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]

March 25th, 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 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, January 22nd and February 15th 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 is 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 including 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 that 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 that change relatively slowly in population impact, a rapid population-wide introduction of effective vaccine can result in an index that underestimates 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 that 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 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	**Salta**	**0.942**	**21/1/12 - 21/3/14**	**21/5/14**
Argentina	La Rioja	0.694		
Argentina	Mendoza	0.626		
Argentina	Tucuman	0.601		
Argentina	Jujuy	0.572		

Argentina	Cordoba	0.565		
Argentina	Formosa	0.544		
Argentina	Corrientes	0.532		
Argentina	Catamarca	0.471		
Argentina	Misiones	0.435		
Argentina	Santa Fe	0.403		
Argentina	City of Buenos Aires	0.385		
Argentina	La Pampa	0.373		
Argentina	Santiago del Estero	0.369		
Argentina	Chaco	0.348		
Argentina	Buenos Aires Province	0.343		
Argentina	San Luis	0.283		
Argentina	San Juan	0.28		
Argentina	Chubut	0.121		
Argentina	Tierra del Fuego	0.119		
Argentina	Entre Rios	0.095		
Argentina	Neuquen	0.046		
Argentina	Rio Negro	0.025		
Brazil	**Maranhao**	**0.936**	**21/1/14 - 21/3/17**	**21/5/17**
Brazil	Alagoas	0.934		
Brazil	Ceara	0.933		
Brazil	Acre	0.881		
Brazil	Goias	0.869		
Brazil	Paraiba	0.819		
Brazil	Mato Grosso do Sul	0.776		
Brazil	Tocantins	0.767		
Brazil	Piaui	0.764		
Brazil	Rio Grande do Sul	0.702		
Brazil	Santa Catarina	0.699		
Brazil	Pernambuco	0.683		
Brazil	Mato Grosso	0.676		
Brazil	Sao Paulo	0.661		
Brazil	Espirito Santo	0.658		
Brazil	Para	0.656		
Brazil	Amapa	0.646		
Brazil	Minas Gerais	0.634		
Brazil	Rio Grande do Norte	0.627		
Brazil	Roraima	0.598		
Brazil	Distrito Federal	0.595		
Brazil	Parana	0.591		
Brazil	Rondonia	0.582		
Brazil	Sergipe	0.56		
Brazil	Bahia	0.447		

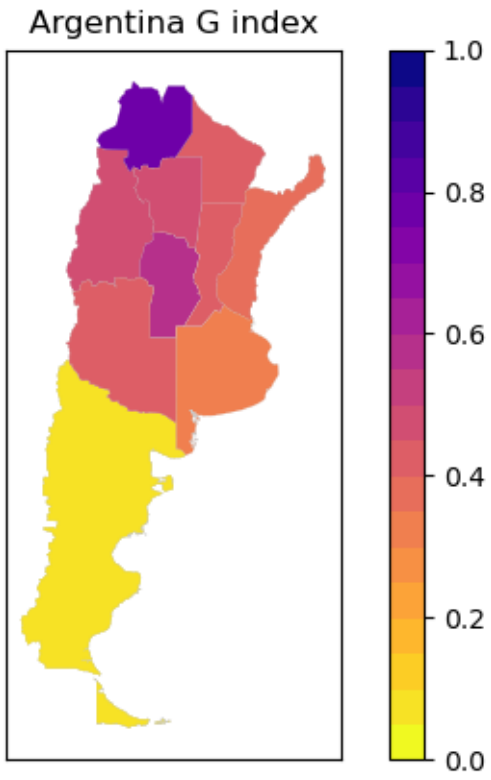
Brazil	Rio de Janeiro	0.384		
Brazil	Amazonas	0.174		
Colombia	**Magdalena**	**0.684**	**21/1/14 - 21/3/17**	**21/5/17**
Colombia	Atlantico	0.539		
Colombia	La Guajira	0.513		
Colombia	Sucre	0.498		
Colombia	Risaralda	0.481		
Colombia	Cordoba	0.468		
Colombia	Antioquia	0.464		
Colombia	Guainia	0.463		
Colombia	Caldas	0.45		
Colombia	Amazonas	0.412		
Colombia	Cesar	0.4		
Colombia	Choco	0.394		
Colombia	Valle del Cauca	0.375		
Colombia	Quindio	0.347		
Colombia	Narino	0.294		
Colombia	Bolivar	0.285		
Colombia	Bogota	0.273		
Colombia	Meta	0.262		
Colombia	Vichada	0.252		
Colombia	Cauca	0.228		
Colombia	Putumayo	0.208		
Colombia	Vaupes	0.203		
Colombia	Casanare	0.202		
Colombia	Guaviare	0.16		
Colombia	Cundinamarca	0.126		
Colombia	Arauca	0.121		
Colombia	Tolima	0.116		
Colombia	Huila	0.111		
Colombia	Boyaca	0.104		
Colombia	Norte de Santander	0.1		
Colombia	Caqueta	0.042		
Colombia	San Andres y Providencia	0.04		
Colombia	Santander	0.003		
The Gambia	**The Gambia**	**0.69**	**21/1/12 - 21/3/14**	**21/5/14**
India	**Punjab**	**0.939**	**21/1/14 - 21/3/17**	**21/5/17**
India	Maharashtra	0.909		
India	Madhya Pradesh	0.826		
India	Dadra and Nagar Haveli	0.823		
India	Haryana	0.803		
India	Chandigarh	0.802		
India	NCT of Delhi	0.757		

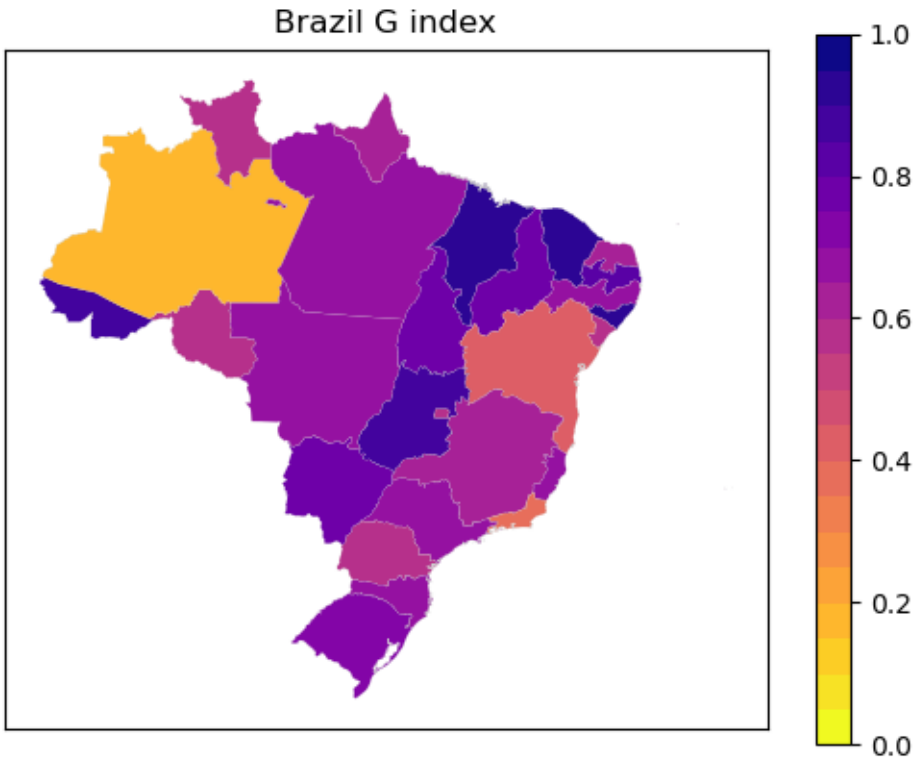
India	Jammu and Kashmir	0.734		
India	Gujarat	0.733		
India	Rajasthan	0.73		
India	West Bengal	0.716		
India	Himachal Pradesh	0.715		
India	Assam	0.683		
India	Jharkhand	0.678		
India	Lakshadweep	0.674		
India	Andhra Pradesh	0.652		
India	Karnataka	0.647		
India	Telangana	0.635		
India	Uttarakhand	0.616		
India	Goa	0.598		
India	Tripura	0.559		
India	Uttar Pradesh	0.552		
India	Chhattisgarh	0.54		
India	Ladakh	0.524		
India	Puducherry	0.493		
India	Sikkim	0.492		
India	Odisha	0.456		
India	Andaman and Nicobar	0.444		
India	Meghalaya	0.434		
India	Bihar	0.355		
India	Nagaland	0.199		
India	Kerala	0.161		
India	Arunachal Pradesh	0.097		
India	Manipur	0.076		
India	Mizoram	0.049		
Malawi	**Malawi**	**0.074**	**21/1/12 - 21/3/14**	**21/5/14**
Mexico	**Mexico City**	**0.105**	**21/1/2 - 21/3/4**	**21/5/4**
Mexico	Jalisco	0.05		
Mozambique	**Mozambique**	**0.193**	**21/1/12 - 21/3/14**	**21/5/14**
Pakistan	**Sindh**	**0.066**	**21/1/12 - 21/3/14**	**21/5/14**
United Kingdom	**Scotland**	**0.239**	**21/1/14 - 21/3/17**	**21/5/17**
United Kingdom	North East and Yorkshire	0.146		
United Kingdom	Northern Ireland	0.125		
United Kingdom	East of England	0.099		
United Kingdom	North West	0.095		

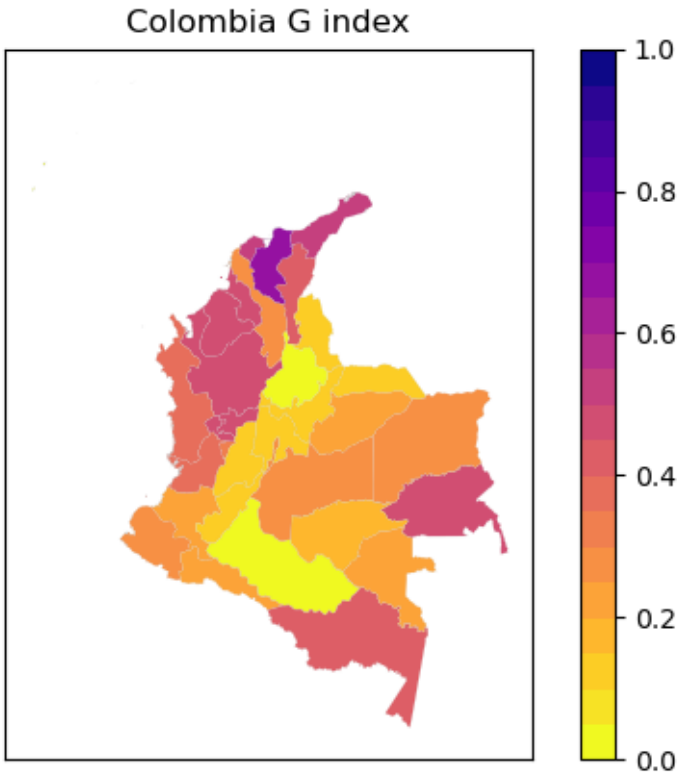
United Kingdom	Midlands	0.094		
United Kingdom	Wales	0.081		
United Kingdom	South East	0.079		
United Kingdom	London	0.055		
United Kingdom	South West	0.05		

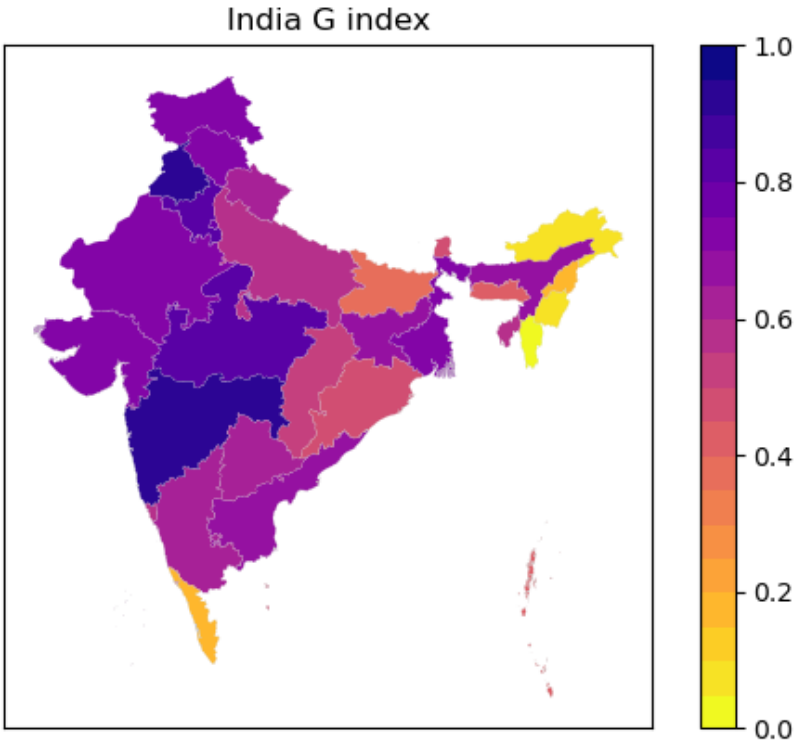
Figures section 1: Spatial heterogeneity in G index.

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.

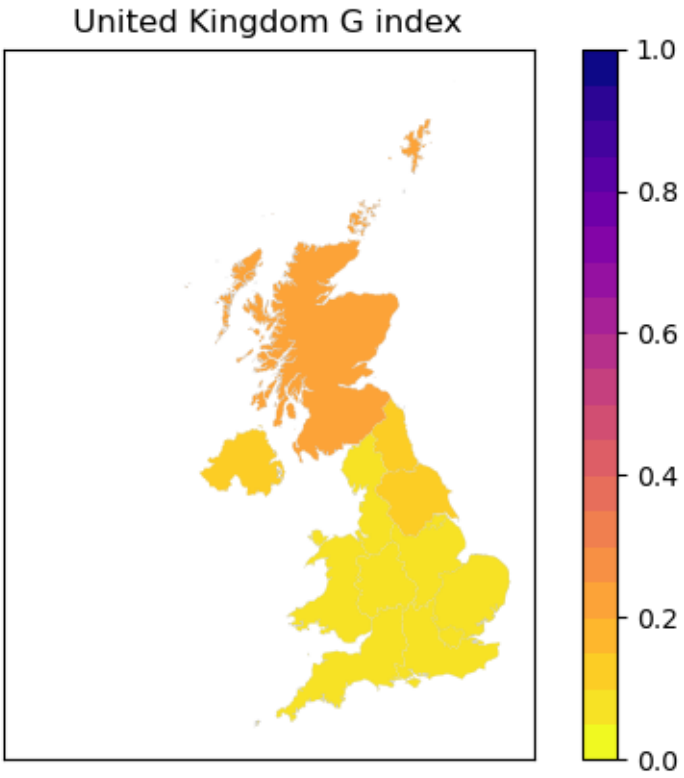



Brazil G index
1.0
0.8
0.6
0.4
0.2
0.0


Colombia G index
1.0
0.8
0.6
0.4
0.2
0.0



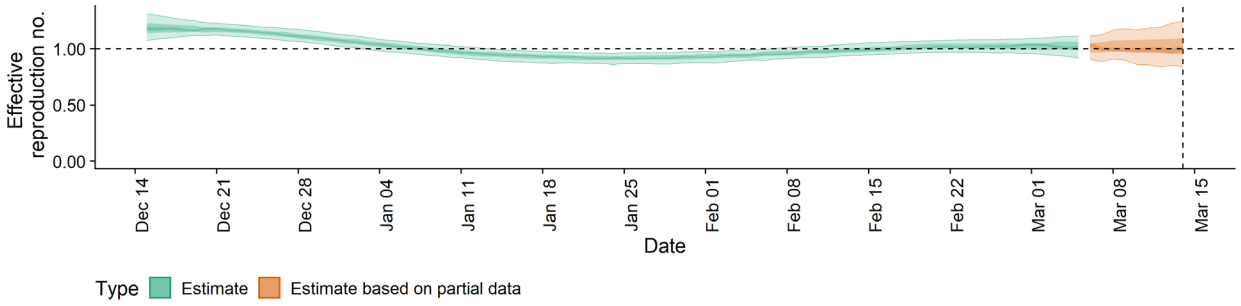
*** No data for Tamil Nadu in this update.**


United Kingdom G index
1.0
0.8
0.6
0.4
0.2
0.0

Figures section 2: Historical Rt estimates

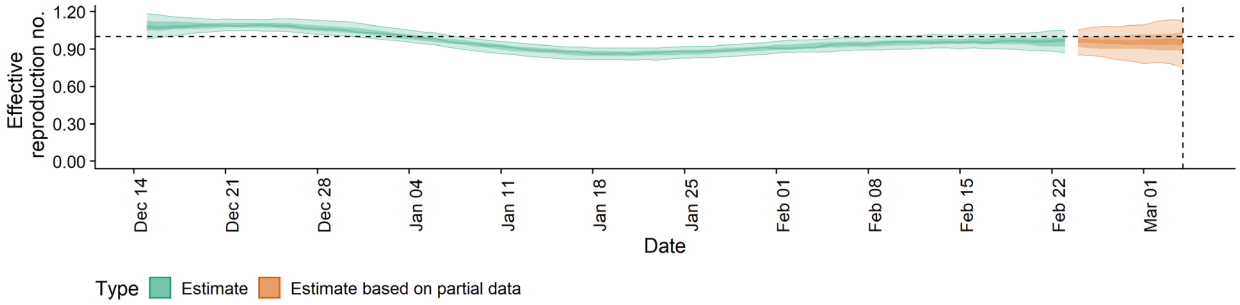
Regional estimates of Rt for Brazil, Colombia, the United Kingdom and India available at Epiforecasts.io.

Buenos Aires Province (Argentina):



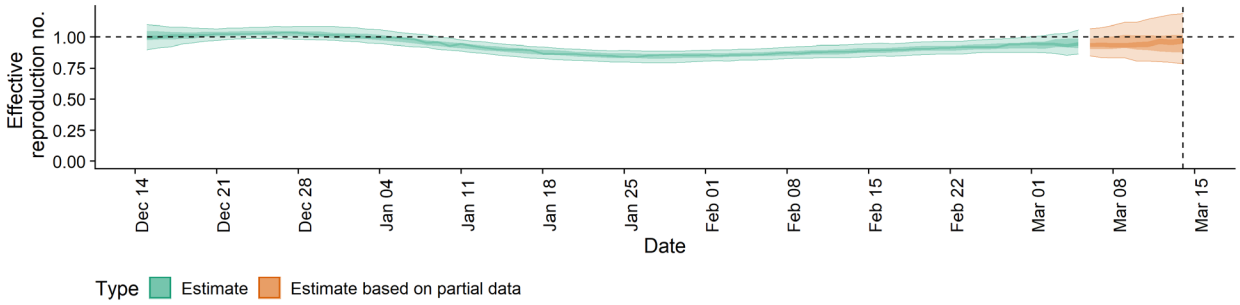
Estimated Rt values for Buenos Aires Province showing 50% and 90% credible intervals.

Mexico City (Mexico):



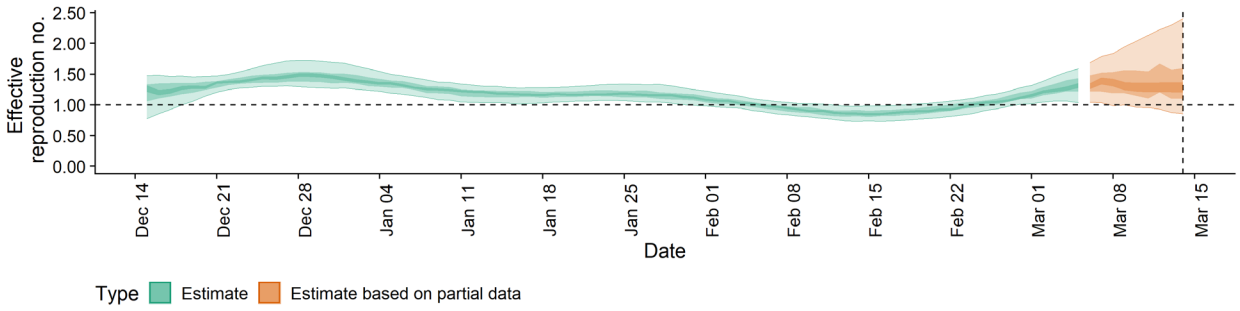
Estimated Rt values for Mexico City showing 50% and 0% credible intervals.

Sindh Province (Pakistan):



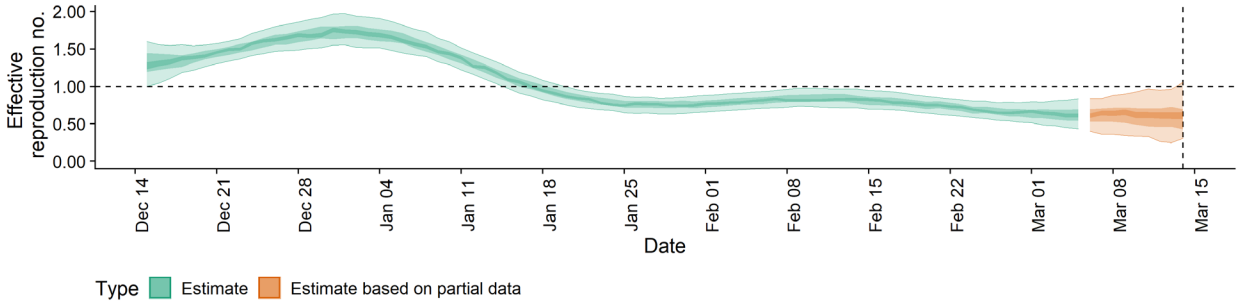
Estimated Rt values for Sindh province showing 50% and 90% credible intervals.

The Gambia:



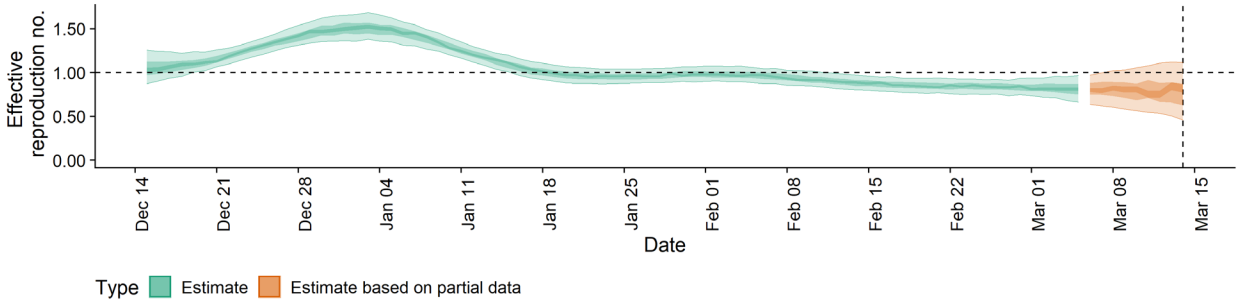
Estimated Rt values for The Gambia showing 50% and 90% credible intervals.

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.