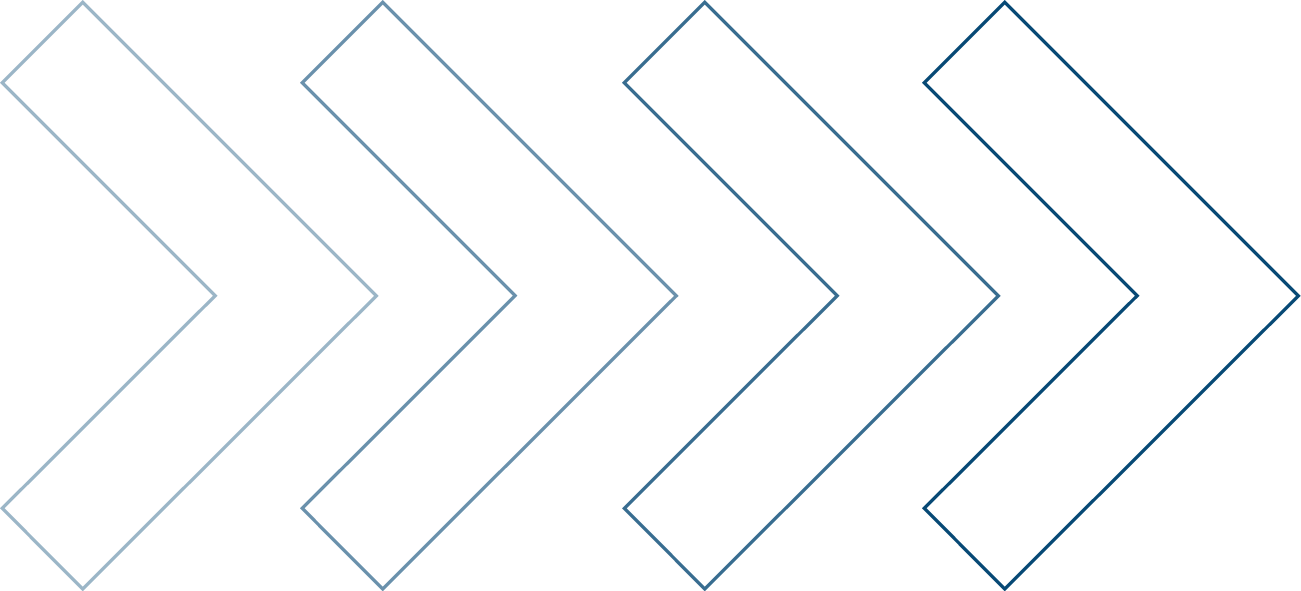
An Overview of Population Models of Epidemics



April 2020

An Overview of Population Models of Epidemics

|  |  |  |  |
| --- | --- | --- | --- |
| AUTHOR |  |  |  |
|  |  |  |

Table of Contents

[Importance of Modeling Epidemics 4](#_Toc38890801)

[The Challenges of Modeling an Outbreak Caused by a Novel Pathogen 4](#_Toc38890802)

[Key Terms that Describe Virus Tranmission and Virulence 8](#_Toc38890803)

[Overview of Modeling Approaches for Projecting Outbreaks Forward in Time 9](#_Toc38890804)

[A Closer Look at Population Models of Epidemics 10](#_Toc38890805)

[Validating and Updating Population Models 10](#_Toc38890806)

[About The Society of Actuaries 11](#_Toc38890807)

# Importance of Modeling Epidemics

The 1918 flu pandemic, oftened referred to as the “Spanish Flu”, infected an estimated 500 million persons and resulted in at least 50 million deaths, equivalent to about 3% of the world’s population at that time.[[1]](#footnote-1) While there have been major advances in medicine over the last 100 years, and vaccines have reduced the risk of outbreaks of transmissible diseases, the 1918 pandemic serves as a reminder that a novel pathogen can rapidly infect a large percentage of the human population over a short time period, with devasting consequences.

New virus strains arise on a regular basis through mutation. In addition, a virus that previously existed only in animal hosts may jump to the human population. The jump could be facilitated by a mutation that increases the virus’ capacity to enter human cells, or via the merger or “recombination” of genetic material from two separate viruses that, by chance, simultaneously infect a single animal cell, leading to the creation of an entirely new virus.[[2]](#footnote-2)

A new virus that finds its way into the human population holds a temporary competitive advantage: its novel structure makes it less likely to encounter a strong immune defense. Few people may possess immunity to the virus, or perhaps none at all. Meeting little or no resistence, a new virus has the potential to spread rapidly, causing an epidemic or pandemic.[[3]](#footnote-3)

In the event of an outbreak, simulation models can provide policymakers, governments and citizens with a means to assess strategic options. One course of action is to do nothing, letting the outbreak run its natural course. If simulations indicate that this could lead to an unacceptably high loss of life, additional simulations can be run to evaluate the impact of actions intended to decelerate or halt the outbreak. Models can provide a rough sense of the effect of options such as school closures, closures of bars and restaurants, travel restrictions, and stay-at-home or shelter-in-place orders. For each option, a forecast of the daily demand for hospital services can be compared against hospital capacity, the goal being to avoid a scenario in which hospitals are overwhelmed with an abrupt surge in the number of new patients.

When forecasting an outbreak caused by a novel pathogen, model builders must contend with many unknowns. The speed of an outbreak may leave the scientific community scrambling to collect and analyze data needed to understand the pathogen’s risk characteristics. Data limitations compel modelers to use assumptions that may have a wide range of uncertainty, and, as a consequence, outbreak forecasts also have a wide range of uncertainty. A particular forecast could, for example, indicate that anywhere from 0.3% to 1.0% of a region’s population could die in the absence of interventions to slow the spread of an infection. A wide range of predicted outcomes may frustrate policymakers and citizens who seek a more precise quantification of risk. However, a range of possible outcomes is preferable to the alternative of “flying blind”, without any guidance and insight from data-driven forecasts. Without forecasts, the human race could be caught completely offguard, deprived of the opportunity to reshape its fate via timely interventions.

# The Challenges of Modeling an Outbreak Caused by a Novel Pathogen

For an outbreak caused by a novel pathogen, the starting point for researchers is usually a small dataset (perhaps gathered from a single city or region) that is likely be inadequate to confidentally provide answers to the following key questions:

* How easily or rapidly is the virus transmitted, and under what conditions?
* What is the average length of the infectious period, and how does this period vary from one individual to another?
* How long can the virus survive outside of a human host (for example, on a kitchen counter)?
* What is the risk of death for those who are infected?
* What factors influence the risk of death, such as age, gender and comorbities?
* Are there individuals who have been infected but who exhibit no symptoms, or whose symtoms are so mild that they don’t need to visit a doctor?
* Is immunity conferred to those who survive an infection?
* Are there country, region and city-specific factors that influence the rate of transmission and/or the fatality rate?
* Does weather influence the transmission rate? For example, is transmission correlated with temperature or humidity levels?

The passage of time gradually increases the body of evidence available to researchers, and, eventually, reliable answers can generally be provided to these questions. But when faced with an outbreak that is rapidly progressing, time is a luxury that decision-makers do not have. In the early stages of an outbreak, a virus may spread at a geometric rate (Table 1), leaving little time for model builders to gather data and refine their forecasts.

Table

number of days required for a virus to infect 50% of a population,

Assuming A constant geometric rate of growth

|  |  |
| --- | --- |
| Days Required for Cumulative Number of Infected Persons to Double | Number of Days for Infection to  Spread to 50% of a Region’s Population,  Given an Initial Infection Rate of 1 out of 1 Million |
|
| 2 | 37.9 |
| 3 | 56.8 |
| 4 | 75.7 |
| 5 | 94.7 |
| 6 | 113.6 |
| 7 | 132.5 |

Source : authors’ calculations

Consider an outbreak in which the number of infected persons doubles every 3 days. Assuming an initial infection rate of just one out of a million persons, in just 57 days the infection will have spread to 50% of the population. Of course, geometric growth cannot continue indefinitely due to the finite size of a population, and an outbreak naturally decelerates after a large percentage of a population has been exposed. Nevertheless, the results in Table 1 roughly illustrate the severe time constraints under which model builders and decision-makers must operate.

With respect to SARS-CoV-2 – the coronavirus that causes the illness known as COVID-19 – data from Johns Hopkins University indicates indicates that, in many countries, the rate of growth of reported cases was rapid during the early stages of the outbreak (Table 2). In the United States and Italy, for example, the total number of cases doubled from 1 per 100 thousand to 2 per 100 thousand persons in just 2.4 days. In the United Kingdom, Canada and France, this increase occurred in about 3.2 days.

Table 2

Doubling Time for total Reported Covid-19 Cases,

after first reaching a threshold of one case per 100,000 of general population

|  |  |  |
| --- | --- | --- |
| Country | Date on Which Cumulative Cases First Exceeded  One Per 100,000 Persons | Number of Days for Reported Cases to Double to Two Per 100,000 Persons |
| Iran | Mar-01 | 1.6 |
| Spain | Mar-07 | 1.8 |
| Italy | Feb-27 | 2.4 |
| US | Mar-15 | 2.4 |
| Korea, South | Feb-23 | 2.8 |
| United Kingdom | Mar-13 | 3.1 |
| France | Mar-06 | 3.2 |
| Canada | Mar-16 | 3.2 |
| Russia | Mar-29 | 3.4 |
| Germany | Mar-08 | 4.0 |
| China | Feb-02 | 4.8 |

Source : authors’ calculations using Johns Hopkins University’s COVID-19 database

Thus, outbreak modelers operate under extreme time pressure, and with much at stake. Delaying a model’s release by a month to refine its key parameters could render it useless to decision-makers. Model forecasts are most valuable to decision-makers early in an outbreak. The earlier an intervention such as social distancing is implemented, the greater will be its likely impact. A delay of merely a few days can greatly reduce the positive impact of a particular intervention.

Time pressure forces outbreak modelers to make due with the limited data that is available early in an outbreak caused by a novel pathogen. Data limitations make it challenging to assess the threat posed by a novel virus. Recent history provides some sense of the magnitude of this challenge. In the spring of 2009, a novel influenza virus emerged and spread rapidly throughout the world. Referred to as “Swine Flu”, researchers scrambled to process emerging data and develop estimates of the danger posed by this virus. In a report published in 2013, four years after the outbreak, the National Institute of Health (NIH) reviewed 77 estimates of Swine Flu case fatality risk from 50 published studies.[[4]](#footnote-4) Many of the 77 estimates were produced in the first nine months of the pandemic. The NIH found “substantial heterogeneity in the published estimates, ranging from less than 1 to more than 10,000 deaths per 100,000 cases or infections.” The report concludes that “our review highlights the difficulty in estimating the seriousness of infection with a novel influenza virus”.[[5]](#footnote-5)

If humans were clones of each other – with identical levels of health, identical immune systems and an identical chance of succumbing to any particular virus – and if all cities (or regions) were copies of each other, with identical mobility patterns, identical household structures, identical patterns of social contact and identical health care systems – the task of quantifying the risk posed by a new virus would be far easier. In this simplified world, a virus that claimed the lives of 0.5% of the population in city “A” would be likely to have the same effect in city “B”. Correspondingly, if closing schools in city “A” reduced the fatality rate from 0.5% to 0.2%, then a similar effect could be expected in city “B”.

Obviously, this “clone” world is not the real world, where individuals differ greatly with respect to health status, immune response, and social behavior, and where cities and regions vary with respect to population density, age distribution, household structures, mobility patterns and health care services. These differences can impact both the rate of virus transmission as well as the the risk a virus poses to those who become infected.

Thus, outbreak modelers must contend with the possibility that data gathered from region “A” may not be directly applicable to region “B”. To some extent, it may be possible to address this issue. For example, if virus transmission and/or lethality are correlated with age, then differences in age distributions between region “A” and “B” can be factored into forecasts. However, in the early stages of an outbreak, many key factors that affect virus transmission and lethality may go undetected. Suppose, for example, that air pollutants influence the transmission of a particular virus, but the initial data is drawn solely from one city. Without data drawn from a range of cities with varying air quality, researchers will lack the covariates required to statistically link air pollutants to virus transmission.

Even in the absence of differences between region “A” and “B”, modelers must contend with the problem of sample bias. As with any statistical analysis, a biased sample may produce a disorted set of conclusions that are not applicable to the population as a whole. A common source of sample bias in an outbreak arises as follows: some individuals become very sick and must be hospitalized, and these individuals are captured in the data; in contrast, other individuals experience mild symptoms or perhaps no symptoms whatsoever, and these individuals are not captured in the data. Thus, the data includes only those individuals who were severely affected by the virus, and, as a consequence, fatality rates estimated from this data will overstate the risk posed to the general population.

In addition, even if a sample is unbiased, it may lack the size required to develop narrow confidence intervals for key parameters. For the sake of argument, suppose that all individuals are identical clones that share the same health status and immune response, and with the same probability of succumbing to any particular virus. Consider a cruise ship with 500 “clone” passengers. Suppose that all passengers become infected by a virus, and that 2% die. Even in this simple case, substantial uncertainty exists regarding the fatality rate. Assuming that the outcome for each individual is an independent Bernoulli trial, and using the Central Limit Theorem to infer that the sum of independent Bernoulli trials is an approximately normal distribution, the 90% confidence interval for the case fatality rate is quite wide, running from 0.97% to 3.03%.

Lastly, modelers must contend with data reporting methods that may vary from city to city, region and region, and country to country. For example, a particular locality may require a positive laboratory test (for a particular type of virus) before entering a case into its database. Other localities may depend on the judgment of a patient’s doctor (or coroner) to assess whether an individual is (or was) infected. Even if laboratory tests are used as opposed to human judgment, the tests may not yet be 100% reliable (because the pathogen is novel), producing false positive and false negative results. Moreover, if a patient with poor health status dies from a virus, it may be unclear whether the virus was the primary cause of death, or if the patient would have died anyway from another condition.

# Key Terms that Describe Virus Tranmission and Virulence

Outbreak modelers and epidiemologists use various terms and statistics to describe the risk posed by a virus. Key terms are as follows:

**Basic reproduction number (R0)**

The average number of persons to whom an infected person transmits the virus, assuming that all individuals in a population are susceptible to infection. This value is often refered to as “R0” or “R naught”. A value of 2, for example, indicates that, on average, an infected person will transmit the virus to two additional persons. R0 depends not only on the characteristics of a virus, but also upon social and environmental factors. Thus, a particular virus might have an R0 of 2.5 when estimated for a dense urban area, but only 1.8 when measured in a rural area.

All else equal, the larger the value of R0, the greater will be the speed of an outbreak. But R0 by itself does not provide a measurement of the rate at which an outbreak spreads. For example, both a slowly evolving outbreak and a quickly evolving outbreak could have an R0 of 2.0. In the slow outbreak, an individual’s infectious period would be relatively long, but the risk of transmission on any particular day of that period would be relatively low. In the case of the fast outbreak, the infectious period would be shorter, but the risk of transmission on any particular day would be larger.

If R0 is less than 1.0, the number of new infections will decline across time. Conversely, if R0 is greater than 1.0, then an infection could, in theory, spread through a population. In general, the greater the value of R0, the harder it will be to control an outbreak.

**Effective reproduction number (Re)**

The average number of persons to whom an infected person transmits the virus, given the current state of the population. At the outset of an outbreak caused by a novel pathogen, the entire population is susceptible, and R0 and Re are identical. As an outbreak progresses, the susceptible population declines because some individuals will already have been exposed to the virus. As a result, Re declines across time. If, for example, the susceptible population is half of the total population, then Re will be half of R0.

**Incubation period**

The period between an individual’s exposure and the onset of clinical symptoms.

**Infectious period**

The period during which an infected individual can transmit the virus to another person. Depending on the virus, and to some extent the individual, the infectious period may begin immediately after infection, or there may be a lag between the time of infection and the time at which the individual becomes infectious.

**Generation time**

If a person transmits a virus to another person, the time between the onset of symptoms in the first person and the onset of symptoms in the second person is refered to as “generation time”. In general, the shorter the average generation time, and the larger the basic reproduction number, the greater the rate at which an infection propogates through a population.

**Case Fatality Rate (CFR)**

The number of virus-related deaths divided by the number of diagnosed cases. In the early stages of an outbreak, the CFR typically captures only the most severe cases, and thus may fail to be a good indicator of the risk posed to the population that is not yet infected.

Note that a CFR is not identical to a mortality rate. A mortality rate is typically expressed as an annual rate, while a CFR is independent of time. For example, consider a fast-acting virus that runs its course in merely one week, resulting in death in 10% of cases. This virus has the same CFR as a slower-acting virus that kills 10% of infected persons, but does so across an average period of illness of 6 months.

**Infection Fatality Rate (IFR)**

Like the CFR, the IFR is equal to deaths divided by cases. But the IFR attempts to capture not only the most severe cases, but also mild and “subclinical” cases that do not typically lead to visits to a physician. An IFR is lower than the corresponding CFR.

# Overview of Modeling Approaches for Projecting Outbreaks Forward in Time

Roughly speaking, there are two main types of outbreak forecasting models: 1) statistical models and (2) mechanistic models. The Institute for Health Metrics and Evaluation (IHME) model[[6]](#footnote-6), which has been frequently cited by the media as well as by the White House[[7]](#footnote-7), is an example of a statistical model[[8]](#footnote-8), while the Imperial College of London’s model[[9]](#footnote-9),[[10]](#footnote-10) is an example of a susceptible-infected-recovered (SIR) model, which is a type of mechanistic model.

A statistical model uses correlations or patterns in data to forecast the propagation of a virus. A common approach is to focus on the time series of virus-related deaths, separately by city or geographic region, fitting this data to a curve the describes the anticipated rise, peak and fall of the number of daily deaths. The curve is extracted from cities or regions that have already passed through the outbreak (such as Wuhan, China, with respect to COVID-19). The assumption is that, in each different region, the outbreak will follow a similar “shape”, curve or pattern across time. A model may tweak or adjust the assumed outbreak shape to account for region or city-specific factors, such as delays associated with implementing social distancing measures.

In contrast to statistical models, mechanistic models focus on the dynamic processes through which a virus propagates through a population. Estimates of the transmissibility and lethality of the virus are used to simulate the progression of an outbreak across time. A SIR model, for example, projects shifts in the population from “susceptible” (i.e. not yet infected) to “infected”, and from “infected” to either “recovered” or deceased. Some SIR models are quite simple, assuming that all persons have an equal chance of becoming sick, that infected persons are equally likely to transmit the virus, and that infected persons share the same probability of death. More complicated SIR models subdivide the population into groups, each group having distinct characteristics with respect to risk of infection, risk of transmission, and risk of death. Some SIR models go a step further, using an agent-based method to simulate unique individuals (as opposed to groups of individuals), each interacting with other unique simulated individuals.

# A Closer Look at Population Models of Epidemics

Types of population models: SIR, SEIR, SEIRS

Number of “compartments” and subgroups

Homogenous vs non-homogenous mixing models

Difficulty of determining models parameters during the early stage of an epidemic: basic reproduction number, IFR

Effective reproduction number can vary from country to country and region to region

Effective reproduction number is affected by interventions such as social distancing

# Validating and Updating Population Models

ronavirus, there are four periods associated with the disease:

Infected, but not contagious and not symptomatic.

Infected and contagious, but not symptomatic.

Infected, contagious, and symptomatic.

Recovering (assuming survival), where you may still have symptoms but are no longer contagious.

Without a widespread test available to us all, pa

To project hospital visits, models use data to assess the probability that an infected person will become seriously ill. Some models, such as the IHME’s model, assume that hospital services are a function of the projected number of deaths. Many models not only forecast the demand for hospital services, but also compare that demand to available hospital beds in each geographic region, thereby developing a sense of where hospital strain is most likely to occur.

Relative to statistical models, SIR models are generally more “data hungry”, but they provide greater modeling flexibility. For example, a SIR model’s key parameters can be sensitivity-tested to produce a range of simulated outcomes, such as low, most-likely and high estimates for the number of hospitalizations and deaths. In addition, the impact of social distancing measures can be estimated by adjusting the basic reproduction number (“R0”) which describes the rate of virus transmission.

Both statistical models and mechanistic models can quickly become “stale” during the early stages of an outbreak. With little data to draw upon, initial modeling efforts necessitate the use of assumptions that have a wide range of uncertainty. As an outbreak progresses, the pool of available data expands, providing researchers with valuable information that can be used to revise their models.

Inevitably, model revisions result in shifts in outbreak forecasts. Large shifts could potentially undermine the public’s faith in a model. However, revisions of forecasts do not, in general, arise from a lack of modeling expertise, but rather from data limitations that are part and parcel of dealing with a new pathogen. Revisions to forecasts are a sign that modelers are paying attention to the continuous influx of new data produced by researchers around the world, and diligently adjusting their models to reflect the most current available information about the pathogen.

# About The Society of Actuaries

With roots dating back to 1889, the [Society of Actuaries](https://www.soa.org/) (SOA) is the world’s largest actuarial professional organizations with more than 31,000 members. Through research and education, the SOA’s mission is to advance actuarial knowledge and to enhance the ability of actuaries to provide expert advice and relevant solutions for financial, business and societal challenges. The SOA’s vision is for actuaries to be the leading professionals in the measurement and management of risk.

The SOA supports actuaries and advances knowledge through research and education. As part of its work, the SOA seeks to inform public policy development and public understanding through research. The SOA aspires to be a trusted source of objective, data-driven research and analysis with an actuarial perspective for its members, industry, policymakers and the public. This distinct perspective comes from the SOA as an association of actuaries, who have a rigorous formal education and direct experience as practitioners as they perform applied research. The SOA also welcomes the opportunity to partner with other organizations in our work where appropriate.

The SOA has a history of working with public policymakers and regulators in developing historical experience studies and projection techniques as well as individual reports on health care, retirement and other topics. The SOA’s research is intended to aid the work of policymakers and regulators and follow certain core principles:

**Objectivity:** The SOA’s research informs and provides analysis that can be relied upon by other individuals or organizations involved in public policy discussions. The SOA does not take advocacy positions or lobby specific policy proposals.

**Quality:** The SOA aspires to the highest ethical and quality standards in all of its research and analysis. Our research process is overseen by experienced actuaries and nonactuaries from a range of industry sectors and organizations. A rigorous peer-review process ensures the quality and integrity of our work.

**Relevance:** The SOA provides timely research on public policy issues. Our research advances actuarial knowledge while providing critical insights on key policy issues, and thereby provides value to stakeholders and decision makers.

**Quantification:** The SOA leverages the diverse skill sets of actuaries to provide research and findings that are driven by the best available data and methods. Actuaries use detailed modeling to analyze financial risk and provide distinct insight and quantification. Further, actuarial standards require transparency and the disclosure of the assumptions and analytic approach underlying the work.

Society of Actuaries

475 N. Martingale Road, Suite 600

Schaumburg, Illinois 60173

[www.SOA.org](http://www.SOA.org)

1. https://www.npr.org/2020/04/02/826358104/the-1918-flu-pandemic-was-brutal-killing-as-many-as-100-million-people-worldwide [↑](#footnote-ref-1)
2. An overview of virus mutation is available here: <https://www.ncbi.nlm.nih.gov/books/NBK8439/>. An overview of how viruses can jump from one species to another is available here: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2546865/> [↑](#footnote-ref-2)
3. <https://www.dictionary.com/e/epidemic-vs-pandemic/> [↑](#footnote-ref-3)
4. Wong, Jessica Y., et al. (2013, November 24). “Case Fatality Risk of Influenza A(H1N1pdm09): A Systematic Review.” *Epidemiology (Cambridge, Mass.), 24(6),* 830–841. [*https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3809029/*](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3809029/). [↑](#footnote-ref-4)
5. With the passage of time and the expansion of available data, researchers were able to produce increasingly reliable estimates of the mortality risk posed by (H1N1)pdm09, the virus that caused the Swine Flu outbreak of 2009. Today, more than ten years after the outbreak, the Centers for Disease Control and Prevention (CDC) estimates that 61 million Americans were infected with the virus during the one-year period beginning in April 2009, of which 12.5 thousand persons died. This translates into an infection mortality rate of 0.02%. This is less than the CDC’s estimate of 0.1% for the mortality rate associated with seasonal flu. [↑](#footnote-ref-5)
6. Murray, Christopher JL. (2020, March 30). “Forecasting COVID-19 Impact on Hospital Bed-Days, ICU-Days, Ventilator-Days and Deaths by U.S. State in the Next Four Months.” MedRxiv. [*https://www.medrxiv.org/content/10.1101/2020.03.27.20043752v1*](https://www.medrxiv.org/content/10.1101/2020.03.27.20043752v1)*.* [↑](#footnote-ref-6)
7. Aizenman, Nurith. (2020, April 1). “Five Key Facts Not Explained in White House COVID-19 Projections.” NPR. [*https://www.npr.org/sections/health-shots/2020/04/01/824744490/5-key-facts-the-white-house-isnt-saying-about-their-covid-19-projections*](https://www.npr.org/sections/health-shots/2020/04/01/824744490/5-key-facts-the-white-house-isnt-saying-about-their-covid-19-projections)*.* [↑](#footnote-ref-7)
8. While the IMHE model uses a statistical approach to project the number of deaths, the component of the model that projects hospital service utilization is best described as mechanistic. Thus, the IMHE model has both statistical and mechanistic components. [↑](#footnote-ref-8)
9. Flaxman, Seth, et al. (2020, March 30). “Report 13—Estimating the Number of Infections and the Impact of Non-Pharmaceutical Interventions on COVID-19 in 11 European Countries.” Imperial College London. [*https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-13-europe-npi-impact/*](https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-13-europe-npi-impact/). [↑](#footnote-ref-9)
10. Adam, David. (2020, April 2). “Special Report: The Simulations Driving the World’s Response to COVID-19.” Nature. [*https://www.nature.com/articles/d41586-020-01003-6*](https://www.nature.com/articles/d41586-020-01003-6)*.* [↑](#footnote-ref-10)