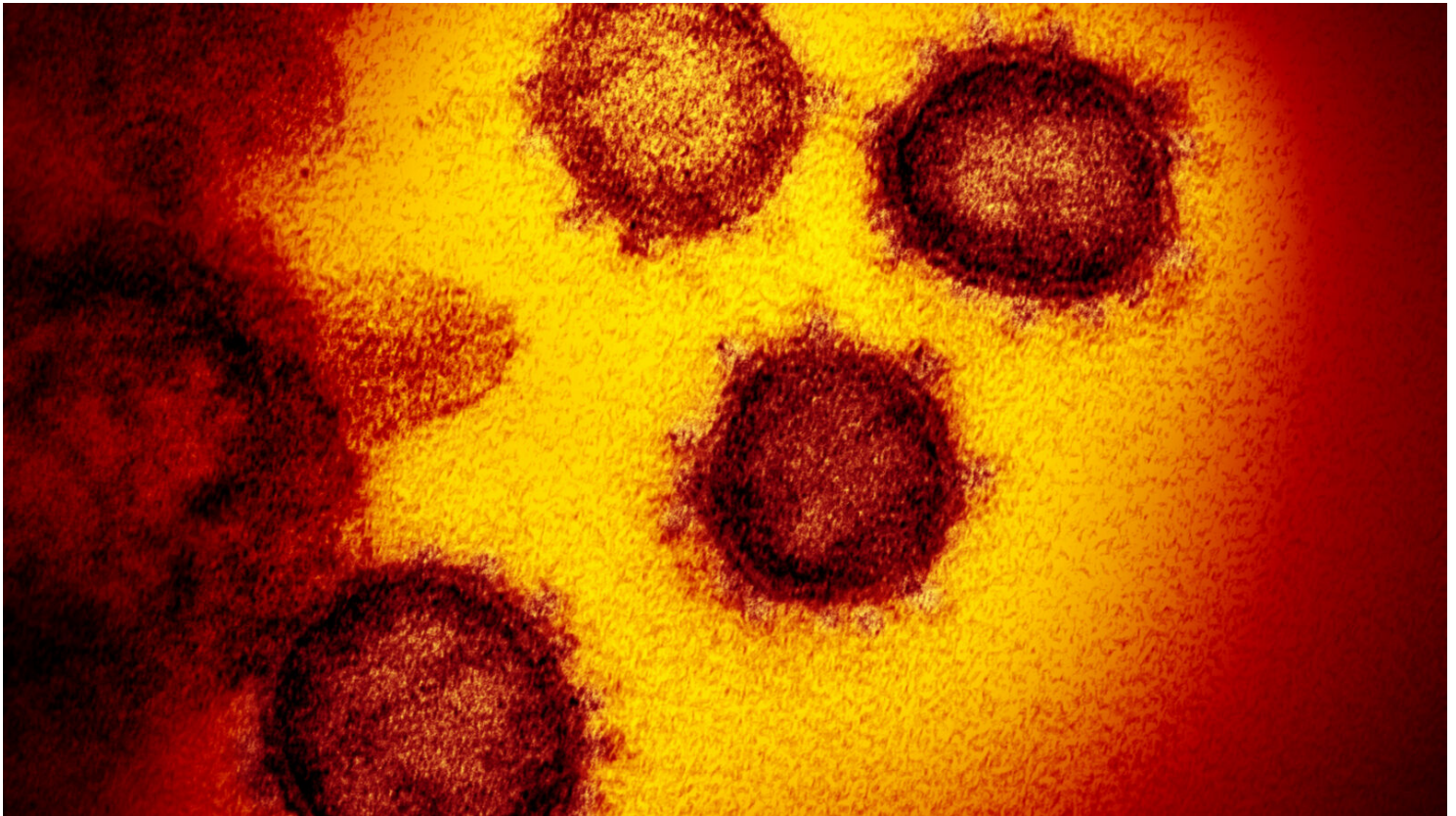


STAT¹

Disease modelers gaze into their computers to see the future of Covid-19, and it isn't good

By [Sharon Begley](#)² [@sxbegle](#)³

February 14, 2020



SARS-CoV-2, the virus that causes Covid-19. *Courtesy NIAID-RML*

At least 550,000 cases. Maybe 4.4 million. Or something in between.

Like weather forecasters, researchers who use mathematical equations to project how bad a disease outbreak might become are used to uncertainties and incomplete data, and [Covid-19](#)⁴, the disease caused by the new-to-humans coronavirus that began circulating in Wuhan, China, late last year, has those everywhere you look. That can make the mathematical models of outbreaks, with

their wide range of forecasts, seem like guesswork gussied up with differential equations; the eightfold difference in projected Covid-19 cases in Wuhan, [calculated](#)⁵ by a team from the U.S. and Canada, isn't unusual for the early weeks of an outbreak of a never-before-seen illness.

But infectious-disease models have been approximating reality better and better in recent years, thanks to a better understanding of everything from how germs behave to how much time people spend on buses.

“Year by year there have been improvements in forecasting models and the way they are combined to provide forecasts,” said physicist Alessandro Vespignani of Northeastern University, a leading infectious-disease modeler.

[Related:](#)⁶

[Experts envision two scenarios if the new coronavirus isn't contained](#)⁶

That's not to say there's not room for improvement. The key variables of most models are mostly the same ones epidemiologists have used for decades to predict the course of outbreaks. But with greater computer power now at their disposal, modelers are incorporating more fine-grained data to better reflect the reality of how people live their lives and interact in the modern world — from commuting to work to jetting around the world. These more detailed models can take weeks to spit out their conclusions, but they can better inform public health officials on the likely impact of disease-control measures.

Models are not intended to be scare machines, projecting worst-case possibilities. (Modelers prefer “project” to “predict,” to indicate that the outcomes they describe are predicated on numerous assumptions.) The idea is to calculate numerous what-ifs: What if schools and workplaces closed? What if public transit stopped? What if there were a 90% effective vaccine and half the population received it in a month?

“Our overarching goal is to minimize the spread and burden of infectious disease,” said Sara Del Valle, an applied mathematician and disease modeler at

Los Alamos National Laboratory. By calculating the effects of countermeasures such as social isolation, travel bans, vaccination, and using face masks, modelers can “understand what’s going on and inform policymakers,” she said. For instance, although many face masks are too porous to keep viral particles out (or in), their message of *possible contagion here!* “keeps people away from you” and reduces disease spread, Del Valle said. “I’m a fan of face masks.”

The clearest sign of the progress in modeling comes from flu forecasts in the U.S. Every year, about two dozen labs try to [model](#)⁷ the flu season, and have been coming ever closer to accurately forecasting its timing, peak, and short-term intensity. The U.S. Centers for Disease Control and Prevention determines which model did the best; for 2018-2019, it was [one](#)⁸ from Los Alamos.

Los Alamos also nailed the course of the 2003 outbreak of SARS in Toronto, [including](#)⁹ when it would peak. “And it was spot on in the number of people who would be infected,” said Del Valle: just under [400](#)¹⁰ in that city, of a global total of about 8,000.

The computers that run disease models grind through calculations that reflect researchers’ best estimates of factors that two Scottish researchers [identified](#)¹¹ a century ago as shaping the course of an outbreak: how many people are susceptible, how many are infectious, and how many are recovered (or dead) and presumably immune.

That sounds simple, but errors in any of those estimates can send a model wildly off course. In the autumn of 2014, modelers at CDC [projected](#)¹² that the Ebola outbreak in West Africa could reach 550,000 to 1.4 million cases in Liberia and Sierra Leone by late January if nothing changed. As it happened, heroic efforts to isolate patients, trace contacts, and stop unsafe burial practices kept the number of [cases](#)¹³ to 28,600 (and 11,325 deaths).

To calculate how people move from “susceptible” to “infectious” to “recovered,” modelers write equations that include such factors as the number of secondary infections each infected person typically causes and how long it takes from when

one person gets sick to when the people she infects does. “These two numbers define the growth rate of an epidemic,” Vespignani said.

The first number is called the basic reproduction number. Written R_0 (“R naught”), it varies by virus; a strain that spreads more easily through the air, as by aerosols rather than heavier droplets released when an infected person sneezes or coughs, has a higher R_0 . It has been a central focus of infectious disease experts in the current outbreak because a value above 1 portends sustained transmission. When the R_0 of Covid-19 was estimated several weeks ago to be above 2, social media exploded with “[pandemic](#)¹⁴ is coming!” hysteria.

But while important, worshipping at the shrine of R_0 “belies the complexity that two different pathogens can exhibit, even when they have the same R_0 ,” the Canadian-U.S. team argues in a paper posted to the preprint site medRxiv. Said senior author Antoine Allard of Laval University in Quebec, “the relation between R_0 , the risk of an epidemic, and its potential size becomes less straightforward, and sometimes counterintuitive in more realistic models.”

To make models more realistic, he and his colleagues argue, they should abandon the simplistic assumption that everyone has the same likelihood of getting sick from Covid-19 after coming in contact with someone already infected. For SARS, for instance, that likelihood clearly varied.

“Bodies may react differently to an infection, which in turn can facilitate or inhibit the transmission of the pathogen to others,” Allard said. “The behavioral component is also very important. Can you afford to stay at home a few days or do you go to work even if you are sick? How many people do you meet every day? Do you live alone? Do you commute by car or public transportation?”

When people’s chances of becoming infected vary, an outbreak is more likely to be eventually contained (by tracing contacts and isolating cases); it might reach a cumulative 550,000 cases in Wuhan, Allard and his colleagues concluded. If everyone has the same chance, as with flu (absent vaccination), the probability of containment is significantly lower and could reach 4.4 million there. Or as the researchers warn, “the outbreak almost certainly cannot be contained and we must

prepare for a pandemic”

[Related:](#) ¹⁵

[CDC director: More person-to-person coronavirus infections in U.S. likely, but containment still possible](#) ¹⁵

Modelers are also incorporating the time between when one person becomes ill and someone she infects does. If every case infects two people and that takes two days, then the epidemic doubles every two days. If every case infects two people and they get sick four days after the first, then the epidemic doubles every four days.

This “serial time” is related to how quickly a virus multiplies, and it can have a big effect. For a [study](#) ¹⁶ published this month in Annals of Internal Medicine, researchers at the University of Toronto created an [interactive tool](#) ¹⁷ that instantly updates projections based on different values of R_0 and serial interval.

Using an R_0 of 2.3 and serial interval of seven days, they project 300,000 cases by next week. If the serial interval is even one day less, the number of cases blasts past 1.5 million by then. But if the countermeasures that China introduced in January, including isolating patients, encouraging people to wear face masks, and of course quarantining Wuhan, reduce the effective reproduction number, as has almost certainly happened, those astronomical numbers would plummet: to 100,000 and 350,000 cases, respectively.

Just as public health officials care how long someone can be infected without showing symptoms (so they know how long to monitor people), so do modelers. “When people are exposed but not infected, they tend to travel and can’t be detected,” Vespignani said. “The more realistic you want your model to be, the more you should incorporate” the exposed-but-not-ill population. This “E” has lately become a fourth category in disease models, joining susceptible, infectious, and recovered.

At Los Alamos, Del Valle and her colleagues are using alternatives to the century-

old susceptible/infectious/recovered models in hopes of getting a more realistic picture of an outbreak's likely course. A bedrock assumption of the traditional models is "homogeneous mixing," Del Valle said, meaning everyone has an equal chance of encountering anyone. That isn't what happens in the real world, where people are more likely to encounter others of similar income, education, age, and even religion (church pews can get crowded).

"Ideally, you'd break the population into many groups" and estimate the likelihood of each one's members interacting with each other and with every kind of outsider, Del Valle said. "Your model would become more accurate."

Called "agent-based models," they simulate hypothetical individuals, sometimes tens of millions of them, as they go about their day. That requires knowing things like how many people commute from where to where for work or school, how they travel, where and how often they shop, whether it's customary to visit the sick, and other key details. Computers then simulate everyone's movements and interactions, for instance by starting with one infected person leaving home in the morning, chatting with other parents at school drop-off, continuing to work on a bus, standing 2 feet from customers and colleagues, and visiting a pharmacy for her migraine prescription.

The models keep track of people second by second, said Los Alamos computer scientist Geoff Fairchild, "and let you assess the impact of different decisions, like closing schools during flu season." (Some [research](#)¹⁸ shows that can dampen an outbreak.) Although "agent-based models can simulate reality better," he said, they are less widely used because they require enormous computing power. Even on the Los Alamos supercomputer, a single run of a complicated model can take days or even weeks — not counting the weeks of work modelers spend writing equations to feed the computer.

[Related:](#) ¹⁴

[**Understanding pandemics: What they mean, don't mean, and what comes next with the coronavirus**](#) ¹⁴

The Los Alamos researchers are still wrestling with their Covid-19 model, which is showing — incorrectly — the outbreak “exploding quite quickly in China,” Del Valle said. It is overestimating how many susceptible people become infected, probably because it’s not accurately accounting for social isolation and other countermeasures. Those seem to have reduced R_0 toward the lower range of 2-to-5 that most modelers are using, she said.

In the current outbreak, researchers are building models not only to peek into the future but also to reality-check the present. Working backwards from confirmed infections in countries other than mainland China, researchers at Imperial College London who advise the World Health Organization [estimated](#)¹⁹ that Wuhan had 1,000 to 9,700 symptomatic cases as of Jan. 18. Three days later, all of mainland China had officially reported 440 cases, supporting the concerns of global health officials that China was undercounting.

In a more recent model run, Jonathan Read of England’s University of Lancaster and his colleagues [estimated](#)²⁰ “that only about 1 in 20 infections were being detected” in late January, Read said: There were probably 11,090 to 33,490 infections in Wuhan as of Jan. 22, when China reported 547 cases. “It highlights how difficult it is to track down and identify this virus,” Read said, especially with residents of quarantined Wuhan being turned away from overwhelmed hospitals and clinics without being tested for the virus. Using a similar approach, modelers led by Dr. Wai-Kit Ming of Jinan University in Guangzhou [estimated](#)²¹ that through Jan. 31, China probably had 88,000 cases, not the 11,200 reported.

Read’s group is updating its model to estimate the fraction of true cases in February; China’s cumulative cases topped 60,000 on Thursday.

For modelers, a huge undercount can corrupt the data they base their equations on. But even with that disadvantage the Covid-19 models “are doing quite well, despite a lot of complicated dynamics on the ground,” said Los Alamos’s Fairchild. While it’s not clear yet if they’ve nailed the true numbers of cases, they are correctly projecting the outbreak’s basic shape: increasing exponentially, the number of cases growing more quickly the more cases there are.

About the Author



[Sharon Begley](#)²

Senior Writer, Science and Discovery

Sharon covers science and discovery.

sharon.begley@statnews.com²²
[@sxbegle](#)³

Links

1. <https://www.statnews.com/category/health/>
2. <https://www.statnews.com/staff/sharon-begley/>
3. <https://twitter.com/sxbegle>
4. <https://www.statnews.com/2020/02/11/disease-caused-by-the-novel-coronavirus-has-name-covid-19/>
5. <https://www.medrxiv.org/content/10.1101/2020.02.10.20021725v1.full.pdf>
6. <https://www.statnews.com/2020/02/04/two-scenarios-if-new-coronavirus-isnt-contained/>
7. <https://www.cdc.gov/flu/weekly/flusight/index.html>
8. <https://www.lanl.gov/discover/news-release-archive/2019/October/1022-flu-forecasting-challenge.php>
9. <https://www.sciencedirect.com/science/article/abs/pii/S0022519303002285>
10. <https://www.ncbi.nlm.nih.gov/books/NBK92467/>
11. <https://royalsocietypublishing.org/doi/10.1098/rspa.1927.0118>
12. <https://www.cdc.gov/mmwr/volumes/65/su/su6503a12.htm>
13. <https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html>
14. <https://www.statnews.com/2020/02/12/understanding-pandemics-what-they-mean-coronavirus/>
15. <https://www.statnews.com/2020/02/12/cdc-director-more-person-to-person-coronavirus-infections-in-u-s-likely-but-containment-still-possible/>
16. <https://annals.org/aim/fullarticle/2760912/reporting-epidemic-growth-reproduction-numbers-2019-novel-coronavirus-2019-ncov>
17. https://art-bd.shinyapps.io/nCov_control/
18. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004681>
19. <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/2019-nCoV-outbreak-report-22-01-2020.pdf>
20. <https://www.medrxiv.org/content/10.1101/2020.01.23.20018549v2>
21. <https://www.biorxiv.org/content/10.1101/2020.01.27.922443v2>
22. <https://www.statnews.com/2020/02/14/disease-modelers-see-future-of-covid-19/mailto:sharon.begley@statnews.com>

23. <https://www.statnews.com/tag/china/>
24. <https://www.statnews.com/tag/coronavirus/>
25. <https://www.statnews.com/tag/global-health/>
26. <https://www.statnews.com/tag/public-health/>
27. <https://www.statnews.com/article-report-license-request/>

© 2020 STAT