medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Noname manuscript No.
(will be inserted by the editor)

COVID-19 epidemic scenarios into 2021 based on observed key superspreading events

Mario Santana-Cibrian · M. Adrian Acuña-Zegarra · Carlos E. Rodríguez · Ramsés H. Mena · Jorge X. Velasco-Hernandez

Received: date / Accepted: date

Abstract Key high transmission dates for the year 2020 are used to create scenarios of the evolution of the COVID-19 pandemic in several states of Mexico for 2021. These scenarios are obtained through the estimation of a time-dependent contact rate, where the main assumption is that disease behavior is heavily determined by the mobility and social activity of the population during holidays and other important calendar dates. First, changes in the effective contact rate on predetermined dates of 2020 are estimated. Then, this information is used to propose different scenarios for the number of cases and deaths for 2021. The fundamental assumptions behind this methodology are that the effective contact rate incorporates the main superspreading transmission events during last year, that each region has an independent epidemic not explicitly interconnected with other regions and, finally, that there are no new highly transmissible SARS-CoV-2 vari-

Mario Santana-Cibrian
CONACYT-Instituto de Matematicas UNAM
Queretaro, 76230, Mexico
E-mail: msantana@im.unam.mx

M. Adrian Acuña-Zegarra
Departamento de Matemáticas
Universidad de Sonora, Sonora, 83000, Mexico
E-mail: adrian.acuna@unison.mx

Carlos E. Rodríguez
Department of Probability and Statistics, IIMAS UNAM
CDMX, 01000, Mexico
E-mail: carloserwin@sigma.iimas.unam.mx

Ramsés H. Mena
Department of Probability and Statistics, IIMAS UNAM
CDMX, 01000, Mexico
E-mail: ramses@sigma.iimas.unam.mx

Jorge X. Velasco-Hernández
Instituto de Matemáticas UNAM
Queretaro, 76230, Mexico
E-mail: jx.velasco@im.unam.mx

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ants active during the timeline of the forecasts. Also, several levels of vaccination are considered to analyze their impact on the projections of the epidemic curve. The objective is to generate a range of scenarios that could be useful to evaluate the possible evolution of the epidemic and its likely impact on incidence and mortality.

Keywords Superspreading · SARS-CoV-2 epidemic · Kermack-McKendrick models · Epidemic projections · Bayesian inference

1 Introduction

As a consequence of the COVID-19 epidemic's impact on the regional and global economy, one of the most relevant problems that every country faces is to decide how and when businesses, public centers, tourism, schools, and universities can safely reopen [21]. Therefore, it is of the highest importance to postulate plausible scenarios for the evolution of the SARS-CoV-2 pandemic to design effective reopening strategies for decision-makers. This knowledge is even more pressing in countries that lack of infrastructure (i.e., testing, contact tracing) to acquire a more precise or, perhaps we should say, a less uncertain idea of the behavior of the pandemic on days or, ideally, weeks into the future.

During 2020, much effort was centered on projecting the fate of the COVID-19 pandemic and evaluating the efficacy of the mitigation strategies adopted to contain it [7, 14]. The transmission dynamics of the pandemic has been modeled using many different methodologies, several of them centered on estimating the effective reproduction number R_t or using some version of the well-known Kermack-McKendrick model [8, 15, 28, 33].

Around the world, nonpharmaceutical interventios (NPIs) vary in strength. Examples include lockdowns, use of face masks, regulations on the number of people allowed in meetings, and so forth. They range from strict and mandatory enforcement by the governments to a voluntary personal decision. The rationale behind any particular version of a mitigation strategy followed in each country or region is generally based on local or national conditions that intermix public health, economic and political factors and perspectives [4]. One factor, however, that has shown to be decisive in the success or failure of a given containment or mitigation strategy and that has show remarkable difficulty to control or project is human behavior. Recently, some efforts in modeling this topic have been reported [3].

In Mexico, states present different shapes of epidemic curves. For example, Fig. 1 shows the epidemic curves of Mexico City, Queretaro state, and Tamaulipas state until May 2021. There are some features in these curves worth describing. One is the initial slow growth of the epidemic that seems almost linear in Mexico City and Queretaro state with a shorter growth period in Queretaro state; the second is the long plateau that develops in Mexico City and a much shorter one in Queretaro state, and the third an epidemic curve with two peaks in Tamaulipas.

The initial quasi-linear growth of the epidemic cannot be explained by classical models. Previously it has been have argued that the initial slow growth was a consequence of an early application of largely voluntary mitigation measures [1]. The first recorded case in Mexico occurred by the end of February 2020, and the

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

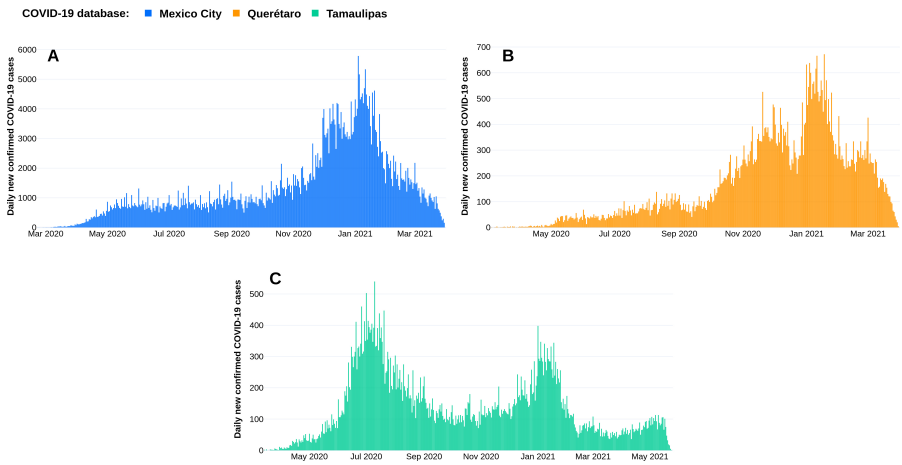

Fig. 1 Daily number of COVID-19 confirmed cases by symptoms onset until May 24, 2021 for A) Mexico City, B) Queretaro state and C) Tamaulipas state.

first mitigation measures were applied 25 days later, on March 23, 2020. Moreover, there were superspreading events in Mexico City during Easter celebrations (April 6-12) and early May (April 30 to May 10) that shifted the day of maximum incidence to the end of May [1,26], and pushed the epidemic into a quasi-stationary state characterized by values of $R_t \approx 1$. This behavior, observed around the world, has been explored in [16,27,31]. In particular,
[31] claim that this quasi-linear growth and the maintenance of the effective reproduction number around $R_t \approx 1$ for sustained periods involves critical changes in the structure of the underlying contact network of individuals.

Superspreading events are indeed associated to heightened population mobility and social activity [2,5,9,18,24,25,29,32,34]. It is a phenomenon that has shown to be determinant for the time evolution of many infectious diseases [19]. The SARS-CoV-2 epidemic dynamic observed in Mexico City and, in general, in the country, has been driven by events associated with heightened mobility and increased social activities during holidays and other important calendar key dates. The characteristic plateau of Mexico City reached by the end of May, 2020, for example, has been explained by a succession of events that were essentially superspreading events [1,26].

Now that we have more than one year of data regarding the evolution of the COVID-19 pandemic, the following question arises: can we use the past history of the epidemic curve to project the main likely tendencies in the behavior of COVID-19 for the rest of 2021? The objective of the present work is to provide an answer to that question for the case of the Mexican epidemics. In order to do that, a characterization of the transmission dynamics of COVID-19 is presented based on the assumption that human behavior during specific dates is the main factor driving the evolution of the pandemic. It is postulated that a modification of the Kermack-McKendrick model with a time-dependent effective contact rate can accurately describe the observed number of confirmed cases and deaths of COVID-19. The effective contact rate will change in calendar dates associated with religious or civic holidays, commercial incentives and government interven-

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

tions, when superspreading events or changes in tendency occurred. Then, this characterization of the transmission dynamics of the disease is used to develop a methodology to generate epidemic scenarios that use the known history of the disease recorded during 2020 and early 2021. The effect of these change points or key calendar dates on the contact rate occurs at a much faster scale than changes due to the implementation or lifting of NPIs [12,17]. All these dates are known in advance, and activities can be planned (short vacations, family visits, among others). We focus on the epidemic developing in Mexico, taking as examples three particular states (Mexico City, Queretaro state and Tamaulipas state).

Our approach is similar to the one presented in [12]. However, to determine changes in the effective contact rate, besides looking at the actual effects of governmental mitigation measures, we look at particular events on dates related to civic, religious, and official vacation periods known in advance each year.

The paper is organized as follows. Section 2 presents the methodology used. Section 3 describes the results of the estimation process and the proyected scenarios. Finally, Section 4, contains the discussion about this work.

2 Methodology

This section describes the underlying mathematical model, the estimation of the time-varying effective contact rate and the basic assumptions for the projection of scenarios.

2.1 Mathematical model

A compartmental model is used to describe the evolution of the COVID-19 epidemic. Fig. 2 shows the corresponding model diagram. The model considers three

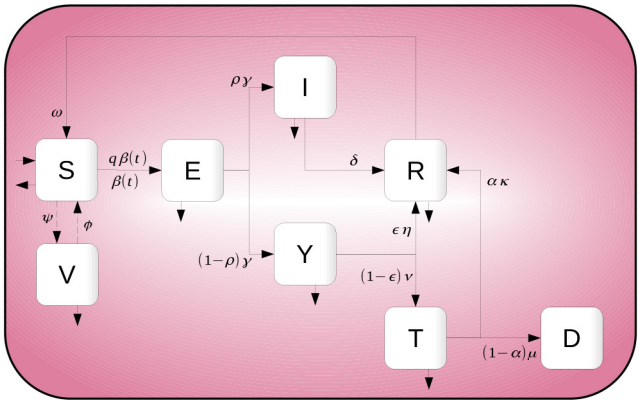

Fig. 2 Diagram for the mathematical model. There are three types of infectious individuals: asymptomatic (I), symptomatic (Y) and reported (T), but this last group does not play a role in the transmission. The other state variables represent: susceptible (S), vaccinated (V), exposed (E), recovered (R) and dead (D).

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

classes of infected individuals: asymptomatic (I), symptomatic (Y), and reported (T). Once reported, infected individuals are effectively isolated and are no longer participants in the transmission process. The three infected classes can recover (R) but only reported symptomatic individuals can die (D) since we consider the ideal case where all severe cases are reported. Susceptible (S) individuals can be vaccinated (V) with an effective vaccination rate ψ. Both vaccinated and reported individuals will eventually return to the susceptible state after a certain, possibly different, immunity period. Vital dynamics are also included since this work is aimed to produce mid-term scenarios for the epidemic. The full set of equations is given by

$$\begin{aligned}
S' &= \omega R + \phi V - \psi S - \beta(t)(Y + qI)\frac{S}{N} + \chi M - \chi S, \\
V' &= \psi S - \phi V - \chi V, \\
E' &= \beta(t)(Y + qI)\frac{S}{N} - \gamma E - \chi E, \\
I' &= \rho\gamma E - \delta I - \chi I, \\
Y' &= (1-\rho)\gamma E - \epsilon\eta Y - (1-\epsilon)\nu Y - \chi Y, \\
T' &= (1-\epsilon)\nu Y - \alpha\kappa T - (1-\alpha)\mu T - \chi T, \\
R' &= \delta I + \epsilon\eta Y + \alpha\kappa T - \omega R - \chi R, \\
D' &= (1-\alpha)\mu T,
\end{aligned} \tag{1}$$

where $N = S + V + E + I + Y + R$ is the population that participates in the infectious process and $M = N + T$ is the total population at time t.

The effective vaccination rate is deduced as

$$\phi = -\ln(1 - a \times b)/c, \tag{2}$$

where a is the target coverage (% of the population targeted to receive the vaccine), b the vaccine efficacy and c the time horizon in which a should be achieved. In what follows ϕ defines plausible scenarios because in Mexico vaccination roll-outs are regionally heterogeneous and vaccine availability is limited.

The basic reproduction number is given by

$$\mathcal{R}_0 = \beta_0\Big(\frac{q\rho\gamma}{(\delta+\chi)(\gamma+\chi)} + \frac{(1-\rho)\gamma}{(\epsilon\eta + (1-\epsilon)\nu + \chi)(\gamma+\chi)}\Big).$$

The parameter β_0 is the effective contact rate at the start of the epidemic. Furthermore, the effective reproduction number is calculated as

$$R_t = \mathcal{R}(t)S(t)/N(t), \tag{3}$$

where

$$\mathcal{R}(t) = \beta(t)\Big(\frac{q\rho\gamma}{(\delta+\chi)(\gamma+\chi)} + \frac{(1-\rho)\gamma}{(\epsilon\eta + (1-\epsilon)\nu + \chi)(\gamma+\chi)}\Big).$$

Table 1 shows a description of all model's parameters and their preset values.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Parameter	Description	Value	Units
$\beta(t)$	Effective contact rate	estimated	days^{-1}
ψ	Effective vaccination rate	variable	days^{-1}
ϕ	Vaccination immunity rate	0.005556	days^{-1}
ω	Natural immunity rate	0.005556	days^{-1}
γ	Incubation rate	0.196078	days^{-1}
δ	Asymptomatic recovery rate	0.142857	days^{-1}
η	Symptomatic recovery rate	0.071429	days^{-1}
κ	Reported recovery rate	0.1	days^{-1}
ν	Screening rate	estimated	days^{-1}
μ	Death rate	estimated	days^{-1}
ρ	Proportion of asymptomatic	0.2	%
ϵ	Proportion of symptomatic recovered	0.85	%
α	Proportion of reported recovered	estimated	%
q	Asymptomatic infectiousness reduction	0.45	%

Table 1 Description of the parameters involved in model 1 (see [1,26] for sources).

2.2 Effective contact rate estimation

The main feature of the model is the time-dependent effective contact rate $\beta(t)$. This function is defined by a piecewise cubic Hermite interpolating polynomial [13]. We set b_k as the effective contact rate at the k-th time change point, i.e. $\beta(t_k) = b_k$, $k = 1, \ldots, K$. Fig. 3 shows a diagram of the structure of $\beta(t)$. Hermite interpolation is used to guarantee that $\beta(t)$ remains positive for all times. We

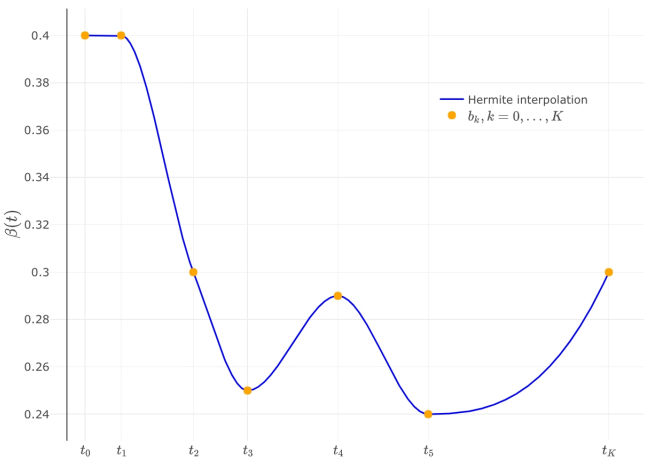

Fig. 3 Construction of the effective contact rate $\beta(t)$. It is defined by a piecewise cubic Hermite interpolating polynomial, parametrized by K change points at fixed times $t_1, t_2, \ldots, t_K$ and their corresponding values $b_1, \ldots, b_K$.

emphasise that, instead of choosing the change times t_k equally spaced along the observation period, they are located at predetermined calendar dates associated to

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

superspreding events, government interventions or other events that could affect the evolution of the epidemic curve.

The function $\beta(t)$ (parameters $t_1, t_2, \ldots, t_K, b_1, b_2, \ldots, b_K$) must be estimated from data. We use federal national records from daily reported cases and deaths to create point and interval estimates of these parameters, within a Bayesian inference framework. Details about the inference can be found in Appendix A.

2.3 Projected scenarios

Our main point is that the effective contact rate $\beta(t)$ in 2020 and early 2021 provides information that can be used to create scenarios for the evolution of the epidemic curve for the rest of 2021. There exists key calendar events or fixed dates where the Mexican population has family gatherings, political or civic activities every year, that during 2020 were associated to heightened transmission, heightened hospital demand, and increases in deaths. If this observation holds then, there are several ways in which we can project trends and define scenarios. One way is by simply prolonging for a few weeks or months the last observed trend of $\beta(t)$ and see the effect on the number of cases and deaths using the model. This is a common approach to create short-term forecasts. It usually works as long as that last observed trend does not change in the data, which could occur in a few days, weeks, or even months. However, we can improve this methodology using our estimation of $\beta(t)$: it is possible to know which dates are associated with changes in the trend and the magnitude of the contact rate. Therefore, we can assume that the percentage change observed in $\beta(t)$ between two consecutive dates will be repeated in the same period the next year. To exemplify this idea, suppose that $\beta(t)$ on December 24, 2020, was 0.2 and on December 31, 2020, was 0.3, implying that the contact rate increased by 50%. Then, to predict the change in the same period of 2021, the estimated contact rate on December 24, 2021, lets say 0.05, will increase a 50% to reach 0.075. Starting this process on the last date of the estimated contact rate, it is possible to generate projections that incorporate the history of the epidemic.

3 Results

To exemplify this methodology, the epidemic data of several states of Mexico will be used to estimate the effective contact rate $\beta(t)$ and then to propose scenarios for the rest of 2021 and early 2022.

3.1 Key superspreading events

The overall perspective of the epidemic management in Mexico can be fairly summarized by saying that a) all non-pharmaceutical interventions have been and are, to date, non-mandatory but voluntary except for some policies regarding closure or reopening of businesses; and b) the population increased mobility and social activities during certain periods is strongly associated to nation-wide holidays. The list of dates that we consider important for every state is given in Table 2.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Date	Description	Date	Description
2020-03-13	First COVID-19 case	2020-12-24	Christmas
2020-03-23	NPI start	2020-12-31	New year's eve
2020-04-30	Children's day	2021-01-06	Wise men day
2020-05-10	Mother's day	2021-02-14	Valentine's day
2020-06-21	Father's day	2021-03-14	Benito Juarez birth
2020-09-16	Independence day	2021-04-04	Easter
2020-11-02	Day of the Dead	2021-04-30	Children's day
2020-11-21	Buen Fin (Black Friday equivalent)	2021-05-10	Mother's day
2020-12-12	Day of the Virgin of Guadalupe		

Table 2 Key calendar dates associated to superspreading events used as change points for the effective contact rate $\beta(t)$ in the state of Tamaulipas. Except for the date of the first COVID-19 case (by symptoms onset), all the other dates are the same for each state analyzed. Mexico City reported its first case on February 20, 2020 and Querétaro state reported its first case on March 5, 2020.

To provide evidence on the key calendar dates superspreading hypothesis, several indicators are analyzed such as positivity, hospitalizations, mortality, among others. Only positivity is presented but other indicatores can be found in the supplementary material. When analyzing raw counts of the number of cases by day, the variability between days makes it difficult to determine the behavior of the data curve. To deal with this problem, we use a simple moving average with a window size of seven days, making it easier to identify overall trends. The positivity rate is defined as the ratio of positive tests to suspected cases.

Figure 4 shows the positivity rate for three states: Mexico City, Queretaro and Tamaulipas. Estimations are noisy but still it can be appreciated that, depending on the length of the holiday, there is an increase either immediately after or even during the key event. Also, changes in the trend can be associated to several key dates. As an example, December celebrations such as Christmas cause an increase in the positivity rate, which starts decreasing after New Year.

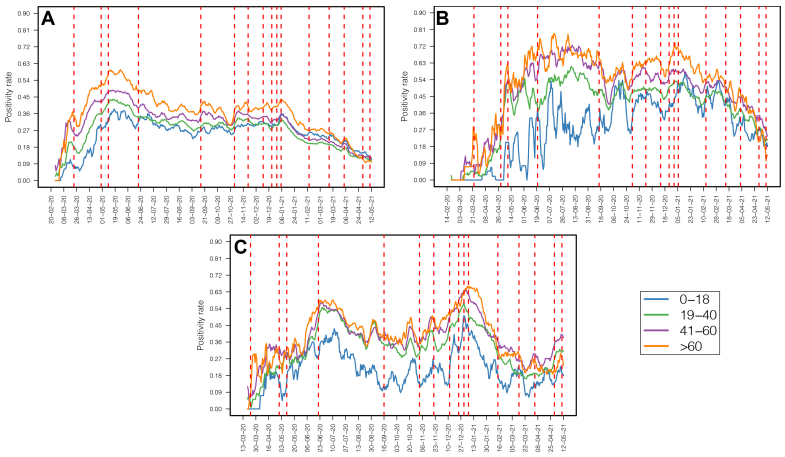

Fig. 4 Positivity rate by age group for A) Mexico City, B) Queretaro state and C) Tamaulipas state. Dashed red vertical lines give the dates listed in Table 2.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

It must be pointed out that changes in these indicators are not exclusive to the key dates we have selected (Table 2). Also, not necessarily all the dates chosen have the same impact for each state. However, our main hypothesis is that significant changes are indeed associated to the dates previously selected and they are enough to drive the general epidemic trends in each state.

3.2 Effective contact rate estimates for Mexico

As mentioned in Section 2.2, the effective contact rate is determined by an interpolating function and parametrized by $(t_0, t_1, \ldots, t_K, b_0, b_1, \ldots, b_K)$. Times $t_0, \ldots, t_K$ are determined by the key dates in table 2. We point out that $\beta(t_0)$ and $\beta(t_1)$, b_0 and b_1 respectively, are equal. Given that t_1 represents the time then the first mitigation measures were implemented, setting $b_0 = b_1$ implies that the effective contact rate is constant right before the mitigation interventions that altered the natural progression of the epidemic. All other parameters $b_1, \ldots, b_K$ are estimated. It is assumed that, at the beginning of the pandemic, there are no vaccinated individuals, no reported infected individuals, no recovered cases and no deaths. The number of exposed, asymptomatically infected and symptomatically infected individuals are also estimated. Finally, the COVID-19 death rate, the screening rate, and the proportion of reported recovered individuals can vary from one state to another and therefore are also estimated. See Appendix A for more details.

Fig. 5 shows the point-wise posterior median estimate of $\beta(t)$ for several states in Mexico, including the three examples previously discussed: Mexico City, Queretaro and Tamaulipas. Although each place has a different curve there are, nonetheless, common patterns. Notice that the rate remains low between June 2020 and September 2020 and the effect of the winter holidays is clear in each state. It is interesting that a few states showed an increase in the contact rate after the start of mitigation measures such as Queretaro, Guanajuato and Campeche. Overall, we can group states that show very similar behavior, a type of information that could potentially be used to classify epidemic curves. This will be explored in a forthcoming work.

3.3 Scenarios

As described above, we modify the last estimate of the effective contact rate, $\beta(t)$, corresponding to April 30, 2021, with the purpose of generating projections supported on the effective contact rate patterns observed on the same dates in 2020. However, the one aspect of the epidemic that is clearly different from last year's is the availability of vaccines, and thus this factor requires especial handling. Vaccination in places like Mexico is difficult to model since many different vaccines are applied, each with very different characteristics, such as the ones produced by Pfizer, Moderna, AstraZeneca and CanSino, to mention some. Moreover, the national vaccination strategy contemplates several stages depending on the stock of vaccines, which at the moment is very limited. By the end of May 2021, 16% of the population has been vaccinated, although only 9% has a complete vaccination scheme [20]. Nationally, vaccination started at the end of December 2020, but there are differences in the vaccination roll out in each state and limited information

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

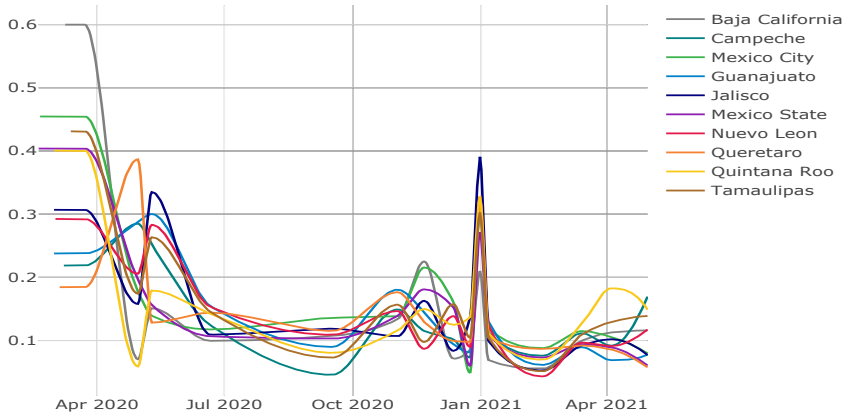

Fig. 5 Effective contact rate $\beta(t)$ estimated for several Mexican states from the beginning of the pandemic (late February to early March) in 2020 until April 30, 2021.

about it. For the purposes of this work, it is assumed that vaccination started on February 15, 2021, which is the date when most of the states started vaccination for the general population above 60 years old. The efficacy of the vaccines is set to 0.95 and that the expected coverage of the vaccines is 30%, 50% and 70% of the population after the first year. These three scenarios are considered to analyze the effect of vaccination.

Importantly, the variability of the estimates of $\beta(t)$ is high, especially at the beginning of the epidemic. This is mainly caused by the initial conditions that are also estimated. Kermack-McKendrick models are very sensitive to changes in these conditions. As a consequence, a subset of projections for some states is extremely high and biologically unfeasible as they imply very high values of the effective reproduction number R_t. For that reason, the predictions shown in this work are limited to those with estimates of R_t for the prediction period below the maximum observed R_t in the past. We think this restriction is reasonable.

Fig. 6 shows the results for Tamaulipas state. The model correctly describes the main trends during the estimation period. The effect of the vaccination is to decrease the number of new reported cases and deaths, although qualitatively, the behavior of the predictions is the same. Thus, if similar conditions to 2020 occur in 2021, the most likely scenarios indicate that the number of new infections will be low and decreasing for the rest of the year. Nonetheless, there is still the possibility of an important increase of the number of infections in the following months.

Fig. 7 shows the scenarios for Mexico City. In this case the most likely trend is that the number of cases and deaths decrease in time with very low levels on infections for the rest of the year. There is a small probability of an outbreak at the end of 2021 and the beginning of 2022, and its magnitude could by similar to the one observed for the same period in 2020 if only 30% is vaccinated by that date. Fig. 8 shows the scenarios for Queretaro state. In this case the number of cases and deaths go down very fast and there will be no more outbreaks.

Black bars in Figs. 6, 7 and 8 show two weeks of available data on reported cases and deaths for the last three weeks that was not used for in the estimation.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

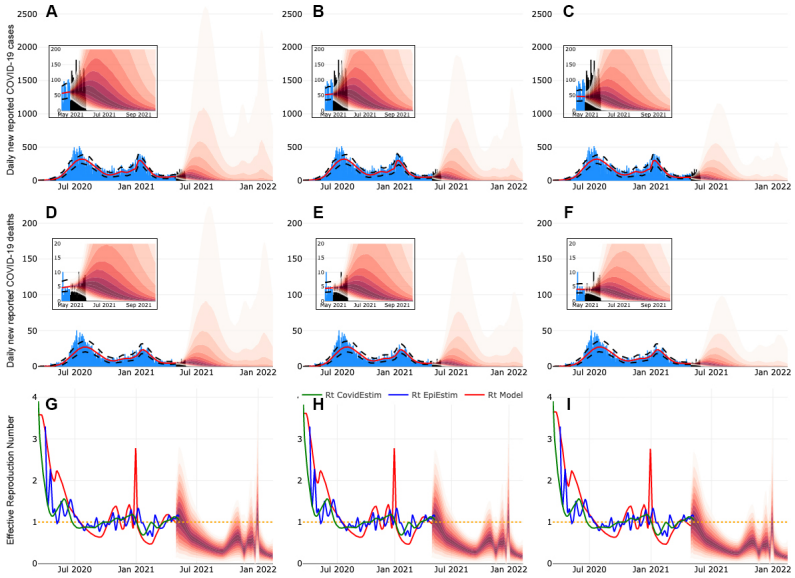

Fig. 6 Projected scenarios for Tamaulipas state. First, second and third columns show scenarios when target vaccination coverage are 30%, 50% and 70%, respectively. These coverages are achieved in a year. The insets show an enlarged view of the first months of each projection.

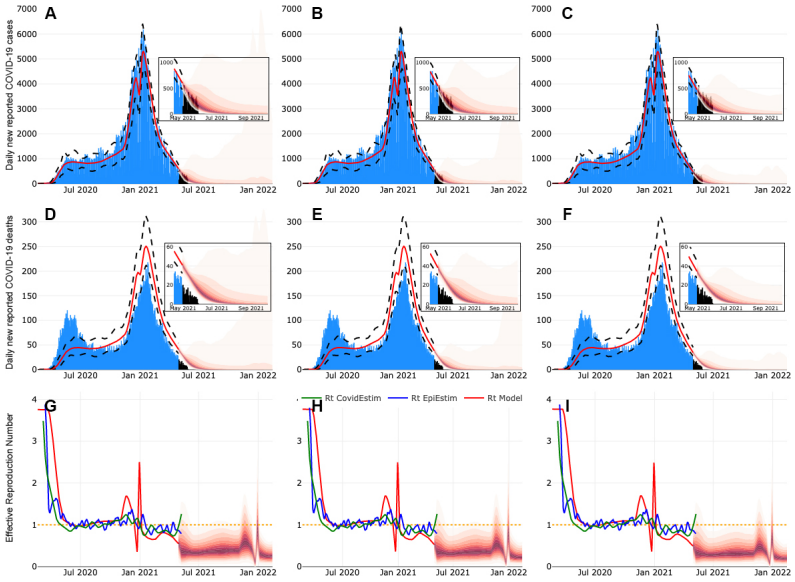

Fig. 7 Projected scenarios for Mexico City. First, second and third columns show scenarios when target vaccination coverage are 30%, 50% and 70%, respectively. These coverages are achieved in a year. The insets show an enlarged view of the first months of each projection.

Projections agree with the observed data. These figures also show 3 estimates of the effective reproduction number for each scenario according to three different

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

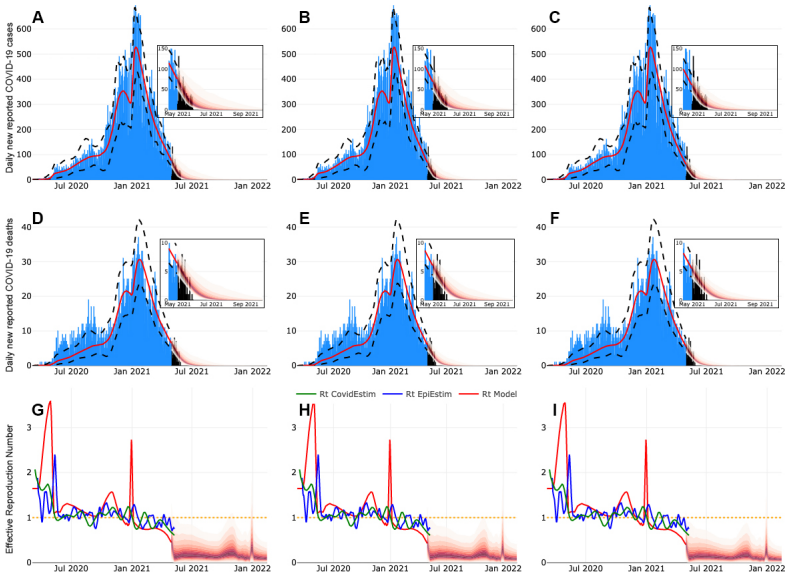

Fig. 8 Projected scenarios for Queretaro state. First, second and third columns show scenarios where target vaccination coverage are 30%, 50% and 70% during the first year, respectively. The insets show an enlarged view of the first months of each projection.

methodologies, one derived form the equation 3 and the other two obtained using EpiEstim [6,30] and Covidestim [10,22].

4 Discussion

Our objective is to generate a range of scenarios using the known history of the epidemic, that could be useful to evaluate its possible evolution and its likely impact on incidence and mortality. The fundamental assumptions behind this methodology are that the effective contact rate incorporates the main superspreading transmission events during last year, that each region has an independent epidemic not explicitly interconnected with other regions and, finally, that there are no new highly transmissible variants active during the timeline of the forecasts.

The magnitude and changes of the contact rates on key dates on 2020 together with the magnitude of R_t, constitute two indicators that provide an approximation of what could happen during 2021. The main assumptions are that these dates will again constitute superspreading events and that the evolution of the epidemic curve in 2021 will still be highly influenced by the same events. The idea is simple but the use of key calendar dates creates a practical basis, in the absence of significant testing and contact tracing, to anticipate changes in the transmission of the disease that can provide mid to long term outbreak risk projections. There is no doubt that superspreading events will still be important for the COVID-19 pandemic however, it is not clear what the magnitude of their impact will be. The methodology presented here is a step towards the solution of this problem.

It could be argued that this procedure will not be useful because the conditions of the previous year will not be the same for the next one. However, in places like

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Mexico the proposed procedure is sensible because: 1) lockdowns and other non-pharmaceutical interventions are not mandatory, 2) there is very limited testing and practically no contact tracing, and 3) significant superspreading events occurred mostly during holidays that happen every year. Nonetheless, we are aware that, for now, only the short-term predictions can be evaluated.

New events that were not present during 2020 can have an important impact in the evolution of the epidemic curve during 2021. For example, on June 6, 2021, there will be Federal Elections in the country, which could potentially constitute a superspreading event. If that were the case, the current projections may fail. However, it is possible to adjust the estimates to accommodate this date. The process will be less immediate than the one used here because there were no elections during 2020. Nonetheless, scenarios can be created based on the changes observed in other superspreading events. As was mentioned before, there are several ways of using the information contained in $\beta(t)$.

Our model does not explicitly introduces seasonality that could start acting on the epidemic by October 2021 under the assumption that COVID-19 behaves similarly to influenza. However, since the pattern of changes of the observed contact rate for 2020 is being used, any seasonal effect with impact on last year's contact rate will be inherited into the projected contact rates of 2021.

Even without the projections, the estimation of a time dependent effective contact rate alone provides interesting information. The evolution of the epidemic curves of different locations can be compared, and possibly grouped, using these estimations. In the case of Mexico, a common feature to all states is that the effective contact rate had one fast period of growth on the winter holidays (December 12 to December 31, 2020), supporting the idea that superspreading events did occur in the proposed dates. Even when not all the key dates considered in the analysis have an significant effect in all the states, the results suggest that it is important to closely look at the behavior of the epidemic on these dates.

Finally, this study highlights the importance of enforcing the application of NPIs in Mexico and countries of similar characteristics. In all our scenarios, the reduction in $\beta(t)$ can be achieved only through them, given the low vaccination rate that exists at the end of May. This is not surprising, a recent study [23] has shown that even with high efficacy vaccines and a greater coverage than the current one in Mexico, vaccination alone cannot achieve a significant reduction in cases if NPIs are relaxed too quickly.

Acknowledgements

JXVH and MSC acknowledge support from UNAM PAPIIT DGAPA IN115720 and IV100220 grants. CERHV and RHM are grateful for the support of UNAM PAPIIT grant IG100221. MAAZ acknowledges support from PRODEP Programme (No. 511-6/2019-8291).

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Author contributions statement

MSC, MAAZ and JXVH conceived the idea; CERHV and RHM analyzed data, MSC and MAAZ performed the model projections and parameter estimation, all authors discussed, reviewed and wrote the paper.

Additional information

Further analysis for other Mexican states is shown in the site:
https://github.com/ProjectsEpi/KeySuperspreadingEvents.git

Competing interests

The authors report no competing interests.

References

1. Acuña-Zegarra, M.A., Santana-Cibrian, M., Velasco-Hernandez, J.X.: Modeling behavioral change and COVID-19 containment in Mexico: A trade-off between lockdown and compliance. Mathematical biosciences **10**, 108370 (2020). DOI 10.1016/j.mbs.2020.108370
2. Adam, D.C., Wu, P., Wong, J.Y., Lau, E.H., Tsang, T.K., Cauchemez, S., Leung, G.M., Cowling, B.J.: Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nature Medicine **26**(11), 1714–1719 (2020). DOI 10.1038/s41591-020-1092-0
3. Agusto, F.B., Erovenko, I.V., Fulk, A., Abu-Saymeh, Q., Romero-Alvarez, D., Ponce, J., Sindi, S., Ortega, O., Saint Onge, J.M., Peterson, A.T.: To isolate or not to isolate: The impact of changing behavior on COVID-19 transmission. medRxiv (2020). DOI 10.1101/2020.08.30.20184804
4. Allain-Dupre, D., Chatry, I., Kornprobst, A., Michalun, M.V.: The territorial impact of COVID-19: Managing the crisis across levels of government. Tech. rep., OECD (2020)
5. Althouse, B.M., Wenger, E.A., Miller, J.C., Scarpino, S.V., Allard, A., Hebert-Dufresne, L., Hu, H.: Superspreading events in the transmission dynamics of SARS-CoV-2: Opportunities for interventions and control. PLoS Biology **18**(11), 1–13 (2020). DOI 10.1371/journal.pbio.3000897
6. Anne Cori Neil M. Ferguson, C.F., Cauchemez, S.: A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. American Journal of Epidemiology **178**(9) (2013). DOI 10.1093/aje/kwt133
7. Askitas, N., Tatsiramos, K., Verheyden, B.: Estimating worldwide effects of non-pharmaceutical interventions on COVID-19 incidence and population mobility patterns using a multiple-event study. Scientific Reports **11**(1), 1–13 (2021). DOI 10.1038/s41598-021-81442-x
8. Bubar, K.M., Kissler, S.M., Lipsitch, M., Cobey, S., Grad, Y.H., Larremore, D.B., Reinholt, K.: Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. Science **371**(February), 916–921 (2020). DOI 10.1101/2020.09.08.20190629
9. Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D., Leskovec, J.: Mobility network models of COVID-19 explain inequities and inform reopening. Nature **589**(7840), 82–87 (2021). DOI 10.1038/s41586-020-2923-3
10. Chitwood, M.H., Russi, M., Gunasekera, K., Havumaki, J., Pitzer, V.E., Warren, J.L., Weinberger, D.M., Cohen, T., Menzies, N.: Bayesian nowcasting with adjustment for delayed and incomplete reporting to estimate COVID-19 infections in the United States. medRxiv (2020). DOI 10.1101/2020.06.17.20133983
11. Christen, J.A., Fox, C.: A general purpose sampling algorithm for continuous distributions (the t-walk). Bayesian Analysis **5**(2), 263–281 (2010). DOI 10.1214/10-BA603

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

12. Dehning, J., Zierenberg, J., Spitner, F., Wibral, M., Neto, J., Wilczek, M., Priesemann, V.: Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science **369**(6500), 1–15 (2020). DOI 10.1126/science.abb9789
13. Fritsch, F.N., Carlson, R.E.: Monotone piecewise cubic interpolation. SIAM Journal on Numerical Analysis **17**(2), 238–246 (1980). DOI 10.1137/0717021. URL https://doi.org/10.1137/0717021
14. INSP: Recomendaciones. In: Reflexiones sobre la respuesta de Mexico ante la pandemia de COVID19 y sugerencias para enfrentar los próximos retos. Instituto Nacional de Salud Pública, Instituto Nacional de Salud Pública, Mexico, Mexico City (2020)
15. Kermack, W., McKendrick, A.: A Contribution to the Mathematical Theory of Epidemics. Proceedings of the Royal Society A **115**(772) (1927). DOI 10.1098/rspa.1927.0118
16. Komarova, N.L., Schang, L.M., Wodarz, D.: Patterns of the covid-19 pandemic spread around the world: exponential versus power laws. Journal of the Royal Society Interface **17**(170), 20200518 (2020). DOI 10.1098/rsif.2020.0518
17. Li, Y., Campbell, H., Kulkarni, D., Harpur, A., Nundy, M., Wang, X., Nair, H., for COVID, U.N., et al.: The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. The Lancet Infectious Diseases **21**(2), 193–202 (2021). DOI 10.1016/S1473-3099(20)30785-4
18. Liu, Y., Eggo, R.M., Kucharski, A.J.: Secondary attack rate and superspreading events for sars-cov-2. The Lancet **395**(10227), e47 (2020). DOI 10.1016/S0140-6736(20)30462-1
19. Lloyd-Smith, J.O., Schreiber, S.J., Kopp, P.E., Getz, W.M.: Superspreading and the effect of individual variation on disease emergence. Nature **438**(7066), 355–359 (2005). DOI 10.1038/nature04153
20. Mathieu, E., Ritchie, H., Ortiz-Ospina, E., Roser, M., Hasell, J., Appel, C., Giattino, C., Rodés-Guirao, L.: A global database of COVID-19 vaccinations. Nature Human Behaviour (2021). DOI 10.1038/s41562-021-01122-8
21. May, B.: Research Briefing — Global World: GDP revisions. Tech. rep., Oxford Economics, Oxford (2021)
22. McGough, S.F., Johansson, M.A., Lipsitch, M., Menzies, N.A.: Nowcasting by Bayesian smoothing: A flexible, generalizable model for real-time epidemic tracking. PLoS Computational Biology **16**(4), 1–20 (2020). DOI 10.1371/journal.pcbi.1007735
23. Moore, S., Hill, E.M., Tildesley, M.J., Dyson, L., Keeling, M.J.: Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study. The Lancet Infectious Diseases (2021). DOI 10.1016/s1473-3099(21)00143-2
24. Popa, A., Genger, J.W., Nicholson, M.D., Penz, T., Schmid, D., Aberle, S.W., Agerer, B., Lercher, A., Endler, L., Colaço, H., Smyth, M., Schuster, M., Grau, M.L., Martínez-Jiménez, F., Pich, O., Borena, W., Pawelka, E., Keszei, Z., Senekowitsch, M., Laine, J., Aberle, J.H., Redlberger-Fritz, M., Karolyi, M., Zoufaly, A., Maritschnik, S., Borkovec, M., Hufnagl, P., Nairz, M., Weiss, G., Wolfinger, M.T., von Laer, D., Superti-Furga, G., Lopez-Bigas, N., Puchhammer-Stöckl, E., Allerberger, F., Michor, F., Bock, C., Bergthaler, A.: Genomic epidemiology of superspreading events in Austria reveals mutational dynamics and transmission properties of SARS-CoV-2. Science Translational Medicine **12**(573) (2020). DOI 10.1126/scitranslmed.abe2555
25. Rader, B., Scarpino, S.V., Nande, A., Hill, A.L., Adlam, B., Reiner, R.C., Pigott, D.M., Gutierrez, B., Zarebski, A.E., Shrestha, M., et al.: Crowding and the shape of covid-19 epidemics. Nature medicine **26**(12), 1829–1834 (2020). DOI 10.1038/s41591-020-1104-0
26. Santana-Cibrian, M., Acuña-Zegarra, M.A., Velasco-Hernandez, J.X.: Lifting mobility restrictions and the effect of superspreading events on the short-term dynamics of COVID-19. Mathematical Biosciences and Engineering **17**(5), 6240–6258 (2020). DOI 10.3934/mbe.2020330
27. Singer, H.M.: The COVID-19 pandemic: growth patterns, power law scaling, and saturation. Physical biology **17**(5), 055001 (2020). DOI 10.1088/1478-3975/ab9bf5
28. Subramanian, R., He, Q., Pascual, M.: Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology and testing capacity. PNAS **118**(9), e2019716118 (2020). DOI 10.1073/pnas.2019716118
29. Sun, K., Wang, W., Gao, L., Wang, Y., Luo, K., Ren, L., Zhan, Z., Chen, X., Zhao, S., Huang, Y., Sun, Q., Liu, Z., Litvinova, M., Vespignani, A., Ajelli, M., Viboud, C., Yu, H.: Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. Science **371**(6526), eabe2424 (2021). DOI 10.1126/science.abe2424

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

30. Thompson, R., Stockwin, J., van Gaalen, R.D., Polonsky, J., Kamvar, Z., Demarsh, P., Dahlqwist, E., Li, S., Miguel, E., Jombart, T., et al.: Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics **29**, 100356 (2019). DOI 10.1016/j.epidem.2019.100356
31. Thurner, S., Klimek, P., Hanel, R.: A network-based explanation of why most COVID-19 infection curves are linear. Proceedings of the National Academy of Sciences of the United States of America **1**(37), 22684–22689 (2020). DOI 10.1073/pnas.2010398117
32. Wong, F., Collins, J.J.: Evidence that coronavirus superspreading is fat-tailed. Proceedings of the National Academy of Sciences p. 202018490 (2020). DOI 10.1073/pnas.2018490117
33. Wu, J.T., Leung, K., Leung, G.M.: Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan , China : a modelling study. The Lancet **395**(20), 689–697 (2020). DOI 10.1016/S0140-6736(20)30260-9
34. Zhang, Y., Li, Y., Wang, L., Li, M., Zhou, X.: Evaluating transmission heterogeneity and super-spreading event of COVID-19 in a metropolis of China. International Journal of Environmental Research and Public Health **17**(10) (2020). DOI 10.3390/ijerph17103705

A Statistical inference

In order to create predictions for the evolution of the COVID-19 pandemic during 2021, we focus on the estimation of the time dependent contact rate $\beta(t)$ during 2020 and the start of 2021 as was described in Section 3. To simplify the estimation process, $\beta(t)$ is assumed to be a piecewise Hermite interpolating polinomial that only changes at preset times $t_1, t_2, \ldots, t_k$, and it is constant before t_1. Then, instead of estimating a continuous function, we only need to estimate the values of the effective contact rate $b_1, b_2, \ldots, b_k$ at $t_1, t_2, \ldots, t_k$.

The dates where the effective contact rate changes are described in Table 2. Those dates are converted into a numeric scale to get $t_1, t_2, \ldots, t_k$ simply by calculating the number of days from the first reported case (by symptoms onset). For each Mexican state, the starting date for the analysis is different.

The initial number of exposed (E_0), asymptomatic (I_0) and symptomatic (Y_0) individuals will also be estimated as these are important unknown quantities. Parameters α, μ and ν depend on the state analyzed. Parameter $1/\mu$ is directly estimated from the records of confirmed cases by calculating the average difference in days between the date of symptoms onset and the date of death, while $1/\nu$ is calculated as the average difference in days between the date of symptoms onset and the date of hospital registration. Only parameter α is included in the inference process.

Let $\boldsymbol{\theta} = (E_0, I_0, Y_0, \alpha, b_1, b_2, \ldots, b_k)$ the vector of parameters that will be estimated using a Bayesian inference approach. All the other parameters needed to solve model (1) are fixed and their values can be found in Table 1.

Let X_j and Y_j be the random variables that count the number of daily COVID-19 reported cases and deaths at time t_j, respectively, for $j = 1, 2, ...n$, where t_j represent the number of days since the first reported case by symptoms onset. We assume that the probability distribution of X_j and Y_j, conditional on the vector of parameters $\boldsymbol{\theta}$, is a Poisson distribution such that

$$E[X_j] = \mu_X(t_j|\boldsymbol{\theta}) = A(t_j|\boldsymbol{\theta}) - A(t_{j-1}|\boldsymbol{\theta}), \tag{4}$$
$$E[Y_j] = \mu_Y(t_j|\boldsymbol{\theta}) = D(t_j|\boldsymbol{\theta}) - D(t_{j-1}|\boldsymbol{\theta}), \tag{5}$$

with $A(t|\boldsymbol{\theta})$ ad $D(t|\boldsymbol{\theta})$ being the cumulative number of reported cases and deaths according to model (1). Assuming that variables $X_1, X_2, \ldots, X_n, Y_1, \ldots, Y_n$ are conditionally independent, then the likelihood function is given by

$$\pi(x_1, \ldots, x_n, y_1, \ldots, y_n|\boldsymbol{\theta}) = \prod_{j=1}^{n} \frac{\mu_X(t_j|\boldsymbol{\theta})^{x_j} \exp\{-\mu_X(t_j|\boldsymbol{\theta})\}}{x_j!} \frac{\mu_Y(t_j|\boldsymbol{\theta})^{y_j} \exp\{-\mu_Y(t_j|\boldsymbol{\theta})\}}{y_j!}. \tag{6}$$

The joint prior distribution for vector $\boldsymbol{\theta}$ is a product of independent probability distributions. For all the initial conditions E_0, I_0, Y_0, the prior distribution is Uniform(0,20) and for all the

medRxiv preprint doi: https://doi.org/10.1101/2021.04.14.21255436; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

contact rates $b_1, \ldots, b_k$ the prior is Uniform(0,5). The prior for α is a Beta distribution with parameters $c_1 = 20.0$ and $c_2 = 2.2$ which has an expected value of 0.9. Then

$$\pi(\boldsymbol{\theta}) = \pi(E_0)\pi(I_0)\pi(Y_0)\prod_{i=1}^{k}\pi(b_i) \propto \alpha^{c_1-1}(1-\alpha)^{c_2-1}.$$

The posterior distribution of the parameters of interest is

$$\pi(\boldsymbol{\theta}|y_1, \ldots, y_n) \propto \pi(y_1, \ldots, y_n|\boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

and it does not have an analytical form since the likelihood function depends on the numerical solution of the ODE system (1). We analyze the posterior distribution using an MCMC algorithm called *t-walk* [11]. For each state, three chains of 1,000,000 iterations are run, from which 100,000 are discarded as burn in. At the end, only 2000 iterations are used to create the estimations presented in this work. The posterior predictive distribution is used to create the scenarios for the rest of 2021 under the three specific values of the vaccination rate ϕ.