How mathematical approaches could help decision-making to epidemic control? The successful experience against COVID-19 in Cuba.

Lizet Sanchez[1,2], Patricia Lorenzo-Luaces[1,2], Carlos Sebrango[1,3], Antonio Torres-Reyes[1,4], Claudia Fonte[1,5], Marelis Crespo [1,5], Jose E. Valdes [1,5], Waldemar Baldoquin[16,], Paula Fonte[1,5], Yudivian Almeida[1,5], Suilán Estévez[1,5], Alejandro Piad[1,5], Wilfredo Morales, [1,5] Nancy Pérez[1,4] , Rafael Araujo[1,7], Maria del Carmen Pria[1,8], Ariadna Corral[1,8], Vivian Noriega[1,8], Maria Vidal[1,8], Agustin Lage[2], Raul Guinovart[1,5], Pedro Mas[1,6]
[1] Cuban National Group for Modelling of the COVID-19 pandemics
[2] Center of Molecular Immunology
[3]University of Sancti Spiritus "Jose Marti Perez"
[4]Faculty of Geography, University of Havana
[5] Faculty of Mathematics and Computer Sciences, University of Havana
[6] Tropical Medicine Institute "Pedro Kouri"
[7]Demographic Studies Center, CEDEM, University of Havana
[8] National School of Public Health

Corresponding author: Lizet Sánchez, PhD. Cuban National Group for Modelling of the COVID-19 pandemics. Clinical Research Division, Center of Molecular Immunology, Calle 216 esq 15. Atabey, Havana 11600, Cuba; email: lsanchez@cim.sld.cu, tel: 5372143145.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

ABSTRACT

Background. There is a gap for the effective use of mathematical models for real-time decision-making. We aimed to illustrate with the Cuban experience to control the COVID-19, how mathematical models can be put in place to answer key decision-makers´ questions.

Methods. A science-policy partnership was created to mutually define questions, communicate results and facilitate the translation of modeling advice into actions. For forecasting and planning at national level mechanistic models and machine learning based on the epidemic patterns in other countries were used. Statistical models to explain the variability of transmission was used to stratify control actions. The effect of interventions was assessed using branching process models, time varying reproduction number (R_t) and social mixing patterns by location, and by age group.

Findings. The mathematical approach implemented contribute to successful control of the COVID-19 in Cuba. The urbanization, living conditions and the economic index explain the 73% of the variability of the transmission at provincial level. Increased risk of transmission were identified in 33 municipalities mostly in densely populated urban areas with high aging index. Control intervention reduced the transmission from R_0=2.84 (95% CI: 1.52 - 4.76) to R_t=0.6 (95% CI:0.2-2.38). The highest transmission was detected among adolescents and from people older than 60 years.

Conclusions. Understanding the key questions for decision-making at all times, translating problems into a mathematical language, integrating different approaches to their solution and being able to present the results in an easy-to-understand way is vital to have a timely impact on controlling the epidemic.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

INTRODUCTION

Since the beginning of the Covid-19 epidemic, it showed to have mechanisms of spread that favoured its rapid expansion. On March 11, 2020, the World Health Organization (WHO) declared it as a pandemic.[1] In the absence of a COVID-19 vaccine, the control was aimed at reducing contact rates in the population and thereby reducing transmission of the virus. WHO launched a strategy to promote strengthening readiness to rapidly identify, diagnose and treat cases; identification and tracking of contacts; infection prevention and control in healthcare settings; implementation of health measures for travellers; and awareness raising in the population though risk communication and community engagement.[2]

The Cuban health system has been hailed as among the successfully stories in infectious diseases control.[3] Specifically in the COVID-19 control, Cuba's strategy was to apply everything learned over the years. It differs from the strategies followed by other countries in four fundamental elements: the participation of the primary health care medical teams in the follow-up of people suspected of being infected in the community, hospitalization of all confirmed cases, isolation of all contacts, and early treatment to prevent progression to severe forms of the diseases. Additionally, in this epidemic, multidisciplinary team, including mathematicians, epidemiologists, computer scientists and geographer, was created to work together with the health authorities and with the government at all levels to decide the best strategy at all time. Mathematical modelling was, as never before, of crucial importance in all decisions related to the epidemic control.

Models yield quantitative projections, and qualitative analysis, in the short- and long-term allowed allocating resources, planning or evaluating interventions. Numerous mathematical models are reported to forecast the future of COVID-19 disease,[4,5] to estimate important epidemiological parameters,[6] to evaluate the risks,[7,8] and to assess control interventions.[9-11] However, although there are some examples of real impact in the government policies, such as the case of United Kindong,[12] most of them, express academic visions with little role in health policy.[13] There is a gap between the model development and their effective use for decision-making in real time. This is given mainly by the use of technical language, the contradictory messages obtained by different models or the visualization of the results in non-illustrative terms for practical purposes. On the other hand, although it is well know that a model cannot cover all aspects of a problem, a large part of the studies try to answer decision-making questions through a single modelling approach.

In this study, we present the Cuban experience to illustrate how mathematical models can be put in place to answer key decision-makers' questions as the epidemic progresses.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

METHODS

Context

Cuba has a total population of 11.2 million of people distributed between 15 provinces, divided in turn into 168 municipalities. A National Action Plan for Epidemics was activated as early as January.[14] Surveillance at all ports, airports and marinas was increased. In February community polyclinics and hospitals were reorganized for attending infectious disease patients. Molecular biology laboratories in the western, central, and eastern region were activated for COVID-19 testing, supplementing lab at Havana's Pedro Kourí Tropical Medicine Institute (national reference centre).

On March 9, first prime-time television "Roundtable" broadcast on COVID-19 announced series of measures; frequent hand washing, and physical distancing was recommended. Additional 1322 beds in 11 hospitals and 824 beds in 10 isolation centres around the country, were set aside for COVID-19 cases, contacts and suspected cases. On March 13, only 2 days after the confirmation of the first cases, massive events and artistic shows were suspended throughout the country. Cuba joined the international call for social distancing to eliminate chains of transmission. The international travel restrictions were adopted by the country in March 24, but the effective border shutdowns was effective on April 2. Active surveillance accompanied by prevention and health promotions activities was implemented. On March 30 occurred the cessation of the face-to-face activities in schools. On April 11, the interprovincial and then urban transport was interrupted. At the same time, Cuban biopharmaceutical industry guarantees the production of medications for preventive, immunomodulatory and therapeutic treatment of Covid-19 included in the clinical protocol established by the Health authorities. Clinical trials evaluating new therapies were accelerated launched.

A website that provides real-time information on the outbreak (https://covid19cubadata.github.io) facilitates real-time monitoring and modelling of the epidemic.

Data source

The daily reported confirmed cases data of COVID-2019 from March 11, 2020 to May 31, 2020 by the 15 provinces and 168 municipalities were used. For each province and each municipality, the numbers of cumulative cases as well as the population counts in 5-year age groups were recorded. Demographic data on the population sizes, population density, urbanization, masculinity index and aging index were provided by the National Office of Statistics. We used the mean estimates of the population for 2019. All data was related to geographic coordinates by a unique municipality code. The latitude and longitude from the

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

centre of each municipality was extracted from the GEOCUBA cartographic data on a 1:25000 scale.

The follow-up of the epidemiological chain of cases was obtained from the epidemiological surveys carried out by health professionals at local level. For most of the patients it was possible to discover the source of infection. It was identified if the transmission occurred in the province or was imported from other places. The official data from the statistical departments of the Ministry of Health were used for this purpose.

The demographic characteristics of the countries used for the comparison with Cuba in the study were obtained from the website: http://www.woldometer.info. The confirmed reported cases for the countries were obtained from http://pomber.github.io/covid19.

Modelling approaches

Several approaches could be found in the literature. Particularly in Cuba, artificial intelligence, mechanistic and statistical modeling strategies based on the previous experience were used. A science-policy partnership was created to mutually define policy questions, communicate results and facilitate the translation of modeling advice into actions. The question relevant for policymakers, the mathematical model approach used and it implications by epidemic stages are shown in the box. In this study we will focused on the first four directions. The mathematical models used for the study of the epidemic severity and mortality, together with the evaluation of pharmacological interventions, and the post-epidemic scenarios will be part of another study.

Directions	Questions	Mathematical models used in Cuba	Implications
1. Forecasting and planning at national level	What is the maximum number of cases expected in the epidemic? When will the epidemic peak occur? How many patients are expected to be hospitalized? How many will require therapy services?	Mechanistic models (SIR and SEIR models)	Resource planning and allocation at all levels, organization of health services and isolation sites
2. Identification of possible scenarios for Cuba based on the epidemic patterns in other countries	Which countries have an epidemic behavior similar to Cuba? What number of cases will be in the next days based on similar country patterns? In which epidemic scenario, of the predicted ones, will the country move?	Similarity functions. Machine learning models.	Short term planning and resource allocation. Scenarios identification for decision making.
3. Transition from early to establish transmission	What is the transmissibility in the country context? Are there differences between regions? What does it depend on? Which areas and population groups are most at risk?	Basic reproduction number, turning point and final size estimation by provinces, phenomenological and regression models. Spatial analysis.	Stratification of control measures. Intensification of actions in vulnerable areas and groups

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

4. Effectiveness of case detecting strategy and control interventions	What would have happened if the measures had not been implemented? What is the impact of non pharmacological interventions? How efficient are the methods of detecting infected people?	branching process simulation models Time varying effective reproduction number estimation, transmission indexes estimation	Changes in the testing policy. Redirection of communication strategies. Modifications of non-pharmacological interventions
5. Severity and mortality (ongoing)	Who is at increased risk of becoming serious? Who are most at risk of dying? Is therapy being effective in reducing risks?	Classification methods and survival models.	Changes in clinical protocols to focus on individuals at risks
6. Evaluation of pharmacological interventions (ongoing)	What is the impact of pharmacological interventions? How does mortality change with the improvement of clinical protocols and medical care? Which therapy is the best for each patient?	Generative Bayesian models and Causal inferences models.	Changes in clinical protocols to reduce severity cases and mortality
7. Plan de de-escalation and the post-epidemic period (ongoing)	How and in what order should interventions be de-escalated at national or local levels? How and when to demobilize the capacities and resources created or retained to face crisis? How to identify the conditions that cause new outbreaks?	Mechanistic models (SIR and SEIR models) and Bayesian semi-mechanistic models.	Successive de-escalation of interventions at the regional and national levels. Redirection of communication strategies. Create indicators to identify alerts of new outbreaks.

Forecasting and planning at national level

For the long-term forecast a susceptible- infectious-recover (SIR) model was implemented in order to generate the curves of active cases at the national level. Here, we consider a SIR model with the demography.[15] The selection is based on the fact that this model includes a parameter which helps measure the effectiveness of the state intervention. This model allows estimating the existence of endemic points and the possibility of re-growth.

Identification of possible scenarios for Cuba based on the epidemic patterns in other countries

In order to identify countries with an epidemic behavior similar to Cuba, both their demographic characteristics and the information of COVID-19 transmission were taken into account. Initially, a similarity function was defined for each country according to the population size, population density, fertility, average age and urban population. With this function, a group of countries demographically similar to Cuba was obtained. Subsequently, a similarity function with displacement was constructed to select countries with demographics and curves of confirmed cases similar to Cuba in at least 7 days. Then, with the data of confirmed cases from these countries, several learning models were trained. From them, the best result, based on the prediction error for the Cuban data, was obtained with Lasso Regression model. It was selected to predict in the short term (from 7 to 14 days) the behavior of confirmed cases for Cuba at the beginning of the epidemic. Additionally, it was used to predict in which scenery, from the scenarios defined using the SIR models, Cuba would probably move.

Transition from early to establish transmission

Estimating the strength of epidemics is of great concern for policymakers when the transmission starts. The basic reproduction number (R_0) is the most common measure of this strength. It represents the average number of new cases that occur after the introduction of one infected individual in an entirely susceptible population. This parameter determines whether

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

the transmission will terminate naturally ($R_0 <1$) or evolve into an epidemic ($R_0> 1$). Most of the models used in epidemiology offer an estimate of this value. We used the initial exponential growth rate of the epidemic to infer the R_0.[16] For the generation time, parameter used for R_0 estimation, we assumed it follows a Gamma distribution with a mean of 6.87 days and a standard deviation of 5.42 days, as was estimated from a Cuban study.

Once the transmission was established, we study the progress of the epidemic at three spatial levels: national, provincial, and municipality levels. At the national level, short-term (7 days) forecasts approach was carried out in real time using phenomenological models. Richards model, three parameter logistic model (3P logistic) and the Gompertz models were adjusted to the cumulative daily cases. Thus, three different sceneries were predicted. Due to the demographic and connectivity differences between the capital and the rest of the country, separate analyses were carried out for Cuba, Havana and the rest of the country's provinces. In previous studies was proved that reasonably accurate prediction for all outbreak could be obtained as long as the data include the inflection point and a time interval shortly after.[17]

The turning point, the final size of the outbreak and its confidence interval, were directly derived from the parameters of the models. The turning point indicates the moment in which the epidemic reaches the maximum number of cases per day, while the final size of the epidemics is the total number of infested people in a population.

At the provincial level, we analyzed the strength of the transmission through R_0. At local level, usually the reported cumulative cases are modeled separately for each province in a particular country. However, when the interest lays on the variability among and within provinces, a nonlinear mixed effects model is recommended to accounts for heterogeneity between areas.[18]

To explain the variability of the transmissibility at this level, a multiple regression analysis was performed. Demographic factors, living conditions, and the comprehensive development index along with their biological, ecological, economic, and behavioral dimensions were included in the model. The life condition index and the integral development index were obtained from previous studies.[19]

At municipality level, a spatial analysiswas carried out to identify vulnerable areas where actions must be intensified.[20] Smoothed estimates of relative risk (RR) were obtained from Conditional Autoregressive (CAR) Bayes approach. The posterior probability of each municipality'sRR exceeding unity was calculatedand mapped providinga measure of the strength of (statistical) evidence. High probabilities can beinterpreted as providing clear evidence of an excess risk,whereas low probabilities are related to regions with no risk.For a better representationof the province Havana, the capital of Cubawith more than 2 million inhabitants and including 15municipalities, a zoom was made. We used Moran's I statistic to measure the spatial autocorrelation.

Effectiveness of case detecting strategy and control interventions

At this point, we use a combination of several mathematical tools. To evaluate the effect of the interventions in the first month of the epidemic, we simulate a situation in which no measures are taken to contain the transmission. To illustrate this we used the data from Havana. A model based on a branching process was used to simulate the transmission in the first month.[21] The estimated $R_0=2.14$ for Havana and the gamma distribution (shape=1.61, scale=4.27) estimated for the serial interval were used. For the initial value of the infected people we considered the 15 COVID-19 positive travelers confirmed in Havana until March 24. 1000 process paths were generated. The Python language was used for programming.

In Cuba, there are not previous studies exploring social mixing patterns. However, using the information on the epidemiological chain of transmission, two transmission matrixes could be built. The first was used to evaluate the effectiveness of the interruption of interprovincial transportation and the effectiveness in controlling imported cases. The second to determine

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

transmission pattern by age groups. The elements m_{ij} in the transmission matrix represents the contacts of province j made by an infectious individual from the province i, divided by the total number of infected individuals in the province i. In the same way, the transmission matrix was constructed for each 10 years ages groups.

Next, we evaluate the measures that were taken through the changes in the effective reproductive number (Rt). It is an extension of the concept of R_0, that is, it is the effective number of secondary cases generated by one primary case in the real population in the time t. The monitoring of the reproduction number over time provides feedback on the effectiveness of interventions and on the need to intensify control efforts. For this purpose, we used the methodology proposed by Cori et al.[22]

RESULT

Forecasting and planning at national level

The figure 1 shows the predictions obtained in two different times. The first is the analysis carried out at the beginning of the epidemic, the dots represent the real cases occurred. The second corresponds to the prediction made in late May. The possibility of a second small wave was predicted and alerted to the health authorities.

<Insert Figure 1>

Identification of possible scenarios for Cuba based on the epidemic patterns in other countries

The figure 2 shows the forecast of confirmed cases to Cuba taken into account the epidemic behaviour in countries similar to Cuba, both demographically and epidemiologically.

<Insert Figure 2>

Transition from early to establish transmission

Before the border shutdowns, 156 infected travellers coming from outside of the country. The majority where detected in Havana (102 patients; 65.4%) and in Villa Clara (17 patient, 10.8%). However, the spread of the diseases occurred in all provinces. Figure 3 present the cumulative, daily new cases and forecast model estimation for Havana, outside of Havana and for Cuba. Cumulative reported COVID-19 cases by province are presented in supplementary information. At the moment of this report, only a significant increase in the number of cases is expected in the next 7 days for Havana and Matanzas. The prediction for all provinces could be found in the supplementary information.

<Insert Figure 3>

The R_0, the turning point and the final size estimations are presented in Table 1. The R_0 was 2.71 (1.90, 3.77) for Cuba. However, a huge variation was observed between provinces, the estimates ranged from 0.81 to 2.37. Multiple linear regression analysis was used to explain the variation of R_0 at provincial level. The best fit was obtained with the model containing the urbanization, living conditions index and the economic dimension obtained from the

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

Comprehensive Development Index. This model was able to account for 73% of the variance in R_0, F(3,11)=10.0 , p=0.002, R^2=0.73.

Table 1. Basic reproduction numbers, turning point, final size, and its 95% confidence intervals estimates. Cuba, March 11 to May 21, 2020.

Provinces	R_0 (95%CI)	Turningpoint**	Final size**
Pinar del Río	1.62(0.49, 3.56)	42.99 (39.04,44.66)	50 (48,53)
Artemisa	1.74 (0.19, 5.37)	25.91 (24.90,26.91)	30 (29,31)
Isla de la Juventud	1.64 (1.02 , 2.46)	40.07 (38.33,41.81)	42 (41,42)
La Habana	2.14 (1.36 , 3.17)	41.44 (40.65,42.23)	1092 (1022,1161)
Mayabeque	1.84 (0.53, 4.15)	40.43 (39.36,41.50)	48 (46, 50)
Matanzas	2.37 (1.16 , 4.82)	39.50 (37.22,41.79)	166 (140,192)
Cienfuegos	0.81 (0.04, 2.58)	48.27(44.65,50.51)	24 (23,26)
Villa Clara	2.35 (1.49, 3.58)	36.32 (35.23,37.41)	200 (195,204)
Sancti Spiritus	1.58 (0.74, 2.86)	33.51 (31.52,35.50)	63 (61,65)
Ciego de Ávila	1.74 (1.31, 2.26)	25.24 (23.75,26.73)	87 (84,89)
Camagüey	1.84 (0.97, 3.16)	24.58(21.67, 27.48)	41 (39, 43)
Las Tunas	0.98 (0.01, 4.99)	39.02 (37.34,40.71)	15 (15,16)
Holguín	2.04 (1.13, 3.43)	31.03 (29.73,32.34)	91 (88, 93)
Granma	0.87 (0.11, 2.30)	19.88(16.22,21.29)	10 (10,11)
Santiago de Cuba	1.97 (1.10, 3.22)	23.88(21.96, 25.80)	50 (49, 52)
Guantánamo	1.84 (0.08, 7.24]	33.10 (31.42, 34.77)	17 (16, 17)

The posterior CAR-smoothed RR estimates for COVID-19 in the 168 municipalities of Cuba range from 0.03 to 9.18. IRs and CAR-smoothed RRs for localities having risk estimations of above 1.2, or a posterior probability (based on the smoothed model) of at least 0.90 that their estimated RR would be higher than 1, were selected (see Table in the supplementary material). These criteria were used to select a reasonable number of municipalities (from a total of 168) with some evidence of an excess risk. Estimates of the RR show less variation than the observed IR as can be expected after Bayesian smoothing. Although in crude IRs no clear patterns are elucidated in the distribution of risk, the smoothed maps provide well-delimited homogeneous areas in which different geographical risk configurations could be distinguished.

A region with a moderately higher risk (CAR smoothed RR varying from 1.45 to 7.84) is observed in almost all municipalities of Havana province. Three other clusters with high risk are observed in the province of Matanzas, Villa Clara and Ciego de Avila (Table 2). The global Moran's I statistics was 0.36 (p-value = 2.686e-14). This confirms findings described with the CAR model.

Since the focus of our study is to alert the authorities about the risk to intensify actions, more important result the analysis of the posterior probability of having an increased risk of transmission. Figure 4 present the distribution of the posterior probabilities of relative risks exceeding unity for COVID-19 and the distribution of demographics characteristics in the country. A low, but significant correlation, was found between the IRs and demographic characteristics (urbanization: 0.34; p<0.001 and population density: 0.37; p<0.001). Given that the severity of the COVID-19 disease is associated with aging, the distribution of this index was shown in the figure 5. Note that the central region has municipalities with high risk of transmission and at the same time the highest rates of aging.

<Insert Figure 4>

Effectiveness of case detecting strategy and control interventions

Figure 5 illustrate the results of the application of different mathematical tools to evaluate the effectiveness of the case detecting strategy and to assess control interventions. The average curve together with the 90% of confidence region resulting from the simulation model and the real number of cumulative confirmed cases are showed in Figure 5a. The simulation corresponds to a situation in which no measures are taken to contain the transmission, but assume the border shutdowns. Note that, according to the model, between 1 000 and 2 500

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

people should have been infected in the first month of transmission. However, in April 11, only 321 confirmed cases were reported. The confirmed cases are below the confidence region almost from the beginning of the epidemic. That could be a consequence of two factors, probably a combination of them. The first could be the presence of infected cases not detected by the health system. In the panel b of the figure we show the distribution of the symptomatic and asymptomatic confirmed cases by age group (Figure 5b). The bi-modality in the age distribution of the asymptomatic cases, not observed in symptomatic reported cases, alerts the possibility to have not detected infected people, mainly in the ages between 30 and 50 years old. Second, it can also be the result of the effectiveness of the implemented interventions, which is in accordance with the non-explosion of the lethality values (3.7%).

In Figure 5c we can see the transmission matrix by province (on the left) and by age (on the right). At the time of the study, the source of infection had been identified in 89.7% of confirmed cases. The matrix resulted from the contact analysis at provincial level has an approximately diagonal structure, indicating that almost all transmission occurred inside the provinces. However, Havana received the highest amount of infected people from outside. The provinces in the eastern region, together with Pinar del Rio and Artemisa, in the west region of the country, were more effective to control the imported cases. The greatest dispersion of transmission occurred in the provinces located from Havana to Sancti Spiritus (marked in the box). The social mixing pattern analysis by age group showed that the highest transmission occurred among adolescents and from people older than 60 years to people between 20 to 60 years old, mainly families and caregivers.

Another important indicator in evaluating the effectiveness of control measures was the R_t. Estimates of R_t maintain values above 1 in the first month of the epidemic. From April 2, there is a gradual decrease in the values. This coincides with the effective border shutdowns (Figure 6). From April 18, one week after the suspension of all urban and inter-municipal transport, the R_t reaches values below one. At the end of May, consistently with the prediction of the SIR model, an increase is observed in R_t, indicating the possibility of a second wave of transmission, that was alerted to the authorities.

<Insert Figure 6>

DISCUSSION

The integration of different approaches to mathematical modelling allowed answering key questions for decision-making in the course of the COVID-19 epidemic in Cuba. Analyzes were carried out at the national, provincial and municipality levels in correspondence with the questions that arise at these levels. Having a mathematical modelling strategy complementing real-time prediction, risk assessment, the analysis of transmission determinants, and the identification of vulnerable areas and groups, allowed targeting of actions and better planning of resources. Additionally, having indicators for monitoring the effectiveness of actions was important to correct it on time. Translating the results of the models into a clear visual language for decision-makers was important in order to make a timely impact in improving the epidemic control strategies.

Throughout the epidemic, different types of models were used: dynamical models, models based on artificial intelligence, statistical models, and branching process models. The first were used to plan different scenarios and to assess possible responses to the epidemic, even before it started. Models based on artificial intelligence use data from other countries with similar characteristics to Cuba. This allowed a prediction that takes into account the peculiarities of the virus, common to different countries. Once the epidemic spread, at a more advanced stage, the opportunity appeared to use statistical models that are based on directly from the data. Each of these models also has weak points; -the possibility of including a measure in the model, the

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

impossibility of capturing the peculiarities of a given country and the bias produced by the data and detection policies, to mention one in each case. For that reason, an integrated use of each one of them is recommended. Moreover, to ccombine mathematical models, exploring the extent and real time prediction of the outbreak, with spatial risk assessment at small area levels could give a more complete picture to analysis the evidence and to inform strategic efforts at the local level to improve health outcomes by managing COVID-19 infections in high risk areas.

We acknowledge that models and simulations are only part of the input influencing decisions. The experience of confronting COVID-19 confirms the important contribution offered by close and interactive collaboration between scientists and policymakers. Specifically in Cuba, one a week there was a meeting between the modelling team and the highest government authorities of the country.[23] As results the transmission has been kept under control in the majority of the national territory in two months. The actions allowed moving from the worst case to the favourable scenario. Changes in case-finding strategies and further intensification of control actions were necessary to reduce the transmission in the municipalities identified with excess risk. The effectiveness of global and local measures for the reduction of mobility and social contact was shown. However, the model result revealed that we had, and still have, a way to go.

Case finding in Cuba was based on the screening of patients with symptoms, investigation of contacts and screening of groups at risk such as health care workers and elderly people homes. The results of this study evidenced that the detection strategy could be improved. A massive screening to detect asymptomatic infection had been implemented in the high risk areas identified. Also, a national longitudinal survey was planned to estimate the seroprevalence and to detect active infection.

In our study areas, the basic reproduction numbers were comparable with the obtained in other countries.[24,8] However, the turning point was attained more quickly, which could reflect the impact of the implemented control measures. The limited spread of the COVID-19, to the Havana and Central Provinces, evidence the effect of the opportune interruption of interprovincial and then urban transport. A plan to de-escalate the measures can be implemented taking into account the risk analysis and the proven preparation of the health system capacities in the territory for the control of the epidemic. Educational outreach to promote risk fighting and good behaviours in the population should be also differentiated. According to our results, we need to intensify communication strategies specifically targeted on adolescents, families and individuals in contact with the people older than 60 years to reduce the physical contact and the transmission in this groups.

The clustering of adjacent counties with high and low incidence rates suggests that important factors influencing COVID-19 rates may exist at the municipality level. Municipalities with the highest and lowest rates might be particularly influenced by demographic, economic and living conditions varying characteristics. In urban areas with increased concentration of population, more frequent and complex interaction occurred. In addition to the probability of excess of risk, social and demographic factors should be taking into account in risk assessments. Urbanization and population density may be elements that determine the extent of transmission. In the analysis carried out at provincial level these determinants seem to explain the observed geographical variations. The population density, age distribution, and the urban index are factors with very similar figures among various countries and though they play a role in a pandemic cannot be considered specific for Cuba. In Europe, an integrative overview of factors influencing the sharp increases of the outbreak also emphasizes the role of demographic factors and population health index.[7]

Future research may focus on a zoomed local level, a look up into municipalities may elucidate, geographic characteristic related to risk, hidden into municipality scale. In addition, at individual level, mathematical modelling of severity should be integrated to the analysis

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

performed here. Mathematical model could be used to identify and validate prognostic biomarkers of the clinical evolution of patients. Also, mathematicians could be involved on the evaluation of the effectiveness of treatment protocols and on the impact assessment of the preventive and supportive therapies for older adults. Future researches should delve into these aspects, but also to develop mathematical models to guide the successive de-escalation of interventions at the regional and national levels.

In conclusion, The real-time predictions and risk assessment carried out in the midst of the COVID-19 epidemic course in Cuba helped guide actions at the national and local levels. These findings helped local public authorities effectively reduce transmission by targeting resources and efforts. Understanding the key questions for decision-making at all times, translating problems into a mathematical language, integrating different mathematical approaches to their solution and being able to present the results in an easy-to-understand way is vital to achieve a correct interpretation of the results that may have a timely impact on controlling the epidemic.

Author contributions

LS, RGD and PM conceived the study and coordinated the modeling work. LS, CF, RGD, CS, PLL contributed to writing the manuscript. Dynamical models was led by RGD, who programmed the model with help from WM. Machine learning models were implemented by YA, SE and AP. Risk analysis was carried out by PLL, PF, LS, ATR and NP. Statistical models were developed by CS, WB, CF, MCP, AC and LS. Branching process models and simulation were developed by JV and MC. PM, VN, LS participate in the epidemiologic interpretation of the models. MV participated in bibliographic searches and information management. All authors interpreted the study findings, contributed to the manuscript, and approved the work and the final version for publication.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

1. WHO. Coronavirus disease (COVID-2019) situation Reports. *World Health Organization*. 2020. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situationreports/ [Accessed May 21, 2020].

2. WHO. Strategic preparedness and response plan. 2020;(February).

3. Keck CW, Reed GA. The Curious Case of Cuba. *American Journal of Public Health*. 2012;102(8):13-22.

4. Jewell NP, Lewnard JA, Jewell BL. Predictive Mathematical Models of the COVID-19 Pandemic Underlying Principles and Value of Projections. *JAMA*. 2020;323(19).

5. Shinde GR, Kalamkar AB, Mahalle PN, Dey N, Chaki J. Forecasting Models for Coronavirus Disease (COVID - 19): A Survey of the State - of - the - Art. *Science*. 2020:1-15.

6. Boldog P, Tekeli T, Vizi Z, et al. Risk Assessment of Novel Coronavirus COVID-19 Outbreaks Outside China. *Journal of Clinical Medicine*. 2020;9:571.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

7. Goumenou M. COVID - 19 in Northern Italy : An integrative overview of factors possibly influencing the sharp increase of the outbreak (Review). *Molecular Medicine*. 2020;22:20-32.

8. Huang R, Liu M, Ding Y. Spatial-temporal distribution of COVID-19 in China and its prediction : A data-driven modeling analysis. *J Infect Dev Ctries*. 2020;14(3):246-253.

9. Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of COVID-19 : a mathematical modelling study. *Lancet infec Dis*. 2020;20:553-558.

10. Ferguson NM, Laydon D, Nedjati-gilani G, et al. Report 9 : Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Response*. 2020;(March).

11. Hellewell J, Abbott S, Gimma A, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health*. 2020;8:488-496.

12. Walker PGT, Whittaker C, Watson O, et al. Report 12 : The Global Impact of COVID-19 and Strategies for Mitigation and Suppression. *Response*. 2020;(March):1-19.

13. Enserink BM, Kupferschmidt K. With COVID-19, modeling takes on life and death importance. *Science*. 2020;367:6485-6486.

14. Plan C, Educa- H, Security S, Trade D, Defense C. Cuba ’ s COVID-19 strategy : Main actions thrhough April 23, 2020. *Medicc Rev*. 2020;22(2):50-52.

15. Wang W. Backward bifurcation of an epidemic model with treatment. *Mathematical Biosciences*. 2006;201:58-71.

16. White LF, Pagano M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. 2008;(December 2007):2999-3016.

17. Sebrango, Carlos Rafael; Martines D.A.; Sanchez L; Thilakarathne PJ, Fava, Emmanuel; Van der Stuyft P, Shkedy Z. Real-time parameter estimation of Zika outbreaks using model averaging. *Epidemiol Infect*. 2017;145:2313-2323.

18. Sebrango C, Sanchez L, Shkedy Z. Modeling dengue outbreak data using nonlinear mixed effects model. 2017;38(3):252-258.

19. Corral A, Pria MC. Designing a Living Conditions Index and Classification of the National Territory. *Revista Cubana De Medicina*. 2015;31(3):323-332.

20. Lorenzo-luaces P, Guerra-yi ME, Faes C. Spatial analysis of breast and cervical cancer incidence in small geographical areas in Cuba , 1999 – 2003. *European Journal of Cancer Prevention*. 2003:1999-2003.

21. Jacob C. Branching Processes : Their Role in Epidemiology. *International Journal of Environmental Research and Public Health*. 2010:1186 - 1204.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

22. Cori A, Ferguson NM, Fraser C, Cauchemez S. Practice of Epidemiology A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology*. 2013;178(9):1505-1512.

23. Díaz-canel M, Núñez-Jover J. Gestión gubernamental y ciencia cubana en el enfrentamiento a la COVID-19. *Anales de la Academia de Ciencias de Cuba*. 2020;10(2).

24. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*. 2020;382(13):1199-1207.

Figures

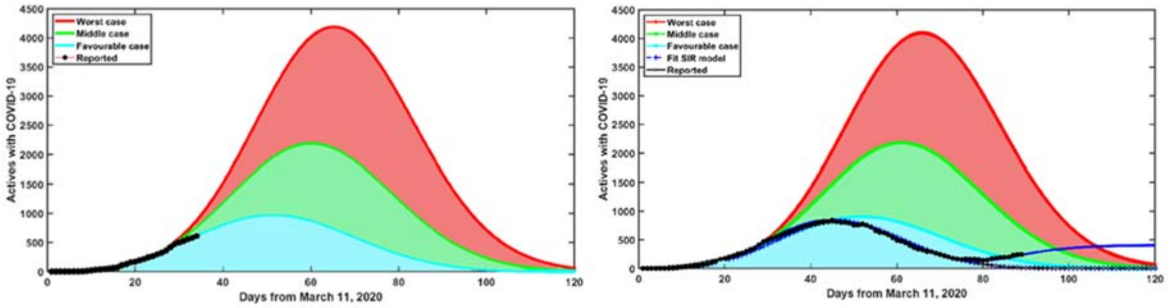

Figure 1 Active cases reported (dot) and predicted (line) by day in the worst (red), middle (green) and favoured (cyan) scenery. a) Forecast and real data until March 11, b) Real confirmed cases and forecast using the data until May 31.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

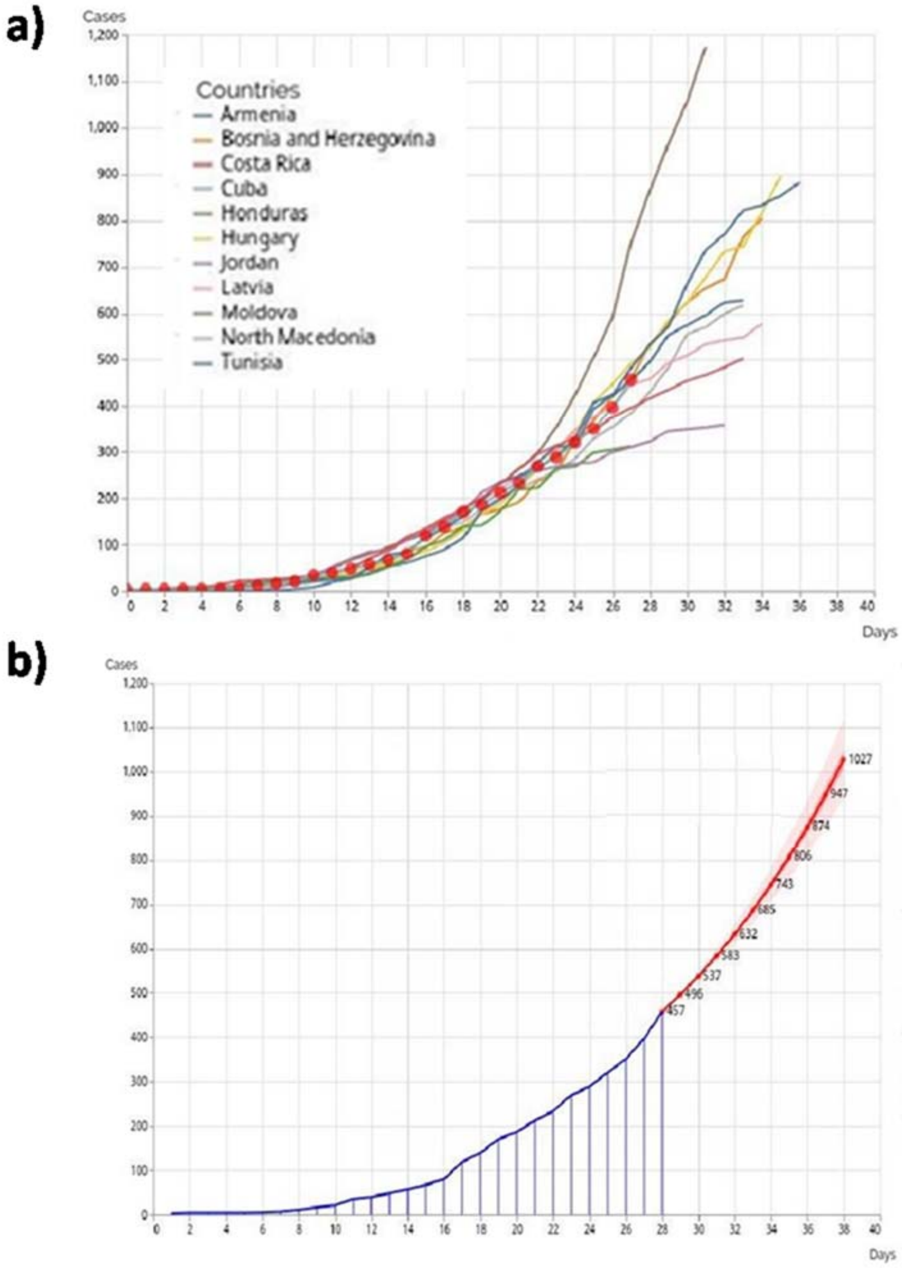

Figure 2. a) Cumulative cases by selected countries with demographics and epidemic patterns similar to Cuba until epidemic day 28 and b) Prediction of confirmed cases for Cuba until April 18, 2020 (day 38).

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

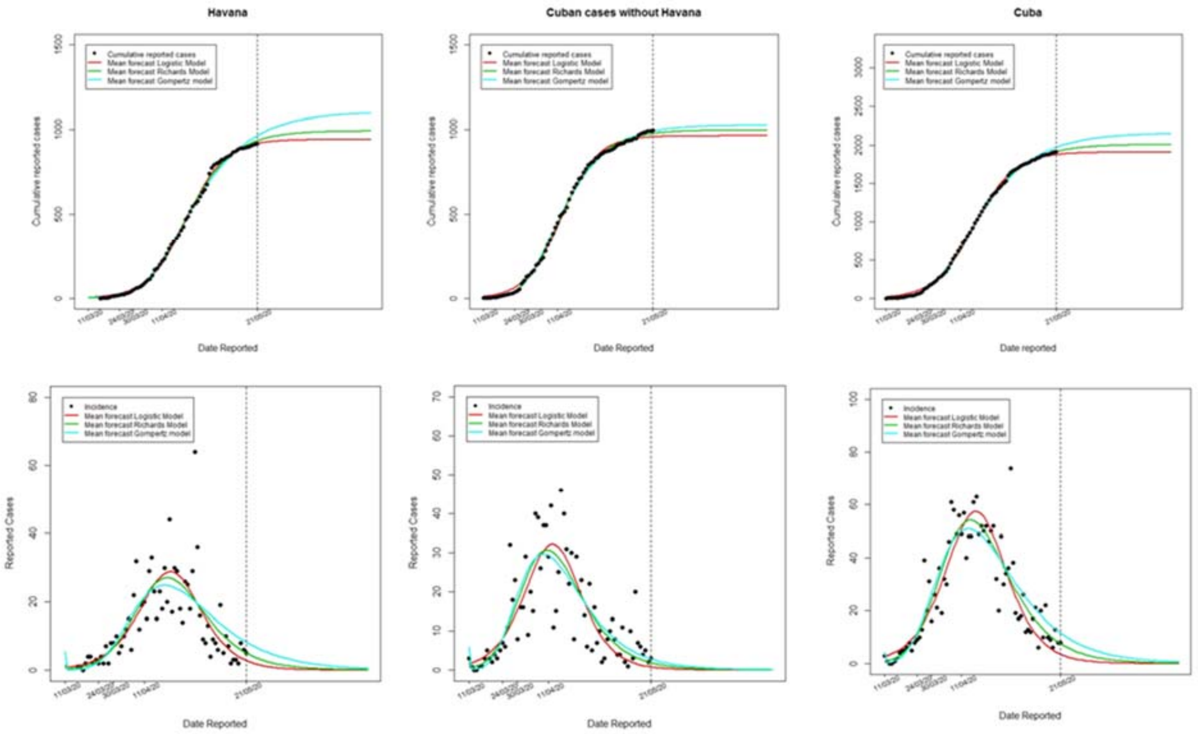

Figure 3. Cumulative, daily reported cases and forecast model estimation. a) Cumulative cases for Havana, b) cumulative cases outside Havana, c) cumulative cases for Cuba, d) incidence cases for Havana, e) incidence cases outside Havana and f) incidence cases for Cuba. Red, green and blue lines correspond to the Logistic, Richards and Gompertz models

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

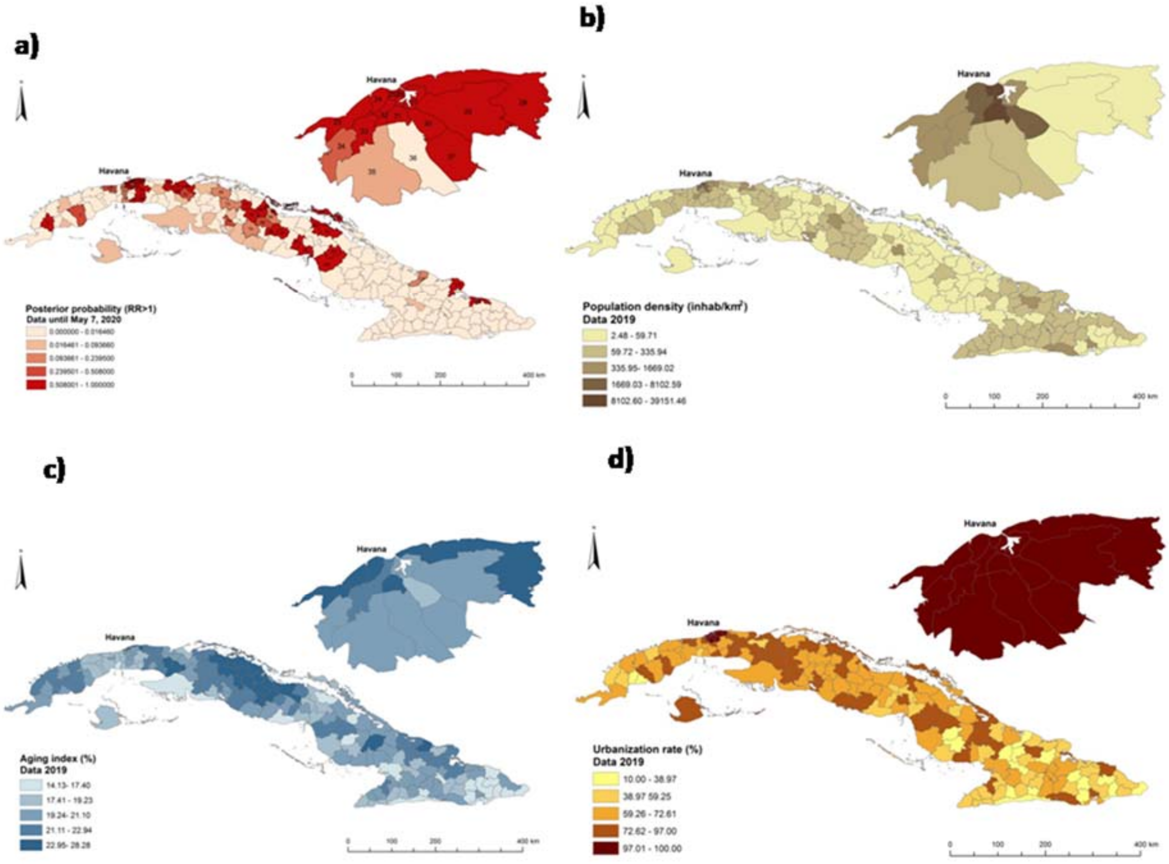

Figure 4. Posterior probabilities of relative risks exceeding unity for COVID-19 and demographics characteristics. (a) Posterior probabilities of excess risk for each municipality (b) population density (c) Aging index: number of elders (≥ 60 years) per 100 persons (d) urbanization (fraction of urban population per 100 persons).

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

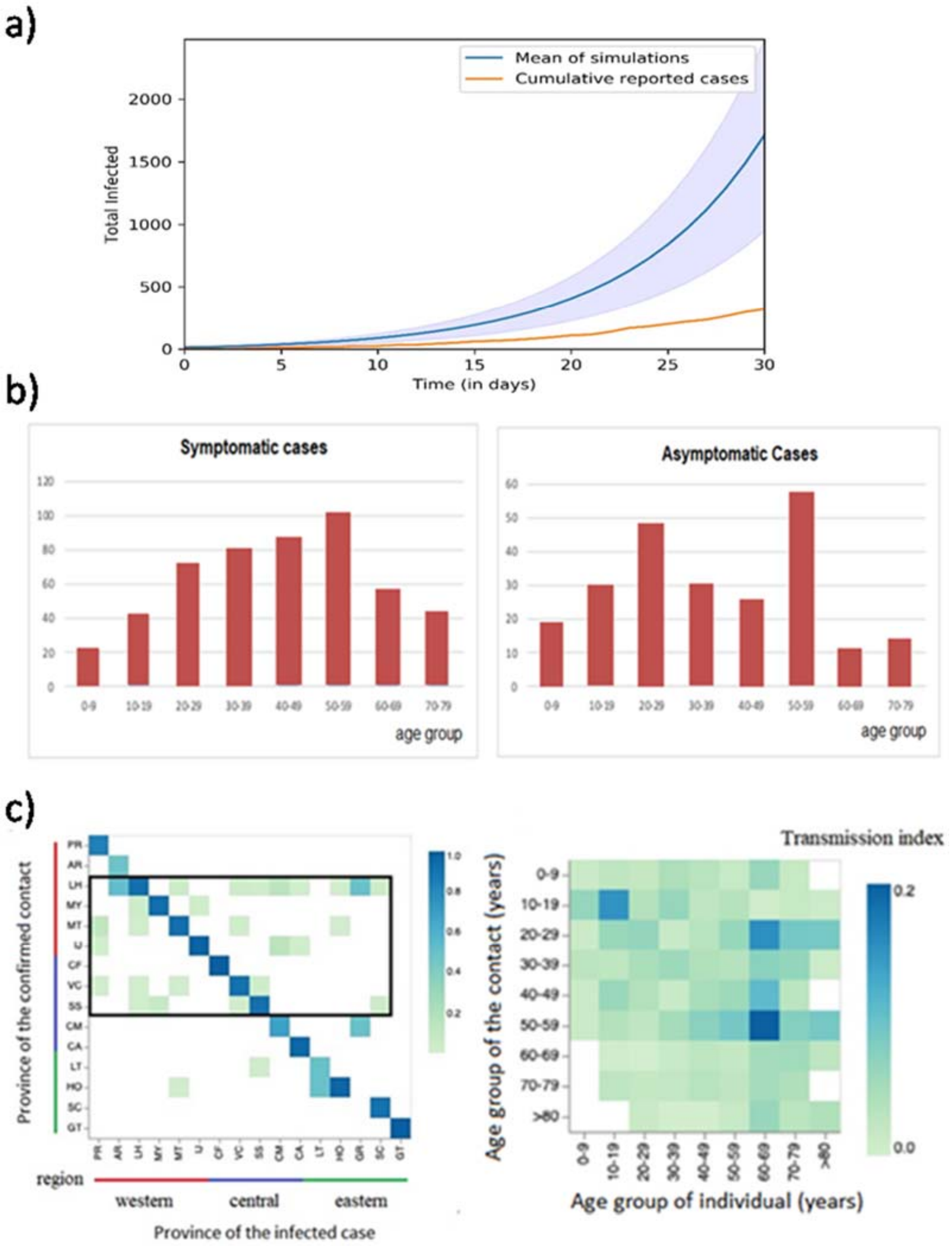

Figure 5. Effectiveness of case detecting strategy and control interventions: a) Predicted infected cases by the simulation model and cumulative confirmed cases at the beginning of the epidemic, b) age distribution of symptomatic (left) and asymptomatic (right) confirmed cases c) Transmission matrix by province (left) and by age (right). PR: Pinar del Rio, AR: Artemisa, LH: Havana, MY: Mayabeque, MT: Matanzas, IJ: Isla de la Juventud, CF: Cienfuego, VC: Villa Clara, SS: Sancti Spiritus, CM: Camaguey, CA: Ciego de Avila, LT: Las Tunas, HO: Holguin, GR: Granma, SC: Santiago de Cuba, GT: Guantanamo. Data until April 11.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

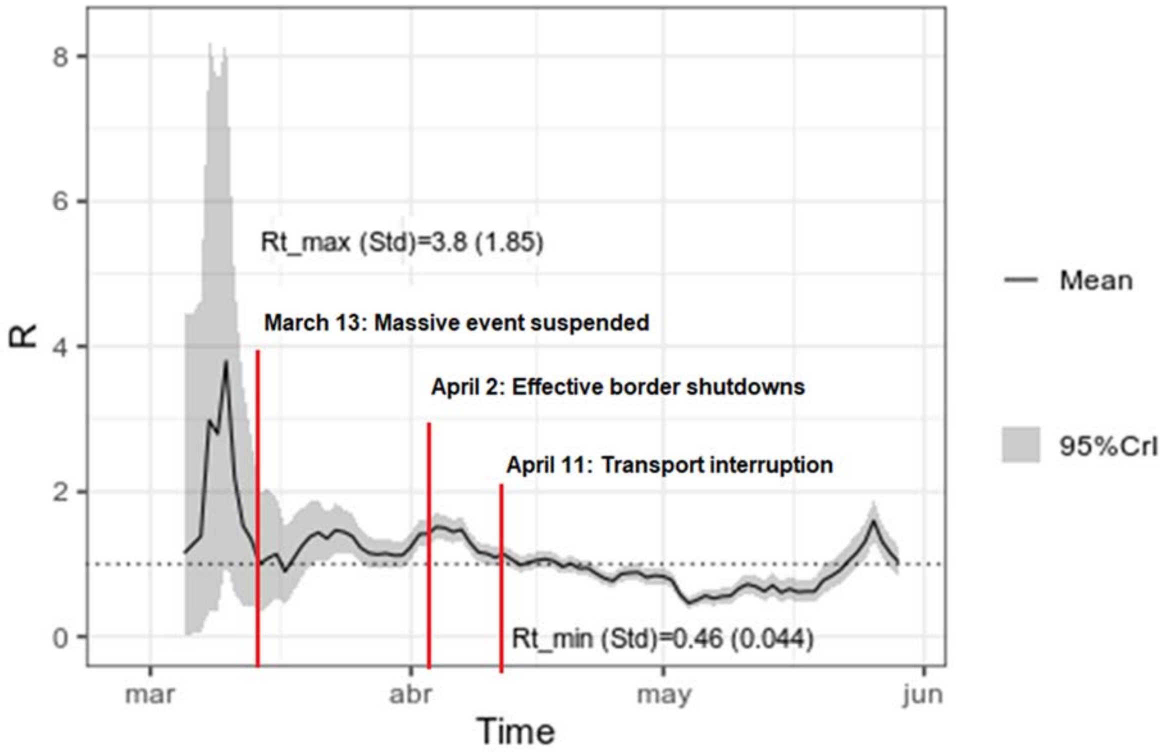

Figure 6. Time dependent reproduction number (Rt) estimated for Cuba and control interventions. Data until May 31, 2020.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

Supplementary information

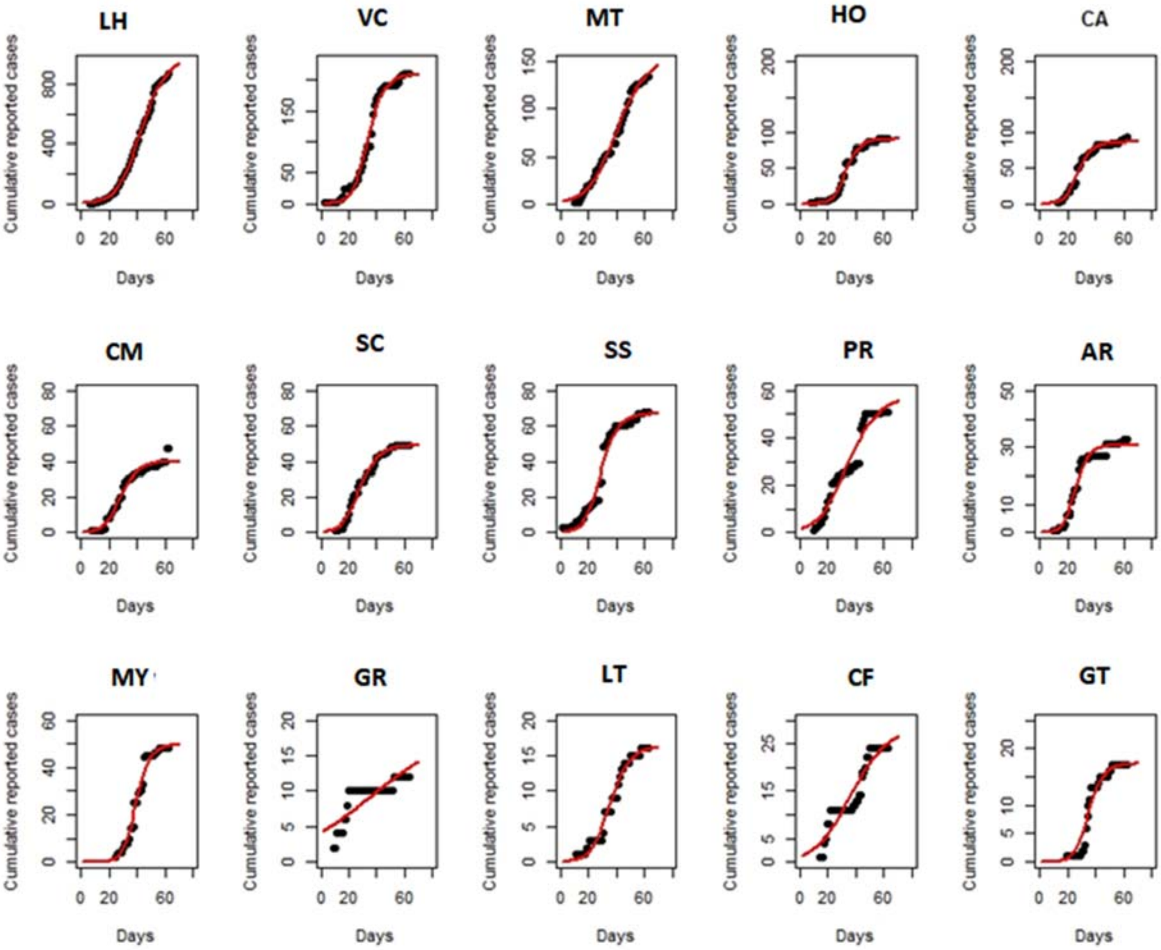

Figure. Cumulative reported COVID19 cases by province real (dot) and fitted (red lines) using non-linear mixed model. The point represents the real data and red line the estimation by the Richards Model. Data from March 11, 2020 to May 21, 2020.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271

Table. Municipalities with excess risk of COVID-19 (CAR-smoothed RR > 1.20 or posterior probability of smoothed relative riskgreater than unity of at least 0.90). Cuba, March 11 to May 21, 2020.

Province and Municipalities (code map)	Crude IR (95%CI)	CAR-smoothed RR (95%CI)	PP (RR>1)
Pinar del Rio			
Guane (11)	3.51 (1.75-5.46)	3.28 (1.95-4.97)	0.9999
Havana*			
Cotorro (37)	7.89 (6.25-9.62)	7.84 (6.37-9.46)	0.9999
Centro Habana (25)	4.79 (3.8-5.82)	4.77 (3.88-5.75)	0.9999
Cerro (32)	3.93 (2.99-4.91)	3.90 (3.07-4.83)	0.9999
Regla (27)	3.00 (1.53-4.61)	2.95 (1.8-4.38)	0.9999
Habana del Este (28)	2.9 (2.22-3.62)	2.88 (2.28-3.56)	0.9999
San miguel del Padron (30)	2.63 (1.94-3.36)	2.62 (2.01-3.3)	0.9999
Plaza de la Revolucion (24)	2.5 (1.81-3.24)	2.49 (1.89-3.18)	0.9999
Habana Vieja (26)	2.46 (1.52-3.49)	2.45 (1.67-3.39)	0.9999
Marianao (33)	2.31 (1.61-3.06)	2.3 (1.7-2.99)	0.9999
Diez de Octubre (31)	2.04 (1.51-2.6)	2.03 (1.57-2.56)	0.9999
Guanabacoa (29)	1.94 (1.26-2.67)	1.93 (1.36-2.6)	0.9999
Playa (23)	1.74 (1.22-2.3)	1.74 (1.29-2.26)	0.9999
La Lisa (34)	1.45 (0.91-2.05)	1.45 (0.99-1.99)	0.9929
Mayabeque			
Melena del Sur (46)	2.9 (0.68-5.45)	2.75 (1.25-4.83)	0.9963
San Jose de las Lajas (39)	2.32 (1.38-3.35)	2.28 (1.51-3.21)	0.9999
Batabano (47)	1.46 (0.05-3.11)	1.4 (0.53-2.73)	0.8119
Matanzas			
Limonar (56)	2.7 (0.86-4.79)	2.52 (1.25-4.25)	0.9973
Cardenas (50)	2.01 (1.4-2.67)	1.99 (1.47-2.59)	0.9999
Matanzas (49) **	1.93 (1.34-2.55)	1.91 (1.4-2.48)	0.9999
Villa clara			
Santa Clara (70)**	3.2 (2.61-3.82)	3.19 (2.64-3.78)	0.9999
Camajuani (66)	2.74 (1.56-4.03)	2.66 (1.72-3.82)	0.9999
Cifuentes (71)	2.15 (0.51-4.03)	1.99 (0.91-3.52)	0.9757
Caibarien (67)	1.3 (0.24-2.53)	1.24 (0.54-2.25)	0.7730
Santi spiritus			
Taguasco (85)	4.08 (2.14-6.21)	3.9 (2.4-5.75)	0.9999
Cabaiguan (86)	2.98 (1.81-4.25)	2.9 (1.95-4.03)	0.9999
Ciego de Avila			
Majagua (97)	5.16 (2.64-7.94)	4.83 (2.91-7.24)	0.9999
Venezuela (99)	4.09 (1.83-6.6)	3.81 (2.15-5.94)	0.9999
Bolivia (93)	2.63 (0.09-5.61)	2.29 (0.81-4.56)	0.9612
Moron (92)	2.06 (1.11-3.11)	1.99 (1.24-2.91)	0.9991
Camaguey			
Carlos M. de Cespedes (101)	9.59 (6.09-13.38)	9.18 (6.31-12.56)	0.9999
Florida (109)	1.48 (0.66-2.4)	1.4 (0.79-2.19)	0.9365
Holguin			
Frank Pais (133)	2.6 (0.61-4.88)	2.17 (0.93-3.95)	0.9785
Banes (124)	1.76 (0.93-2.69)	1.65 (1.01-2.46)	0.9919

*Capital of the country , **Capital of the province, IR: standard incidence rate, RR: Relative Risk, PP: Posterior probability.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=3629271