Tracking the emergence of disparities in the subnational spread of COVID-19 in Brazil using an online application for real-time data visualisation.

Paul Mee[1,2,3], Neal Alexander[1,3], Philippe Mayaud[4], Felipe de Jesus Colon Gonzalez[1,2], Sam Abbott[1,2], Andreza Aruska de Souza Santos [4], Andre Luis Acosta [5], Kris V. Parag [6], Rafael H. M. Pereira [7], Carlos A. Prete Jr [8], Ester C. Sabino [9], Nuno R. Faria [6,9,10] , LSHTM Centre for Mathematical Modelling of Infectious Disease COVID-19 working group [11] , Oliver J Brady [1,2]

Affiliations

[1] Faculty of Epidemiology and Public Health, London School of Hygiene and Tropical Medicine, London, , United Kingdom.
[2] Centre for Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom.
[3] International Statistics and Epidemiology Group, London School of Hygiene and Tropical Medicine, London, United Kingdom.
[4] Oxford School of Global and Area Studies, University of Oxford, Oxford, United Kingdom.
[5] Instituto de Biociências, Universidade de São Paulo, Rua do Matão, Trav. 14, n. 321, 05508-900 São Paulo, São Paulo, Brazil
[6] MRC Centre for Global Infectious Disease Analysis, J-IDEA, Imperial College London, London, United Kingdom.
[7] Instituto de Pesquisa Econômica Aplicada (Ipea), Brasilia, Brazil.
[8] Department of Electronic Systems Engineering, University of São Paulo, São Paulo, Brazil.
[9] Departamento de Molestias Infecciosas e Parasitarias and Instituto de Medicina Tropical da Faculdade de Medicina da Universidade de São Paulo (FMUSP), São Paulo, SP 05403–000, Brazil.
[10] Department of Zoology, University of Oxford, Oxford, United Kingdom.
[11] See acknowledgements for the list of CMMID COVID-19 working group members at the time of publication

Abstract

Background - Brazil is one of the countries worst affected by the COVID-19 pandemic with over 14 million cases and 375,000 deaths reported. Comparison of real-time local COVID-19 data between areas is essential for understanding transmission, measuring the effects of interventions and predicting the course of the epidemic, but are often challenging due to different population sizes and structures.

Methods – Here we describe the development of a new app for the real-time visualisation of COVID-19 data in Brazil at the municipality level. In the CLIC-Brazil app, daily updates of case and death data are downloaded, age standardised and used to estimate reproduction number (R_t). We show how such platforms can perform real-time regression analyses to identify factors associated with the rate of initial spread and early reproduction number. We also use survival methods to predict the likelihood of occurrence of a new peak of COVID-19 incidence.

Findings – After an initial introduction in São Paulo and Rio de Janeiro states in early March 2020, the epidemic spread to Northern states and then to highly populated coastal regions and the Central-West. Municipalities with higher metrics of social development experienced earlier arrival of COVID-19 (decrease of 3.8 days (95% CI:1.8-5.8) for each 10% increase in social development index). Differences in the initial epidemic intensity (mean *Rt*) were largely driven by geographic location and the date of local onset.

Interpretation - This study demonstrates that platforms that monitor, standardise and analyse the epidemiological data at a local level can give useful real-time insights into outbreak dynamics that can be used to better adapt responses to the current and future pandemics.

Funding - This project was supported by a Medical Research Council-São Paulo Research Foundation (FAPESP) CADDE partnership award (MR/S0195/1 and FAPESP 18/14389-0)

Research in context panel

Evidence before this study

Brazil is one of the countries worst affected by the COVID-19 epidemic. The spread of the epidemic across the country has been highly heterogeneous and this unfolding pattern can potentially reveal important real-time information about epidemiological risk factors and the effectiveness of non-pharmaceutical interventions. A variety of COVID-19 data dashboards provide real time insights, but few present data down to local levels or perform the data standardisation procedures necessary to reveal the true patterns of spread and the factors driving the differences in severity between places. Several retrospective analyses using reported case and population-based serological datasets have been performed, but only for specific areas of Brazil and only at particular times in the epidemic. As a result, such studies have, unsurprisingly, found conflicting risk factors e.g. socio-economic status has gone from a risk to a protective factor.

Added value of this study

This study describes the development of the COVID-19 Local Information Comparison (CLIC Brazil) app (https://cmmid.github.io/visualisations/lacpt) and demonstrates the scientific insights possible with such local, standardised, real-time platforms. The app provides access to standardised case and death time-series of epidemic data for each of the 5570 municipalities in Brazil and is updated daily.

We found that the COVID-19 epidemic established earliest in municipalities with higher population density, higher social development index (SDI) and greater percentage of residences with piped sanitation, and latest in areas geographically distant from the main population centres. By looking at local early epidemic progression rates (R_t) we were able to map heterogeneity in transmission intensity across Brazil and show its correlation with socioeconomic and connectivity measures.

Implications of all the available evidence

The app provides a useful tool for comparing epidemic severity between places and over time at a finer grained spatial resolution. It shows how rapid insights can be gained into the unfolding risk factors of epidemics which is increasingly important when such risk factors rapidly evolve due to non-pharmaceutical interventions, emergence of new variants and vaccination.

Brazil is currently experiencing a second wave of the epidemic, driven by the emergence of a more transmissible variant (P1). Continual tracking of this phase of the epidemic using methods demonstrated in our analyses will be important to assess similarities and differences in the spatial spread of different COVID-19 epidemic waves.

Introduction

COVID-19 is a new respiratory and multi-organ illness caused by infection with the severe acute respiratory syndrome coronavirus type-2 (SARS-CoV-2) which emerged in December 2019 in Wuhan, China. As of 20th April 2021 over 142 million COVID-19 cases and over 3,000,000 deaths have been reported worldwide[1]. Brazil is one of the worst affected countries with over 14 million cases and 375,000 deaths reported [1]. Heterogenous patterns of propagation of the virus across the country have been driven by a complex intersection of causative factors including; continued movements of people between urban centres throughout the epidemic, differential imposition of interventions designed to reduce transmission and relative isolation of municipalities from the major population centres [2]. Currently the country is experiencing a second wave of the epidemic driven by a viral variant which arose in the Amazonas region and has quickly spread throughout the country [3]. Comparisons of incidence between different local areas can give important insights into the patterns of spread and the burden of an epidemic and help to separate generalisable from context specific transmission trends. Such comparisons are complicated by differences in the characteristics of local populations that affect the risk of disease even if levels of infection are equal. In particular, age is a major risk factor for infection with SARS-CoV-2 and subsequent COVID-19 disease. Consequently, differences in the age distribution need to be taken into account when comparing local areas [4–6]. The differences in epidemic severity between places could also be driven by sociodemographic factors, ethnicity, the relative isolation of different regions and the levels of implementation and effectiveness of non-pharmaceutical interventions. A serosurvey conducted in May and June of 2020 in cities across Brazil found evidence that prevalence of SARS-Cov-2 antibodies , an indicator of prior infection, was higher for those; living in crowded conditions, of non-white ethnicity and those in the lowest socio-economic groups [7]. In contrast a study of data from the early stages of the epidemic in Brazil , up to May 2020, found that the those in higher socio-economic groups were more likely to have a positive test for COVID-19 [8]. This may imply a changing risk profile over time or may reflect differential access to testing among socio-economic groups.

With the increased roll-out of vaccinations, local differences in vaccine uptake also need to be considered [9]. All of these aspects could impact on the rate of spread and onset time of the epidemic in a given locality. Quantifying the role of these components, and the interplay between them, is important for understanding patterns of past infection and the likely severity of future waves of infection.

The field of real-time analysis of infectious disease data is rapidly expanding, in part due to greater automation, digitisation and online sharing of data [10]. Projects such as the Johns Hopkins University Covid-19 Dashboard [11] aim to provide a global overview of cases and deaths with the goal of making international comparisons [12] and a number of sub-national-level real-time data dashboards have also been established for finer scale domestic comparisons such as that for Italy [13]. Such dashboards are useful for rapid situation reports, yet direct comparison between regions with differing age distributions and onset times offers limited epidemiological insight into the rate of spread and local burden of the epidemic.

Websites such as EpiForecasts [14,15] , the Covid-19 Forecasting Hub [16–18] and 'Short-term forecasts for multiple countries' [19] aim to make and compare short-term projections of disease incidence using mathematical and statistical models. As part of this process, some models aim to estimate the effective reproduction number R_t, an estimate of the average number of new infections that will occur from each infected person. Real-time estimates of R_t over time [20,21] are useful for planning interventions to mitigate the impact of the epidemic [22] . Accurate estimation of R_t is complicated by the need to correct for the delays between infection and reporting of cases of disease and under-ascertainment of cases. Due to the computational resources required to run these forecasting

models , most existing analysis dashboards only give predictions at the national or first administrative level (e.g. State in Brazil) and are not updated daily to reflect the latest situation [23]. There is a need for a new class of dashboards that are able to perform basic data standardisation to account for differing population age structures and allow for regional comparisons. Additionally, such dashboards should provide local summaries of key epidemiological parameters and support rapid data analyses of outbreak dynamics whilst retaining the contemporary focus of real-time data streams.

In response to this we have developed an online application for the real-time visualisation of COVID-19 cases and death data in Brazil at the municipality (second-level administrative division) level. This allows real-time comparisons of the development of the epidemic in Brazil to be made at a local level to allow local decision makers to track and compare epidemic progression rates between different areas. The COVID-19 Local Information Comparison (CLIC Brazil) [https://cmmid.github.io/visualisations/lacpt] has been active since May 2020, early in the Brazilian epidemic. The data underlying the app are updated daily and relevant local data summaries and analytics re-computed. Here we describe the CLIC Brazil application and the insights about the early evolution of the COVID-19 outbreak in Brazil that it has helped reveal.

Methods

Data Sources

The numbers of COVID-19 cases and deaths, aggregated by municipality were automatically downloaded daily from the Brazil.io COVID-19 project repository [24] .This repository contains data extracted from the bulletins of state health secretariats. Data on the distribution of the population by age and the sociodemographic characteristics of each municipality were obtained from themost recent national demographic census, run by the Instituto Brasileiro de Geografia e Estatística (IBGE) in 2010 [25]. Data on the age distribution of COVID-19 cases were derived from case reports throughout Brazil between 2nd February and 25th March 2020, collected by the Brazilian Ministry of Health and were used with their permission. The travel time in hours from each municipality to the most populous metropolitan area in the state was calculated using WorldPop population data [26,27] and the Malaria Atlas Project travel time friction surface using the Malaria Atlas [27] accumulated cost route finding algorithm within the malariaAtlas R package [28,29] . The socio-demographic Index (SDI) is a composite average of the rankings of the incomes per capita, average educational attainment, and fertility rates scaled between 0 (lowest) and 1 (highest) [30] . The geographic region in which each place was located was assigned using a standardised designation which groups the States and the Federal District of Brasilia into five macro regions of national planning [31] **(Fig 1)**.Data on COVID-19 cases and deaths were censored on January 14th 2021. Data on the types of non-pharmaceutical interventions implemented, and the dates of their announcement were extracted from data collated by the Cepal Observatory with edits and updates on timing of interventions at the municipality and state level by de Souza Santos [3,8]. Full details of the data sources and data processing are included in the **supplementary materials.**

Development of the application

The COVID-19 Local Information Comparison (CLIC Brazil) application was developed using the R package "shiny", version 1.5.0 [32]. CLIC Brazil provides users with options for graphical display of information and all computation required is handled remotely. Plots are generated using the R package "ggplot2" version 3.3.2 [33] . Spatial data presented in the form of maps are portrayed using the "leaflet" R package version 2.0.3 [34]. Screenshots from the app are shown in **Figures S3 to S5** in the Supplementary materials. All code described in the paper can be downloaded from this github repository - https://github.com/Paul-Mee/clic_brazil .

Calculating the comparable measure of standardised incidence

To enable comparison of COVID-19 case counts between municipalities with different age structures and thus different probability of disease given infection, we standardised each municipality's incidence to the national-level age structure. (See **supplementary materials** for details)
When comparing the progression of the epidemic in different municipalities over time, a comparable outbreak start criterion has to be established. We defined the arrival of COVID-19 in a given municipality as the day in which cumulative standardised incidence first exceeded 1 case per 10,000 residents.

R_t Estimation

The raw case count data was adjusted to account for the differential probabilities of COVID-19 cases being reported on each day of the week (heaping). (See **supplementary materials** for details)
The EpiFilter algorithm was used to estimate R_t [35], this method uses a recursive Bayesian filter to derive estimates from a time series of all incident cases. The serial interval (SI), defined as the time between the onset of symptoms in the source of infection and in the recipient, was modelled as a gamma distribution with a mean of 6.5 days with a standard deviation of 4.03 days [36] .A fixed value of 10 days was used for the delay between symptom onset and case reporting , this was consistent with an average value of 10.2 days for 2,420,904 suspected COVID-19 cases reported between March 1st and August 18th 2020 in all the state capitals and Federal District of Brazil in the e-SUS notification system [37] .Predictions were only made for municipalities with more than 30 days of data and more than 200 COVID-19 cases reported, to allow sufficient data for the algorithm to give reliable estimates[38]. The resultant R_t curves were plotted to enable comparison between selected cities in the CLIC Brazil app.

Regression analyses

Using data from the CLIC Brazil app, regression models were developed to quantify how the impact of COVID-19 on municipalities could be explained by sociodemographic characteristics and spatial connectivity.

Analysis of factors associated with time for the epidemic to arrive in a municipality.

Here we sought to assess which characteristics of a municipality could be associated with earlier arrival of COVID-19. For this purpose, the arrival of COVID-19 in Brazil was set at 31st March 2020, which is the date at which the first municipality exceeded our arrival threshold incidence of 1 per 10,000 residents. The time to arrival for each municipality was calculated as the difference in days between the national arrival date and the day the standardised incidence first exceeded this threshold incidence.

A Tobit regression analysis [39] as implemented in the R package VGAM [40] was carried out as the outcome was right censored due to some municipalities not yet having reached the outbreak threshold incidence. Initially a series of univariable regression models were developed to test whether each covariate was individually associated with the outcome. Following this, multivariable models were developed. Geographic region was included in all models as a fixed effect to allow the variations in the outcome unexplained by other covariates to be partially accounted for. Both forward and backward stepwise regression approaches were used to develop multivariate models. In the forward stepwise approach variables were added and retained if the likelihood ratio test statistic indicated that they significantly improved the model fit. In the backward stepwise method all variables were included and dropped iteratively to assess whether their omission worsened the model fit. The final model was selected as the 'best' model from the two methods. Sensitivity analyses were carried out to assess the extent to which the results were dependent on the definitions of the epidemic start and threshold incidence.

Analysis of factors associated with the rate of growth in the early stages of the epidemic.

We calculated the mean value of R_t over a window between 30 and 150 days from the beginning of the epidemic in each municipality. Day 0 was defined by our outbreak threshold, standardised incidence greater than 1 case per 10,000 individuals in the particular municipality. Calculating R_t using data within the window of 30-150 days since the arrival of COVID-19 in each municipality balanced the need to include enough data for accurate estimation [38] with the need to estimate R_t before the build-up of substantial immunity or reactive interventions (therefore approximating R_0). A series of univariable linear regression models were developed to test whether individual covariates were associated with the outcome . In order to develop a model in which the effect estimates had a multiplicative effect the outcome was transformed as log(mean *Rt*) and the model estimates were then back transformed (exponentiated) .

Initial analyses indicated that the geographic region in which the municipality was located (Central-West, North, North-East, South, South-East), and the start date of local epidemic, divided into 3 roughly equal categories : (19th March to 26th April, 27th April to 12th May and 13th May to 17th July) , were significantly associated with mean R_t, with relatively large effect estimates, . Hence these two variables and the interaction between them were included in all subsequent multivariable models. For the interaction terms, estimated marginal means for all combinations of covariates were calculated for the final model. Sensitivity analyses were carried out to assess the effect of varying the end point of the time-window over which mean R_t was calculated.

Predicting whether a new maximum incidence will occur

We used Cox regression as implemented in the "coxph" package in R [41] to estimate the probability of each municipality surpassing its previous maximum daily standardized incidence (i.e. a new "record" incidence) within the following 30 days . The analysis time was the number of days since the start of the epidemic (cumulative standardised incidence exceeded 1 case per 10,000). The event of interest was the setting of a new record incidence, which in general occurs more than once. (Full details are included in the **supplementary materials**)

The reporting guidelines in the reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement were used [42].

Results

To compare the spatial progression of the COVID-19 epidemic in Brazil, we mapped cumulative standardised disease incidence in **Fig. 1**

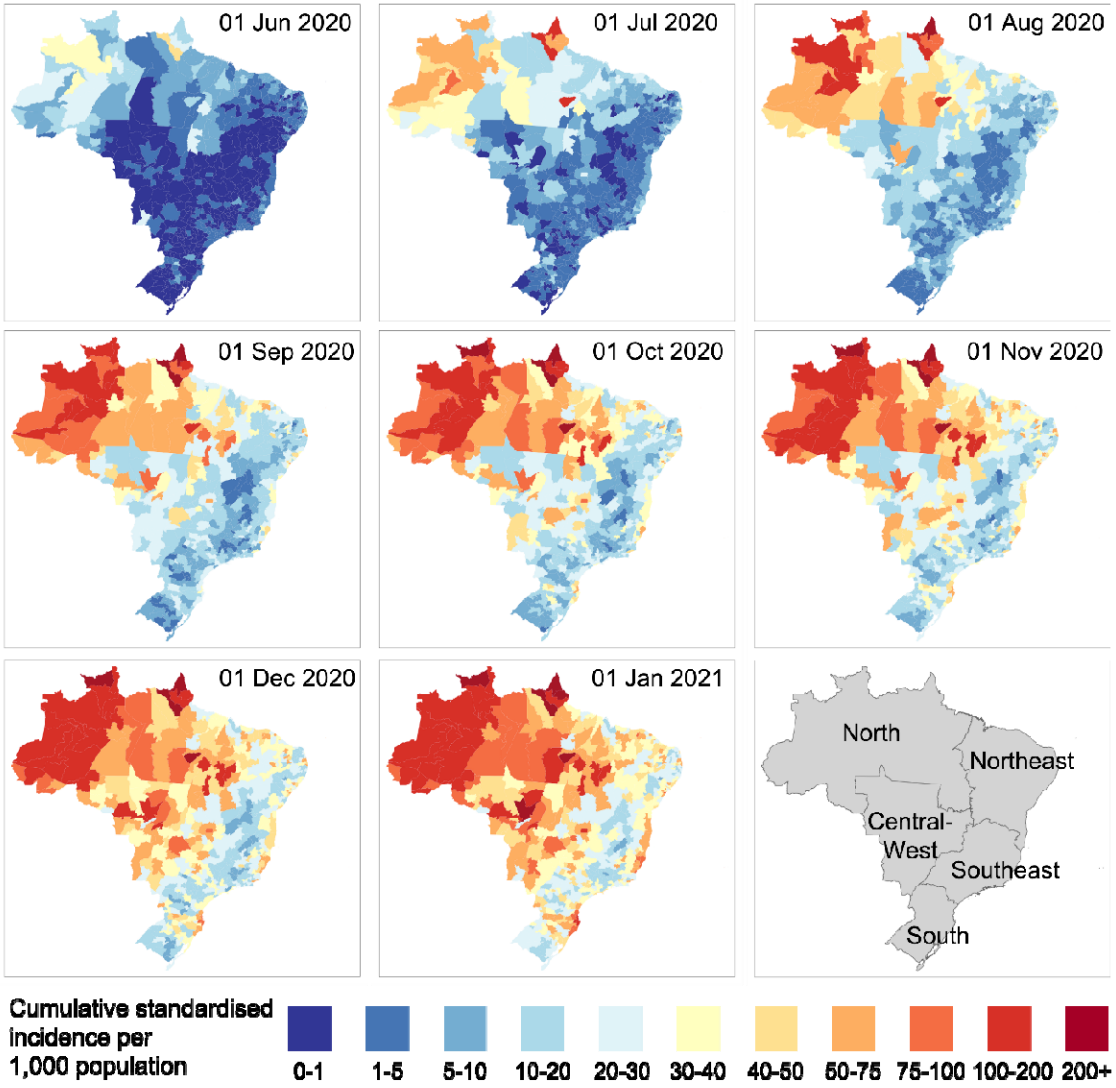


Figure 1: Spatial progression of the COVID-19 epidemic in Brazil. Standardised case incidence is a measure that allows comparison between municipalities with different population sizes and age structures. Standardised incidence for all 5570 municipalities has been aggregated to microregion (n = 557) by population-weighted averaging for visualisation purposes. The bottom right panel includes a map of the six macro regions of Brazil for reference.

Despite the first introductions and local SARS-CoV-2 transmission events occurring in São Paulo and Rio de Janeiro states in early March 2020 [43] the focus of the outbreak quickly shifted to the North region of the country where the first big outbreaks occurred in late May/early June, particularly in the border states of Amazonas, Roraima and Amapá (Fig. 1). By July 2020, COVID-19 transmission was widespread across the North region and began to spread to major coastal cities, particularly in the Northeast. By August, transmission had spread along the highly populated coastal areas and into the Central West region. Through August and September SARS-CoV-2 spread to the final transmission free areas in sparsely populated inland regions and in the far South. By October 1st, transmission was widespread throughout the whole country. During November-December 2020 the re-emergence of large outbreaks occurred throughout the Central-West region alongside renewed

growth in coastal cities of the South East and South that, as of 1st January 2021, have cumulative standardised incidence rates comparable to some of the worst affected areas in the North.
To compare the trajectory of COVID-19 outbreaks in local areas once the first wave of epidemic had begun, we plotted cumulative standardised case counts per municipality in different states and regions in Brazil (Fig. 2). This revealed that the outbreak was comparatively faster growing and reached higher cumulative incidence in the North region of Brazil (Fig. 2A and F). Within this region, the most northerly states of Amazonas, Amapá and Roraima were the most severely affected with some municipalities experiencing cumulative case incidence rates as high as 45%. Areas in the Southeast and, until recently, South regions have suffered comparably smaller slower growing outbreaks (Fig. 2D-F). Outbreak trajectories in the Central-West and Northeast regions were between the high rates observed in the North and low rates observed in the South in the first wave, but have since diverged with rates beginning to rise again in the Central-West region (Fig. 2F). Despite these trends, there was considerable within region and within state heterogeneity in outbreak trajectories, suggesting factors other than just geographical proximity were important in shaping the trajectory of the epidemic. Plots for specific municipalities can be viewed on the CLIC Brazil app [https://cmmid.github.io/visualisations/lacpt]

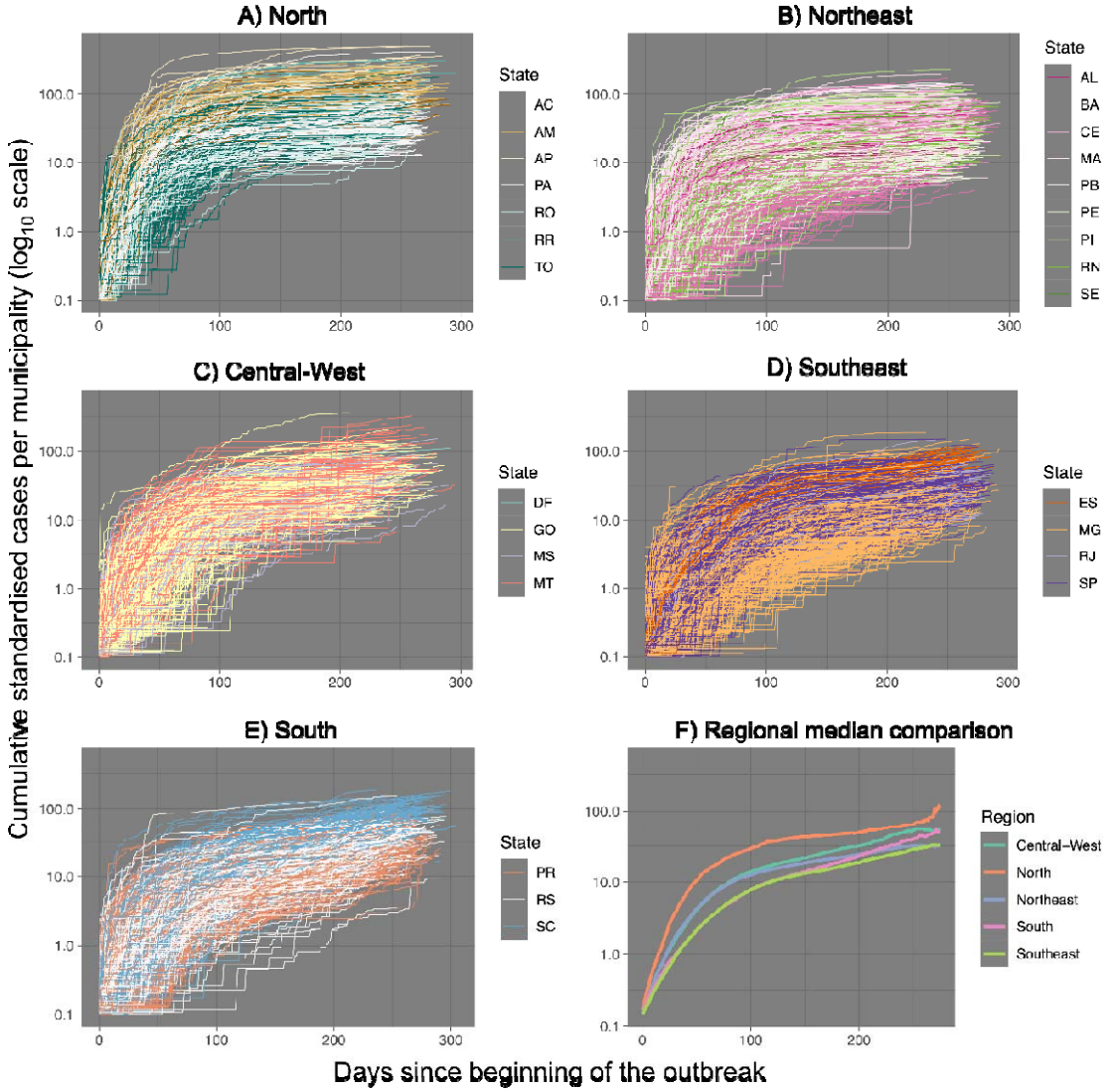


Figure 2: Comparison of outbreak trajectories in different regions of Brazil. Each line represents a municipality within the geographical area. Cases per municipality have been age- and population-

standardised, plotted cumulatively and aligned to the date of detection of the first COVID-19 cases in each municipality (defined as an incidence of greater than 0.1 standardised cases per 1,000 residents). Only municipalities that have reported 50 or more cases are shown. Panel F compares region median trajectories.

Between 28th February and 27th March 2020, a range of state-level restrictions were announced to limit the spread of SARS-CoV2 including declaring an emergency, industry, retail, service and transport restrictions and school closures. At the time these interventions were announced, only a very small number of municipalities had reported a single COVID-19 case (51 of 5,570 municipalities). This indicates that the announcement of interventions in Brazil occurred before the first COVID-19 cases appeared in the majority of municipalities (mean of 52.1 days before the first case was reported in pre-emptive municipalities). Even in municipalities where interventions were announced reactively there was only a mean 2.2 days between reporting the first case and their announcement (red area in **Fig. 3).**

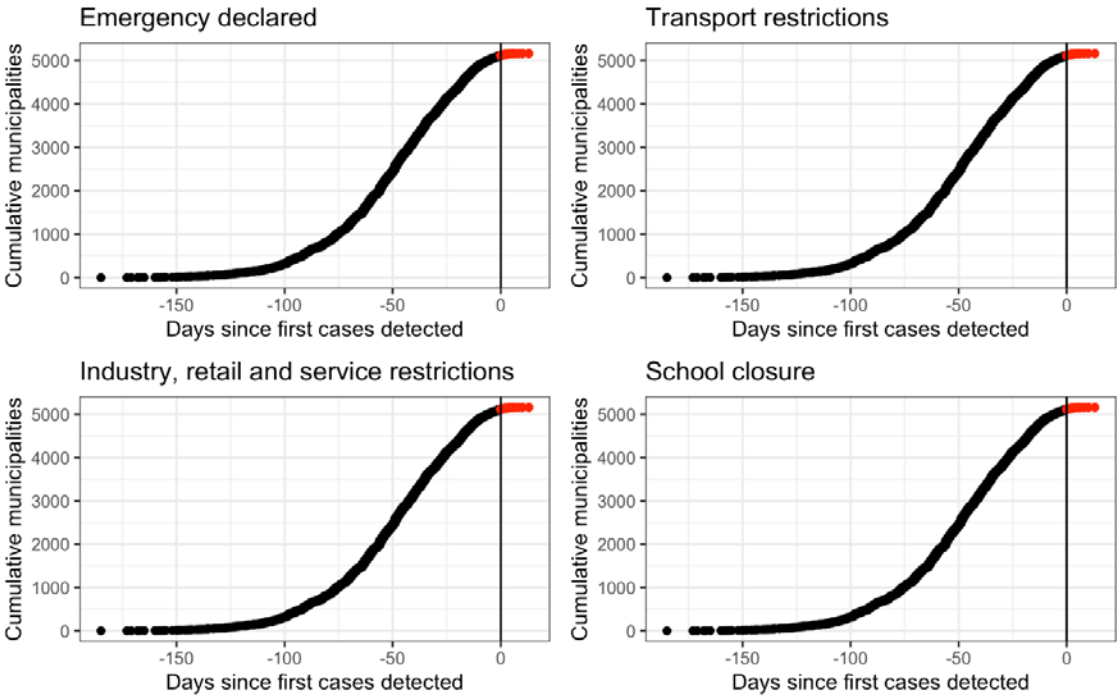


Figure 3: Timing of announcement of interventions relative to the first COVID-19 cases being reported (black vertical line). Dots in black represent municipalities where restrictions were announced prior to any cases being reported, while dots in red show municipalities where restrictions were announced after cases were reported. 5,160 municipalities were included in this analysis.

Table 1 Tobit regression analysis of time to outbreak for each municipality (days since 31st March 2020). - Starting incidence 1 case per 10,000 residents - Epidemic onset incidence = 10 cases per 10,000 residents.

Characteristic of the municipality		Frequency	Median value (Interquartile range)	Unadjusted (univariable) model estimates [95% CI]	Adjusted (multivariable) model estimates [95% CI]
Geographic region	Central West	437		-	-
	North	448		-31.7 [-36.1,-27.3]	-37.1 [-41.4,-32.8]
	North East	1754		-15.4 [-18.9,-11.9]	-4.1 [-7.8,-0.4]
	South	1140		3.3 [-0.4,7.0]	19.7 [15.9 ,23.4]
	South-East	1605		2.9 [-0.7,6.4]	25.2 [21.5,29.0]
Population density (log_n) (population/km^2)			3.21 (2.47 - 3.98)	-5.4 [-6.1,-4.8]	-8.0 [-8.8 ,-7.3]
Percentage of residences with piped water			72.41 (56.47 - 84.73)	0.0 [0.0,0.0]	0.0 [0.0 ,0.1]
Percentage of residences with piped sewage or septic tanks			37.73 (12.81 - 70.62)	0.0 [0.0,0.0]	-0.2 [-0.2 ,-0.1]
Travel time (log_n) by road to most populous municipality in the state (hours)			2.96 (2.33 - 3.44)	-1.7 [-3.1,-0.2]	2.9 [1.4 ,4.3]
Social Development Index (SDI)			0.25 (0.22 - 0.28)	14.3 [-4.5,33.1]	-38.1 [-58.3 ,-17.9]

Our "outbreak" threshold (standardised incidence of 10 cases per 10,000) was first exceeded on April 12th 2020. By the censoring date for this study of January 14th 2021 nearly all municipalities (5,382 out of a total of 5,384 for which covariate data was available) had exceeded the outbreak threshold incidence.

Consistent with the patterns of observed spread in **Figures 1 and 2**, the univariable analyses suggested that municipalities in the North region exceeded the outbreak threshold earlier, followed by those in the Central-West and Northeast regions and finally the South and Southeast regions. **(Table 1).** After adjusting for geographic region, the epidemic can be seen to have arrived earliest in those municipalities with higher population density, higher social development index (SDI) and greater percentage of residences with piped sanitation. There was evidence that municipalities further from the main population centres in the state had a later arrival of the epidemic. Considering the magnitudes of the effect estimates, a 10% increase in the population density shortened the arrival time by 0.8 days [95% CI:0.7-0.9]. An increase of 10% in the travel time to the largest city in the State was associated with a delay in arrival of 0.3 days [95% CI 0.1 -0.4] . An increase of 0.1 units in the SDI was associated with a decrease of 3.8 days [95% CI 1.8 -5.8] in the time to arrival.

Sensitivity analyses were carried out to investigate the effect of changing the threshold incidence for epidemic arrival from 5 to 15 cases per 10,000 **(Tables S1 and S2 Supplementary materials).** The interpretation of the direction of the effect of the included covariates remained the same within this range, whilst there were variations in the magnitude of the effect **(Table S3 Supplementary materials)**.

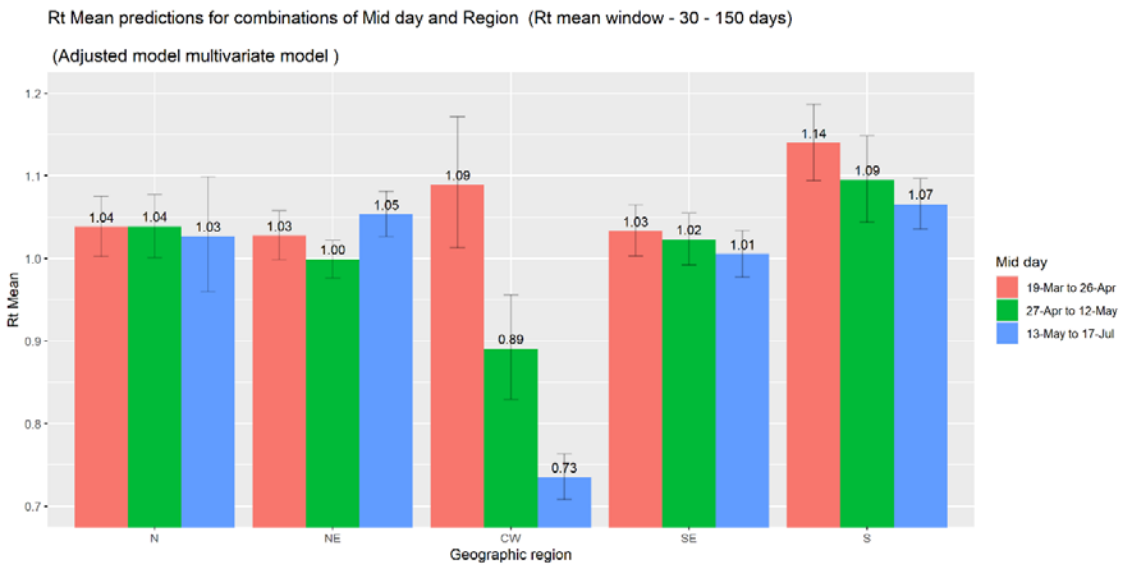
Determinants of rate of growth of the COVID-19 epidemic

Table 2. Summary of unadjusted and adjusted multivariate mean *Rt* linear regression model. Mean *Rt* is calculated over a window from 30 to 150 days since the arrival (standardised incidence > 1 case per 10,000) of COVID-19 in the respective municipality.

Characteristic of the municipality		Median value (Interquartile range)	Freq	Unadjusted (univariable) model estimates (95% CI)	Adjusted (multivariable) model estimates (95% CI)
Geographic Region	Central West		124	1	- [1]
	North		224	1.214 [1.170,1.260]	- [1]
	North East		562	1.245 [1.204,1.287]	- [1]
	South		247	1.374 [1.326,1.426]	- [1]
	South-East		462	1.32 [1.276,1.365]	- [1]
Date of local epidemic start (standardised incidence > 1 case per 10000)	19-Mar to 26-Apr		520	1	- [1]
	27-Apr to 12 May		532	0.961 [0.940,0.982]	- [1]
	13-May to 17-Jul		567	0.944 [0.923,0.965]	- [1]
Population density (log_n) (population/km^2)		3.76 (2.77 - 4.76)		1.025 [1.021,1.030]	1.000 [0.994,1.007]
Percentage of residences with piped water		76.67 (59.41 - 89.02)		1.002 [1.001,1.002]	1.001 [1.000,1.001]
Percentage of residences with piped sewage or septic tanks		46.28 (16.54 - 78.16)		1.002 [1.001,1.002]	1.001 [1.000,1.001]
Travel time (log_n) by road to most populous municipality in the state (hours)		2.98 (2.31 - 3.48)		0.981 [0.967,0.993]	0.975 [0.962,0.989]
Social Development Index (SDI)		0.25 (0.22 - 0.30)		1.711 [1.449,2.020]	0.903 [0.743,1.096]

[1] As model contains an interaction between geographic region and start day the effect estimates for the interaction terms are shown in the expanded table below

Figure 4 Estimated marginal mean values of *Rt* mean in the final adjusted multivariate linear regression model [*].



[*] The Codes for geographic region are ; North (N), Northeast (NE), Central-West (CW), Southeast (SE) and South (S) see map in **Figure1.**

To measure the intensity of the epidemic in each municipality after arrival we calculated the mean reproduction number (R_t) over the early phase of the epidemic. A total of 1,619 municipalities contained sufficient data for R_t calculation (ie, at least 30 days of data from the first case report and more than 500 cumulative cases) with estimated values ranging from 0.181 to 1.323.

From the unadjusted analyses **(Table 2),** it can be observed that the early epidemic was more intense on average in all other regions than the Central West, with regression coefficients indicating that mean R_t ranged from an increase by a factor of 1.21 in the North region to 1.37 in the Southern region. Epidemic intensity decreased in the two latter time periods compared to the first, though the effect was significantly smaller than that for geographic region; the regression coefficients for the latter two time periods indicated that the effect of time was to decrease mean R_t by factors of 0.96 and 0.94 respectively . In the multivariate model, a high degree of heterogeneity across space and time in mean R_t was seen **(Fig 4).** In the Central West region mean R_t decreased from 1.09 to 0.73, over the three time periods. There was no statistical evidence for a change in the other regions, with an increase of the point estimates from 1.03 to 1.05 in the North-East region and a small decrease in all other regions

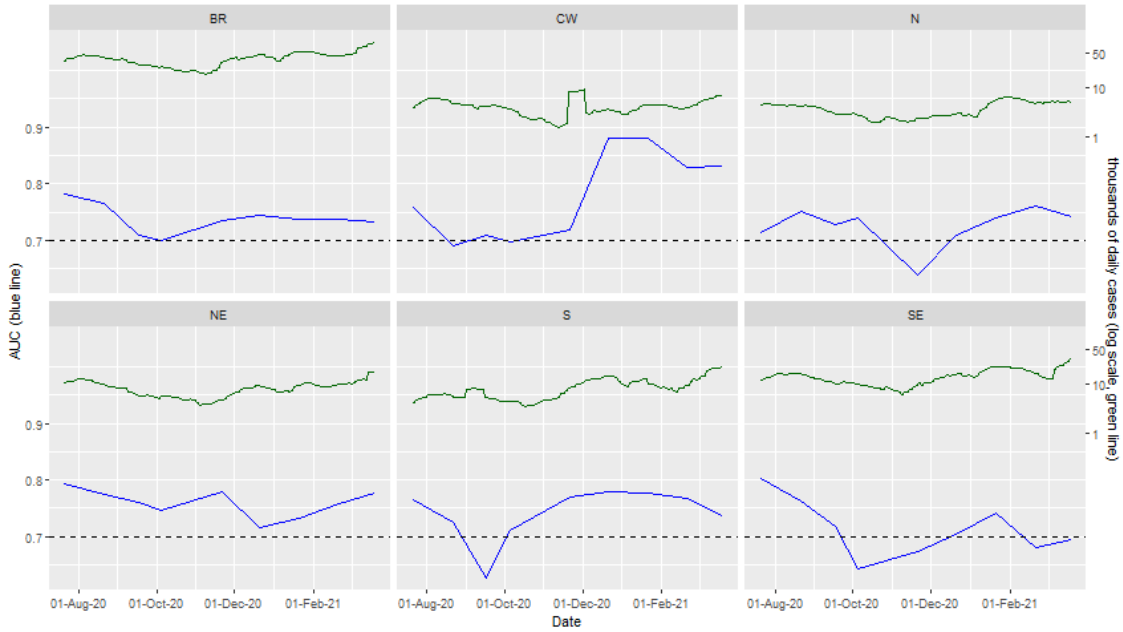
The effect on mean R_t for factors other than geographic region were relatively small indicating that a large amount of the variation was not explained by these factors. From the univariable analysis it was seen that municipalities that were more densely populated, with higher levels of provision of piped water and sanitation or a higher SDI had higher mean R_t values **(Table 2),** whilst those further from the main population centre in the state had lower values. In the adjusted multivariable analysis, the statistical evidence for an association with population density and SDI was lost and the association with the percentage of residences with piped water and piped sanitation weakened.

There remained strong evidence that mean Rt decreased with increasing travel time to the most populous municipality in the state coefficient = 0.975 [95% CI 0.962,0.989] indicating that a 10% increase in population density would be associated with an increase in mean Rt of 0.098.

A sensitivity analysis was carried out in which the end date for the calculation of mean R_t was adjusted over a range from 100 to 180 days. The unadjusted and adjusted models for the extreme values are presented in **Tables S4 and S5 in the Supplementary materials**. With the end of the time window set to 100 days, there was some evidence from the multivariate model for an association with population density, however the magnitude of the effect was small (coefficient = 0.991 [95% CI 0.985,0.996])

Predicting new maximum values of incidence

Figure 5 shows the values of the area under the ROC curve (AUC) for the ability to predict a new maximum incidence in the following 30 days. The values are generally between 0.70 and 0.80, corresponding to accuracy described previously as "useful for some purposes" [44]. For the Central West region, AUCs are high (between 0.8 and 0.9) after a spike in incidence in late 2020, indicating that the lack of subsequent peaks was predictable. Overall, the AUC values are associated with incidence, suggesting that the method is more able to learn across municipalities when higher or lower rates are propagating across the country, i.e. that increases elsewhere helped predict peaks in each index municipality. From each ROC curve, values of sensitivity and specificity were chosen to maximize the sum of these two parameters. For sensitivity, averaging over time, the regions had similar values, between 71 and 73%. For specificity, the average values ranged from 63% for the South-East region to 72% in the Central-West.



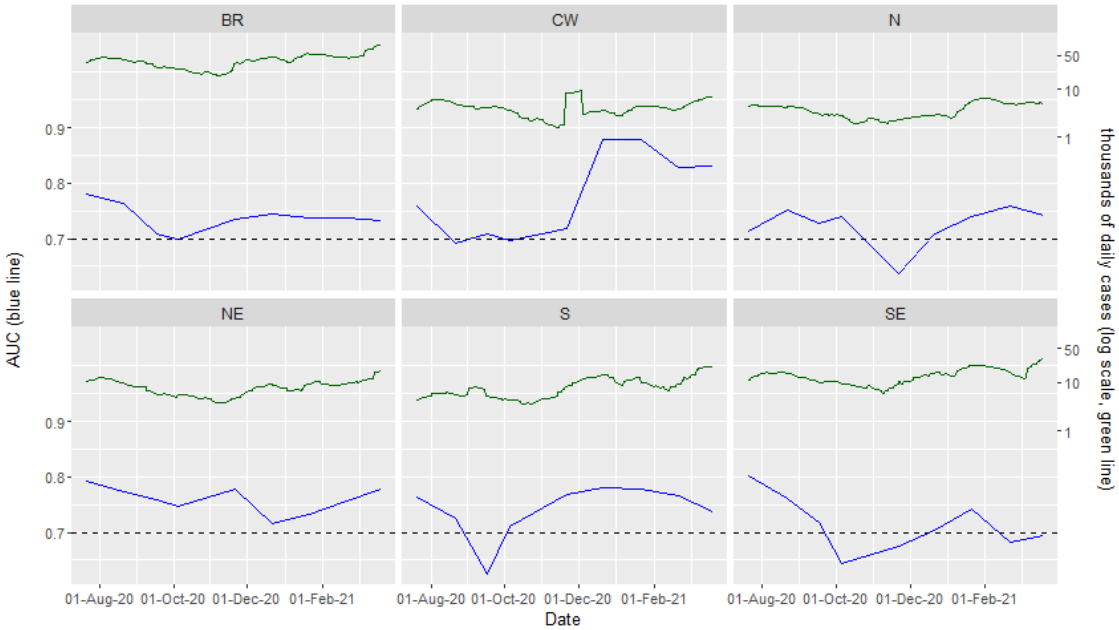


Figure 5. Area under the curve (AUC) for prediction of a new maximum incidence in the following 30 days (left vertical axis, blue line) against calendar time (horizontal axis). The daily number of incident cases (rolling fortnightly average) is shown in the green line, right axis. The dashed line shows the lower end of the range 0.70-0.90. There is one panel for each region (Central-West, North, North-East, South and South-East), with the "BR" panel showing all Brazil.

Discussion

We describe the development of an online application, CLIC Brazil, that allows comparison of the spread and impact of the COVID-19 epidemic in Brazil between local areas (municipalities). We show how basic analyses, largely available through the application can be used to identify pathways and determinants of spread. Further we identify and explain heterogeneities in burden and assess the relative timeliness of reactive interventions.

Our analyses show that despite an initial identification of SARS Cov-2 in the large South Eastern cities of São Paulo and Rio de Janeiro [45,46] the early focus of the epidemic quickly shifted to the Northern region before spreading to North Eastern coastal cities and then to the Southern and South Eastern regions of the country. In general, places with higher metrics of social development experienced earlier onset and faster growing epidemics, although the magnitude of the effect explained by these factors was low compared to that due to space and, to a lesser extent, time. Places that were more distant from major population centres experienced later onset and slower growth. The factors underlying these trends may be related to transport connectivity as well as inequities in access to testing and treatment [47]and higher prevalence of other health conditions in less developed places.

There were large differences in the mean *Rt* between local areas and over time that could not be explained by the covariates included in our analysis, These differences may be related to differences in patterns of social mixing or the imposition of, and the level of adherence to, non-pharmaceutical interventions (NPIs), and the extent to which the roll-out of public measure designed to curb the spread of the epidemic was devolved to a local level by national and State authorities [48].

It has been shown that in the early phases of the epidemic, municipalities with a higher degree of socioeconomic vulnerability (poor housing conditions and low levels of formal employment) had higher COVID-19 death rates but this association was lost over time [47] . This was thought to be due in part to the ability of these municipalities to introduce non-pharmaceutical interventions that were effective in mitigating the spread of the epidemic and also their ability to expand healthcare capacity. Our finding that municipalities with higher social demographic indices experienced an earlier arrival and more rapid propagation of the epidemic may be due to these places having a stronger healthcare infrastructure and so provide infected individuals with more effective treatment.

There was evidence, at least for the Central-West region, that the reproduction number was lower in places where the epidemic arrived later. This may be explained by an increased awareness of the need to adhere to social distancing and other NPIs that developed over time.

There was considerable variation between geographic regions and over time in the predictive accuracy of the algorithm developed to predict new incidence maxima for municipalities. This was driven in part by regional and national incidence levels.

Among the limitations of this analysis are the likely inconsistencies in reporting of cases and deaths over time and between places. Also, the covariates included as proxy measures of social development may not fully reflect the impact of wealth and employment type on the ability of individuals in different areas to adopt social distancing and lessen their risk of infection. Another limitation is that places on the border regions of a state may be distant from the most populous place in that state but closer to a large city in a neighbouring state. However, given the small effect sizes for this covariate the effect on the outcome would be small.

Since December 2020 cases have begun to increase again, initially focussed on cities such as Manaus in the Northern Amazonas [3,49,50] .There is emerging evidence this new variant of concern (P1/

B.1.1.28.1) has higher transmissibility and exhibits the ability to evade that the neutralising effect of antibodies to previous infection. The P1 variant has rapidly spread to become the dominant strain in Amazonas state and throughout the country [50]. Analyses of factors associated with early development of the epidemic may be useful in predicting the spread of this new variant. Continual tracking of this next phase of the epidemic using methods demonstrated in our analyses will be important to assess similarities and differences in the spatial spread of different COVID-19 epidemic waves. These analyses should include spatially disaggregated data on vaccine coverage [9] to assess the impact of the epidemic on vaccination on the development of the epidemic. Pairing these outputs with phylogenetic analyses [3,49] of SARS-CoV-2 virus samples would enable a more detailed picture of past and present subnational spread of the epidemic.

Conclusion

This study demonstrates that by monitoring, standardising and analysing the development of an epidemic at a local level, insights can be gained into spatial and temporal heterogeneities. Such insights are often impossible using raw case counts or when data are aggregated over larger areas. We show the utility of using age-standardised incidence as a comparable epidemiological metric for a variety of analyses and have developed an on-line application that allows a range of stakeholders to simply compare and contrast the evolution of the COVID-19 epidemic in different areas. This approach could prove useful for real-time local monitoring and analysis of a range of other emerging infectious disease outbreaks.

Acknowledgements

This project was supported by a Medical Research Council (MRC-UK) and the São Paulo Research Foundation (FAPESP) Newton partnership award (MR/S0195/1 and FAPESP 18/14389-0) for the CADDE project (http://caddecentre.org/). N.R.F. is supported by a Wellcome Trust and Royal Society Sir Henry Dale Fellowship: (204311/Z/16/Z). OJB was supported by a Wellcome Trust Sir Henry Wellcome Fellowship (206471/Z/17/Z) . CAPJ was supported by FAPESP (2019/21858-0), Fundação Faculdade de Medicina and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

The following authors were part of the Centre for Mathematical Modelling of Infectious Disease COVID-19 working group. Each contributed in processing, cleaning and interpretation of data, interpreted findings, contributed to the manuscript, and approved the work for publication: Naomi R Waterlow, C Julian Villabona-Arenas, James D Munday, Graham F. Medley, Rachel Lowe, Yang Liu, Amy Gimma, Kevin van Zandvoort, Joel Hellewell, Damien C Tully , Megan Auzenbergs, Gwenan M Knight, Adam J Kucharski, Rosanna C Barnard, William Waites, W John Edmunds, Nikos I Bosse , Akira Endo, Emilie Finch, Timothy W Russell, Yung-Wai Desmond Chan, Matthew Quaife, Rosalind M Eggo, Kiesha Prem, Rachael Pung, Thibaut Jombart, Billy J Quilty, Samuel Clifford, Mihaly Koltai, Hamish P Gibbs, Sam Abbott, Christopher I Jarvis, Yalda Jafari, Petra Klepac, Fabienne Krauer, Fiona Yueqian Sun, Sebastian Funk, Frank G Sandmann, Emily S Nightingale, Jiayao Lei, Sophie R Meakin, Alicia Rosello, Carl A B Pearson, David Hodgson, Ciara V McCarthy, Anna M Foss, Katherine E. Atkins,

References

1. WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus Disease (COVID-19) Dashboard. Accessed March 24, 2021. https://covid19.who.int/
2. Castro MC, Kim S, Barberia L, et al. Spatiotemporal pattern of COVID-19 spread in Brazil. *Science*. Published online April 14, 2021:eabh1558. doi:10.1126/science.abh1558
3. Faria NR, Mellan TA, Whittaker C, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. Published online April 14, 2021:eabh2644. doi:10.1126/science.abh2644
4. Dowd JB, Andriano L, Brazel DM, et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*. 2020;117(18):9696-9698. doi:10.1073/pnas.2004911117
5. Dudel C, Riffe T, Acosta E, van Raalte A, Strozza C, Myrskylä M. Monitoring trends and differences in COVID-19 case-fatality rates using decomposition methods: Contributions of age structure and age-specific fatality. Masquelier B, ed. *PLOS ONE*. 2020;15(9):e0238904. doi:10.1371/journal.pone.0238904
6. Ferrante L, Steinmetz WA, Almeida ACL, et al. Brazil's policies condemn Amazonia to a second wave of COVID-19. *Nature Medicine*. 2020;26(9):1315. doi:10.1038/s41591-020-1026-x
7. Hallal PC, Hartwig FP, Horta BL, et al. SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys. *The Lancet Global Health*. 2020;8(11):e1390-e1398. doi:10.1016/S2214-109X(20)30387-9
8. de Souza WM, Buss LF, Candido D da S, et al. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nature Human Behaviour*. 2020;4(8):856-865. doi:10.1038/s41562-020-0928-4
9. Painel da vacinação COVID-19. Accessed April 29, 2021. https://apps.kauebraga.dev/shiny/painel_vacinacao_covid/
10. Jombart T, Ghozzi S, Schumacher D, et al. Real-time monitoring of COVID-19 dynamics using automated trend fitting and anomaly detection. *medRxiv*. Published online September 3, 2020:2020.09.02.20186502. doi:10.1101/2020.09.02.20186502
11. COVID-19 Map - Johns Hopkins Coronavirus Resource Center. Accessed March 9, 2021. https://coronavirus.jhu.edu/map.html
12. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*. 2020;20(5):533-534. doi:10.1016/S1473-3099(20)30120-1
13. COVID-19 ITALIA - Desktop. Accessed March 9, 2021. https://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaac82fe38d4138b1
14. EpiForecasts - We forecast infectious disease outbreaks. Accessed March 9, 2021. https://epiforecasts.io/
15. Abbott S, Hellewell J, Thompson RN, et al. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Research*. 2020;5:112. doi:10.12688/wellcomeopenres.16006.1
16. Viz - COVID-19 Forecast Hub | COVID-19. Accessed March 9, 2021. https://viz.covid19forecasthub.org/
17. Ray EL, Wattanachit N, Niemi J, et al. Ensemble forecasts of Coronavirus Disease 2019 (COVID-19) in the U.S. *medRxiv*. Published online August 22, 2020:2020.08.19.20177493. doi:10.1101/2020.08.19.20177493
18. Cramer EY, Ray EL, Lopez VK, et al. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US. *medRxiv*. 2021;46:2021.02.03.21250974. doi:10.1101/2021.02.03.21250974
19. Short-term forecasts of COVID-19 deaths in multiple countries. Accessed April 29, 2021. https://mrc-ide.github.io/covid19-short-term-forecasts/

20. Gostic KM, McGough L, Baskerville E, et al. Practical considerations for measuring the effective reproductive number, Rt. *medRxiv*. Published online June 23, 2020:2020.06.18.20134858. doi:10.1101/2020.06.18.20134858
21. Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology*. 2013;178(9):1505-1512. doi:10.1093/aje/kwt133
22. Cauchemez S, Boëlle PY, Thomas G, Valleron AJ. Estimating in real time the efficacy of measures to control emerging communicable diseases. *American Journal of Epidemiology*. 2006;164(6):591-597. doi:10.1093/aje/kwj274
23. Covid-19: National and Subnational estimates for Brazil. Accessed March 9, 2021. https://epiforecasts.io/covid/posts/national/brazil/
24. Brazil.IO. Brazil.IO: COVID-19 Source: Health Departments of the Federative Units, data process.
25. Instituto Brasileiro de Geografia e Estatisitica (IBGE). IBGE - Census 2010. https://sidra.ibge.gov.br/tabela/3107
26. WorldPop. Accessed April 29, 2021. https://www.worldpop.org/
27. Sorichetta A, Hornby GM, Stevens FR, Gaughan AE, Linard C, Tatem AJ. High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Scientific Data*. 2015;2(1):1-12. doi:10.1038/sdata.2015.45
28. Kraemer MUG, Sadilek A, Zhang Q, et al. Mapping global variation in human mobility. *Nature Human Behaviour*. 2020;4(8):800-810. doi:10.1038/s41562-020-0875-0
29. Pfeffer DA, Lucas TCD, May D, et al. MalariaAtlas: An R interface to global malariometric data hosted by the Malaria Atlas Project. *Malaria Journal*. 2018;17(1):352. doi:10.1186/s12936-018-2500-5
30. Wang H, Naghavi M, Allen C, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*. 2016;388(10053):1459-1544. doi:10.1016/S0140-6736(16)31012-1
31. Pereira, R.H.M; Gonçalves CN; et. al. geobr: Loads Shapefiles of Official Spatial Data Sets of Brazil. GitHub repository. Published 2019. https://github.com/ipeaGIT/geobr.
32. Chang W, Cheng J, Allaire JJ, et al. shiny: Web Application Framework for R. Published online 2021.
33. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2016.
34. Cheng J, Karambelkar B, Xie Y. leaflet: Create Interactive Web Maps with the JavaScript "Leaflet" Library. Published online 2021.
35. Parag KV. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *medRxiv*. Published online September 18, 2020:2020.09.14.20194589. doi:10.1101/2020.09.14.20194589
36. Ferguson NM. Report 9 - Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand | Faculty of Medicine | Imperial College London. Accessed April 22, 2021. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/
37. Lima FET, Albuquerque NLS de, Florencio S de SG, et al. Intervalo de tempo decorrido entre o início dos sintomas e a realização do exame para COVID-19 nas capitais brasileiras, agosto de 2020*. *Epidemiologia e Serviços de Saúde*. 2021;30(1):e2020788. doi:10.1590/s1679-4974202100010002
38. Parag K V., Donnelly CA. Using information theory to optimise epidemic models for real-time prediction and estimation. Ferrari M (Matt), ed. *PLOS Computational Biology*. 2020;16(7):e1007990. doi:10.1371/journal.pcbi.1007990
39. Tobin J. Estimation of Relationships for Limited Dependent Variables. *Econometrica*. 1958;26(1):24. doi:10.2307/1907382

40. VGAM. Accessed April 29, 2021. https://www.stat.auckland.ac.nz/~yee/VGAM/
41. GitHub - therneau/survival: Survival package for R. Accessed April 29, 2021. https://github.com/therneau/survival
42. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Medicine*. 2015;12(10):e1001885. doi:10.1371/journal.pmed.1001885
43. Candido DS, Claro IM, de Jesus JG, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science (New York, NY)*. 2020;369(6508):1255-1260. doi:10.1126/science.abd2161
44. Swets JA. Measuring the accuracy of diagnostic systems. *Science Science*. 1988;240(4857):1285-1293. doi:10.1126/science.3287615
45. de Jesus JG, Sacchi C, Candido D da S, et al. Importation and early local transmission of covid-19 in brazil, 2020. *Revista do Instituto de Medicina Tropical de Sao Paulo*. 2020;62. doi:10.1590/S1678-9946202062030
46. Candido DDS, Watts A, Abade L, et al. Routes for COVID-19 importation in Brazil. *Journal of Travel Medicine*. 2020;27(3):1-3. doi:10.1093/jtm/taaa042
47. Rocha R, Atun R, Massuda A, et al. Effect of socioeconomic inequalities and vulnerabilities on health-system preparedness and response to COVID-19 in Brazil: a comprehensive analysis. *The Lancet Global Health*. 2021;9(6):e782-e792. doi:10.1016/s2214-109x(21)00081-4
48. de Souza Santos AA, Candido D da S, de Souza WM, et al. Dataset on SARS-CoV-2 non-pharmaceutical interventions in Brazilian municipalities. *Scientific Data*. 2021;8(1):73. doi:10.1038/s41597-021-00859-1
49. Sabino EC, Buss LF, Carvalho MPS, et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *The Lancet*. 2021;397(10273):452-455. doi:10.1016/S0140-6736(21)00183-5
50. Naveca F, Souza V, Corado A, et al. COVID-19 epidemic in the Brazilian state of Amazonas was driven by long-term persistence of endemic SARS-CoV-2 lineages and the recent emergence of the new Variant of Concern P.1. Published online February 25, 2021. doi:10.21203/rs.3.rs-275494/v1