

## AI-ENHANCED PREDICTIVE MODELS TO COMBAT THE NEXT COVID WAVE



Existent models to track the virus have been based on hazy assumptions and flawed data. With the threat of a second wave of Covid-19 imminent, decision makers need to leverage a new model based on AI to manage the virus going forward. Such a continuously learning model would be based on real-time, robust data, providing governments with a soft and directed plan to reopen the economy.

## How quickly should things open up?

Critical information is needed to decide how countries should restart their economies. Robust data on the ease of transmission of the virus, extent of its development in new locations, potential for mutation and real-time identification of potential hot spots are all needed.

However, the data science models that have been developed thus far suffer from data points that are constantly evolving and sometimes unreliable. <sup>1</sup> Many early statistical models were developed in a rush <sup>2</sup> and flawed by data based on limited testing capacity. <sup>3/4</sup>

A different approach is needed, one built on cutting-edge artificial intelligence that is adaptive, built to scale and automated. This model would be superior to current models, because it requires less data to learn and predict the path of the virus when a second wave occurs. The hope is that public sector leaders could use this model to predict COVID-19 dynamics more effectively, formulate better pandemic control procedures and devise a robust exit strategy for the current wave and beyond.

## Why is the current data failing us?

Good data is needed to track and respond to a pandemic of this magnitude. This data must be accurate and statistically significant. Things such as fatality rates, R-values, virus transmission rates and herd immunity levels have so far been impossible to estimate across geographies. Reliable data at the beginning of an outbreak caused by a new virus is rare, and the predictions of these values have been uncertain. <sup>5</sup>

Good observational data is also needed. This has been difficult in the

U.S. and in other countries, where mass testing has been challenging. Another problem is the difference in sensitivity and specificity rates caused by a variance in immunodiagnostic test types and specimen collection techniques. With such high variability, it is difficult to understand both the number of people infected and those who have already contracted the virus and recovered.

Further, antibodies that are used to detect a past infection can also be generated by other forms of coronavirus, including the common cold and seasonal influenza.

With these uncertainties, it is very difficult to determine who is immune, who is asymptomatic or who has mild symptoms. Some people may be symptomatic but have yet to be tested. These carriers could easily spark a new wave of the pandemic if things are opened up too soon.

It is also imperative to understand who is especially vulnerable to the virus. New data shows that obesity, diabetes, heart disease and vitamin D deficiency make things worse. However, the jury is still out on whether transmission has any seasonal influence or weather dependency. Current models working on so many unknown variables are highly likely to be problematic.

## The problem with the current models

Existent computational models to track the virus have been based on insufficient data and hazy assumptions.

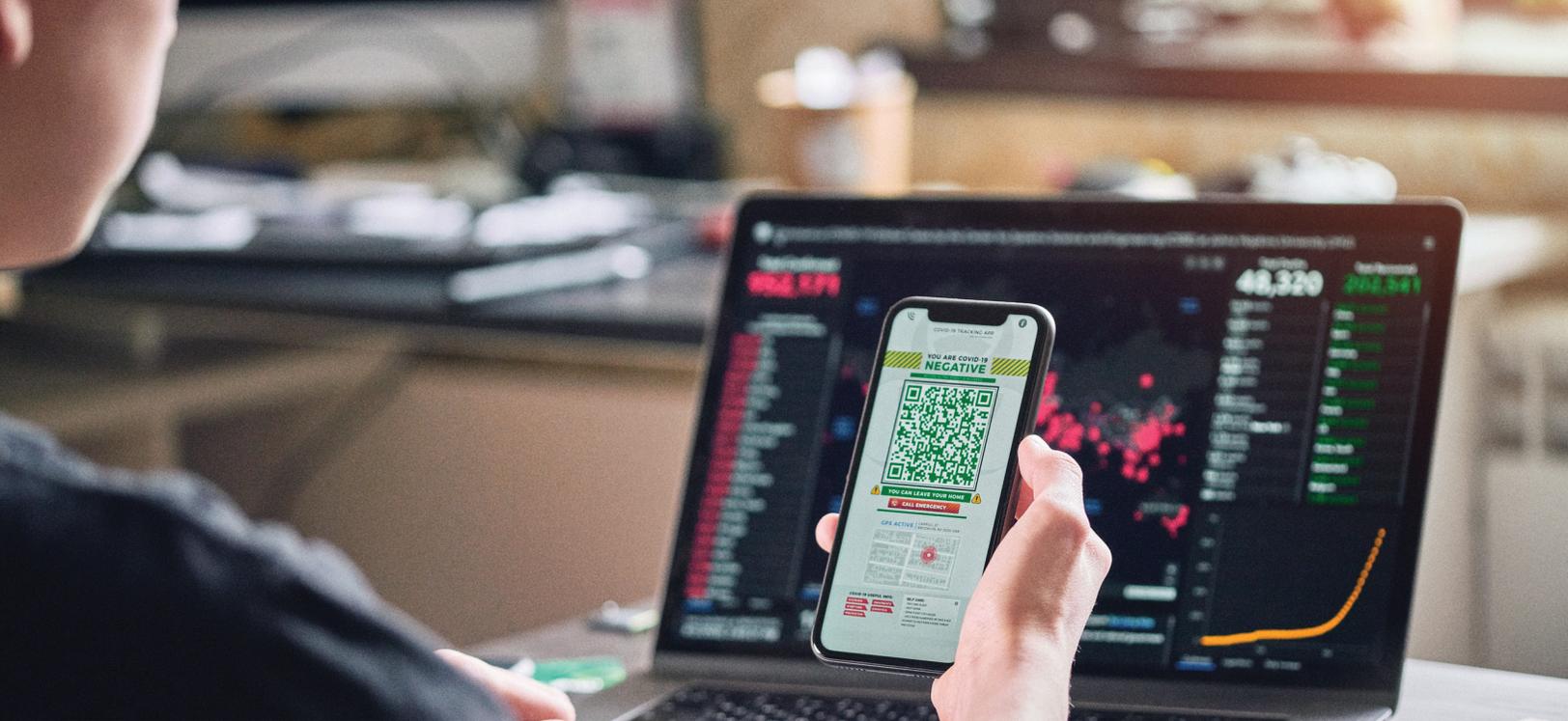
A model built by the Institute for Health Metrics and Evaluation — and used by the White House — assumed that the progression of the pandemic would follow in the footsteps of that seen in China, Spain and Italy, among other countries. But it failed to take into account any differences in key regional parameters.

Population characteristics, availability and variance in COVID-19 testing, access to critical care facilities, levels of quarantine, and the date when social distancing came into effect were all based on what was seen elsewhere in the world. Such a nonmechanistic model suffers from the fallacy of Farr's law, <sup>6</sup> which says that all epidemics are symmetrical and scale uniformly. No wonder, then, that the IHME's projections led to confusion with policymakers and the wider public in the U.S. <sup>7</sup>

Many existing models are based on faulty data and fail to take into account key regional parameters

Some universities took a different route. These early models incorporated estimates of contagion, transmission and R-values, along with factors that increase the risk of serious illness or death. They also integrated the time frame from infection to actual clinical recovery. This working model is known as SEIR, which categorizes the population into four buckets (susceptible, exposed, infected, resistant). Though more precise than IHME-like empirical models, these models don't factor in evolving knowledge of the virus and risk to the population. They also underestimate the spread of the disease as the influence of population behavior, local environmental factors and political decisions are not factored in. Moreover, these models don't clearly mention key assumptions that have been included in their genesis. Inaccurate assumptions can lead to errors in the working model. <sup>8</sup>

The Imperial College London model that predicted high infection and fatality rates actually failed to infer the obvious change in population behavior that would still arise even in the absence of government-mandated interventions. It also lacked



an understanding of how the virus reproduction ( $R_0$ ) number would change due to this behavior.

According to the University of Minnesota, other models that were based on other coronaviruses — such as SARS-CoV-1 (which causes severe acute respiratory syndrome) and Middle East Respiratory Syndrome coronavirus, known as MERS-CoV — have similarly failed to provide useful guidance on what to predict from COVID-19.<sup>9</sup> The most recent UMass Amherst's ensemble type model (commonly used in weather predictions) triangulates a comparative prediction from multiple models. This approach may incorporate the uncertainty and errors already integrated into existing models it has referenced to develop their prediction hypothesis.

## Modeling the right exit strategy

What all this means is that there is neither the right data nor the right model to tell governments when and how they should open up. This has the knock-on effect of triggering

social confusion and destabilizing markets. Even now, some leaders have moved ahead by easing controls without much preparation for the potential fallout.

An AI model will enable leaders to plan ahead based on fail-safe, real-time data

The World Health Organization warns that new outbreaks are likely to occur as a result of abruptly ending lockdowns around the world.<sup>10</sup> Some experts say that a new wave could hit in late fall or early winter of 2020, or even sooner, depending on how reopening is handled now.

In such indeterminate times, a new model is needed, one that helps government and task force leaders to determine the plan ahead based on fail-safe, real-time data. This model (and the plan it determines) should factor in the likelihood of a second wave and where that wave is likely to be. This will inform businesses and governments regarding how and where to procure essential supplies,

such as ventilators and personal protective equipment. It will also tell governments how to manage successive hospitalizations and ICU admissions while providing guidance on how the economy can return to a semblance of normalcy with less economic disruption should a second wave occur.

Of course, some models have been more successful than others in the first instance. In New Zealand, to get the economy back up and running, good recommendations have been made regarding increased testing, social distancing and protection of vulnerable people. But on a wider scale, decision-making at a more granular level is missing. For example, should lockdowns continue in certain locations in order to protect vulnerable populations? How should the population be stratified to determine who emerges from the lockdown first and what social distancing measures will be required in the future in order to curb the pandemic below the health systems' capacity? Finally, what strategy is needed to implement contact tracing at scale and ensure health care is sufficient should another wave occur?

## An AI model to navigate the next wave

Public health decision-makers need better information and better insights into early warning indicators to navigate the various post-lockdown scenarios. Many of the strongest new models are under scientific scrutiny.

But they could do better, overcoming some of the shortcomings of earlier models.

One idea is to adopt AI, where artificial neural networks and deep learning techniques could augment the existing epidemiological model, making it more dynamic and more responsive in real-time. This AI model would be adaptive, built to scale, automated and use “semi-supervised” or unsupervised learning. More comprehensive results would still accrue from such a model, even without a universally accepted and large-scale COVID-19 testing report being immediately available to make a reliable denominator.<sup>11</sup>

This AI model would be self-sustaining. It would require a reduced amount of data to learn and predict compared with many current models (which have long learning curves and demand that all training data is rigorously and correctly labeled — a difficult feat because of limited testing and the underreporting of COVID-19 deaths outside hospitals). The AI system would feature continuous adjustment to input parameters, would continuously learn and wouldn't suffer from the inevitable “adjustment delay” of current models. Further, with deep learning, AI could discover complex patterns, auto-detect anomalies, and self-learn and self-heal automatically. It would also be able to judge the accuracy of the variables (e.g., variances in test results), producing much more reliable results than anything we have currently seen.

Because the spread of the disease is different in different places, the

key parameters in the AI model would include regional population characteristics such as age distribution, socioeconomic status, population risk factors (e.g., smoking, obesity, drug dependencies) and the percentage of older adults with comorbidities. It would also include regional and environmental factors such as population size, density, individual mobility and social-distancing effects.

Further, the model will look beyond the assumption that every individual in a population subset has the same chance of catching the infection. Instead, the input data would be considered — by population's socio-economic, demographic, education level, unhealthy habits and health status. The number of infected individuals who quarantined and no longer could spread the infection would be incorporated into the model as well.

AI would enable each government to implement a soft and directed plan to reopen the economy

Where could we get this data? The pipeline would consist of large regional data sets similar to what Medicare and Medicaid collect and include an individual's demographic details, along with clinical variables such as COVID-19 test reports and clinical data (treatment records, health risks and lab results). All this should be augmented with social determinants of health and integrated with epidemiological factors like infection rate over time, contact rate, case fatality reports and hygiene factors (hand washing, face mask adoption, social distancing).

Developed with these guiding principles, this AI model would provide realistic insights into the progression of the disease. It would

also model individual behavior in response to the changing effects of the disease. Such insights would then be used to guide task force leaders to proactively detect, track and quarantine COVID-19-vulnerable individuals to keep infection transmission at a manageable level. Through early prediction of the virus epicenter and disease hot spot trends, the model would help in the optimal allocation of PPE. It would also ensure regions are ready when a second wave occurs. Instead of a harsh and wide quarantine policy, this sort of AI-driven intelligence would enable each government to implement a soft and directed plan for reopening their economy.

## A well-informed exit strategy

COVID-19 is an unknown entity and more dangerous than anything we've seen since the Spanish flu.<sup>12</sup>

Until governments have standard testing mechanisms, better test coverage and a proven treatment regime, models need to be used to stay on top of the situation and make data-based decisions. Such decisions need to be unbiased, consistent and realistic. Unfortunately, most existing models do not provide granular guidance on how to manage the pandemic going forward. Now is the time to adopt a new deep-learning AI model built with an ideal blend of epidemiology, bioinformatics and health economics. Such a model would provide more statistically relevant forecasts on the actual nature and course of the virus. If developed quickly, such intelligence would be a boon for governments, enabling them to make rational decisions in “next to real time,” mitigating the effect of the pandemic and providing the economy with a robust and well-informed exit strategy.

## References

- <sup>1</sup> [Who Is John Ioannidis?](#), Edward Peter Stringham, April 19, 2020, American Institute for Economic Research
- <sup>2</sup> [Special report: The simulations driving the world's response to COVID-19](#), David Adam, April 2, 2020, Nature.com
- <sup>3</sup> [COVID-19 Testing: Challenges, Limitations and Suggestions for Improvement](#), Hu, E., April 9, 2020, Preprints.org
- <sup>4</sup> [Covid-19 testing: overcoming challenges in the next phase of the epidemic](#), Arjun K. Manrai & Kenneth D. Mandl, March 31, 2020, Stat news
- <sup>5</sup> [Forecasting the novel coronavirus COVID-19](#), Fotios Petropoulos & Spyros Makridakis, March 31, 2020, Plos One
- <sup>6</sup> [Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic](#), Nicholas P. Jewell & Joseph A. Lewnard & Britta L. Jewell, April 14, 2020, Annals of Internal Medicine
- <sup>7</sup> [Influential Covid-19 model uses flawed methods and shouldn't guide U.S. policies, critics say](#), Sharon Begley, April 17, 2020, Stat news
- <sup>8</sup> [Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections](#), Nicholas P. Jewell & Joseph A. Lewnard & Britta L. Jewell, April 16, 2020, JAMA
- <sup>9</sup> [COVID-19: The CIDRAP Viewpoint](#), April 30, 2020, CIDRAP
- <sup>10</sup> [Coronavirus: 'Deadly resurgence' if curbs lifted too early, WHO warns](#), April 10, 2020, BBC News
- <sup>11</sup> [Denominator matters in estimating COVID-19 mortality rates](#), Bamba Gaye & Anouar Fanidi & Xavier Jouven, April 7, 2020, European Heart Journal
- <sup>12</sup> [Experts predict 'significant COVID-19 activity' in US for up to 2 years](#), Amanda Woods, May 1, 2020, New York Post

## Authors

### Eric Paternoster

Senior Vice President and CEO –  
Infosys Public Services  
eric\_paternoster@infosys.com

### Dr. Suman De

Head of Government Healthcare Analytics –  
Infosys Public Services  
suman.de@infosys.com

---

## About Infosys Knowledge Institute

The Infosys Knowledge Institute helps industry leaders develop a deeper understanding of business and technology trends through compelling thought leadership. Our researchers and subject matter experts provide a fact base that aids decision making on critical business and technology issues.

To view our research, visit Infosys Knowledge Institute at [infosys.com/IKI](https://infosys.com/IKI)

---

For more information, contact [askus@infosys.com](mailto:askus@infosys.com)



---

© 2020 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.