Update: An efficient, objective index for predictive disease incidence ranking of COVID-19 vaccine trial sites

Authors: Bradley G. Wagner[1*], Prashanth Selvaraj[1] and Stewart T. Chang[1]

February 15th, 2021

1. Institute for Disease Modeling, Global Health, Bill and Melinda Gates Foundation

* Contact author: **bwagner@idmod.org**

Purpose of document:

The aim of this report is to provide ongoing updates to predictive subnational results for a selection of vaccine-trial sites based on future COVID-19 disease incidence. The output of the analysis is a normalized, ranking index between 0 and 1 we denote as "G". This value predicts trial sites in terms of confidence in COVID-19 case incidence beginning after a two-month lag from the selection date (corresponding roughly to site prep time). This report provides updated results for the November 13th, December 16th, and January 22nd documents of the same name. For discussion regarding interpretation of results as well as methodology and validation of methods, please refer to the original document. A short additional discussion of the implications for model predictions of emergence and circulation of novel SARS-CoV-2 variants, which originally appeared in the January 22nd update, is included for reference. A short additional section has been added explaining the limitations of the predictive index when incorporating the impact of rapid scale-up of approved COVID-19 vaccines in the entire population.

Document structure and usage:

Countries for which results have been updated are indicated in Table 1, which includes the geographic level of analysis. The time period of data used in the analysis as well as the corresponding future trial start dates for which the analysis is relevant are denoted in Table 2. Tables and figures sections correspond to those in the original document. As described in the original document, the modeling methodology does not account for future introduction of novel interventions that may reduce transmission. As such, these methods should only be used to evaluate future vaccine trial sites in regions where there **is not yet**

population-wide vaccine coverage. Note that predictive index values will change with temporal updates and such changes may be related to epidemiology and immunity, behavior, public policy, or other factors. The method is based on case data (it makes no assumptions regarding case detection rate), but it is not mechanistic, and as such makes no claims as to the underlying drivers of prediction changes.

As described in more detail in the original document, the method described computes a normalized index (**G-index**, ranging from 0 to 1) designed to rank prediction trial sites in terms of confidence in COVID-19 case incidence beginning after a two-month lag from the selection date (corresponding roughly to site prep time). Higher values indicate more confidence in sustained transmission; values greater than 0.5 indicate the epidemic is more likely than not to have been in a growth phase during the historical lookback period used to construct the index.

Emergence of novel SARS-CoV-2 variants

The predictive index presented does not explicitly consider emergence of (potentially more transmissible) novel variants of SARS-CoV-2. However, since the index is based on historical growth in cases such variants are implicitly accounted for in the computation. It should be noted that, as discussed in the original document, the index uses the past as a predictor of the future and therefore cannot predict either outbreaks where there is no history of circulation or the future emergence of more transmissible variants. The index takes into account the impact of novel variants only after there is a history of population-level circulation, and therefore one should be cautious in interpreting predictions in the lag phase between emergence of known higher transmissibility variants and their fixation in the population.

Effects of vaccine rollout on model predictions

The predictive index G is based on case data and is therefore agnostic to the types of interventions which reduce transmission. As such it will reflect the impact of vaccines or natural immunity in the population. However, there is a lag between introduction of interventions and reflection in case data. While this lag has less effect on the index for interventions which change relatively slowly in population impact, a rapid population-wide introduction of effective vaccine can result in an index which under-estimates near future intervention impact. In practice, the predictive index is intended to inform vaccine trial site selection, and there are numerous reasons not to conduct phase 3 vaccine trials against a rapid scale-up of another approved COVID-19 vaccine. Therefore, issues related to vaccine rollout, with respect to index computation, should pose less of a practical concern. For a backdrop of vaccine which has already been

rolled out (near its maximum coverage) and whose effect has been seen in the population for at least a few months the predictive index will not suffer from these lag issues.

Table 1: Countries and regions represented in the trial site analysis and geographic level of analysis. References indicate where the collated disease incidence data that underlies the model was obtained, if applicable.

Country	Geographic level of analysis
Argentina	subnational[2]
Brazil	subnational[3]
Colombia	subnational[3]
Gambia	national[4]
Malawi	national[4]
Mozambique	national[4]
India	subnational[3]
Mexico	subnational[4]
Pakistan	subnational[4]
United Kingdom	subnational[3]

Table 2: Index values by region:

G-index values computed by region. The lookback period used is indicated as well as target trial start date. Here the target trial start date is 2 months from the decision point (the last data collected) as was empirically validated (see the "Validation" section and "Definitions" in the original document). Maps showing the geographic distribution of the index by country are given in section Figures: Spatial heterogeneity. Historical Rt estimates for each region are shown in section Figures: Historical Rt estimates. Note that lookback periods and target trial dates were chosen at the country level.

Country	Region	Index Value (G)	Lookback Period	Target Trial Start Date
Argentina	**Catamarca**	**0.878**	**20-12-77 – 21-2-7**	**21-4-7**
Argentina	Jujuy	0.791		
Argentina	Misiones	0.768		
Argentina	Salta	0.754		
Argentina	La Rioja	0.678		

Argentina	San Luis	0.607		
Argentina	Neuquen	0.543		
Argentina	Santiago del Estero	0.535		
Argentina	Province of Buenos Aires	0.527		
Argentina	City of Buenos Aires	0.523		
Argentina	Mendoza	0.512		
Argentina	Cardoba	0.494		
Argentina	Entre Rios	0.486		
Argentina	Formosa	0.484		
Argentina	Corrientes	0.469		
Argentina	Rio Negro	0.461		
Argentina	Chaco	0.414		
Argentina	Tucuman	0.391		
Argentina	La Pampa	0.368		
Argentina	Santa Fe	0.352		
Argentina	Chubut	0.350		
Argentina	Tierra del Fuego	0.150		
Argentina	San Juan	0.097		
Brazil	Ceara	0.914	**20-12-12 – 21-2-12**	**21-4-12**
Brazil	Paraba	0.897		
Brazil	Alagoas	0.886		
Brazil	Acre	0.802		
Brazil	Maranhao	0.732		
Brazil	Pernambuco	0.729		
Brazil	Roraima	0.728		
Brazil	Para	0.657		
Brazil	Goias	0.626		
Brazil	Rio Grande do Norte	0.608		
Brazil	Mato Grosso	0.590		
Brazil	Bahia	0.559		
Brazil	Sao Paulo	0.557		
Brazil	Rondania	0.544		
Brazil	Santa Catarina	0.512		
Brazil	Amazonas	0.506		
Brazil	Tocantins	0.496		
Brazil	Rio de Janeiro	0.485		
Brazil	Piaui	0.484		
Brazil	Minas Gerais	0.476		
Brazil	Distrito Federal	0.354		
Brazil	Amapa	0.294		
Brazil	Sergipe	0.228		
Brazil	Parana	0.199		
Brazil	Mato Grosso do Sul	0.147		

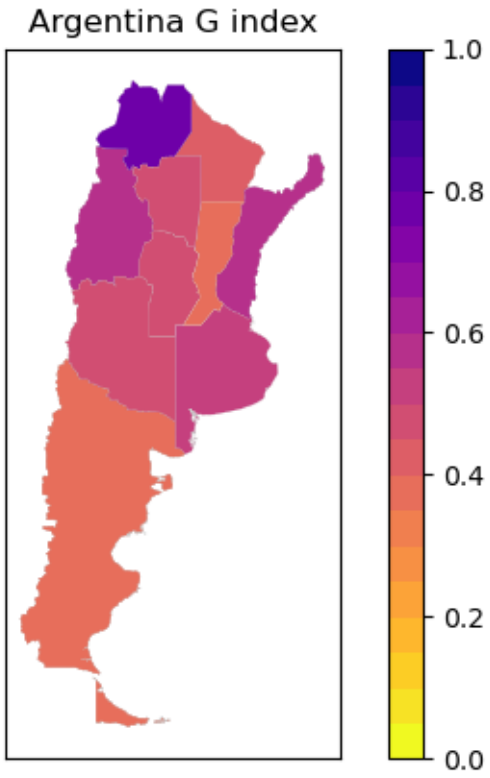
Brazil	Rio Grande do Sul	0.142		
Brazil	Espirito Santo	0.031		
Colombia	**Amazonas**	**0.926**	**20-12-12 – 21-2-12**	**21-4-12**
Colombia	Cordoba	0.784		
Colombia	Sucre	0.643		
Colombia	Guainia	0.563		
Colombia	Vichada	0.541		
Colombia	Choco	0.502		
Colombia	Huila	0.463		
Colombia	Cauca	0.454		
Colombia	Valle (including Cali)	0.451		
Colombia	Cundinamarca	0.441		
Colombia	Guaviare	0.436		
Colombia	Boyaca	0.433		
Colombia	Casanare	0.408		
Colombia	Arauca	0.406		
Colombia	Narino	0.398		
Colombia	Caldas	0.395		
Colombia	Putumayo	0.394		
Colombia	San andres y providencia	0.380		
Colombia	Magdalena	0.375		
Colombia	Bogota	0.358		
Colombia	Cesar	0.354		
Colombia	Santander	0.345		
Colombia	Atlantico	0.341		
Colombia	Antioquia	0.335		
Colombia	Meta	0.300		
Colombia	Tolima	0.292		
Colombia	Vaupes	0.223		
Colombia	Caqueta	0.204		
Colombia	Quindio	0.186		
Colombia	Risaralda	0.140		
Colombia	Bolivar	0.116		
Colombia	La Guajira	0.115		
Colombia	Norte de santander	0.036		
The Gambia	**Gambia (country-level)**	**0.859**	**20-12-7 – 21-2-7**	**21-4-7**
India	**Maharashtra**	**0.523**	**20-12-12 – 21-2-12**	**21-4-12**
India	Kerala	0.481		
India	Punjab	0.416		
India	Nagaland	0.373		
India	Meghalaya	0.368		
India	Manipur	0.280		
India	Haryana	0.258		

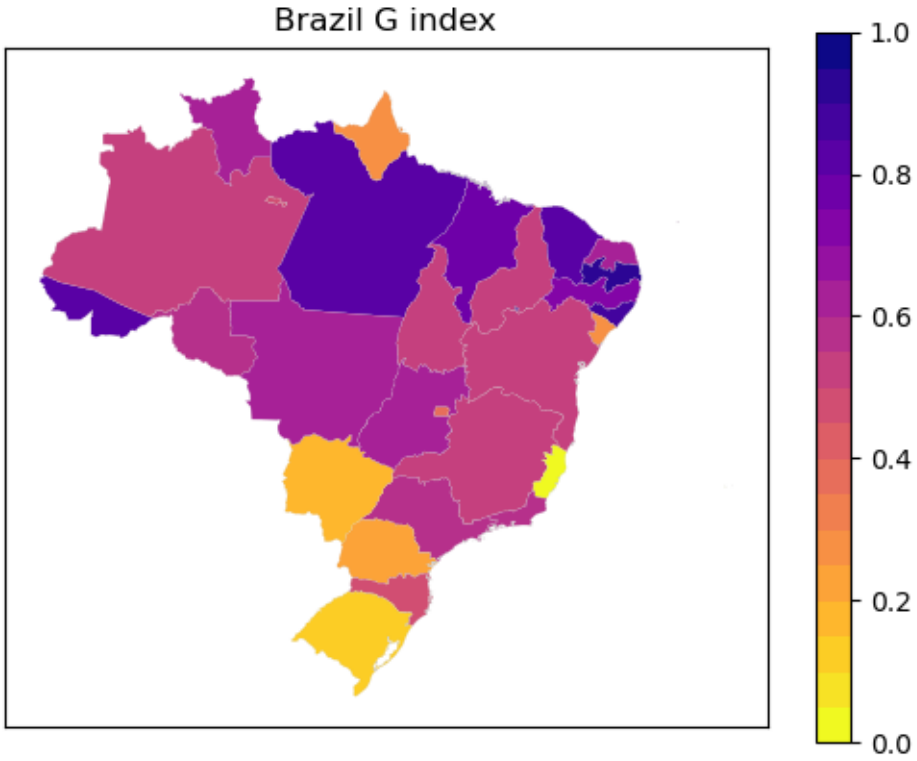
India	Madhya Pradesh	0.253		
India	Puducherry	0.229		
India	Jammu and Kashmir	0.207		
India	Himachal Pradesh	0.201		
India	Gujarat	0.189		
India	Andaman and Nicobar	0.163		
India	NCT of Delhi	0.152		
India	Rajasthan	0.142		
India	Uttarakhand	0.137		
India	Jharkhand	0.136		
India	West Bengal	0.116		
India	Ladakh	0.106		
India	Goa	0.086		
India	Karnataka	0.079		
India	Odisha	0.076		
India	Dadra and Nagar Haveli	0.058		
India	Chandigarh	0.043		
India	Sikkim	0.023		
India	Telangana	0.020		
India	Mizoram	0.009		
India	Bihar	0.006		
India	Assam	0.004		
India	Arunachal Pradesh	0.003		
India	Chhattisgarh	0.002		
India	Uttar Pradesh	0.001		
India	Tripura	0.001		
India	Andhra Pradesh	0.000		
Mexico	**Jalisco**	**0.502**	**20-11-25 – 21-1-25**	**21-3-25**
Mexico	Mexico City	0.478		
Malawi	**Malawi (country-level)**	**0.684**	**20-12-7 – 21-2-7**	**21-4-7**
Mozambique	**Mozambique (country-level)**	**0.671**	**20-12-7 – 21-2-7**	**21-4-7**
Pakistan	**Sindh**	**0.242**	**20-12-7 – 21-2-7**	**21-4-7**
United Kingdom	**Scotland**	**0.355**	**20-12-12 – 21-2-12**	**21-4-12**
United Kingdom	Midlands	0.341		
United Kingdom	North West	0.337		
United Kingdom	South West	0.324		
United Kingdom	North East	0.310		
United Kingdom	Northern Ireland	0.270		

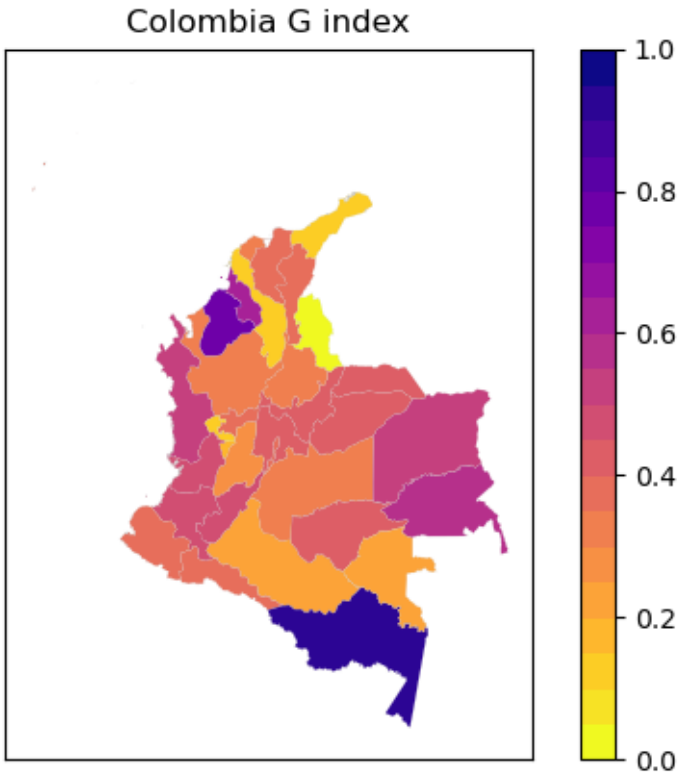
United Kingdom	South East	0.231		
United Kingdom	East of England and Yorkshire	0.208		
United Kingdom	London	0.203		
United Kingdom	Wales	0.031		

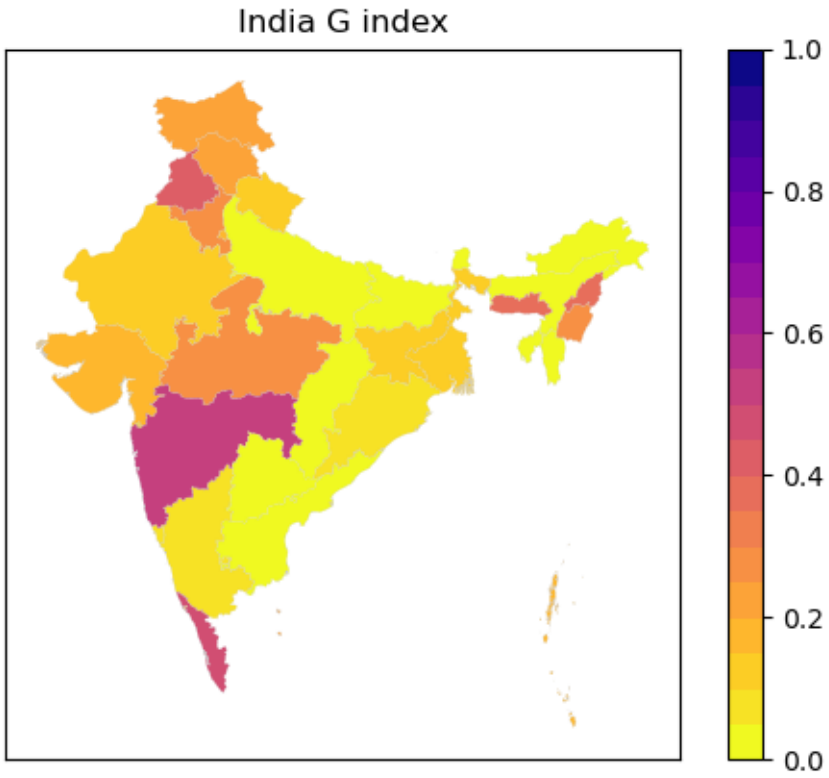
Figures section 1: Spatial heterogeneity in G index.

Note the color scale is identical for all maps shown. See Table 2 for values by region, lookback period used in input data, and target trial dates. Note that for Pakistan (Sindh), Mexico (Jalisco, Mexico City) and The Gambia, values are given in Table 2.

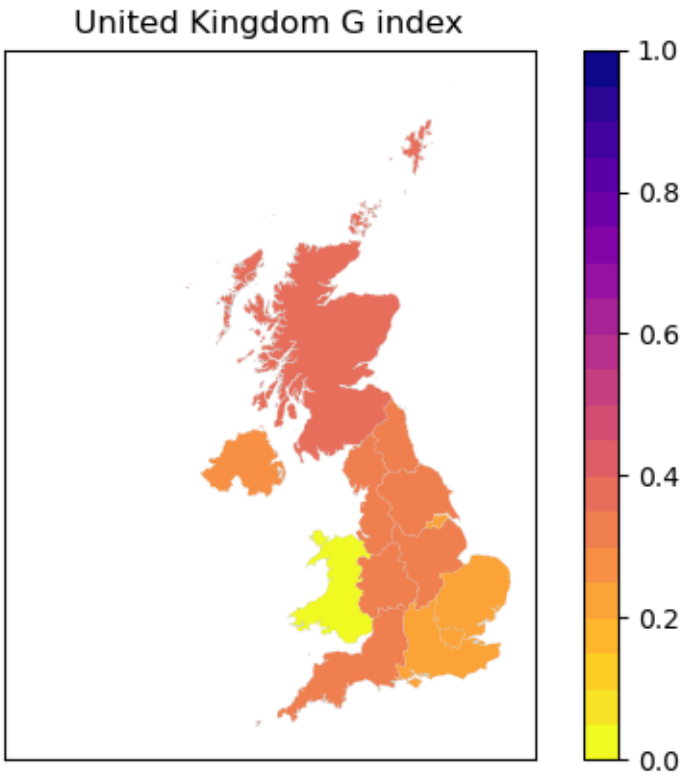



Brazil G index
1.0
0.8
0.6
0.4
0.2
0.0


Colombia G index
1.0
0.8
0.6
0.4
0.2
0.0



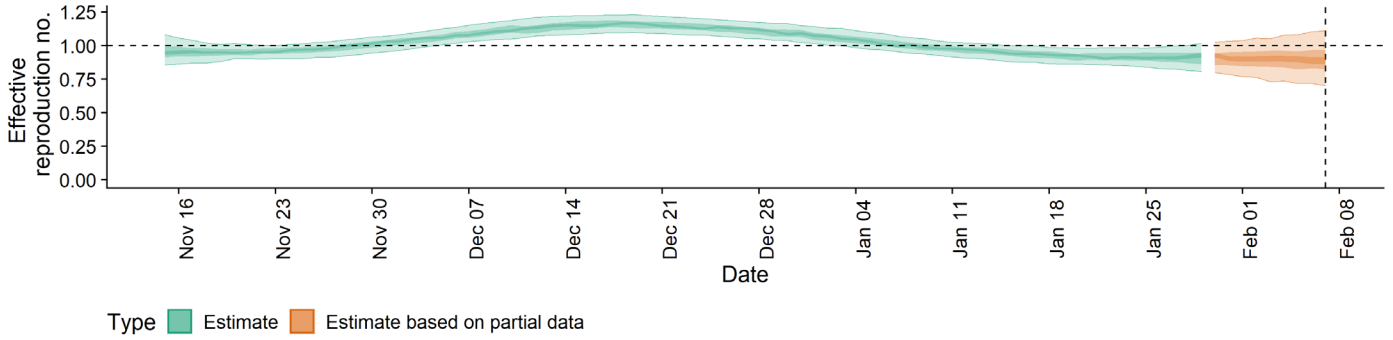
*** No data for Tamil Nadu in this update.**


United Kingdom G index
1.0
0.8
0.6
0.4
0.2
0.0

Figures section 2: Historical Rt estimates

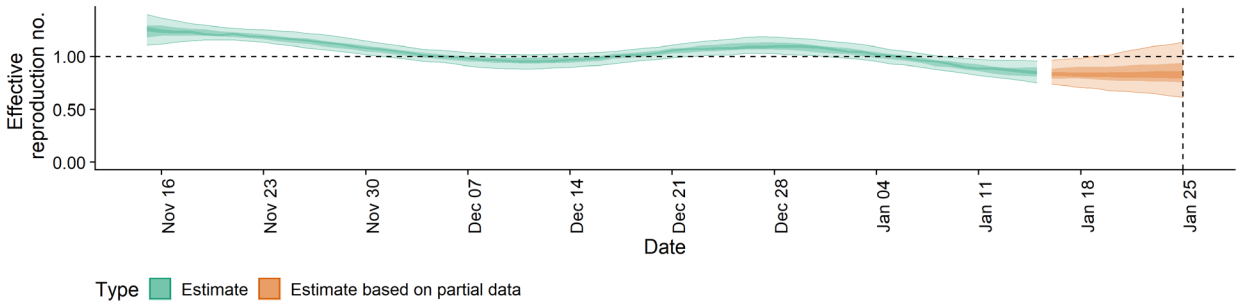
Regional estimates of Rt for Brazil, Colombia, the United Kingdom and India available at Epiforecasts.io.

Buenos Aires Province (Argentina):



Estimated Rt values for Buenos Aires Province showing 50% and 90% credible intervals.

Mexico City (Mexico):

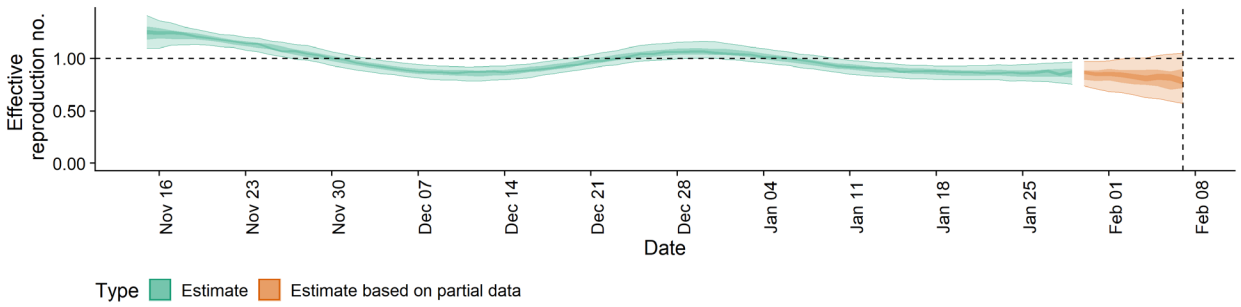


Estimated Rt values for Mexico City showing 50% and 0% credible intervals.

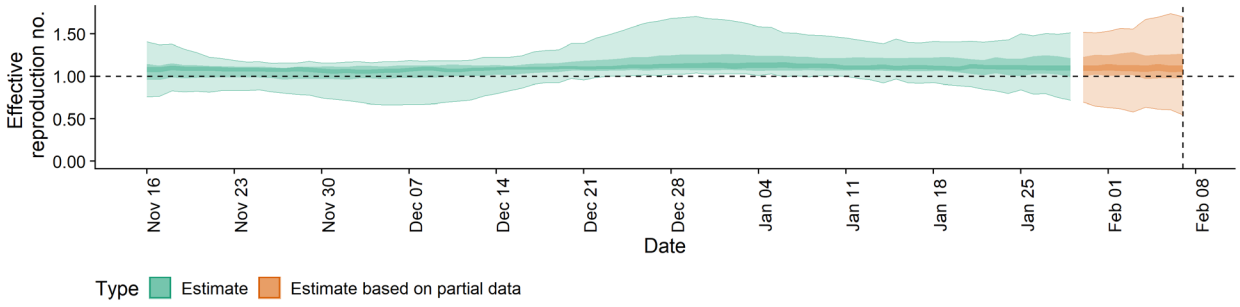
**** Note that there was an underlying retrospective change in reporting data (beyond issues of partial data) for Mexico City which accounts for changes between the January 22nd results and the results shown above. Jalisco data was unaffected beyond expected changes for partial data.**

Sindh Province (Pakistan):

Estimated Rt values for Sindh province showing 50% and 90% credible intervals.

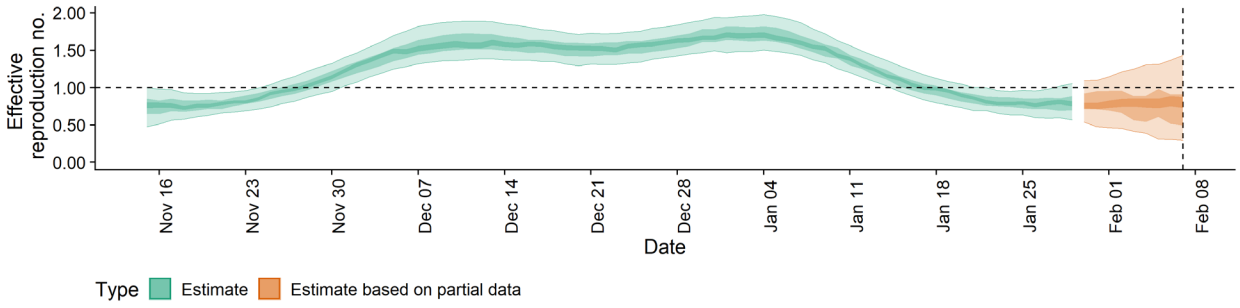


The Gambia:



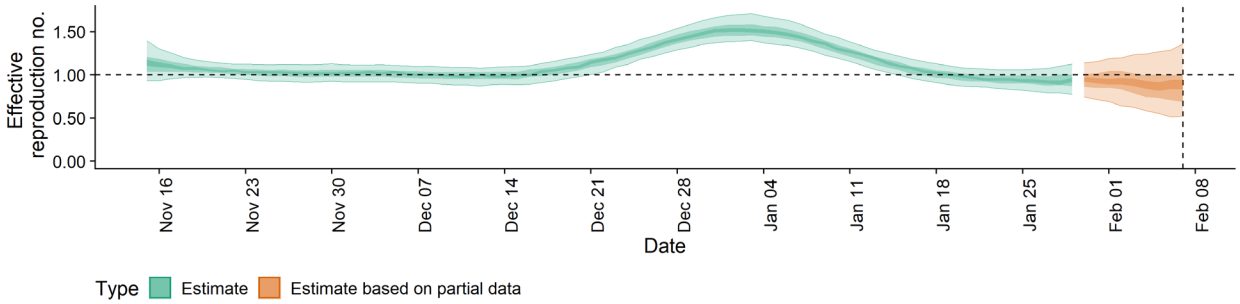
Estimated Rt values for The Gambia showing 50% and 90% credible intervals.

Malawi



Estimated Rt values for Malawi showing 50% and 90% credible intervals.

Mozambique:



Estimated Rt values for Mozambique showing 50% and 90% credible intervals.

References:

1. Abbott, S. *et al.* Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Res.* **5**, 112 (2020).

2. Google-research/open-covid-19-data: Open source aggregation pipeline for public COVID-19 data, including hospitalization/ICU/ventilator numbers for many countries. https://github.com/google-research/open-covid-19-data.

3. Epiforecasts/covid-rt-estimates: National and subnational estimates of the time-varying reproduction number for Covid-19. https://github.com/epiforecasts/covid-rt-estimates.

4. IHME, COVID-19. https://covid19.healthdata.org/global?view=total-deaths&tab=trend.

5. Thompson, R. N. *et al.* Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* **29**, 100356 (2019).

6. EpiNow2. https://cran.r-project.org/web/packages/EpiNow2/EpiNow2.pdf.

7. Stan - Stan. https://mc-stan.org/.

8. Ganyani, T. *et al.* Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance* **25**, (2020).

9. Lauer, S. A. *et al.* The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **172**, 577–582 (2020).

10. Covid-projections/covid-projections: Code powering Covid Act Now - A site urging Public leaders & health officials to take action now to prevent the spread of COVID-19. https://github.com/covid-projections/covid-projections.