medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

Mathias Peirlinck[a], Kevin Linka[a], Francisco Sahli Costabal[b], Jay Bhattacharya[c], Eran Bendavid[c], John P.A. Ioannidis[c,d], Ellen Kuhl[a]

[a]*Department of Mechanical Engineering, Stanford University School of Engineering, Stanford, California, United States*
[b]*Department of Mechanical and Metallurgical Engineering and Institute for Biological and Medical Engineering, Schools of Engineering, Biology and Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile*
[c]*Department of Medicine, Stanford University School of Medicine, Stanford, California, United States*
[d]*Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, California, United States*

Abstract

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 computes, 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 populations. Our study quantifies the sensitivity of the outbreak dynamics of COVID-19 to three parameters: the effective reproduction number, the ratio between the symptomatic and asymptomatic populations, and the infectious periods of both groups. For nine distinct locations, our model estimates the fraction of the population that has been infected and recovered by Jun 15, 2020 to 24.15% (95% CI: 20.48%-28.14%) for Heinsberg (NRW, Germany), 2.40% (95% CI: 2.09%-2.76%) for Ada County (ID, USA), 46.19% (95% CI: 45.81%-46.60%) for New York City (NY, USA), 11.26% (95% CI: 7.21%-16.03%) for Santa Clara County (CA, USA), 3.09% (95% CI: 2.27%-4.03%) for Denmark, 12.35% (95% CI: 10.03%-15.18%) for Geneva Canton (Switzerland), 5.24% (95% CI: 4.84%-5.70%) for the Netherlands, 1.53% (95% CI: 0.76%-2.62%) for Rio Grande do Sul (Brazil), and 5.32% (95% CI: 4.77%-5.93%) for Belgium. Our method traces the initial outbreak date in Santa Clara County back to January 20, 2020 (95% CI: December 29, 2019 - February 13, 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, managing community transmission through increasing population awareness, promoting physical distancing, and encouraging behavioral changes could become more relevant.

Keywords: COVID-19; epidemiology; asymptomatic transmission; machine learning; uncertainty quantification; Bayesian inference

1. Motivation

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 [9]. In the early stages of a pandemic, doctors, researchers, and political decision makers mainly focus on symptomatic individuals that come for testing and address those who require the most urgent medical attention [14]. 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 [34]. In this manuscript, we collectively use the term "asymptomatic" for

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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, 8, 11, 13, 17, 23, 28, 30, 37, 46, 47, 52, 53, 56, 54, 55, 59, 60, 62, 64, 66, 68]. 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 [15]. 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 [41]. As of today, more than 50 studies have reported an asymptomatic population, twenty-three of them with a sample size of at least 500 [27], 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, 6, 59]. To no surprise, the reported undercount varies hugely, ranging from 3.5 and 5.0 in Luxembourg [56] and Germany [59] to 543 and 627 in Iran [53] and Japan [8] respectively. 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 [68], a study of 400 health care workers in London with an undercount of 35.0 [65], a community spreading study of 131 patients with influenza-like symptoms in Los Angeles with an undercount of 100.0 [58], and a seroprevalence study in Los Angeles county with an undercount of 43.5 [57]. 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 [15], and second, to reliably predict the success of surveillance and control strategies, e.g., contact tracing or vaccination [18].

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 [34]. Many classical epidemiology models base their predictions on compartment models in which individuals pass through different stages as they experience the disease [29]. A popular model to simulate the outbreak dynamics of COVID-19 is the SEIR model [14], 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 [42] and to illustrate how asymptomatic individuals have facilitated the rapid spread of COVID-19 throughout China [34], South Korea [63], and Italy [19]. 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 [45], 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 [36]. 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.

2. Methods

2.1. 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 1 illustrates our SEIIR

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

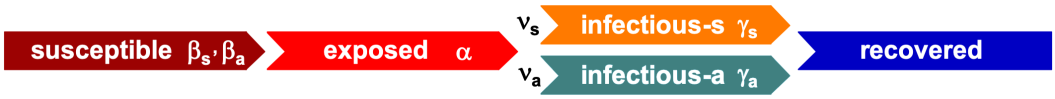

Figure 1: **SEIIR epidemiology model.** The SEIIR model contains five compartments for the susceptible, exposed, symptomatic infectious, asymptomatic infectious, and recovered populations. The transition rates between the compartments, β, α, and γ are inverses of the contact period $B = 1/\beta$, the 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 $B_s = 1/\beta_s$ and $B_a = 1/\beta_a$ and individual infectious periods $C_s = 1/\gamma_s$ and $C_a = 1/\gamma_a$. 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.

model, which is governed by a set of five ordinary differential equations,

$$\begin{aligned}
\dot{S} &= -S\,[\beta_s I_s + \beta_a I_a] \\
\dot{E} &= +S\,[\beta_s I_s + \beta_a I_a] - \alpha E \\
\dot{I}_s &= + \nu_s \alpha E - \gamma_s I_s \\
\dot{I}_a &= + \nu_a \alpha E - \gamma_a I_a \\
\dot{R} &= + \gamma_s I_s + \gamma_a I_a\,,
\end{aligned} \tag{1}$$

where the fractions of all five populations add up to one, $S + E + I_s + I_a + R = 1$. We assume that both the symptomatic group I_s and the asymptomatic group I_a can generate new infections. We introduce these two groups by fractions ν_s and ν_a of the subjects transferring from the exposed group E. We postulate that the two infectious groups I_s and I_a have the same latent period $A = 1/\alpha$, but can have individual contact periods $B_s = 1/\beta_s$ and $B_a = 1/\beta_a$ to mimic their different community spreading, and individual infectious periods $C_s = 1/\gamma_s$ and $C_a = 1/\gamma_a$ to mimic their different likelihood of isolation. From the infectious fractions, we can derive the overall contact and infectious rates β and γ from their individual symptomatic and asymptomatic counterparts, β_s, β_a, γ_s, and γ_a,

$$\beta = \nu_s \beta_s + \nu_a \beta_a \quad \text{and} \quad \gamma = \nu_s \gamma_s + \nu_a \gamma_a\,. \tag{2}$$

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,

$$B = \frac{B_a B_s}{\nu_s B_a + \nu_a B_s} \quad \text{and} \quad C = \frac{C_a C_s}{\nu_s C_a + \nu_a C_s}\,. \tag{3}$$

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,

$$R_0 = \frac{C}{B} = \frac{C_a C_s}{B_a B_s}\,\frac{\nu_s B_a + \nu_a B_s}{\nu_s C_a + \nu_a C_s} = \frac{\nu_s \beta_s + \nu_a \beta_a}{\nu_s \gamma_s + \nu_a \gamma_a} = \frac{\beta}{\gamma}\,. \tag{4}$$

For a large asymptomatic group $\nu_a \to 1$, the basic reproduction number approaches the ratio between the infectious and contact periods of the asymptomatic population, $R_0 \to C_a/B_a$, which could be significantly larger than the basic reproduction number for the symptomatic group, $R_0 = C_s/B_s$, 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 [36], but is the same for the symptomatic and asymptomatic groups,

$$\beta = \beta_s = \beta_a = \beta(t) \quad \text{such that} \quad R(t) = \frac{C}{B(t)} = \frac{C_a C_s}{[\nu_s C_a + \nu_a C_s]\,B(t)} = \frac{\beta(t)}{\nu_s \gamma_s + \nu_a \gamma_a} = \frac{\beta(t)}{\gamma}\,. \tag{5}$$

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., $\beta_s = \beta_a = \beta$ and $\gamma_s = \gamma_a = \gamma$, we can translate the SEIIR model (1) into the classical SEIR model (6) with four compartments, the susceptible, exposed, infectious, and recovered populations [35]. For this special case,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 2: **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 $I_s = \nu_s I$ and $I_a = \nu_a I$.

we can back-calculate the symptomatic and asymptomatic groups from equation (6.3) as $I_s = \nu_s\, I$ and $I_a = \nu_a\, I$. Figure 2 illustrates the SEIR model, which is governed by a set of four ordinary differential equations [25],

$$\begin{aligned} \dot{S} &= -\beta\, S\, I \\ \dot{E} &= +\beta\, S\, I \;-\; \alpha\, E \\ \dot{I} &= \qquad\quad +\; \alpha\, E \;-\; \gamma\, I \\ \dot{R} &= \qquad\qquad\qquad +\; \gamma\, I\,. \end{aligned} \tag{6}$$

2.2. *Seroprevalence studies.*

We consider seroprevalence studies where the representation ratio, the ratio between the amount of seroprevalence samples and the representative population, is larger than 0.02%. Subsequently, we draw the daily number of confirmed reported cases for Heinsberg, Ada County, New York City, Santa Clara County, Denmark, Geneva Canton, Netherlands, Rio Grande do Sul, and Belgium from online data repositories [5, 10, 24, 39, 40, 61]. The symptomatic fraction ν_s follows from the ratio between the reported confirmed cases for the studied region and the amount of seroprevalence-estimated cases on the last day of the seroprevalence study. Table 1 summarizes the study date, number of samples, representation ratio, population, and symptomatic fraction of each location.

Table 1: **Seroprevalence studies with a representative population larger than** 0.02%**, location, study date, number of samples, representation ratio, population, and symptomatic fraction** ν_s**.**

Location	Study date	Samples	Population	Repr.ratio	Symptomatic fraction ν_s	Ref.
Heinsberg, NRW, Germany	March 30 - April 6	919	41946	2.191%	20.00% (95%CI: 16.32% - 25.20%)	[59]
Ada County, ID, USA	Late April	4856	481587	1.008%	7.90% (95%CI: 6.40%-9.82%)	[4]
New York City, NY, USA	May 5 - Jun 5	28523	8398748	0.340%	5.76% (95%CI: 5.68%-5.84%)	[46]
Santa Clara County, CA, USA	April 2 - 3	3330	1781642	0.187%	1.77% (95%CI: 1.05% - 3.82%)	[3]
Denmark	April 6 - 17	9496	5824857	0.163%	6.95% (95%CI: 5.13%-13.12%)	[13]
Geneva Canton, Switzerland	April 20-27	576	504031	0.114%	10.34% (95%CI: 7.66%-16.44%)	[60]
Netherlands	April 1-15	7361	17282163	0.043%	17.31% (95%CI: 15.14%-19.95%)	[55]
Rio Grande do Sul, Brasil	April 25 - 27	4188	11377239	0.037%	8.10% (95%CI: 3.72%-22.07%)	[54]
Belgium	April 14 - May 13	2700	11492641	0.023%	10.21% (95%CI: 8.61%-12.18%)	[52]

2.3. *Uncertainty quantification.*

Our SEIIR model uses the following set of parameters, $\boldsymbol{\vartheta} = \{A, C_s, C_a, \nu_s, \beta(t), E_0, I_{s0}, I_{a0}, \sigma\}$. To reduce the set of unknowns, we fix the latency period $A = 2.5$ days and the symptomatic infectious period $C_s = 6.5$ days [32, 33, 50]. Since the asymptomatic infectious period C_a is unreported, we study three cases with $C_a : C_s = 0.5$, 1.0, and 2.0, resulting in infectious periods of $C_a = 3.25$, 6.5, and 13.0 days. We impose the symptomatic fraction of the infectious group, ν_s to follow the distributions from Table 1. 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 $I_{a0} = [\,1 - \nu_s\,]/\nu_s\, I_{s0}$, and approximate the initial exposed population as $E_0 = [\,A/C_s\,]\,[\,I_{s0} + I_{a0}\,]$ [12]. Table 2 summarizes the choice of our priors.

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, $D(t) = I_s(t) + R_s(t)$ with $\dot{R}_s(t) = \gamma_s\, I_s(t)$. By scaling

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 2: **Prior distributions for SEIIR model parameters.**

Parameter	Interpretation	Distribution
ν_s	symptomatic fraction	Table 1
A	latent period	fixed; 2.5 days [33, 32, 45]
C_s	symptomatic infectious period	fixed; 6.5 days [33, 50]
C_a	asymptomatic infectious period	fixed; 3.25 or 6.5 or 13.0 days [33, 50]
$\log(\beta(t))$	dynamic contact rate	Gaussian random walk(μ_{RW}, σ_{RW})
μ_{RW}	drift	normal($\mu = 0.0, \sigma = 1.0$)
σ_{RW}	daily step width	half-normal($\sigma = 0.02$)
E_0	initial exposed	determinstic([A/C_s] [$I_{s0} + I_{a0}$]) [12]
I_{s0}	initial symptomatic	log-normal($\hat{D}(t = 0)$, 1.0)
I_{a0}	initial asymptomatic	deterministic([$1 - \nu_s$]/$\nu_s I_{s0}$)
σ	likelihood width	half-Cauchy($\beta = 1$)

the confirmed reported cases with the total population N, we obtain the relative detected population $\hat{D}(t)$. The day on which each location went into lockdown marks day 0 and the beginning of our simulation. From this day on, we calculate the simulated detected population, $D(t) = I_s(t) + R_s(t)$ with $\dot{R}_s(t) = \gamma_s I_s(t)$. We quantify the likelihood of the parameter set and model outcome $D(t)$ in correlation to the reported cases $\hat{D}(t)$ [51, 40, 24], using Student's t-distribution,

$$p(\hat{D}(t) \,|\, D(t, \boldsymbol{\vartheta})) \sim \text{student}\,\mathrm{T}_{\nu=4}(\,\text{mean} = D(t, \boldsymbol{\vartheta}); \text{width} = \sigma\,). \tag{7}$$

We choose this distribution because it resembles a Gaussian distribution and makes the Markov-Chain Monte Carlo more robust with respect to outliers [7, 31]. Here, σ represents the width of the likelihood $p(\hat{D}(t) \,|\, \boldsymbol{\vartheta})$ between the time-varying reported and the modeled symptomatic populations. Using Bayes' rule, we compute the posterior distribution of the parameters [44, 48] to account for the prior knowledge on the parameters and the reported confirmed cases themselves,

$$p(\boldsymbol{\vartheta} \,|\, \hat{D}(t)) = \frac{p(\hat{D}(t) \,|\, D(t, \boldsymbol{\vartheta}))\; p(\boldsymbol{\vartheta})}{p(\hat{D}(t))}\,. \tag{8}$$

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) [26], which is a type of Hamiltonian Monte Carlo algorithm as implemented in PyMC3 [49]. We use four chains, and the first 4 times 500 samples serve to tune the sampler and are later discarded. We use the subsequent 4 times 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_{s0} and I_{a0}, 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, $S(t)$, $E(t)$, $I_s(t)$, $I_a(t)$, and $R(t)$ and report their values with the associated 95% credible interval.

2.4. Hierarchical asymptomatic infectious period C_a estimation.

To estimate the asymptomatic infectious period C_a, we combine the confirmed case data from all locations and set up a hierarchical model for C_a. Here we postulate that each location-specific $C_a^{\text{loc i}}$ is drawn from a hyperdistribution C_a. We assume that the initial reproduction number, which is disease-specific, is the same for all locations [16] and is constant during a two-week time window around each local lockdown date: 10 days prior till 3 days ($\sim A$, the latent period) after each local lockdown date. The priors for these distributions are summarized in Table 3. We perform

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Bayesian inference as explained above and subsequently use the hierarchical posterior C_a distribution to quantify the time-varying asymptomatic infectious population size at each of the locations after the lockdowns.

Table 3: **Prior distributions for SEIIR-based hierarchical C_a estimation.**

Parameter	Interpretation	Distribution
ν_s	symptomatic fraction	Table 1
A	latent period	fixed; 2.5 days [33, 32, 45]
C_s	symptomatic infectious period	fixed; 6.5 days [33, 50]
C_a	asymptomatic infectious period	normal($\mu = 6.5, \sigma = \sigma_{C_a}$)
σ_{C_a}		half-normal($\sigma = 2.0$)
$C_a^{\text{loc i}}$		normal($\mu = C_a, \sigma = 2.0$)
R_0	initial reproduction number	normal($\mu = 3.87, \sigma = 0.44$) [16]
E_0	initial exposed	determinstic([A/C_s] [$I_{s0} + I_{a0}$]) [12]
I_{s0}	initial symptomatic	log-normal($\hat{D}(t = 0)$, 1.0)
I_{a0}	initial asymptomatic	deterministic([$1 - \nu_s$]/ν_s I_{s0})
σ	likelihood width	half-Cauchy($\beta = 1$)

2.5. 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_{a0} to estimate the date of the very first COVID-19 case in Santa Clara County [51]. Specifically, for each parameter set, we create an SEIIR model and assume that the outbreak begins with one single asymptomatic infectious individual. We fix the latency and symptomatic infectious periods to $A = 2.5$ days and $C_s = 6.5$ days and impose the hierarchically estimated infectious period C_a. For each posterior sample for the exposed, symptomatic infectious, and asymptomatic infectious population size on the lockdown date (March 16, 2020), we use the Nelder-Mead optimization method [20] 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 16, 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.

3. Results

3.1. Outbreak dynamics of COVID-19 in Santa Clara County.

Figure 3 illustrates the outbreak dynamics of COVID-19 in Santa Clara County. To explore the effects of C_a = on the initial effective reproductive number, we started the simulations 10 days before the local lockdown date: March 6, 2020. 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 $C_s = 6.5$ days, and for three asymptomatic infectious periods, $C_a = 3.25$, 6.5, and 13.0 days, from left to right. The gray and green-purple 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 fraction ν_s, and the initial exposed and infectious populations E_0, I_{s0}, and I_{a0}. The red, orange, and green histograms on the middle row display the inferred initial exposed and infectious populations, E_0, I_{s0}, and I_{a0}, for the three different asymptomatic infectious periods. The graphs confirm that our dynamic SEIIR epidemiology model is capable of correctly capturing the gradual flattening and deflattening of the

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

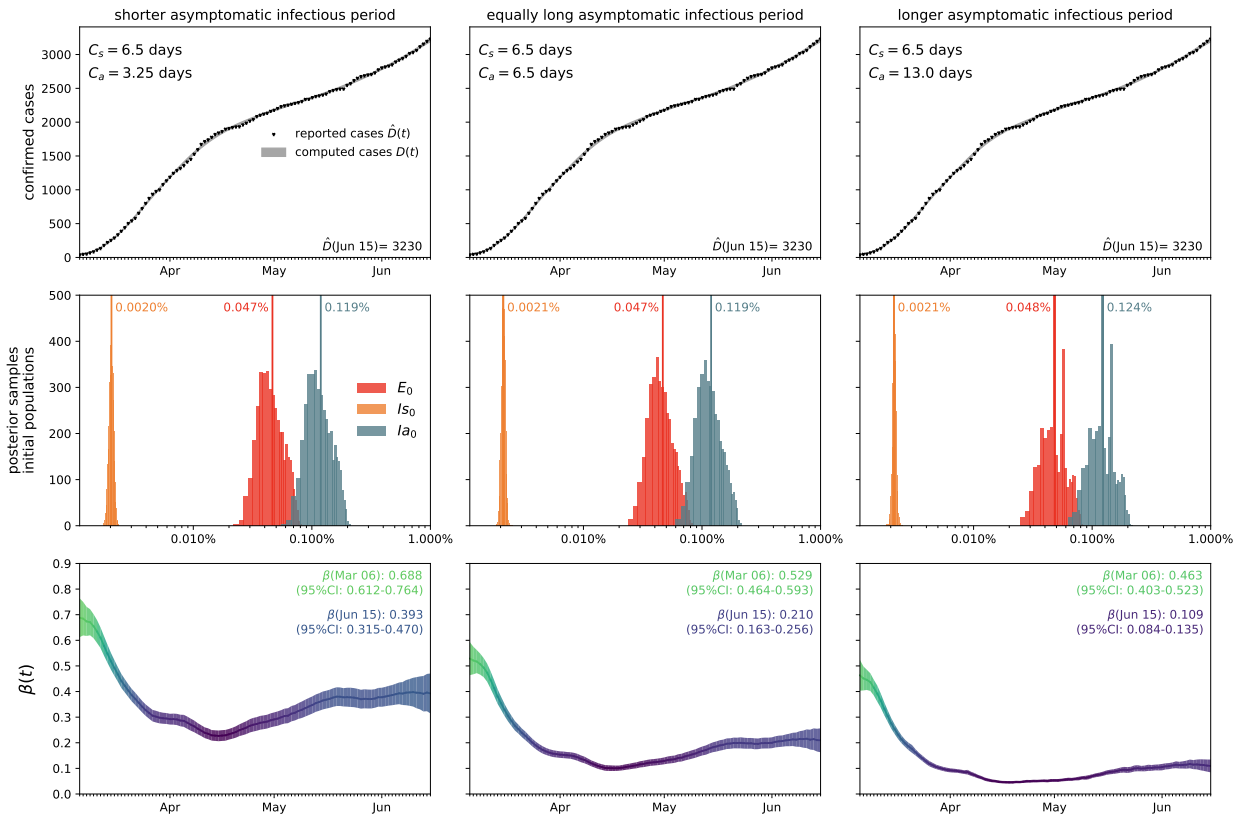

Figure 3: **Outbreak dynamics of COVID-19 in Santa Clara County.** The simulation learns the time-varying contact rate $\beta(t)$ for fixed latent and symptomatic infectious periods $A = 2.5$ days and $C_s = 6.5$ days, and for three asymptomatic infectious periods $C_a = 3.25$ days, 6.5 days, and 13.0 days (from left to right). Computed and reported confirmed cases in Santa Clara County, $D(t) = I_s(t) + R_s(t)$ and $\hat{D}(t)$ (top), initial exposed and infectious populations, E_0, I_{s0}, and I_{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 $\nu_s = I_s/I$, and the initial exposed and infectious populations E_0, I_{s0}, and I_{a0}.

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 $\beta(t)$ after the lockdown date (March 16, 2020) 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_a: For larger asymptomatic infectious periods C_a, from left to right, to explain the same number of confirmed cases $D(t) = I_s(t) + R_s(t)$, the contact rate $\beta(t)$ has to decrease. On March 6, 2020, the mean contact rate $\beta(t)$ was 0.688 (95% CI: 0.612 - 0.764) for an infectious period of C_a=3.25 days, 0.529 (95% CI: 0.464 - 0.593) for C_a= 6.5 days, and 0.463 (95% CI: 0.403 - 0.523) for C_a=13.0 days. By March 16, 2020, the day Santa Clara County announced the first county-wide shelter-in-place order in the entire United States, these contact rates $\beta(t)$ were 0.491 (95% CI: 0.462 - 0.523) for an infectious period of C_a=3.25 days, 0.328 (95% CI: 0.305 - 0.352) for C_a= 6.5 days, and 0.252 (95% CI: 0.234 - 0.271) for C_a=13.0 days.

3.2. *Effect of asymptomatic transmission of COVID-19 in Santa Clara County.*

Figure 4 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 $C_a = 3.25$ 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 $R(t) = C_s\beta(t)/[\,\nu_s + \nu_a C_s/C_a\,]$, as C_a increases, C_s/C_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

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

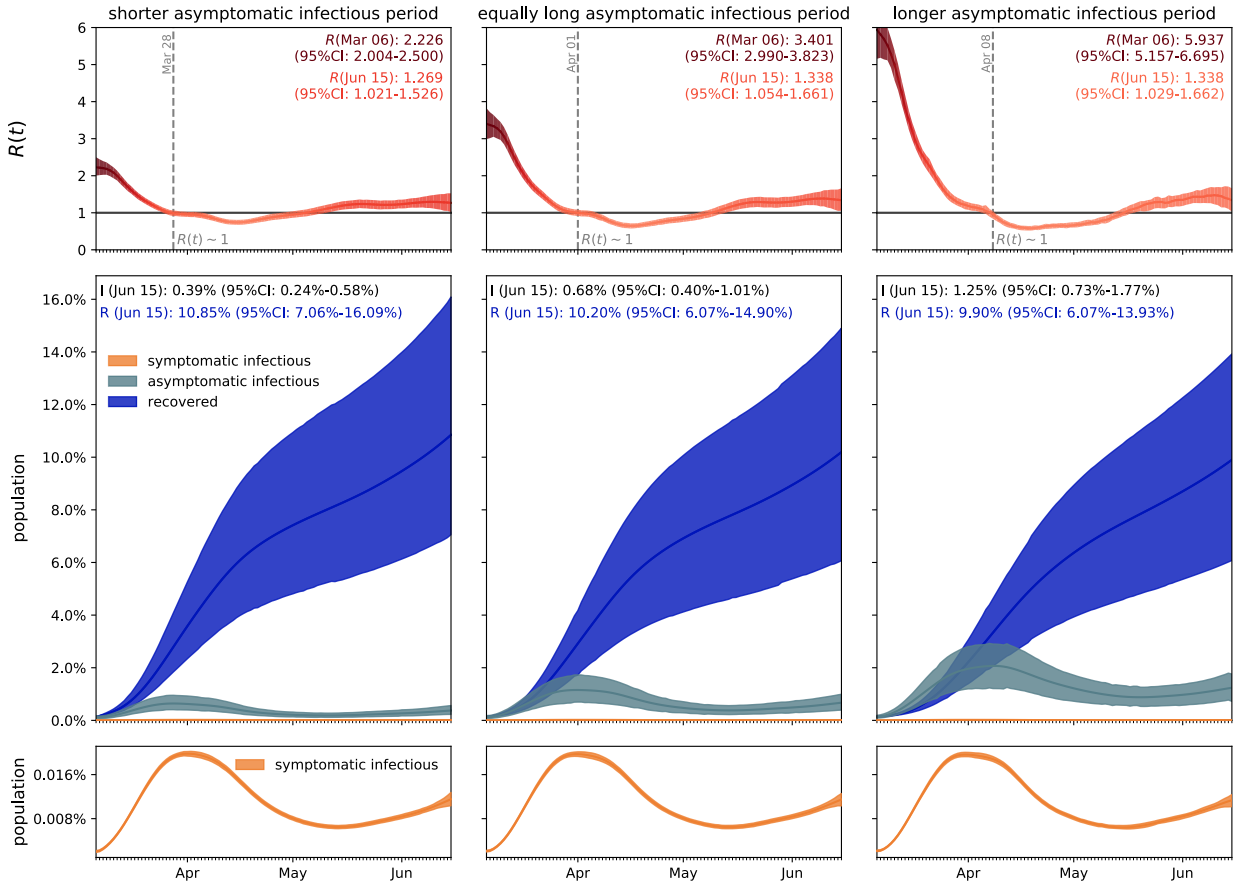

Figure 4: **Effect of asymptomatic transmission of COVID-19 in Santa Clara County.** The simulation learns the time-varying contact rate $\beta(t)$, and with it the time-varying effective reproduction number $R(t)$, for fixed latent and symptomatic infectious periods $A = 2.5$ days and $C_s = 6.5$ days, and for three asymptomatic infectious periods $C_a = 3.25$ 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 $\nu_s = I_s/I$, the initial exposed population E_0 and the initial infectious populations I_{s0} and I_{a0}.

single infectious individual infects less than one new individual, which indicates that the outbreak decays. The dashed vertical lines indicate the date $R(t) = 1$ 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, it took until March 28 before Santa Clara County managed to get $R(t)$ below 1 for the first time after the outbreak. For C_a=6.5 days, this only occurred by April 1 and for C_a=13.0 days, this occurred on April 8, 2020. This confirms our intuition that, the larger the asymptomatic infectious period C_a, for example because asymptomatic individuals will not 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 6, 2020 amounts to 0.70% (95% CI: 0.43%-0.97%) on March 28, 2020 for $C_a = 3.25$ days, 1.23% (95% CI: 0.72%-1.75%) on April 2, 2020 for $C_a = 6.5$ days, and 2.10% (95% CI: 1.25%-2.94%) on April 7, 2020 for $C_a = 13.0$ days. For larger asymptomatic infectious periods C_a, from left to right, the recovered population R decreases. Specifically, on June 15, 2020, the recovered population R amounts to 10.85%

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

(95% CI: 7.06%-16.09%) for an infectious period of C_a =3.25 days, 10.20% (95% CI: 6.07%-14.90%) for C_a =6.5 days, and 9.90% (95% CI: 6.07%-13.93%) for C_a =13.0 days. Similarly, and important when considering different exit strategies, the total infectious population, $I = I_s + I_a$, on June 15, 2020 is estimated to 0.39% (95% CI: 0.24%-0.58%) for C_a =3.25 days, 0.68% (95% CI: 0.40%-1.01%) for C_a =6.5 days, and 1.25% (95% CI: 0.73%-1.77%) for C_a =13.0 days.

3.3. Outbreak dynamics of COVID-19 worldwide.

Figure 5 illustrates the hierarchical asymptomatic infectious period C_a estimation for nine locations that reported COVID-19 antibody prevalence in a representative sample of the population: Heinsberg (NRW, Germany), Ada County (ID, USA), New York City (NY, USA), Santa Clara County (CA, USA), Denmark, Geneva Canton (Switzerland), Netherlands, Rio Grande do Sul (Brazil) and Belgium. This estimate assumes an initial fixed reproduction number R_0 = 3.87 (95%CI: 3.01-4.66) during a two-week window before lockdown which results in an asymptomatic infectious period of C_a = 5.76 (95%CI: 3.59-8.09) days. With this value, we simulate the outbreak dynamics of COVID-19 in all nine locations. Figure 6 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 nine locations. Here, we assume fixed latent and infectious periods of A = 2.5 days, C_s = 6.5 days, and the hierarchical asymptomatic infectious period C_a = 5.76 (95%CI: 3.59-8.09) days from Figure 5. For all nine locations, the calculated metrics display the fine balance between the dynamics of the effective reproduction number and the control of the epidemic outbreak. The downward trend of the effective reproductive number $R(t)$ quantifies how fast each location managed to control the spreading of COVID-19. The dashed vertical line indicates the first time each location managed to lower the effective reproduction below $R(t)$ = 1 after lockdown. This critical transition occurred on March 20 for Heinsberg, April 4 for Ada County, April 28 for New York City, March 30 for Santa Clara County, April 8 for Denmark, April 2 for Geneva, April 14 for the Netherlands, June 8 for Rio Grande do Sul and April 11 for Belgium. Based on our simulations, the maximum infectious population size amounted to 3.54% (95% CI: 2.96%-4.11%) on March 19 in Heinsberg, 0.47% (95% CI: 0.41%-0.55%) on May 5 in Ada County, 6.11% (95% CI: 5.99%-6.21%) on April 13 in New York City, 1.11% (95% CI: 0.69%-1.65%) on March 29 in Santa Clara County, 0.38% (95% CI: 0.27%-0.50%)

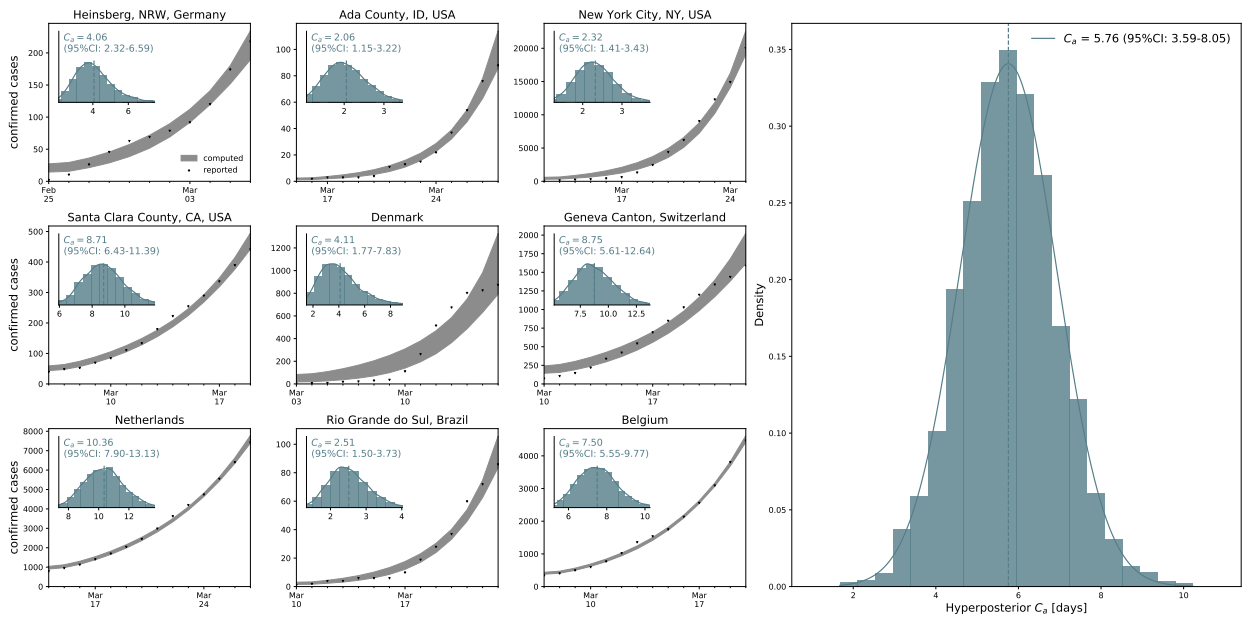

Figure 5: **Hierarchical asymptomatic infectious period C_a estimation.** Assuming an initial fixed reproduction number R_0 = 3.87 (95%CI: 3.01-4.66) [16], the simulation generates histograms of the asymptomatic infectious period C_a for each location based on the location-specific symptomatic fraction ν_s. The black dots and grey regions represent the reported and simulated detected cases $\hat{D}(t)$ and $D(t)$ respectively. The hierarchical hyperdistribution for the asymptomatic infectious period results in C_a = 5.76 (95%CI: 3.59-8.09) days, right histogram.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

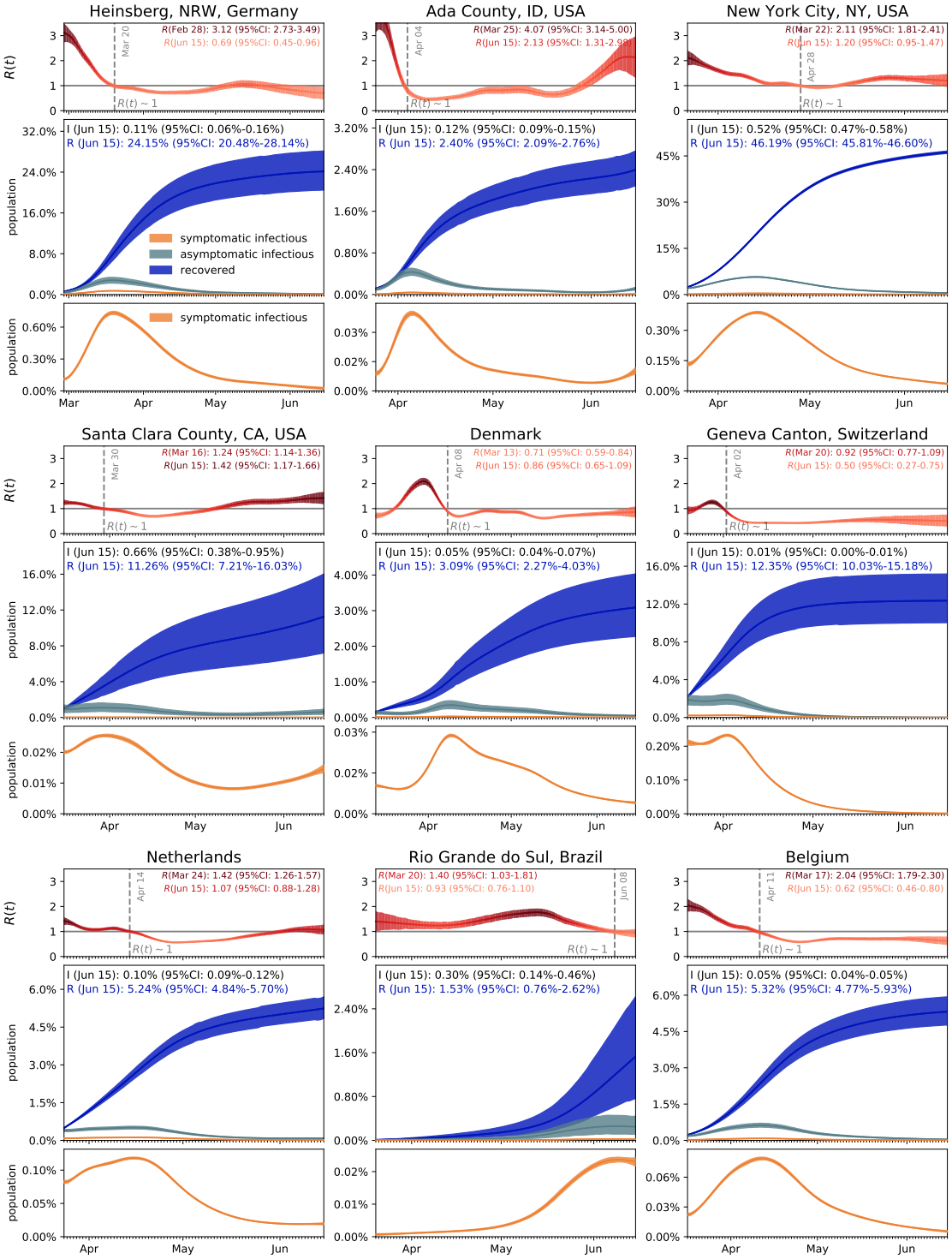

Figure 6: **Outbreak dynamics of COVID-19 worldwide.** Dynamic effective reproduction number R(t) and symptomatic, asymptomatic, and recovered populations at all nine locations. The simulation learns the time-varying contact rate $\beta(t)$, and with it the time-varying effective reproduction number $R(t)$, to predict the symptomatic infectious, asymptomatic infectious, and recovered populations I_s, I_a, and R, for fixed latent and infectious periods $A = 2.5$ days, $C_s = 6.5$ days, the hierarchical asymptomatic infectious period $C_a = 5.76$ (95%CI: 3.59-8.09) days from Figure 5. The dashed vertical lines mark the the first time each location managed to lower the effective reproduction below $R(t) = 1$ after lockdown. The colored regions highlight the 95% credible interval for the effective reproductive number $R(t)$ (top), the symptomatic and asymptomatic populations I_s and I_a, and the recovered population R (bottom plots), for uncertainties in the number of confirmed cases D, the fraction of the symptomatic infectious population ν_s, the initial exposed population E_0, and the initial infectious populations I_{s0} and I_{a0}.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

on April 9 in Denmark, 2.10% (95% CI: 1.65%-2.69%) on April 2 in Geneva, 0.63% (95% CI: 0.57%-0.68%) on April 15 in the Netherlands, 0.28% (95% CI: 0.14%-0.48%) on June 8 in Rio Grande do Sul, and 0.70% (95% CI: 0.62%-0.78%) on April 11 in Belgium. On Jun 15, 2020, the estimated recovered population reached 24.15% (95% CI: 20.48%-28.14%) in Heinsberg, NRW, Germany 2.40% (95% CI: 2.09%-2.76%) in Ada County, ID, USA 46.19% (95% CI: 45.81%-46.60%) in New York City, NY, USA 11.26% (95% CI: 7.21%-16.03%) in Santa Clara County, CA, USA 3.09% (95% CI: 2.27%-4.03%) in Denmark 12.35% (95% CI: 10.03%-15.18%) in Geneva Canton, Switzerland 5.24% (95% CI: 4.84%-5.70%) in Netherlands 1.53% (95% CI: 0.76%-2.62%) in Rio Grande do Sul, Brazil 5.32% (95% CI: 4.77%-5.93%) in Belgium.

3.4. Estimating the outbreak date.

Figure 7 shows the estimated outbreak date of COVID-19 in Santa Clara County. For fixed latent and symptomatic infectious periods $A = 2.5$ days and $C_s = 6.5$ days, and for the hierarchically estimated asymptomatic infectious periods $C_a = 5.76$ (95%CI: 3.59-8.09) days, the graph highlights the estimated date of the first COVID-19 case in the county. Based on the reported case data from March 16, 2020 onward, and taking into account uncertainty on the fraction of the symptomatic infectious population ν_s, on the initial exposed population E_0, and on the initial symptomatic and asymptomatic infectious populations I_{s0} and I_{a0}, we systematically backtracked the date of the first undetected infectious individual. Our results suggest that the first case of COVID-19 in Santa Clara County dates back to January 20, 2020 (95% CI: December 29, 2019 - February 13, 2020).

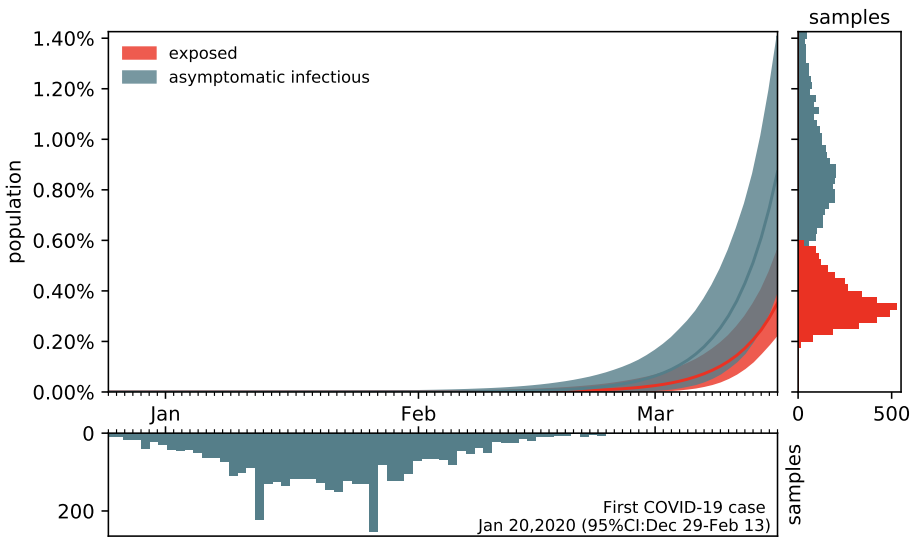

Figure 7: **Estimating the outbreak date of COVID-19 in Santa Clara County varying asymptomatic infectious periods C_a.** Estimated date of the first COVID-19 case in Santa Clara County for fixed latent and symptomatic infectious periods $A = 2.5$ days and $C_s = 6.5$ days, and for the hierarchical asymptomatic infectious period $C_a = 5.76$ (95%CI: 3.59-8.09) days from Figure 5. The colored regions in the main plot highlight the 95% credible interval for the time evolution of the exposed and asymptomatic infectious populations E and I_a estimated based on the reported cases $\hat{D}(t)$ from March 16, 2020 onward and taking into account uncertainties on the fraction of the symptomatic infectious population $\nu_S = I_S/I$, and the exposed and asymptomatic infectious populations E_0 and I_{a0} on March 16, 2020 (right plot). The bottom plot histogram shows the distribution of the most probable origin dates to January 20, 2020 (95% CI: December 29, 2019 - February 13, 2020).

4. 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,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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, $I = I_s + I_a$, 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. 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 [67], our study shows that the overall outbreak dynamics can be represented by the classical SEIR model [25] using equations (6). 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 $I = I_s/\nu_s$ and $R = R_s/\nu_s$ 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 $\nu_a = [1 - \nu_s]$, 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 [15]. 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 [27].

For different dynamics, the overall outbreak dynamics depend on both size and infectiousness 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 [42]. 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 $C_a < C_s$ and a lack of early isolation of the asymptomatic population through a larger infectious period $C_a > C_s$, while, for simplicity, keeping the latent period A and contact rate β similar across both groups [33]. Our study shows that the overall reproduction number, $R(t) = [C_a\, C_s]/[\nu_s\, C_a + \nu_a\, C_s]\, \beta(t)$, 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, we report the results for three different scenarios where the asymptomatic group is half as infectious, $C_a = 0.5\, C_s$, equally infectious, $C_a = C_s$, and twice as infectious $C_a = 2.0\, C_s$ as the symptomatic group for Santa Clara County. The second case, the middle column in Figures 3 and 4, corresponds to the special case with comparable dynamics and similar parameters. Our learnt asymptomatic infectious periods of $C_a = 5.76$ days in Figure 5 suggest that C_a is smaller than the symptomatic infectious period of $C_s = 6.5$ days and that the asymptomatic population is slightly less infectious as the symptomatic population. This can be a combined effect of less viral shedding as opposed to the symptomatic individuals, whilst concomitantly having more contacts because the asymptomatic individual does not realize he/she is spreading the disease.

Dynamic contact rates are a metric for the efficiency of public health interventions. 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 [25]. 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 [45]. To address this limitation, we introduce a time-dependent contact rate $\beta(t)$, which we learn dynamically from the reported case data. Figures 3 and 6 demonstrate 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 [7] or used sliding windows over the amount of novel reported infections [43] to motivate dynamic contact rates. As such, our framework provides a model-based method for statistical infer-

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

ence 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 (4) and (5) and our simulations in Figure 4 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 infectious period C_a of the asymptomatic group, the larger the initial effective reproduction number, $R(t) = C_s\beta(t)/[\nu_s + \nu_a C_s/C_a]$, 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 [34]. 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 4 and 6 demonstrate that we can successfully learn the critical time window until $R(t)$ drops below 1, which, in Santa Clara County, took till March 28 for C_a =3.25 days, till April 1 for C_a = 6.5 days and till April 8 for C_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 [42]. 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. 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 [27]. Two major challenges drive the interest in these studies: estimating the severity of the outbreak, e.g., hospitalization and mortality rates [15], and predicting the success of surveillance and control efforts, e.g., contact tracing or vaccination [18]. 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 4 suggests, these transmission dynamics can have a significant effect on the size of the asymptomatic population: For infectious periods of C_a = 0.5, 1.0, and 2.0 C_s, the maximum infectious population varies from 0.70% to 1.23% and 2.10%. 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 nine different locations at which COVID-19 antibody prevalence was measured in a representative sample of the population, Heinsberg (NRW, Germany) [59], Ada County (ID, USA) [4], New York City (NY, USA) [46], Santa Clara County (CA, USA) [3], Denmark [13], Geneva Canton (Switzerland) [60], Netherlands [55], Rio Grande do Sul (Brasil) [54], and Belgium [52]. The fraction of the symptomatic population in these nine locations is ν_s = 20.00%, 7.90%, 5.76%, 1.77%, 6.95%, 10.34%, 17.31%, 8.10%, and 10.21% respectively, broadly representing the range of reported symptomatic versus estimated total cases worldwide [3, 4, 8, 11, 13, 17, 23, 28, 30, 37, 46, 47, 52, 53, 56, 54, 55, 59, 60, 62, 64, 66, 68]. Of the nine locations we analyzed here, Heinsberg tested IgG and IgA, Ada County tested IgG, New York City tested IgG, Santa Clara County tested IgG and IgM, Denmark tested IgG and IgM, Geneva tested IgG, the Netherlands tested IgG, IgM and IgA, Rio Grande do Sul tested IgG and IgM, and Belgium tested IgG. While we did include reported uncertainty on the seroprevalence data, seroprevalence would likely have been higher if all 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 6 display remarkably similar trends: In most locations, the effective reproduction number $R(t)$ drops rapidly to values below one within a window of about three weeks after the lockdown date. However, the maximum infectious population, a value that is closely monitored by hospitals and health care systems, varies significantly ranging from 0.28% and 0.38% in Rio Grande do Sul and Denmark respectively to 3.54% and 6.11% in Heinsberg and New York City respectively. This is consistent with the reported 'superspreader' events in these last two locations. 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 [22]. While it is difficult to measure the effects of the unreported

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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 [34].

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 dates back to January 20 (Figure 7). This back-calculated early outbreak date is 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 after lockdown [51].

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 $\beta_s(t)$ and $\beta_a(t)$, we currently do not have data on the temporal evolution of the hidden asymptomatic infectious population $I_a(t)$ and longitudinal large population antibody studies would be needed to appropriately calibrate $\beta_a(t)$. Second, the ratio between the symptomatic and asymptomatic populations $\nu_s : \nu_a$ 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 $\nu_s(t)$ and $\nu_a(t)$ 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 increases in most 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, and predict an even earlier outbreak date. When longitudinal antibody studies become available, additional methodologies can be developed to correct for this limitation. Fourth, while we have included uncertainty in the seroprevalence data, the nine locations we analyzed here tested different types of antibodies and had different sampling procedures. Seroprevalence could have been higher if all 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.

5. 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 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

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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. A large asymptomatic population is associated with a high risk of community spread and could require a conscious shift from containment to mitigation induced by behavior changes. Our study suggests that, until vaccination and treatment become available, increasing population awareness, encouraging increased hygiene, mandating the use of face masks, restricting travel, and promoting physical distancing could be the most successful strategies to manage the impact of COVID-19 on both our economy and our health care system.

Acknowledgements

This work was supported by a Stanford Bio-X IIP Seed Grant (M.P. and E.K.), by a DAAD Fellowship (K.L.), and by the Stanford COVID-19 Seroprevalence Study Fund (J.B. and E.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: Jun 15, 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, M.H. Wener, S.L. Fink, C. Morishima, A. Chaudhary, K.R. Jerome, P.C. Mathias, A.L. Greninger Performance Characteristics of the Abbott Architect SARS-CoV-2 IgG Assay and Seroprevalence in Boise, Idaho J Clin Microbiol. 2020 doi:10.1128/JCM.00941-20

[5] Painel de casos de doença pelo coronavírus 2019 (COVID-19) no Brasil pelo Ministerio da Saude `www.covid.saude.gov.br/` assessed: Jun 15, 2020.

[6] 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: Jun 15, 2020.

[7] 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

[8] 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.

[9] 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.

[10] European Centre for Disease Prevention and Control. 2020. Situation update worldwide. `www.ecdc.europa.eu/en/geographical-distribution- 2019-ncov-cases` assessed: Jun 15, 2020.

[11] M. Emmenegger, E. De Cecco, D. Lamparter, R.P.B. Jacquat, D. Ebner, M.M. Schneider, I. Condado Morales, D. Schneider, B. Dogancay, J. Guo, A. Wiedmer, J. Domange, M. Imeri, R. Moos, C. Zografou, C. Trevisan, A. Gonzalez-Guerra, A. Carrella, I.L. Dubach, C.K. Xu, G. Meisl, V. Kosmoliaptsis, T. Malinauskas, N. Burgess-Brown, R. Owens, J. Mongkolsapaya, S. Hatch, G.R. Screaton, K. Schubert, J.D. Huck, F. Liu, F. Pojer, K. Lau, D. Hacker, E. Probst-Mueller, C. Cervia, J. Nilsson, O. Boyman, L. Saleh, K. Spanaus, A. von Eckardstein, D.J. Schaer, N. Ban, C. Tsai, J. Marino, G.F.X. Schertler, J. Gottschalk, B.M. Frey, R. Reimann, S. Hornemann, A.M. Ring, T.P.J. Knowles, I. Xenarios, D.I. Stuart, A. Aguzzi Population-wide evolution of SARS-CoV-2 immunity tracked by a ternary immunoassay. medRxiv 2020, doi: 10.1101/2020.05.31.20118554

[12] 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.

[13] C. Erikstrup, C.E. Hother, O.B.V Pedersen, K. Molbak, R.L. Skov, D.K. Holm, S. Saekmose, A.C. Nilsson, P.T. Brooks, J.K. 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.

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

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

[16] S. Flaxman, S. Mishra, A. Gandy, H. Unwin, H. Coupland, T. Mellan, H. Zhu, T. Berah, J. Eaton, P. Perez Guzman, N. Schmit, L. Cilloni, K. Ainslie, M. Baguelin, I. Blake, A. Boonyasiri, O. Boyd, L. Cattarino, C. Ciavarella, L. Cooper, Z. Cucunuba Perez, G. Cuomo-Dannenburg, A. Dighe, A. Djaafara, I. Dorigatti, S. Van Elsland, R. Fitzjohn, H. Fu, K. Gaythorpe, L. Geidelberg, N. Grassly, W. Green, T. Hallett, A. Hamlet, W. Hinsley, B. Jeffrey, D. Jorgensen, E. Knock, D. Laydon, G. Nedjati Gilani, P. Nouvellet, K. Parag, I. Siveroni, H. Thompson, R. Verity, E. Volz, C. Walters, H. Wang, Y. Wang, O. Watson, P. Winskill, X. Xi, C. Whittaker, P. Walker, A. Ghani, C. Donnelly, S. Riley, L.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Okell, M. Vollmer, N. Ferguson, S. Bhatt Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries Spiral Imperial College London 2020, doi: 10.25561/77731

[17] 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.

[18] 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.

[19] 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, doi:10.1073/pnas.2004978117.

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

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

[22] 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.

[23] H. Hallal, F. Hartwig, B. Horta, G.D. Victora, M. Silveira, C. Struchiner, L.P. Vidaletti, N. Neumann, L.S. Pellanda, O.A. Dellagostin, M.N. Burattini, A.M. Menezes, F.C. Barros, A.J. Barros, C.G. Victora Remarkable variability in SARS-CoV-2 antibodies across Brazilian 1 regions: nationwide serological household survey in 27 states. medRxiv 2020, doi: 10.1101/2020.05.30.20117531

[24] Heinsberg 2020. Aktuelles aus dem Kreishaus. Coronavirus im Kreis Heinsberg. `www.kreis-heinsberg.de/aktuelles`. assessed: Jun 15, 2020.

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

[26] 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.

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

[28] I. Jerkovic, T. Ljubic, Z. Basic, I. Kruzic, N. Kunac, J. Bezic, A. Vuko, A. Markotic, S. Andjelinovic SARS-CoV-2 antibody seroprevalence in industry workers in Split-Dalmatia and Sibenik-Knin County, Croatia. medRxiv 2020, doi: 10.1101/2020.05.11.20095158

[29] 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.

[30] V. Kraehling, M. Kern, S. Halwe, H. Mueller, C.P.C. Rohde, M. Savini, M. Schmidt, J. Wilhelm, S. Becker, S. Ciesek, R. Gottschalk Epidemiological study to detect active SARS-CoV-2 infections and seropositive persons in a selected cohort of employees in the Frankfurt am Main metropolitan area. medRxiv 2020, doi: 10.1101/2020.05.20.20107730

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] D. Ng, G. Goldgof, B. Shy, A. Levine, J. Balcerek, S.P. Bapat, J. Prostko, M. Rodgers, K. Coller, S. Pearce, S. Franz, L. Du, M. Stone, S. Pillai, A. Sotomayor-Gonzalez, V. Servellita, C. Sanchez-San Martin, A. Granados, D.R. Glasner, L.M. Han, K. Truong, N. Akagi, D.N. Nguyen, N. Neumann, D. Qazi, E. Hsu, W. Gu, Y.A. Santos, B. Custer, V. Green, P. Williamson, N.K. Hills, C.M. Lu, J.D. Whitman, S. Stramer, C. Wang, K. Reyes, J. Hakim, K. Sujishi, F. Alazzeh, L. Pharm, C. Oon, S. Miller, T. Kurtz, J. Hackett, G. Simmons, M.P. Busch, C.Y. Chiu SARS-CoV-2 seroprevalence and neutralizing activity in donor and patient blood from the San Francisco Bay Area. medRxiv 2020, doi: 10.1101/2020.05.19.20107482

[38] 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

[39] New York City Health Department. NYC Coronavirus Disease 2019 (COVID-19) Data www.github.com/nychealth/coronavirus-data, assessed: Jun 15, 2020.

[40] New York Times. An ongoing repository of data on coronavirus cases and deaths in the U.S. www.github.com/nytimes/covid-19-data, assessed: Jun 15, 2020.

[41] 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.

[42] 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.

[43] S.W. Park, K. Sun, C. Viboud, B.T. Grenfell, J. Dushoff. Potential roles of social distancing in mitigating the spread of coronavirus disease

medRxiv preprint doi: https://doi.org/10.1101/2020.05.23.20111419; this version posted June 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2019 (COVID-19) in South Korea. medRxiv 2020.03.27.20045815, doi 10.1101/2020.03.27.20045815

[44] 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.

[45] 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.

[46] J. Reifer, N. Hayum, B. Heszkel, I. Klagsbald, V.A. Streva SARS-CoV-2 IgG Antibody Responses in New York City medRxiv doi:10.1101/2020.2005.2023.20111427.

[47] E.S. Rosenberg, J.M. Tesoriero, E.M. Rosenthal, R. Chung, M.A. Barranco, L.M. Styer, M.M. Parker, S.J. Leung, J.M. Morne, D. Greene, D.R. Holtgrave, D. Hoefer, J. Kumar, T. Udo, B. Hutton, H.A. Zucker Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York. medRxiv 2020, doi: 10.1101/2020.05.25.20113050

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

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

[50] 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

[51] Santa Clara County COVID-19 Cases and Hospitalizations Dashboard. `www.sccgov.org` assessed: Jun 15, 2020.

[52] Sciensano: National Public Health Research Institute of Belgium Weekly Epidemiological Bulleting from May 29,2020 `https://covid-19.sciensano.be` assessed: Jun 15, 2020.

[53] 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.

[54] M. Silveira, A. Barros, B. Horta, L. Pellanda, G. Victora, 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. Hartwig, A.M. Menezes, F.C. Barros, P. Hallal, C.G. Victora Repeated population-based surveys of antibodies against SARS-CoV-2 in Southern Brazil medRxiv doi: 10.1101/2020.05.01.20087205

[55] 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.

[56] C.J. Snoeck, M. Vaillant, T. Abdelrahman, V.P. Satagopam, J.D. Turner, K. Beaumont, C.P.C. Gomes, J.V. Fritz, V.E. Schröder, A. Kaysen, L. Pavelka, L. Stute, G. Ramos Meyers, L. Pauly, M. Hansen, C. Pauly, G.A. Aguayo, M. Perquin, A. Hanff, S. Ghosh, M. Gantenbein, L. Huiart, M. Ollert, R. Krüger Prevalence of SARS-CoV-2 infection in the Luxembourgish population: the CON-VINCE study. medRxiv doi: 10.1101/2020.05.11.20092916

[57] 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.

[58] 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.

[59] 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.

[60] S. Stringhini, A. Wisniak, G. Piumatti, A.S. Azman, S.A. Lauer, H. Baysson, D. De Ridder, D. Petrovic, S. Schrempft, K. Marcus, 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 seroprevalence of anti-SARS-CoV-2 IgG antibodies in a population-based sample from Geneva, Switzerland medRxiv doi:10.1101/2020.05.02.20088898

[61] Swiss Federal Office of Public Health FOPH COVID19 case numbers Cantons of Switzerland and Principality of Liechtenstein. `www.github.com/openZH/covid19/`, assessed: Jun 15, 2020.

[62] M. Takita, T. Matsumura, K. Yamamoto, E. Yamashita, K. Hosoda, T. Hamaki, E. Kusumi, Regional Difference in Seroprevalence of SARS-CoV-2 in Tokyo: Results from the community point-of-care antibody testing. medRxiv 2020, doi: 10.1101/2020.06.03.20121020

[63] 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.

[64] 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.

[65] 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.

[66] L. Valenti, A. Bergna, S. Pelusi, F. Facciotti, A. Lai, M. Tarkowski, A. Berzuini, F. Caprioli, L. Santoro, G. Baselli, C. Della Ventura, E. Erba, S. Bosari, M. Galli, G. Zehender, D. Prati SARS-CoV-2 seroprevalence trends in healthy blood donors during the COVID-19 Milan outbreak. medRxiv 2020, doi: 10.1101/2020.05.11.20098442

[67] 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.

[68] 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.