,]\,[\,I_{\mathrm{s0}} + I_{\mathrm{a0}}\,]$ [9]. Table 1 summarizes the choice of our priors.

Table 1: **Prior distributions for SEIIR model parameters.**

Parameter	Interpretation	Distribution	Ref.
ν_{s}	symptomatic fraction	normal(0.0243,0.0071); bounds [0.0105,0.0382] normal(0.1015,0.0193); bounds [0.0637,0.1393] normal(0.2076,0.0227); bounds [0.1632,0.2520]	3 5 46
A C_{s} C_{a}	latent period infectious period infectious period	fixed; 2.5 days fixed; 6.5 days fixed; 3.25 or 6.5 or 13.0 days	25,26,36 26,39 26,39
$\log(\beta(t))$ μ_{RW} σ_{RW}	dynamic contact rate drift daily step width	Gaussian random walk($\mu_{\mathrm{RW}}, \sigma_{\mathrm{RW}}$) normal($\mu = 0.0, \sigma = 1.0$) half-normal($\sigma = 0.02$)	
E_0 I_{s0} I_{a0}	initial exposed initial symptomatic initial asymptomatic	deterministic($[\,A/C_{\mathrm{s}}\,]\,[\,I_{\mathrm{s0}} + I_{\mathrm{a0}}\,]$) log-normal($\hat{D}(t=0)$, 1.0) deterministic($[\,1-\nu_{\mathrm{s}}\,]/\nu_{\mathrm{s}}\,I_{\mathrm{s0}}$)	
σ	likelihood width	half-Cauchy($\beta = 1$)	

For each parameter set, we deterministically calculate the time series of each compartment using an explicit time integrator. For each point in time, we calculate the detected cases, $D(t) = I_{\mathrm{s}}(t) + R_{\mathrm{s}}(t)$ with $\dot{R}_{\mathrm{s}}(t) = \gamma_s\,I_{\mathrm{s}}(t)$. We quantify the likelihood of the parameter set and model outcome in correlation to the reported cases $\hat{D}(t)$ [19,31,40], using Student's t-distribution,

$$p(\hat{D}(t)\,|\,D(t,\boldsymbol{\vartheta})) \sim \mathrm{student\,T}_{\nu=4}(\,\mathrm{mean} = D(t,\boldsymbol{\vartheta}); \mathrm{width} = \sigma\,). \tag{8}$$

We choose this distribution because it resembles a Gaussian distribution and makes the

Markov-Chain Monte Carlo more robust with respect to outliers [6,24]. Here, σ represents the width of the likelihood $p(\hat{D}(t) \mid \boldsymbol{\vartheta})$ between the time-varying reported and the modeled symptomatic populations. Using Bayes' rule, we compute the posterior distribution of the parameters [35,37] to account for the prior knowledge on the parameters and the reported confirmed cases themselves,

$$p(\boldsymbol{\vartheta} \mid \hat{D}(t)) = \frac{p(\hat{D}(t) \mid D(t, \boldsymbol{\vartheta}))\, p(\boldsymbol{\vartheta})}{p(\hat{D}(t))}\,. \tag{9}$$

Since we cannot describe the posterior distribution over the model parameters $\boldsymbol{\vartheta}$ analytically, we adopt approximate-inference techniques to calibrate our model on the available data. We use the NO-U-Turn sampler (NUTS) [21], which is a type of Hamiltonian Monte Carlo algorithm as implemented in PyMC3 [38]. We use four chains, and the first 500 samples serve to tune the sampler and are later discarded. We use the subsequent 1000 samples as the posterior distribution for the parameters $\boldsymbol{\vartheta}$. From the converged posterior distribution, we sample multiple combinations of parameters that describe the time evolution of reported cases. Using these posterior samples, we quantify the uncertainty of each parameter based on the reported case data. As such, each parameter set provides a set of values for the initial exposed population E_0, the initial symptomatic and asymptomatic populations $I_{\mathrm{s}0}$ and $I_{\mathrm{a}0}$, the symptomatic fraction ν_{s}, and the time evolution of the contact rate $\beta(t)$. From these values, we quantify the effective reproductive number $R(t)$ and the time evolution of the susceptible, exposed, symptomatic infectious, asymptomatic infectious and recovered populations, $S(t)$, $E(t)$, $I_{\mathrm{s}}(t)$, $I_{\mathrm{a}}(t)$, and $R(t)$ and report their values with the associated 95% credible interval.

Forecasting. From the posterior distributions, for the three asymptomatic infectious periods $C_{\mathrm{a}} = 3.25$, 6.5 and 13.0 days, we predict the COVID-19 outbreak dynamics from mid May to mid June. As a forecasting scenario, we assume a gradual relaxation of the current public health interventions that translates into a 10% increase in the current contact rate $\beta(t)$. With this projected new $\beta(t)$, we predict the daily increments in each SEIIR compartment. To quantify the number of hospital and intensive care unit beds needed for this scenario, we introduce the fraction of symptomatic individuals that requires hospitalization ν_{h} and the fraction of hospitalized individuals that requires intensive care ν_{c} [30,36]. Since the hospitalization and intensive care fractions ν_{h} and ν_{c} strongly depends on the local testing frequency and age demographics, we learn these fractions as location-dependent parameters from the reported hospitalizations and intensive care unit needs [40]. Specifically, we assume a mean hospitalized period of four days and a mean intensive care unit period of ten days [39] and use a Nelder-Mead optimization algorithm [16] to find the most probable ν_{h} and ν_{c} fractions.

Estimating the outbreak date. For each sample from the posterior distribution, we use the estimated initial exposed population E_0 and the estimated initial asymptomatic infectious population $I_{\mathrm{a}0}$ to estimate the date of the very first COVID-19 case in Santa Clara County [40]. Specifically, for each parameter set, we create an SEIIR model and assume that the outbreak begins with one single infectious individual. We fix the latency and symptomatic infectious periods to $A = 2.5$ days and $C_{\mathrm{s}} = 6.5$ days and the three asymptomatic infectious periods $C_{\mathrm{a}} = 3.25$, 6.5 and 13.0 days. For each C_{a} case and posterior sample

for the exposed, symptomatic infectious, and asymptomatic infectious population size at day 0, on March 2, 2020, we use the Nelder-Mead optimization method [16] to find the most probable outbreak origin date. Specifically, we solve the SEIIR model forward in time using an explicit time integration, starting from various start dates before March 2, 2020, and iteratively minimize the difference between the computed exposed, symptomatic, and asymptomatic infectious populations and the sample's actual exposed, symptomatic, and asymptomatic infectious populations. We concomitantly fit a static contact rate parameter β which is bounded between zero and the posterior sample's estimated contact rate on day 0, $\beta(0)$. Repeating this process for each sample of the Bayesian inference generates a distribution of possible origin dates. From this distribution, we compute the most probable origin date and its uncertainty.

Acknowledgements

This work was supported by a DAAD Fellowship (K.L.), by a Stanford Bio-X IIP Seed Grant (M.P. and E.K.), and by the Stanford COVID-19 Seroprevalence Study Fund (E.B. and J.B.).

References

1. E. Allday, M. Kawahara. First known U.S. coronavirus death occurred on Feb. 6 in Santa Clara County. San Francisco Chronicle, April 22, 2020. https://www.sfchronicle.com/health/article/First-known-U-S-coronavirus-death-occurred-on-15217316.php assessed: May 13, 2020.

2. J. L. Aron, I. B. Schwartz. Seasonality and period-doubling bifurcation in an epidemic model. Journal of Theoretical Biology 110 (1984) 665-679.

3. E. Bendavic, B. Mulaney, N. Sood, S. Shah, E. Ling, R. Bromley-Dulfano, C. Lai, Z. Weissberg, R. Saavedra-Walker, J. Tedrow, D. Tversky, A. Bogan, T. Kupiec, D. Eichner, R. Gupta, J.P.A. Ioannidis, J. Bhattacharya. COVID-19 antibody seroprevalence in Santa Clara County, California. medRxiv doi:10.1101/2020.04.14.20062463.

4. A. Bryan, G. Pepper G, M.H. Wener, S.L. Fink, C. Morishima, A. Chaudhary, K. Jerome, P.C. Mathias, A. Greninger. Performance characteristics of the abbott architect SARS-CoV-2 IgG assay and seroprevalence testing in Idaho. medRxiv doi:10.1101/2020.04.27.20082362.

5. A.M. Cuomo. Amid ongoing COVID-19 pandemic, governor cuomo announces results of completed antibody testing study. https://www.governor.ny.gov/news/amid-ongoing-covid-19-pandemic-governor-cuomo-announces-results-completed-antibody-testing. assessed: May 13, 2020.

6. J. Dehning, J. Zierenberg, F.P Spitzner, M. Wibral, J.P. Neto, M. Wilczek, V. Priesemann. Inferring COVID-19 spreading rates and potential change points for case number forecasts arXiv (2020) preprint:2004.01105

7. A. Doi, K. Iwata, H. Kuroda, T. Hasuike, A. Kanda, T. Nagao, H. Nishioka, K. Tomii, T. Morimoto, Y. Kihara. Seroprevalence of novel coronavirus disease (COVID-19) in

Kobe, Japan. medRxiv doi:10/1101/2020.04.26.20079822.

8. E. Dong, L. Gardner. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infectious Disease (2020) doi:10.1016/S1473-3099(20)30120-1.

9. R. Engbert, F.R. Drepper. Chance and chaos in population biology–Models of recurrent epidemics and food chain dynamics. Chaos, Solutions & Fractals 4 (1994) 1147-1169.

10. C. Erikstrup, C.E. Hother, O.B. Vestager Pedersen, K. Molbak, R.L. Skov, D. Kinggaard Holm, S. Saekmose, A.C. Nilsson, P. Terrence Brooks, J. Kjaergaard Boldsen, C. Mikkelsen, M. Gybel-Brask, E. Sorensen, K.M. Dinh, S. Mikkelsen, B.K. Moller, T. Haunstrup, L. Harritshoj, B.A. Jensen, H. Hjalgrim, S.T. Lillevang, H. Ullum. Estimation of SARS-CoV-2 infection fatality rate by real-time antibody screening of blood donors. medRxiv doi:10.1101/2020.04.24.20075291.

11. Y. Fang, Y. Nie, M. Penny. Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions. J. Med. Virol. (2020) 1-15.

12. A.S. Fauci, H.C. Lane, R.R. Redfield. Covid-19–Navigating the uncharted. New England Journal of Medicine 382 (2020) 1268-1269.

13. A.Fontanet, L. Tondeur, Y. Madec, R. Grant, C. Besombes, N. Jolly, S. Fernandes Pellerin M.N. Ungeheuer, I. Cailleau, L. Kuhmel, S. Temmam, C. Huon, K.Y. Chen, B. Crescenzo, S. Munier, C. Demeret, L. Grzelak, I. Staropoli, T. Bruel, P.

Gallian, S. Cauchemez, S. van der Werf, O. Schwartz, M. Eloit, B. Hoen. Cluster of COVID-19 in northern France: A retrospective closed cohort study. medRxiv 10.1101/2020.04.20071134.

14. C. Fraser, S. Riley, R.M. Anderson, N.M. Ferguson. Factors that make an infectious disease outbreak controllable. Proceedings of the National Academy of Sciences 101 (2004) 6146-6151.

15. M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, A. Rinaldo. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. Proceedings of the National Academy of Sciences; in press, doie:10.1073/pnas.2004978117.

16. F. Gao, L. Han. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. Computational Optimization and Applications 51(2012) 259-277.

17. A. Gelman, ,J. Hill. Data Analysis using Regression and Multilevel/Hierarchical Models. Cambridge University Press, 2006.

18. A. Grifoni, D. Weiskopf, S.I. Ramirez, J. Mateus, J.M. Dan, C. Rydyznski Moderbacher, S.A. Rawlings, A. Sutherland, L. Premkumar R.S. Jadi, D. Marrama, A.M. de Silva, A. Frazier, A. Carlin, J.A. Greenbaum, B. Peters. F. Krammer. D.M. Smith, S. Crotty, A. Sette. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. Cell, doi: 10.1016/j.cell.2020.05.015.

19. Heinsberg 2020. Aktuelles aus dem Kreishaus. Coronavirus im Kreis Heinsberg. https://www.kreis-heinsberg.de/aktuelles/aktuelles/?pid=5149. assessed: May 13, 2020.

20. H. W. Hethcote. The mathematics of infectious diseases. SIAM Review 42 (2000) 599-653.

21. M.D. Hoffman, A. Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research (2014), 15(1), 1593-1623.

22. J.P.A. Ioannidis. The infection fatality rate of COVID-19 inferred from seroprevalence data. medRxiv. doi:10.1101/2020.05.13.20101253.

23. W. O. Kermack, G. McKendrick. Contributions to the mathematical theory of epidemics, Part I. Proceedings of the Royal Society London Series A 115 (1927) 700-721.

24. K.L. Lange, R.J.A. Little, M.G. Taylor. Robust statistical modeling using the t distribution. Journal of the American Statistical Association 84 (1989) 881-896.

25. S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, J. Lessler. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Annals of Internal Medicine (2020) doi:10.7326/M20-0504.

26. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K.S.M. Leung, E.H.Y. Lau, J.Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T.T.Y. Lam, J.T. Wu, G.F. Gao, B.J. Cowling, B. Yang, G.M. Leung, Z. Feng. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. New England Journal of Medicine 382 (2020) 1199-1207.

27. R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science 368 (2020) 489-493.

28. K. Linka, M. Peirlinck, F. Sahli Costabal, E. Kuhl. Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. Computer Methods in Biomechanics and Biomedical Engineering (2020) in press; doi:10.1080/10255842.2020.1759560.

29. K. Linka, M. Peirlinck, E. Kuhl. The reproduction number of COVID-19 and its correlation with public health interventions. medRxiv doi:10.1101/2020.05.01.20088047.

30. N.B. Noll, I. Aksamentov, V. Druelle, A. Badenhorst, B. Ronzani, G. Jefferies, J. Albert, R. Neher. COVID-19 Scenarios: an interactive tool to explore the spread and associated morbidity and mortality of SARS-CoV-2. medRxiv doi: 10.1101/2020.05.05.20091363

31. New York Times. An ongoing repository of data on coronavirus cases and deaths in the U.S. https://github.com/nytimes/covid-19-data, assessed: May 13, 2020.

32. X. Pan, D. Chen, Y. Xia, X, Wu, T. Li, X. Ou, L. Zhou, J. Liu. Asymptomatic cases in a family cluster with SARS-CoV-2 infection. Lancet Infectious Diseases 20 (2020) 410-411.

33. S.W. Park, D.M. Cornforth, J. Dushoff, J.W. Weitz. The time scale of asymptomatic transmission affects estimates of epidemic potential in the COVID-19 outbreak. medRxiv 2020.03.09.20033514, doi 10.1101/2020.03.09.20033514.

34. S.W. Park, K. Sun, C. Viboud, B.T. Grenfell, J. Dushoff. Potential roles of social distancing in mitigating the spread of coronavirus disease 2019 (COVID-19) in South Korea. medRxiv 2020.03.27.20045815, doi 10.1101/2020.03.27.20045815

35. M. Peirlinck, F. Sahli Costabal, K.L. Sack, J.S. Choy, G.S. Kassab, J.M. Guccione, M. De Beule, P. Segers, E. Kuhl. Using machine learning to characterize heart failure across the scales. Biomechanics and Modeling in Mechanobiology 18 (2019) 1987-2001.

36. M. Peirlinck, K. Linka, F. Sahli Costabal, E. Kuhl. Outbreak dynamics of COVID-19 in China and the United States. Biomechanics and Modeling in Mechanobiology (2020) in press, doi:10.1101/2020.04.06.20055863.

37. F. Sahli Costabal, K. Matsuno, J. Yao, P. Perdikaris, E. Kuhl. Machine learning in drug development: Characterizing the effect of 30 drugs on the QT interval using Gaus-

sian process regression, sensitivity analysis, and uncertainty quantification. Computer Methods in Applied Mechanics and Engineering 348 (2019) 313-333.

38. J. Salvatier, T.V. Wiecki, C. Fonnesbeck. Probabilistic programming in Python using PyMC3. Peer Journal Computational Science 2 (2016) e55.

39. S. Sanche, Y.T. Lin, C. Xu, R. Romero-Severson, N. Hengartner, R. Ke. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. Emerging Infectious Disease (2020) doi:10.3201/eid2607.200282

40. Santa Clara County COVID-19 Cases and Hospitalizations Dashboard. www.sccgov.org assessed: May 13, 2020.

41. M. Shakiba, S. Nazari, F. Mehrabian, S.M. Rezvani, Z. Ghasempour, A. Heidarzadeh. Seroprevalence of COVID-19 virus infection in Guilan province, Iran. medRxiv doi:10.1101/2020.04.26.20079244.

42. M.F. Silveira, A.J.D. Barros, B.L. Horta, L.C. Pellanda, O. Dellagostin, C. Struchiner, M. Burattini, A. Valim, E. Berlezi, J. Mesa, M.L. Ikeda, M. Mesenburg, M. Mantesso, M. Dall'Agnol, R. Bittencourt, F.P. Hartwig, A.M. Menezes, F.C. Barros, P. Hallal, C.G. Victora. Repeated population-based surveys of antibodies against SARS2-CoV-2 in Southern Brazil. medRxiv doi:10.1101/2020.05.01.20087205.

43. E. Slot, B.M. Hogema, C.B.E.M. Reusken, J.H. Reimerink, M. Molier, H.M. Karregat, J. Ijist, V.M.J. Novotny, R.A.W. van Lier, H.L. Zaaijer. Herd immunity is

not a realistic exit strategy during a COVID-19 outbreak. Research Square 2020. https://dx.doi.org/10.21203/rs.3.rs-25862/v1.

44. N. Sood, P. Simon, P. Ebner, D. Eichner, J. Reynolds, E. Bendavid, J. Bhattacharya. Seroprevalence of SARS-CoV-2-specific antibodies among adults in Los Angeles County, California, on April 10-11, 2020. Journal of the American Medical Association. doi:10.1001/jama.2020.8279.

45. B. Spellberg, M. Haddix, R. Lee, S. Butler-Wu, P. Holtom, H. Yee, P. Gounder. Community prevalence of SARS-CoV-2 among patients with influenzalike illnesses presenting to a Los Angeles medical center in March 2020. Journal of the American Medical Association. doi:10.1001/jama.2020.4958.

46. H. Streeck, B. Schulte, B.M. Kümmerer, E. Richter, T. Höller, C. Fuhrmann, E. Bartok, R. Dolscheid, M. Berger, L. Wessendorf, M. Eschbach-Bludau, A. Kellings, A. Schwaiger, M. Coenen, P. Hoffmann, B. Stoffel-Wagner, M.M. Nöthen, A.M. Eis-Hübinger, M. Exner2, R.M. Schmithausen, M. Schmid, G. Hartmann. Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. medRxiv doi:10.1101/2020.05.04.20090076.

47. S. Stringhini, A. Wisniak, G. Piumatti, A.S. Azman, S.A. Lauer, H. Baysson, D. De Ridder, D. Petrovic, S. Schrempft, M. Kailing, I. Arm-Vernez, S. Yerly, O. Keiser, S. Hurst, K. Posfay-Barbe, D. Trono, D. Pittet, L. Getaz, F. Chappuis, I. Eckerle, N. Vuilleumier, B. Meyer, A. Flahault, L. Kaiser, I. Guessous. Repeated seropreva-

lence of anti 1 -SARS-CoV-2 IgG antibodies in a population-based sample. medRxiv doi:10.1101/2020.05.02.20088898.

48. B. Tang, F. Xia, N.L. Bragazzi, Z. McCarthy, X. Wang, S. He, X. Sun, S. Tang, Y. Xiao, J. Wu. Lessons drawn from China and South Korea for managing COVID-19 epidemic: insights from a comparative modeling study. Bulletin of the World Health Organization. doi: http://dx.doi.org/10.2471/BLT.20.257238.

49. C. Thompson, N. Grayson, R.S. Paton, J. Lourenco, B.S. Penman, L. Lee. Neutralising antibodies to SARS coronavirus 2 in Scottish blood donors – a pilot study of the value of serology to determine population exposure. medRxiv doi:10.1101/2020.04.13.20060467.

50. T.A. Treibel, C. Manisty, M. Burton, A. McKnight, J. Lambourne, J.B. Augusto, X. Couto-Parada, T. Cutino-MOguel, M. Noursadeghi, J.C. Moon. COVID-19: PCR screening of asympotmatic healthcare workers at London hospital. Lancet. doi:10.1016/S0140-6736(20)31100-4.

51. A. Viguerie, G. Lorenzo, F. Auricchio, D. Baroli, T.J.R. Hughes, A. Patton, A. Reali, T.E. Yankeelov, A. Veneziani. Simulating the spread of COVID-19 via a spatially-resolved SEIRD model with heterogeneous diffusion. Oden Institute Report 20-09, Austin.

52. X. Wu, B. Fu, L. Chen, Y. Feng. Serological tests facilitate identification of asymptomatic SARS-CoV-2 infection in Wuhan, China. Journal of Medical Virology. doi:10.1002/jmv.25904.