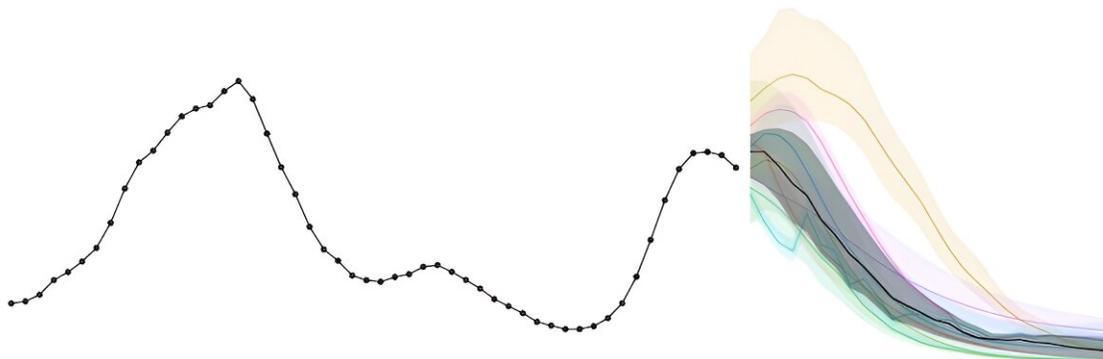# Pooling multiple models during COVID-19 pandemic provides more reliable projections about an uncertain future: Study

November 20 2023, by Emily Howerton, Cecile Viboud and Justin Lessler

Collecting projections from multiple independent models provides a fuller picture of possible futures – as in this graph of potential hospitalizations – and allows researchers to generate an ensemble. Credit: COVID-19 Scenario Modeling Hub, CC BY-ND

How can anyone decide on the best course of action in a world full of unknowns? There are few better examples of this challenge than the COVID-19 pandemic, when officials fervently compared potential outcomes as they weighed options like whether to implement lockdowns or require masks in schools. The main tools they used to compare these futures were epidemic models.

But often, models included numerous unstated assumptions and considered only one scenario—for instance, [that lockdowns would continue](link). Chosen scenarios were rarely consistent across models. All this variability made it difficult to compare models, because it's unclear whether the differences between them were due to different starting assumptions or scientific disagreement.

In response, we came together with colleagues to found the [U.S. COVID-19 Scenario Modeling Hub](link) in December 2020. We provide [real-time](link), long-term projections in the U.S. for use by [federal agencies](link) such as the Centers for Disease Control and Prevention, local health authorities and the public.

We work directly with public health officials to identify which possible futures, or scenarios, would be most helpful to consider as they set policy, and we convene multiple independent modeling teams to make projections of public health outcomes for each scenario. Crucially, having multiple teams address the same question allows us to better envision what could possibly happen in the future.

Since its inception, the Scenario Modeling Hub has generated 17 rounds of projections of COVID-19 cases, hospitalizations, and deaths in the U.S. across varying stages of the pandemic. In a recent study published in the journal *Nature Communications*, we looked back at all these projections and [evaluated how well they matched the reality](link) that unfolded. This work provided insights about when and what kinds of model projections are most trustworthy—and, most importantly, supported our strategy of combining multiple models into one ensemble.

Multiple models are better than just one

A founding principle of our Scenario Modeling Hub is that multiple models are more reliable than one.

From tomorrow's temperature on your weather app to predictions of interest rates in the next few months, you will likely use the combined results of multiple models all the time. Especially in times like the COVID-19 pandemic when uncertainty abounds, combining projections from multiple models into an ensemble provides a fuller picture of what could happen in the future. Ensembles have become ubiquitous in many fields, primarily because they work.

Our analysis of this approach with COVID-19 models resoundingly showed the strong performance of the Scenario Modeling Hub ensemble. Not only did the ensemble give us more accurate predictions of what could happen in the future overall, it was substantially more consistent than any individual model throughout the different stages of the pandemic. When one model failed, another performed well, and by taking into account results from all of these varying models, the ensemble emerged as more accurate and more reliable.

Researchers have previously shown performance benefits of ensembles for short-term forecasts of influenza, dengue and SARS-CoV-2. But our recent study is one of the first times researchers have tested this effect for long-term projections of alternative scenarios.

A combination of many models is most useful

The Scenario Modeling Hub ran 16 different rounds of COVID-19 models over two years of the pandemic, each trying to project what the future would hold. Different models, named A through J, performed better or worse over the various rounds. But the ensemble built from multiple models was more accurate and more reliable over time.

Better Average Worse

Round	Ensemble	A	B	C	D	E	F	G	H
1	**1.1**	0.5	1.8					1.4	0.8
2	**0.5**	0.7	2.1	0.3				4.7	0.8
3	**0.8**	1.6	0.8	1.4					0.7
4	**0.6**	1.5	0.6	0.4	0.9			15.2	0.3
5	**0.5**	0.8	1.1	3.0	1.1			0.6	0.8
6	**0.9**	1.2	0.8	1.2	0.8	1.0		1.0	1.0
7	**0.6**	0.9	0.6	0.7	1.1	1.3		2.4	0.8
9	**0.7**	1.1	0.6	2.0	0.9	1.6			1.0
11	**0.4**	2.5	0.7	1.4		2.2			0.7
12	**0.4**	2.4	1.1	0.4		2.2			0.9
13	**0.8**	0.9	0.8	0.9		1.4			1.4
14	**0.5**	1.7	0.7	1.5			1.2		1.3
15	**0.5**	0.8	1.1	0.8			1.3		1.5
16	**0.4**	1.8	0.7	0.7			1.6		1.6
Overall	**0.6**	1.1	0.7	0.9	0.9	1.3	1.3	2.3	0.9

A rating of 1 means the model performed "on average" compared to all the other models, with a lower number being better than average and a higher number being worse.

Table: The Conversation, CC-BY-ND • Source: Howerton et al, Nature Communications 2023 • Created with Datawrapper

Credit: The Conversation

A 'hub' makes multimodel projections possible

While scientists know combining multiple models into an ensemble improves predictions, it can be tricky to put an ensemble together. For example, in order for an ensemble to be meaningful, model outputs and key assumptions need to be standardized. If one model assumes a new COVID-19 variant will gain steam and another model does not, they will come up with vastly different results. Likewise, a model that projects cases and one that projects hospitalizations would not provide comparable results.

Many of these challenges are overcome by convening as a "hub." Our modeling teams meet weekly to make sure we're all on the same page about the scenarios we model. This way, any differences in what individual models project are the result of things researchers truly do not know. Retaining this scientific disagreement is essential; the success of the Scenario Modeling Hub ensemble arises because each modeling team takes a different approach.

At our hub we work together to design our scenarios strategically and in close collaboration with public health officials. By projecting outcomes under specific scenarios, we can estimate the impact of particular interventions, like vaccination.

For example, a scenario with higher vaccine uptake can be compared with a scenario with current vaccination rates to understand how many lives could potentially be saved. Our projections have informed recommendations of COVID-19 vaccines for children and bivalent boosters for all age groups, both in 2022 and 2023.

In other cases, we design scenarios to explore the effects of important unknowns, such as the impact of a new variant—known or hypothetical. These types of scenarios can help individuals and institutions know what they might be up against in the future and plan accordingly.

Although the hub process requires substantial time and resources, our results showed that the effort has clear payoffs: The information we generate together is more reliable than the information we could generate alone.

Past reliability, confidence for future

Because Scenario Modeling Hub projections can inform real public health decisions, it is essential that we provide the best possible information. Holding ourselves accountable in retrospective evaluations not only allows us to identify places where the models and the scenarios can be improved but also helps us build trust with the people who rely on our projections.

Our hub has expanded to produce scenario projections for influenza, and we are introducing projections of respiratory syncytial virus, or RSV. And encouragingly, other groups abroad, particularly in the EU, are replicating our setup.

Scientists around the world can take the hub-based approach that we've shown improves reliability during the COVID-19 pandemic and use it to support a comprehensive public health response to important pathogen threats.

Provided by The Conversation