medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Research Article Summary: Inferring change points in the COVID-19 spreading reveals the effectiveness of interventions

Jonas Dehning,[1*] Johannes Zierenberg,[1*] F. Paul Spitzner,[1*]
Michael Wibral,[2] Joao Pinheiro Neto,[1] Michael Wilczek,[1*] Viola Priesemann[1*]

[1] Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany,
[2] Campus Institute for Dynamics of Biological Networks, University Göttingen, Germany

[*] Authors contributed equally. Correspondence should be addressed to viola.priesemann@ds.mpg.de.

Introduction. When faced with the outbreak of a novel epidemic like COVID-19, rapid response measures are required by individuals as well as by society as a whole to mitigate the spread of the virus. During this initial, time-critical period, neither the central epidemiological parameters, nor the effectiveness of interventions like cancellation of public events, school closings, and social distancing are known.

Rationale. As one of the key epidemiological parameters, we infer the spreading rate λ from confirmed COVID-19 case numbers at the example of Germany by combining Bayesian inference based on Markov-Chain Monte-Carlo sampling with a class of SIR (Susceptible-Infected-Recovered) compartmental models from epidemiology. Our analysis characterizes the temporal change of the spreading rate and, importantly, allows us to identify potential change points and to provide short-term forecast scenarios based on various degrees of social distancing. A detailed description is provided in the accompanying paper, and the models, inference, and predictions are available on github. While we apply it to Germany, our approach can be readily adapted to other countries or regions.

Results. In Germany, interventions to contain the outbreak were implemented in three steps over three weeks: Around March 9, large public events like soccer matches were cancelled. On March 16, schools and childcare facilities as well as many non-essential stores were closed. One week later, on March 23, a far-reaching contact ban ("Kontaktsperre"), which included the prohibition of even small public gatherings as well as the further closing of restaurants and non-essential stores, was imposed by the government authorities.

From the observed case numbers of COVID-19, we can quantify the impact of these measures on the disease spread (Fig. 0). Based on our analysis, which includes data until April 21, we have evidence of three change points: the first changed the spreading rate from $\lambda_0 = 0.43$ (95 % credible interval (CI: $[0.35, 0.51]$)) to $\lambda_1 = 0.25$ (CI: $[0.20, 0.30]$), and occurred around March 6 (CI: March 2 to March 9); the second change point resulted in $\lambda_2 = 0.15$ (CI: $[0.12, 0.20]$), and occurred around March 15 (CI: March 13 to March 17). Both changes in λ slowed the spread of the virus, but still implied exponential growth (Fig. 0, red and orange traces). To contain the disease spread, and turn from exponential growth to a decline of new cases, a further decrease in λ was necessary. Our analysis shows that this transition has been reached by the third change point that resulted in $\lambda_3 = 0.09$ (CI: $[0.06, 0.12]$) around March 23 (CI: March 20 to March 25).

With this third change point, λ transitioned below the critical value where the spreading rate λ balances the recovery rate μ, i.e. the effective growth rate $\lambda^* = \lambda - \mu \approx 0$ (Fig. 0, gray traces). Importantly, $\lambda^* = 0$ presents the watershed between exponential growth or decay. Given the delay of approximately two weeks between an intervention and first inference of the induced changes in λ^*, future interventions such as lifting restrictions warrant careful consideration.

Our detailed analysis shows that, *in the current phase*, reliable short- and long-term forecasts are very difficult as they critically hinge on how the epidemiological parameters change in response to interventions: In Fig. 0 already the three example scenarios quickly diverge from each other, and consequently span a considerable range of future case numbers. Thus, any uncertainty on the magnitude of our social distancing in the past two weeks can have a major impact on the case numbers in the next two weeks. Beyond two weeks, the case numbers depend on our future behavior, for which we have to make explicit assumptions. In the main paper we illustrate how the precise magnitude and timing of potential change points impact the forecast of case numbers (Fig. 2).

Conclusions. We developed a Bayesian framework to infer central epidemiological parameters and the timing and magnitude of intervention effects. Thereby, the efficiency of political and individual intervention measures for social distancing and containment can be assessed in a timely manner. We find evidence for a successive decrease of the spreading rate in Germany around March 6 and around March 15, which significantly reduced the magnitude of exponential growth, but was not sufficient to turn growth into decay. Our analysis also shows that a further decrease of the spreading rate occurred around March 23, turning exponential growth into decay. Future interventions and lifting of restrictions can be modeled as additional change points, enabling short-term forecasts for case numbers. In general, our analysis code may help to infer the efficiency of measures taken in other countries and inform policy makers about tightening, loosening and selecting appropriate rules for containment.

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/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

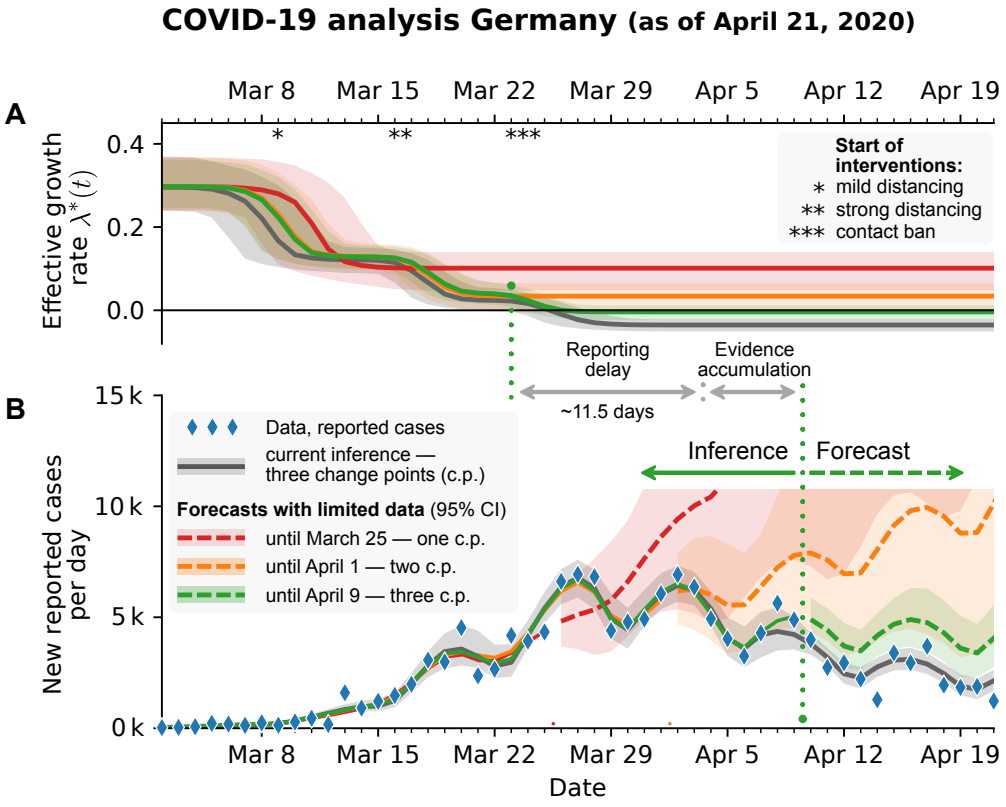

FIG. 0. Bayesian analysis of the German COVID-19 data (blue diamonds) reveals three change points that match the timing of publicly announced interventions. **A:** The inferred effective growth rate (difference between spreading and recovery rate, $\lambda^* = \lambda - \mu$) for an SIR model with weekly reporting modulation and reporting delay that includes scenarios with one, two or three change points (red, orange, green; fitted to case reports until March 25, April 1 and April 9, respectively). The timing of the inferred change points in growth rate is consistent with the timing of German governmental interventions (depicted as $*$, $**$, and $***$). **B:** Comparing inferred models with the actual new reported cases per day reveals the effectiveness of governmental interventions. After the first two interventions, the number of new cases still grew exponentially (red, orange); only after the third intervention did the number of new cases start to saturate (green) or even to decline (gray, data until April 21). This illustrates that the future development strongly depends on our distancing behavior. Note the delay between a change point and the observation of changes in the number of new cases of almost two weeks — a combination of reporting delay and a minimal period of evidence accumulation.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

Inferring change points in the COVID-19 spreading reveals the effectiveness of interventions

Jonas Dehning[1*], Johannes Zierenberg[1*], F. Paul Spitzner[1*], Michael Wibral[2], Joao Pinheiro Neto[1], Michael Wilczek[1*], and Viola Priesemann[1*]

[1] *Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany and*

[2] *Campus Institute for Dynamics of Biological Networks, University of Göttingen, Germany*

(Dated: May 4, 2020)

As COVID-19 is rapidly spreading across the globe, short-term modeling forecasts provide time-critical information for decisions on containment and mitigation strategies. A main challenge for short-term forecasts is the assessment of key epidemiological parameters and how they change when first interventions show an effect. By combining an established epidemiological model with Bayesian inference, we analyze the time dependence of the effective growth rate of new infections. Focusing on the COVID-19 spread in Germany, we detect change points in the effective growth rate that correlate well with the times of publicly announced interventions. Thereby, we can quantify the effect of interventions, and we can incorporate the corresponding change points into forecasts of future scenarios and case numbers. Our code is freely available and can be readily adapted to any country or region.

INTRODUCTION

During the initial outbreak of an epidemic, reliable short-term forecasts are key to estimate required medical capacities, and to inform and advice the public and decision makers [1]. During this initial phase, three tasks are of particular importance to provide time-critical information for crisis mitigation: (1) establishing central epidemiological parameters, such as the basic reproduction number, that can be used for short-term forecasting; (2) simulating the effects of different possible interventions aimed at the mitigation of the outbreak; (3) estimating the actual effects of the measures taken — to rapidly adjust them and to adapt short-term forecasts. Tackling these tasks is challenging due to the large statistical and systematic errors that are present during the initial stages of an epidemic with its low case numbers. This is further complicated by the fact that mitigation measures are taken rapidly, while the outbreak unfolds, but they take an effect only after an a priori unknown delay. To obtain reasonable parameter estimates for short-term forecasting and policy evaluation despite these complications, any prior knowledge available needs to be integrated into modeling efforts to reduce uncertainties. This includes knowledge about basic mechanisms of disease transmission, recovery, as well as preliminary estimates of epidemiological parameters from other countries, or from closely related pathogens. The integration of prior knowledge, the quantitative assessment of the remaining uncertainties about epidemiological parameters, and the principled propagation of these uncertainties into forecasts is the domain of Bayesian modeling and inference [2, 3].

Here, we draw on an established class of models for epidemic outbreaks: The Susceptible-Infected-Recovered (SIR) model [4–7] specifies the rates with which population compartments change over time, i.e., with which susceptible people become infectious, or infectious people recover. This simple model can be formulated in terms of coupled ordinary differential equations (in mean field), which enable analytical treatment [8, 9] or fast evaluation (ideally suited for Bayesian inference). Accordingly, SIR-like models have been used to model epidemic spreads, from Bayesian Markov-Chain Monte Carlo (MCMC) parameter estimation [10–12] to detailed scenario discussions [13–16]. Recently, this family of models also played a dominant role in the analyses of the global corona virus (SARS-CoV-2) outbreak, from inference [17–19] to scenario forecast [20–27] to control strategies [28, 29].

We combine the SIR model (and generalizations thereof) with Bayesian parameter inference and augment the model by a time-dependent spreading rate. The time dependence is implemented via potential change points that reflect changes in the spreading rate driven by governmental interventions. Based on three distinct measures taken in Germany, we detect three corresponding change points from reported COVID-19 case numbers. Already on April 1 we had reported evidence for the first two change points, and predicted the third one [30]. Now, with data until April 21, we have evidence for all three change points. First, the spreading rate decreased from 0.43 (with 95% credible interval, CI $[0.35, 0.51]$) to 0.25 (CI $[0.20, 0.30]$), with this decrease initiated around March 6 (CI [2, 9]). This matches the cancellation of large public events such as trade fairs and soccer matches. Second, the spreading rate decreased further to 0.15 (CI $[0.12, 0.20]$) initiated around March 15 (CI [13, 17]). This matches the closing of schools, childcare facilities, and non-essential stores. Third, the spreading rate decreased further to 0.09 (CI $[0.06, 0.13]$) initiated around

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

March 23 (CI [20, 25]). This corresponds well to the strict contact ban, which was announced on March 22. While the first two change points were not sufficient to switch from growth of novel cases to a decline, the third change point brought this crucial reversal.

Our framework is designed to infer the effectiveness of past measures and to explore potential future scenarios along with propagating the respective uncertainties. In the following, we demonstrate the potential impact of timing and magnitude of change points, and report our inference about the three past governmental interventions in Germany. Our framework can be readily adapted to any other country or region. The code (already including data sources from many other countries), as well as the figures are all available on Github [31].

BACKGROUND: INFERENCE OF CENTRAL EPIDEMIOLOGICAL PARAMETERS AND THE EFFECTS OF INTERVENTIONS

In order to simulate the general effect of different possible interventions on the spread of COVID-19 in Germany, we first focus on the initial phase of the outbreak when no serious mitigation measures were implemented. In the absence of interventions, an epidemic outbreak can be described by SIR models with constant spreading rate (Methods). In Germany, first serious interventions occurred around March 9 and affected the case reports with an observation delay (a combination of incubation period with median 5–6 days [32]) and a test delay (time until doctor is visited plus test-evaluation time) that we assume to be both about 2–3 days. Hence, in order to infer central epidemiological parameters, we consider as the initial phase the time period from March 2 to March 15. In order to simulate the effect of different possible interventions, we then model the effects of interventions as change points in the spreading rate (Methods).

Bayesian inference for central epidemiological parameters during the initial phase of the outbreak

We perform Bayesian inference for the central epidemiological parameters of an SIR model using Markov-Chain Monte Carlo (MCMC) sampling (Fig. 1). The central parameters are the spreading rate λ, the recovery rate μ, the reporting delay D and the number of initially infected people I_0. We chose informative priors based on available knowledge for λ, μ and D, and we chose uninformative priors for the remaining parameters (Methods). Also, we intentionally kept the informative priors as broad as possible such that the data would constrain the parameters (Fig. 1).

As median estimates, we obtain for the spreading rate $\lambda = 0.41$, $\mu = 0.12$, $D = 8.6$, and $I_0 = 19$ (see Fig. 1 C–H for the posterior distributions and the 95% credible intervals). Converted to the basic reproduction number $R_0 = \lambda/\mu$, we find a median $R_0 = 3.4$ (CI [2.4, 4.7]), which is consistent with previous reports that find median values between 2.3 and 3.3 [18, 33, 34]. Overall, the model shows good agreement with both new cases (Fig. 1 A) and cumulative cases (Fig. 1 B) that show the expected exponential growth (linear in log-lin plot). The observed data are clearly informative about λ, I_0 and σ (indicated by the difference between the priors (gray line) and posteriors (histograms) in Fig. 1 D,E,F). However, μ and D are largely determined by our prior choice of parameters (histograms match gray line in Fig. 1 C,H). This is to be expected for the initial phase of an epidemic outbreak, which is dominated by exponential growth.

In order to quantify the impact of possible interventions, we concentrate on the effective growth of active infections before and after the intervention. As long as the number of infections and recoveries are small compared to the population size, the number of active infections per day can be approximated by an exponential growth (Fig. 1A,B) with effective growth rate $\lambda^* = \lambda - \mu$ (see Methods). As a consequence, λ and μ cannot be estimated independently. This is further supported by a systematic scan of the model's log-likelihood in the λ–μ space that shows an equipotential line for the maximum likelihood (Fig. 1 I). This strongly suggests that the growth rate λ^* is the relevant free parameter with a median $\lambda^* = 28\%$ (Fig. 1 G). The control parameter of the dynamics in the exponential phase is thus the (effective) growth rate: If the growth rate is larger than zero ($\lambda > \mu$), case numbers grow exponentially; if the growth rate is smaller than zero ($\lambda < \mu$), the recovery dominates and the new cases decrease. The two different dynamics (supercritical and subcritical, respectively) are separated by a critical point at $\lambda^* = 0$ ($\lambda = \mu$) [35].

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

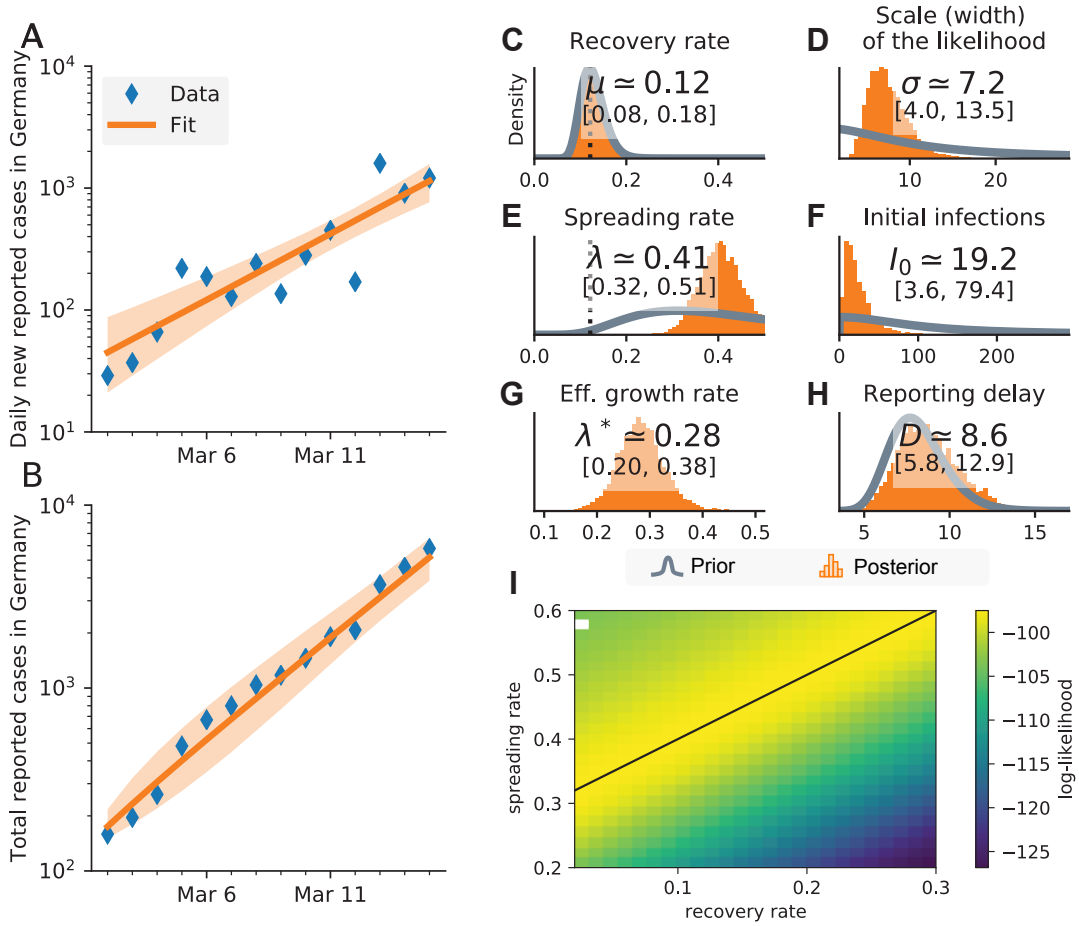

FIG. 1. Inference of central epidemiological parameters of the SIR model during the initial onset period, March 2–15. **A:** The number of new cases and **B:** the total (cumulative) number of cases increase exponentially over time. **C–H:** Prior (gray) and posterior (orange) distributions for all model parameters: estimated spreading rate λ, recovery rate μ, reporting delay D between infection date and reporting date, number of cases I_0 at the start of the simulation, scale-factor σ of the width of the likelihood distribution, and the effective growth rate $\lambda^* = \lambda - \mu$. **I:** Log-likelihood distribution for different combinations of λ and μ. A linear combination of λ and μ yields the same maximal likelihood (black line). White dot: Inference did not converge.

Magnitude and timing of interventions matter for the mitigation of the outbreak

We simulate different, hypothetical interventions following the initial phase in order to show that both, the amount of change in behavior (leading to a change in spreading rate λ, Fig. 2 A) and the exact timing of the change (Fig. 2 B) determine the future development. Hypothetical interventions build on the inferred parameters from the initial phase (Fig. 1, in particular median $\lambda_0 = 0.41$ and median $\mu = 0.12$) and were implemented as change points in the spreading rate from the inferred λ_0 to a new value λ_1. With such a change point, we model three potential scenarios of public behavior: **(I) No social distancing;** Public behavior is unaltered and the spread continues with the inferred rate ($\lambda_1 = \lambda_0$ with median $\lambda_1 = 0.41 > \mu$). **(II) Mild social distancing;** The spreading rate decreases to 50%, ($\lambda_1 = \lambda_0 / 2$ with median $\lambda_1 = 0.21 > \mu$). Although people effectively reduce the number of contacts by a factor of two in this second scenario, the total number of reported cases continues to grow alongside scenario (I) for the time period of the reporting delay D (median $D = 8.6$ from initial phase, see below for a more constrained estimation). Also, we still observe an exponential increase of new infections after the intervention becomes effective, because the growth rate remains positive, $\lambda_1^* = \lambda_1 - \mu > 0$. **(III) Strong social distancing;** Here, the spreading rate decreases to 10%, ($\lambda_1 = \lambda_0 / 10$ with median $\lambda_1 = 0.04 < \mu$). The assumptions here are that contacts are severely limited, but even when people stay at home as much as possible, *some* contacts are still unavoidable. Even under such drastic policy changes, no effect is visible until the reporting delay D is over. Thereafter, a quick decrease in daily new infections manifests within two weeks (delay plus change point duration), and the total number of cases reaches a stable plateau.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

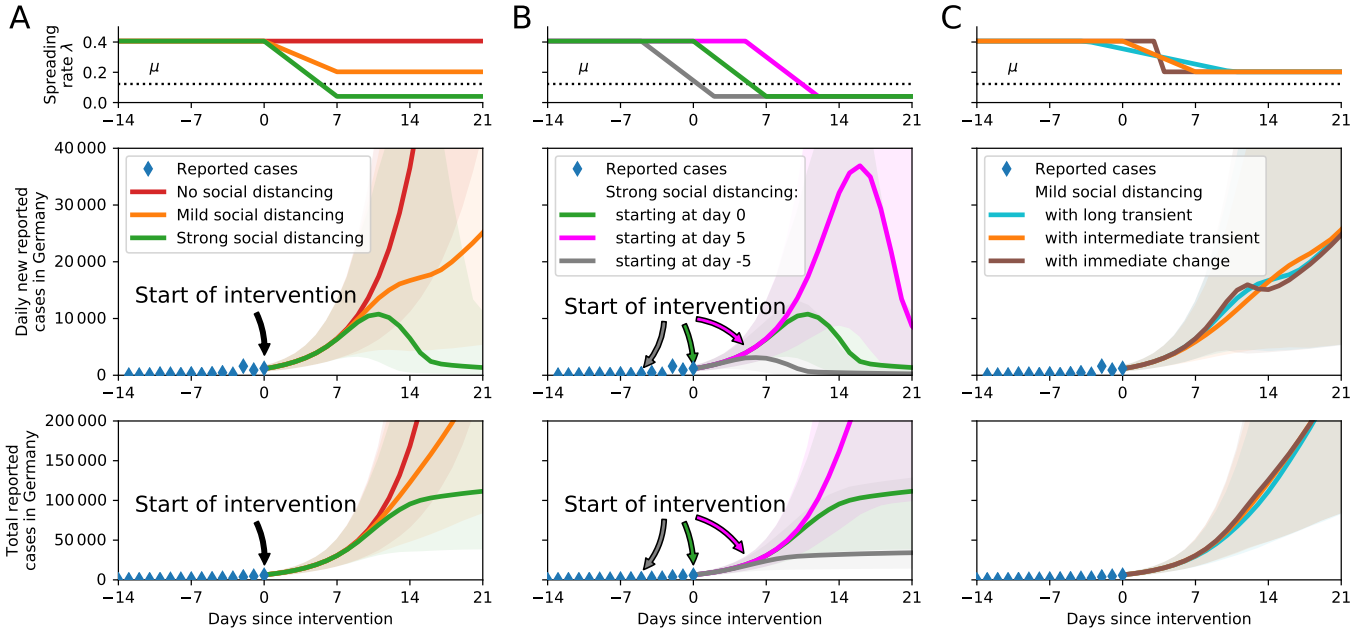

FIG. 2. The timing and effectiveness of interventions strongly impact future COVID-19 cases. **A:** We assume three different scenarios for interventions starting on March 16: (I, red) no social distancing, (II, orange) mild social distancing, or (III, green) strict social distancing. **B:** Delaying the restrictions has a major impact on case numbers: strict restrictions starting on March 16 (green), five days later (magenta) or five days earlier (gray). **C:** Comparison of the time span over which interventions ramp up to full effect. For all ramps that are *centered around the same day*, the resulting case numbers are fairly similar. However, a sudden change of the spreading rate can cause a temporary decrease of daily new cases, although $\lambda > \mu$ at all times (brown).

Only in this last scenario a plateau is reached, because here the growth rate becomes negative, $\lambda^* < 0$, which leads to decreasing numbers of new infections.

Furthermore, the timing of an intervention matters: Apart from the strength of an intervention, its *onset* time has great impact on the total case number (Fig. 2 B,C). For example, focusing on the strong intervention (III) — where a stable plateau is reached — the effect of advancing or delaying the change point by just five days leads to more than a three-fold difference in cumulative cases.

While we find that the timing of an intervention has great effect on case numbers, the duration over which the change takes place has only minor effect — if the intervals of change are centered around the same date. In Fig. 2 C we illustrate the adjustment of $\lambda_0 \to \lambda_1$ for mild social distancing with durations of 14, 7 and 1 day(s). The change point duration is a simple way to incorporate variability in individual behavior, and is not linked to the reporting delay D. As an interesting effect, a sudden change in the spreading rate can lead to a temporary decrease of new case numbers, despite the fact that the effective growth rate remains positive at all times.

RESULTS

In order to model real-world data, we further refined the SIR model. Most importantly, we account for systematic variations of case reports throughout the week, in particular lower case numbers on the weekend, by explicitly modeling a weekly reporting modulation (see Methods). Indeed, model comparisons confirm that models with this correction outperform those without (see Table S2). In the supplemental material, we further generalize our model to include an explicit incubation period (SEIR-like, Fig. S3) that yields results consistent with our main model.

We incorporate the effect of governmental interventions into our models by introducing flexible change points in the spreading rate (see Methods). During the COVID-19 outbreak in Germany, governmental interventions occurred in three stages from (i) the cancellation of large events with more than 1000 participants (around March 9), through (ii) closing of schools, childcare facilities and the majority of stores (in effect March 16), to (iii) the contact ban and closing of all non-essential stores (in effect March 23). The aim of all these interventions was to reduce the (effective) growth rate $\lambda^* = \lambda - \mu$. As soon as the growth rate becomes negative ($\lambda^* < 0$), the number of new confirmed infections decreases after the respective reporting delay.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

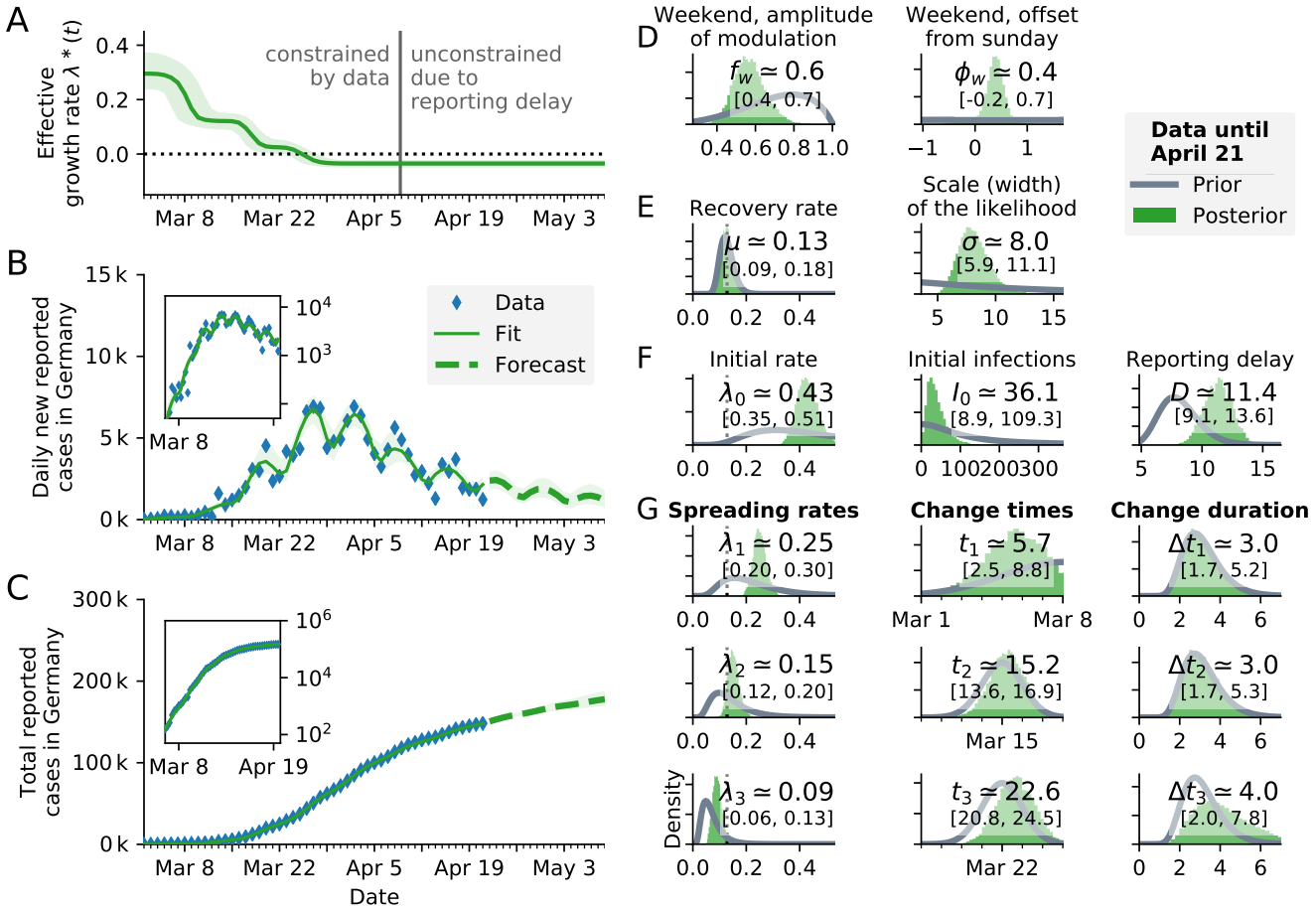

FIG. 3. Bayesian analysis of the German COVID-19 data (blue diamonds) until April 21 reveals three change points that are consistent with three major governmental interventions. **A:** Time-dependent model estimate of the effective spreading rate $\lambda^*(t)$. **B:** Comparison of daily new reported cases and the model (green solid line for median fit with 95% credible intervals, dashed line for median forecast with 95% CI); **inset:** same data in log-lin scale. **C:** Comparison of total reported cases and the model (same representation as in B). **D–F:** Priors (gray lines) and posteriors (green histograms) of all model parameters; inset values indicate the median and 95% credible intervals of the posteriors. For the same model with one or two change points, please see the corresponding figures in the SI (Fig. S1, S2, Table S2).

Detecting change points in the spreading rate — and quantifying the amount of change as quickly as possible — becomes a central modeling challenge when short-term forecasts are required. To address this challenge, we assume an initial spreading rate λ_0 (the exponential growth phase, cf. Fig. 1) and up to three potential change points motivated by the German governmental interventions: The first change point ($\lambda_0 \to \lambda_1$) is expected around March 9 (t_1) as a result of the official recommendations to cancel large events. A second change point ($\lambda_1 \to \lambda_2$) is expected around March 16 (t_2), when schools and many stores were closed. A third change point ($\lambda_2 \to \lambda_3$) is expected around March 23 (t_3), when all non-essential stores were closed, and a contact ban was enacted. We model the behavioral changes that are introduced at these change points to unfold over a few days Δt_i, but the changes in duration can be partly compensated by changes in the onset time t_i (see Fig. 2 C, scenarios). We chose priors for all parameters based on the information available to us up to March 28 (see Methods). In addition, we performed a sensitivity analysis by employing wider priors in the supplemental material (Figs. S5, S6, S7, Table S2), which yielded consistent results. On March 28, the data were already informative about the first change point, and thereby helped to inform our forecast scenarios.

The inferred parameters for the models with change points are consistent with the inferred parameters from the exponential onset phase (Figs. 1, 3 & Figs. S1, S2). In particular, all estimated λ_0-values from models with multiple change points are compatible with the value of the model without change points (during the exponential onset phase, $\lambda_0 = 0.41$, CI $[0.32, 0.51]$, assuming a stationary λ until March 15, Fig. 1 E). Also the scale factor σ and the number

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

TABLE I. Model comparison shows that the three-change-point model describes the data best: leave-one-out (LOO) cross-validation for main models (SIR with weekend correction) and a different number of change points. Lower LOO-scores represent a better match between model and data.

Model	LOO-score	Effective number of parameters (pLOO)
zero change points	927 ± 9	8.31
one change points	819 ± 16	13.46
two change points	796 ± 17	12.53
three change points	787 ± 17	13.42

of initial infections I_0 for the models with change points are consistent with the initial model inference during the exponential onset phase (Fig. 1 D,F).

The models with two or three change points fit the observed data better than those with fewer change points

The models with three change points describe the data better than models with fewer change points, as indicated by the leave-one-out (LOO) cross-validation-based Bayesian model comparison [36] (lowest LOO-score in Table I). However, the LOO-scores of the model with two and three change points differ by less than one standard error. This originates from an extended duration of the second change point in the two-change-point model, which partially captures the effect of the third intervention. As expected, the models with none or a single change point have LOO-scores that are at least one standard deviation higher (worse) than those of the best models, and we will not consider them further.

When comparing our inference based on three change points to the number of confirmed cases, we find them to largely match (Fig. 3 B,C). The dominant periodic change in the daily new reported cases (Fig. 3 B) is well described by the weekday modulation. In addition to the periodic change, the daily new case numbers also reflect the fairly sudden change of the spreading rate at the change points (cf. Figs. 2 and S4 for the effect of change points without the modulation). Most importantly, the cumulative effect of change points manifests in an overarching decay in new case numbers that is visible after April 5 and follows the third change point (with reporting delay).

Change point detection quantifies the effect of governmental interventions on the outbreak of COVID-19 in Germany

Ideally, detected changes can be related to specific mitigation measures, so that one gains insights into the effectiveness of different measures (Fig. 3). Indeed, we found clear evidence for three change points in the posterior distributions of the model parameters: First, $\lambda(t)$ decreased from $\lambda_0 = 0.43$ (with 95% credible interval, CI $[0.35, 0.51]$) to $\lambda_1 = 0.25$ (CI [0.20, 0.30]). The date of the change point was inferred to be March 6 (CI [2, 9])]; this inferred date matches the timing of the first governmental intervention including cancellations of large events, as well as increased awareness. After this first intervention, the (effective) growth rate $\lambda^*(t) = \lambda(t) - \mu$ decreased by more than a factor 2, from median $\lambda_0 - \mu = 0.3$ to median $\lambda_1 - \mu = 0.12$, given that the recovery rate was inferred as $\mu = 0.13$ (CI [0.09, 0.18]). Second, $\lambda(t)$ decreased from $\lambda_1 = 0.25$ to $\lambda_2 = 0.15$ (CI [0.12, 0.20]), which is larger than our prior assumption. The date of the change point was inferred to be March 15 (CI [13, 17])]; this inferred date matches the timing of the second governmental intervention including closing schools and some stores. After this second intervention, the median growth rate became $\lambda^*(t) = \lambda_2 - \mu = 0.02 \approx 0$ and is thus in the vicinity of the critical point, yet still positive. The first two interventions in Germany thereby mitigated the spread by drastically reducing the growth rate, but the spread of COVID-19 remained exponential. Third, $\lambda(t)$ decreased from $\lambda_2 = 0.15$ to $\lambda_3 = 0.09$ (CI [0.06, 0.13]). The date of the change point was inferred to be March 23 (CI [20, 25])]; this inferred date matches the timing of the third governmental intervention including contact ban and closing of all non-essential stores. Only after this third intervention, the median (effective) growth rate, $\lambda^*(t) = \lambda_3 - \mu = -0.03 < 0$ (CI $[-0.05, -0.02]$)], finally became negative, indicating a decrease in the number of new infections. We can thus clearly relate the change points to the governmental interventions and quantify their mitigation effect.

DISCUSSION

We presented a Bayesian approach for a timely monitoring of the effect of governmental interventions on the spread of an epidemic outbreak. At the example of the COVID-19 outbreak in Germany, we applied this approach to infer

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

the central epidemiological parameters and three change points in the spreading rate from the number of reported cases. We showed that change points in the spreading rate affect the confirmed case numbers with a delay of about two weeks (median reporting delay of $D = 11.4$ days plus a median change-point duration of 3 days). Thereby, we were able to relate the inferred change points to the three major governmental interventions in Germany: We found a clear reduction of the spreading rate related to each governmental intervention (Fig. 3), (i) the cancellation of large events with more than 1000 participants (around March 9), (ii) the closing of schools, childcare centers and the majority of stores (in effect March 16), and (iii) the contact ban and closing of all non-essential stores (in effect March 23).

Our results suggest that the full extent of governmental interventions was necessary to stop exponential growth. The first two governmental interventions brought a reduction of the growth rate λ^* from 30% to 12% and down to 2%, respectively. However, these numbers still implied exponential growth. Only with the third intervention — the contact ban — we found that we have crossed the transition in new case numbers from growth to decay. However, the decay rate of about -3% (CI $[-5\%, -2\%]$) remains close to zero. Hence, even a minor increase in the spreading rate may again switch the dynamics to the unstable regime with exponential growth.

We used a formal Bayesian model comparison in order to validate the presence of change points. Our model comparison ruled out models with fewer than two change points (Table I, S2). While this may seem trivial, it has important consequences for making short-term forecasts that decision makers rely on: Demonstrating and quantifying the effect of past change points can be used to formulate priors for the effects of future, similar change points. These priors help to project the effects of more recent change points into future forecasts, even when these change points are not apparent in the reported case numbers yet. Consequently, it is important to look out for and identify potential change points as early as possible to incorporate them into forecasts.

The detection of change points and their interpretation depend crucially on an accurate estimate of the reporting delay D. Therefore, the validity of its estimate should be evaluated. In our model, D contains at least three distinct factors: the biological incubation period (median 5–6 days) [32], an additional delay from first symptoms to symptoms motivating a test (1–3 days) and a possible delay before a testing results come in (1–4 days). The sum of these delays seems compatible with our inferred median delay of $D = 11.4$ days, especially given the wide range of reported incubation periods.

We chose to keep our main model comparatively simple, because of the small number of data points initially available during an epidemic outbreak. With such a low number of data points, only a limited number of parameters can be effectively constrained. Hence, we chose to approximate a time-dependent spreading rate $\lambda(t)$ by episodes of constant spreading rates λ_i that are separated by three change points where a transition occurs. Our results show that this main model is currently sufficient for Germany: While we introduced fairly broad priors on the spreading rates, we obtained comparably narrow posterior distributions for each spreading rate λ_i (Fig. 3). We additionally evaluated extensions of our main model with three change points, e.g., by explicitly taking into account the incubation period (Fig. S3). These models yield consistent results for the three change points, and all have LOO scores within one standard error of each other. Thus, we consider our main model to be sufficient for case numbers in Germany at present.

Our framework can be easily adapted to other countries and enables one to incorporate future developments. For other countries, or for forecasts within smaller communities (e.g. federal states or cities), additional details may become important, such as explicit modeling of incubation time distributions [17, 37] (i.e. as done in Fig. S3), spatial heterogeneity [17, 21], isolation effects [20, 37], subsampling effects hiding undetected cases even beyond the reporting delay [38, 39], or the age and contact structure of the population [26]. In countries where drastic changes in test coverage are expected, this will have to be included as well. The methodology presented here is capable in principle of incorporating such details. It also lends itself to modeling of continuous drifts in the spreading rate, e.g. reflecting reactions of the public to news coverage of a catastrophic situation, or people growing tired of mitigation measures. Such further adaptations, however, can only be performed on a per-country basis by experts with an intimate knowledge of the local situation. Our code provides a solid and extensible base for this. For Germany, several developments in the near future may have to be included in the model. First, people may have transiently changed their behavior over the Easter holidays; second, we expect a series of change points, as well as continuous drifts, with governments trying to ease and calibrate mitigation measures. Third, extensions to hierarchical models will enable regional assessments, e.g. on the level of federal states.

Even after the three major governmental interventions in Germany, effective growth rates remain close to zero and warrant careful consideration of future measures. At present, estimates of effective growth rates dropped to -3% and thereby remain close to zero – the watershed between exponential growth or decay. Together with the delay of approximately two weeks between infection and case report this warrants caution in lifting restrictions for two reasons: First, lifting restrictions too much will quickly lead to renewed exponential growth; second, we would be effectively blind to this worsened situation for nearly two weeks in which it will develop uninhibited. This may result in unwanted growth in case numbers beyond the level that the health system can cope with – especially when the active cases have

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

not gone down close to zero before lifting restrictions, thus re-initiating growth from a high base level. Therefore, it is important to consider lifting restriction only when the number of active cases are so low that a two-week increase will not pose a serious threat.

In conclusion, the presented Bayesian approach allows to detect and quantify the effect of recent governmental interventions and – combined with potential subsequent interventions – to forecast future case number scenarios. Our analysis highlights the importance of precise timing and magnitude of interventions for future case numbers. It also stresses the importance of including the reporting delay D between the date of infection and the date of the confirmed case in the model. The delay D, together with the time required to implement interventions causes a total delay between an intervention and its visibility in the case numbers of about two weeks for COVID-19 in Germany. This means that changes in our behavior today can only be detected in confirmed cases in two weeks. Combined with the current spreading rate that is still around zero, the inferred spreading and observation dynamics warrant an extremely careful planning of future measures.

MATERIALS AND METHODS

As a basis for our Bayesian inference and the forecast scenarios, we use the differential equations of the well-established SIR (Susceptible-Infected-Recovered) model. We also test the robustness of our results by means of more sophisticated models, in particular an SEIR-like model that explicitly incorporates an incubation period (Fig. S3). While the SIR model-dynamics is well understood in general, here our main challenge is to estimate model parameters specifically for the COVID-19 outbreak, and to use them for forecasting. To that end, we combined a Bayesian approach — to incorporate prior knowledge — with Markov Chain Monte Carlo (MCMC) sampling — to compute the posterior distribution of the parameters and to sample from it for forecasting. Put simply, we first estimate the parameter distribution that best describes the observed situation, and then we use many samples from this parameter distribution to evolve the model equations and thus forecast future developments.

The data used comes from the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) dashboard [40]. The JHU CSSE provides up-to-date data on COVID-19 infections, usually a few days ahead of official German sources. The exact version of the data and code is available at [31]. Data were incorporated until April 21. Note that after this cutoff date, additional modeling of the effects of behavioral changes over the Easter holidays becomes necessary.

Simple model: SIR model with stationary spreading rate

We consider a time-discrete version of the standard SIR model. In short, we assume that the disease spreads at rate λ from the infected population compartment (I) to the susceptible compartment (S), and that the infected population compartment recovers (R) at rate μ. This well-established model for disease spreading can be described by the following set of (deterministic) ordinary differential equations (see, e.g., Refs [5, 6, 20]). Within a population of size N,

$$\begin{aligned} \tfrac{\mathrm{d}S}{\mathrm{d}t} &= -\lambda \tfrac{SI}{N} \\ \tfrac{\mathrm{d}I}{\mathrm{d}t} &= \lambda \tfrac{SI}{N} - \mu I \\ \tfrac{\mathrm{d}R}{\mathrm{d}t} &= \mu I\,. \end{aligned} \tag{1}$$

As a remark, during the onset phase of an epidemic only a very small fraction of the population is infected (I) or recovered (R), and thus $S \approx N \gg I$ such that $S/N \approx 1$. Therefore, the differential equation for the infected reduces to a simple linear equation, exhibiting an exponential growth

$$\frac{dI}{dt} = (\lambda - \mu) I \quad \text{solved by} \quad I(t) = I(0)\; e^{(\lambda-\mu)t}\,. \tag{2}$$

Because our data set is discrete in time (Δt =1 day), we solve the above differential equations with a discrete time step ($\mathrm{d}I/\mathrm{d}t \approx \Delta I/\Delta t$), such that

$$\begin{aligned} S_t - S_{t-1} &= -\lambda \Delta t \tfrac{S_{t-1}}{N} I_{t-1} &&=: -I_t^{\text{new}} \\ R_t - R_{t-1} &= \mu \Delta t I_{t-1} &&=: R_t^{\text{new}} \\ I_t - I_{t-1} &= \left(\lambda \tfrac{S_{t-1}}{N} - \mu\right) \Delta t I_{t-1} &&= I_t^{\text{new}} - R_t^{\text{new}}\,. \end{aligned} \tag{3}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

Importantly, I_t models the number of all (currently) active infected people, while I_t^{new} is the number of new infections that will eventually be reported according to standard WHO convention. Importantly, we explicitly include a reporting delay D between new infections I_t^{new} and newly reported cases (C_t) as

$$C_t = I_{t-D}^{\text{new}}. \tag{4}$$

We begin our simulations at time $t = 0$ with I_0 infected cases and start including real-word data of reported cases $\hat{C}_t$ from day $t > D$ (see below for a parameterization).

In our model we do not explicitly incorporate the inflow of additional infected people by travel for two reasons. First, we implicitly model the initial surge of infections with I_0. Second, previous work showed that travel during the outbreak has only modest effects on the dynamics, e.g., travel restrictions in China merely delayed the exponential spread if not combined with reductions of spreading [41].

Full model: SIR model with weekly reporting modulation and change points in spreading rate

Our change point detection builds on a generalization of the simple SIR model with stationary spreading rate. We now assume that the spreading rate λ_i, $i = 1, ..., n$, may change at certain time points t_i from λ_{i-1} to λ_i, linearly over a time window of Δt_i days. Thereby, we account for policy changes to reduce λ, which were implemented in Germany step by step. Thus, the parameters t_i, Δt_i, and λ_i are added to the parameter set of the simple model above, and the differential equations are augmented by the time-varying λ_i.

In addition, we include a weekly modulation to account for lower case reports around the weekend which subsequently accumulate during the week. To model the systematic variation of case reports during the week, we adapted the newly reported cases by a reporting fraction

$$\begin{aligned} C_t &= I_{t-D}^{\text{new}}\,(1 - f(t))\,, \qquad \text{with} \\ f(t) &= (1 - f_w) \cdot \left(1 - \left|\sin\left(\tfrac{\pi}{7}t - \tfrac{1}{2}\Phi_w\right)\right|\right)\,, \end{aligned} \tag{5}$$

where f_w and Φ_w will also be constrained by the data.

Estimating model parameters with Bayesian MCMC

We estimate the set of model parameters $\theta = \{\lambda_i, t_i, \mu, D, \sigma, I_0, f_w, \Phi_w\}$ using Bayesian inference with Markov-chain Monte-Carlo (MCMC). The parameter σ is the scale factor for the width of the likelihood $P\big(\hat{C}_t \big| \theta\big)$ between observed data and model (see below). Our implementation relies on the python package PyMC3 [42] with NUTS (No-U-Turn Sampling) [43] using multiple, independent Markov chains. The structure of our approach is the following:

Initialization of the Markov chains via variational inference. The posterior is approximated by Gaussian distributions ignoring correlations between parameters through automatic differentiation variational inference (ADVI) [44], which is implemented in PyMC3. From this distribution, four starting points for four chains are sampled.

Burn-in phase: Each chain performs 1000 burn-in (tuning) steps using NUTS, which are not recorded. This serves as equilibration in order to sample from an equilibrium distribution in the next step.

Sampling phase: Each chain performs 4000 steps, which are used to approximate the posterior distribution. To ensure that the Chains are equilibrated and sampled from the whole posterior distribution (ergodicity), we verified that the R-hat statistic is below 1.05, which is implemented in PyMC3. The rank normalized R-hat diagnostic tests for lack of convergence by comparing the variances within chains and between chains: For identical within-chain and between-chain variances R-hat becomes 1, indicating convergence. For well-converged chains the resulting samples will describe the real-world data well, so that consistent forecasts are possible in the forecast phase.

Forecast using Monte Carlo samples. For the forecast, we take all samples from the MCMC step and continue time integration according to different forecast scenarios explained below. Note that the overall procedure yields an ensemble of forecasts — as opposed to a single forecast that would be solely based on one set of (previously optimized) parameters.

MCMC sampling details Each MCMC step requires to propose a new set of parameters θ, to generate a (fully deterministic) time series of new infected cases $C(\theta) = \{C_t(\theta)\}$ of the same length as the observed real-world data $\hat{C} = \left\{\hat{C}_t\right\}$, and to accept or reject θ. In our case, the NUTS implementation (in PyMC3) first proposes a new set of

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

parameters θ based on an advanced gradient-based algorithm and subsequently accepts or rejects it such that the resulting samples reflect the posterior distribution

$$p(\theta|\hat{C}) \propto p(\hat{C}|\theta)p(\theta),$$

where $p(\hat{C}|\theta)$ is the likelihood for the data given the parameters and $p(\theta)$ is the prior distribution of the parameters (see below). The likelihood quantifies the similarity between model outcome and the available real-world time series. Here, the likelihood is the product over local likelihoods

$$p\big(\hat{C}_t\big|\theta\big) \sim \mathrm{StudentT}_{\nu=4}\left(\mathrm{mean} = \mathrm{C_t}(\theta),\ \mathrm{width} = \sigma\sqrt{\mathrm{C_t}(\theta)}\right).$$

quantifying the similarity between the model outcome for one time point t, $C_t(\theta)$, and the corresponding real-world data point $\hat{C}_t$. We chose the Student's t-distribution because it resembles a Gaussian distribution around the mean but features heavy tails, which make the MCMC more robust with respect to outliers [45], and thus reporting noise. The case-number-dependent width is motivated by observation noise through random subsampling [38], resulting in a variance proportional to the mean. Our likelihood neglects any noise in the dynamic process, as the SIR model is deterministic, but could be in principle extended to incorporate typical demographic noise from stochastic spreading dynamics [35, 46].

Priors that constrain model parameters

As short-term forecasts are time-critical at the onset of an epidemic, the available real-world data is typically not informative enough to identify all free parameters, or to empirically find their underlying distributions. We therefore chose informative priors on initial model parameters where possible and complemented them with uninformative priors otherwise. Our choices are summarized in Tab. II for the simple model, i.e. the SIR model with stationary spreading rate for the exponential onset phase, and in Tab. III for the full model with change points, and are discussed in the following.

TABLE II. Priors on the model parameters for the SIR model with stationary spreading rate.

Parameter	Variable	Prior distribution
Spreading rate	λ	$\mathrm{LogNormal}(\log(0.4), 0.5)$
Recovery rate	μ	$\mathrm{LogNormal}(\log(1/8), 0.2)$
Reporting delay	D	$\mathrm{LogNormal}(\log(8), 0.2)$
Initially infected	I_0	$\mathrm{HalfCauchy}(100)$
Scale factor	σ	$\mathrm{HalfCauchy}(10)$

Priors for the simple model (Table II): In order to constrain our simple model, an SIR model with stationary spreading rate for the exponential onset phase, we chose the following informative priors. Because of the ambiguity between the spreading and recovery rate in the exponential onset phase (see description of simple model), we chose a narrow log-normal prior for the recovery rate $\mu \sim \mathrm{LogNormal}(\log(1/8), 0.2)$ with median recovery time of 8 days [20]. Note that our implementation of μ accounts for the recovery of infected people and isolation measures because it describes the duration during which a person can infect others. For the spreading rate, we assume a broad log-normal prior distribution $\lambda \sim \mathrm{LogNormal}(\log(0.4), 0.5)$ with median 0.4. This way, the prior for $\lambda - \mu$ has median 0.275 and the prior for the base reproduction number ($R_0 = \lambda/\mu$) has median 3.2, consistent with the broad range of previous estimates [18, 33, 34]. In addition, we chose a log-normal prior for the reporting delay $D \sim \mathrm{LogNormal}(\log(8), 0.2)$ to incorporate both the incubation time between 1–14 days with median 5 [32] plus a delay from infected people waiting to contact the doctor and get tested.

The remaining model parameters are constrained by uninformative priors, in practice the Half-Cauchy distribution [47]. The half-Cauchy distribution $\mathrm{HalfCauchy}(\mathrm{x}, \beta) = 2/\pi\beta[1 + (\mathrm{x}/\beta)^2]$ is essentially a flat prior from zero to $O(\beta)$ with heavy tails beyond. Thereby, β merely sets the order of magnitude that should not be exceeded for a given parameter. We chose for the number of initially infected people in the model (16 days before first data point) $I_0 \sim \mathrm{HalfCauchy}(100)$ assuming an order of magnitude $O(100)$ and below. In addition, we chose the scale factor of the width of the likelihood function as $\sigma \sim \mathrm{HalfCauchy}(10)$; this choice means that the variance in reported numbers may be up to a factor of 100 larger than the actual reported number.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

TABLE III. Priors on the model parameters for the SIR model with change points and weekly reporting modulation.

Parameter	Variable	Prior distribution
Change points	t_1	Normal$(2020/03/09, 3)$
	t_2	Normal$(2020/03/16, 1)$
	t_3	Normal$(2020/03/23, 1)$
Change duration	Δt_i	LogNormal$(\log(3), 0.3)$
Spreading rates	λ_0	LogNormal$(\log(0.4), 0.5)$
	λ_1	LogNormal$(\log(0.2), 0.5)$
	λ_2	LogNormal$(\log(1/8), 0.5)$
	λ_3	LogNormal$(\log(1/16), 0.5)$
Recovery rate	μ	LogNormal$(\log(1/8), 0.2)$
Reporting delay	D	LogNormal$(\log(8), 0.2)$
Weekly modulation amplitude	f_w	Beta$(\text{mean} = 0.7, \text{std} = 0.17)$
Weekly modulation phase	Φ_w	vonMises$(\text{mean} = 0, \kappa = 0.01)$ (nearly flat)
Initially infected	I_0	HalfCauchy(100)
Scale factor	σ	HalfCauchy(10)

Priors for the full model (Table III): In order to constrain our full model, an SIR model with weekly reporting modulation and change points in the spreading rate, we chose the same priors as for the simple model but added the required priors associated with the change points. In general, we assume that each set of governmental interventions (together with the increasing awareness) leads to a reduction (and not an increase) of λ at a change point. As we cannot know yet the precise reduction factor, we adhere to assume a reduction by $\approx 50\%$, but always with a fairly wide uncertainty, so that in principle even an increase at the change point would be possible. We model the time dependence of λ as change points, and not as continuous changes because the policy changes were implemented in three discrete steps, which were presumably followed by the public in a timely fashion.

For the spreading rates, we chose log-normal distributed priors as in the simple model. In particular, we chose for the initial spreading rate the same prior as in the simple model, $\lambda_0 \sim \text{LogNormal}(\log(0.4), 0.5)$; after the first change point $\lambda_1 \sim \text{LogNormal}(\log(0.2), 0.5)$, assuming the first intervention to reduce the spreading rate by 50% from our initial estimates ($\lambda_0 \approx 0.4$) with a broad prior distribution; after the second change point $\lambda_2 \sim \text{LogNormal}(\log(1/8), 0.5)$, assuming the second intervention to reduce the spreading rate to the level of the recovery rate, which would lead to a stationary number of new infections. This corresponds approximately to a reduction of λ at the change point by 50%; and after the third change point $\lambda_3 \sim \text{LogNormal}(\log(1/16), 0.5)$, assuming the third intervention to reduce the spreading rate again by 50%. With that intervention, λ_3 is smaller than the recovery rate μ, causing a decrease in new case numbers and a saturation of the cumulative number of infections.

For the timing of change points, we chose normally distributed priors. In particular, we chose $t_1 \sim \text{Normal}(2020/03/09, 3)$ for the first change point because on the weekend of March 8, large public events, like soccer matches or fairs, were cancelled. For the second change point, we chose $t_2 \sim \text{Normal}(2020/03/16, 1)$, because on March 15, the closing of schools and other educational institutions along with the closing of non-essential stores were announced and implemented on the following day. Restaurants were allowed to stay open until 6 pm. For the third change point, we chose $t_3 \sim \text{Normal}(2020/03/23, 1)$, because on March 23, a far-reaching contact ban ("Kontaktsperre"), which includes the prohibition of even small public gatherings as well as complete closing of restaurants and non-essential stores was imposed by the government authorities. Further policy changes may occur in future; however, for now, we do not include more change points.

The change points take effect over a certain time period Δt_i for which we choose $\Delta t_i \sim \text{LogNormal}(\log(3), 0.3)$ with a median of 3 days over which the spreading rate changes continuously as interventions have to become effective. The precise duration of the transition has hardly any affect on the cumulative number of cases (Fig. 2 E-F). We assumed a duration of three days, because some policies were not announced at the same day for all states within Germany; moreover, the smooth transition also can absorb continuous changes in behavior.

The number of tests that are performed and reported vary regularly over the course of a week and are especially low during weekends. To account for this periodic variation, we modulated the number of inferred cases by the absolute value of a sine function with in total a period of 7 days. We chose this function as it is a non-symmetric oscillation, fitting the weekly variation of cases on a phenomenological level. For the amplitude of the modulation we chose a weakly informative Beta prior: $f_w \sim \text{Beta}(\text{mean} = 0.7, \text{std} = 0.17)$ and for the phase a nearly flat circular distribution: $\Phi_w \sim \text{vonMises}(\text{mean} = 0, \kappa = 0.01)$.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

Model comparison

Since change point detection entails evaluating models with different numbers of parameters, some form of fair model comparison is needed. This is necessary to compensate for the higher flexibility of more complex models, as this flexibility carries the risk of overfitting and overconfident forecasts. The standard approach to avoid over-fitting in machine learning is cross-validation, and cross validation has recently also been advocated for Bayesian model comparison (e.g. [3, 36]), especially for models employed for predictions and forecasts. Thus, one would ideally like to compare the models with different numbers of change points by the probability they assign to previously unobserved data points. Technically this is measured by their out-of-sample prediction accuracy, i.e. their log pointwise predictive density (lppd):

$$\text{lppd} = \sum_{i=1}^{N} \log \left(\int p(y_i^{\text{os}}|\theta) p_{\text{post}}(\theta) d\theta \right) \ , \tag{6}$$

where the vector $[y_1^{\text{os}}, \ldots, y_1^{\text{os}}]$ is a an out-of-sample dataset of N new data points, and where $p_{\text{post}}(\theta) = p_{\text{post}}(\theta|y, M_j)$ is the posterior distribution of the parameters, given the in-sample data y and the model M_j. In practice, the integral is approximated using a sufficient amount of samples from $p_{\text{post}}(\theta)$. However, this approach is only reasonable if a sufficient amount of out-of-sample data is available, which is not the case in the early stages of a disease outbreak. Therefore, the pointwise out-of-sample prediction accuracy was approximated using Leave-one-out cross-validation (LOO) in PyMC3 to compute equation 6 individually for each left out data point based on the model fit to the other data points. The sum of these values, multiplied by a factor of -2 then yields the leave-one-out cross-validation (LOO-CV) score. Thus, lower LOO-CV scores imply better models.

* viola.priesemann@ds.mpg.de; Authors contributed equally

[1] M. Enserink, K. Kupferschmidt, *Science* **367**, 1414 (2020).
[2] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).
[3] A. Gelman, *et al.*, *Bayesian Data Analysis, Third Edition* (CRC Press, 2013).
[4] W. O. Kermack, A. G. McKendrick, G. T. Walker, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **115**, 700 (1927).
[5] H. Hethcote, *SIAM Rev.* **42**, 599 (2000).
[6] J. Anderson, I. Lampl, I. Reichova, M. Carandini, D. Ferster, *Nat. Neurosci.* **3**, 617 (2000).
[7] N. C. Grassly, C. Fraser, *Nat Rev Microbiol* **6**, 477 (2008).
[8] R. Parshani, S. Carmi, S. Havlin, *Phys. Rev. Lett.* **104**, 258701 (2010).
[9] T. Harko, F. S. N. Lobo, M. K. Mak, *Applied Mathematics and Computation* **236**, 184 (2014).
[10] T. Britton, P. D. O'Neill, *Scand J Stat* **29**, 375 (2002).
[11] J. Lourenço, *et al.*, *eLife* **6**, e29820 (2017).
[12] N. R. Faria, *et al.*, *Sci Rep* **7**, 1 (2017).
[13] B. Shulgin, L. Stone, Z. Agur, *Bull. Math. Biol.* **60**, 1123 (1998).
[14] O. N. Bjørnstad, B. F. Finkenstädt, B. T. Grenfell, *Ecol. Monogr.* **72**, 169 (2002).
[15] L. Hufnagel, D. Brockmann, T. Geisel, *Proc. Natl. Acad. Sci. USA* **101**, 15124 (2004).
[16] A. Pandey, *et al.*, *Science* **346**, 991 (2014).
[17] R. Li, *et al.*, *Science* (2020).
[18] A. J. Kucharski, *et al.*, *The Lancet Infectious Diseases* (2020).
[19] J. Lourenco, *et al.*, *medRxiv* p. 2020.03.24.20042291 (2020).
[20] B. F. Maier, D. Brockmann, *Science* (2020).
[21] P. Bittihn, R. Golestanian, *arXiv:2003.08784* (2020).
[22] R. M. Anderson, H. Heesterbeek, D. Klinkenberg, T. D. Hollingsworth, *The Lancet* **395**, 931 (2020).
[23] J. R. Fauver, *et al.*, *medRxiv* p. 2020.03.25.20043828 (2020).
[24] A. Arenas, *et al.*, *medRxiv* p. 2020.03.21.20040022 (2020).
[25] O. Mitjà, *et al.*, *The Lancet* (2020).
[26] S. L. Chang, N. Harding, C. Zachreson, O. M. Cliff, M. Prokopenko, *arXiv:2003.10218* (2020).
[27] W. Bock, *et al.*, *medRxiv* p. 2020.03.25.20043109 (2020).
[28] C. Gros, R. Valenti, K. Valenti, D. Gros, *arXiv:2004.00493* (2020).
[29] V. Zlatić, I. Barjašić, A. Kadović, H. Štefančić, A. Gabrielli, *arXiv:2003.08479* (2020).
[30] J. Dehning, *et al.*, *ArXiv200401105 Version 1* (2020).
[31] J. Dehning, *et al.*, https://github.com/Priesemann-Group/Covid19_inference_forecast v2020.05.01, Zenodo (2020).
[32] S. A. Lauer, *et al.*, *Ann Intern Med* (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

[33] J. Zhang, *et al.*, *medRxiv* p. 2020.03.19.20039107 (2020).
[34] Y. Liu, A. A. Gayle, A. Wilder-Smith, J. Rocklöv, *J Travel Med* **27** (2020).
[35] M. A. Muñoz, *Rev. Mod. Phys.* **90**, 031001 (2018).
[36] A. Vehtari, A. Gelman, J. Gabry, *Stat Comput* **27**, 1413 (2017).
[37] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, *arXiv:2002.06563* (2020).
[38] J. Wilting, V. Priesemann, *Nat. Commun.* **9**, 2325 (2018).
[39] Y.-C. Chen, P.-E. Lu, C.-S. Chang, T.-H. Liu, *arXiv:2003.00122* (2020).
[40] E. Dong, H. Du, L. Gardner, *The Lancet Infectious Diseases* **0** (2020).
[41] M. Chinazzi, *et al.*, *Science* **368**, 395 (2020).
[42] J. Salvatier, T. V. Wiecki, C. Fonnesbeck, *PeerJ Comput. Sci.* **2**, e55 (2016).
[43] M. D. Hoffman, A. Gelman, *J. Mach. Learn. Res.* **15**, 1593 (2014).
[44] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, D. M. Blei, *J. Mach. Learn. Res.* **18**, 430 (2017).
[45] K. L. Lange, R. J. A. Little, J. M. G. Taylor, *J. Am. Stat. Assoc.* **84**, 881 (1989).
[46] S. di Santo, P. Villegas, R. Burioni, M. A. Muñoz, *Phys. Rev. E* **95** (2017).
[47] A. Gelman, *Bayesian Anal.* **1**, 515 (2006).
[48] Johns Hopkins data snapshot is available at https://raw.githubusercontent.com/CSSEGISandData/COVID-19/bda67e3db0e8dca4540297633d431a8021c035c8/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv.

ACKNOWLEDGMENTS

We thank Tim Friede, Theo Geisel, Knut Heidemann, Moritz Linkmann, Matthias Loidolt, and Vladimir Zykov for carefully and promptly reviewing our work internally. We thank the Priesemann group - especially Daniel Gonzalez Marx, Fabian Mikulasch, Lucas Rudelt & Andreas Schneider - for exciting discussions and for their valuable comments. We thank the colleagues of the Göttingen Campus, with whom we were discussing the project and the COVID-19 case forecast in the past weeks very intensively: Heike Bickeböller, Eberhard Bodenschatz, Wolfgang Brück, Alexander Ecker, Andreas Leha, Ramin Golestanian, Helmut Grubmüller, Stephan Herminghaus, Reinhard Jahn, Norbert Lossau & Simone Scheithauer. We thank Nils Bertschinger for stimulating discussion on Bayesian modeling and model comparison. **Funding:** All authors received support from the Max-Planck-Society. JD and PS acknowledge funding by SMARTSTART, the joint training program in computational neuroscience by the VolkswagenStiftung and the Bernstein Network. JZ received financial support from the Joachim Herz Stiftung. MW is employed at the Campus Institute for Dynamics of Biological Networks funded by the VolkswagenStiftung. **Author contributions:** JD, JZ, MW, MW, VP designed research. JD, JZ, PS conducted research. JD, JZ, PS, JPN, MW, MW, VP analyzed the data. JD, PS, JPN created figures. All authors wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Both data [48] and analysis code [31] are available online. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

LIST OF SUPPLEMENTARY MATERIALS:

Table S1 – S2
Fig S1 – S7

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

SUPPLEMENTARY MATERIALS:

TABLE S1. Overview of model parameters.

Variable	Parameter
λ	Spreading rate
μ	Recovery rate
$\lambda^* = \lambda - \mu$	Effective spreading rate
λ_i	Spreading rate after i-th intervention
t_i	Time of i-th intervention
f_w	Amplitude of weekend corrections
Φ_w	Phase shift of weekend correction
σ	Scale factor of the width of Student's t-distribution
N	Population size (83.000.000)
S_t	Susceptible at time t
I_t	Infected at time t
R_t	Recovered at time t
Δt	Time step
$R_t^{\text{new}} = \mu \Delta t I_{t-1}$	New recoveries at time t
$I_t^{\text{new}} = \lambda \Delta t \frac{S_{t-1}}{N} I_{t-1}$	New infections at time t
$C_t = I_{t-D}^{\text{new}}$	New reported cases at time t
D	Delay of case detection

TABLE S2. Model comparison: Using leave-one-out (LOO) cross-validation, we compare our original "SIR main" model that features a weekend modulation (to account for fewer reported cases during weekends) with other model variants. Remarkably, the median inferred effective growth rate λ^* after the last change-point is very similar for all model variants. For full details on the model variants, see the figure captions in the SI: (i) SEIR-like with explicit incubation time, Fig. S3. (ii) SIR excluding the weekend modulation, Fig. S4. (iii) Sensitivity analysis by applying wider priors to different parameters, Figs. S5, S6, S7. Lower LOO-scores represent a better match between model and data (pLOO is the effective number of parameters). †For the SEIR-like model, the magnitude of the effective growth rate is not directly comparable because of the explicit incubation period.

Model	# c-pts.	λ_0^*	λ_1^*	λ_2^*	λ_3^*	LOO-score	pLOO
SIR main	0	0.03				927 ± 9	8.31
SIR main	1	0.21	-0.03			819 ± 16	13.46
SIR main	2	0.30	0.11	-0.03		796 ± 17	12.53
SIR main	3	0.30	0.12	0.02	-0.03	787 ± 17	13.42
SIR without weekend modulation	3	0.31	0.13	0.04	-0.03	807 ± 17	9.65
SIR with wider delay prior	3	0.30	0.12	0.02	-0.04	787 ± 17	14.15
SIR with wider change point priors	3	0.30	0.12	0.02	-0.04	787 ± 17	14.29
SIR with wider transient length prior	3	0.29	0.12	0.02	-0.04	785 ± 17	13.92
SEIR-like†	3	1.95	0.45	0.05	-0.12	782 ± 17	10.71

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

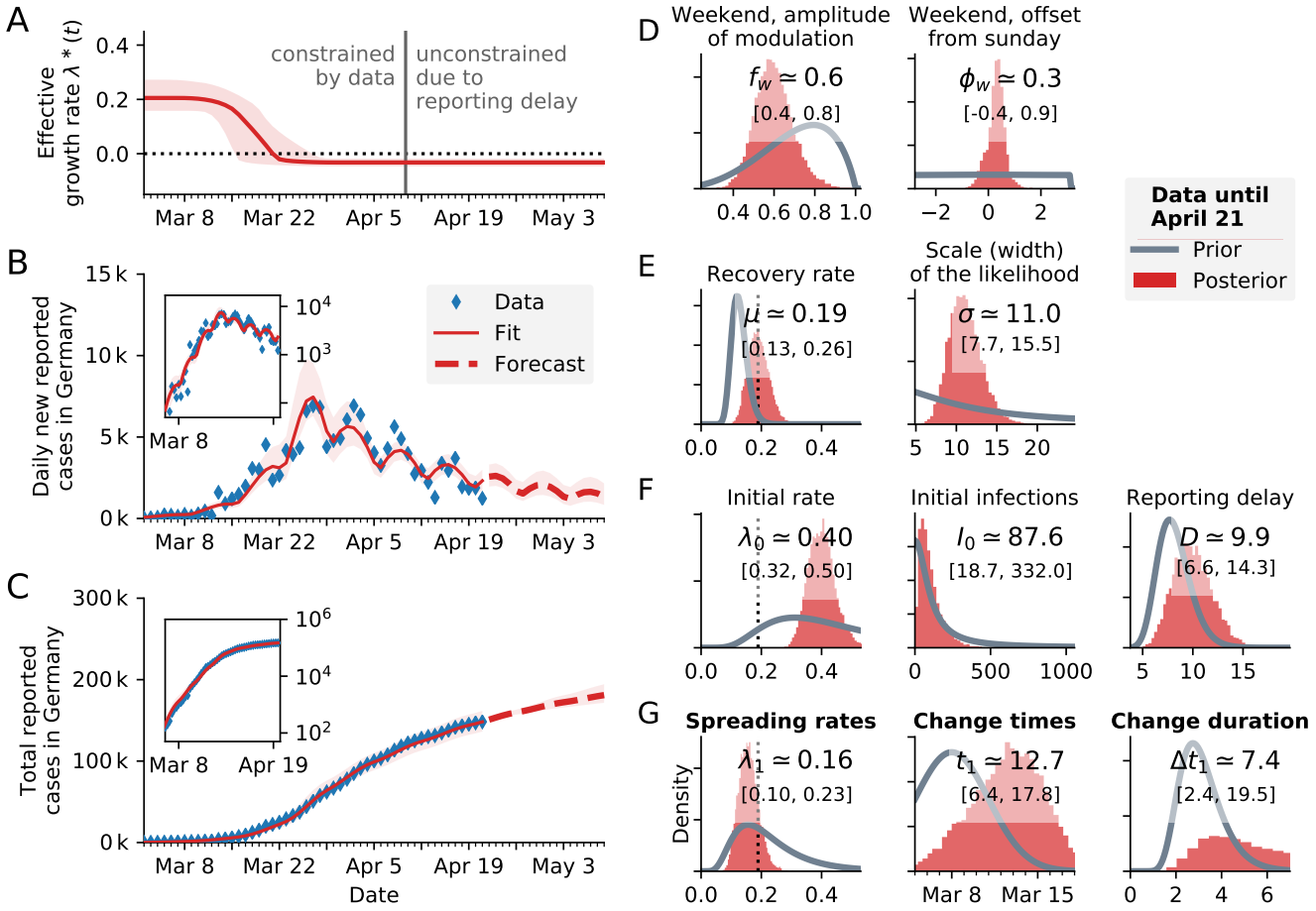

FIG. S1. **Model comparison:** Change-point detection as in Fig. 3 (three change points, main text) with the **same model** "SIR main" but only **one change point**. With only one change point, the model cannot describe the data well after April 1. **A:** Time-dependent model estimate of the effective growth rate $\lambda^*(t)$. **B:** Forecast and comparison of daily new reported cases with the model fit, **inset:** Same on log-lin scale. **C:** Same as B but for cumulative (total) cases. **D–G:** Posterior distributions from the change point detection (red) compared to prior distributions (gray). Please refer to Fig. 1 (main text) for a more detailed description of the distributions.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

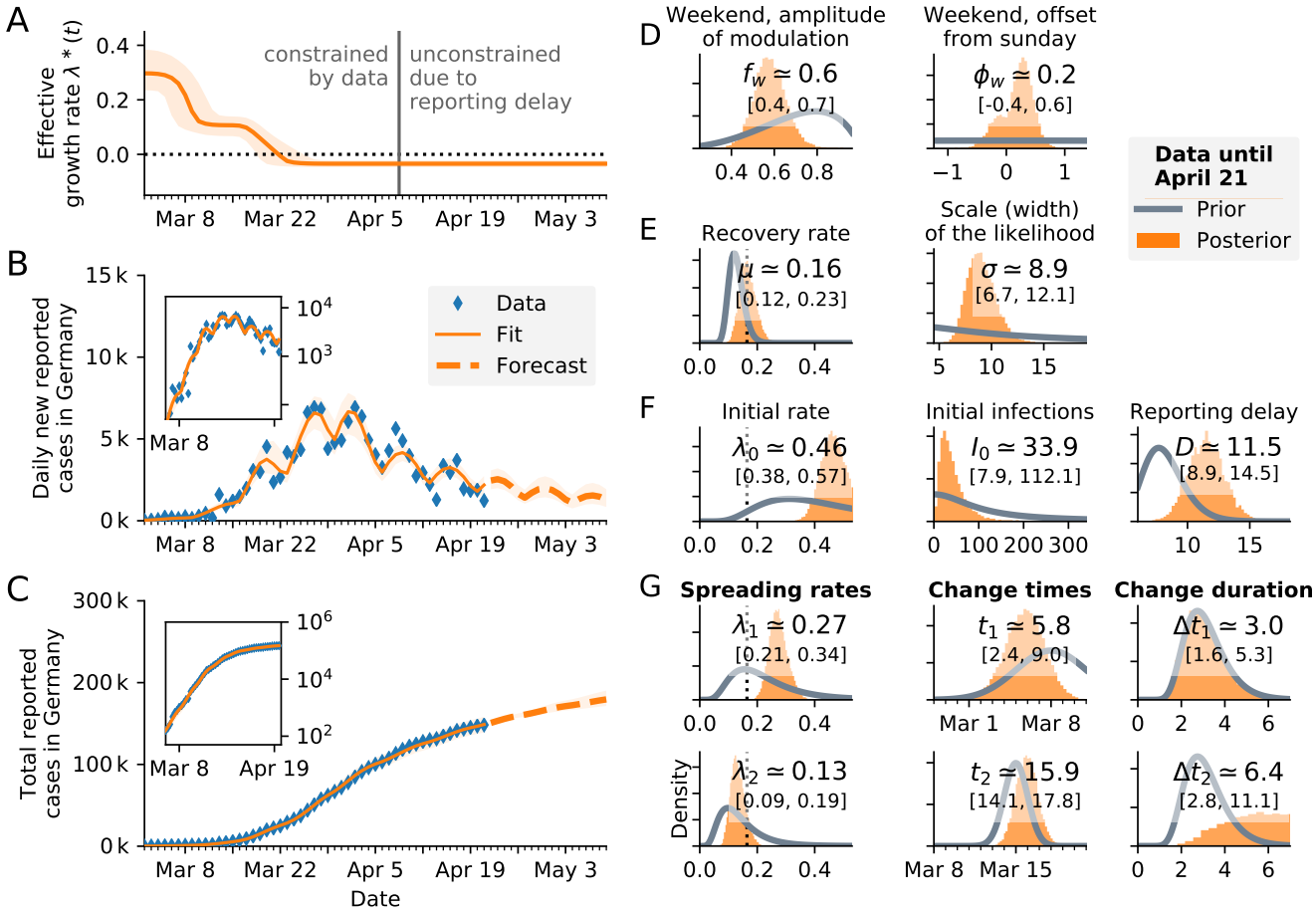

FIG. S2. **Model comparison:** Change-point detection as in Fig. 3 (three change points, main text) with the **same model** "SIR main" but with **two change points**. With two change points, the onset time of the second change point is close to the (middle) one inferred when using the 3-c.p. model. However, the effective growth rate λ^* after the respective last c.p. is the same in both cases, $\lambda^* = -0.03$ (see also Table S2). **A:** Time-dependent model estimate of the effective growth rate $\lambda^*(t)$. Two change points are clearly visible. **B:** Forecast and comparison of daily new reported cases with the model fit, **inset:** Same on log-lin scale. **C:** Same as B but for cumulative (total) cases. **D–G:** Posterior distributions from the change point detection (orange) compared to prior distributions (gray). Please refer to Fig. 1 (main text) for a more detailed description of the distributions.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

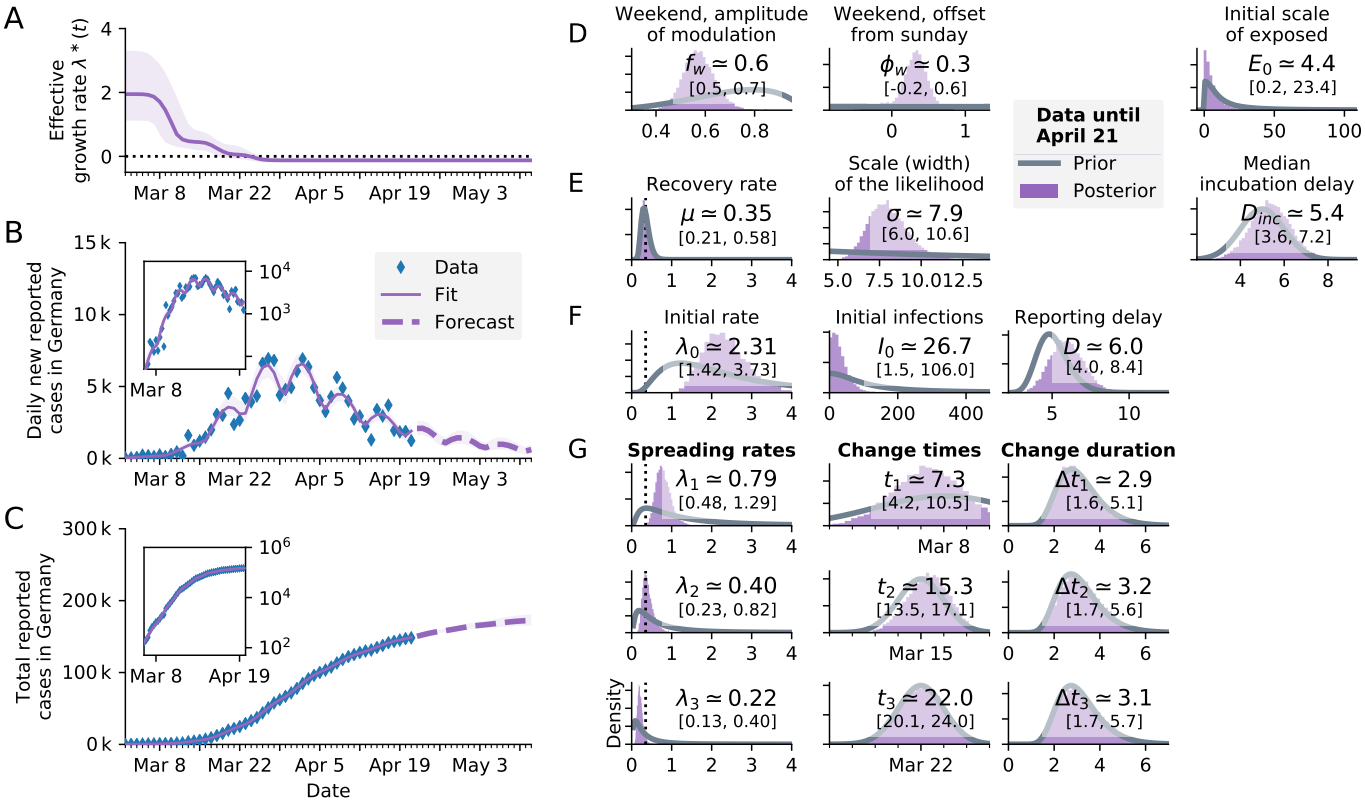

FIG. S3. **Model comparison:** Change-point detection as in Fig. 3 (three change points, main text) but with a more involved **SEIR-*like*** model and **three change points**, code available online [31]. Arguably, the SEIR-like model provides are more realistic (but also more complex) description of virus propagation. It yields a slightly better (lower) LOO-score than our "SIR main" model in the cross-validation, Table S2. However, inferred parameters are compatible with the simpler model and the inferred dynamics are similar. **Model details**: The SEIR-like model builds on our "SIR main" model (which includes a weekend correction, see Methods). The SEIR-like model features an explicit log-normal incubation period and a lognormal reporting delay. The incubation period is implemented as a discrete convolution of multiple exposed pools with a lognormal kernel. The discrete lognormal kernel is parameterized as follows (to match the characteristic incubation time of COVID-19 [32]): median D_{inc}, scale parameter 0.418 and normalized to 1. The median is a free parameter with prior Normal$(5, 1)$ (days) [32]. The reporting delay is implemented in a similar manner: as a convolution of the number of new cases with a lognormal kernel and the scale parameter is fixed to 0.3. In order to match the total delay of the main model (between the infection and the observation), the median D is a free parameter with prior LogNormal$(5, 0.2)$ (days). Note that because of the lognormal-distributed incubation period in the SEIR-like model, the spreading and recovery rates are not directly comparable to the SIR models. We correspondingly adapted the respective priors [17]: We adapted the prior median of the recovery rate μ to $1/3$ with a scale factor of 0.3, and the prior median for λ_0, λ_1, λ_2, and λ_3 to 2.0, 1.0, 0.5 and 0.25, respectively, with a scale parameter of 1 each. **A:** Time-dependent model estimate of the effective growth rate $\lambda^*(t)$. **B:** Comparison of daily new reported cases and the model (inset: log-lin scale). **C:** As B but for total (cumulative) reported cases. **D–G**: Prior and posterior distributions of all free parameters.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

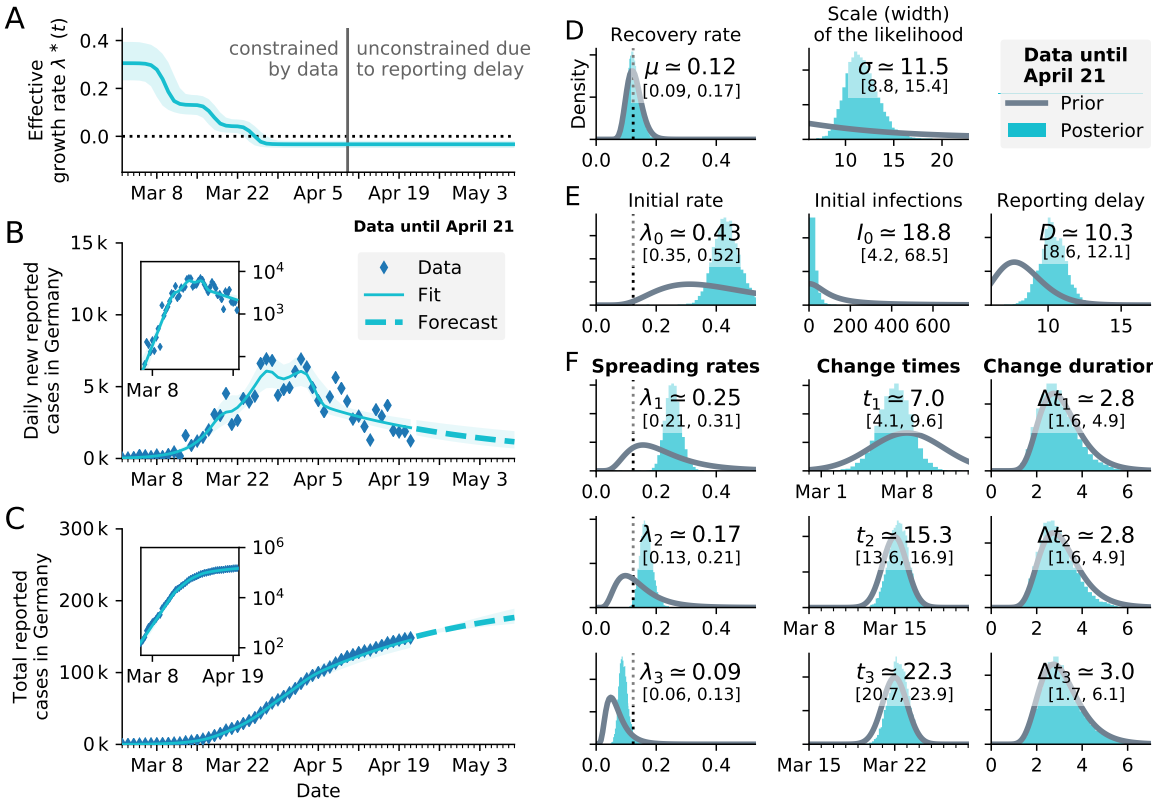

FIG. S4. **Model comparison:** Change-point detection as in Fig. 3 (three change points, main text) with an **SIR model that excludes the weekend modulation** but features the **three change points**. In this version of the model, we excluded the assumption that daily new reported cases depend on the weekday (which is modeled as an absolute sine with an amplitude and a phase shift as inferred parameters in the main model). While the inferred parameters from the model that excludes the weekend modulation (in particular rates and onset times of change points) match the model that includes the modulation, the LOO-scores of the cross-validation are worse, Table S2. Especially in panel B it becomes clear that the (empirically motivated) dependence of reported on cases on the weekday is justified. Without the modulation, the model fit does not capture the periodic changes in the data. **A:** Time-dependent model estimate of the effective growth rate $\lambda^*(t)$. **B:** Comparison of daily new reported cases and the model (inset: log-lin scale). **C:** As B but for total (cumulative) reported cases. **D–F**: Prior and posterior distributions of all free parameters.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

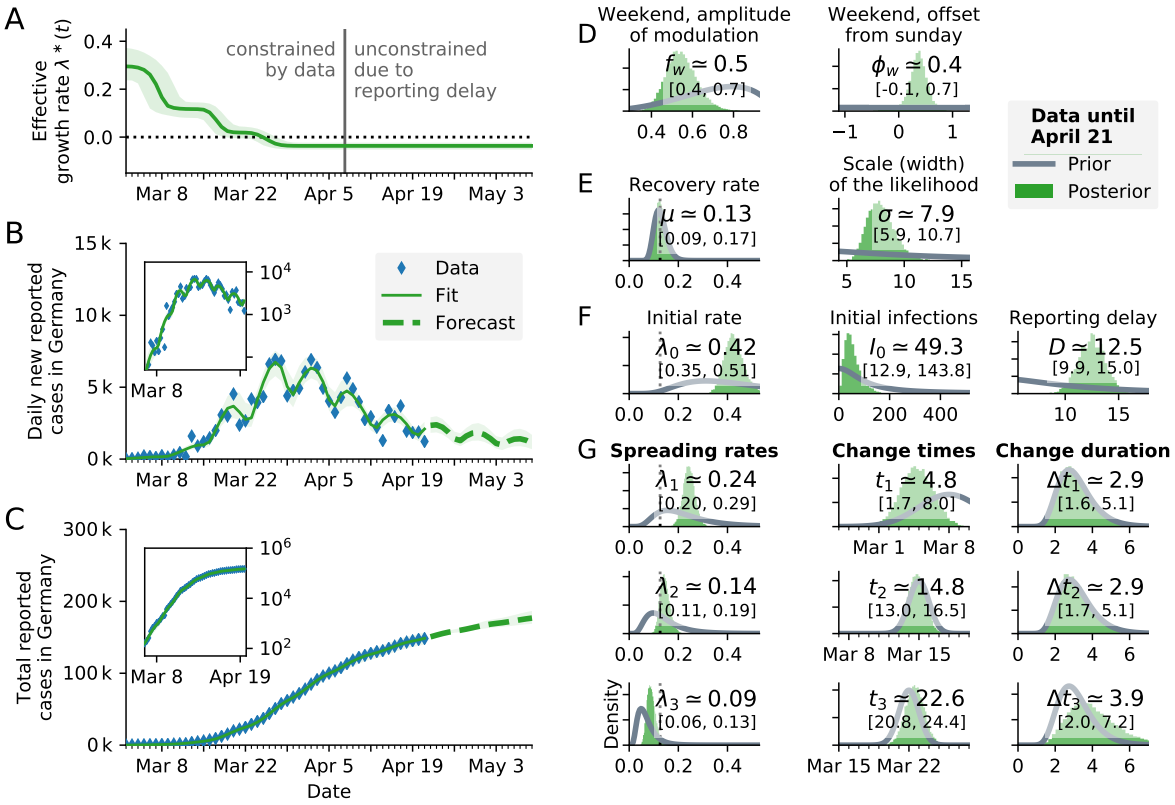

FIG. S5. **Sensitivity analysis:** Change-point detection as in Fig. 3 (main text, three change points, same model) but with **a prior for the reporting delay that is 4 times wider** (panel F, third column). All parameters and change points are constrained by data. In particular, the posterior distribution of the reporting delay is slightly wider than with the original prior, but it is well constrained by data. **A:** Time-dependent model estimate of the effective growth rate $\lambda^*(t)$. **B:** Comparison of daily new reported cases and the model (inset: log-lin scale). **C:** As B but for total (cumulative) reported cases. **D–G**: Prior and posterior distributions of all free parameters.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

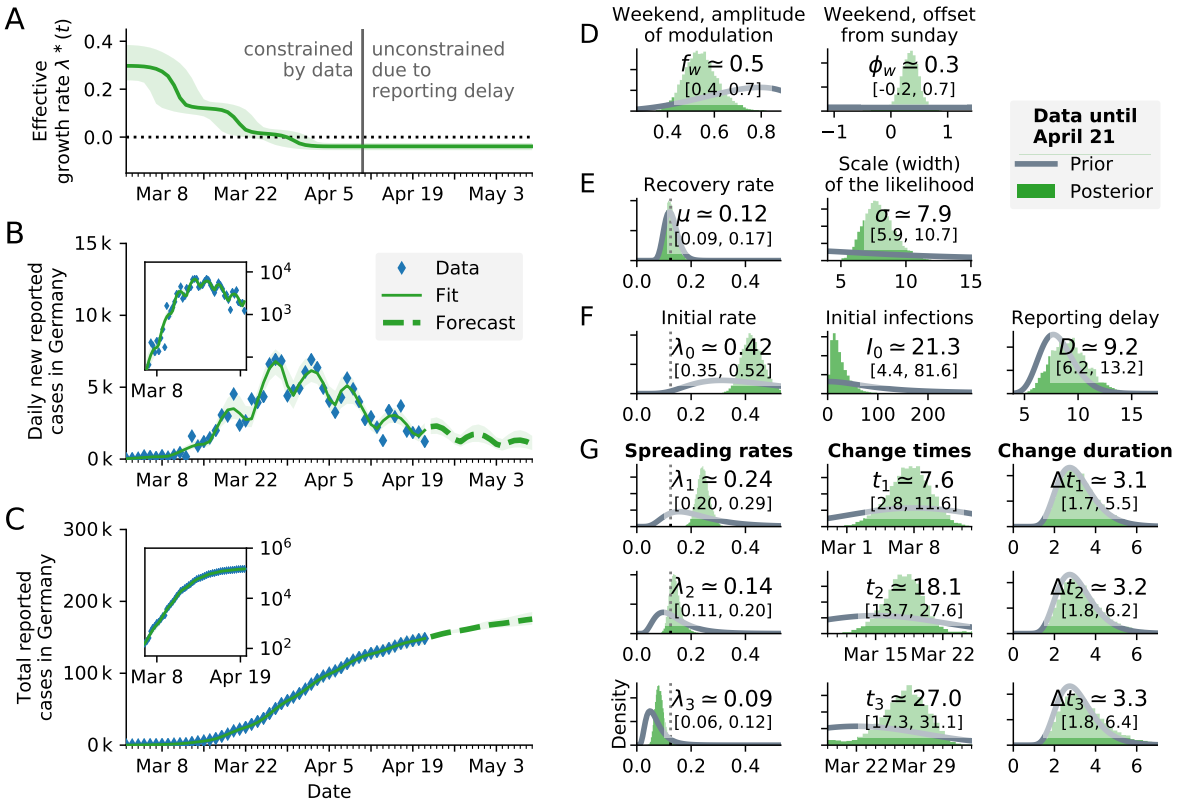

FIG. S6. **Sensitivity analysis:** Change-point detection as in Fig. 3 (main text, three change points, same model) but with **a prior for the change times that is 14 days wide**, instead of ~ 2 days, (panel G, second column). All parameters and change points are constrained by data but the change points occur at later times compared to the original priors. **A:** Time-dependent model estimate of the effective growth rate $\lambda^*(t)$. **B:** Comparison of daily new reported cases and the model (inset: log-lin scale). **C:** As B but for total (cumulative) reported cases. **D–G**: Prior and posterior distributions of all free parameters.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.02.20050922; this version posted May 6, 2020. 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. All rights reserved. No reuse allowed without permission.

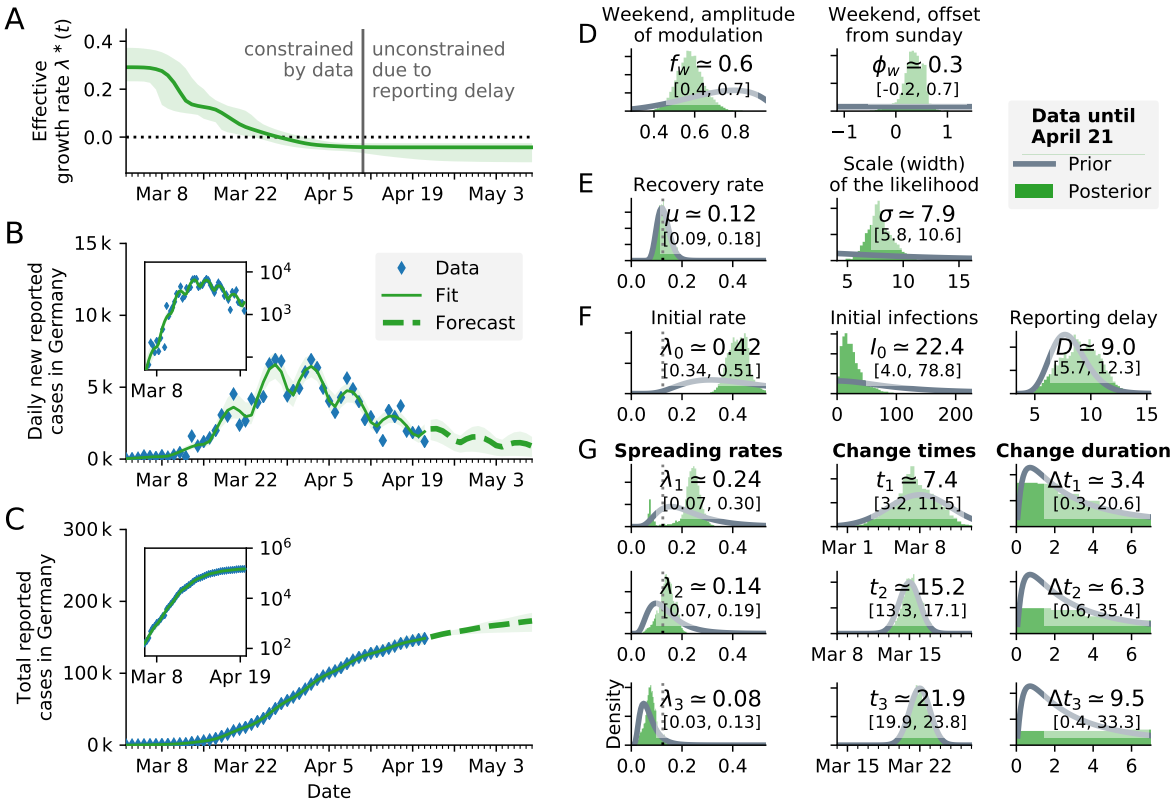

FIG. S7. **Sensitivity analysis:** Change-point detection as in Fig. 3 (main text, three change points, same model) but with **a prior for the change duration that is 4 times wider** (panel G, third column). The duration of the first change point Δt_1 is robust to the wider prior; the data constrains the posterior. The durations of the second and third change point, Δt_2 and Δt_3 are not constrained by the data but they depend on our chosen priors. This is also visible in the lack of plateaus in the effective growth, panel A. However, the inferred spreading rate λ and the forecast are *not sensitive* to the wider priors. **A:** Time-dependent model estimate of the effective growth rate $\lambda^*(t)$. **B:** Comparison of daily new reported cases and the model (inset: log-lin scale). **C:** As B but for total (cumulative) reported cases. **D–G**: Prior and posterior distributions of all free parameters.