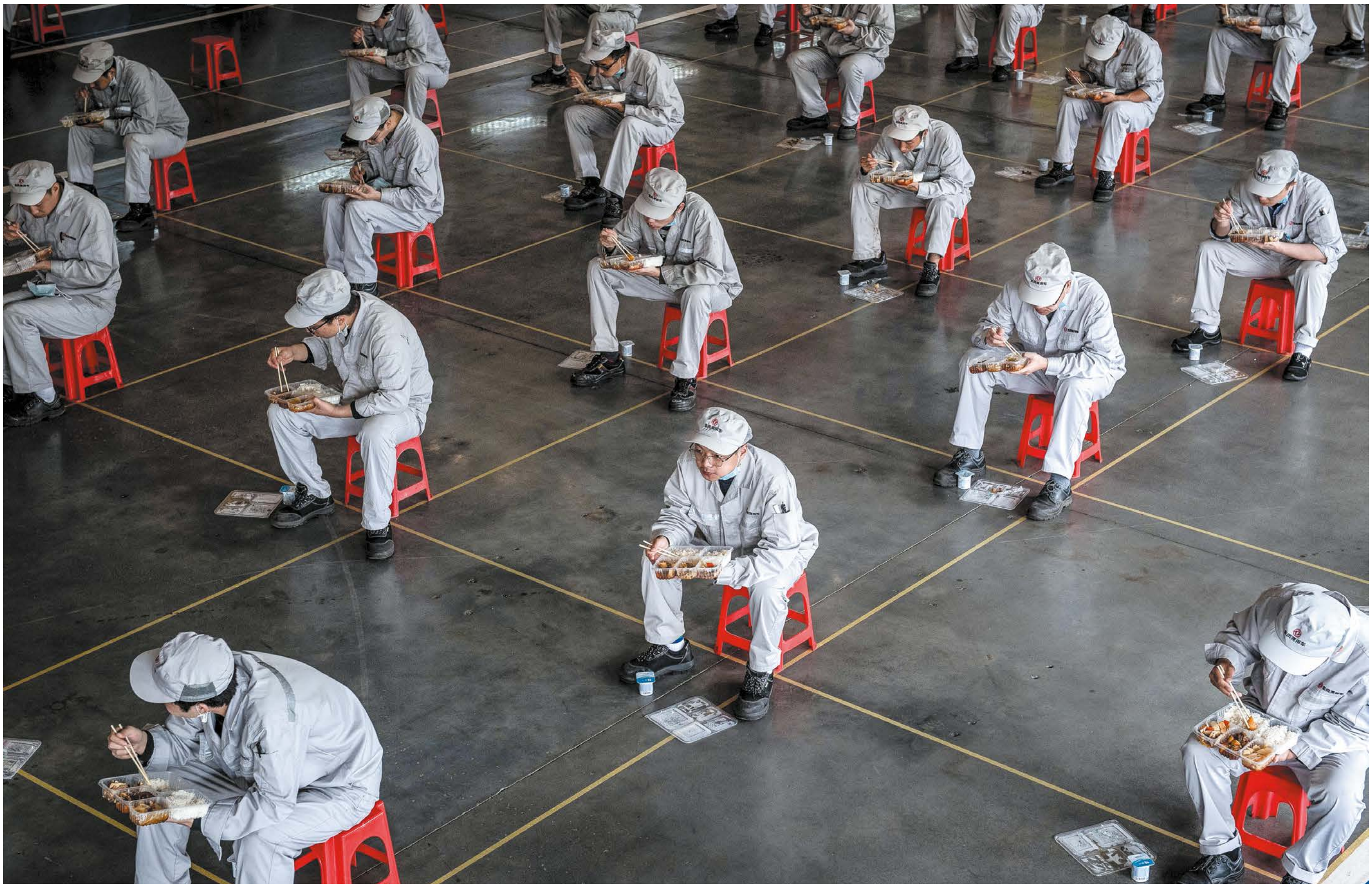


AFP/GETTY

Staff at a car-manufacturing plant in Wuhan, China, observe social-distancing measures during their lunch break.

MODELLING THE PANDEMIC

The simulations driving the world's response to COVID-19. By David Adam

When Neil Ferguson visited the heart of British government in London's Downing Street, he was much closer to the COVID-19 pandemic than he realized. Ferguson, a mathematical epidemiologist at Imperial College London, briefed officials in mid-March on the latest results of his team's computer models, which simulated the rapid spread of the coronavirus SARS-CoV-2 through the UK population. Less than 36 hours later, he announced on Twitter that he had a fever and a cough. A positive test followed. The disease-tracking scientist had become a data point in his own project.

Ferguson is one of the highest-profile faces in the effort to use mathematical models that predict the spread of the virus – and that show how government actions could alter the course of the outbreak. "It's been an immensely intensive and exhausting few months," says Ferguson, who kept working throughout his relatively mild symptoms of COVID-19. "I haven't really had a day off since mid-January."

Research does not get much more policy-relevant than this. When updated data in the Imperial team's model[1] indicated that the United Kingdom's health service would soon be overwhelmed with severe cases of COVID-19, and might face more than 500,000 deaths if the government took no action, Prime Minister Boris Johnson almost immediately announced stringent new restrictions on people's movements. The same model suggested that, with no action, the United States might face 2.2 million deaths; it was shared with the White House and new guidance on social distancing quickly followed (see 'Simulation shock').

Governments across the world are relying on mathematical projections to help guide decisions in this pandemic. Computer simulations account for only a fraction of the data analyses that modelling teams have performed in the crisis, Ferguson notes, but they are an increasingly important part of policymaking. But, as he and other modellers warn, much information about how SARS-CoV-2 spreads is still unknown and must be estimated or assumed – and that limits the precision of forecasts. An earlier version of the Imperial model, for instance, estimated that SARS-CoV-2 would be about as severe as influenza in necessitating the hospitalization of those infected. That turned out to be incorrect.

The true performance of simulations in this pandemic might become clear only months or years from now. But to understand the value of COVID-19 models, it's crucial to know how they are made and the assumptions on which they are built. "We're building simplified representations of reality. Models are not crystal balls," Ferguson says.

Coronavirus models: the basics

Many of the models simulating how diseases spread are unique to individual academic groups that have been developing them for years. But the mathematical principles are similar. They are based on trying to understand how people move between three main states, and how quickly: individuals are either susceptible (S) to the virus; have become infected (I); and then either recover (R) or die. The R group is presumed to be immune to the virus, so can no longer pass on the infection. People with natural immunity would also belong to this group.

The simplest SIR models make basic assumptions, such as that everyone has the same chance of catching the virus from an infected person because the population is perfectly and evenly mixed. More-advanced models, which make the quantitative predictions policymakers need during an emerging pandemic, subdivide people into smaller groups — by age, sex, health status, employment, number of contacts, and so on — to set who meets whom, when and in which places.

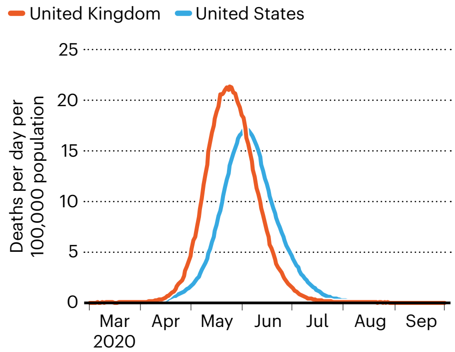
Using detailed information on population size and density, how old people are, transport links, the size of social networks and health-care provision, modellers build a virtual copy of a city, region or an entire country using differential equations to govern the movements and interactions of population groups in space and time. Then they seed this world with an infection and watch how things unfold.

But that, in turn, requires information that can be only loosely estimated at the start of an epidemic, such as the proportion of infected people who die, and the basic reproduction number (R_0) — the number of people, on average, to whom one infected person will pass the virus. The modellers at Imperial, for instance, estimated in their 16 March report[1] that 0.9% of people infected with COVID-19 would die (a figure adjusted to the United Kingdom's specific demographics); that the R_0 was between 2 and 2.6; and that SARS-CoV-2 takes 5.1 days to incubate in an infected person. Those figures depended on other kinds of modelling: rough estimates by epidemiologists who tried to piece together the virus's basic properties from incomplete information in different countries during the pandemic's early stages.

Some parameters, meanwhile, must be entirely assumed. The Imperial team had to surmise, for instance, that there is no natural immunity to COVID-19 — so the entire population starts out in the susceptible group — and that people who recover from COVID-19 are

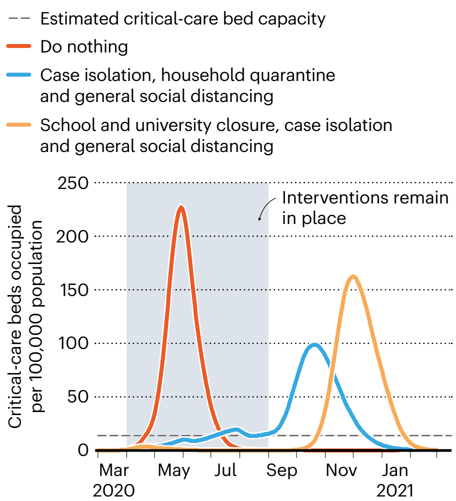
SIMULATION SHOCK

A model by Imperial College London in mid-March predicted a total of more than 500,000 UK deaths from COVID-19, and more than 2.2 million in the United States if no action was taken to stop the virus spreading in those countries.



A SECOND WAVE

In the United States, implementing measures to contain the virus could stop people with COVID-19 from immediately overwhelming the country's critical-care hospital-bed capacity, a simulation from Imperial College London suggests. But a second wave of the pandemic might be expected later in the year.



SOURCES: REF. 1

immune to reinfection in the short term.

A simulation run using these parameters would always give the same forecast. But simulations called stochastic models inject randomness: rolling a virtual dice to see whether someone in the I group infects an S person when they meet. This gives a range of possibilities when the model is run multiple times.

Modellers also simulate people's activities in different ways. In 'equation-based' models, individuals are sorted into population groups. But as the groups are broken into smaller, more-representative social subsets to better reflect reality, the models get increasingly complicated. An alternative approach is to use an 'agent-based' method in which each individual moves around and acts according to their own specific rules — rather like the simulated characters in the video-game series The Sims.

"You have a couple of lines of code, and those drive how your agents act, how they go about their day," says Elizabeth Hunter, who works on models of disease transmission at Technological University Dublin.

Agent-based models build the same kinds of virtual world as the equation-based ones, but each person can behave differently on a given day or in an identical situation. "These very specific models are extremely data hungry," says Kathleen O'Reilly, an epidemiologist at the London School of Hygiene and Tropical Medicine (LSHTM). "You need to collect information on households, how individuals travel to work and what they do at the weekend."

Which model to choose?

The Imperial team has used both agent-based and equation-based models in this pandemic. The 16 March simulations that the team ran to inform the UK government's COVID-19 response used an agent-based model built in 2005 to see what would happen in Thailand if H5N1 avian flu mutated to a version that could spread easily between people[2]. Ferguson told *Nature* in 2005 that collecting detailed data on Thailand's population was harder than writing the programming code for the model. That code was not released when his team's projections on the coronavirus pandemic were first made public, but the team is working with Microsoft to tidy up the code and make it available, Ferguson says.

On 26 March, Ferguson's team released global projections of the impact of COVID-19 that use the simpler equation-based approach[3]. It divides people into four groups: S, E, I and R, where 'E' refers to those who have been exposed, but who are not yet infectious. "They give broadly similar overall numbers," says epidemiologist Azra Ghani, who is also in the Imperial group. For instance, the global projections suggest that, had the United States taken no action against the virus, it would have seen 2.18 million deaths. The earlier agent-based simulation, run with the same assumptions on mortality rate and reproduction number, estimated 2.2 million US deaths[1].

The different kinds of model have their own strengths and weaknesses, says Vittoria Colizza, a modeller at the Pierre Louis Institute of Epidemiology and Public Health in Paris, who is advising the French government during the current emergency. Being able to bunch one group into a compartment inside an equation-based model makes things simpler — and quicker — because the model doesn't need to run at the high-resolution level of treating everyone as an individual. When Colizza and her team wanted to test the effects on infection rates of compelling large parts of the French population to work from home, for example, she used an equation-based model. "We didn't need to track each individual separately and see if they were spending some time at work or some time at school," she says.

Although projections might not diverge wildly depending on the approach chosen, it's natural to wonder how reliable any of the simulations are. Unfortunately, during a pandemic it is hard to get data – such as on infection rates – against which to judge a model's projections.

"You can project forwards and then compare against what you get. But the problem is that our surveillance systems are pretty rubbish," says John Edmunds, who is a modeller at the LSHTM. "The total numbers of cases reported, is that accurate? No. Accurate anywhere? No."

"Forecasts made during an outbreak are rarely investigated during or after the event for their accuracy, and only recently have forecasters begun to make results, code, models and data available for retrospective analysis," Edmunds and his team noted last year in a paper[4] that assessed the performance of forecasts made in a 2014–15 Ebola outbreak in Sierra Leone. They found that it was possible to reliably predict the epidemic's course one or two weeks ahead of time, but no longer.

To minimize the impact of incomplete data and incorrect assumptions, modellers typically carry out hundreds of separate runs, with the input parameters tweaked slightly each time. This 'sensitivity analysis' tries to prevent model outputs swinging wildly when a single input changes. And to avoid too much reliance on one model, Ferguson says, the UK government took advice from a number of modelling groups, including teams at Imperial and the LSHTM (see, for example, ref. 5). "We all reached similar conclusions," he says.

Updating the simulation

Media reports have suggested that an update to the Imperial team's model in early March was a critical factor in jolting the UK government into changing its policy on the pandemic. The researchers initially estimated that 15% of hospital cases would need to be treated in an intensive-care unit (ICU), but then updated that to 30%, a figure used in the first public release of their work on 16 March. That model showed the UK health service, with just over 4,000 ICU beds, would be overwhelmed.

Government officials had previously talked up a theory of allowing the disease to spread while protecting the oldest in society, because large numbers of infected people would recover and provide herd immunity for the rest. But they changed their course on seeing the new figures, ordering social-distancing measures. Critics then asked why social distancing hadn't been discussed earlier, why widespread testing hadn't happened, and why modellers had even chosen the 15% figure, given that a January paper showed that more than 30% of a small group of people with COVID-19 in China needed treatment in ICUs[6].

Ferguson says the significance of the model update might have been exaggerated. Even before that, he says, models already indicated that COVID-19, if left entirely unmitigated, could kill in the order of half a million UK citizens over the next year and that ICUs would be stretched beyond capacity. Advisory teams

"We're building simplified representations of reality. Models are not crystal balls."

UK epidemiologist Neil Ferguson.

IMPERIAL COLLEGE LONDON

had discussed suppressing the pandemic by social distancing, but officials were worried that this would only lead to a bigger second outbreak later in the year. Widespread testing of the kind seen in South Korea was not considered; but, in part, says Ferguson, this was because Britain's health agency had told government advisers that it would not be able to scale up testing fast enough.

As for the Chinese data on ICUs, clinicians had looked at them, but noted that only half the cases seemed to need invasive mechanical ventilators; the others were given pressurized oxygen, so might not need an ICU bed. On the basis of this and their experience with viral pneumonia, clinicians had advised modellers that 15% was a better assumption.

The key update came the week before Ferguson briefed government officials at Downing Street. Clinicians who had been talking to horrified colleagues in Italy said that pressurized oxygen wasn't working well and that all 30% of the severe hospitalized cases would need invasive ventilation in an ICU. Ferguson says the updated models' mortality projections didn't change hugely, because many predicted deaths are likely to occur in the community rather than in hospitals. But the understanding of how health services would be overwhelmed, and the experience of Italy, led to a "sudden focusing of minds", he says: government officials swiftly pivoted to social-distancing measures.

Testing needed

As researchers discover more about the virus, they are updating many other key variables. In the 26 March report[3] on the global impact of COVID-19, the Imperial team revised its 16 March estimate of R_0 upwards to between 2.4 and 3.3; in a 30 March report[7] on the spread of the virus in 11 European countries, the researchers put it somewhere in the range of 3 to 4.7. But some crucial information remains hidden from the modellers. A reliable test to see who has been infected without showing symptoms would be a game changer for modellers, and might significantly alter the predicted path of the pandemic.

To stress the need for such a test, a team at the University of Oxford, UK, led by theoretical epidemiologist Sunetra Gupta, has suggested that the pattern of recorded UK deaths might fit a range of SIR models, including one that assumes millions of people have already been infected but haven't shown any symptoms[8]. Only tests that reveal such infections can show what's going on in reality.

If all countries adopt strategies of strict social distancing, testing and isolation of infected cases before their deaths reach 0.2 per 100,000 people per week, the Imperial team says, then the global death total from COVID-19 could be cut to less than 1.9 million by the end of the year. And the British response, Ferguson said on 25 March, makes him "reasonably confident" that total deaths in the United Kingdom will be held below 20,000.

Ferguson says that nationwide lockdowns across Europe are already working to reduce the transmission of SARS-CoV-2. But how long social distancing will have to stay in place is a big question for countries worried about their economies and the mental and physical health of their cooped-up citizens. Social distancing will reduce the spread of the virus for now, but lifting these measures might allow a second wave of the pandemic later on, an Imperial model has suggested[1] (see 'A second wave').

Ferguson says he hopes that, in practice, countries can follow the example of South Korea, which has managed to impose a less-rigid version of social distancing by rolling out high levels of testing and tracing the contacts of those infected. Only close monitoring of regions as they lift lockdown restrictions, as China's Hubei province is now doing, will provide modellers with the information needed to forecast the longer-term toll of the pandemic.

David Adam is a science journalist based in London. Additional reporting by Richard Van Noorden.

1. Ferguson, N. M. *et al*. Preprint at Spiral https://doi.org/10.25561/77482 (2020).
2. Ferguson, N. M. *et al*. *Nature* **437**, 209–214 (2005).
3. Walker, P. G. T. *et al*. Preprint at Spiral https://go.nature.com/2yqz47x (2020).
4. Funk, S. *et al*. *PLoS Comput. Biol*. **15**, e1006785 (2019).
5. Prem, K. *et al*. *Lancet Public Health* https://doi.org/10.1016/S2468-2667(20)30073-6 (2020).
6. Huang, C. *et al*. *Lancet* **395**, 497–506 (2020).
7. Flaxman, S. *et al*. Preprint at Spiral https://doi.org/10.25561/77731 (2020).
8. Lourenço, J. *et al*. Preprint at medRxiv https://doi.org/10.1101/2020.03.24.20042291 (2020).