

Why Predictive Modeling is Critical in the Fight against COVID-19

| DIGITAL TRANSFORMATION TOOLKIT
KNOWLEDGE TOOLS

20

PAHO



Pan American
Health
Organization



World Health
Organization
REGIONAL OFFICE FOR THE
Americas

Why Predictive Modeling is Critical in the Fight against COVID-19?

IMPORTANT NOTE: Stay informed with timely information on the coronavirus disease (COVID-19), available on the [PAHO](#) and [WHO](#) websites and through your national and local public health authorities

Introduction. A number of predictive models and forecasting exercises have been developed by various organizations, such as research groups, academic institutions, hospitals, and consulting companies, with the main aim to support health systems in with COVID-19 strategic decision making, planning, and health policy formulation that help in the fight against COVID-19. Predictive models are helpful for estimating the number of COVID-19 cases and deaths; the resources required, e.g., such as hospital patient beds and ICU beds; and the demand for supplies, such as personal protective equipment (PPE). Because predictive models for COVID-19 must rely on a rapidly changing situation and underlying data, they produce results that may change repeatedly as data are updated and revised. Nevertheless, the predictive models are meaningful and can offer crucial insights to policymakers. It is important that we understand the strengths and weaknesses of predictive models in order to use them judiciously as support and reference tools for COVID-19 planning and action.

What is predictive analytics?

Predictive analytics is statistical analysis that uses data mining, machine learning, and algorithms based on historical data series to identify behavior patterns and trends to predict future scenarios. Although predictive analytics is a well-known analytical method, recently it has been enhanced by the availability of large data resources or Big Data, increased computational capacity, and modern analytical mechanisms.¹

Why is predictive analytics important in the fight against COVID-19?

Predictive analytics allows us to estimate the pandemic's behavior within an acceptable degree of uncertainty by establishing when and under which conditions countries can expect increases, peaks, and reductions in new cases (incidence) and deaths (mortality). With this information, we can calculate the demand for acute medical services; determine the timeframes for partially or completely lifting containment measures (i.e., lockdowns); and even predict renewed needs for subsequent waves of the pandemic. Estimating health care demand allows planning for required health technologies (PPE, ventilators, etc.) to ensure adequate end-to-end supply chain and distribution, and managing human resources for an appropriate and timely response.

What is modeling and how can it inform COVID-19 intervention policy decisions?

Modeling is a process within predictive analytics that allows us to perform single or comparative analyses of relevant approaches to a problem, considering benefits and

¹ <https://ieeexplore.ieee.org/document/8612393/>

limitations, and aiming to make the best course of action most explicit. Mathematical models are widely used to inform policymaking in public health because they can pretest the relative effectiveness of various public health strategies (e.g., types of health care services or the number and expertise of human resources). Models can also be used to explore how different public health strategies (e.g., lifting of restrictions, lockdown) may impact the pandemic's trajectory of the pandemic among various groups of people (e.g., gender, age, etc.). Irrespective of the quality of available data, policymakers will need to make important decisions that will have substantial implications for COVID-19 transmission and mortality. Predictive models can evaluate different “what if” scenarios and guide decision-making.

What is machine learning and why is it important in the fight against COVID-19?

Machine learning (ML) is a subset of artificial intelligence (AI); both are tools that can increase the power of epidemiological and decision-making models within public health. ML uses computers to automate the discovery of patterns in very large datasets. The computer is said to be “learning” because it is programmed to improve its own performance.² Several ML approaches have been identified to help in the COVID-19 pandemic response: early detection of symptoms and risk factors, improved diagnosis of atypical cases, monitoring of existing cases and new outbreaks in different populations, contact tracing of people with confirmed or suspected infection, as well as accelerating vaccine development and identifying new treatments or new uses for existing treatments for other illnesses.³

To date, ML and AI have not helped to predict how COVID-19 will spread, for two reasons: 1. AI requires substantial amounts of data for algorithm “training” and there is little historical COVID-19 clinical data available to date; 2. There are pitfalls for Big Data and ML in the context of infectious diseases, e.g., cleaning “messy” data harvested from social media. Currently, most models used for forecasting and tracking do not use ML methods, but rather, established epidemiological models. These are mainly *mathematical models* that have distributional assumptions and try to predict forward in time or *stochastic “epidemiological” models* that use basic parameters to explore the spread of disease through an artificially constructed population.

What are the main differences between predictive modeling and machine learning?

Predictive modeling seeks to estimate a certain, specific result within a closer time range. ML is a methodology of great use in analytical predictions⁴ and can be applied to both types of models to improve forecast and prediction accuracy. Both approaches assist in forecasting scenarios since both estimate possible future events over broad periods of time by analyzing past and present data.

What should policymakers expect to see reported in predictive models for COVID-19?

- Precise statement of the problem and the data used (parameters, inputs, and outputs)
- Sufficient detail on data quality, especially determining the data
- Explicit information on the modeling methodology (identifying the appropriate ML, data mining system, and other data analysis processes)
- Determination of whether or not the prediction process met the expected objectives⁵

² https://doi.org/10.1007/978-1-4842-3069-5_4

³ <https://www.sciencedirect.com/science/article/pii/S1871402120300771>

⁴ <https://powerpivotpro.com/2017/08/difference-forecasting-predictive-analytics/> | <https://www.onemodel.co/blog/ai-academy-forecasting-vs-predictive-modeling/>

⁵ <https://econsultancy.com/predictive-analytics-four-prerequisites-of-an-effective-strategy/>

What are the **main limitations** of predictive models?

There is an inherent **uncertainty** to predictive modeling that can be introduced at any step of the model-building process. It is difficult to exactly quantify the amount of uncertainty in a model, but there are specific types of uncertainties that can affect a model's performance. From the onset, the model *structure* itself and the set of assumptions under which the model is built can contribute to its uncertainty (e.g., what diseases will be included/excluded in the analysis?). *Methodological* uncertainty can be introduced when defining the scope of the analysis (e.g., time horizon). Units of analysis, such as population or patient subgroups, can have a broad degree of *heterogeneity*, which requires different sets of assumptions to be applied for each group. The degree of accuracy around a statistical estimate and how much deviation from that estimate is the *parameter uncertainty* (statistical error). Finally, the model's results can be evaluated by the ability to generalize them to the wider target population (*generalizability*). Performing "**sensitivity analyses**" is key to better understanding uncertainty. It is a technique used in statistics to evaluate the impact that a particular dependent variable, under a given set of assumptions, might have on the overall result. Helpful predictive models *must undergo* thorough deterministic and probabilistic sensitivity analyses. Uncertainty can be reduced by increasing the sample size and improving the quality of the data used to populate the model.

Where can I **find out more** about COVID-19 modeling and forecasting?

- <https://covid19.healthdata.org/>
- <https://www.covid19sim.org/>
- <https://www.covidanalytics.io/>

Contact information

Tel.: +1 (202) 974-3399 • Correo electrónico: eoc@paho.org

PAHO highlights and appreciates the support of the Spanish Agency for International Development Cooperation (AECID), the United States Agency for International Development (USAID), and the Government of Canada.

Acknowledgments

This factsheet was prepared in collaboration with the following institutions:

- Inter-American Development Bank (IDB) Social Protection and Health Division
- Center for Health Informatics, School of Information Sciences, Grainger College of Engineering, The University of Illinois
- The George Alleyne Chronic Disease Research Centre, A unit of the Caribbean Institute for Health Research, The University of the West Indies
- Institute for Health Metrics and Evaluation (IHME)
- Hospital Italiano de Buenos Aires, Department of Medical Informatics
- University of Buenos Aires, Laboratory of Technological Innovation in Public Health
- Central American Health Informatics Network (RECAINSA)
- Massachusetts General Hospital Institute for Technology Assessment, Harvard Medical School
- Center for Health and Humanitarian Systems, Georgia Institute of Technology
- PAHO Information Systems for Health (IS4H) Network of Experts

PAHO/EIH/IS/COVID-19/20-0007

© **Pan American Health Organization, 2020**

Some rights reserved. This work is available under license [CC BY-NC-SA 3.0 IGO](https://creativecommons.org/licenses/by-nc-sa/3.0/)