

# Detection of transmission change points during unlock-3 and unlock-4 measures controlling COVID-19 in India

**Manisha Mandal**

*Department of Physiology, MGM Medical College, Kishanganj-855107, India*

Email: debmanisha@rediffmail.com, ORCID: <https://orcid.org/0000-0002-9562-5534>

**Shyamapada Mandal\***

*Department of Zoology, University of Gour Banga, Malda-732103, India*

Email: samtropmed@gmail.com, ORCID: <https://orcid.org/0000-0002-9488-3523>

\*Corresponding author: Email: samtropmed@gmail.com

## ABSTRACT

Documentation in scientific literature is not available on prospective evaluation of the efficiency of the unlock measure related to COVID-19 transmission change points in India, projecting the infected population, planning suitable measures related to future interventions and lifting of restrictions so that the economic settings are not damaged beyond repair. We have applied SIR model and Bayesian approach combined with Monte Carlo Markov algorithms on the Indian COVID-19 daily new infected cases from 1 August 2020 to 30 September 2020. We showed that the COVID-19 epidemic declined after implementing unlock-4 measure and the identified change-points were consistent with the timelines of announced unlock-3 and unlock-4 measure, on 1 August 2020 and 1 September 2020, respectively, effectiveness of which were quantified as the change in both effective transmission rates (100% reduction) and the basic reproduction number attaining 1, implying measures taken to control and mitigate the COVID-19 epidemic in India managed to flatten and recede the epidemic curve.

**Key words:** COVID-19, epidemiological parameters in India, unlock-3 and unlock-4, SIR model, Bayesian inference, Monte Carlo Markov sampling.

## 1. Introduction

To contain COVID-19 spread in India, strong phasic lockdowns were implemented leading to reduction of human contact to a maximum 55%, and 34% at the end of lockdown, followed by stratified unlock measures with gradual return to activities, controlling social contacts to 19% reduction, as on 30 September, 2020 (IHME, 2020). India, currently is the world's second-worst-hit country with nearly 11.7 million COVID-19 infections including more than 98,000 deaths, as on 30 September 2020 (COVID19India, 2020). Until COVID-19 is completely eradicated, and effective treatment or vaccine become available, non-pharmaceutical intervention policies are the key public health options to control the epidemics (Varghese et al., 2020). During the evolution of COVID-19, India implemented lockdowns in four phases from 24 March to 31 May 2020 as containment and mitigation measure, followed by unlocks in four phases from 1 June to 30 September 2020, featured by conditional relaxations of restrictions outside containment zones in graded manner, to minimise the negative economic and social consequences of strict lockdown measures (MOHWF, GoI, 2020). With the escalating case numbers and prolonged COVID-19 epidemic situation in India, the present study is an investigation to determine, if within the currently implemented non-pharmaceutical strategies, taking into account the simulated stochastic SIR model of transmission dynamics, is effective in curbing the spread of COVID-19. For this purpose, we made posterior inference on transmission rate  $\lambda$ , recovery rate  $\mu$ , reproduction number  $R_0$ , number of initially infected people  $I_0$ , reporting delay  $D$ , width of India's epidemic curve observed daily infected cases only its best fit and simulated effective transmission rate  $\lambda^* = \lambda - \mu$ , based on data-driven likelihood updates of prior settings. We determined also the change-points in disease transmission and investigated the effectiveness of unlock-3 and unlock-4 measures,

with respect to their strength, timing and duration. We used an open source probabilistic programming in Python code PyMC3 with theano to compute gradients via automatic differentiation variational inference (ADVI), and followed model interpretation on German COVID-19 data (Dehning et al., 2020), based on GitHub repository (Priesemann-Group, 2020), to analyse the recent COVID-19 pandemic situation in India with emphasis on unlock-3 and unlock-4 measure.

## 2. Methods

### 2.1. Data sources

The data on ongoing new daily and cumulative COVID-19 cases in India were retrieved from Johns Hopkins University Centre for Systems Science and Engineering dashboard, up to September 30, 2020 (CSSE, 2020). The codes for this research article pertaining to the analysis of unlock-3 and unlock-4 situation in India was run on Jupyter notebook using PymC3=3.8 and was based on GitHub repository (Priesemann-Group, 2020), by importing python-based data analysis toolkit (pandas); libraries for working with arrays (numpy), plotting (matplotlib), scientific and technical computing (scipy), and multi-dimensional arrays (theano); modules including Basic date and time types (datetime), System-specific parameters and functions (sys), and Python object serialization (pickle); package for Bayesian statistical modeling and probabilistic machine learning with MCMC and ADVI algorithms (PymC3) (Kucukelbir et al., 2007).

### 2.2. SIR model

The SIR model was based on time-varying cumulative number of COVID-19 cases, where the total population size ( $N$ ) was categorized into three mutually exclusive infection levels, assuming that any infectious person ( $I$ ), is likely to contact any susceptible person ( $S$ ), and later recovered ( $R$ ), so that  $N = S + I + R$ . The dynamics of the pandemic in India was modelled using the following three differential equations:

$$\dot{S} = -\frac{\lambda SI}{N}, \dot{I} = \frac{\lambda SI}{N} - \mu I, \dot{R} = \mu I, \quad (1)$$

where  $\lambda$  represents the transmission rate of the infected people to infect susceptible people and  $\mu$  denotes the recovery rate of the infected people to recover (Kermack et al., 1927). This is solved by using a forward finite-difference scheme (Carcione et al., 2014):

$$S^{n+1} = S^n + dt\left(-\frac{\lambda S^n I^n}{N^n}\right), I^{n+1} = I^n + dt\left(\frac{\lambda S^n I^n}{N^n} - \mu I^n\right), R^{n+1} = R^n + dt(\mu I^n), \quad (2)$$

where  $n$  is a natural number which divides time  $t$  in  $n$  discrete  $dt$  time steps,  $t = ndt$ .

The fraction of maximum number of infected people,  $(I_{max}) = 1 + \frac{1}{R_0} \left( \ln \frac{1}{R_0} - 1 \right)$ , where  $R_0 = \frac{\lambda}{\mu}$  and the fraction of people remaining susceptible to infection ( $S_{inf}$ ) is related to  $R_0$  by:  $R_0 = \frac{\ln S_{inf}}{S_{inf} - 1}$ . The overall infection attack rate ( $IAR$ ) defined as the fraction of the population that eventually becomes infected is related to  $R_0$  by:  $R_0 = -\frac{\ln(1-IAR)}{IAR}$ . (3)

### 2.3. SEIR model

The SEIR model is an extension of the SIR with an added exposure ( $E$ ) period due to the reported incubation period of COVID-19 during which individuals are not yet infectious (Hethcote, 2000). The SEIR models the total population size ( $N$ ) divided into four mutually exclusive infection stages,  $N = S + E + I + R$ , and is based on following differential equations:

$$\dot{S} = -\frac{\lambda SI}{N}, \dot{E} = \frac{\lambda SI}{N} - \sigma E, \dot{I} = \sigma E - \mu I, \dot{R} = \mu I, \quad (4)$$

where,  $\sigma$  is the rate at which individuals in incubation become infectious. The differential equation is solved as (Carcione et al., 2014):

$$S^{n+1} = S^n + dt\left(-\frac{\lambda S^n I^n}{N^n}\right), E^{n+1} = E^n + dt\left(\frac{\lambda S^n I^n}{N^n} - \sigma E^n\right), \\ I^{n+1} = I^n + dt(\sigma E^n - \mu I^n), R^{n+1} = R^n + dt(\mu I^n) \quad (5)$$

## 2.4. Model inclusions

A reporting delay  $D$ , was incorporated in becoming infected ( $I^{new}$ ) and being reported, such that the daily reported cases  $R_t$  at any time  $t$  was given by (Dehning et al., 2020),  $R_t = I_{t-D}^{new}$ . To examine if there was any weekend effect on daily reported case numbers, a periodic sine function was assigned to the reporting fraction  $f(t)$  expressed as,

$$R_t = I_{t-D}^{new}(1 - f(t)), \quad (6)$$

$$f(t) = (1 - f_w) \cdot (1 - |\sin \frac{\pi t}{7} - 0.5\phi_w|) \quad (7)$$

where  $f_w$  and  $\phi_w$  are the weekly modulation amplitude and phase respectively (Dehning et al., 2020).

## 2.5. Bayesian Inference

For a statistical model,  $p(x|\theta)$ , that reflects our beliefs about  $x$  given  $\theta$ , with the prior distribution  $\pi(\theta)$ , on an observed data  $D_n = \{X_1, \dots, X_n\}$ , the posterior distribution is expressed as (Bayes, 1763; Box et al., 1992):  $p(\theta, X_1, \dots, X_n) = \frac{p(X_1, \dots, X_n, \theta)\pi(\theta)}{p(X_1, \dots, X_n)}$

$$\frac{L_n(\theta)\pi(\theta)}{c_n} \propto L_n(\theta)\pi(\theta) \quad (8)$$

where,  $L_n(\theta) = \prod_{i=1}^n p(X_i, \theta)$  is the likelihood function.

The likelihood is a measure of the goodness of fit between model prediction and the observed data on reported case numbers, applied hereby using Student-t distribution. The evidence

$$c_n = p(X_1, \dots, X_n) = \int p(X_1, \dots, X_n, \theta)\pi(\theta)d\theta = \int L_n(\theta)\pi(\theta)d\theta. \quad (9)$$

The Bayesian posterior interval estimate for  $a$  and  $b$ ,  $\alpha \in (0,1)$ ,  $C = (a, b)$ , is given by,

$$P(\theta \in C|D_n) = \int_a^b p(\theta|D_n)d\theta = 1 - \alpha \quad (10)$$

The Bayesian predictive distribution is

$$p(x, D_n) = \int p(x, \theta)p(\theta|D_n)d\theta \quad (11)$$

The inferences about a function  $\tau = g(\theta)$ , so that cumulative distribution function for  $\tau$  is  $H(t, D_n) = P(g(\theta) \leq t, D_n)$

$$= \int_A p(\theta|D_n)d\theta \quad (12)$$

where  $A = \{\theta: g(\theta) \leq t\}$ ; the posterior density is  $p(\tau, D_n) = H'(\tau, D_n)$ .

## 2.6. Priors

The prior distribution settings for the model parameter estimation were made by incorporating LogNormal values of  $\lambda$ ,  $\mu$ , and  $D$  and half-Cauchy distribution for  $I_0$ , and  $\sigma$  (Table 1 and 2). The priors on change points in transmission rate were based on announcements of applied intervention including unlock-3 on 1 August 2020 and unlock-4 on 1 September 2020.

## 2.7. Markov Chain Monte Carlo (MCMC) sampling

This method is essentially a Monte Carlo sampling with multiple Markov chains, used to approximate the posterior distribution of model parameters by including ADVI, 1000 tuning steps with NUTS (No U Turn Sampling) algorithm (Hoffman et al., 2014) for each of four chains, and R-hat diagnostics for equilibrated chain convergence of model parameters (Vehtari et al., 2019). A sequence of random variables  $\{X_1, \dots, X_n\}$ , on a discrete state space is called a Markov chain if

$$p(X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_1 = x_1) = p(X_t = x_t, X_{t-1} = x_{t-1}) \quad (13)$$

We wanted to find a setting of a parameter  $x \in R$ , such that the expectation  $h(x) \equiv E_t(H_t, x) = 0$ , the updates were applied as (Hoffman et al., 2014):

$$x_{t+1} \leftarrow \mu - \frac{\sqrt{t}}{\gamma} \frac{1}{t+t_0} \sum_{i=1}^t H_i; \bar{x}_{t+1} \leftarrow \eta_t x_{t+1} + (1 - \eta_t)x_t, \quad (14)$$

where  $t$  is iteration,  $\eta_t$  is the step size schedule,  $\mu$  is a freely chosen point that the iterates  $x_t$  are shrunk towards,  $\gamma > 0$  is a free parameter that controls the amount of shrinkage towards  $\mu$ ,  $t_0 \geq 0$  is a free parameter that stabilizes the initial iterations of the algorithm,  $\eta_t \equiv t^{-\kappa}$  is a step size schedule satisfying

the conditions,  $\sum_t \eta_t = \infty$ ;  $\sum_t \eta_t^2 < \infty$ . The step size parameter was set for NUTS using stochastic optimization with vanishing adaptation of the primal-dual algorithm. For each iteration we defined the statistic  $H_t^{NUTS}$  and its expectation when the chain reached equilibrium as (Hoffman et al., 2014):

$$H_t^{NUTS} \equiv \frac{1}{|B_t^{final}|} \sum_{\theta, r \in B_t^{final}} \min \left\{ 1, \frac{p(\theta, r)}{p(\theta^{t-1}, r^{t,0})} \right\}; h^{NUTS} \equiv E_t[H_t^{NUTS}]; \quad (15)$$

where  $B_t^{final}$  is the set of all states explored during the final doubling of iteration  $t$  of the Markov chain;  $\theta^{t-1}, r^{t,0}$  are the initial position and sampled momentum for the  $t^{th}$  iteration of the Markov chain;  $H^{NUTS}$  is the average acceptance probability. We applied in the above updates equation:  
 $H_t \equiv \delta - H^{NUTS}$  and  $x \equiv \log \epsilon$  for the step size  $\epsilon$  to combine  $h^{NUTS} = \delta$  for any  $\delta \in (0, 1)$ .

## 2.8. Model comparison

For model fit and comparison using MCMC, following computations were made using *LOO* (leave one out) package in PyMC3: the Bayesian *LOO* estimate of the expected log pointwise predictive density (*ELPD-LOO*) for a new point, standard error (*SE*) of *ELPD-LOO*, the difference between *ELPD-LOO* and the non-cross-validated log posterior predictive density (*pLOO*) interpreted as the effective number of parameters (Vehtari et al., 2017). Lower *LOO* scores indicated better consistency between models. *LOO* scores with  $SE < 1$  represented model compatibility while *LOO* scores  $SE > 1$  indicated mismatch between the models.

For data  $D_n = (X_1, \dots, X_n)$ ,  $ELPD \sum_{i=1}^n \int p_t(\sim X_i) \log p(\sim X_i | X) d\sim X_i$ , where  $p_t(\sim X_i)$  is the distribution of the true data generating process for  $\sim X_i$ , which is approximated by cross-validation with  $LPD = \sum_{i=1}^n \log p(X_i | X) = \sum_{i=1}^n \log \int p(X_i | \theta) p(\theta | X_i) d\theta$  (16)

The *LPD* computed with  $S$  draws from a posterior distribution =  $\widehat{LPD}$  = computed log pointwise predictive density =  $\sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(X_i | \theta^s) \right)$  (17)

The *ELPD-LOO*, calculated by cross-validation by running the model  $n$  times is  $\sum_{i=1}^n \log p(X_i, X_{-i}) d\theta$ , where  $p(X_i, X_{-i}) = \int p(X_i, \theta) p(\theta, X_i) d\theta$ . (18)

## 3. Results and Discussion

The median posterior distribution of COVID-19 epidemiological parameters using SIR model (from Eqs. 1 and 2) combined with Bayesian inference (generated by Eqs. from 8 to 15) during unlock-3 were  $\lambda = 0.15$  (0.11 – 0.22),  $\mu = 0.14$  (0.10 – 0.21),  $R_0 = 1.07$  (1.05 – 1.10),  $I_0 = 327139$  (220028 – 450505),  $D = 8$  days (5.5 – 11.8),  $\sigma = 18.7$  (13.4 – 26.1),  $\lambda^* = 1(1 - 2)$ , the values in bracket indicate 95% confidence-intervals *CI*s (Fig. 1), according to Eqs. from 16 to 18,  $ELPD-LOO = -305.07$ ,  $pLOO = 3.04$ , computed from 2000 by 30 log-likelihood matrix (Table 3). The corresponding values during unlock-4 were  $\lambda = 0.15$  (0.10 – 0.20),  $\mu = 0.15$  (0.10 – 0.20),  $R_0 = 1.00$ ,  $I_0 = 642500$  (455336 – 922684),  $D = 8.0$  days (5.4 – 11.9),  $\sigma = 19.1$  (13.5 – 26.8),  $\lambda^* = (-1 - 0)$  (Fig. 2),  $ELPD-LOO = -299.14$ ,  $pLOO = 2.74$ , computed from 2000 by 29 log-likelihood matrix (Table 3). The median estimates of  $R_0$  decreased from 1.07 in the unlock-3 period to 1.00 in the unlock-4 period, with SIR model, indicating slowing down of the spread of the disease. The  $R_0$  was reported as 1.14 at the end of August 2020, and 1.12 in the mid of September 2020 in India, with stationary-time-series auto regressive integrated moving average model (Yadav et al., 2020). The mathematical models help to determine the effect of preventive policies against COVID-19, primarily by maintaining the reproduction number  $R_0 < 1$  to inhibit further spread of infection, whereas  $R_0 > 1$  indicate continuation of the epidemics, which fade away when the transmissibility is reduced by  $(1 - \frac{1}{R_0})$  (Ferguson et al., 2020; Mandal et al., 2020; Pan et al., 2020).

The median daily COVID-19 infected cases in India at the end of unlock-3 reached  $\approx 80,000$ , that increased about 1.09 times, to  $\approx 87,500$  at the end of unlock-4 (Figs. 1A and 2A), while the median cumulative infected cases increased 1.8 fold to 6.3 million at the end of unlock-4 from 3.5 million at

the end of unlock-3 (Figs. 1B and 2B). The difference in daily infected cases showed approximately 1.125 fold change to 5,625 during unlock-4 on 30 September, 2020, from 5000 during unlock-3 on 31 August, 2020 (Figs. 1C and 2C). The COVID-19 daily cases increased during unlock-3 but decreased during unlock-4, though with a higher end point (Figs. 1A and 2A). The real-time daily and cumulative cases were consistent with SIR-model displaying linear-semi-logarithmic variation. A decreasing first order differences of the logarithm of the cumulative cases over time indicated exponential growth as found for India, whereas a constant trend indicated logarithmic growth of the epidemic curve as seen in the US, for the period from 17 September 2020 to 1 October 2020 (Baruah, 2020).

The priors and posteriors exhibited different  $\lambda$ ,  $I_0$ , and  $\sigma$  (Figs. 1D, 1F, 1H, 2D, 2F, and 2H) implicating informative feature of the observed data; and matched  $\mu$  and  $D$ , indicating dependency of the observed data on prior informatives (Figs. 1E, 1G, 2E, and 2G), in unlock-3 and unlock-4 exhibiting exponential transmission rates. The  $\lambda^*$  became zero in unlock-4, from 10% in unlock-3 (Figs. 1I and 2I), indicating post unlock-4 effect through inhibition of new infections. The  $\sigma$ , a measure of goodness of fit between  $\lambda$  and  $\mu$ , showed equipotential line for the maximum likelihood (Figs. 1J and 2J), implicating  $\lambda^*$  as an important and independent regulator of COVID-19 transmission dynamics.

Both daily and cumulative infected cases remain unaltered until the duration of  $D$  and change-point were over, beyond which both continued to rise, build on the hypothesis that ongoing unlock phase and its post-effect prevailed, but continuation of pre-unlock situation caused decline in both, the effect being more significant in the daily cases (Fig. 3A and 3D). The daily as well as cumulative case numbers showed rising trend in post unlock-3 scenario (Fig. 3A), however, the onset timings of intervention had no effect on case numbers (Fig. 3B). With declining new cases and rising cumulative cases (Fig. 3D), during unlock-4, advancing or delaying change-point onset by five days showed insignificant difference in cumulative cases (Fig. 3E). The cumulative cases remained unchanged with the change in transient duration of intervention, the new cases showed similar variation as  $\lambda$  (Fig. 3C and 3F).

The SIR-model parameters with two identified change-points without weekend effect showed that the first change-point matched with the timelines of publicly announced strategies around 1 August 2020, when unlock-3 began, coinciding with continued closure of educational institutions and banned social gatherings, permission of interstate transport, besides release of night curfews. This change-point featured  $\lambda_1 = 0.17$  (0.12 – 0.21) that unfolded over 2.9 days (1.6 – 5.4) (Fig. 4F). The second change-point was detected around 1 September 2020, which coincided with the announcement of unlock-4, featured by lockdown measures remaining in force in containment zones, some activities permitted outside containment zones with reopening of metro-rail in graded manner, small gatherings permitted, continued compulsion of face-masking in public. The second change point had  $\lambda_2 = 0.15$  (0.10 – 0.19) that unfolded over 3.3 days (1.7 – 6.7) (Fig. 4F). With  $\lambda_0 = 0.16, \sigma = 15.6, D = 9.6$  days,  $\mu = 0.15$ , the change-points were quantified as  $\lambda_1^* = \lambda_1 - \mu = 0.02, R_0 = 1.13$  during unlock-3 and  $\lambda_2^* = \lambda_2 - \mu = 0, R_0 = 1$  during unlock-4, implying effectiveness of unlock-4 measure bringing 100% reduction of  $\lambda^*$  and decline of the COVID-19 epidemic (Fig. 4). Thus the effectiveness of an intervention modelled as Bayesian change points could help us interpret the impact of different control measures and to include them into forecasts. Previously, Bayesian inference of COVID-19 change points correlating with social distancing restrictions were applied using the example of Germany (Dehning et al., 2020). Similar model was used to detect and assess latent events associated with spreading rates in South Africa (Mbuva et al., 2020) and the US (Jiang et al., 2020).

Compared to the two-change-point model without weekend-correction ( $ELPD-LOO = -687.67, SE = 8.00, pLOO = 6.71$ , computed from 2000 by 68 log-likelihood matrix) (Table 3), the  $ELPD-LOO$  for the SIR-model, also computed from 2000 by 68 log-likelihood matrix, with two change-points and weekend-modulation was higher by 30.76 ( $ELPD-LOO = -656.91, SE = 6.82, pLOO = 7.43$ ) (Table 3), [ $\lambda_0 = 0.16, \lambda_1 = 0.17, \lambda_2 = 0.15, \mu = 0.15, \sigma = 10.2, D = 12.8$  days,  $\lambda_1^* = \lambda_1 - \mu = 0.02, \lambda_2^* = \lambda_2 - \mu = 0, f_w = 0.8, \phi_w = 2.7$ ] (Fig. 5). Sampling were run for 4 chains with 1000 tunes and 500 draw iterations so that a total of 6000 draws occurred. Thus, in our

study the lower  $LOO$  in the former, differing by  $1.18SE$ , indicated higher consistency of the two-change-points SIR model with data excluding the weekend-factor that implied homogenous reporting of daily new cases through the entire week irrespective of weekend effect. However, higher number of COVID-19 daily new cases were reported during weekdays compared to weekends in Germany, substantiated by lower  $LOO$  score in weekend-effect model compared to that without weekend-effect model (Dehning et al., 2020).

Application of SEIR model based on Eqs. 4 and 5 with two-change-points and weekend-modulation as per Eqs. 6 and 7 ( $f_w = 0.8$ ,  $\phi_w = 2.7$ ), exhibited greater negativity of the effective transmission rate with unlock-4 compared to unlock-3 measure ( $\lambda_1^* = \lambda_1 - \mu = -0.03$ ,  $\lambda_2^* = \lambda_2 - \mu = -0.07$ ) (Fig. 6), and higher transmission and recovery rates ( $\lambda_0 = 0.36$ ,  $\lambda_1 = 0.30$ ,  $\lambda_2 = 0.26$ ,  $\mu = 0.33$ ) (Fig. 6) compared to the SIR model (Fig. 5). Cross-validation showed the  $ELPD-LOO$  for the SEIR-model ( $-650.94$ ,  $SE = 7.16$ ,  $pLOO = 13.66$ , computed from 2000 by 68 log-likelihood matrix) slightly greater (by 5.97) than that of the corresponding SIR-model (Fig. 5), with  $1SE(0.34)$  higher variation (Table 3), indicating considerably greater evidence for SIR model with respect to SEIR in explaining current COVID-19 data in Indian context. Similarly, SIR model displayed superior goodness of fit to the SEIR on South African data whereas SEIR produced a slightly better  $LOO$  score than the SIR main model on German data (Dehning et al., 2020; Mbuva et al., 2020).

Estimating the effect of change-points is vital for priors settings that help to anticipate the effects of any impending change points and accordingly make future projections. The SIR-model with one change point (Fig. 7) centred around the implementation of unlock-3 on 1 August 2020 showed superior goodness of fit with the COVID-19 observed data in India compared to the model with two change-points announcement of unlock-3 and unlock-4 on 1 August 2020 and 1 September 2020 respectively (Fig. 5); both models examined over the period from 25 July 2020 to 30 September 2020, with weekend-modulation. This was evident from the lower (by 43.14)  $ELPD-LOO$  score with one change point ( $-700.05$ ,  $SE = 6.14$ ,  $pLOO = 6.24$ ) (Table 3), ( $\lambda_0 = 0.17$ , and  $\lambda_1 = 0.16$ ,  $\mu = 0.15$ ,  $\sigma = 17.6$ ,  $f_w = 0.8$ ,  $\phi_w = 2.6$ ,  $D = 8.7$  days,  $\lambda_1^* = \lambda_1 - \mu = 0.01$ ) (Fig. 7) that fitted the observed data better compared to that with two change-points (Fig. 5) with  $< 1 SE (= 0.68)$  lower difference (Table 3). This was also clear from the simulation effect of hypothetical inventions on future COVID-19 cases in India, which showed that continuation of pre-unlock situation would have caused further decrease in both daily new and cumulative infected cases (Fig. 3A and 3D), implying that extension of stricter social-distancing measures would have been advantageous in reducing cases. Association of COVID-19 transmission in India in the context of containment measures demonstrated lower  $ELPD-LOO = -305.39$  ( $SE = 4.58$ ,  $pLOO = 3.45$ ) for unlock-3, computed from 2000 by 30 log-likelihood matrix, and for the unlock-4 computed from 2000 by 29 log-likelihood matrix as  $-299.04$  ( $SE = 4.08$ ,  $pLOO = 2.75$ ) (Table 3), using Eqs. 16 to 18. SIR models with three change-points described the data better than fewer change points on German data and SIR model with two change points was the best fit on South African data, as exhibited by the  $LOO$  cross-validation and all change points coinciding with respective government interventions (Dehning et al., 2020; Mbuva et al., 2020).

Surveillance of COVID-19 pandemic involved a reporting delay factor (range: 7 – 13 days) that was composed of testing delay between the incubation period of the virus (time period for the symptoms to develop following infection with the virus, with median estimates of 5 – 6 days) and the testing date (1 – 3 days); an additional delay occurred between the testing date and results date (1 – 4 days) (Lauer et al., 2020). The  $D$  extrapolated from SIR-model with two change-points (12.8 days) (Fig. 5F) versus one change point (8.7 days) with weekend modulation (Fig. 7F), SIR model with two change points without weekend modulation (10.2 days) (Fig. 5F), SEIR model with two change points with weekend modulation (reporting plus incubation delay 10.9 days) (Fig. 6F), indicate consistency with the above mentioned summated delay factor; the inherent difference within the obtained values might be due to the experimental conditions and model types adopted. The median change duration in all such situations were estimated to be around 3 days that were necessary to enact interventions, in the form of continuing

and lifting of restrictions based on containment areas. Thus, the reporting delay combined with the change duration ranged from 11 to 16 days, which represented the time gap required to identify any change points in infected case numbers that in conjunction with the effective COVID-19 transmission rate, help to determine pertinent containment measures. The India COVID-19 daily infected case numbers using SIR and SEIR models were estimated to be around 75,000 and 72,000 respectively; the cumulative infected case numbers using both models were estimated at 8000,000 as of October 18, 2020 (for the period from 25 July 2020 to 30 September 2020); the effective transmission rate stabilized at less than zero,  $R_0 = 1$  and 0.79 using SIR and SEIR models respectively (Figs. 5 and 6).

#### 4. Conclusion

Overall, the SIR model, including weekend-modulation and one-change point, with continuation of intervention similar to the unlock 3 situation was favoured over other models. However, the finding from the SIR model including weekend-modulation and two-change points that,  $\mu$  was 0.15,  $\lambda^*$  was 10% and 'zero' during unlock-3 and unlock-4 respectively, implied unlock-4 measure brought 100% reduction of  $\lambda^*$ , beginning around 5 September, 2020 (Fig. 5A-G), indicating new recoveries exceeding the new infections. Therefore, the epidemic curve is expected to decline to the baseline level, when the effective transmission rate becomes remarkably negative leading to sustained dwindling of new infections, provided no re-infection occurs and non-pharmaceutical interventions such as voluntary face-masking, physical-distancing, in addition to government measures including graded lockdown intervention in containment zones are maintained.

**Author's contribution:** Manisha Mandal and Shyamapada Mandal jointly designed the study, analysed and interpreted the data, discussed and wrote the manuscript.

**Funding Source:** Nil

**Declaration of competing interests:** There is no conflict of interest by the authors.

#### References

1. Baruah, H.K., 2020. A Numerical Study of the Current COVID-19 Spread Patterns in India, the USA and the World. MedRxiv Preprint. Available at. <https://doi.org/10.1101/2020.10.05.20206839>.
2. Bayes, T., 1763. Bayesian inference of phylogeny. *Phil. Trans. Roy. Soc.* 330.
3. Box, G.E.P., Tiao, G.C., 1992. Bayesian inference in statistical analysis. Wiley-Interscience, USA, 1-608.
4. Carcione, J. M., 2014. In wave Fields in Real Media: Theory and numerical simulation of wave propagation in anisotropic, anelastic, porous and electromagnetic media (3<sup>rd</sup> edn) Elsevier Science, Amsterdam, 38, 1-690
5. CSSE (Center for Systems Science and Engineering), 2020. COVID-19 Dashboard. Johns Hopkins University. Available at. <https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>.
6. COVID19India, 2020. Available at. <https://www.covid19india.org/>
7. Dehning, J., Zierenberg, J., Spitzner, F.P., Wibral, M., Neto, J.P., Wilczek, M., Priesemann, V., 2020. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science.* 369, 161-171.
8. Ferguson, N., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., ..., 2020. Report 9: Impact of non-pharmaceutical interventions (NPIS) to reduce COVID-19 mortality and healthcare demand. Imperial College London, <https://doi.org/10.25561/77482>.
9. Hethcote, H.W., 2000. The mathematics of infectious diseases. *SIAM Rev.* 42, 599-653.

10. Hoffman, M.D., Gelman, A., 2014. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Machine Learning Res.* 30, 1351-1381.
11. Institute for Health Metrics and Evaluation (IHME), 2020. COVID-19 mortality, infection, testing, hospital resource use, and social distancing projections. University of Washington, USA.
12. Jiang, S., Zhou, Q., Zhan, X., Li, Q., 2020. BayesSMILES: Bayesian segmentation modeling for longitudinal epidemiological studies. *MedRxiv Preprint*. Available at <https://doi.org/10.1101/2020.10.06.20208132>.
13. Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond.* 115(172), 700-721.
14. Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., Blei, D.M., 2017. Automatic differentiation variational inference. *J. Mach. Learn. Res.* 18, 1-45.
15. Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., ..., 2020. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* 172, 577-582.
16. Mandal, M., Mandal, S., 2020. COVID-19 pandemic scenario in India compared to China and rest of the world: a data driven and model analysis. *MedRxiv Preprint*. Available at <https://doi.org/10.1101/2020.04.20.20072744>.
17. Mbuva, R., Marwala, T., 2020. Bayesian inference of COVID-19 spreading rates in South Africa. *PLoS ONE*. 15(8): e0237126. <https://doi.org/10.1371/journal.pone.0237126>.
18. Ministry of Health and Family Welfare. Government of India, 2020. Available at <https://www.mohfw.gov.in/>
19. Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., Huang, J., ..., 2020. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA* 323(19), 1889-1890.
20. Priesemann-Group, 2020. Available at [https://github.com/Priesemann-Group/covid19\\_inference\\_forecast](https://github.com/Priesemann-Group/covid19_inference_forecast).
21. Varghese, G.M., John, R., 2020. COVID-19 in India: Moving from containment to mitigation. *Indian J. Med. Res.* 151(2), 136-139.
22. Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics Computing*. 27(5), 1413-1432.
23. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., Bürkner, P-C., 2020. Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC. *Bayesian Analysis*. Available at <https://doi.org/10.1214/20-BA1221>.
24. Yadav, S., Yadav, P.K., 2020. The peak of COVID-19 in India. *MedRxiv Preprint*. Available at <https://doi.org/10.1101/2020.09.17.20197087>.

**Table 1. Prior distribution settings for India unlock-3 and unlock-4 SIR model parameters with fixed transmission rate**

| Parameter          | Variable  | Prior distribution                  |
|--------------------|-----------|-------------------------------------|
| Transmission rate  | $\lambda$ | <i>LogNormal</i> [ $\log(0.4)$ , 0. |
| Recovery rate      | $\mu$     | <i>LogNormal</i> [ $\log(0.125)$    |
| Reporting delay    | $D$       | <i>LogNormal</i> [ $\log(8)$ ; 0.2  |
| Initially infected | $I_0$     | <i>HalfCauchy</i> (100)             |
| Scale factor       | $\sigma$  | <i>HalfCauchy</i> (10)              |

**Table 2. Prior distribution settings for SIR model parameters with changing transmission rates and weekend reporting factor**

| Parameter                   | Variable     | Prior distribution   |
|-----------------------------|--------------|--|
| Change points               | $t_1$        | 01 August 2020   |
|                             | $t_2$        | 01 September 2020  |
| Change duration             | $\Delta t_i$ | <i>LogNormal</i> [ $\log(3)$ , 0.3]                          |
| Spreading rate              | $\lambda_0$  | <i>LogNormal</i> [ $\log(0.4)$ , 0.5]                        |
|                             | $\lambda_1$  | <i>LogNormal</i> [ $\log(0.16)$ , 0.5]                       |
|                             | $\lambda_2$  | <i>LogNormal</i> [ $\log(0.15)$ , 0.5]                       |
| Recovery rate               | $\mu$        | <i>LogNormal</i> [ $\log(0.125)$ , 0.2]                      |
| Reporting delay             | $D$          | <i>LogNormal</i> [ $\log(8)$ , 0.2]                          |
| Initially infected          | $I_0$        | <i>HalfCauchy</i> (100)                                      |
| Scale factor                | $\sigma$     | <i>HalfCauchy</i> (10)                                       |
| Weekly modulation amplitude | $f_w$        | $\beta(\text{mean} = 0.7, \text{standard deviation} = 0.17)$ |
| Weekly modulation phase     | $\phi_w$     | <i>VonMises</i> ( $\text{mean} = 0, k = 0.01$ )              |

**Table 3. Bayesian LOO estimate of the ELPD for model comparison**

| Model   | ELPD-<br>LOO | p-LOO | StandardE<br>rror | n<br>log-likelihood<br>matrix |
|---|--------------|-------|-------------------|-------------------------------|
| SIR during unlock-3                                 | -305.07      | 3.04  | 4.39              | 30                            |
| SIR during unlock-4                                 | -299.14      | 2.70  | 3.90              | 29                            |
| SIR during unlock-3 and intervention evaluation     | -305.39      | 3.45  | 4.58              | 30                            |
| SIR during unlock-4 and intervention evaluation     | -299.04      | 2.75  | 4.08              | 29                            |
| SIR with 2 change points and without weekend factor | -687.67      | 6.71  | 8.00              | 68                            |
| SIR with 2 change points and with weekend factor    | -656.91      | 7.43  | 6.82              | 68                            |
| SEIR with 2 change points and with weekend factor   | -650.94      | 13.66 | 7.16              | 68                            |
| SIR with 1 change point and with weekend factor     | -700.05      | 6.24  | 6.14              | 68                            |

## Figure legends:

### Fig. 1. Bayesian inference of India SIR COVID-19 epidemic model parameters during unlock-3 from 1 to 31 August 2020.

Exponential growth of (A) daily infected cases  $y = 53694.96e^{0.01050x}$  and (B) cumulative infected cases  $y = 1702723.32e^{0.025x}$  with decreasing variation in rate of change of the logarithmic cumulative case, average 0.0108 (0.0104 – 0.0116); (C) difference in daily infected cases between fit and data; priors (red) based posterior (cyan) inference of (D) transmission rate  $\lambda$ , (E) recovery rate  $\mu$ , (F) number of initially infected people  $I_0$ , (G) reporting delay  $D$ , (H) scale factor of likelihood  $\sigma$  between observed daily infected cases (blue) and its median fit with 95% CI (orange), (I) effective transmission rate  $\lambda^* = \lambda - \mu$ ; (J) log-likelihood combination of transmission and recovery rates with maximum value (black line) and data non-convergence (white rectangle). Serial interval (SI) across India during unlock-3 was 27.43 days (26.06 – 29.25); recovery rate (RR) 71.9% (76.35 – 79.16); fatality rate (FR) 1.93% (1.91 – 1.97); reproduction number  $R_0$  1.07; IAR 14.2% people per million population,  $I_{max} = 0.00218$  during unlock-3, and  $S_{inf} = 0.9345$  after unlock-3, computed from Eqn. 3.

### Fig. 2. Bayesian inference of India SIR COVID-19 epidemic model parameters during unlock-4 from 1 to 30 September 2020.

Implementation of unlock-4 measure was consistent with exponential (A) decay of daily infected cases  $y = 87133.43e^{-0.00049x}$  and (B) growth of cumulative infected cases  $y = 3700990.43e^{0.018x}$  with diminished rate and continued dwindling in rate of change of the logarithmic cumulative case, average 0.0080 (0.0075 – 0.0084); (C) difference in daily infected cases between fit and data, with declining width of likelihood; priors based posterior inference of (D)  $\lambda$ , (E)  $\mu$ , (F)  $I_0$ , (G)  $D$ , (H)  $\sigma$ , (I)  $\lambda^*$ ; (J) log-likelihood between  $\lambda$  and  $\mu$ . Compared to unlock-3, unlock-4 strategies featured higher SI 37.35 days (36.43 – 42.21), enhanced RR 79.23% (78.44 – 80.01), reduced FR 1.64% (1.62 – 1.67), decreased  $R_0$  1.00, increased IAR at 18.6%,  $I_{max}$  decreased to zero during unlock-4;  $S_{inf}$  increased to 1.075 after unlock-4.

### Fig. 3. Effect of strength, timing and duration of unlock-3 and unlock-4 measures on infected case numbers.

Each unlock measure featured re-opening of activities outside Containment Zones in phased manner and strict lockdown in containment zones only. Specifically unlock-3 removed night curfews, reopened recreational centres like gymnasiums and yoga centres. Unlock4 reopened metro rail in graded manner, and permitted limited gatherings. Under the extended relaxations, social distancing were hypothesized to be  $\sim 0.9$  factor stronger and  $\sim 0.9$  factor milder respectively, as a pre- and post-effect of unlock measure. (A) With respect to the strength of unlock-3, the transmission rate remained nearly the same but the daily and cumulative infected cases increased. Perpetuation of pre-unlock scenario would have caused decline in all. (B) Delaying the onset timings of unlock-3 measure showed insignificant change in cumulative case numbers: unlock-3 starting on 1 August 2020 (green), 5 days later (magenta), or 5 days earlier (gray). (C) The cumulative cases remained unchanged with change in transient [immediate (brown), intermediate (green), long (cyan)] duration of unlock-3, the new cases showed similar variation as  $\lambda$ . (D), (E), (F) same effect as (A), (B), (C) but with declining daily infected cases in unlock-4 starting 1 September 2020.

**Fig. 4. Identification and quantification of change points in COVID-19 transmission rate of India during unlock-3 and unlock-4 phase using SIR model without weekend effect.**

Time-series SIR model fit estimates of the (A) effective transmission rate  $\lambda^*(t)$ , (B) daily infected cases compared to the observed data, and (C) cumulative infected cases compared to the observed data. Inset shows semi-log plots. Underreporting factor on 18 October 2020, for daily and cumulative infected cases were and respectively, using SIR model. (D-G) Priors and posterior distribution of model parameters, values are expressed in median and 95% CIs of the posteriors. The SIR exhibited 5.97 lower LOO score than SEIR, both models with 2 change points and a weekend factor. Thus SIR model represented better consistency with observed data compared to SEIR model.

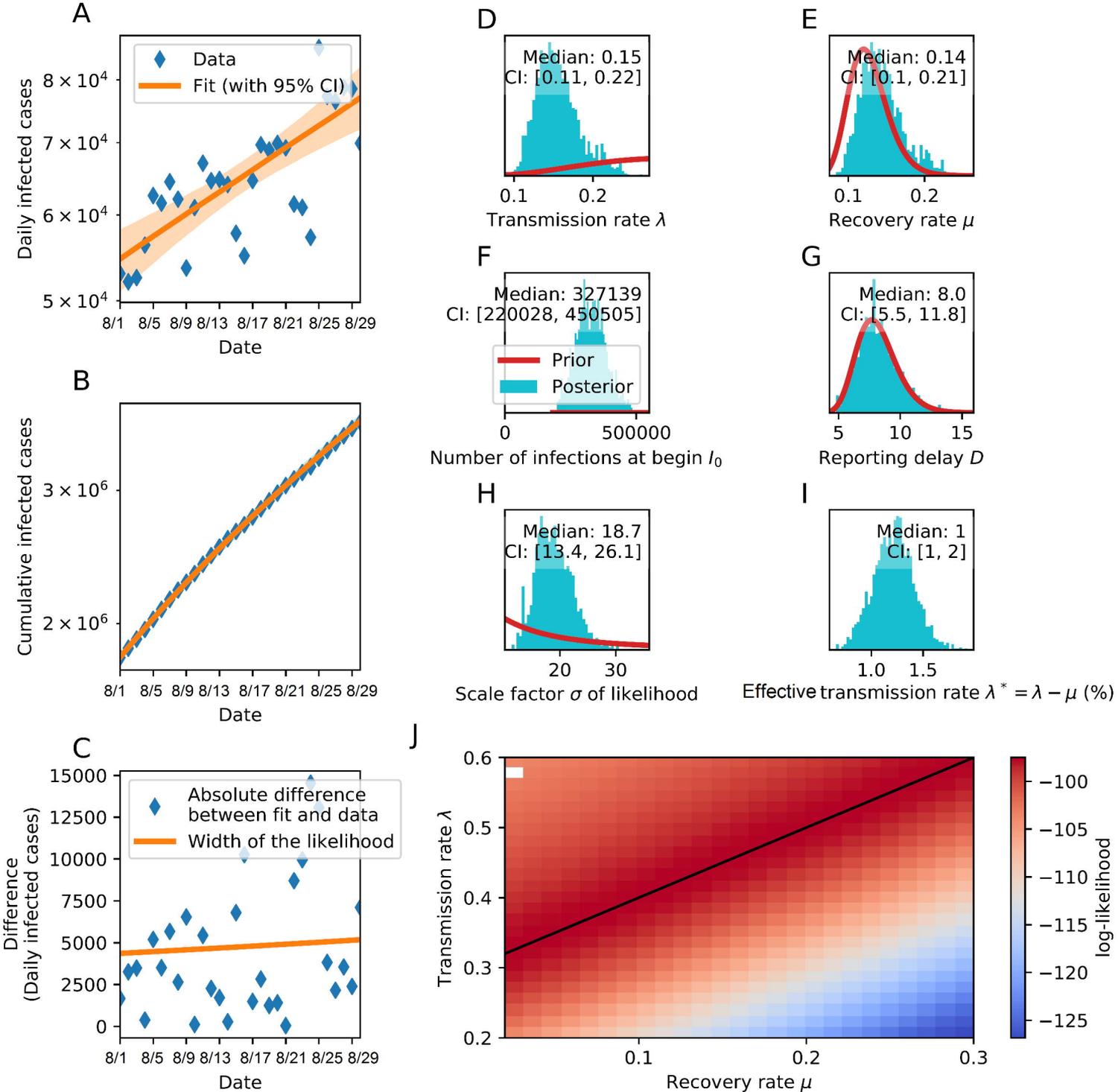
**Fig. 5. Identification and quantification of change points in COVID-19 transmission rate of India during unlock-3 and unlock-4 phase using SEIR model with weekend effect.**

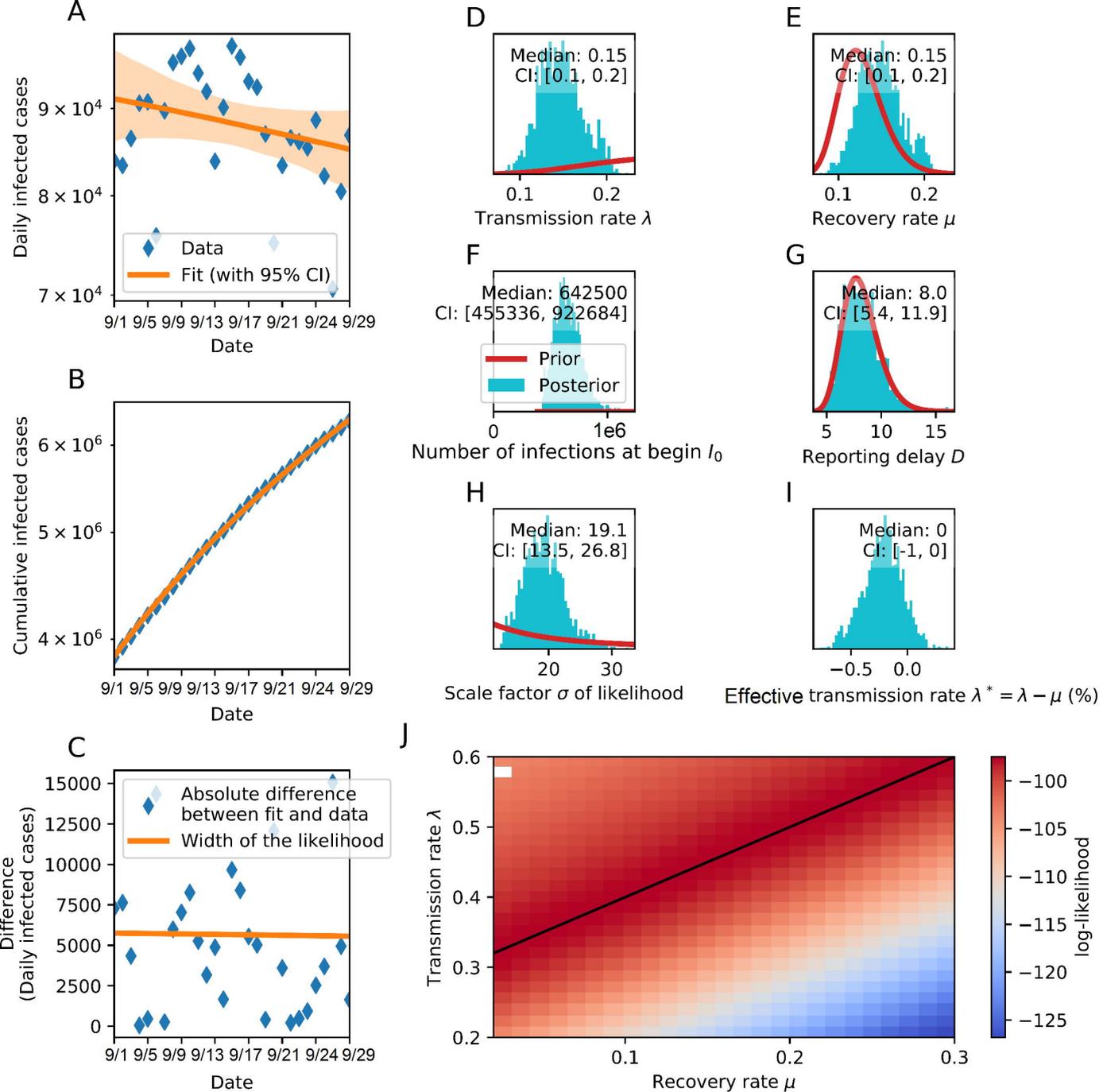
Time-series SIR model fit estimates of the (A)  $\lambda^*(t)$ , (B) daily infected cases compared to the observed data, and (C) cumulative infected cases compared to the observed data. Inset shows semi-log plots. Underreporting factor on 18 October 2020, for daily infected cases were higher using SEIR model compared to SIR model whereas the underreporting factor for cumulative infected cases were same as SIR model. (D-G) Priors and posterior distribution of model parameters, values are expressed in median and 95% CIs of the posteriors. The SEIR model featured an additional incubation period  $D_{inc}$  with prior lognormal (5, 1) scale parameter 0.418, and an initial exposed function  $E_0 \sim \text{HalfCauchy}(10)$ . The corresponding priors for reporting delay  $D$  were (5, 0.2),  $\lambda_0$  (2, 0.7),  $\mu$  (0.3, 0.3), and the other priors were same as SIR model.

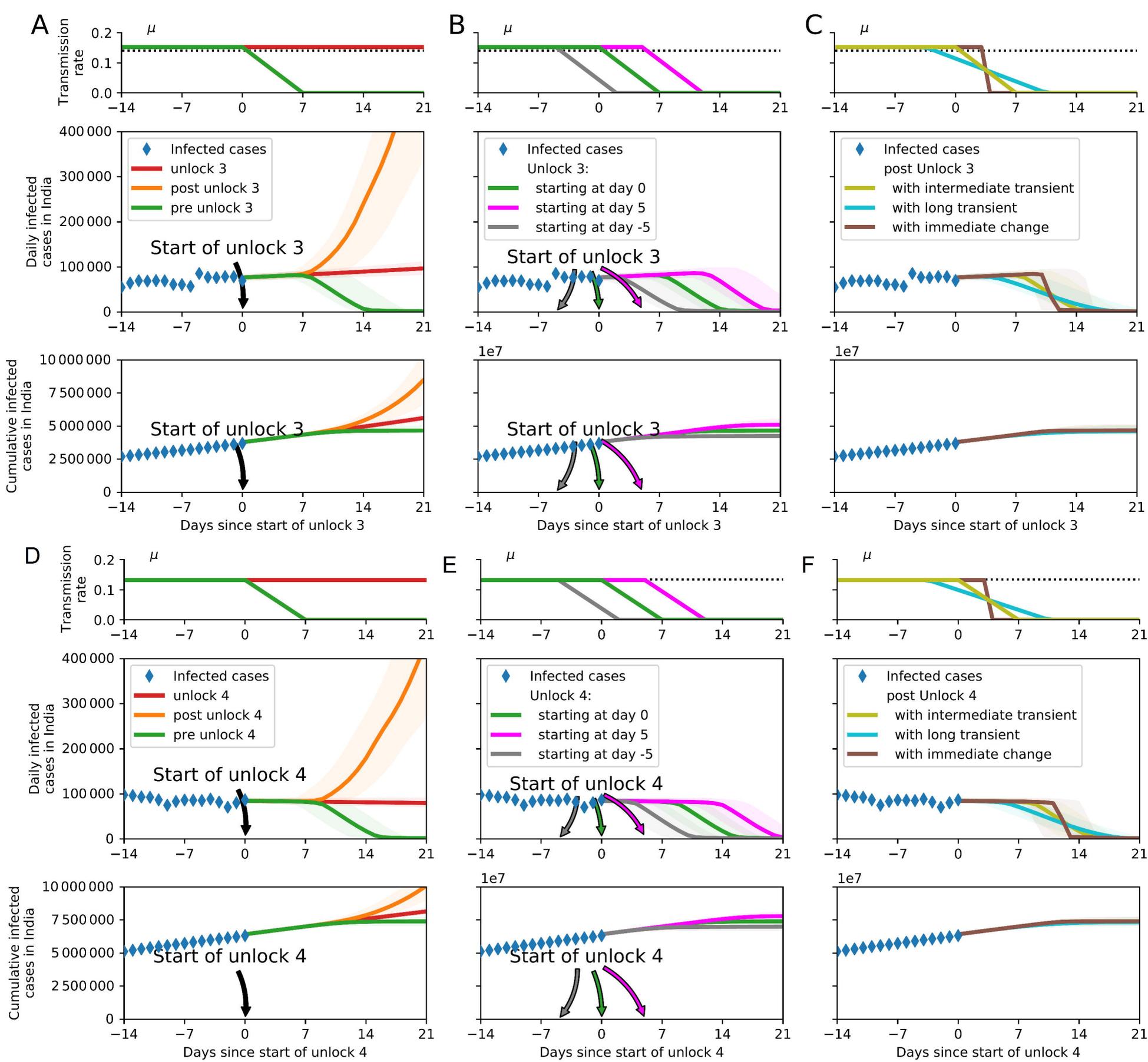
**Fig. 6. SIR model with two change points and weekend effect.** Same as Fig. 4 but with weekend effect and higher LOO score.

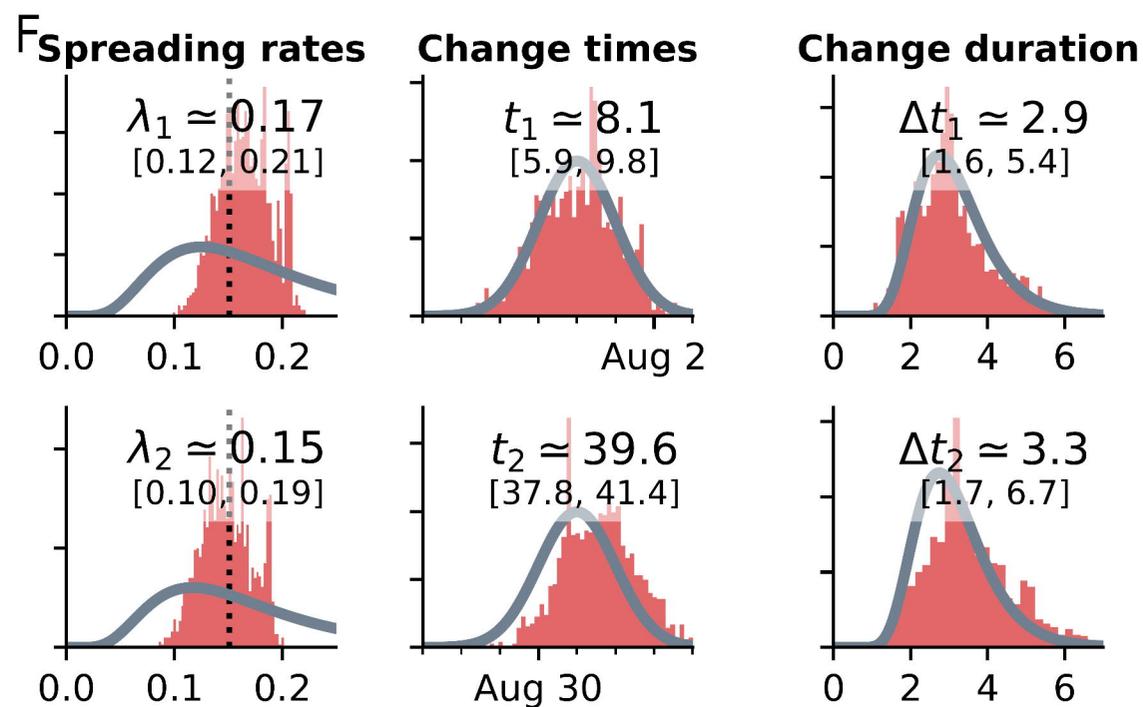
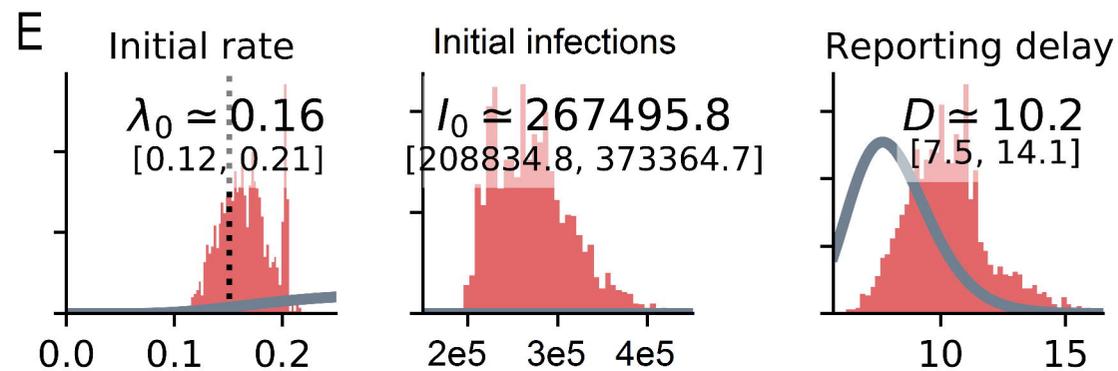
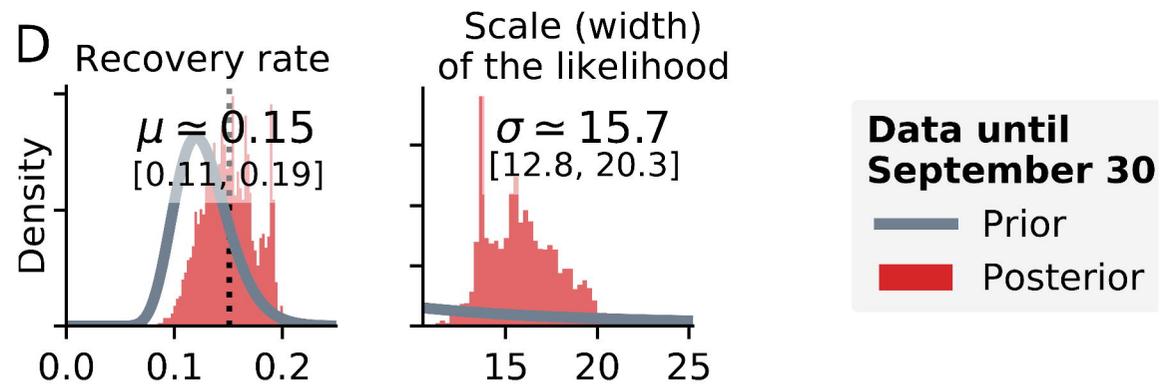
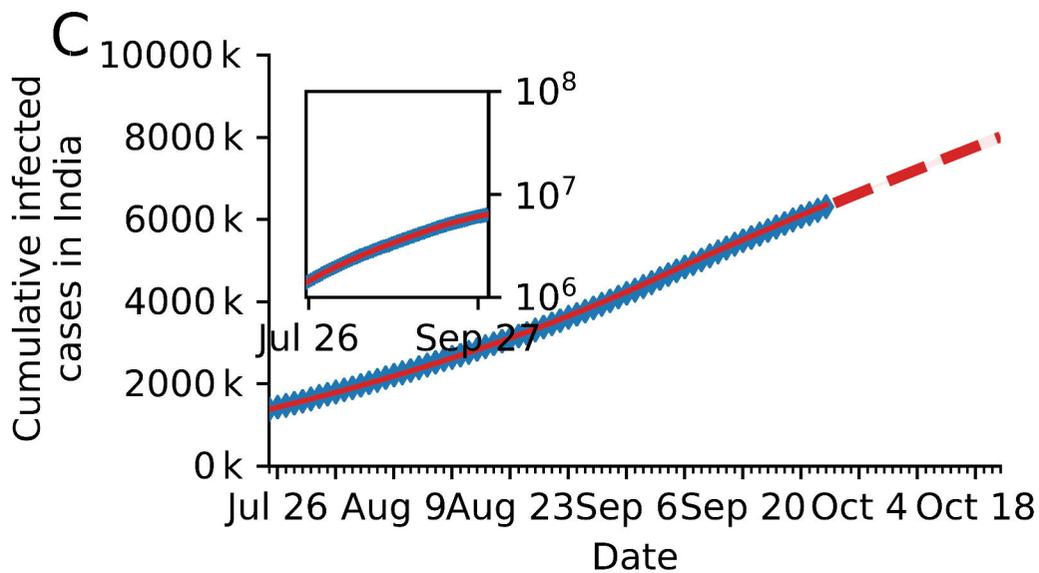
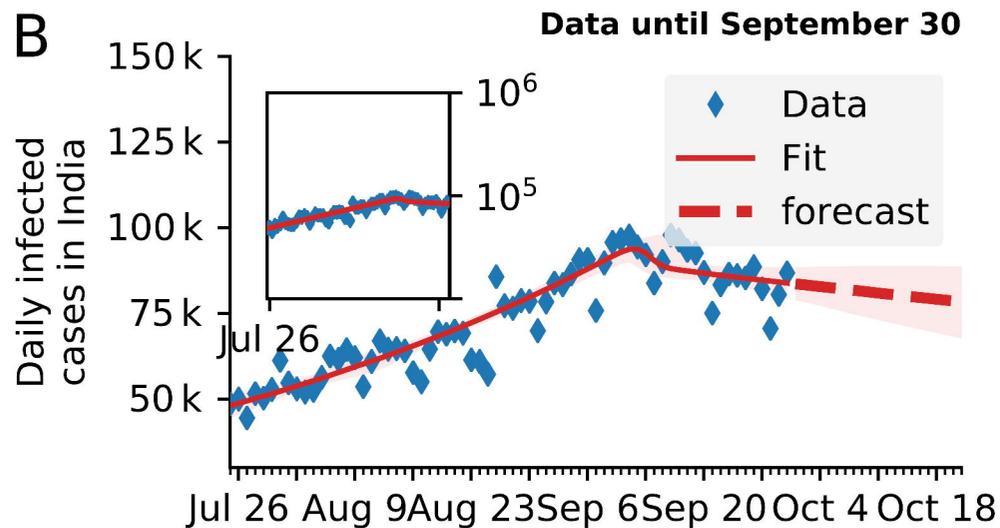
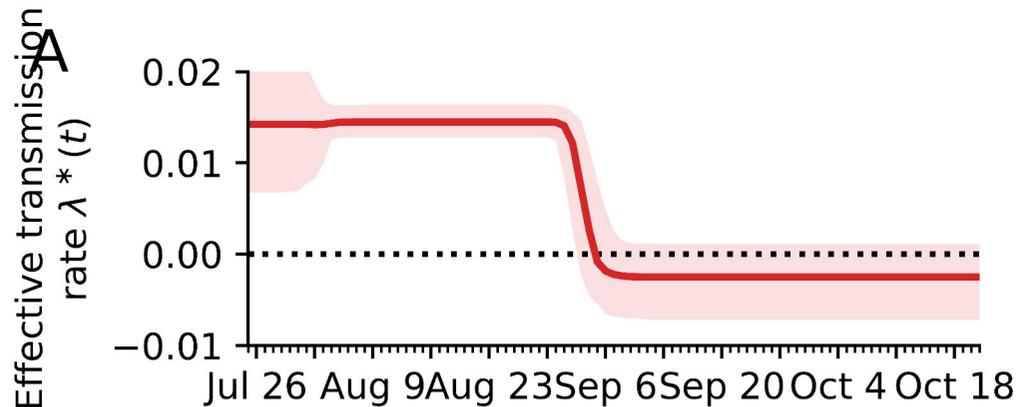
**Fig. 7. SIR model with one change point with weekend effect.**

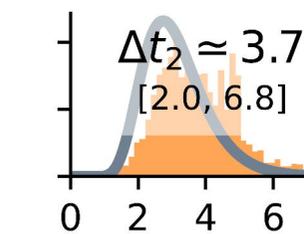
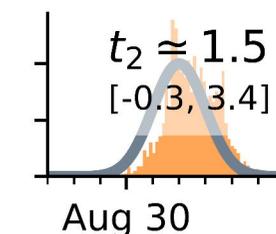
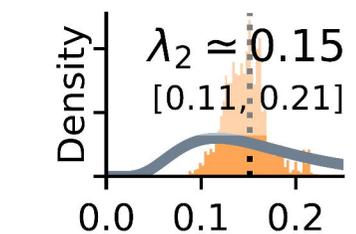
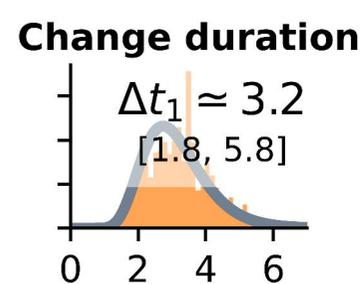
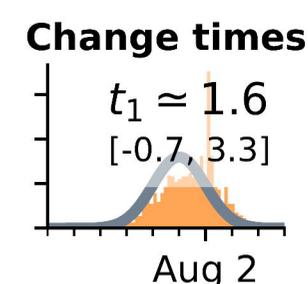
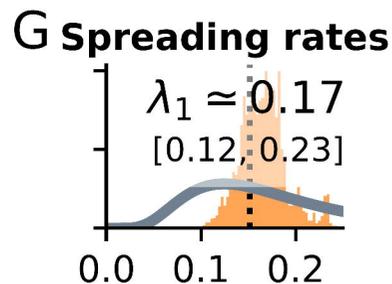
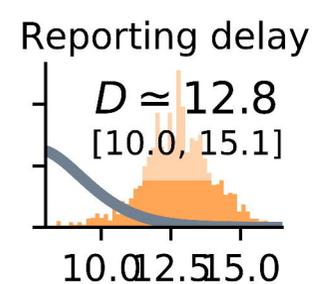
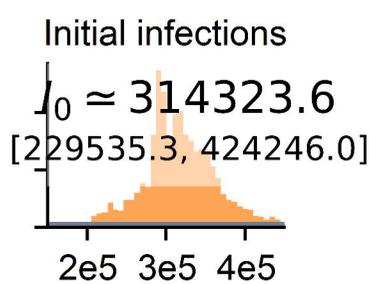
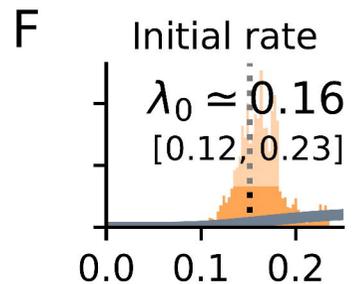
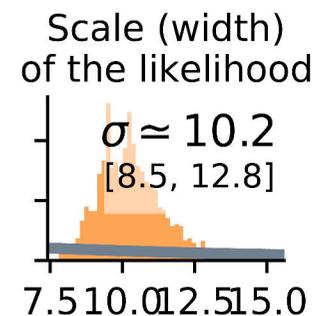
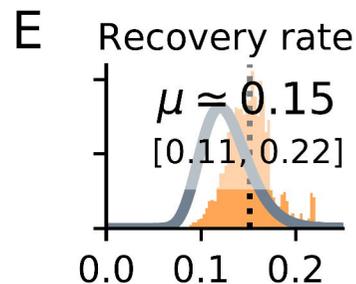
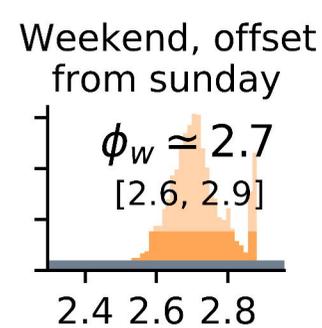
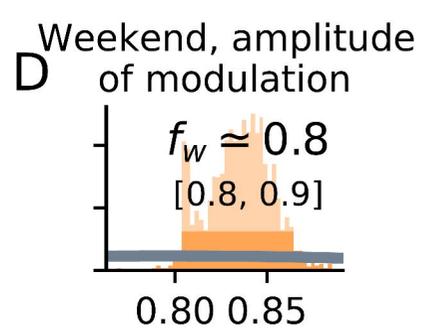
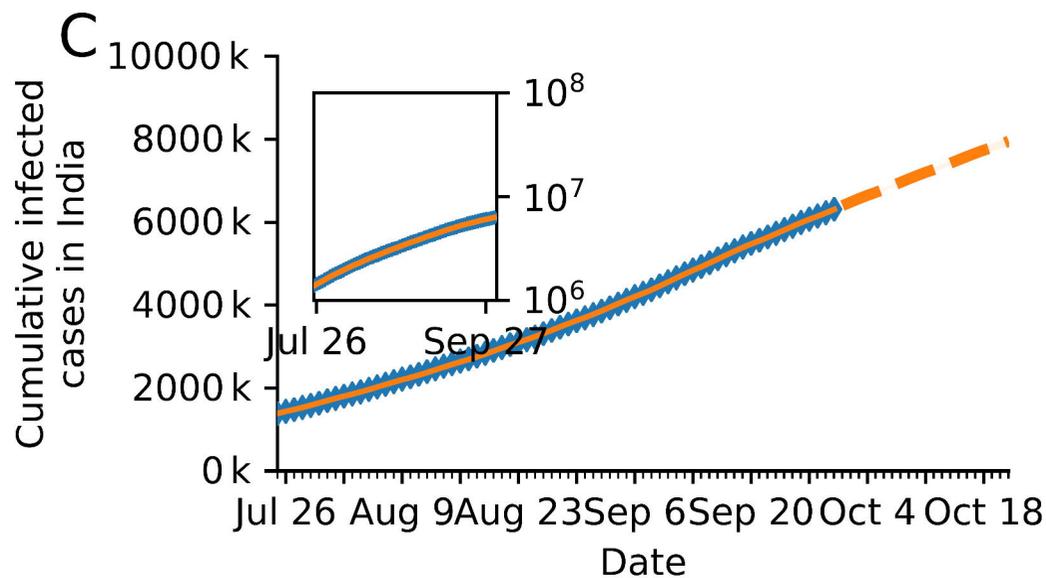
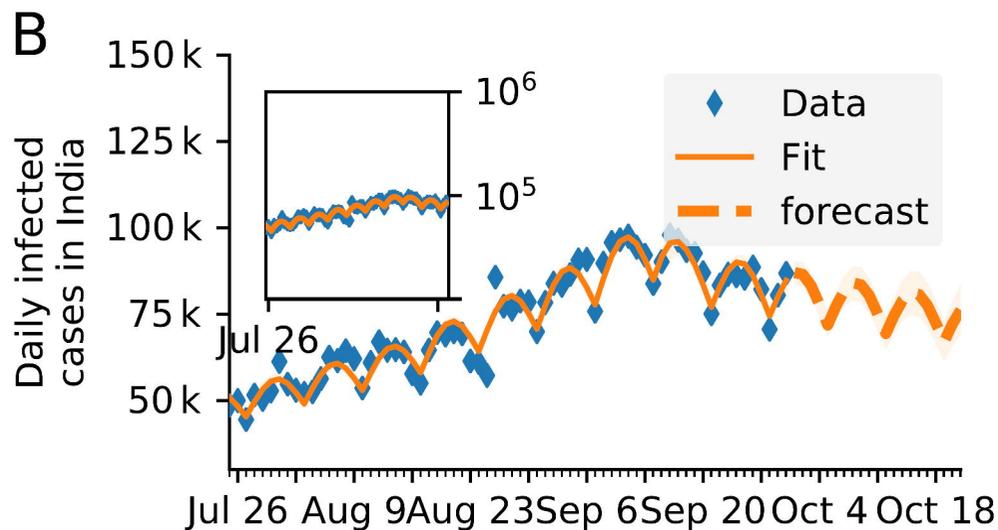
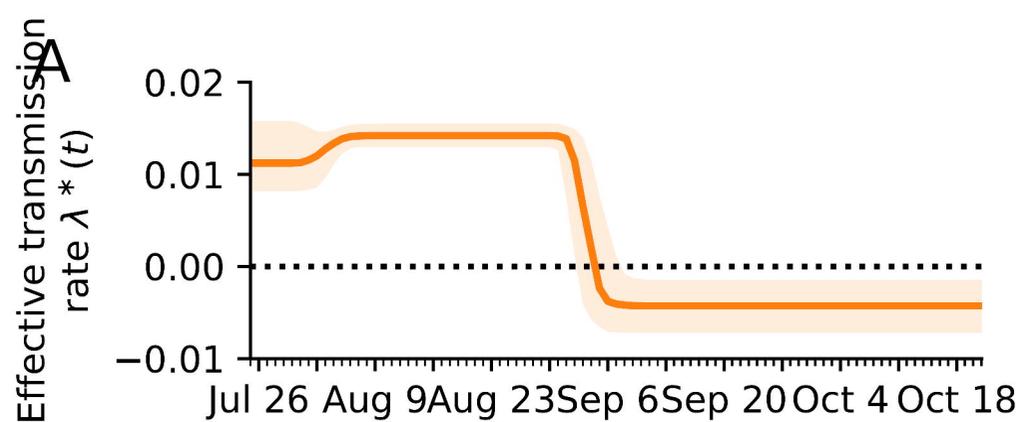
Same as Fig. 4 but with one change point and weekend effect and lower LOO score. Most favoured model. Underreporting factor on 18 October 2020, for daily infected cases were higher 2.1 (=130,000/61,871) with this model compared to SIR model in Fig. 4.





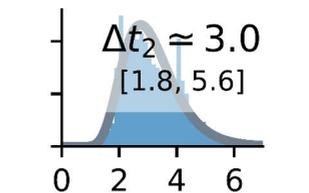
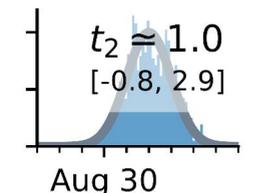
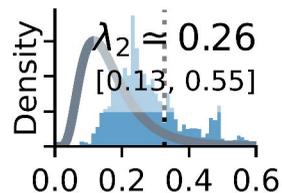
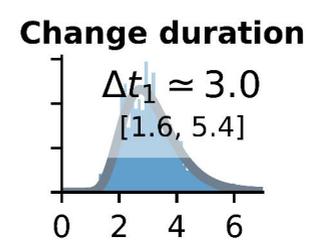
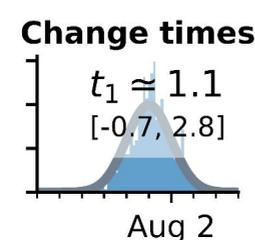
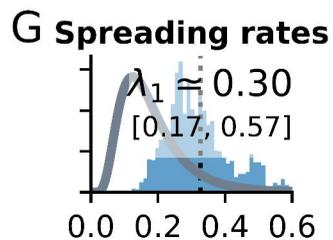
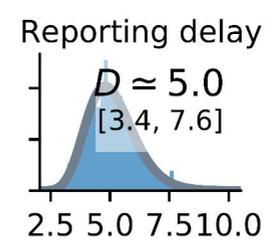
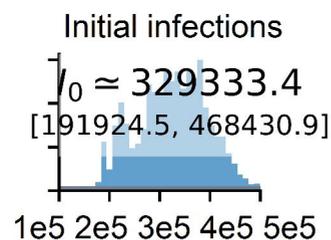
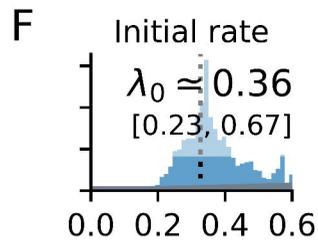
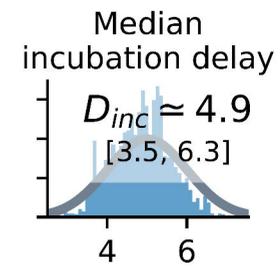
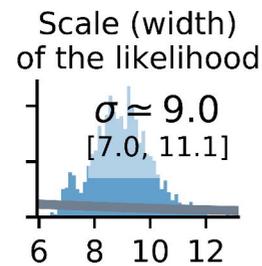
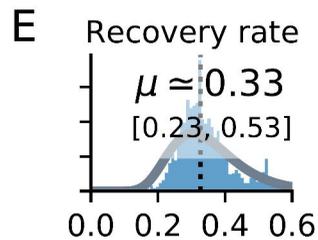
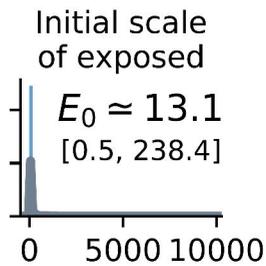
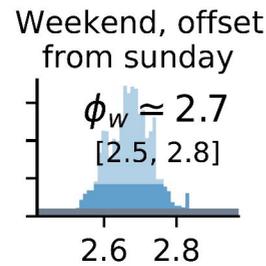
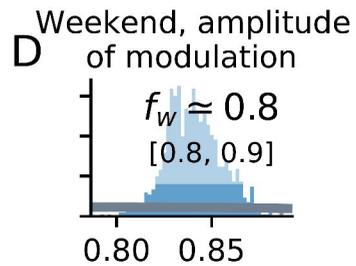
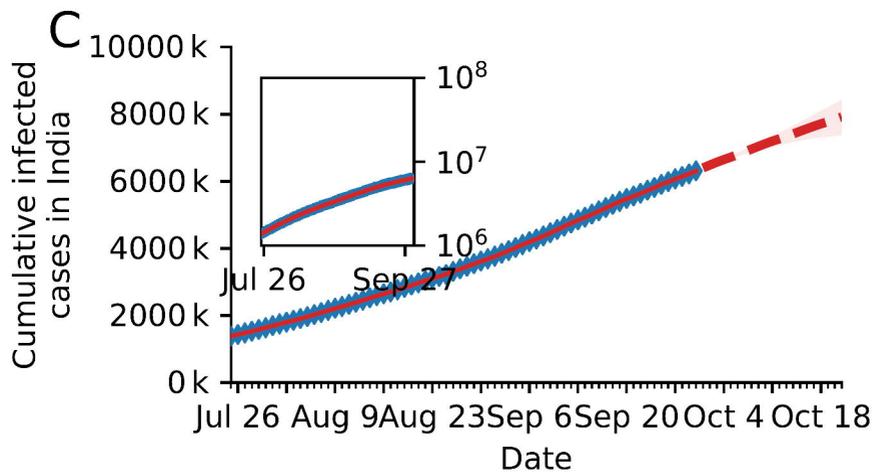
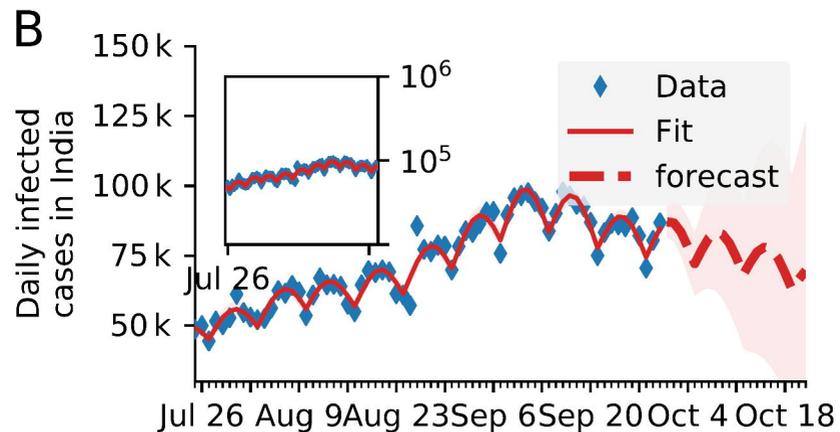
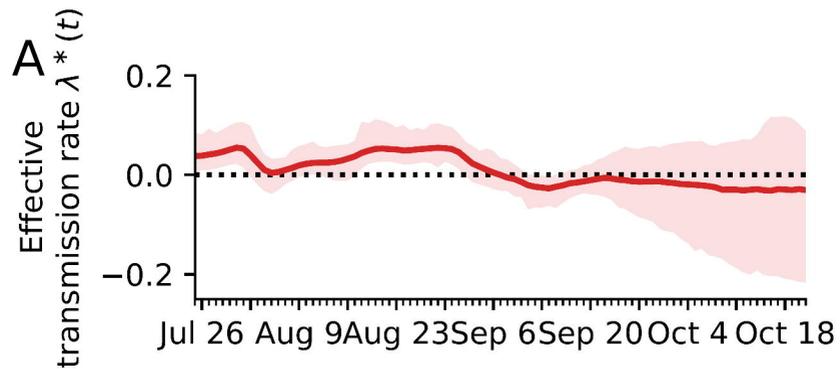






**Data until September 30**

— Prior  
— Posterior



**Data until September 30**

— Prior  
— Posterior

