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]

January 22nd, 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 and December 16th 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 is included in the final section as well as a short discussion of some country level changes in estimates from previous reports.

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.

Qualitative changes at the country level

We note that at the country level, The Gambia has shown sustained transmission over a 2-month or longer time period (G= 0.59) for the first time dating back to July 2020 data. As discussed in the original document qualitative changes can be an artifact of changes in testing; however, in this case such observed increases would be less likely to be sustained over long time periods. In fact, the parsimonious definition of the index **G—**as the probability of the reproductive number to be above 1 over a 2-month time period **—**is motivated by the desire to be robust to rapid changes in testing which may appear as rapid increases and decreases in the effective reproduction number. For a regional, country-level comparison we included analysis for Mozambique and Malawi (see Figures section 2). Both countries show, to some degree, a similar pattern of qualitative change in effective sustained effective reproduction number above 1 (estimated from case data) with G=0.84, 0.68 respectively.

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	**Misiones**	**0.937**	**20-11-15 – 21-1-15**	**21-3-15**
Argentina	Catamarca	0.759		
Argentina	City of Buenos Aires	0.758		
Argentina	Rio Negro	0.732		
Argentina	Province of Buenos Aires	0.724		
Argentina	Formosa	0.722		

Argentina	Entre Rios	0.700		
Argentina	Neuquen	0.683		
Argentina	Chubut	0.666		
Argentina	La Pampa	0.648		
Argentina	Jujuy	0.645		
Argentina	Santiago del Estero	0.509		
Argentina	Cordoba	0.493		
Argentina	Salta	0.488		
Argentina	Santa Fe	0.482		
Argentina	Chaco	0.456		
Argentina	Mendoza	0.449		
Argentina	Corrientes	0.441		
Argentina	La Rioja	0.412		
Argentina	Tucuman	0.376		
Argentina	Tierra del Fuego	0.281		
Argentina	San Luis	0.272		
Argentina	San Juan	0.124		
Brazil	**Mato Grosso**	**0.969**	**20-11-15 –21-1-15**	**21-3-15**
Brazil	Rondania	0.960		
Brazil	Sao Paulo	0.942		
Brazil	Amazonas	0.911		
Brazil	Minas Gerais	0.909		
Brazil	Paraba	0.862		
Brazil	Para	0.820		
Brazil	Rio de Janeiro	0.810		
Brazil	Alagoas	0.798		
Brazil	Tocantins	0.739		
Brazil	Distrito Federal	0.686		
Brazil	Sergipe	0.645		
Brazil	Parana	0.628		
Brazil	Pernambuco	0.628		
Brazil	Ceara	0.617		
Brazil	Goias	0.609		
Brazil	Bahia	0.592		
Brazil	Amapa	0.585		
Brazil	Rio Grande do Norte	0.578		
Brazil	Espirito Santo	0.549		
Brazil	Piaui	0.543		
Brazil	Rio Grande do Sul	0.535		
Brazil	Mato Grosso do Sul	0.509		
Brazil	Acre	0.488		
Brazil	Maranhao	0.472		
Brazil	Roraima	0.430		

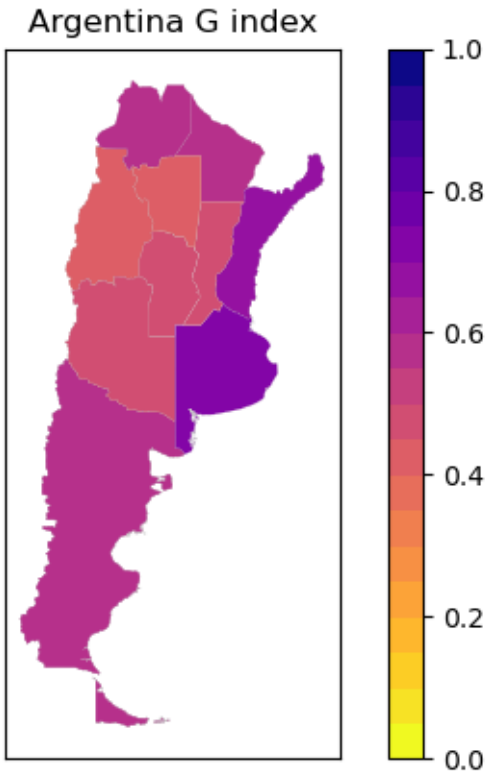
Brazil	Santa Catarina	0.354		
Colombia	**Cundinamarca**	**1.000**	**20-11-15 – 21-1-15**	**21-3-15**
Colombia	Bogota	0.993		
Colombia	Narino	0.990		
Colombia	Santander	0.972		
Colombia	Putumayo	0.971		
Colombia	Valle del Cauca (incl Cali)	0.952		
Colombia	Tolima	0.950		
Colombia	Atlantico	0.941		
Colombia	Cauca	0.903		
Colombia	Choco	0.896		
Colombia	Magdalena	0.864		
Colombia	Antioquia	0.836		
Colombia	Meta	0.802		
Colombia	Sucre	0.779		
Colombia	Boyaca	0.776		
Colombia	Risaralda	0.761		
Colombia	Cesar	0.756		
Colombia	Huila	0.751		
Colombia	Casanare	0.728		
Colombia	Cordoba	0.708		
Colombia	Amazonas	0.647		
Colombia	Caldas	0.622		
Colombia	Bolivar	0.603		
Colombia	Quindio	0.585		
Colombia	Guaviare	0.579		
Colombia	Norte de Santander	0.563		
Colombia	Arauca	0.531		
Colombia	Caqueta	0.412		
Colombia	Vichada	0.380		
Colombia	La Guajira	0.372		
Colombia	San Andres y Providencia	0.350		
Colombia	Vaupes	0.169		
Colombia	Guainia	0.093		
The Gambia	**Gambia (country level)**	**0.591**	**20-11-10 – 21-1-10**	**21-3-10**
Malawi	**Malawi (country level)**	**0.831**	**20-11-10 – 21-1-10**	**21-3-10**
Mozambique	**Mozambique (country level)**	**0.680**	**20-11-10 – 21-1-10**	**21-3-10**
India	**Kerala**	**0.521**	**20-11-15 – 21-1-15**	**21-3-15**
India	Nagaland	0.293		
India	Meghalaya	0.262		
India	Jharkhand	0.260		
India	Maharashtra	0.253		

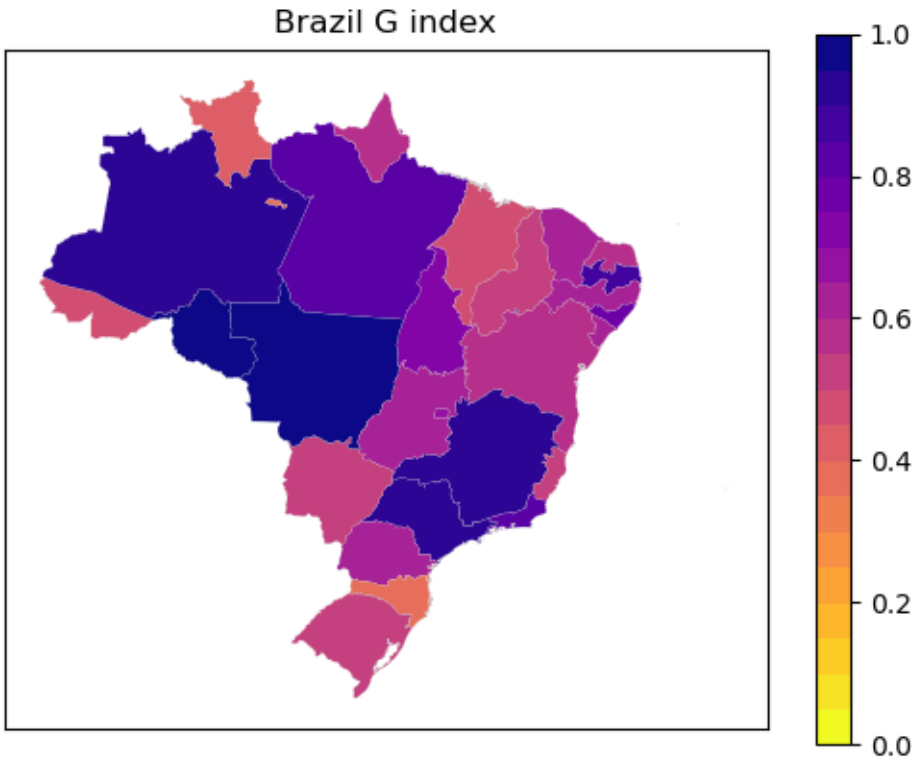
India	Uttarakhand	0.252		
India	Puducherry	0.216		
India	Bihar	0.187		
India	Odisha	0.168		
India	Rajasthan	0.151		
India	Goa	0.148		
India	Sikkim	0.108		
India	Mizoram	0.105		
India	Gujarat	0.102		
India	Chandigarh	0.099		
India	Chhattisgarh	0.086		
India	Dadra and Nagar Haveli	0.069		
India	Jammu and Kashmir	0.067		
India	Punjab	0.054		
India	Madhya Pradesh	0.044		
India	Assam	0.033		
India	Himachal Pradesh	0.032		
India	Telangana	0.029		
India	Manipur	0.028		
India	Karnataka	0.017		
India	Ladakh	0.016		
India	Andaman and Nicobar	0.010		
India	Tripura	0.008		
India	Tamil Nadu	0.004		
India	West Bengal	0.002		
India	Haryana	0.001		
India	Uttar Pradesh	0.001		
India	Andhra Pradesh	0.000		
India	NCT of Delhi	0.000		
India	Arunachal Pradesh	0.000		
Mexico	**Jalisco**	**0.841**	**20-11-10 – 21-1-10**	**21-3-10**
Mexico	Mexico City	0.369		
Pakistan	**Sindh**	**0.533**	**20-11-10 – 21-1-10**	**21-3-10**
United Kingdom	**Northern Ireland**	**0.606**	**20-11-15 – 21-1-15**	**21-3-15**
United Kingdom	North West	0.603		
United Kingdom	East Midlands	0.598		
United Kingdom	South West	0.598		
United Kingdom	South East	0.597		

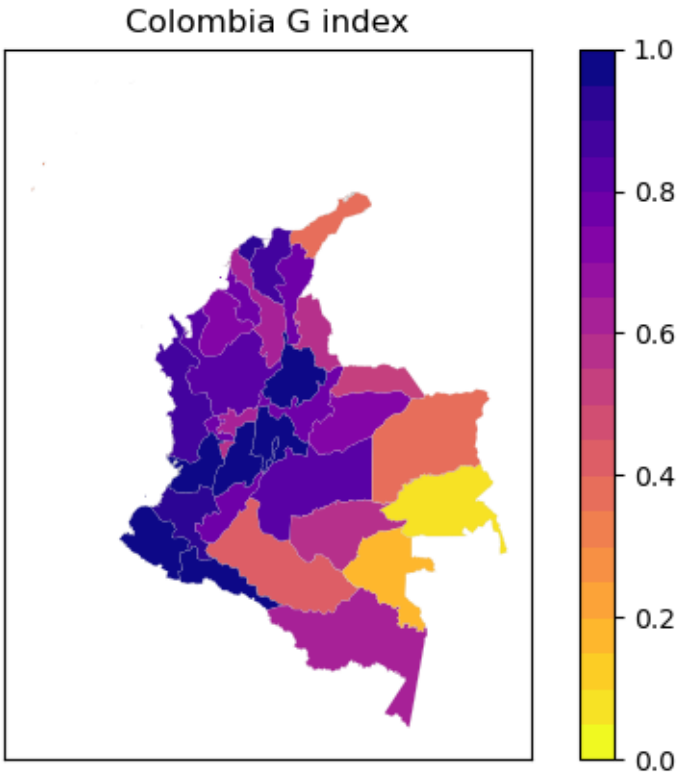
United Kingdom	East of England	0.586		
United Kingdom	London	0.574		
United Kingdom	Scotland	0.548		
United Kingdom	Yorkshire	0.526		
United Kingdom	Wales	0.431		
United Kingdom	West Midlands	0.598		
United Kingdom	North East	0.526		

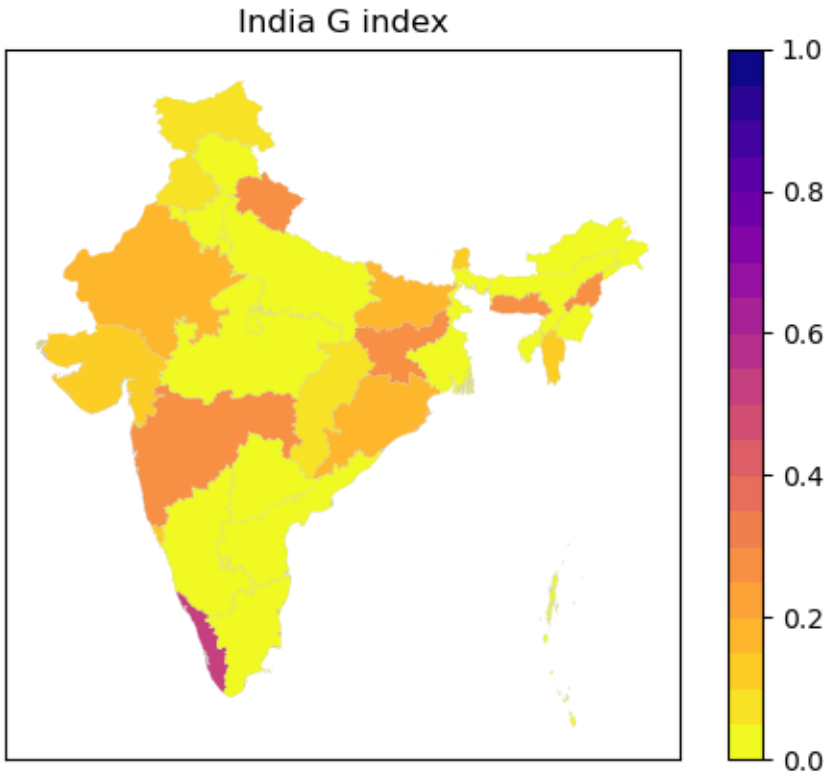
Figures section 1: Spatial heterogeneity in G index.

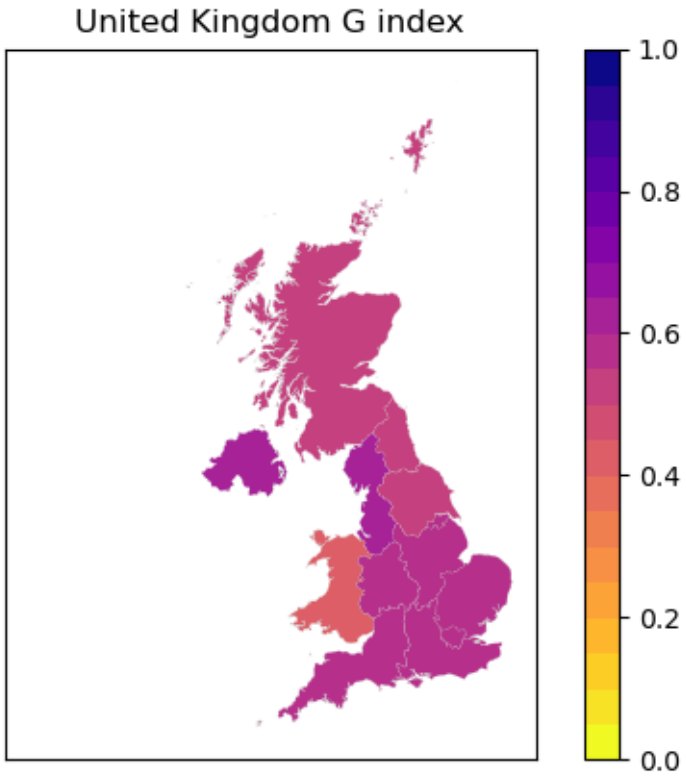
Note the color scale is identical for all maps shown. See Table 2 for values by region, lookback period used in input data, and target trial dates. Note that for Pakistan (Sindh), Mexico (Jalisco, Mexico City) and The Gambia, values are given in Table 2.




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

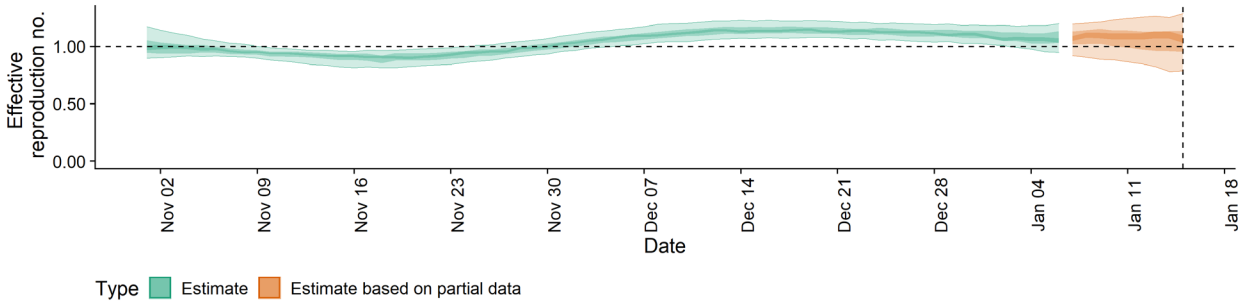

India G index
1.0
0.8
0.6
0.4
0.2
0.0


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.

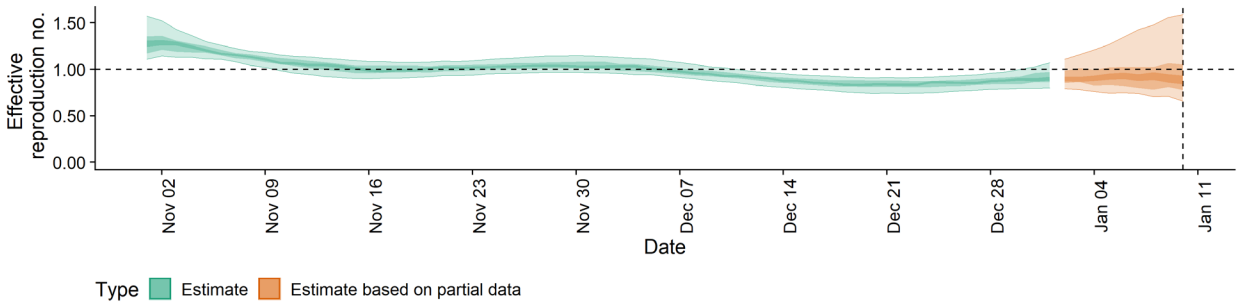
Buenos Aires Province (Argentina):



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

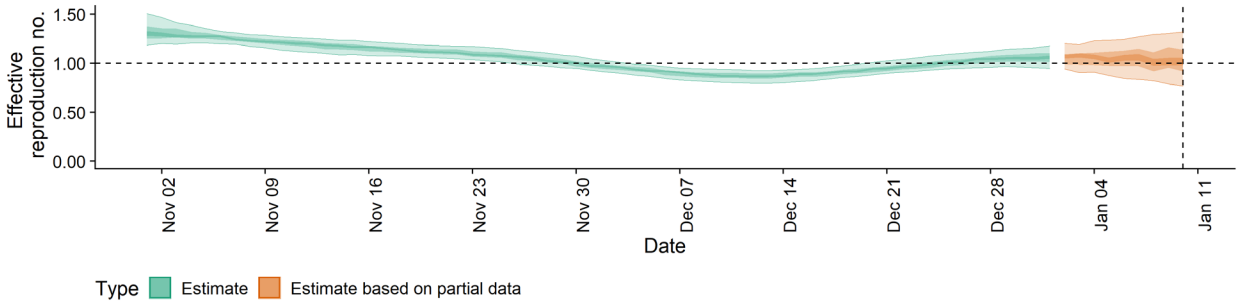
Mexico City (Mexico):

9



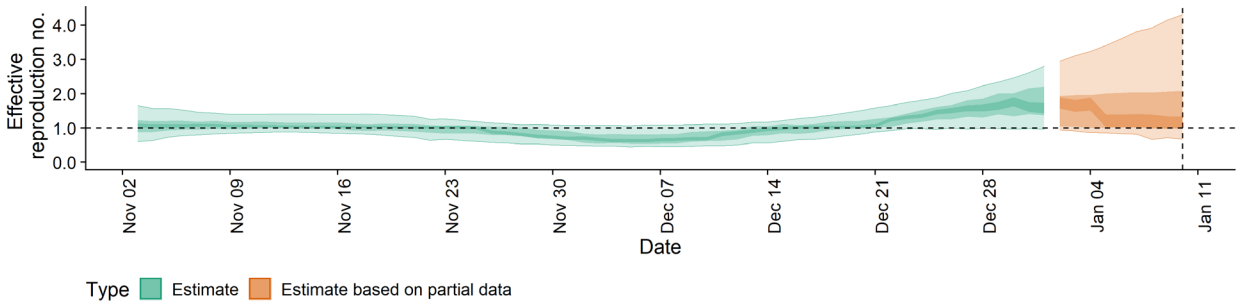
Estimated Rt values for Mexico City showing 50% and 0% credible intervals.

Sindh Province (Pakistan):



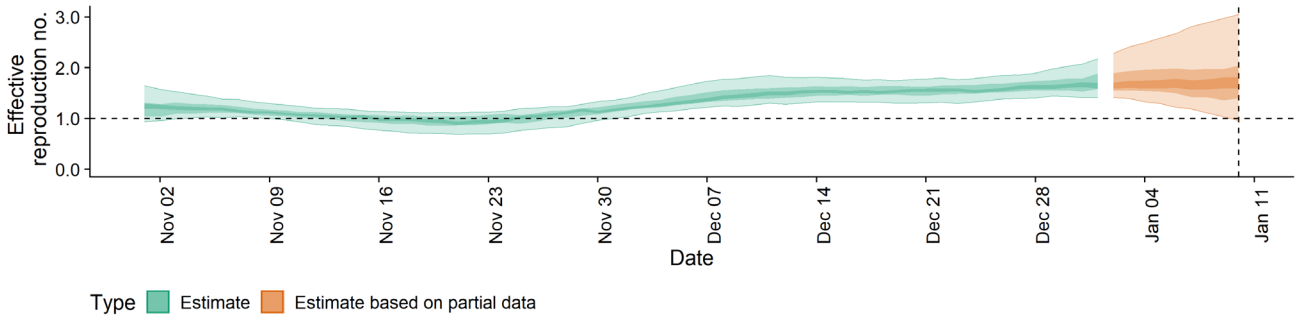
Estimated Rt values for Sindh province showing 50% and 90% credible intervals.

The Gambia:



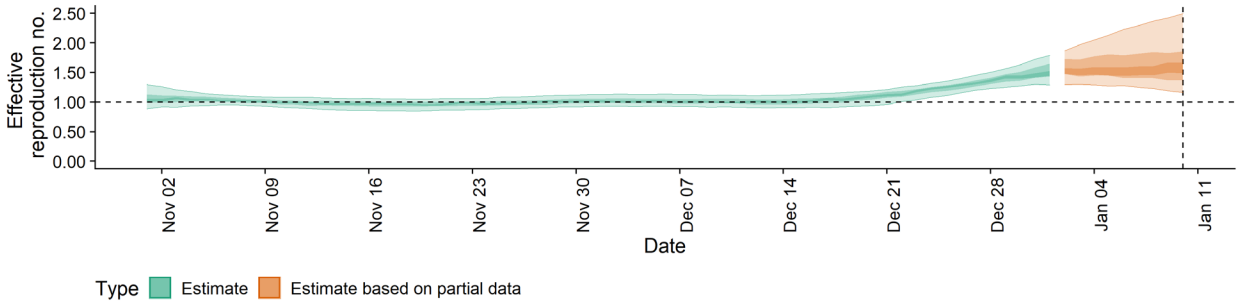
Estimated Rt values for The Gambia showing 50% and 90% credible intervals.

Malawi



Estimated Rt values for Malawi showing 50% and 90% credible intervals (G = 0.84).

Mozambique:



Estimated Rt values for Malawi showing 50% and 90% credible intervals (G = 0.68).

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.