CovidVisualized: Visualized compilation of international updating models' estimates of COVID-19 pandemic at global and country levels

Farshad Pourmalek (pourmalek_farshad@yahoo.com)
https://orcid.org/0000-0002-2134-0771

Data Note

Keywords: COVID-19, pandemic, epidemic, models, visualization, global, Canada, Iran

DOI: https://doi.org/10.21203/rs.3.rs-768714/v1

Abstract

Objectives:

To identify international and periodically updated models of the COVID-19 epidemic, compile and visualize their estimation results, at the global and country levels, and periodically update the compilations. When one or more model predicts an increase in daily cases or infections and deaths in the next one to three months, this can be used as an early alarm for technical advisors to the national and subnational decision-makers to consider suggesting augmentation of preventive interventions.

Data description:

Five international and periodically updated models of the COVID-19 pandemic were identified, created by: (1) Massachusetts Institute of Technology, Cambridge, (2) Institute for Health Metrics and Evaluation, Seattle, (3) Imperial College, London, (4) Los Alamos National Laboratories, Los Alamos, and (5) University of Southern California, Los Angeles. Estimates of these five identified models were gathered, combined, and graphed at global and two country levels. Canada and Iran were chosen as countries with and without subnational estimates respectively. Compilations of results are periodically updated. Three Github repositories contain the codes and results: “CovidVisualizedGlobal” for the global level, “CovidVisualizedCountry” for a country with subnational estimates, Canada, and “covir2” for a country without subnational estimates, Iran.

Objective

Objectives and rationale: The objectives are to identify international and periodically updated models of COVID-19 epidemic, compile and visualize their estimations’ results, at the global and country levels, and periodically update the compilations. The ultimate objective is to provide an early warning system for technical advisors to the decision makers. When the predictions of one or more model show an increase in daily cases or infections, hospitalizations, or deaths in near future, technical advisors to the national and subnational decision-makers may consider suggesting augmentation of non-pharmacologic preventive interventions and vaccination. In doing so, strengths and weaknesses of individual models need to be considered, as well as those of this work. Models’ estimates demonstrate the trajectory of COVID-19 deaths, cases or infections, and hospital-related outcomes in one to three months into the future.

Eligibility criteria: Eligibility criteria for the target COVID-19 models to be included are (1) an international model scope, and (2) periodic updates. “International model” denotes a model that estimates COVID-19 cases or infections and deaths for all countries of the world, with global-level estimates that equate the sum of the national-level estimates. “Periodically updating” denotes a model with a record of periodically updated estimates since its first release, with continued updates in 2021.

Finding the eligible models: The eligible models were found within the literature search of a previously publication, "Rapid review of COVID-19 epidemic estimation studies for Iran" [1], and the results were verified by comparison with models found in a recently published study on "Predictive performance of international COVID-19 mortality forecasting models" [2]. While non-updating or one-time models can be contemporaneously useful, their results would not sustain up-to-dateness in the long run, especially with emergence of new variants of concern and various degrees of uncertainties in progression of vaccination coverage.

Identified eligible models: Five international and periodically updated models of the COVID-19 pandemic were identified: (1) DELPHI[1], Massachusetts Institute of Technology, Cambridge (abbreviation used in this work: DELP) [3], (2) Institute for Health Metrics and Evaluation, Seattle (IHME) [4], (3) Imperial College, London (IMPE) [5], (4) Los Alamos National Laboratories, Los Alamos (LANL) [6], (5) University of Southern California, Los Angeles, by Srivastava, Ajitesh (SRIV) [7]. Official reports of countries to World Health Organization, curated by Johns Hopkins University Coronavirus resource center (JOHN) [8] were also used for comparison.

The COVID-19 epidemic model by Youyang Gu [9], and the model by University of California, Los Angeles model [10] could not be categorized as international and periodically updating models. The COVID-19 International Modelling Consortium (CoMo Consortium) model, created by researchers at the University of Oxford and Cornell University [11], and CovidSim (COVID Simulation) model, created by researchers at Imperial College, London [12], provide templates for researchers to model the future of epidemic trajectory at national and subnational levels of their choice, through adjusting the model inputs and setting the time horizon into future for the estimations. Unlike the above-mentioned five international and periodically updating models, the latter two models are not intended for periodic updates by their original creators. The CoMo Consortium has engaged a number of countries, including Iran, but not Canada. There is no evidence of either model being used on a periodically updating basis in Iran or Canada.

[1] DELPHI: Differential Equations Lead to Predictions of Hospitalizations and Infections (model)

Data Description

The data is described under the following items: (1) repositories for codes and data sharing, (2) data management, and (3) periodical uptakes.

Repositories for codes and data sharing

GitHub (https://github.com) is used for sharing the codes and data. Global and country levels were chosen for demonstration of results. For the national level, one country with and one country without subnational estimates – Canada, and Iran respectively – were chosen, based on personal bounds.

Four of the five identified models share codes and estimates updates via GitHub repositories, and the IHME estimates that are released on IHME's own web site [4]. Three Github repositories were created for this project: “CovidVisualizedGlobal” [13] for the global level, “CovidVisualizedCountry” [14] for Canada, and “covir2” [15] for Iran. These are referred to as CovidVisualized GitHub repositories hereon. The “covir2” repository was created as “COVID Iran Review number 2” in March 2021, as an update after a first-iteration review was published [1]. The idea and work were further developed toward applicability to any location in the world, with case application for the global level, as well as Canada, and Iran.

GitHub repositories allow others to view and/or download, scrutinize, and verify integrity of the codes and data. It is also possible to minimally modify the codes to recreate similar responsories for any other country that reports COVID-19 cases and deaths to World Health Organization. Such use of the codes and data in GitHub is free of charge and bound to the pertinent licenses.

Data management

Data management template: A data management template was created for assigning comparable variable names to various outcomes from different models. Comparable and common variable names consist of generic parts (positions in the variable name) denoting the following items: (1) daily or total, (2) deaths, cases or infections, or other outcomes, (3) mean estimate, or lower, or upper uncertainty limit, (4) raw or smoothed estimate, (5) individual model, and (6) scenarios within each model. This template is described in detail in “variable name structure” in the CovidVisualized GitHub repositories [16]. Stata SE 14.2 (Stata Statistical Software. StataCorp. College Station, Texas) was used to write and run the codes on macOS Big Sur, and test run on Microsoft Windows 10.

Data management: Data acquisition, management, and graphing were performed via Stata codes. Stata codes download the models’ output files from their respective web sites, edit them according to the data management template, store each model’s estimates in a single file, and create graphs for all outcomes produced by each model. Then, the models’ single data files are compiled into one single final file, and graphs for all common outcomes are created for this compilation of all models. These graphs are shown on the pages of the three CovidVisualized GitHub repositories [13-15].

Outcome types: Besides daily deaths, daily (incident) cases or infections, total deaths, and total cases or infections, other outcomes estimated by one or more individual models, included prevalence, active or prevalent cases, recovered cases, hospital admissions, regular beds needed, ICU (Intensive Care Unit) beds needed, ventilated cases, seroprevalence, and effective reproduction number. All outcome types were graphed.

Secondary variables: Secondary variables are those created in this work using the primary variables released by the individual models. Secondary variables include case fatality rate (CFR), infection fatality rate (IFR), cases per deaths, estimated to reported deaths, and estimated to reported cases. CFR and IFR have only a daily version, and the other secondary variables have both daily and total (cumulative) versions.

Uncertainty, scenarios, variants, and vaccines: For each model, and for each outcome, both the point (mean) estimates, and the interval estimates (95% uncertainty limits) were graphed where available. Similarly, both the reference scenario (aka status quo) and alternative scenarios (i.e., better, and worse scenarios) were graphed for models with more than one scenario (i.e., IHME and IMPE). Assumptions about, and empirical inputs from, distributions of variants and vaccinations across space and time have been progressively included in models and scenarios of IHME and IMPE.

Subnational estimates: The DELP and IHME models, provide subnational-level estimates for countries that report both national and subnational level COVID-19 outcomes. Graphs were created for national and all subnational-level locations (i.e., provinces in Canada) available in DELP and IHME model outputs.

Periodical uptakes

A set of conventions were created for the periodic uptake of the models' estimates updates. The two models with least frequency of periodic updates of estimates are IHME and IMPE, which get updated on approximately a weekly and bi-weekly basis, respectively. With the release of each update of the either of these two models, the whole set of the five included models are updated in all the three CovidVisualized GitHub repositories, in which the most recent update of each model is used. These updates of CovidVisualized repositories are labelled as "uptakes" to differentiate them from models' estimates updates. These conventions for periodical uptake are described in detail in the CovidVisualized GitHub repositories [17]. R software via RStudio 1.4 (Integrated Development for R. RStudio. PBC, Boston, Massachusetts) was used for semi-automatization of the uptakes' execution. Estimates of the LANL model get updated about every 3-4 days, and DELP and SRIV models get updated daily. Uptaking the models' estimates updates with every update of these latter three models is not expected to depict a much more informative profile of the trajectory of the epidemic in future, when compared against the current convention of uptakes. The IHME, IMPE, and SRIV models provide estimates for about three months into the future with each update release, the DELP model for about two months, and the LANL model for about one month.

With each uptake, a directory is created in the root of the main branch of each of the three CovidVisualized repositories, and named with the uptake date (e.g., 20211027). Uptakes are also created retrospectively, compiling the results of the previous updates of model's results. As such, available uptakes for Iran and for the global level have been created going back to April 2021, and for Canada to June 2021. Under each dated uptake directory, there are two directories for "code" and for "output". Under each, there are located directories with the abbreviated name of the models. For example, "DELP" Directories under "code" store the Stata code files (.do), and those under "output" contain the outputs form executing the codes: the single data file for model estimates (in .dta and .csv formats), Stata log file (.smcl), and the graphs in PDF format. The directory "master" contains the master Stata do-file which executes all the other do files, and the directory "merge" (under "code") contains the code for creating the single final merged file of all models, and the directory "merge" (under "output") stores the created single final merged file of all compiled models, as well as the graphs that contain all the models. Selected graphs that contain all the models are visible in the root page of each uptake directory and are also stacked in a reverse chronological order in the main page of each repository. With each uptake, selected graphs of estimated outcomes are added on the start of the main page for each repository, and Situation Reports are created (and share with Ministry of Health) for Iran (and will be created for Canada and its provinces). Table 1 demonstrates an overview of the data sets (the three CovidVisualized GitHub repositories).

Similar work: The "covidcompare" tool [18] provides graph visualization of latest estimates of daily and total deaths from international and periodically updating COVID-19 models for countries of the world and US states, along with historical forecasts and model performance, based on IHME's "Predictive performance of international COVID-19 mortality forecasting models" [2].

Table 1: Overview of data files/data sets.

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data set 1	CovidVisualizedGlobal, COVID-19 estimates at the global level [13]	Stata code (.do), log (.smcl), data (.dta); R code (.R); data (.csv), graph (.pdf)	GitHub repository http://doi.org/10.5281/zenodo.5019030
Data set 2	CovidVisualizedCountry, COVID-19 estimates at the country level: Canada [14]	Stata code (.do), log (.smcl), data (.dta); R code (.R); data (.csv), graph (.pdf)	GitHub repository http://doi.org/10.5281/zenodo.5019482
Data set 3	covir2, COVID-19 estimates at the country level: Iran [15]	Stata code (.do), log (.smcl), data (.dta); R code (.R); data (.csv), graph (.pdf)	GitHub repository http://doi.org/10.5281/zenodo.5020797

Limitations

Limitations: Limitations of this work include the programming languages, automatization of uptakes, and choice of web site for presentation of the results. Stata programming language constitutes about 99% of the codes. Whereas Stata is a commercial software package, use of non-commercial package such as R and/or Python can increase accessibility and adoptability of the codes for other researchers. Further use of R and/or Python can also make the uptakes almost fully automatized. Some health researchers may not be familiar with GitHub and use of GIT programming, and additional use of a dedicated web site that is more visible to and accessible for health researchers and technical advisors to the public health decision makers can increase the reach and effect of this work. Strengths and weaknesses of individual international and periodically updating COVID-19 pandemic models are not mentioned here, but they have been discussed elsewhere [1-2].

Strengths: Strengths of this work include usability for informing technical advisors to the decision-makers, adaptability for use in other countries, and automatized data acquisition.

Tested usability for informing technical advisors to the decision-makers at the country level: Results of the GitHub repository “covir2” [15] were used to present the predictions of the five international and periodically updating models of COVID-19 pandemic, about possibility, timing, slope, height, and drivers of a potential fifth wave of the epidemic in Iran. This was done using the results of the covir2 repository along with results of an e-mail survey of more than 40 epidemiologist and public health specialists. The predictions and results were presented and described in a live online session for three Deputy Ministers of Health and six epidemiologists selected by the Ministry of Health and Medical Education (MOHME) of Iran. Periodical situation reports based on each uptake are also shared with MOHME.

Adaptability of the codes for use in other countries or regions in the world: The codes available in GitHub repositories “CovidVisualizedCountry” [14] and “covir2” [15] can be slightly modified by any researcher in order to be used for countries with and without subnational estimates respectively. “CovidVisualizedCountry” can be modified for use for any type of regionalization of the countries of the world, e.g., World Health Organization regions.

Automatized data acquisition: The Stata codes in these repositories automatically download the estimates data from the five included models once executed. There is no additional need for users to locate, download, and edit the estimated data of individual models prior to running the codes. This automatic data acquisition further enhances computational reproducibility – “obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis” [19].

Further research: Further research directions include using an “ensemble” method for combination of estimates from included models, as well as retrospective assessment of models’ predictive performance. In ensemble methods, individual models are evaluated for minimum requirements of quality and reporting and are combined using specific relative weights for each model, where the weights reflect the comparative accuracy of each model. Such ensemble methods are used by European Centre for Disease Prevention and Control [20] and US COVID-19 Forecast Hub [21]. The ensemble models have been empirically shown to be more accurate than any of the individual models used in the ensemble method [22]. Retrospective assessment of models’ predictive performance includes using statistical and graphical methods to estimate and visualize accuracy of models’ estimations [2].

Declarations

Ethics approval and consent to participate

All the used and produced data are at non-individual and aggregate level; publicly available on the Internet; and under pertinent licenses and copyrights for non-commercial use, reproduction, and distribution for scientific research, provided that the conditions mentioned in their respective licenses and copyrights are met, as provided in [23]. Therefore, no ethics approval or consent to participate were applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data described in this Data note can be freely and openly accessed on (1) GitHub repository "CovidVisualizedGlobal" under (http://doi.org/10.5281/zenodo.5019030) [13], (2) GitHub repository "CovidVisualizedCountry" under (http://doi.org/10.5281/zenodo.5019482) [14], and (3) GitHub repository "covir2" under (http://doi.org/10.5281/zenodo.5020797) [15]. Please see table 1 and references [13-15] for details and links to the data.

Competing interests

The author worked as a post-graduate research fellow in Institute for Health Metrics and Evaluation from 2009 to 2011 and continues voluntary collaboration as a Global Burden of Disease study collaborator without employment or financial relation. The author declares that he has no competing interests.

Funding

There were no sources of funding for this research.

Authors' contributions

Not applicable.

Acknowledgements

Not applicable.

Abbreviations

CoMo (Consortium): COVID-19 International Modelling Consortium

CovidSim: COVID Simulation (model)

CovidVisualizedCountry: Covid Visualized Country. Repository created in this work for visualization of COVID-19 epidemic models' estimates at country level – countries with subnational estimates, e.g., Canada

CovidVisualizedGlobal: Covid Visualized Global. Repository created in this work for visualization of COVID-19 epidemic models' estimates at global level

covir2: COVID Iran Review Number 2. Repository created in this work for visualization of COVID-19 epidemic models' estimates at country level – countries without subnational estimates, e.g., Iran

DELP: DELPHI. Differential Equations Lead to Predictions of Hospitalizations and Infections. COVID-19 pandemic model named DELPHI by Massachusetts Institute of Technology, Cambridge

IHME: Institute for Health Metrics and Evaluation. COVID-19 pandemic model by Institute for Health Metrics and Evaluation, Seattle

IMPE: Imperial. COVID-19 pandemic model by Imperial College, London

JOHN: Johns Hopkins. Coronavirus resource center, Johns Hopkins University, Baltimore

LANL: Los Alamos National Laboratories. COVID-19 pandemic model by Los Alamos National Laboratories, Los Alamos

MOHME: Ministry of Health and Medical Education (of Iran)

SRIV: Srivastava, Ajitesh. COVID-19 pandemic model by University of Southern California, Los Angeles

References

1. Pourmalek F, Rezaei Hemami M, Janani L, Moradi-Lakeh M. Rapid review of COVID-19 epidemic estimation studies for Iran. BMC Public Health. 2021 Feb 1;21(1):257. doi: 10.1186/s12889-021-10183-3.

2. Friedman J, Liu P, Troeger CE, Carter A, Reiner RC Jr, Barber RM, Collins J, Lim SS, Pigott DM, Vos T, Hay SI, Murray CJL, Gakidou E. Predictive performance of international COVID-19 mortality forecasting models. Nat Commun. 2021 May 10;12(1):2609. doi: 10.1038/s41467-021-22457-w.

3. COVID Analytics. DELPHI epidemiological case predictions. Cambridge: Operations Research Center, Massachusetts Institute of Technology. https://www.covidanalytics.io/projections and https://github.com/COVIDAnalytics/website/tree/master/data/predicted Accessed 23 June 2021.

4. Institute for Health Metrics and Evaluation (IHME). COVID-19 mortality, infection, testing, hospital resource use, and social distancing projections. Seattle: Institute for Health Metrics and Evaluation

(IHME), University of Washington. http://www.healthdata.org/covid/ and http://www.healthdata.org/covid/data-downloads Accessed 23 June 2021.

5. MRC Centre for Global Infectious Disease Analysis (MRC GIDA). Future scenarios of the healthcare burden of COVID-19 in low- or middle-income countries. London: MRC Centre for Global Infectious Disease Analysis, Imperial College London. https://mrc-ide.github.io/global-lmic-reports/ and https://github.com/mrc-ide/global-lmic-reports/tree/master/data Accessed 23 June 2021.

6. Los Alamos National Laboratory (LANL). COVID-19 cases and deaths forecasts. Los Alamos: Los Alamos National Laboratory (LANL). https://covid-19.bsvgateway.org Accessed 23 June 2021.

7. Srivastava, Ajitesh. University of Southern California (USC). COVID-19 forecast. Los Angeles: University of Southern California. https://scc-usc.github.io/ReCOVER-COVID-19 and https://github.com/scc-usc/ReCOVER-COVID-19/tree/master/results/historical_forecasts Accessed 23 June 2021.

8. Johns Hopkins University. Coronavirus resource center. https://coronavirus. jhu.edu/map.html and https://github.com/CSSEGISandData/COVID-19 Accessed 23 June 2021.

9. Gu, Youyang. COVID-19 Projections Using Machine Learning. https://covid19-projections.com and https://github.com/youyanggu/covid19_projections Accessed 23 June 2021.

10. Statistical Machine Learning Lab, Computer Science Department, University of California, Los Angeles. Combating COVID-19. https://covid19.uclaml.org/info.html and https://github.com/uclaml/ucla-covid19-forecasts/tree/master/current_projection Accessed 23 June 2021.

11. The COVID-19 International Modelling Consortium (CoMo Consortium), University of Oxford and Cornell University. https://www.medsci.ox.ac.uk/news/como-consortium-the-covid-19-pandemic-modelling-in-context and https://github.com/ocelhay/como Accessed 23 June 2021.

12. MRC Centre for Global Infectious Disease Analysis, Imperial College London. CovidSim. COVID-19 Scenario Analysis Tool. https://covidsim.org/v5.20210616/?place=ca and https://covidsim.org/v5.20210616/?place=ir Accessed 23 June 2021.

13. Pourmalek, F. GitHub repository “CovidVisualizedGlobal”: Combine and visualize international periodically updating estimates of COVID-19 at the global level. Version 1.1, Released June 23, 2021. http://doi.org/10.5281/zenodo.5019030 https://github.com/pourmalek/CovidVisualizedGlobal Accessed 23 June 2021.

14. Pourmalek, F. GitHub repository “CovidVisualizedCountry”: Combine and visualize international periodically updating estimates of COVID-19 at the country level, countries with subnational level estimates: Canada, national level, provinces, and territories. Version 1.1, Released June 23, 2021.

http://doi.org/10.5281/zenodo.5019482 https://github.com/pourmalek/CovidVisualizedCountry Accessed 23 June 2021.

15. Pourmalek, F. GitHub repository “covid2”: Combine and visualize international periodically updating estimates of COVID-19 at the country level, countries without subnational level estimates: Iran. Version 2.2, Released June 23, 2021. http://doi.org/10.5281/zenodo.5020797 https://github.com/pourmalek/covir2 Accessed 23 June 2021.

16. Pourmalek, F. “covid2”: Combine and visualize international periodically updating estimates of COVID-19 at the country level: Iran. Version 2.2, Released June 23, 2021. Variable name structure. http://doi.org/10.5281/zenodo.5020797 https://github.com/pourmalek/covir2/blob/main/Variable%20name%20structure.md Accessed 23 June 2021.

17. Pourmalek, F. “covir2”: Combine and visualize international periodically updating estimates of COVID-19 at the country level: Iran. Version 2.2, Released June 23, 2021. Setup. http://doi.org/10.5281/zenodo.5020797 https://github.com/pourmalek/covir2/tree/main/setup Accessed 23 June 2021.

18. Friedman J, Liu P, Akre S. The covidcompare tool. https://covidcompare.io/about Accessed 23 June 2021.

19. National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science. Washington, DC: The National Academies Press. 2019. https://doi.org/10.17226/25303 Accessed 23 June 2021.

20. European Centre for Disease Prevention and Control. European Covid-19 Forecast Hub. https://covid19forecasthub.eu/background.html and https://github.com/epiforecasts/covid19-forecast-hub-europe Accessed 23 June 2021.

21. COVID-19 Forecast Hub. https://covid19forecasthub.org/doc/ensemble and https://github.com/reichlab/covid19-forecast-hub Accessed 23 June 2021.

22. Cramer EY, Ray EL, Lopez VK, Bracher J, Brennen A, et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US. medRxiv preprint. Posted February 05, 2021. https://www.medrxiv.org/content/10.1101/2021.02.03.21250974v1 Accessed 23 June 2021.

23. Pourmalek, F. GitHub repository “covid2”: Combine and visualize international periodically updating estimates of COVID-19 at the country level, countries without subnational level estimates: Iran. Version 2.2, Released June 23, 2021. Licenses / Copyrights of data and / or graphs used in this repository. http://doi.org/10.5281/zenodo.5020797 https://github.com/pourmalek/covir2#licenses–copyrights-of-data-and–or-graphs-used-in-this-repository Accessed 23 June 2021.

Covidvisualized Tools

covir2	https://github.com/pourmalek/covir2
CovidVisualizedCountry	https://github.com/pourmalek/CovidVisualizedCountry
CovidVisualizedGlobal	https://github.com/pourmalek/CovidVisualizedGlobal