

The turning point and end of an expanding epidemic cannot be precisely forecast

Mario Castro^{a,b}, Saúl Ares^{a,c}, José A. Cuesta^{a,d,e,f}, and Susanna Manrubia^{a,c,1}

^aGrupo Interdisciplinar de Sistemas Complejos, 28911 Madrid, Spain; ^bInstituto de Investigación Tecnológica, Universidad Pontificia Comillas, 28015 Madrid, Spain; ^cDepartamento de Biología de Sistemas, Centro Nacional de Biotecnología, 28049 Madrid, Spain; ^dDepartamento de Matemáticas, Universidad Carlos III de Madrid, 28911 Leganes, Spain; ^eInstituto de Biocomputación y Física de Sistemas Complejos, Campus Río Ebro, Universidad de Zaragoza, 50018 Zaragoza, Spain; and ^fUniversidad Carlos III de Madrid–Santander Big Data Institute, 28903 Getafe, Spain

Edited by Eugene V. Koonin, National Institutes of Health, Bethesda, MD, and approved September 11, 2020 (received for review April 23, 2020)

Epidemic spread is characterized by exponentially growing dynamics, which are intrinsically unpredictable. The time at which the growth in the number of infected individuals halts and starts decreasing cannot be calculated with certainty before the turning point is actually attained; neither can the end of the epidemic after the turning point. A susceptible–infected–removed (SIR) model with confinement (SCIR) illustrates how lockdown measures inhibit infection spread only above a threshold that we calculate. The existence of that threshold has major effects in predictability: A Bayesian fit to the COVID-19 pandemic in Spain shows that a slowdown in the number of newly infected individuals during the expansion phase allows one to infer neither the precise position of the maximum nor whether the measures taken will bring the propagation to the inhibition regime. There is a short horizon for reliable prediction, followed by a dispersion of the possible trajectories that grows extremely fast. The impossibility to predict in the midterm is not due to wrong or incomplete data, since it persists in error-free, synthetically produced datasets and does not necessarily improve by using larger datasets. Our study warns against precise forecasts of the evolution of epidemics based on mean-field, effective, or phenomenological models and supports that only probabilities of different outcomes can be confidently given.

predictability | epidemics | forecast | Bayesian

In 1972, Edward Norton Lorenz delivered a legendary-by-now talk titled *Predictability: Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas?* (1). Lorenz had stumbled onto chaos and uncovered its major consequence for weather prediction. By means of a simple model (2), he had shown that one can never be certain of whether, one week from now, we will have a sunny or a rainy day. Half a century later, we are used to listening to the weather forecast in terms of percentages, probability of rain, intervals for temperature and wind speed, and so on. It is just fuzzy information, but usually is sufficient to make up our minds on what to do next weekend. The key point is that, as far as weather is concerned, we accept that we are bound to cope with uncertainty. The mechanism behind that uncertainty has to do with the exponential amplification of small initial differences prototypical of chaotic systems. It turns out that other systems with exponentially growing variables also display an analogous behavior: They are sensitive to small variations in parameters and amplify small differences, potentially leading to quantitatively and qualitatively different outcomes. Though their dynamics are not chaotic, this is the case of epidemics.

The worldwide ongoing COVID-19 pandemic is triggering multiple attempts at modeling the progression and immediate future of epidemic spread (3–8). Many of the formal approaches used are based on simple, mean-field compartmental models (9, 10) with a different number of classes for the individuals in a population: susceptible (S), infected asymptomatic (E), infected symptomatic (I), recovered (R), dead (D), and several other possible intermediate stages, such as quarantined, hospitalized, or at the intensive care unit (ICU). Beyond their clear interpreta-

tion and ease of use, a main motivation to apply such models relies on trying to estimate the forthcoming stages of the epidemic and on quantifying the effects of nonpharmaceutical measures toward “flattening the curve,” “reaching the peak,” estimating the total number of infected people when the epidemic ends, or controlling the number of ICUs required at a time from now.

Susceptible–infected–removed (SIR)-like models, therefore, are not only employed to drive intuitions and expectations, but are also applied to derive quantitative predictions. Further, the family of compartmental models lies at the basis of more sophisticated attempts to numerically describe the current spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the effect of contention measures (8, 11), where they are mostly used at the local level (12–14).

Here, we consider a variant of an SIR model with reversible confinement of susceptible individuals that we call SIR model with confinement (SCIR). But before we get into further details, we would like to clarify what this paper is not: It is *not* another paper with a simple model aiming to predict the evolution of the epidemic. On the contrary, we intend to show that predicting with models like this one is severely limited by strong instabilities with respect to parameter values. The reason we work with such a minimal model is that we can obtain analytical expressions for the dynamics under sensible approximations and make the point more clear. Also, it allows us to derive simple facts about the epidemics, such as the existence of a threshold that separates

Significance

Susceptible–infected–removed (SIR) models and their extensions are widely used to describe the dynamics of infection spreading. Certain generic features of epidemics are well-illustrated by these models, which can be remarkably good at reproducing empirical data through suitably chosen parameters. However, this does not assure a good job anticipating the forthcoming stages of the process. To illustrate this point, we accurately describe the propagation of COVID-19 in Spain using one such model and show that predictions for its subsequent evolution are disparate, even contradictory. The future of ongoing epidemics is so sensitive to parameter values that predictions are only meaningful within a narrow time window and in probabilistic terms, much as what we are used to in weather forecasts.

Author contributions: M.C., S.A., J.A.C., and S.M. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

See [online](#) for related content such as Commentaries.

¹To whom correspondence may be addressed. Email: smanrubia@cnb.csic.es.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2007868117/-DCSupplemental>.

First published October 1, 2020.

“mild” confinement measures, causing mitigation from stronger measures, leading to the inhibition of infection propagation.

The parameters of the model can be estimated within a relatively narrow range using data available from the COVID-19 pandemic. Yet, unavoidable uncertainties in those parameters, which determine the time at which growth is halted or the overall duration of the pandemic, propagate to the predicted trajectories, preventing reliable prediction of the intermediate and late stages of epidemic spread. This is the main message of this article, because it transcends the model of choice.

Attempts similar to the one performed here are common these days, and, often, strong predictions regarding the number of casualties, the position of the peak, or the duration of the epidemic are drawn. Our model does an excellent job in reproducing past data, but, instead of taking most likely parameter values (or empirically evaluated values) to draw a prediction, we estimate compatible ranges of variations in the parameters. Especially when the process is close to the mitigation-inhibition threshold, predictions of the next few days become extremely sensitive to changes in the parameters and to the addition of subsequent empirical data. Altogether, it turns out that quantitative predictions made in any similar framework are not reliable if not accompanied by their likelihood. The main conclusion we reach is that the deterministic nature of SIR-like models is misleading if aimed at describing the actual course of any pandemic: Prediction of the past is achieved through suitable fitting of data, and different functions may work, but prediction of the future in the midterm cannot be trusted.

SCIR: An SIR Model with Confinement

The SCIR model includes the usual states of an SIR model plus a class C for individuals sent to confinement that are susceptible, but not infected (Fig. 1).

In a sufficiently large population, the number of infected individuals at the initial stages of the infection is well below the population size. Under certain conditions, it may stay small in comparison to the number of susceptible individuals remaining. This seems to be the case in July 2020 for most countries in the time elapsed since COVID-19 started to spread (15, 16). If we assume that $I(t)/N \ll 1$, then we can neglect the nonlinear term in the equation for the number of susceptible individuals and solve the model analytically (*Materials and Methods* and *SI Appendix, section A*). Within this approximation, the number of infected individuals at time t is given by

$$I(t) = I_0 e^{[R_0^*(t) - 1](r + \mu)t}, \quad [1]$$

where

$$R_0^*(t) = \frac{R_0}{q + p} \left[p + q \frac{1 - e^{-(q+p)t}}{(q+p)t} \right], \quad R_0 \equiv \frac{\beta}{r + \mu}, \quad [2]$$

is the effective basic reproduction number modulated by the confinement— R_0 being its value at the beginning of the epidemic. All of the behavior of the epidemic is enclosed in this magnitude. At its initial stages, $I(t) \sim \exp[(R_0 - 1)(r + \mu)t]$, so the epidemic spreads when $R_0 > 1$ (as is the case of COVID-19), and the larger R_0 , the faster it does. When confinement sets in, $R_0^*(t)$ gets tamed, eventually dropping to the value $R_0^*(\infty) = R_0 p / (q + p)$. An important epidemiological message follows from this simple fact: Only if the confinement is strong enough (p and q are sufficiently different so that $R_0^*(\infty) < 1$) can the epidemic be controlled; otherwise, it spreads until eventually decaying due to the standard SIR mechanism—the exhaustion of susceptible individuals.

Another interesting result follows from this simple model. Beyond the threshold for inhibition of infection spread, Eq. 2 captures the transient subexponential growth in the number

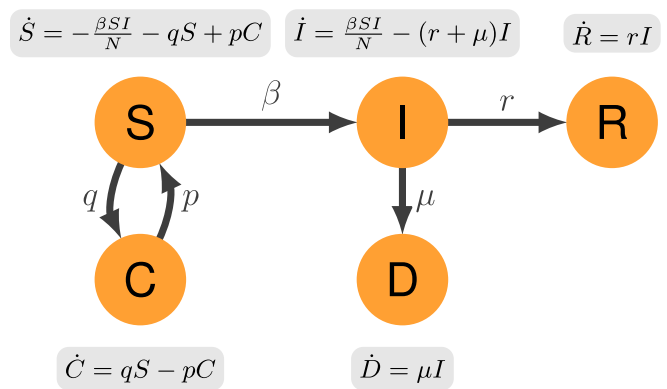


Fig. 1. Diagram of the epidemic model along with the equations ruling the dynamics. Susceptible individuals (S) can enter and exit confinement (C) or become infected (I). Infected individuals can recover (R) or die (D). N is the total population. Rates for each process are displayed in the figure; q depends on specific measures restricting mobility and contacts, while p stands for individuals that leave the confinement measures (e.g., people working at essential jobs like food supply, health care, or policing), as well as for defection. We fit I to data on officially diagnosed cases, which are automatically quarantined: The underlying assumption is that the real, mostly undetected, number of infections is proportional to the diagnosed cases.

of infected individuals. As mentioned above, if global confinement is suppressed ($q = p = 0$), $I(t)$ grows exponentially at a rate $(R_0 - 1)(r + \mu)$. As confinement is turned on, $I(t)$ displays a systematic bending that, for long enough time, will lead to a second exponential regime characterized by the rate $[R_0^*(\infty) - 1](r + \mu)$ —which can be positive or negative depending on the confinement parameters. The bending of the curve is observed in both scenarios, so it cannot be taken as a sign that the epidemic will be eventually controlled.

Fitting COVID-19 data for Spain

In order to illustrate the suitability of our model to reproduce available data, we have used official daily records reported by the Spanish Ministry of Health for all Spanish Autonomous regions since February 28th (17). Strict lockdown permitting only essential trips outside the home was applied on March 14th. However, school and university closure took place on March 11th, so we take this date as the starting point of the confinement. The measure was extended on March 30th to the closure for two weeks of all businesses and companies not providing key services. Between these dates, the data span two different regimes: unconstrained propagation of the epidemic, with $q = p = 0$, and a lockdown phase with effective parameters for the transition to the confined state. Since separated data for the number of recovered and deaths was unreliable, we have merged these two compartments and jointly fitted $r + \mu$, which become a single variable in practice.

We used a Bayesian approach to fit the data, assuming that the numbers of infected and recovered + dead are log-normally distributed with unknown variance and mean given by the expression for $I(t)$ obtained from the model (see *Materials and Methods* for details). At the very early stages of the epidemic (before any recovery or death event), the total number of confirmed cases grows as $e^{\beta t}$ independently of the chosen model (SIR, susceptible-exposed-infectious-recovered [SEIR], etc.). Analyzing this initial growth for every country in the world, it appears that $\beta < 1$ everywhere [doubling times larger than one day are reported in all cases (18)]. Thus, we use informative priors for β and $r + \mu$ (uniform distributions from 0 to 1 days⁻¹) and vague priors for the rates q and p (uniform distributions from 0 to 5 days⁻¹). Also, we use noninformative priors for the variances.

Uncertainty on Peak Occurrence Fitting Prepeak Data. The results of fitting real-time data until March 29th are summarized in Fig. 2. Fig. 2A illustrates the fit to our analytical solution for the aggregated data of all Spanish Autonomous regions, representing country-level progression. Symbols are reported data, and the solid line represents the median of the distribution. Interestingly, quantiles 2.5% and 97.5% provide almost opposite conclusions: Either the epidemic curve “flattens,” or it keeps growing exponentially, albeit at a different rate. This is a consequence of the inherent variability of the fitted parameters—as summarized by their posterior distributions (*SI Appendix*, Fig. S1)—and the exponential character of the epidemic. Similar conclusions can be drawn by inspection of the number of new deaths and recovered cases, $\Delta D + \Delta R$ (*SI Appendix*, Fig. S2). For completeness, we have also considered less realistic assumptions for the prior distributions and show that they lead to less consistent predictions. The obtained fits and posterior distributions are represented in *SI Appendix*, Figs. S3 and S4, respectively.

The systematic bending of the curve (Fig. 2A), due to confinement in the framework of our model, does not guarantee that the epidemic is under control—hence, this information alone can be misleading in interpreting the effects of the measures applied. To emphasize this conclusion, we compute the posterior distribution of the time when the peak of the epidemic occurs. Analytically,

$$t_{\max} = \frac{1}{p+q} \log \left(\frac{\beta q}{(r+\mu)(p+q) - \beta p} \right), \quad [3]$$

which, of course, is only meaningful when the epidemic gets eventually controlled by the confinement measures (i.e., if $(r+\mu)(p+q) > \beta p$). With parameter values inferred from Fig. 2A, confinement measures succeed at inhibiting the epidemic—which is the effect sought—only in 26% of cases, while in 74% of cases, they fail at inhibiting its expansion and only slow it down. Fig. 2B displays the distribution of the day in which the epidemic reaches the maximum, conditional on it actually occurring.

We have also fitted the model to each Spanish Autonomous region and have obtained analogous overall conclusions. Fits to reported data can be seen in *SI Appendix*, Fig. S5; all of them illustrate the goodness of our simple SCIR model at fitting past data. Posterior distributions yield comparable parameter values, within their intervals of definition, though there is significant variability in the percent of trajectories compatible with inhibition of propagation in each Autonomous region. First, the closer a region is to the inhibition threshold, the larger the dispersion of

the forecast. In such Autonomous regions, the epidemic started sooner than in regions that, by the end of March, were clearly in an expanding phase. Secondly, there are multiple regions with a vanishing probability of inhibition under the conditions of the first lockdown applied.

Uncertainty of Epidemic End, Including Postpeak Data. The peak of the epidemic in Spain was actually attained around April 18th. In light of the broad distribution of times returned by fitting prepeak data, and the forecast of only a 26% probability of having a peak, the application of more stringent confinement measures on March 30th seems justified and may have played an important role in the inhibition of COVID-19 propagation. Once the peak has been overcome, however, the uncertainty regarding the end of the epidemic remains high in the midterm. Fig. 3 displays fits of data and forecasts just when the peak was attained and until May 9th using, as above, the analytical result in Eq. 1. The model is able to fit well past data and yields apparently narrower error intervals, as compared to prepeak fits. However, data deviate from the most likely forecast about two weeks into the future. Unfortunately, the publication of data on the number of recovered cases was interrupted on May 17th, making it impossible to know the number of active cases, so we cannot extend the empirical series to the time of writing. As of July 2020, Spain was experiencing a long plateau with a sustained average number of new daily cases well captured by the slow decrease in the $I(t)$ curve. However, even if the progression of the disease were well described by Fig. 3B, the distribution of times at which the epidemic would end is very broad. Fig. 3C shows that there is an uncertainty of around three months in the time needed for the number of confirmed cases to drop below 1,000—hardly an accurate prediction of the epidemic’s end.

Better Data Are Necessary, but Not Sufficient. There are major difficulties in predicting the future of an epidemic with the present or similar models. First, incomplete or noisy data entail uncertain predictions, as can be seen in fits to all Spanish Autonomous regions (*SI Appendix*, Fig. S5). But, even if data were complete and precise, small variations in the parameters bring about growing uncertainties as time elapses. In order to illustrate this point, we have generated a synthetic set of observations through direct integration of the system described in Fig. 1. By construction, uncertainties in the prediction of future trends derived through the Bayesian approach can only be ascribed to the dispersion of posterior distributions (shown in *SI Appendix*, Fig. S8).

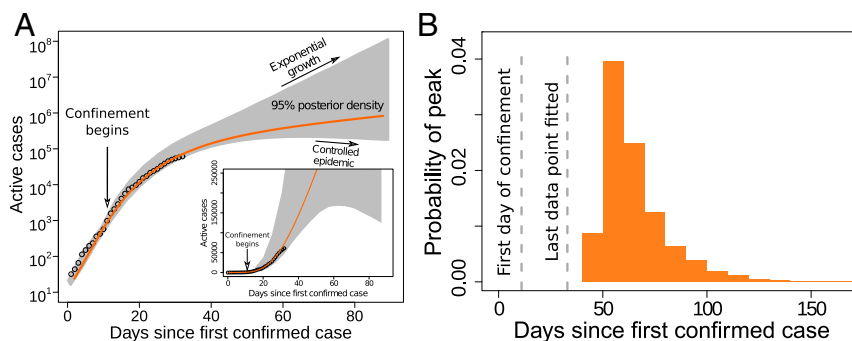


Fig. 2. Fit to data obtained in real time for the daily number of active cases in Spain (from March 1st to March 29th) and peak forecast. (A) Despite the reasonable agreement between model and empirical observations in the growing phase, opposite predictions for the future number of active cases can be derived. The solid line represents the expression for $I(t)$ using the median parameters for each posterior in *SI Appendix*, Fig. S1. The vertical arrow denotes March 11th, the day when schools and universities closed. The shaded area represents the 95% predictive posterior interval: Its increasing width implies that predictability decays exponentially fast. (A, Inset) Same data and curves with linear vertical scale. *SI Appendix*, Figs. S6 and S7 show how this fit and its posteriors evolve as an increasing number of days is included in the fit. An animation is included as *Movie S1*. (B) Posterior distribution of the time to reach the peak of the epidemic, conditioned to actually having a peak (which occurs with probability 0.26). The vertical dashed lines stand for the days when the confinement began and for the date of the last data point used in the fit.

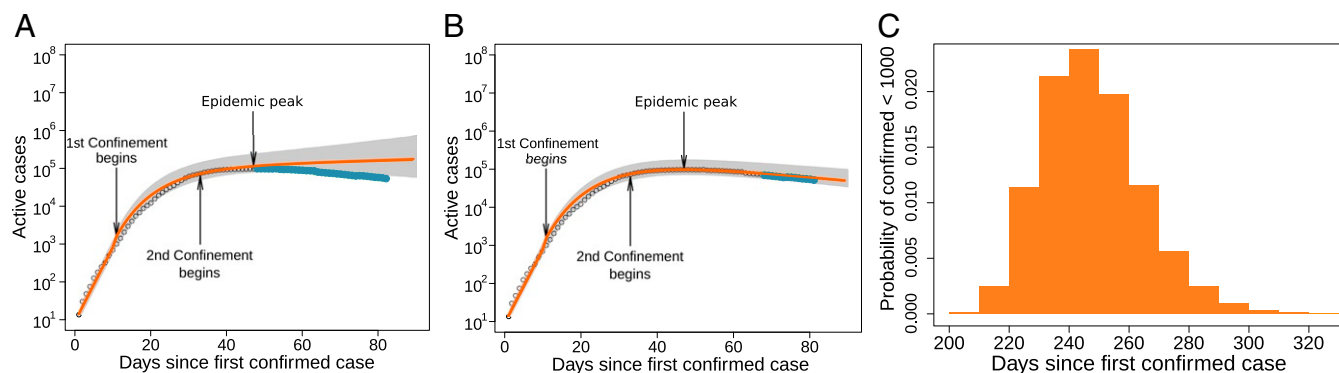


Fig. 3. Fit to postpeak data for the daily number of active cases in Spain. (A) Fit to data up to April 18th (peak day). (B) Fit to data at three weeks postpeak (May 9th). Open symbols represent fitted empirical data, and blue dots correspond to actual measurements until May 17th. (C) Distribution of times until the number of confirmed cases falls below 1,000 for the first time. With about two cases per million inhabitants, this threshold can define the end of the epidemic. The distribution spans about three months, centered around the end of October 2020.

Fig. 4 summarizes the inherent limitations of our model (and any other SIR-like model) to forecast any quantity beyond a limited temporal threshold. Prediction is severely affected by small differences in the fitting parameters, yielding a fan of compatible trajectories and limiting the reliability of the forecast. To complicate things further, the uncertainty of forecasts is not a monotonous function of the number of data points used in the model fits. *SI Appendix, Figs. S9 and S10* illustrate the breadth of the 95% posterior density as a function of the number of data points used, both for empirical and synthetic data. They show that more data do not necessarily entail less uncertainty and improved predictability.

Discussion

Confinement and Turning Points. The implementation of confinement measures to control the expansion of highly transmissible pathogens affects the speed of infection propagation, as measured in the number of newly infected, recovered, or deceased individuals. Confinement bends the progression curve downward, but this bending, which can span a remarkable lapse of time, should not be interpreted as an unequivocal sign that propagation is to be inhibited. Rather, it might represent just a transient, cross-over regime to a new diverging, exponential phase, albeit with a different coefficient. In the simple SCIR model discussed here, these two regimes are clearly identified. The initial growth, before confinement starts, occurs at a rate $(R_0 - 1)(r + \mu)$, which depends on intrinsic properties of the pathogen–host interaction and on contacts between hosts. Sufficiently severe infections, with $R_0 > 1$, cause a pandemic if not controlled. The onset of confinement modifies the long-time trend of the infection by defining a new coefficient for the asymptotically dominating exponential, $(R_0 p(q + p)^{-1} - 1)(r + \mu)$, which includes two important factors: the strength of confinement measures, q , and the lack of adherence of individuals to confinement, p . If $R_0 p(q + p)^{-1} > 1$, the growth slows down before a new asymptotic phase of exponential growth sets in. This phase corresponds to mitigation of infection propagation, but eventual extinction will occur through the usual SIR mechanism of exhaustion of susceptible individuals, while the deceleration observed in between is just a cross-over between two exponentially diverging regimes. If the proportion of infected becomes high enough during the cross-over, our approximation is no longer valid, and the SIR mechanism can kick in, making the second exponential regime unobservable in practice.

Inhibition of infection propagation is achieved only if $R_0 p(q + p)^{-1} < 1$, where a limited fraction of the population—which depends on the confinement strength and collective adherence—

will get the disease. Though the model we have studied here is simple enough so as to allow us to derive exact results and to characterize the nature of the turning point (or cross-over), the qualitative scenario should be shared by any other compartmental models with growing and decreasing phases dominated by exponential behavior and able to implement the effects of confinement.

Phenomenological models do a good job in reproducing the temporal track of single outbreaks that result either from the unconstrained SIR mechanism of exhaustion of susceptibles or from a sustained inhibition of propagation down to extinction. However, models such as the generalized logistic (19) or Gompertz (20) growth curves cannot capture some of the important qualitative behavior in the SCIR class. Those models cannot reproduce long plateaus, such as those extending for months in

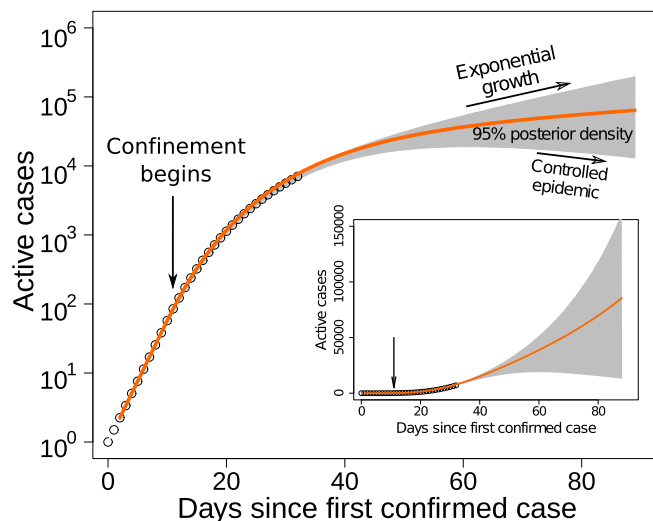


Fig. 4. Data generated through direct simulation of the system described in Fig. 1 are used as input to determine posterior distributions for parameters through a Bayesian approach. Parameter values are $\beta = 0.425$, $p = 0.007$, $q = 0.062$, and $r + \mu = 0.021$ in the mitigation regime taken from the median of the posteriors in *SI Appendix, Fig. S1* (all measured in day^{-1}). Though the dataset is complete and noiseless, consideration of only the growing phase of the epidemic implies a remarkable uncertainty in compatible trajectories. It is worth noting that, albeit those parameters would predict that the epidemic is not controlled, variability still leaves a 3% chance that it actually is. (Inset) Same data and curves with linear vertical scale.

Spain (*SI Appendix*, Fig. S11) or Italy. Phenomenological models do not have instability points such as a mitigation-inhibition threshold, and, for that reason, they cannot embrace the possibility that propagation is slowed down, but not halted (*SI Appendix*). The identification of such thresholds is essential to quantify the effectiveness of nonpharmaceutical interventions (21).

Related SIR-Like Models. There are several models in the literature conceptually analogous to the one described above. Obvious ones are SIR and SEIR (where the R state is understood as “removed” individuals and groups both recovered and dead individuals). In SEIR models, the E state permits to include the effect of a latent period, where individuals are infected but asymptomatic. Depending on the disease it mimics, individuals can be infectious or not. In general, the consideration of the E state brings about a delay in the completion of the disease, but does not entail qualitative changes.

Some models consider the effect of quarantined individuals. Susceptible–infectious–quarantined–recovered (SIQR) and susceptible–infected–quarantined–susceptible (SIQS) models have been introduced and studied earlier (22, 23), but confinement was not considered there. More recent models aiming at including the effects of confinement and quarantine, as applied in Italy and Spain to contain COVID-19—i.e., to class S —have generalized the classical SEIR model (24, 25) in a way different from ours. In one model (24), the possibility of individuals not committing themselves to the confinement (analogous to setting $p \rightarrow 0$ in our model) is discarded, while possible advances in treating the disease through a recovery rate that increases with time and a death rate that decreases with time are included. A variant of one such model was used to draw very precise predictions on the course of epidemics that had to be subsequently revised in the light of new data (26). In another case (25), nonsimultaneous ingoing and outgoing fluxes to the S compartment are considered, thus standing for strict confinement followed by free relaxation. A so-called SIRX model with irreversible quarantine has been shown to recover the systematic subexponential growth in the expansion phase (6). However, the quarantined class acts as an absorbing state and leads to an unrealistic feature of the model: At any strength, quarantine entails inhibition of infection propagation. This property is equivalent to the saturation implicit in phenomenological models, which cannot embrace mitigation of epidemic propagation without full inhibition (*SI Appendix*, section D). Finally, a model quite similar to ours has been proposed, though no results on its dynamics are available as of yet (27).

Still, other models increase the number of different states considered with the aim of becoming more realistic, especially motivated by specific observations of COVID-19. For example, the fact that only a fraction of the actually infected individuals is detected has been included in generalized mean-field models (5), or the different progress of the disease depending on the age group has been taken into account in models that consider stratified, age-structured populations with (11) or without (7) the consideration of physical location. Those models, by definition, have a significantly larger number of parameters. Again, the eventual aim of those models is to draw apparently soundly motivated and precise predictions on the time at which the pandemic will halt.

Compartmental models are appealing for at least two reasons: They are simple to formulate and offer a clear epidemic interpretation. But different models—as one can see by comparing the examples in this section—lead to different predictions. All models use either observational data on the progress of the epidemic or empirically evaluated parameters, or both. However, many of them lack a sensitivity analysis that propagates actual errors in data and parameters to their predictions. It is to be expected that, should they do so, most predictions might turn out to be compat-

ible with different models (with different effective parameters) within their intervals of confidence. These models might return some reliable, probabilistic forecasts in the near future, but only if current conditions for propagation (e.g., confinement measures or the average collective habits of the population) remain unaltered (8). The reconstruction of the Wuhan epidemic, for instance, clearly illustrates that changes in contention measures turn reliable prediction nearly impossible (28).

Effective Parameters and Identifiability. In principle, parameters characterizing the transitions between states in any SIR-like model are related to quantities amenable to empirical estimations. For instance, β quantifies the transmissibility of the virus; q should relate to the fraction of confined population and vary with different nonpharmaceutical measures put in place; p quantifies the adherence of population to confinement rules, and, thus, can be estimated through data on mobile phone location (29); and so on. However, it is important to emphasize that a direct empirical estimation of the parameters need not be the right value to input the model with. Any simple SIR-like model (and even more so any phenomenological model) is, by definition, leaving aside a number of realistic features. When fits of actual data are attempted, those models yield, at best, trajectories close to the actual one, but the inferred parameters are necessarily effective. Including or not the E state, confined individuals, age-structured populations, or any other possible level of detail redefines the precise meaning and values of the corresponding rates. Actually, an SIR model has been shown to reproduce better data from COVID-19 spread in Wuhan than an SEIR model (30), as well as to be “unreasonably effective” in describing the outbreaks in different countries affected by the pandemic (31). Moreover, the choice of which states and how many of them to include in the model is a subjective matter that changes dramatically the quality of the fit and, more importantly, the interpretability and identifiability of the parameters (32).

It could be argued that more realistic models are those with a larger number of states and parameters. But, at the same time, in those models, it is more difficult—often impossible—to assign a unique meaningful value to their parameters. If, rather than using the empirical estimates themselves, we attempt a noninformed fitting to the data, we normally end up in a problem of identifiability. That is, there are many different parameter sets (often a continuum of them) that fit the data, a problem that has been specifically analyzed for SIR-like models applied to COVID-19 (33). Though the use of empirical values from independent studies might anchor the values of a subset of parameters, the remaining degrees of freedom (parameters with values not fixed a priori) are often sufficient to reproduce data. In this sense, multiparametric models might lead to overfitting, losing at the same time explanatory power.

Probabilistic Forecasting. Deterministic epidemiological models convey a false impression of uniqueness of trajectories. It is broadly believed that a model able to reproduce empirical data well should be equally good at predicting future outcomes. This causal fallacy too often prevents a careful evaluation, through sensitivity analyses, of the effects of small variations in the parameters in forecasting. When such analyses are performed, results similar to ours should be obtained.

More detailed models using SIR-like descriptions for metapopulations and adding mobility usually incorporate uncertainty in their predictions. In this way, ranges for different quantities have been derived, such as for the basic reproductive number R_0 and its temporal response to confinement measures (34), for the number of undocumented infected individuals (14) or estimations of the effect of physical-distancing measures in the median number of infections (11).

At the other end of the spectrum, one finds statistical approaches that, in the absence of an underlying mechanistic model, fully rely on past data to predict the near future. Numerical approaches of this kind are intrinsically probabilistic and only yield likelihoods of different scenarios, with intervals of confidence that grow extremely fast as time elapses. A notable example of such approaches is the document elaborated by researchers at Imperial College (4), based on Bayesian estimations only informed by Europe-wide data of the COVID-19 pandemic. The results are compatible with multiple scenarios in most countries in the midterm, since the huge intervals of confidence inherent to their approach limit predictability to the near future. At the same time, that document likely yields the most trustworthy (probabilistic) predictions using prepeak data. The use of recent techniques that incorporate the possibility of obtaining closed mathematical expressions to Bayesian approaches (35) might bridge those techniques and the identification of underlying mechanistic models.

Conclusions

SIR-like models are unable to predict with certainty; at most, they can inform on the different likelihood of a variety of trajectories conditional on specific measures and parameters. Uncertainties in the values of the latter prevent a unique interpretation of the data at the transient. Near the threshold separating mitigation from inhibition, the same set of observations might be compatible with either future outcome. If the aim of control protocols is to minimize the total number of infected individuals and the duration of the confinement period, it seems advisable that the strongest possible measures are applied as early as possible. It has been documented that nonpharmaceutical interventions during the 1918 Flu Pandemic in the United States lowered mortality and mitigated adverse economic consequences (36). Deferral of such application is not justified on the basis of a slowdown of infection propagation.

Lorenz closed his 1972 talk (1) by stating the following:

“[Errors in weather forecasting] arise mainly from our failure to observe even the coarser structure with near completeness, our somewhat incomplete knowledge of the governing physical principles, and the inevitable approximations which must be introduced in formulating these principles as procedures which the human brain or the computer can carry out. These shortcomings cannot be entirely eliminated, but they can be greatly reduced by an expanded observing system and intensive research. It is to the ultimate purpose of making not exact forecasts, but the best forecasts which the atmosphere is willing to have us make that the Global Atmospheric Research Program is dedicated.”

Could COVID-19 trigger a Global Epidemic Research Program, intensive investigation in the topic, and an expanded observing system producing accurate and publicly available data? Only through such a program would it be possible to obtain the best forecasts that epidemic models might yield.

Materials and Methods

Data. We have used two slightly different datasets for studies prepeak and postpeak. The first one was downloaded in real time as the epidemic

progressed in Spain and contains information on number of new infected, dead, and recovered cases, as it was available by the April 1, 2020. The second dataset, used in the postpeak analysis, contains data updated in retrospect by the Spanish Ministry of Health, taking into account modifications in the official reporting introduced since April 2020. All data and codes used in this work, together with explanations on how to retrieve our fits, are publicly available at <https://github.com/mariocastro73/predictability> (17).

Bayesian Fit. We have fitted a parametric Bayesian model with the variables.

$$\begin{aligned} \log I(t) &\sim \mathcal{N}(\log I(t_0) + (\beta - \lambda)(t - t_0), \sigma_I), \quad t_0 < t \leq t_2, \\ \log I(t) &\sim \mathcal{N}(\log I(t_2) + [R_0^*(t - t_2) - 1]\lambda(t - t_2), \sigma_I), \quad t_2 < t, \\ \log X(t) &\sim \mathcal{N}(\log(r + \mu) + \log I(t), \sigma_X), \quad t_1 < t, \end{aligned} \quad [4]$$

where $X(t) = \Delta R(t) + \Delta D(t)$ stands for the change in the number of recovered plus dead cases daily reported. Similarly, we choose the following priors

$$\begin{aligned} \beta &\sim \mathcal{U}(0, 1), & 1/\sigma_I^2 &\sim \Gamma(0.01, 0.01), \\ r + \mu &\sim \mathcal{U}(0, 1), & 1/\sigma_D^2 &\sim \Gamma(0.01, 0.01), \\ p &\sim \mathcal{U}(0, 5), & 1/\sigma_X^2 &\sim \Gamma(0.01, 0.01), \\ q &\sim \mathcal{U}(0, 5), \end{aligned} \quad [5]$$

where \sim stands for *distributed as*, and \mathcal{N} , \mathcal{U} , and Γ stand for normal, uniform, and gamma distributions. For each Spanish Autonomous region, t_0 and t_1 stand for the days (since February 28th) where the first Infected and Recovered+Death cases were reported, respectively.

Note that we distinguish between the epidemic before ($t \leq t_2$) and after ($t > t_2$) any confinement measure was applied. Unlike other implementations of epidemic models (4), our model in Eq. 4 aims to fit directly the solution of the deterministic model in Eq. 1.

The priors for β and $r + \mu$ in Eq. 5 are informative priors derived from the fact that, in every country where cases of coronavirus were detected, the doubling period at the very early stages of the epidemic was never smaller than 2 d. Hence, we have taken for these parameters a prior $\mathcal{U}(0, 1)$. For the other parameters, we assume a noninformative prior $\mathcal{U}(0, 5)$ (we assume that changes faster than 1/5 d are meaningless in any compartmental model). The results are consistent with this assumption.

To profit from the largest amount of available data, we fit simultaneously the number of active cases (I) and the new number of cases in the interval $\Delta t = 1$ d of recovered (R) and dead (D) cases in logarithmic scale—a linear scale would give a biased larger weight to later times.

Data Availability. We have implemented the Bayesian model in R using the *rjags* wrapper of the *JAGS* library (37). The full code is provided in *SI Appendix, section B* and can be downloaded from GitHub, <https://github.com/mariocastro73/predictability> (17).

ACKNOWLEDGMENTS. We are indebted to Damián H. Zanette for his critical reading of a previous version of this work; and to Jacobo Aguirre, Yuxuan Cheng, Robert Endres, Javier Martín-Buldú, David J. Jörg, and Anxo Sánchez for their useful comments. This research has been funded by the Spanish Ministerio de Ciencia, Innovación y Universidades (MICINN)-Fondo Europeo de Desarrollo Regional funds of the European Union support, under Projects FIS2016-78883-C2-2-P and PID2019-106339GB-I00 (to M.C.), PGC2018-098186-B-I00 (to J.A.C.), FIS2017-89773-P (to S.M.), FIS2016-78313-P (to S.A.), and PID2019-109320GB-I00/AEI/10.13039/501100011033 (to S.A.). The Spanish MICINN has also funded the “Severo Ochoa” Centers of Excellence (to Centro Nacional de Biotecnología [CNB]) SEV 2017-0712 and Special Grant Proyecto Intramural Especial 2020-20E079 (to CNB, S.M. and S.A.) entitled “Development of protection strategies against SARS-CoV-2.”

1. E. N. Lorenz, “Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?” in AAAS 139th Meeting (1972). <http://eaps4.mit.edu/research/Lorenz/Butterfly.1972.pdf>. Accessed 7 April 2020.
2. E. N. Lorenz, Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141 (1963).
3. E. Estrada, COVID-19 and SARS-CoV-2: modeling the present, looking at the future. *Phys. Rep.* **869**, 1–51 (2020).
4. S. Flaxman et al., Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. *arXiv:2004.11342* (23 April 2020).
5. B. Ivorra, M. Ferrández, M. Vela-Pérez, A. Ramos, Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the

- undetected infections. The case of China. *Commun. Nonlinear Sci.* **88**, 105303 (2020).
6. B. F. Maier, D. Brockmann, Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science* **368**, 742–746 (2020).
7. V. M. Pérez-García, Relaxing quarantine after an epidemic: A mathematical study of the Spanish COVID-19 case. <https://doi.org/10.13140/RG.2.2.36674.73929/1> (April 2020).
8. G. N. Wong et al., Modeling COVID-19 dynamics in Illinois under non-pharmaceutical interventions. *medRxiv* 2020.06.03.20120691 (17 June 2020).
9. W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**, 700–721 (1927).

10. H. Hethcote, The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653 (2000).
11. K. Prem *et al.*, The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. *Lancet* **5**, e261–e270 (2020).
12. M. Gatto *et al.*, Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10484–10491 (2020).
13. A. Arenas *et al.*, A mathematical model for the spatiotemporal epidemic spreading of COVID-19. medRxiv:2020.03.21.20040022 (23 March 2020).
14. R. Li *et al.*, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **368**, 489–493 (2020).
15. M. Pollán *et al.*, Prevalence of SARS-CoV-2 in Spain (ENE-COVID): A nationwide, population-based seroepidemiological study. *Lancet* **396**, 535–544 (2020).
16. S. Stringhini *et al.*, Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): A population-based study. *Lancet* **396**, 313–319 (2020).
17. M. Castro, Datasets and codes from “The turning point and end of an expanding epidemic cannot be precisely forecast.” GitHub. <https://github.com/mariocastro73/predictability>. Accessed 27 July 2020.
18. Our World in Data. Datasets from the world in data (2020). <https://ourworldindata.org/coronavirus-source-data>. Accessed 12 April 2020.
19. K. Wu, D. Darcet, Q. Wang, D. Sornette, Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world. arXiv:2003.05681 (12 March 2020).
20. M. Levitt, A. Scaiewicz, F. Zonta, Predicting the trajectory of any COVID19 epidemic from the best straight line. medRxiv:2020.06.26.20140814 (30 June 2020).
21. J. Dehning *et al.*, Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* **369**, eabb9789 (2020).
22. Z. Feng, H. R. Thieme, Recurrent outbreaks of childhood diseases revisited: The impact of isolation. *Math. Biosci.* **128**, 93–130 (1995).
23. H. Hethcote, M. Zhien, L. Shengbing, Effects of quarantine in six endemic models for infectious diseases. *Math. Biosci.* **180**, 141–160 (2002).
24. L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic analysis of COVID-19 in China by dynamical modeling. arXiv:2002.06563 (16 February 2020).
25. T. de Camino-Beck, A modified SEIR model with confinement and lockdown of COVID-19 for Costa Rica. medRxiv:2020.05.19.20106492 (26 May 2020).
26. L. López, X. Rodó, A modified SEIR model to predict the COVID-19 outbreak in Spain: Simulating control scenarios and multi-scale epidemics. medRxiv:2020.03.27.20045005 (16 April 2020).
27. MUNQU _ team, Modelización epidemiológica del COVID-19 (2020). <https://covid19.webs.upv.es/index.html>. Accessed 12 April 2020.
28. X. Hao *et al.*, Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* **584**, 420–424 (2020).
29. M. U. G. Kraemer *et al.*, Mapping global variation in human mobility. *Nat. Hum. Behav.* **4**, 800–810 (2020).
30. W. C. Roda, M. B. Varughese, D. Han, M. Y. Li, Why is it difficult to accurately predict the COVID-19 epidemic? *Infect. Dis. Model.* **5**, 271–281 (2020).
31. T. Carletti, D. Fanelli, F. Piazza, Covid-19: The unreasonable effectiveness of simple models. *Chaos Soliton. Fract.* **5**, 100034 (2020).
32. C. A. Beauchemin, T. Miura, S. Iwami, Duration of SHIV production by infected cells is not exponentially distributed: Implications for estimates of infection parameters and antiviral efficacy. *Sci. Rep.* **7**, 42765 (2017).
33. G. Massonis, J. R. Banga, A. F. Villaverde, Structural identifiability and observability of compartmental models of the COVID-19 pandemic. arXiv:2006.14295 (25 June 2020).
34. A. Arenas *et al.*, Derivation of the effective reproduction number R for COVID-19 in relation to mobility restrictions and confinement. medRxiv:2020.04.06.20054320 (8 April 2020).
35. R. Guimerà *et al.*, A Bayesian machine scientist to aid in the solution of challenging scientific problems. *Sci. Adv.* **6**, eaav697 (2020).
36. S. Correia, S. Luck, E. Verner, Pandemics depress the economy, public health interventions do not: Evidence from the 1918 flu. (SSRN, 2020). <http://dx.doi.org/10.2139/ssrn.3561560>. Accessed 28 September 2020.
37. M. Plummer *et al.*, “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling” in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, K. Hornik, F. Leisch, A. Zeileis, Eds. (DSC, 2003), vol. 124, pp. 1–10.