

EML webinar overview:

Data-driven modeling of COVID-19—Lessons learned.

Ellen Kuhl

Department of Mechanical Engineering, Stanford University, Stanford, California, United States

Understanding the outbreak dynamics of COVID-19 through the lens of mathematical models is an elusive but significant goal. Within only half a year, the COVID-19 pandemic has resulted in more than 19 million reported cases across 188 countries with more than 700,000 deaths worldwide. Unlike any other disease in history, COVID-19 has generated an unprecedented volume of data, well documented, continuously updated, and broadly available to the general public. Yet, the precise role of mathematical modeling in providing quantitative insight into the COVID-19 pandemic remains a topic of ongoing debate. Here we discuss the lessons learned from six months of modeling COVID-19. We highlight the early success of classical models for infectious diseases and show why these models fail to predict the current outbreak dynamics of COVID-19. We illustrate how data-driven modeling can integrate classical epidemiology modeling and machine learning to infer critical disease parameters—in real time—from reported case data to make informed predictions and guide political decision making. We critically discuss questions that these models can and cannot answer and showcase controversial decisions around the early outbreak dynamics, outbreak control, and exit strategies. We anticipate that this summary will stimulate discussion within the modeling community and help provide guidelines for robust mathematical models to understand and manage the COVID-19 pandemic. EML webinar speakers, videos, and overviews are updated at <https://imechanica.org/node/24098>.

Keywords: COVID-19; data-driven modeling; Bayesian inference; epidemiology; extreme diffusion; extreme growth.

“The most astonishing thing about the pandemic was the complete mystery which surrounded it.” [The Lessons of the Pandemic, G.A. Soper, Science, 1919]

Motivation.

Would you take a measles vaccine to protect yourself against COVID-19? Most likely not. Why then, should we trust models that were initially designed for the measles to predict the dynamics of COVID-19?

In the early stages of the COVID-19 pandemic, under an enormous pressure to deliver results, the obvious solution seemed to be to recycle existing infectious disease models [Anderson & May, 1982] and adapt them to simulate the outbreak dynamics of COVID-19 [Cobey, 2020]. In retrospect, this was an obvious mistake. Many elements of the current pandemic—although similar at first sight—are inherently different from infectious diseases early in the early 20th century [Kermack & McKendrick, 1927]: We now have a longer life expectancy, we are more globally connected, and we travel more; but we also have better access to hygiene, to health care, and to massive amounts of data about the disease.

During the early onset of the outbreak, all eyes were on mathematical modeling with the general expectation that COVID-19 models could precisely predict the trajectory of the pandemic [Hsu, 2020]. Mathematical modeling rapidly became front and center to understanding the exponential increase of infections, the shortage of ventilators, and the limited capacity of hospital beds—too rapidly as we now know. Bold and catastrophic predictions not only initiated a massive press coverage, but also a broad anxiety in the general population. However, within only a few weeks, the vastly different predictions and conflicting conclusions began to create the impression that all mathematical models are in generally unreliable and inherently wrong [Holmdahl & Buckee, 2020]. While the failure of COVID-19 modeling—often by an order of magnitude or more—was devastating for policymakers and public health practitioners, initial mistakes are not new to the modeling community where an iterative cycle of prediction, failure, and redesign is common standard and best practice [Alber et al., 2019]. However, the successful use of mathematical models implies to set the expectations right [Siegenfeld et al., 2020]. Understanding what models can and cannot predict is critical to the Art of Modeling.

Two classes of models have been proposed to understand the outbreak dynamics of COVID-19, statistic and mechanistic models [Holmdahl & Buckee, 2020]. Purely statistic models use machine learning or regression to analyze massive amounts of data and project the number of infections into the future [Institute for Health and Metrics Evaluation, 2020]. Since purely statistic models do not include any disease specific information, their forecasts are reliable only within a short time window. Nonetheless, statistical modeling can be useful, for example, to understand how to allocate resources or make rapid short-term recommendations. Mechanistic models simulate the outbreak through interacting disease mechanisms by using local nonlinear population dynamics and global mixing of populations [Hethcote, 2020]. These models can include disease specific information and potentially make long-term predictions about the outcome of a pandemic. Mechanistic modeling can be useful to explore how the pandemic would change under various assumptions and political interventions. When selecting between statistic and mechanistic models, it is critical to know upfront which questions the model should address [Siegenfeld et al., 2020].

Two interacting features determine the outbreak dynamics of the COVID-19 pandemic: the local epidemiology of the disease and the global mobility of affected individuals [Linka, 2020]. For coronavirus diseases, the local epidemiology is defined by an exponential growth of the outbreak,

where the number of new cases depends exponentially on the growth rate [Cobey, 2020]. From an outbreak dynamics perspective, this implies that the outbreak behaves like a chaotic system for which even small inaccuracies in the prediction can trigger large changes in the number of cases. From an outbreak control perspective, small changes in intervention can alter the current growth rate and convert the dynamics from exponential growth to exponential decay or vice versa [Siegenfeld et al., 2020]. To understand the vulnerability of the model to these small changes, especially in view of the varying reporting practices of the COVID-19 case data, sensitivity analysis and quantifying uncertainty have become critical elements of robust predictive modeling [Jha et al., 2020, Peirlinck et al., 2020a]. A promising technology that integrates statistic and mechanistic approaches and can inherently quantify model uncertainties is data-driven modeling. Throughout the past months, several research groups have started to integrate classical epidemiology models and machine learning to infer critical disease parameters from reported case data and make informed predictions about outbreak dynamics and outbreak control [Jha et al., 2020, Kergssner et al., 2020, Linka et al., 2020, Peirlinck et al., 2020, Viguerie et al., 2020, Wang et al., 2020]. In the reminder of this work, we share the lessons we have learned throughout this process, from the early beginning of the COVID-19 pandemic until the current concerns and challenges in an attempt to safely reopen from lockdown.

Modeling the early outbreak dynamics.

Lesson 1. COVID-19 is spreads exponentially if uncontrolled.

During the early stages of the COVID-19 outbreak, the world stood in awe to see the number of new infections climbing explosively [Johns Hopkins University, 2020]. This rapid increase created a lot of anxiety within both the general population and political decision makers. It seemed natural to turn to mathematical models to understand the rapid spreading of COVID-19 and estimate its consequences [Hsu, 2020]. In fact, the first mathematical models for infectious diseases date back to a smallpox model in the middle of the 18th century [Bernoulli 1760]. Since the 1920s, compartment models have advanced to the method of choice to simulate the epidemiology of infectious diseases [Kermack & McKendrick, 1927]. One of the simplest compartment models is the SEIR model. It represents the timeline of a disease through four compartments, the susceptible, exposed, infectious, and recovered populations [Hethcote, 2000]. The temporal evolution of these compartments is governed by four ordinary differential equations parameterized in terms of the three transition rates between them.

$$\begin{aligned}\dot{S} &= -\beta S I \\ \dot{E} &= +\beta S I - \alpha E \\ \dot{I} &= +\alpha E - \gamma I \\ \dot{R} &= +\gamma I\end{aligned}$$

With only three parameters, α , β , and γ , the classical SEIR model has been successfully used to model previous epidemic outbreaks. The parameters α and γ characterize the transition from the exposed to the infectious state and from the infectious to the recovered state. In fact, they are the inverses of the latent period $A = 1 / \alpha$, the time during which an individual is exposed but not yet infectious, and the infectious period $C = 1 / \gamma$, the time during which an individual can infect others [Hethcote, 2000]. As such, they are disease specific parameters independent of region, state, or country. For COVID-19, in

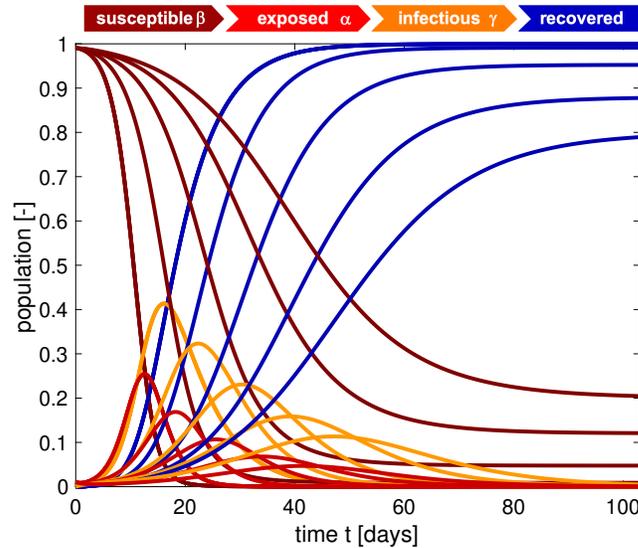


Figure 1. Outbreak dynamics of COVID-19 for varying basic reproduction numbers R_0 . Decreasing the basic reproduction number, from left to right, decreases the exposed and infectious populations, red and orange curves. The susceptible and recovered populations, brown and blue curves, converge to larger and smaller endemic equilibrium values, and converges is slower. The steepest curves correspond to the smallest contact period $B = 0.8$ days and largest basic reproduction number $R_0 = 8.0$ with the maximum infectious population of $I_{max} = 0.412$ after 15 days. Latent period $A = 2.5$ days, infectious period $C = 6.5$ days, initial exposed fraction $E_0 = 0.01$, and contact period $B = 3.3, 2.7, 2.0, 1.4, 0.8$ days, resulting in basic reproduction number $R_0 = C/B = 2.0, 2.4, 3.2, 4.8, 8.0$.

the example of Figure 1, the latent and infectious periods are on the order of $A = 2.5$ days and $C = 6.5$ days [Linka et al., 2020a]. The defining feature of the SEIR model is its nonlinear feedback loop that defines transition from the susceptible to the exposed state. The model typically assumes that this transition scales with the susceptible population S , the infectious population I , and the contact β , the inverse of the contact period $B = 1 / \beta$, between the two. Figure 1 illustrates the dynamics of the susceptible, exposed, infectious, and recovered populations, for varying contact periods of $B = 3.3, 2.7, 2.0, 1.4, 0.8$ days. During the early stages of the COVID-19 pandemic, many research groups have successfully applied the SEIR model to simulate the early period of free exponential growth and estimated the ratio between the infectious period and the contact period to $C/B = 4.8 - 8.0$, associated with the two left-most sets of curves [Liu et al., 2020]. Figure 1 suggests that, for this range of parameters, the outbreak would have peaked after only two weeks, with 30-40% of the population infected, and it would have been over after less than two months. Obviously, this is not what happened [Johns Hopkins University, 2020]. How could a model, that had successfully simulated the measles and chickenpox fail so dramatically in predicting the timeline of COVID-19, despite its initial promise in modeling the early outbreak?

Lesson 2. COVID-19 is as contagious as previous coronaviruses.

An important question to ask, especially during the early stages of the outbreak, is: *How contagious is the new coronavirus?* This closely relates to the question *How does it compare to other coronaviruses or, even more broadly, to other infectious diseases?* A powerful quantitative concept to characterize the contagiousness and transmissibility of the new coronavirus is the basic reproduction number R_0

[Viceconte & Petrosillo, 2020]. This number explains—in simple terms—how many new infections are caused by a single one infectious individual in an otherwise completely susceptible population [Dietz, 1993]. Since the beginning of the new coronavirus pandemic in December 2020, no other number has been discussed more controversially than the reproduction number of COVID-19 [Liu, 2020]. However, it is difficult—if not impossible—to measure R_0 directly. Most basic reproduction numbers of COVID-19 we see in the public media today are estimates of mathematical models that depend critically on the choice of the model, the initial conditions, and numerous other modeling assumptions [Delamater, 2019]. For our SEIR model, the basic reproduction number is simply the product of the contact rate β and the infectious period C , or the ratio between the infectious period C and the contact period B ,

$$R_0 = \beta C = C/B.$$

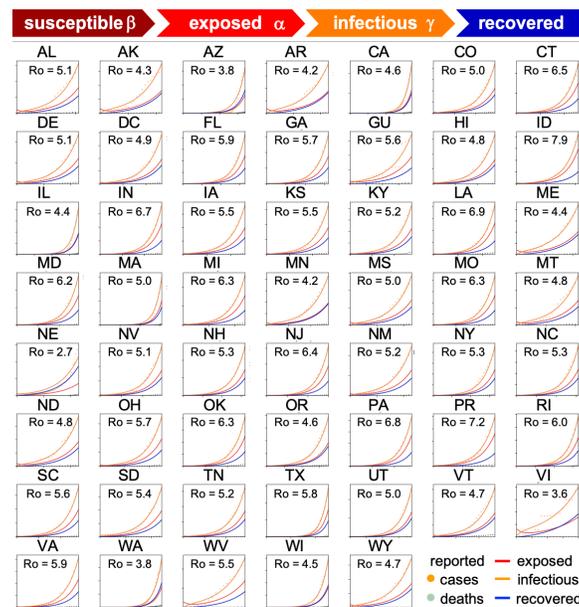


Figure 2. Early outbreak dynamics of COVID-19 in the United States. Reported infectious populations and simulated exposed, infectious, and recovered populations. Simulations are based on a state-specific identification of the contact period BB to define the basic reproduction number $R_0 = C / B$. The mean basic reproduction number was $R_0 = 5.30 \pm 0.95$ [Peirlinck et al., 2020].

Figure 2 shows how we can use reported case data to infer the contact period B and with it the basic reproduction number $R_0 = C / B$ across all United States during the early stages of exponential growth [Peirlinck et al., 2020]. The resulting mean basic reproduction number of 5.30 ± 0.95 for the United States is slightly higher than the mean basic reproduction number of 4.22 ± 1.69 for Europe [Linka et al., 2020a]. Both values agree well with the reported value of 5.7 for the Wuhan outbreak [Sanche et al., 2020] and with a recent review that suggests values from 4.1 to 6.5 from SEIR modeling [Liu et al., 2020]. Compared to traditional infectious diseases, these basic reproduction numbers are lower than the numbers of 18 for measles, 9 for chickenpox, 7 for mumps, and 7 for rubella, but on the order of 5 for poliomyelitis [Anderson & May, 1982]. Compared to the SARS coronavirus with a range from 2 to 5 [Liu et al., 2020], our values for SARS-CoV-2 in Figure 2 are on the higher end, and suggest that the new coronavirus would spread more rapidly than SARS [Wilder et al., 2020]. Knowing the basic reproduction number of COVID-19 is critical to estimate the conditions for herd immunity and predict the success of

vaccination strategies. However, from Figure 1 we would conclude that, for this range of basic reproduction numbers, from $R_0 = 2.0$ to 8.0 , 80-100% of the population would have been infected with the virus within only three months. How useful is the concept of the basic reproduction number if it fails to accurately reproduce the timeline of COVID-19, even in the first few weeks of the outbreak?

Lesson 3: Without vaccination, COVID-19 will be with us for a long time.

The million-dollar question—literally—is: *How long will the COVID-19 pandemic last?* From other infectious diseases including the measles, chickenpox, mumps, polio, rubella, pertussis, and smallpox we know that epidemic outbreaks tend to come to an end before the entire population has been infected [Anderson & May, 1982]. For this class of diseases, the basic reproduction number is larger than one, $R_0 > 1.0$, and an infected individual will initially infect more than one other individual. Figure 1 shows that, under these conditions, the infectious population first increases, then reaches a peak, and decreases toward zero [Kermack & McKendrick, 1927]. As more and more individuals transition from the susceptible through the exposed and infectious states into the recovered state, the susceptible population decreases. Once a large enough fraction of a population has become immune—either through recovery from the infection or through vaccination—this group provides a protection for the susceptible population. The epidemic dies out as the rate of daily new cases, $\beta S I$, decreases [Dietz, 1993]. As such, the classical SEIR model is self-regulating: It naturally converges to an endemic equilibrium, at which either the susceptible group S , or the infectious group I , or both have become small enough to prevent new infections. In epidemiology, this indirect protection is called herd immunity [Fine, 1993]. The concept of herd immunity implies that the converged susceptible population at endemic equilibrium is always larger than zero, $S_\infty > 0$, and its value depends on the basic reproduction number R_0 . In a homogeneous, well-mixed population, herd immunity occurs once a fraction of $[1 - 1/R_0]$ of the population has become immune, either through the disease itself or through vaccination. During the very early stages of the outbreak, political decision makers were actively focusing on answering the question: *When we reach herd immunity?* For the mean basic reproduction numbers of $R_0 = 4.22$ and $R_0 = 5.30$ we found for Europe and the United States in Figure 1, the herd immunity threshold would lie between 76% and 81%. This value is lower than 94% for the measles, 89% for chickenpox with, 86% for mumps and rubella, and on the order of 80% for polio, but significantly higher than the values of 16% to 27% for the seasonal flu [Anderson & May, 1982]. Even the countries with the highest prevalence, Chile with 1.75%, the United States with 1.22%, Brazil with 1.02%, and Sweden with 0.77% [Johns Hopkins University, 2020] do not come close to these values today. Against all initial enthusiasm for reaching herd immunity by the fall, we have now come to realize that we will likely not pass the immunity level threshold any time soon by infection alone. In the near future, vaccination remains the only viable strategy to achieve herd immunity. The basic reproduction number of COVID-19 suggests that at least three fourth of the population, more than 5 billion people worldwide, would have to be vaccinated to achieve herd immunity against COVID-19. While this sounds like a gigantic undertaking, massive vaccination campaigns have successfully controlled deadly contagious diseases such as polio, diphtheria, and rubella and even successfully eradicated smallpox in 1980 [Anderson & May, 1982]. However, until a COVID-19 vaccine is developed and approved, it is crucial to slow the spread of the COVID-19 virus and protect individuals at increased risk of severe illness, including older adults and people of any age with underlying health conditions [Fauci et al., 2020].

Modeling outbreak control.

Lesson 4. We can flatten the curve.

During the early stages of exponential growth, with new case numbers doubling within two or three days, the most urgent question amongst health care providers and political decision makers began to ask was: *Can we reduce the reproduction number R ?* For the broad population, this question became famously and illustratively rephrased as: *Can we flatten the curve?* For the modeling community, this quest for a lower reproduction number all of a sudden meant that the traditional SEIR epidemiology models were no longer suitable to model changes in disease dynamics. While traditional models with static parameters were well-suited to model the outbreak dynamics of unconstrained, freely evolving infectious diseases with fixed basic reproduction numbers in the early 20th century [Hethcote, 2000], they fail capture how behavioral changes and political interventions can modulate the reproduction number to manage the COVID-19 pandemic in the 21st century [Linka et al., 2020a]. *In fact, static reproduction numbers are probably the single most common cause of model failure in COVID-19 modeling.* To simulate a flattening, or rather early bending of the case curve, early modeling approaches adopted an ad hoc strategy that explicitly reduced the total population N to a potentially affected population $N^* = \eta N$. This strategy introduces a scaling coefficient $\eta = N^* / N$, essentially a mere fitting parameter, to indirectly quantify the level of confinement [Arenas et al., 2020]. Averaged over 30 Chinese provinces, the level of confinement was $\eta = 5.19 \cdot 10^{-5} \pm 2.23 \cdot 10^{-4}$ [Peirlinck et al., 2020] and averaged over the 27 countries of the European Union, it was $\eta = 7.67 \cdot 10^{-2} \pm 2.61 \cdot 10^{-1}$ [Linka et al., 2020] suggesting that the effect of COVID-19 was successfully confined to only a very small fraction of the total population. More mechanistic approaches are based on introducing a time-varying effective reproduction number $R(t)$ and on learning it dynamically from the reported case data. For example, we can infer discrete time points at which the contact rates vary [Dehning et al., 2020] or used sliding windows over the amount of novel reported infections [Park et al., 2020]. Strikingly, in many countries, the reported COVID-19 cases data follow a similar characteristic S-shaped pattern in response to political interventions. We can approximate this behavior with an effective reproduction number of hyperbolic tangent type [Linka et al., 2020a],

$$R(t) = R_0 - \frac{1}{2} [1 + \tanh([t - t^*]/T)] [R_0 - R_t] .$$

This ansatz has four physically meaningful parameters that provide valuable information about the responsiveness of the reproduction number to political action: the basic reproduction number R_0 at the beginning of the outbreak, the reduced reproduction number R_t under political interventions, the adaptation time t^* , and the transition time T . We can infer these parameters from the reported case data using Bayesian inference with Markov-Chain Monte-Carlo [Linka et al., 2020a]. This allows us to model an S-shaped case curve that plateaus long before a large fraction of the population has been affected by the disease. Another alternative that allows even more flexibility and does not require $R(t)$ to be monotonic is a Gaussian random walk function [Peirlinck et al., 2020a]. This free form approach naturally captures the effects of public health interventions, however, in a daily varying, rather unpredictable way. The random walk approach is a flexible method to analyze case data retrospectively, but since it does not allow for a closed functional form, it is not very useful to make informed predictions. Taken together, from the failure of traditional static SEIR models, we have learned that we need to introduce dynamic time-varying model parameters if we want to correctly

model behavioral and political changes and reproduce the reported case numbers. This naturally introduces a lot of freedom, a large number of unknowns, and a high level of uncertainty. However, in stark contrast to epidemic outbreaks in the early 20th century, we now have well-documented case data and the appropriate tools to address this challenge [Alber et al., 2019]. The massive amount of COVID-19 case data, all freely available on public data bases [Johns Hopkins University, 2020], has induced a clear paradigm shift from traditional mathematical epidemiology [Hethcote, 2000] towards data-driven physics-based modeling of infectious disease. These new techniques naturally learn the most probable model parameters—in real time—from the time evolution of continuously updated case data, allow us to make projections into the future, and quantify the uncertainty on the estimated predictions.

Lesson 5. Constraining mobility is a drastic but effective mitigation strategy.

Reducing mobility is a controversial but highly effective measure to manage a global pandemic [Zlojutro et al., 2019]. On March 17, 2020, for the first time in history, the European Union closed all its external borders to reduce the spreading of COVID-19 [European Commission, 2020]. In the following two

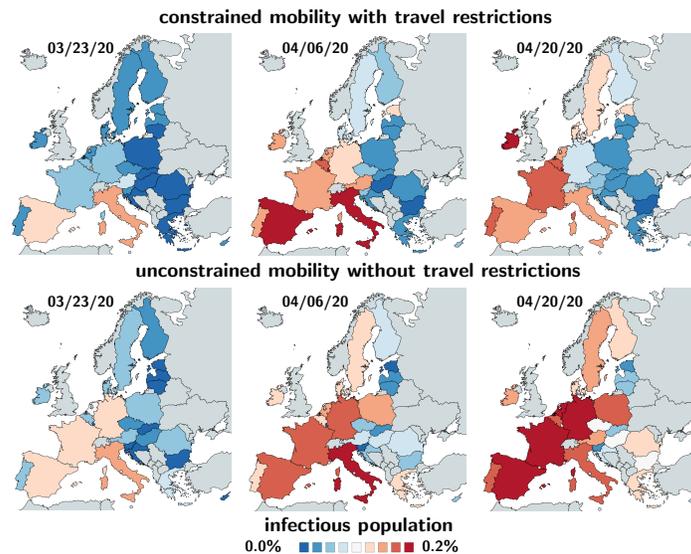


Figure 3. Early outbreak control of COVID-19 in Europe. Infectious population for constrained mobility with travel restrictions versus unconstrained mobility without travel restrictions. Simulations are based on country-specific basic reproduction numbers without and with mobility. The mean basic reproduction number was 4.62 ± 1.32 [Linka et al., 2020].

weeks, the local governments augmented the European regulations with local lockdowns and national travel restrictions. These measures had a dramatic effect on the mobility within the European Union: Within five days, the average passenger air travel in Europe was cut in half, and within two weeks, it was reduced to 5-10% [Eurostat, 2020]. These drastic actions have triggered ongoing debates about the effectiveness of different outbreak strategies and the appropriate level of constraints [Mason Meier et al., 2020]. A simple way to probe the effect of mobility is to model the spreading of COVID-19 through a mobility network of passenger air travel. For the European Union, we can represent this network as a weighted graph G in which the $N = 27$ nodes represent the individual countries and the weighted E edges represent the travel frequency between them. We can estimate the travel frequency

within the graph using passenger air travel statistics before and during the outbreak [Eurostat, 2020]. We record this information in the adjacency matrix A_{IJ} that represents the travel frequency between two countries I and J , and in the degree matrix, $D_{II} = \text{diag} \sum_{J=1, J \neq I}^N A_{IJ}$, that represents the number of incoming and outgoing passengers for each country I . The difference between the degree matrix D_{IJ} and the adjacency matrix A_{IJ} defines the weighted graph Laplacian $L_{IJ} = D_{IJ} - A_{IJ}$. We can then discretize the SEIR model on the weighted graph G and introduce the susceptible, exposed, infectious, and recovered populations S_I, E_I, I_I , and R_I as global unknowns at the nodes of the graph G [Linka et al., 2020]. This results in a set of equations with $4N$ unknowns.

$$\begin{aligned}\dot{S}_I &= -\sum_{J=1}^N L_{IJ} S_J - \beta S_I I_I \\ \dot{E}_I &= -\sum_{J=1}^N L_{IJ} E_J + \beta S_I I_I - \alpha E_I \\ \dot{I}_I &= -\sum_{J=1}^N L_{IJ} I_J + \alpha E_I - \gamma I_I \\ \dot{R}_I &= -\sum_{J=1}^N L_{IJ} R_J + \gamma I_I\end{aligned}$$

Figure 3 highlights the effect of constrained mobility in managing the COVID-19 outbreak across Europe. The top row shows the simulated outbreak under constrained mobility with the imposed travel restrictions and border control in place, the bottom row shows the outbreak under unconstrained mobility without travel restrictions. During the early stages of the pandemic, the predicted outbreak pattern in the bottom row agrees well with the outbreak pattern in the top row. During the later stages, the side-by-side comparison shows a faster spreading of the outbreak under unconstrained mobility with a massive, immediate outbreak in Central Europe and a faster spreading to the eastern and northern countries. Although air travel is certainly not the only determinant of the outbreak dynamics, Figure 3 suggests that mobility is a strong contributor to the global spreading of COVID-19 and supports the decision of the European Union and its local governments to implement rigorous travel restrictions to delay the outbreak of the pandemic [Linka, 2020b].

Lesson 6. Reproduction is correlated to mobility with a time delay of two weeks.

The drastic political measures, travel restrictions, and boarder control during the early stages of the pandemic have stimulated a wave of criticism [Mason Meier et al., 2020]. While it was initially entirely unclear to which extent they would succeed in reducing the number of new infections [Chinazzi et al., 2020], we now know that reducing mobility can effectively flatten the curve. In addition to global air mobility, several studies have proposed to explore correlations between outbreak control and driving, walking, and transit mobility from cell phone data [Apple, 2020]. A important question in outbreak control is: *What is the time delay between intervention and effect?*

Figure 4 summarizes the mobility data from the relative volume of location requests per country, scaled by the baseline volume before the outbreak of the pandemic [Linka, 2020a]. To smoothen the weekday-weekend fluctuations in outbreak and mobility data, we have applied a moving averaging window of seven days. In addition, Figure 4 shows the hyperbolic tangent approximation of the effective reproduction number $R(t)$ inferred from combining the SEIR model with the reported case data in all 27 countries. Interestingly, the drop in global passenger air travel and in local driving, walking, and transit mobility follows a similar hyperbolic tangent type form. For each country, we can extract the time delay Δt between the reduction of air traffic, driving, walking, and transit mobility and the inflec-

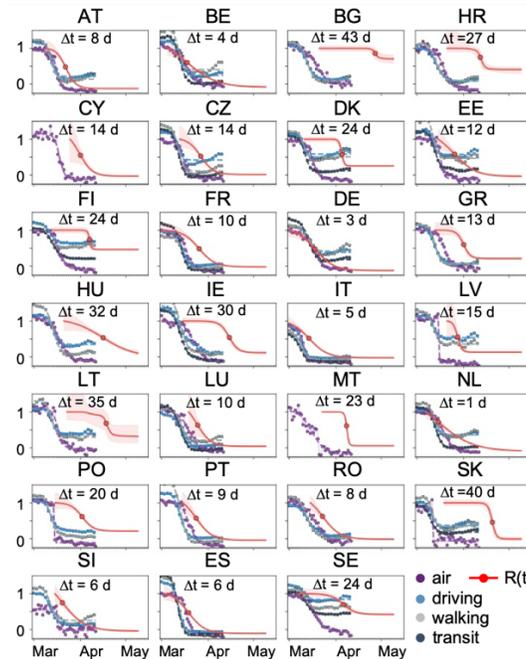


Figure 4. Correlation between reduction in mobility and effective reproduction number of COVID-19 outbreak in Europe. Purple, blue, grey, and black dots represent reduction in air traffic, driving, walking, and transit mobility; red curves show effective reproduction number $R(t)$ with 95% confidence interval. The time delay Δt denotes the delay between reduction in mobility and effective reproduction number. The mean time delay was 17.24 ± 2.00 days [Linka et al., 2020a].

tion point of the reproduction number curve. This time delay is an important socio-economical metric for the response time to political interventions. The population weighted mean time delay across the European Union is $\Delta t = 17.24 \pm 2.00$ days. The country-specific time delay varies hugely across Europe with the fastest response of 0.75 days in the Netherlands, followed by Germany with 3.25 days, Belgium with 4.00 days, and Italy with 5.00 days. These fast response times naturally also reflect early decisions on the national level. For example, Figure 4 clearly showcases the special role of Sweden, where the government focusses efforts on encouraging the right behavior and creating social norms rather than mandatory restrictions: The time delay of 23.75 days is above the European Union average of 17.24 days, and Sweden is one of the few countries where the effective reproduction number has not notably decreased below one. Taken together, from correlating reproduction and mobility, we have learned that, especially during the early stages of an outbreak, controlling mobility can play a critical role in reducing the spread of the pandemic [Linka et al., 2020c]. The time delay between mobility control and reduced reproduction is particularly important to plan exit strategies and estimate risks associated with gradually or radically relaxing local lockdowns and global travel restrictions.

Lesson 7. Most COVID-19 cases are asymptomatic and remain unreported.

In the early stages of the COVID-19 pandemic, doctors, researchers, and political decision makers have mainly focused on symptomatic individuals that came for testing and required urgent medical attention. In the more advanced stages, the interest has shifted towards mildly symptomatic and asymptomatic individuals who, by definition, are difficult to trace and likely to retain normal social and

travel patterns [Li et al., 2020]. As of today, more than 50 studies have reported an asymptomatic population, 23 of them with a sample size of 500 and more [Ioannidis, 2020]. The reported trends are strikingly consistent: A much larger number of individuals display antibody prevalence than we would expect from the reported symptomatic case numbers. In fact, the median undercount across all studies suggests that only one in twenty COVID-19 cases has been noticed and reported [Peirlinck, 2020a].

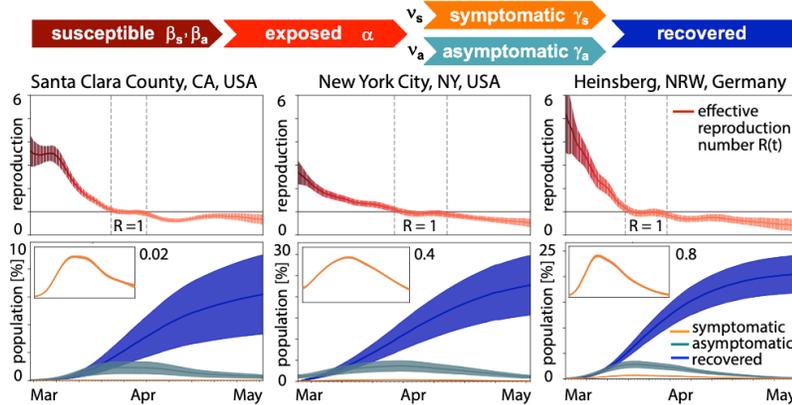


Figure 5. Outbreak dynamics of COVID-19 in Santa Clara County, New York City, and Heinsberg. Dynamic effective reproduction number $R(t)$ and symptomatic, asymptomatic, and recovered populations at three different locations where antibody prevalence studies were performed. The model learns the time-varying effective reproduction number $R(t)$ to predict the symptomatic, asymptomatic, and recovered populations with 95% credible intervals [Peirlinck et al., 2020a].

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. We can model the effects of asymptomatic transmission by extending the classical SEIR model into and SEIIR model with five compartments, the susceptible, exposed, symptomatic infections, asymptomatic infections, and recovered groups [Peirlinck et al., 2020a].

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

Here I_s and I_a denote the symptomatic and asymptomatic groups, which are fractions ν_s and ν_a of the total infectious group I . We postulate that both groups have the same latency rate α , but they can have their own contact rates β_s and β_a and infectious rates γ_s and γ_a . Figure 5 illustrates the resulting model and a representative analysis for three locations where antibody seroprevalence was reported, Santa Clara County with $\nu_s = 1.77\%$ [Bendavid et al., 2020], New York City with $\nu_s = 5.76\%$ [Reifer et al., 2020], and Heinsberg with $\nu_s = 20.00\%$ [Streeck et al., 2020]. The model combines the SEIIR model and Bayesian inference to learn the time-varying effective reproduction number $R(t)$ from the reported

case data and predicts the symptomatic, asymptomatic, and recovered populations with 95% credible intervals [Peirlinck et al., 2020a]. Strikingly, despite notable differences in seroprevalence, the effective reproduction numbers $R(t)$ and the infectious and recovered populations I_s , I_a , and R in Figure 5 follow similar trends: The effective reproduction number $R(t)$ drops rapidly to values below one within a window of about three weeks after the lockdown date, the infectious curves peak, and the recovered curve begins to plateau. Including asymptomatic transmission in the model also allows us to back-calculate the undercount from comparing mortality rates and death counts [Kergassner et al., 2020]. Knowing the exact dimension of the asymptomatic population is critical to truly estimate the severity of the outbreak, e.g., hospitalization or mortality rates [Fauci et al., 2020], and to reliably predict the success of surveillance and control strategies, e.g., contact tracing or vaccination [Fraser, 2004]. Precise knowledge about the asymptomatic population can significantly change our understanding and management of the COVID-19 pandemic: A large asymptomatic population will bring us closer to herd immunity, but will also make isolation, containment, and tracing of individual cases more challenging. Instead, managing community transmission through increasing population awareness, promoting physical distancing, encouraging behavioral changes and massive testing would become more relevant.

Lesson 8. There are massive amounts of data, but they don't always align well with the models.

Within only six months, the COVID-19 pandemic has probably generated more data than any disease in history. New symptomatic cases, recovered cases, and deaths are well documented and publicly shared on numerous websites and reports [Johns Hopkins University, 2020]. Intuitively, we would think that this presents endless opportunities for modeling [Alber et al., 2019]. However, unfortunately, not all the available data align well with the input needed for epidemiology modeling. For example, one question we still cannot address is: *When exactly did the outbreak start?* To accurately model, monitor, and manage the dynamics of COVID-19, it is critical to know precisely when the outbreak first started in a particular region, state, or country. This issue is closely related to selecting appropriate initial conditions for the susceptible, exposed, infectious, and recovered populations S_0 , E_0 , I_0 , and R_0 . For example, the first reported COVID-19 death in the United States was reported in Santa Clara County, California. Although this happened as early as February 6, the case remained unnoticed until April 22 [Allday, 2020]. This unexpected finding suggests that the new coronavirus had been circulating in the Bay Area as early as January. Machine learning allows us to combine our SEIIR model with data from antibody prevalence studies [Bendavid et al., 2020] and reported case numbers [Santa Clara County, 2020], and trace the initial outbreak date back to January 20, 2020 [Peirlinck et al., 2020a]. This early outbreak estimate supports the common intuition that COVID-19 is often present in a population long before the first official case is reported. Knowing the initial outbreak date is critical to trace the origin of the disease, estimate the impact of community spreading, and design successful mitigation strategies.

Modeling exit strategies.

Lesson 9. Selective reopening can be more effective than voluntary quarantine.

A popular strategy to prevent a local outbreak during the COVID-19 pandemic is to restrict incoming travel and locally reduce the case numbers to a manageable dimension. Once a region has successfully contained the disease, the obvious question becomes: *When is it safe to reopen?* There is a legitimate

fear that easing off travel restrictions, even slightly, could trigger a new outbreak and accelerate the spread to an unmanageable degree. As we are trying to identify exit strategies from local lockdowns and global travel restrictions, political decision makers are turning to mathematical models for quantitative insight and scientific guidance [Linka et al., 2020b].

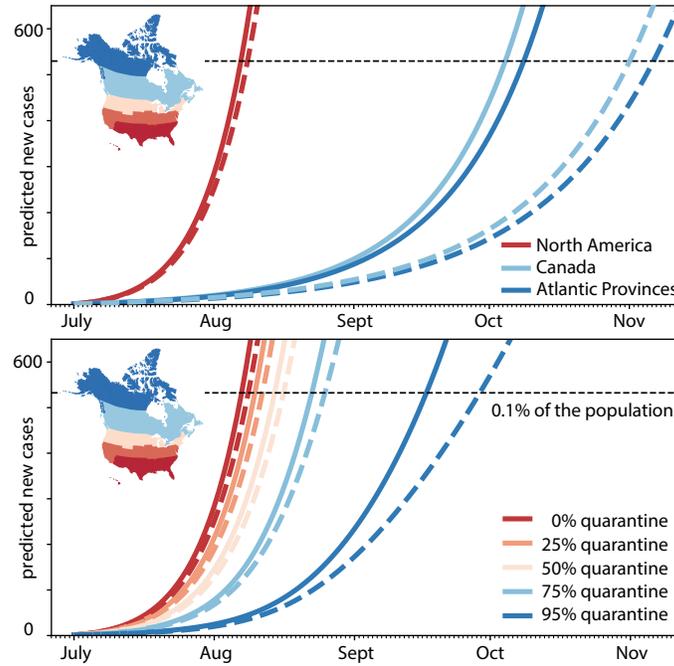


Figure 6. *Outbreak dynamics of COVID-19 in Newfoundland and the effects of restricted travel and quarantine. Reopening forecast for 150-day period with incoming travelers from the Atlantic Provinces, Canada, and all of North America with no quarantine requirements, top, and from all of North America quarantining from 0% to 95%, bottom. Predictions are based on a local SEIR model using the mean effective reproduction number of $R = 1.35$ for all of North America, solid lines, and $R = 1.16$ for Canada, dashed lines, on the day of opening. The black horizontal line marks 0.1% of the population of Newfoundland and Labrador [Linka et al., 2020b].*

Global network mobility models, combined with local epidemiology models, can provide valuable insight into different reopening scenarios. Figure 6 illustrates the effect of different exit strategies for the example of Newfoundland and Labrador, a Canadian province that has enjoyed no new cases since late April, 2020 [Berry et al., 2020]. This analysis combines a network epidemiology model with machine learning to infer parameters and predict the COVID-19 dynamics upon partial and full airport reopening, under perfect and imperfect quarantine conditions. To accurately mimic the incoming susceptible, exposed, infectious, and recovered fractions of travelers at the day of reopening, the model learns the populations of the SEIR model with individual dynamic effective reproduction numbers $R(t)$ for all territories, provinces, and states of North America from the reported case data and weights them with the average daily air travel prior to the outbreak [International Air Transport Association, 2020]. An interesting metric is the estimated number of incoming exposed and infectious travelers upon full reopening, $\Delta E = 0.203/\text{day}$ and $\Delta I = 0.329/\text{day}$. This implies that every five days and every three days, an exposed and an infectious traveler would enter the province of Newfoundland and Labrador. In other words, every other day, a new COVID-19 case would enter the Newfoundland and Labrador via air travel [Linka et al., 2020b]. Since the exposed and early infectious individuals are

still pre-symptomatic, it is impossible to identify and isolate them without strict quarantine requirements [Hellewell et al., 2020]. This raises the question: *What is the best exit strategy?* Is it safer to selectively reopen the province, to only the Atlantic Provinces, to Canada, or to all of North America, or, alternatively, rely on a sufficiently large fraction of the incoming travelers to comply with recommended quarantine requirements? Figure 6 shows that—especially for smaller provinces or states like Newfoundland—tight border control is often easier and more effective than quarantine. Partial reopening, for example within local travel bubbles, is an effective compromise and a reasonable intermediate step towards complete reopening. While relaxing travel restrictions is possible, it would require strict quarantine conditions. Voluntary quarantine, even at an overall rate of 95%, is not enough to entirely prevent a new outbreak. Without comprehensive test-trace-isolate strategies [Fauci et al., 2020], combined with a mandatory quarantine of 100% of COVID-19 positive individuals, reopening can always seed a new exponential outbreak. This concrete example shows that data-driven modeling can provide valuable quantitative insight into the efficacy of travel restrictions to inform political decision making in the controversy of reopening.

Lesson 10. Testing is critical for safe reopening.

Several unexpected trends in the timeline of COVID-19 have recently raised the question: *To which extent does testing manipulate the data?* During the early stages of the pandemic, clearly, only severely symptomatic cases were tested and identified. After gaining a better understanding of the pathophysiology of COVID-19, we now know about the disproportionately high prevalence of asymptomatic transmission. In fact, the estimated undercount during the early stages of the pandemic was on the order of ten or more, meaning that only one in ten infections was detected and reported [Ioannidis, 2020]. As testing is becoming more available and more common, we expect to detect a larger fraction of asymptomatic individuals and see the undercount decrease. From a data science standpoint, the question becomes: *How can we best correct for under- and over-testing?* A simple idea would be to interpret the reported deaths as ground truth, introduce a testing bias, and calibrate the models with respect to the total death count [Holmdahl & Buckee, 2020]. This approach critically depends on a region's current stage within the outbreak and its definition of death, which can vary significantly by country. For example, some countries report death of COVID-19 while others report death with COVID-19. This could explain, at least in part, why some countries like the France, the United Kingdom, and Belgium report death rates of 16.7%, 15.3%, and 14.9%, while others like Denmark, Germany, and Austria report 4.6%, 4.5% and 3.5% [Johns Hopkins University, 2020]. It is now widely recognized that broad testing will not only be important to advance our understanding of the data, but also be a mandatory step for safe reopening. Testing will help us to identify high-risk groups in the population based on age, gender, blood group, and underlying medical conditions. As test-trace-isolate is likely to become the new normal, mathematical models can help us estimate and understand how much and how often we need to test.

Asking the right question.

Throughout the past six months, we have made impressive progress towards understanding the COVID-19 pandemic through data-driven modeling. We now know that the classical epidemiology models that have served us well during the early 20th century are a useful starting point to design models for infectious diseases in the 21st century [Cobey, 2020]. However, since the world is a lot more connected

than a hundred years ago [Soper, 1919], local behavior and global mobility play an equally important role in modeling the outbreak dynamics and outbreak control of COVID-19 today. Against our initial fear of seeing the number of infectious explode beyond control, we now know that we can actually modulate the disease dynamics through behavioral changes and political measures [Linka, 2020a]. Not only have we learned how rapidly the disease curve would grow in the absence of interventions; we also know how long it takes for political measures to effectively bend the curve. In fact, our current COVID-19 models are much better than their public reputation [Holmdahl & Buckee, 2020]. They can predict, interpret, and explain the effects with parameters and numbers [Bar-On et al., 2020].

But we had to learn to set expectations right and to be very specific about asking the right questions [Siegenfeld et al., 2020]. We have learned that generic questions like *What will the disease trajectory look like?* are virtually impossible to answer, especially when projected several weeks or months into the future. Instead, specific questions like *What is the effect of changing this?* are much easier to answer and can be equally insightful for political decision making. And, as a useful by-product, we can even quantify uncertainties and provide confidence intervals on our response.

Six months into the pandemic, there are many more open questions than answers, and we will likely not be able to solve all of them before the disease decays. One of the most pressing questions to complete our understanding of the outbreak dynamics is *What is the true size of the affected population?*, which we can reword into the easier-to-answer questions *How would knowing the rate of asymptomatic transmission change our understanding of the disease?* or *How would this knowledge change with unlimited testing?*. Another important questions that will drive our priorities upon reopening is *How homogeneous is the spread?*, which we could rephrase more quantitatively as *What are the most vulnerable populations?* or *What are scientific metrics to identify superspreading events?*, or even more specifically, *How would our knowledge change if children were more asymptomatic and less infectious than adults?*. Moving into the fall, we will likely want to know *Will there be a second wave?*, but rather we should ask *How do increased mobility during the summer and seasonality during the fall impact the reproduction number?*. Instead of asking the million-dollar question *Can we prevent a resurgence?* we should ask, *What is the limit reproduction number beyond which we can no longer manage the disease through test-isolate-trace strategies?*. These are all questions which data-driven modeling can confidently help to address.

As modelers, it is our ethic responsibility to educate the public to ask the right questions and to communicate the limitations of our answers. One of the most frequent questions the general population is asking today is *When will there be a vaccine?* As modeling community, we should rephrase this question and ask *If we are willing to wait 18 months to find the right vaccines for COVID-19, why don't we allow ourselves at least half of this same time to design the right models?*

Acknowledgments

The discussions and collaborations with Kevin Linka, Mathias Peirlinck, Paris Perdikaris, Francisco Sahli Costabal, and Alain Goriely that have stimulated this review are gratefully acknowledged. This work was supported by the National Institutes of Health Grant U01 HL119578.

References

- M. Alber, A. Buganza Tepole, W. Cannon, S. De, S. Dura-Bernal, K. Garikipati, G. Karniadakis, W.W. Lytton, P. Perdikaris, L. Petzold, E. Kuhl. Integrating machine learning and multiscale modeling: Perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *npj Digital Medicine* 2 (2019) 115.
- E. Allday, M. Kawahara. First known U.S. coronavirus death occurred on Feb. 6 in Santa Clara County. *San Francisco Chronicle*, April 22, 2020.
- R. M. Anderson, R. M. May. Directly transmitted infectious diseases: control by vaccination. *Science* 215 (1982) 1053-1060.
- Apple Mobility Trends. <https://www.apple.com/covid19/mobility>. accessed: July 27, 2020.
- A. Arenas, W. Cota, J. Gomez-Gardenes, S. Gomez, C. Granell, J.T. Matamalas, D. Soriano-Panos, B. Steinegger. Derivation of the effective reproduction number R for COVID-19 in relation to mobility restrictions and confinement. *medRxiv* doi:10.1101/2020.04.06.20054320.
- Y.M. Bar-On, A. Flamholz, R. Phillips, R. Milo. SARS-CoV-2 (COVID-19) by the numbers. *eLife* 9 (2020) e57309.
- E. Bendavid, 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.
- D. Bernoulli. Essay d'une nouvelle analyse de la mortalite causee par la petite verole et des avantages de l'inoculation pour la prevenir. *Mem. Math. Phys. Acad. Roy. Sci. Paris* (1760) 1-45.
- I. Berry, J.P. R. Soucy, A. Tuite, D. Fisman. Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. *Canadian Medical Association Journal* 192 (2020) E420.
- M. Chinazzi, J. T Davis, M. Ajelli, C. Gioanni, ... A. Vespignani. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* (2020) doi:10.1126/science.aba9757.
- S. Cobey. Modeling infectious disease dynamics. *Science* (2020) doi:10.1126/science.aba5659.
- J. Dehning, J. Zierenberg, F.P. Spitzner, M. Wibral, J. Pinheiro Neto, M. Wilczek, V. Priesemann, Inferring COVID-19 spreading rates and potential change points for case number forecasts. *arXiv* (2020) 2004.01105.
- P. L. Delamater, E. J. Street, T. F. Leslie, Y. T. Yang, K. H. Jacobsen. Complexity of the basic reproduction number R_0 . *Emerg. Infect. Disease* 25 (2019) 1-4.
- K. Dietz. The estimation of the basic reproduction number for infectious diseases. *Stat. Meth. Med. Res.* 2 (1993) 23-41.
- European Commission. COVID-19: Temporary restriction on non-essential travel to the EU. Communication from the Commission to the European Parliament, the European Council and the Council. Brussels, March 16, 2020.

Eurostat. Your key to European statistics. Air transport of passengers. <https://ec.europa.eu/eurostat>. accessed: July 28, 2020.

A.S. Fauci, H.C. Lane, R.R. Redfield. Covid-19—Navigating the uncharted. *New Engl. J. Med.* (2020) 382:1268-1269.

P. E. M. Fine. Herd immunity: history, theory, practice. *Epidemiologic Reviews* 15 (1993) 265-302.

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.

J. Hellewell, S. Abbott, A. Gimma, N.I. Bosse, C.I. Jarvis, T.W. Russell, J.D. Munday, A.J. Kucharski, W.J. Edmunds. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Global Health* (2020) 8:e488-496.

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

I. Holmdahl, C. Buckee. Wrong but useful – What Covid-19 epidemiology models can and cannot tell us. *N. Engl. J. Med.* 383 (2020) 303-305.

J. Hsu. Here's how computer models simulate the future spread of new coronavirus. *Scientific American* (2020) February 23, 2020.

Institute for Health and Metrics Evaluation IHME. COVID-19 Projections. <https://covid19.healthdata.org> accessed: July 27, 2020.

International Air Transport Association. <https://www.iata.org>. accessed: July 27, 2020.

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

P.K. Jha, L. Cao, T. Oden. Bayesian-based predictions of COVID-19 evolution in Texas using multispecies mixture-theoretic continuum models. *Computational Mechanics* (2020) in revision.

Johns Hopkins University, Baltimore, 2020. Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering. <https://coronavirus.jhu.edu/map.html>; <https://github.com/CSSEGISandData/covid-19>. assessed: July 28, 2020.

A. Kergassner, C. Burkhardt, D. Lippold, M. Kergassner, L. Pflug, D. Budday, P. Steinmann, S. Budday. Memory-based mesos-scale modeling of Covid-19. medRxiv. doi:10.1101/2020.06.10.20126771.

W. O. Kermack, G. McKendrick. Contributions to the mathematical theory of epidemics, Part I. *Proc. Roy. Soc. London Ser. A* 115 (1927) 700-721.

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.

K. Linka, M. Peirlinck, F. Sahli Costabal, E. Kuhl. Outbreak dynamics of COVID-19 in Europe and the effect of travel restrictions. *Comp. Meth. Biomech. Biomed. Eng;* (2020) online first. doi:10.1080/10255842.2020.1759560.

K. Linka, M. Peirlinck, E. Kuhl. The reproduction number of COVID-19 and its correlation with public health interventions. *Comp. Mech.* (2020a) online first. doi:10.1007/s00466-020-01880-8 .

K. Linka K, P. Rahman, A. Goriely, E. Kuhl. Is it safe to lift COVID-19 travel restrictions? The Newfoundland story. *Comp. Mech.* (2020b) online first. doi: 10.1007/s00466-020-01899-x. medRxiv. doi:10.1101/2020.07.16.20155614.

K. Linka, A. Goriely, E. Kuhl. Global and local mobility as a barometer for COVID-19 dynamics. (2020c) medRxiv doi:10.1101/2020.06.13.20130658.

Y. Liu, A. A. Bayle, A. Wilder-Smith, J. Rocklöv. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Medicine* (2020) 1-4.

B. Mason Meier, R. Habibi, Y. Tony Yang. Travel restrictions violate international law. *Science* 367 (2020) 1436.

S. W. Park, B. M. Bolker, D. Champredon, D. J. D. Earn, M. Li, J. S. Weitz, B. T. Grenfell, J. Dushoff. Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus outbreak. medRxiv (2020) doi:10.1101/2020.01.30.20019877.

M. Peirlinck, K. Linka, F. Sahli Costabal, E. Kuhl. Outbreak dynamics of COVID-19 in China and the United States. *Biomech. Model. Mechanobio.* (2020) doi:10.1007/s10237-020-01332-5.

M. Peirlinck, K. Linka, F. Sahli Costabal, E. Bendavid, J. Bhattacharya, J.P.A. Ioannidis, E. Kuhl. Visualizing the invisible: The effect of asymptomatic transmission on the outbreak dynamics of COVID-19. (2020a) medRxiv doi:10.1101/2020.05.23.20111419.

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

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

Santa Clara County. COVID-19 Cases and Hospitalizations Dashboard. <http://www.sccgov.org>. assessed: July 27, 2020.

A.F. Siegenfeld, N.N. Taleb, Y. Bar-Yam. What models can and cannot tell us about COVID-19. *Proc. Nat. Acad. Sci.* 117 (2020) 16092-16095.

G.A. Soper. The lessons of the pandemic. *Science* XLIX (1919) 501-505.

H. Streeck, B. Schulte, B.M. Kummerer, E. Richter, T. Holler, 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. Nothen, A.M. Eis-Hubinger, 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.

G. Viceconte, N. Petrosillo. COVID-19 R0: Magic number or conundrum? *Infect. Dis. Rep.* 12 (2020) 8516.

A. Viguerie, A. Veneziani, G. Lorenzo, D. Baroli, N. Aretz-Nellesen, A. Patton, T.E. Yankeelov, A. Reali, T.J.R. Hughes, F. Auricchio. Diffusion-reaction compartmental models formulated in a continuum mechanics framework: application to COVID-19, mathematical analysis, and numerical study. *Computational Mechanics* (2020) in revision.

Z. Wang, X. Zhang, G.H. Teichert, M. Carraso-Teja, K. Garikipati. System inference for the spatio-temporal evolution of infectious diseases. *Computational Mechanics* (2020) in revision.

A. Wilder-Smith, C.J. Chiew, V.J. Lee. Can we contain the COVID-19 outbreak with the same measures as for SARS? *Lancet Infect. Dis.* 20 (2020) e102-107.

A. Zlojutro, D. Rey, L. Gardner. A decision-support framework to optimize border control for global outbreak mitigation. *Sci. Rep.* 9 (2019) 2216.