1. Bharti, A. C., Donato, N., Singh, S. and Aggarwal, B. B., Curcumin (diferuloylmethane) down-regulates the constitutive activation of nuclear factor-κ B and IκBα kinase in human multiple myeloma cells, leading to suppression of proliferation and induction of apoptosis. *Blood*, 2003, **101**(3), 1053–1062.
2. Aggarwal, B. B., Kumar, A. and Bharti, A. C., Anticancer potential of curcumin: preclinical and clinical studies. *Anticancer Res*., 2003, **23**(1A), 363–398.
3. Ruby, A. J., Kuttan, G., Dinesh Babu, K., Rajasekharan, K. N. and Kuttan, R., Anti-tumour and antioxidant activity of natural curcuminoids. *Cancer Lett.*, 1995, **94**(1), 79–83.
4. Martín-Cordero, C., López-Lázaro, M., Gálvez, M. and Ayuso, M. J., Curcumin as a DNA topoisomerase II poison. *J. Enzyme Inhib. Med. Chem*., 2003, **18**(6), 505–509.
5. Wang, J. C., Cellular roles of DNA topoisomerases: a molecular perspective. *Nature Rev. Mol. Cell Biol*., 2002, **3**(6), 430–440.
6. Chao-Xin, H. *et al.*, Salvicine functions as novel topoisomerase II poison by binding to ATP pocket. *Mol. Pharmacol*., 2006, **70**, 1593–1601.
7. Cattan, A. R., Levett, D., Douglas, E. A., Middleton, P. G. and Taylorm, P. R., Method for quantifying expression of functionally active topoisomerase II in patients with leukaemia. *J. Clin. Pathol.*, 1996, **49**(10), 848–852.
8. Mueller-Planitz, F. and Herschlag, D., DNA topoisomerase II selects DNA cleavage sites based on reactivity rather than binding affinity. *Nucleic Acids Res*., 2007, **35**(11), 3764–3773.
9. Wei, H., Ruthenburg, A. J., Bechis, S. K. and Verdine, G. L., Nucleotide-dependent domain movement in the ATPase domain of a human type IIA DNA topoisomerase. *J. Biol. Chem*., 2005, **280**(44), 37041–37047.
10. Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E. and Friesner, R. A., A hierarchical approach to all-atom protein loop prediction. *Proteins*, 2004, **55**, 351–367.
11. Halgren, T., Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.*, 2009, **49**, 377–389.
12. Gadakar, P. K., Phukan, S., Dattatreya, P. and Balaji, V. N., Pose prediction accuracy in docking studies and enrichment of actives in the active site of GSK-3β. *J. Chem. Inf. Model.*, 2007, **47**, 1446–1459.
13. Hetenyi, C. and van der Spoel, D., Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Lett.*, 2006, **580**(5), 1447–1450.
14. Sanner, M. F., Python: a programming language for software integration and development. *J. Mol. Graph. Model*., 1999, **17**, 57–61.
15. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J., Automated docking using a lamarckian genetic algorithm and empirical binding free energy function. *J. Comput. Chem.*, 1998, **19**, 1639–1662.
16. Maestro (v7.0.113) − A unified interface for all Schrödinger products, developed and marketed by Schrödinger, LLC, NY, USA. Copyright 2005; http://www.schrodinger.com
17. Friesner, R. A. *et al.*, Extra Precision Glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem*., 2006, **49**, 6177–6196.
18. Thomsen, R. and Christensen, M. H. J., MolDock: A new technique for high-accuracy molecular docking. *Med. Chem.*, 2006, **49**(11), 3315–3321.

ACKNOWLEDGEMENTS. We thank the Director, Indian Institute of Information Technology, Allahabad for support.

Received 18 March 2011; revised accepted 9 September 2011

The transmission dynamics of pandemic influenza A/H1N1 2009–2010 in India

S. R. Gani[1,*], Sk. Taslim Ali[1] and A. S. Kadi[2]

[1]Department of Statistics, Karnatak Arts College, Dharwad 580 001, India
[2]Department of Statistics, Karnatak University, Dharwad 580 003, India

To understand the transmission dynamics of the prevailing pandemic 2009–10, due to the newly emerged pathogen influenza A/H1N1 throughout India, we have analysed the daily reported time-series dataset of the first two waves and a comparative study has been made for different regions of India. In order to quantify the early intensity of the strain, we have estimated basic reproduction number (R_0) through initial intrinsic growth rate method using standard deterministic SEIR model and effective reproduction number (R_t) through the stochastic SIR model using Bayesian inference. The estimate of reproduction number for India is 1.46 with 95% CI (1.15, 1.77), whereas for different states, the reproduction number is between 1.03 and 1.75.

Keywords: Bayesian inference, influenza H1N1, pandemic, reproduction number.

THE presence of the highly pathogenic influenza A/H1N1 virus among human population in several regions of the world, including India, has highlighted the urgent need for taking some efficient preventive measure against the latest pandemic. A novel influenza strain was first detected in Mexico in March 2009, which rapidly spread to different countries of the world[1]. In India, the first case (exogenous) of H1N1 2009–10 pandemic was identified on 17 May 2009 in Hyderabad and then it spread all over the country at varied intensities in different states[2]. As on 17 May 2010, there were 31,924 laboratory-confirmed cases in India and 1525 deaths were reported, i.e. 4.78% of the cases tested positive for influenza A/H1N1 virus[2,3], hence estimating the strength of the epidemic is of great concern.

The basic reproduction number (R_0) is the most common measure of this strength[4,5] to understand the early epidemic situation. In a wholly susceptible population, R_0 is simply the number of secondary cases generated by one primary case. When a fraction p of the population is effectively protected from infection, this quantity is known as the effective reproduction number (R_p) and is related to R_0 by $R_p = (1 - p)R_0$, assuming a well-mixed population[6]. Whereas the effective reproduction number at time t, denoted by R_t, is defined as the actual average number of secondary cases per primary case at time t (for $t > 0$)[7]

*For correspondence. (e-mail: gani_sr@hotmail.com)

and is typically less then R_0. Essentially, to quantify the characteristic or strength of outbreak over time for the whole epidemic span, effective reproduction number (R_t) is commonly used, considering control strategies utilized through public health policies, immunity developed in a population, climatic variations, pathogen evolution, spatial structure of the population along with all other interventions subject to the outbreak. The basic reproduction number quantifies the strength of outbreak at the initial phase, whereas the effective reproduction number explains the dynamics of the whole epidemic span.

The reports published about the estimation of R_0 on the basis of data obtained for different countries differ considerably. For example, R_0 has been variously estimated between 1.21 and 1.35 for Japan[8], 2.2 and 3.0 for Mexico[9], and 1.3 and 1.7 for the United States[10]. The divergence in the estimates for R_0 may be the result of the effect of different methods and models adopted in respective studies.

We have made use of two different approaches through their respective performances, to estimate the reproduction number for influenza A/H1N1 pandemic 2009–10 in India, method-1 – estimation of R_0 using early intrinsic growth rate method through the SEIR deterministic model[11], and method-2 – Bayesian estimation of R_t through stochastic SIR model[12]. We have examined R_t for each segments at the time t = cut-off point (Appendix 1), which is derived for the estimation of R_0. Our estimate of R_0 for India is 1.46 and it ranges from 1.17 to 1.52 for different states. Whereas we observed that the estimate of R_t for India was 1.46 with 95% CI (1.15, 1.77), and for different states, the reproduction number was between 1.03 and 1.75. According to our estimate, a single primary infectious case of H1N1 in India has produced on an average 1.46 secondary infectious cases during the entire infectious period. Also, the estimates show robustness in nature.

The pandemic 2009–10 is known as 'swine flu', caused by the influenza virus A of the sub-type H1N1. Daily reported time-series data for the 2009–10 influenza in India was obtained from the Ministry of Health and Family Welfare, Government of India (2, Figure 1 *a*). In India a total of 31,924 infective cases with 1525 deaths were reported during 366 days of the first two waves of pandemic from 17 May 2009 to 17 May 2010. We have adopted different starting points of the epidemic for different states as well as for the whole of India by considering the first indigenous cases respectively, for the estimation of basic reproduction number. Some statistical adjustment technique has also been employed to manipulate the missing data. The total notified case fatality proportion (CFP) of these two waves in India was found to be 4.78% (Table 1).

Method-1: We have made use of standard deterministic SEIR epidemic model which classifies the individuals into susceptible $S(t)$, latent/exposed $E(t)$, infective $I(t)$ and removed $R(t)$[4]. Susceptible individuals in contact with the virus enter the exposed class at the rate $\beta I(t)/N(t)$, where β is the transmission rate and $N(t) = S(t) + E(t) + I(t) + R(t)$ is the total population at time t. As the time-scale of the epidemic is much faster than the characteristic times for demographic processes (natural births and deaths), these effects are not included in the model under study. Assuming the homogeneous mixing among the individuals, the transmission dynamics of the epidemic can be modelled using the following system of nonlinear differential equations.

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t) = -\frac{\beta S(t)I(t)}{N(t)}, \qquad \frac{\mathrm{d}}{\mathrm{d}t}E(t) = \frac{\beta S(t)I(t)}{N(t)} - kE(t),$$

$$\frac{\mathrm{d}}{\mathrm{d}t}I(t) = kE(t) - \gamma I(t), \qquad \frac{\mathrm{d}}{\mathrm{d}t}R(t) = \gamma I(t), \tag{1}$$

where $1/k$ is the mean latent period and $1/\gamma$ is the mean infectious period.

The key epidemiological parameter that characterizes the transmission potential of an infectious disease is R_0, which can be typically estimated from the early epidemic phase during which the epidemic runs its free course in the absence of interventions and the effects of susceptible depletion are small. Commonly, it is assumed for most of the human infectious diseases that an exponential growth exists in the initial phase of epidemic[4,5]. In such situations, one of the most familiar methods for computing R_0 is based on initial exponential growth rate ($\boldsymbol{r}$) for the cumulative number of cases. This quantity is also known as the intrinsic growth rate or force of infection, estimated by fitting a simple linear regression line ($a + rt$) to the 'BEST' length of its exponential phase in logarithmic scale[11]. The 'BEST' length of its exponential phase (cut-off point, Appendix 1) can be determined by the coefficient of determination, the goodness-of-fit test statistic. R_o is then computed by substituting the estimated intrinsic growth rate $\boldsymbol{r}$ in eq. (2), which is derived from the linearized simple SEIR model[13] and is given by

$$R_0 = 1 + Vr + f(1-f)(Vr)^2 + O[(Vr)^3], \tag{2}$$

where V is the mean serial interval (sum of the mean latent period and mean infectious period), and f is the ratio of the mean infectious period to the mean serial interval.

Method-2: Considering the uncertain behaviour of daily reported cases, we have used Bayesian inferential method applied to a stochastic version of a standard SIR model. One may use a stochastic version of the SEIR model instead, but this leads to more complexity in the computation. This method allows us to estimate R_t directly, through the probabilistic nature of the contagion process. R_t shows time-dependent variation with the decline in susceptible individuals (intrinsic factors) and with the

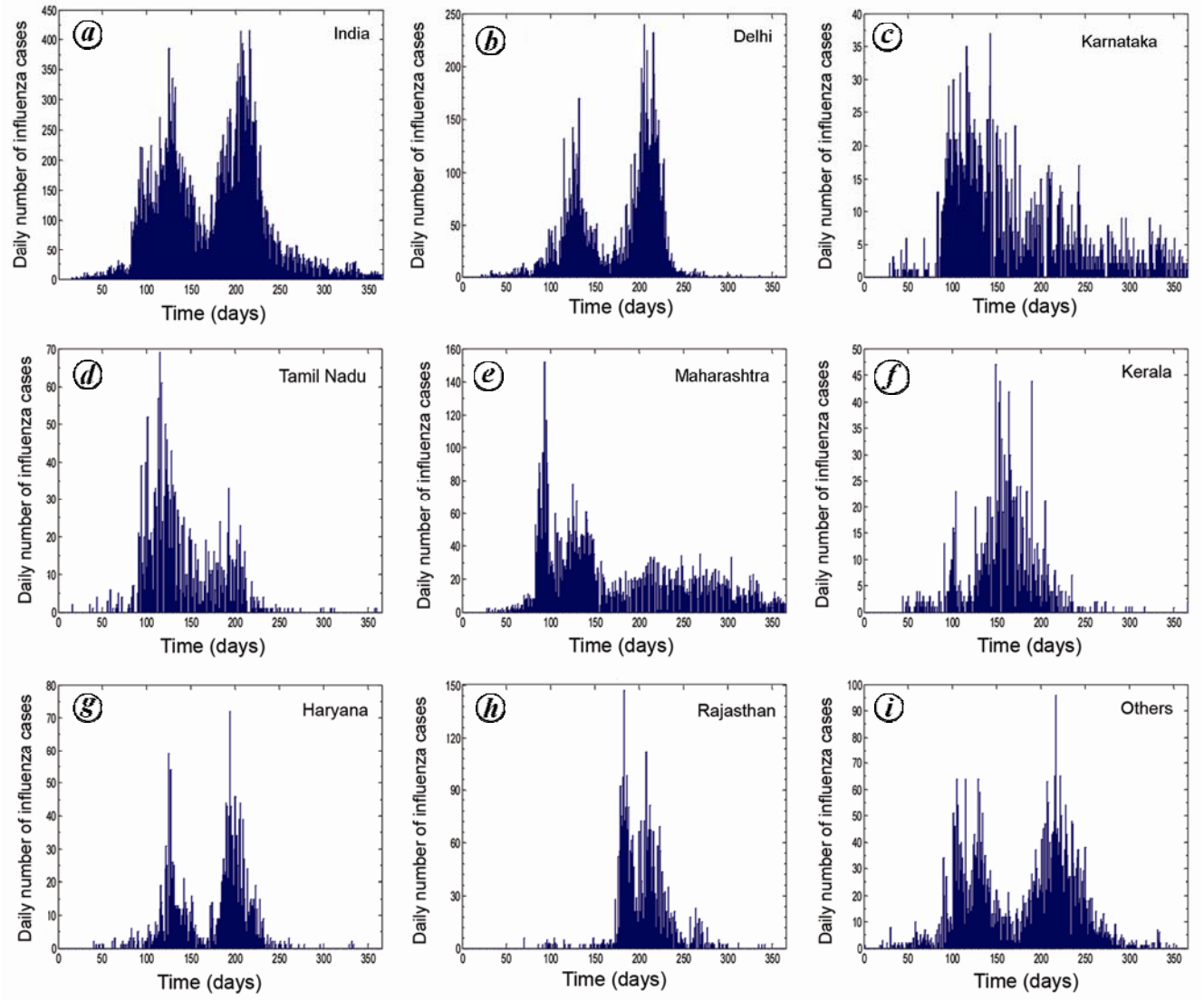

Figure 1. Daily number of influenza A/H1N1 notifications in different segments (fairly affected states) of India during the pandemic 2009–10, shows the different waves patterns. This represents two waves for India, whereas some of the segments do not exhibit the same pattern clearly.

Table 1. Estimates of basic reproduction number and effective reproduction number with 95% CI of the influenza pandemic 2009–10 for India and its different segments (fairly affected states)

State	Total number of infections	Percentage of infections	Total number of deaths	Percentage of deaths	Case fatality proportion (per 100)	Basic reproduction number (R_0)	Effective reproduction number (R_t)* (with 95% CI)	Doubling time
Delhi	9,697	30.38	95	6.23	0.98	1.52	1.24 (0.23, 2.25)	8.89
Karnataka	2,350	7.36	164	10.75	6.98	1.32	1.50 (0.60, 2.41)	13.86
Tamil Nadu	2,090	6.55	7	0.46	0.34	1.50	1.68 (0.79, 2.57)	9.12
Maharashtra	6,283	19.68	461	30.23	7.34	1.49	1.35 (0.25, 2.44)	9.24
Kerala	1,482	4.64	38	2.49	2.56	1.35	1.03 (0.21, 1.85)	12.84
Haryana	1,948	6.10	39	2.56	2.00	1.33	1.31 (0.00, 2.74)	13.33
Rajasthan	3,380	10.59	198	12.98	5.86	1.17	1.75 (0.92, 2.58)	25.67
Others	4,694	14.70	523	34.30	11.14	1.45	1.29 (1.07, 1.51)	10.05
India	31,924	100.00	1525	100.00	4.78	1.46	1.46 (1.15, 1.77)	9.90

*R_t has been calculated for each segment at the time t = cut-off point, which is derived for the estimation of R_0.

implementation of control measures (extrinsic factors). Sometimes, it may be found to increase over time owing to the pathogen evolution or population structure variation.

We have made use of the method developed by Bettencourt and Ribeiro[12], the sequential Bayesian estimation of effective reproduction number through a stochastic SIR model, considering the standard SIR model (omitting removed and death cases for simplicity). Here, the notations are the same as explained earlier.

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t) = -\frac{\beta S(t)I(t)}{N(t)},$$

$$\frac{\mathrm{d}}{\mathrm{d}t}I(t) = \frac{\beta S(t)I(t)}{N(t)} - \gamma I(t). \tag{3}$$

A stochastic version of this model can be formulated by taking the rates on the right-hand side of population equations (3) to determine the mean changes λ over the time τ of the several compartments of population, which usually are evaluated from a probability distribution $P(\lambda)$ with average λ. One may consider the discrete probability distribution P as Poisson distribution (a maximal entropy distribution), when only average is known. If the measure of variation is available too, a more general distribution, such as negative binomial may be adopted instead.

In reality, the information about outbreak is usually not in terms of infectious individuals, but as a collection of reported cases, which at the time of reporting may have progressed. With this limitation, it is worth considering our estimation procedure in terms of the changes in the commutative number of cases, $C(t)$. Then the number of new cases at time t is $\Delta C(t) = C(t) - C(t-\tau)$, where τ denotes the time interval between successive reports. For our study $\tau = 1$ day. As $C(t) = I(t) + R(t) + D(t)$, from eq. (3) the total number of cases up to time t, $C(t)$ obeys the equation $\mathrm{d}C(t)/\mathrm{d}t = \beta S(t)I(t)/N(t)$. To find the expression for the evolution of new cases $\Delta C(t+\tau)$, integrate eq. (3) for $I(t)$ on $(t, t+\tau)$ to obtain

$$I(t+\tau) = I(t)\exp\left[\gamma\int_t^{t+\tau}\left(R_0\frac{S(t')}{N(t')} - 1\right)\mathrm{d}t'\right],$$

$$\approx I(t)\exp[\gamma(R_t - 1)], \tag{4}$$

where $R_0 = \beta/\gamma$ and $R_t = [S(t)/N(t)]R_0$ (the instantaneous reproduction number) is a function of time; and eq. (4) is approximated assuming $S(t)/N(t)$ as constant on $(t, t+\tau)$. Consider $b(R_t) = \exp[\gamma\tau(R_t - 1)]$, then $I(t+\tau) = I(t)b(R_t)$. Now, to obtain the disease progression in terms of reported cases, we discretized the differential equation for the change in total number of cases on $(t, t+\tau)$. Consequently, we have

$$\Delta C(t+\tau) = \Delta C(t)b(R_t), \tag{5}$$

where $b(R_t) = \exp[\gamma\tau(R_t - 1)]$, which is the slope of the straight line passing through each node and origin in the time-delay trajectory plot (Figure 2) of $\Delta C(t)$ versus $\Delta C(t+\tau)$. Then it can be evaluated from the data for each t.

$$R_t = 1 + \log(b(R_t))/\gamma\tau. \tag{6}$$

Equation (5) explains how the initial R_t can be estimated geometrically (without the parameter search or numerical optimization) from an epidemic time-delay plot of surveillance data. But in practice, for emerging infectious diseases, relative variations in case numbers are large (Figure 2, expressing a large fluctuation in new cases). Therefore, this simple geometric approach becomes less realistic. Thus, we need a method which provides a much more robust estimate of the reproduction number. This notion led us to find a stochastic estimation procedure evaluating the probability distribution of R_t instead.

A probabilistic description is decisive for realistic modelling of new cases of emerging infectious diseases, which are typically characterized by large coefficient of variation. This can be achieved by defining the number of new cases, $\Delta C(t+\tau)$ as a discrete stochastic variable generated by a probability distribution with average number of cases given by eq. (5), i.e. $\Delta C(t+\tau) \sim P\{\Delta C(t)b(R_t)\}$, where $P\{\lambda\}$ denotes a discrete probability distribution with mean λ. In other words, for a given R_t (and other parameters like γ) and $\Delta C(t)$, one can predict the distribution of a future case number as $P[\Delta C(t+\tau) \leftarrow \Delta C(t)/R_t] = P\{\lambda\}$, $\lambda = \Delta C(t)b(R_t)$ for the SIR model[12].

With this uncertain measure, the parameter estimation can be achieved using Bayesian approach in the perspective of probabilistic epidemiological models. The probabilistic formulation for future cases is equivalent to the

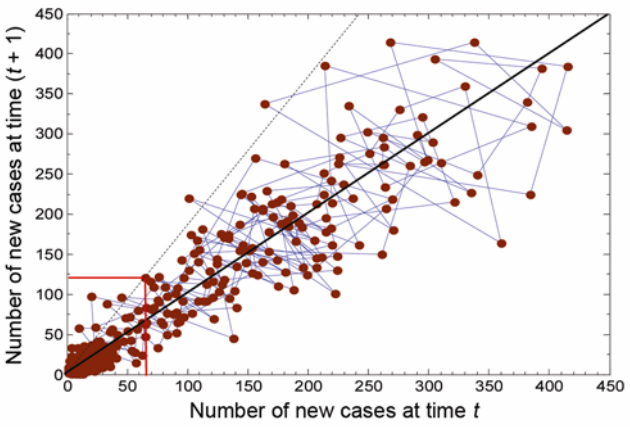

Figure 2. Time-delay trajectory plot of $\Delta C(t)$ versus $\Delta C(t+\tau)$ for the Indian data. The solid diagonal line indicates $b(R_t) = 1$, i.e. $R_t = 1$ (eventually the nodes which are above the line refer to $R_t > 1$ and those below refer to $R_t < 1$). Dotted line indicates in a particular time say $t = 64$, $b(R_t) = 1.88$ (shown as red lines).

estimation of the probability distribution for R_t through Bayes' theorem as

$$P[R_t|\Delta C(t+\tau) \leftarrow \Delta C(t)] = \frac{P[\Delta C(t+\tau) \leftarrow \Delta C(t)/R_t]P[R_t]}{P[\Delta C(t+\tau) \leftarrow \Delta C(t)]}, \quad (7)$$

where $P[R_t]$ is the prior, which reflects any prior knowledge of the distribution of R_t (one can start with uniform distribution); and the denominator is a normalization factor. From successive applications of Bayes' theorem, a sequential estimation scheme that exploits the course of epidemiological observations on real time, can be constructed using the posterior distribution of R_t at time t as a prior in the next estimation step at $t + \tau$. This iteration of eq. (7) results in an updated distribution scheme each time for R_t. From these successive distributions, mean (average) or maximum likelihood estimate (with confidence interval) may be recorded as an estimator of R_t. In this study we have adopted the mean (average) as the best estimator for R_t (Figure 3, for all the regions as well as for the whole of India) over time. We have used the real-time series data of new cases of influenza A/H1N1, $C(t)$ to compute the probability distribution of R_t, adopting the program implementation in MATLAB. We have used an unbiased uniform distribution $U(0, 3)$ as the initial prior (since maximum value of R_t recorded is 3 for this strain). For each subsequent daily iteration, we evaluated full posterior distribution from eq. (7) using the posterior of previous day as a new prior for posterior distribution of R_t; see, Appendix 2. We have made use of other parameter choices reported in the literature for influenza.

The entire population was assumed to be susceptible at the beginning of the epidemic, and at the initial stage of the epidemic the infected fraction of the population at risk raises exponentially when R_0 is greater than 1. Then

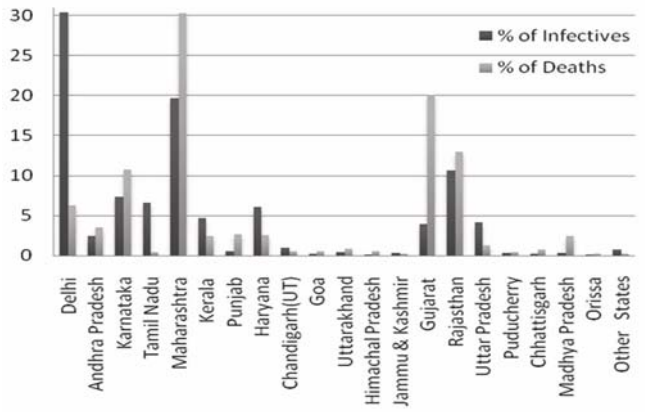

Figure 3. Percentage of infectives and percentage of deaths for different states of India. Gujarat, Maharashtra, Karnataka and Rajasthan have greater death-proportion compared to their respective infective-proportion. 'Other States' includes all less affected states and Union Territories.

the doubling time t_d is defined as the time taken to double the infectious cases[4] and is given by

$$t_d = r^{-1} \ln 2, \quad (8)$$

where r is the initial intrinsic growth rate.

Based on the different model performances (deterministic and stochastic), the reproduction numbers for India (2009–10 pandemic H1N1) and its different states have been estimated from daily reported cases. In method-1, to evaluate the initial exponential growth rate, we have determined the cut-off points considering the first indigenous case as the starting point of the epidemic; the method[11] is described in Appendix 1. For India, we fixed the starting point of the epidemic as 11 June 2009 and the cut-off point was derived as 13 August 2009. The 64 days duration is a fair indication of the best length of exponential growth based on the goodness-of-fit (coefficient of determination, R^2) for India. The estimate of R_0 obtained through this method was 1.46 for India. The mean latent period was fixed to 1.9 days[14] and mean serial interval was 3–6 days[15,16]. Though this method entertains only the initial dataset, we have used the remaining epidemic days for determining the cut-off point; but there were no other indications of any such clear exponential phases. We also estimated the R_0 for each fairly affected state of India and it ranged from 1.17 to 1.52 (Table 1) using method-1. The doubling time was estimated for India as 9.9 days, considering the pure exponential growth at the beginning of epidemic. The respective state-wise doubling time has also been derived (Table 1). Method-2 that estimates effective reproduction number using Bayesian inference is more robust in nature. Considering the same cut-off points as used in method-1, the reproduction number obtained was 1.46 with 95% confidence interval (1.15, 1.77) for India. The same has been calculated for fairly affected state of India and it ranged from 1.03 to 1.75 with the respective confidence intervals (Table 1). Here, one can figure out the strength of the outbreak instantaneously through the value of R_t and also how it is decaying over time (days). We adopted the moving average method for some of the states to overcome the problem of program implementation on real dataset. This led to the distribution patterns of R_t over time being less stochastic in nature (Figure 4 *b*–*h*).

We examined daily reported epidemiological time-series data of the novel influenza 2009–10 strain, which is termed as influenza A of sub-type H1N1. This strain hit India on 17 May 2009 as its first case (exogenous) and till 17 May 2010 had knocked two complete waves of epidemic span. The first wave (May–October 2009) claimed 448 lives from 14,293 infected cases and the second wave (November 2009–May 2010) claimed 1077 lives out of 17,631 infected cases. In addition, the third wave of pandemic is also in its further course of run. A close observation of data from different states reveals that

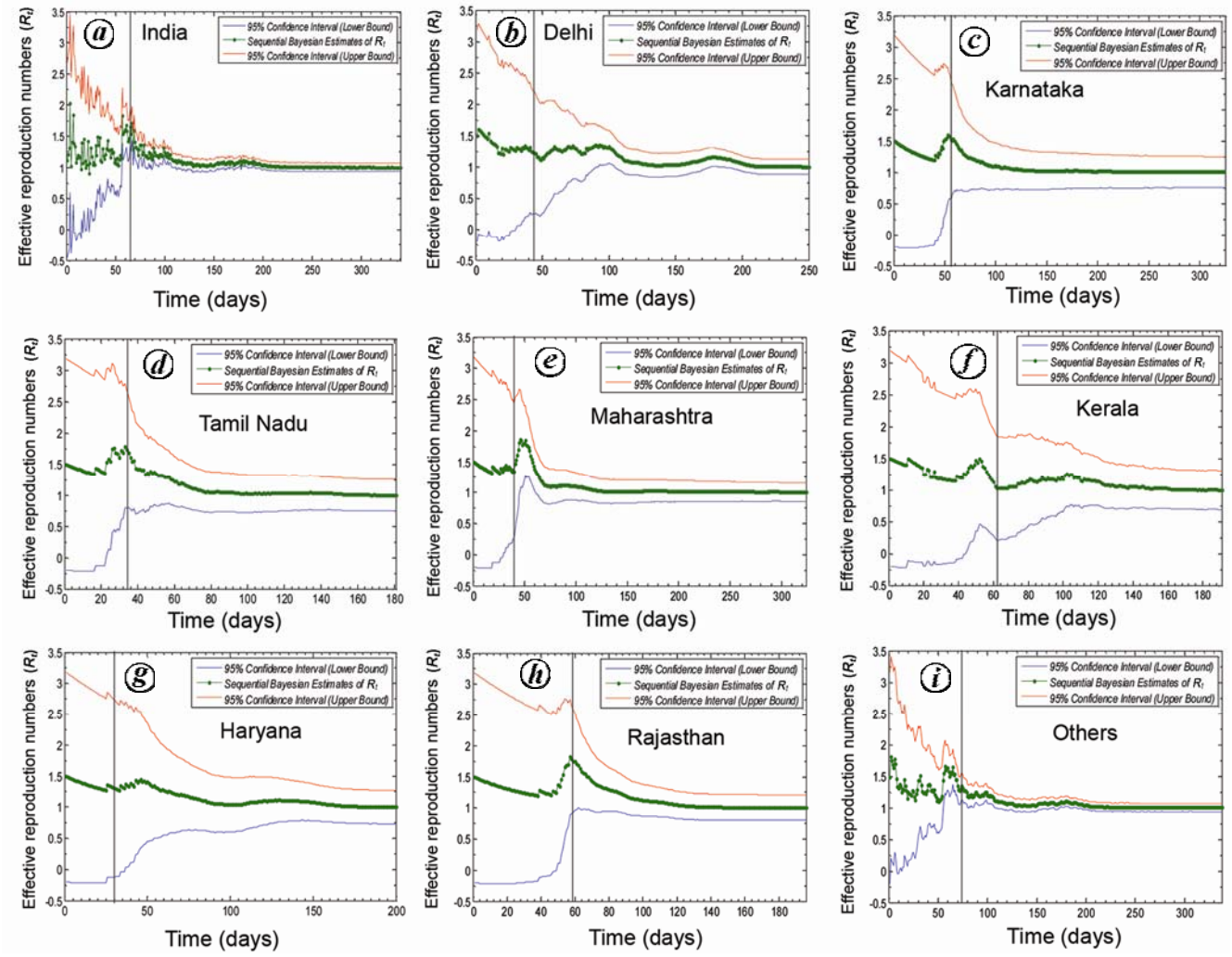

Figure 4. Sequential Bayesian estimation of full distribution leads to the estimation of its expected value (green lines) and 95% CI (red and blue lines). Uncertainty measured by the width of CI decreases with more case observations. The estimate eventually leads to smaller R_t owing to the depliation of susceptibles and later $R_t \to 1$. The vertical lines indicate the cut-off points based on which we estimated R_0 and R_t.

the epidemic started initially in some parts of India and concentrated there and later moved to other parts of the country. For example, the number of infectious cases in the first wave for Rajasthan was notably few and these were mostly exogenous cases. The severity of the epidemic in Rajasthan can be found in the second phase. Analysis of the overall Indian situation forced us to discuss the seven worst-hit states (Table 1). It has also been observed that the present pandemic is comparatively mild in North East states, which account for 60% (about 680 million) of the Indian population. The fatality rate (deaths per 100 infective) due to the recent H1N1 for west and south states was found to be 6.97, whereas it was only 2.0 for north and east states. Also, the insight analysis revealed that there were no two complete waves for each segment of India; Rajasthan and Kerala showed only one clear wave instead (Figure 1).

To get more insight into the dynamics of the pandemic, we have used the initial intrinsic growth rate method and applied it to the standard SEIR deterministic model in order to be commensurate with the paucity of epidemiological data, typically available for this emerging influenza pandemic to measure the intensity of outbreak throughout India. Our estimates of R_0 range from 1.17 to 1.52, with a low level of spatial variation across eight geographical segments of India, which were marginally associated with socio-demographic factors. In contrast, there is a remarkable spatial heterogeneity in the death rates, which range from 0.34% to 7.34% across the country. Interestingly, it has been found that though the intensity of emerging influenza was very high for Delhi (BRN 1.52) and Tamil Nadu (BRN 1.50), CFP was comparatively very less: Delhi (CFP 0.98) and Tamil Nadu (CFP 0.34). This fact leads to the analysis of control policy measures employed for these populations and their structure. Unfortunately, in the 'Other States' segment CFP was very high (11.14), mainly due to its high value in the states in this segment, namely Gujarat, Punjab, Madhya

Pradesh, Chhattisgarh and Himachal Pradesh. Though Punjab, Madhya Pradesh, Chhattisgarh and Himachal Pradesh have less infectious (less than 200), their CFP is comparatively higher. Moreover, Gujarat itself accounted for 24.28 CFP in the case of 'Other States' segment.

Method-1 quantifies the measure of epidemic potential for a pathogen through R_0. In reality, epidemiological data typically permit only the estimation of R_t, which may differ from R_0 due to accured immunity and other factors. In fact, R_t is found to be more informative than R_0, as it reveals the dynamics of the epidemic process over time. For an emerging infectious disease when transmission is at an early stage and the pathogen is adopting to the population, it becomes crucial to monitor quantitative changes of effective reproduction number over time[12]. Thus the detection and tracking of an emerging disease can be formalized in terms of monitoring R_t, as it evolves and approaches the critical threshold $R_t \to 1$. In the case of India, the pandemic H1N1 is not uniform all over the country (Figure 1). Also, the strength of outbreak in some regions like Tamil Nadu has high BRN 1.50 at the beginning, but got infected only 6.55% of the total infected cases in India. Moreover, the dramatic increase in incidence from 20 to 97 occurred from 6 to 7 August 2009, which has a direct effect on the uncertainty of the reproduction number estimates. This phenomenon can be addressed by real-time monitoring of the outbreak through method-2. Moreover, the real-time data limitation accounts for several statistical methodologies, like moving average to avoid no infection in a day (zeros in daily infective); which may decrease the degree of the stochastic nature of the model trajectory. To avoid the complexity and overall data limitation we have adopted the stochastic version of the SIR model (method-2) in our present study on H1N1 to estimate R_t instead of the stochastic SEIR model.

The estimates of reproduction number through both the methods coincide for India, whereas these estimates of R_t are slightly less than R_0 for the sub-regions of India (Table 1), except states like Rajasthan ($R_t = 1.75$, $R_0 = 1.17$), Tamil Nadu ($R_t = 1.68$, $R_0 = 1.50$) and Karnataka ($R_t = 1.50$, $R_0 = 1.32$). This is due to the data smoothening (through moving average), required for the implementation of method-2, incorporationg the fact that the data exhibit very large fluctuations, with some days not showing a single case, while the following days show extremely large number of cases and vice-versa.

In conclusion, our result indicates that the recent pandemic influenza A/H1N1 in India is comparatively mild than that reported in other countries of the world. The second wave in India had more impact on the population compared to the first wave. The early determination of BRN helps to design the optimal control policies to counter the epidemic spread in advance, whereas the real-time estimate of R_t helps understand the future pattern of the disease dynamics. Some regions (like Delhi and Tamil Nadu) had higher intensity at the beginning, but showed comparatively low CFP. On the other hand, it was the reverse for Gujarat, Maharashtra, Karnataka and Rajasthan, which led to the study of the public health facilities as well as pathogen analysis. The varied intensity and fatality measures in different segments of India encouraged us to analyse the entire epidemic span which accounts for the control strategies, population structure and pathogen evolution. Our estimates aim at public health interventions in the case of similar pandemics in the future.

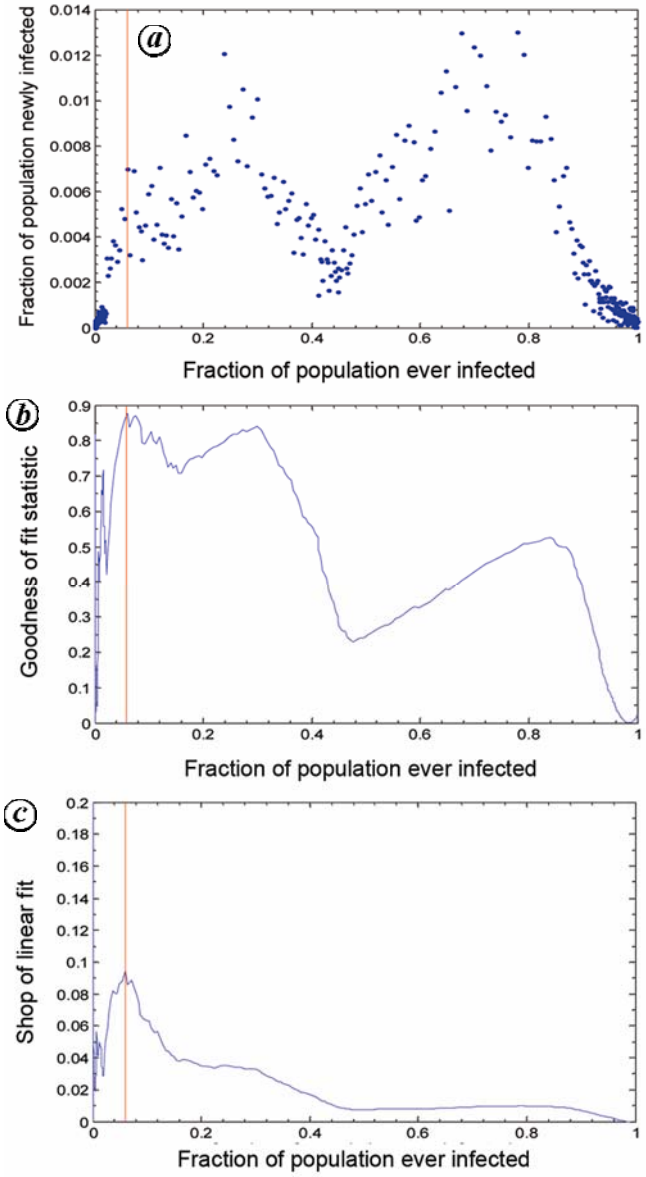

Figure 5. ***a***, Epidemic pattern: fraction of population newly infected against fraction of population ever infected; ***b***, Evaluation of goodness-of-fit statistic (coefficient of variation, R^2) of the successive linear regression; ***c***, Evaluation of the slope. The red vertical line indicates the point, 80th percentile to the left of which is considered as the cut-off point.

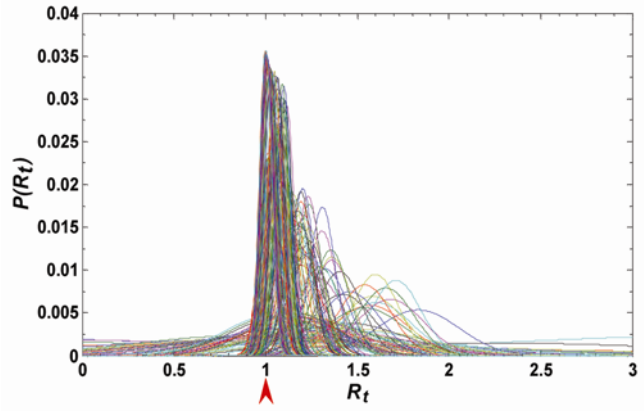

Figure 6. The complete density plot of posterior distribution of R_t through sequential Bayesian estimate from successive daily iteration.

Appendix 1. Determination of cut-off point

Determination of the cut-off point is a crucial part of method-1. The usual process can be figured out through the linear fit[11] to the data, fraction of population newly infected against fraction of population ever infected. The goodness-of-fit statistic (coefficient of determination, R^2) is calculated considering one more data point each time. The slopes can be noted simultaneously for each fit (Figure 5).

We have considered the coefficient of determination R^2 as the goodness-of-fit and have found that the slope fluctuates more at the beginning, but later it was found to be smoothly decreasing in nature. The best estimate of the force of infection ($\boldsymbol{r}$) lies in the intermediate phase and the cut-off point of the initial linear phase should lie therein. Considering the maximum value of R^2, it can be noted that the slope is also around its maximum attainable value. However, choosing the point corresponding to the maximal of R^2 may lead to an underestimation of the reproduction number; hence it is not satisfactory. In this case, we chose the 80th percentile of the slope left to the maximal goodness-of-fit not to the accidentally high value[11].

Appendix 2. Sequential Bayesian estimate of posterior distribution

We computed the full posterior distribution of R_t from eq. (7) through sequential Bayesian estimate from successive daily iteration of observed new cases $\Delta C(t)$ by considering unbiased uniform $U(0, 3)$ as the initial prior. The product of the two probabilities on the right-hand side of eq. (7) was evaluated as a non-parametric function defined in terms of 1000 discrete bins in R between 0 and 3. The distribution of R_t tends to cluster around 1 as the epidemic span increases (Figure 6).

1. Centers for Disease Control and Prevention; 2009 H1N1 Flu: International Situation Update, 9 August 2010; http://www.cdc.gov/h1n1flu/updates/international
2. Ministry of Health and Family Welfare, Government of India, 20 May 2010; http://www.mohfw.nic.in/ and http://pib.nic.in/h1n1/h1n1.asp
3. World Health Organization, Pandemic (H1N1) 2009 – update 96. Available from: http://www.who.int/csr/don/2010_04_16/en/index.html
4. Anderson, R. M. and May, R. M., *Infectious Diseases of Humans*, Oxford University Press, Oxford, 1991.
5. Chowell, G., Nishiura, H. and Bettencourt, L. M. A., Comparative estimation of the reproduction number for pandemic influenza from daily case notification: data. *J. R. Soc. Interface*, 2007, **4**, 155–166.
6. Diekmann, O. and Heesterbeek, J., *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, Wiley, New York, 2000.
7. Nishiura, H., Schwehm, M., Kakehashi, M. and Eichner, M., Transmission potential of primary pneumonic plague: time inhomogeneous evaluation based on historical document of transmission network. *J. Epidemiol. Commun. Health*, 2006, **60**, 642–645.
8. Nishiura, H., Chowell, G., Safan, M. and Castillo-Chavez, C., Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009. *Theor. Biol. Med. Model*, 2010, **7**, 1.
9. Boëlle, P. Y., Bernillon, P. and Desenclos, J. C., A preliminary estimation of the reproduction ratio for new influenza A (H1N1) from the outbreak in Mexico, March–April 2009. *Euro Surveill.*, 2009, **14**(19), pii: 19205.
10. Yang, Y. *et al.*, The transmissibility and control of pandemic influenza A (H1N1) virus. *Science*, 2009, **326**, 729–733.
11. Favier, C. *et al.*, Early determination of the reproduction number for vector-borne diseases: the case of dengue in Brazil. *Trop. Med. Int. Health*, 2006, **11**, 332–340.
12. Bettencourt, L. M. A. and Ribeiro, R. M., Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One*, 2008, **3**(5), e2185.
13. Lipsitch, M. *et al.*, Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 2003, **300**, 1966–1970.
14. Mills, C. E., Robins, J. M. and Lipsitch, M., Transmissibility of 1918 pandemic influenza. *Nature*, 2004, **432**, 904–906.
15. Khakpour, M., Saidi, A. and Naficy, K., Proved viraemia in Asian influenza (Hong Kong variant) during incubation period. *Br. Med. J.*, 1969, **4**, 208–209.
16. Kilbourne, E., Influenza pandemics in perspective. *JAMA*, 1977, **237**, 1225–1228.

ACKNOWLEDGEMENTS. This work has been financially and logistically supported by UGC, New Delhi through Major Research Project (F No. 35-108/2008 (SR), dated 20 March 2009). We thank the anonymous referees for their valuable suggestions.

Received 16 November 2010; revised accepted 26 September 2011