Epidemiological dynamics of the 2009 influenza A(H1N1)v outbreak in India

T. Jesan[1,2], Gautam I. Menon[1] and Sitabhra Sinha[1,*]

[1]The Institute of Mathematical Sciences, CIT Campus, Taramani, Chennai 600 113, India
[2]Health Physics Division, Bhabha Atomic Research Centre, Kalpakkam 603 102, India

We analyse the time-series data for the onset of A(H1N1)v influenza pandemic in India during the period 1 June–30 September 2009. Using a variety of statistical fitting procedures, we obtain a robust estimate of the exponential growth rate $\langle \lambda \rangle \simeq 0.15$. This corresponds to a basic reproduction number $R_0 \simeq 1.45$ for influenza A(H1N1)v in India, a value which lies towards the lower end of the range of values reported for different countries affected by the pandemic.

Keywords: Basic reproduction number, epidemic dynamics, pandemic influenza, swine flu.

A NOVEL influenza strain termed influenza A(H1N1)v, first identified in Mexico in March 2009, has rapidly spread to different countries and is currently the predominant influenza virus in circulation worldwide[1,2]. As of 11 April 2010, it has caused at least 17,798 deaths in 214 countries[3]. The first confirmed case in India, a passenger arriving from USA, was detected on 16 May 2009 in Hyderabad. The initial cases were passengers arriving by international flights. However, towards the end of July, the infections appeared to have spread into the resident population with an increasing number of cases being reported for people who had not been abroad. As of 11 April 2010, there have been 30,352 laboratory-confirmed cases in India (out of 132,796 tested) and 1472 deaths have been reported, i.e. 5% of the cases which tested positive for influenza A(H1N1)v (ref. 4).

To devise effective strategies for combating the spread of pandemic influenza A(H1N1), it is essential to estimate the transmissibility of this disease in a reliable manner. This is generally characterized by the reproductive rate R, defined as the average number of secondary infections resulting from a single (primary) infection. A special case is the basic reproduction number R_0, which is the value of R measured when the overall population is susceptible to the infection, as is the case at the initial stage of an epidemic. Estimates of the basic reproduction number for influenza A(H1N1)v in reports published from data obtained for different countries vary widely. For example, R_0 has been variously estimated to be between 2.2 and 3.0 for Mexico[5], 1.72 for Mexico City[6], between 1.4 and 1.6 for La Gloria in Mexico[7], between 1.3 and 1.7

for the United States[8] and 2.4 for the state of Victoria in Australia[9]. The divergence in the estimates for the basic reproduction number may be a result of under-reporting in the early stages of the epidemic, or due to climatic variations. They may also possibly reflect the effect of different control strategies used in different regions, ranging from social distancing such as school closures and confinement to antiviral treatments.

In this communication, we estimate the basic reproduction number for the infections using the time-series of infections in India extracted from reported data. By assuming an exponential rise in the number of infected cases $I(t)$ during the initial stage of the epidemic when most of the population is susceptible, we can express the basic reproduction number as $R_0 = 1 + \lambda \tau$ (see e.g. Anderson and May[10]), where λ is the rate of exponential growth in the number of infections and τ the mean generation interval, which is approximately equal to 3 days[6]. Using the time-series data, we obtain the slope λ of the exponential growth using several different statistical techniques. Our results show that this quantity has a value of around 0.15, corresponding to $R_0 \simeq 1.45$.

We used data from the daily situation updates available from the website of the Ministry of Health and Family Welfare, Government of India (http://mohfw-h1n1.nic.in/). In our analysis, data up to 30 September 2009 were used, corresponding to a total of 10,078 positive cases. Note that, after 30 September 2009, patients exhibiting mild flu-like symptoms (classified as categories A and B) were no longer tested for the presence of the influenza A(H1N1) virus.

As the data exhibit very large fluctuations, with some days not showing a single case while the following days show extremely large number of cases, it is necessary to smooth the data using a moving window average. We have used an n-day moving average ($n = 2$–10), which removes large fluctuations while remaining faithful to the overall trend.

From the incidence data for the 2009 pandemic influenza in India it appears that the disease has been largely confined to the urban areas of the country. Indeed, six of the seven largest metropolitan areas of India (which together accommodate about 5% of the Indian population) (http://www.world-gazetteer.com/; retrieved on 18 February 2010) account for 7139 infected cases up to 30 September 2009, i.e. 70.8% of the dataset we have used. However, it is possible that this is a result of bias introduced by the easier accessibility to testing facilities for urban populations.

Figure 1 shows the daily number of confirmed infected cases, as well as the 5-day moving average from 1 June to 30 September 2009, for the country as a whole and the six major metropolitan areas which showed the highest incidence of the disease: Hyderabad, Delhi, Bangalore, Mumbai, Chennai and Pune. The adjoining map shows the geographic locations of these six cities. In the period
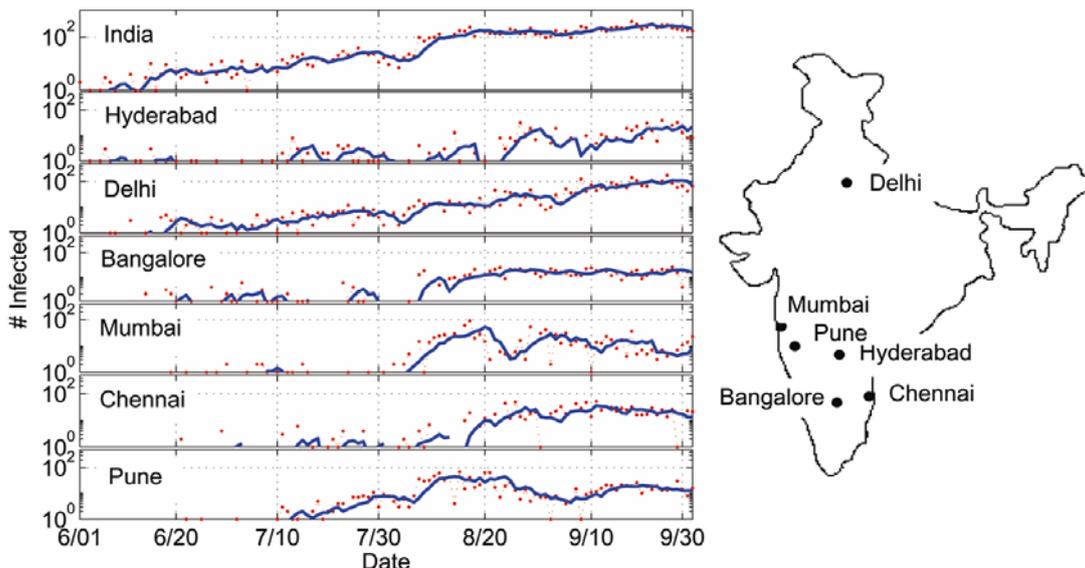
Figure 1. Time-series of the number of infected cases (#Infected), of influenza A(H1N1)v showing the daily data (dotted) as well as the 5-day moving average (solid line) for India and the six metropolitan areas with the highest number of infections (whose geographic locations are shown in the adjoining map). The period shown is from 1 June to 30 September 2009. At the beginning of this period most of the infected people were arriving from abroad, while at the end of it the infection was entrenched in the local population. The data reveal that almost all the cities showed a simultaneous increase in the number of infections towards the end of July and the beginning of August. This is manifested as a sudden rise in '#Infected' for India as a whole (note the semi-logarithmic scale), and can be taken as the period in which the infection started spreading in the resident population.
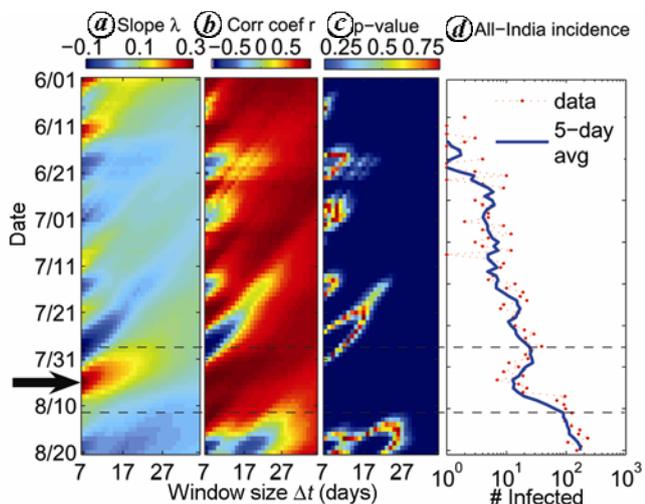

Figure 2. *a*, Exponential slope λ estimated from the time-series data of the number of infected cases (#Infected), averaged over a 5-day period to smoothen the fluctuations (*d*, solid curve). λ is calculated by considering the number of infected cases over a moving window having different sizes (Δt), ranging between 7 and 36 days. By moving the starting point of the window across the period 1 June–20 August (in steps of 1 day) and calculating the best-fit linear slope of the data on a semi-logarithmic scale (i.e. time in normal axis, number of infections in logarithmic axis), we obtain an estimate of λ. The arrow indicates the region between 28 July and 12 August (region within the broken lines), which shows the largest increase in the number of infections within the period under study, corresponding to the period when the epidemic broke out in the resident population. Over this time-interval, the average of λ is calculated for the set of starting dates and window sizes over which (*b*) the correlation coefficient r between log(#Infected) and t, is greater than r_{cutoff} (we consider $0.75 < r_{cutoff} < 1$ in our analysis), and *c*, the measure of significance for the correlation $p < 0.01$.

up to July 2009, infections were largely reported in people arriving from abroad. There is a marked increase in the number of infections towards the end of July and the beginning of August 2009 in all of these cities (note that the ordinate is in logarithmic scale). This is manifest as a sudden rise in the number of infected cases for the country as a whole, implying that the infection started spreading in the resident population in the approximate period of 28 July–August 12.

Figure 2 *a* shows the exponential slope λ estimated in the following way. The time-series of the number of infections is first smoothed by taking a 5-day moving average. The resulting smoothed time-series is then used to estimate λ by a regression procedure applied to the logarithm of the number of infected cases [log(#infected)] across a moving window of length Δt days. The origin of the window is varied across the period 1 June–20 August (in steps of 1 day). We then repeat the procedure by varying the length of the window over the range of 7–36 days. To quantify the quality of regression, we calculate the correlation coefficient r (Figure 2 *b*) between log (#Infected) and time (in days), and its measure of significance p (Figure 2 *c*). The correlation coefficient r is bounded between –1 and 1, with a value closer to 1 indicating a good fit of the data to an exponential increase in the number of infections. The measure of significance of the fitting is expressed by the corresponding p value, which expresses the probability of obtaining the same correlation by random chance from uncorrelated data. The average

of the estimated exponential slope λ is obtained by taking the mean of all values of λ obtained for windows originating between 28 July–12 August and of various sizes, for which the correlation coefficient $r > r_{cutoff}$ (we consider $0.75 < r_{cutoff} < 1$ in our analysis) and the measure of significance $p < 0.01$. For comparison, we show again in Figure 2 d the number of infected cases of H1N1 in India (dotted) together with its 5-day moving average (solid line). The horizontal broken lines running across the figure indicate the period between 28 July and 12 August, which exhibited the highest increase in the number of infections within the period under study (from 1 June to 30 September).

Figure 3 shows the average exponential slope $\langle \lambda \rangle$ as a function of r_{cutoff}, calculated for the original data and for different periods n over which the moving average is taken ($n = 2, 3, 4, 5$ and 10). For $n = 3–5$, the data show a similar profile indicating the robustness of the estimate of the average exponential slope $\langle \lambda \rangle$ with respect to different values of n. The sudden increase in $\langle \lambda \rangle$ around $r_{cutoff} \simeq 0.9$ implies that beyond this region the slope depends sensitively on the cut-off value. Considering the region where the variation is smoother gives an approximate value $\lambda \sim 0.15$, corresponding to a basic reproduction number
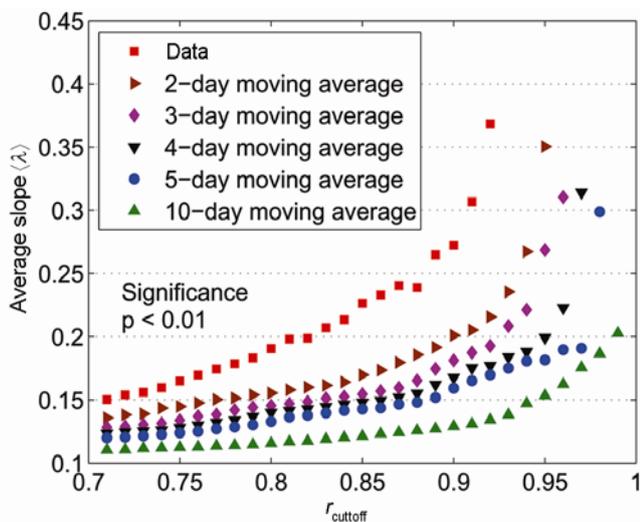
for the epidemic $R_0 = 1 + \lambda\tau \simeq 1.45$, assuming the mean generation interval, $\tau = 3$ days.

We compute the confidence bounds for the estimate of R_0 from the 5-day moving average time-series using the 'confint' function of the scientific software MATLAB (http://www.mathworks.com/). This function generates the goodness-of-fit statistics using the solution of the least squares fitting of log(#Infected) to a linear function. It results in a mean value $\langle \lambda \rangle = 0.16$, with the corresponding 95% confidence intervals calculated as [0.116, 0.206], consistent with our previous estimate of $R_0 \simeq 1.45$.

We have also used bootstrap methods to estimate the exponential slope λ. This involves selecting random samples with replacement from the data such that the sample size equals the size of the actual dataset. The same analysis that was performed on the empirical data is then repeated on each of these samples. The range of the estimated values λ' calculated from the random samples allows determination of the uncertainty in estimation of λ. Figure 4 a shows the average $\langle \lambda' \rangle$, calculated for different periods (with abscissa indicating the starting date and the symbols indicating the duration of the period) from the 5-day moving average time-series data of infected cases. The curves corresponding to the periods of different durations (14–16 days) intersect around 31 July 2009, indicating that the value of the average exponential slope is relatively robust with respect to the choice of the period about this date. The average value of the bootstrap estimates λ' at the intersection of the three curves is 0.15, in agreement with our earlier calculations of λ.
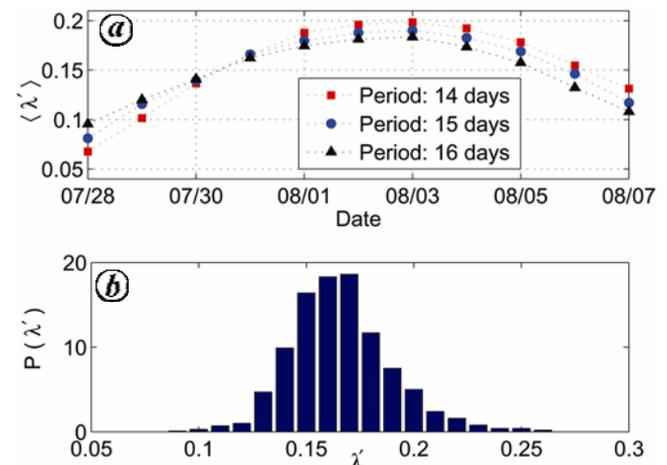

Figure 3. Average slope $\langle \lambda \rangle$ of the variation in log(#Infected) with time t, as a function of the threshold of correlation coefficient, r_{cutoff}, used to filter the data. The averaging is performed for infections occurring within the period 28 July–12 August (for details see caption to Figure 2). Different symbols indicate the actual daily time-series data (squares) and the data smoothed over a moving n-day period, with $n = 2$ (right-pointed triangle), 3 (diamond), 4 (inverted triangle), 5 (circle) and 10 (triangle). The significance of the correlation between log(#Infected) and time t, $P < 0.01$ for all data points used in performing the average. Note that for $n = 3–5$ the data show similar profiles for variation of $\langle \lambda \rangle$ with r_{cutoff}, indicating the robustness of the estimate with respect to different values of n used. The sudden increase in the value of the average slope around $r_{cutoff} \simeq 0.9$, implies that beyond this region the slope depends sensitively on the cut-off value. Considering the region where the variation is more gradual gives us an approximate value of the slope $\lambda \sim 0.15$, corresponding to a basic reproduction number $R_0 \simeq 1.45$.

Figure 4. a, Averages of the bootstrap estimates for the exponential slope, λ', calculated for different periods (with the abscissa indicating the starting date and the symbols indicating the duration) from the 5-day moving average time-series data of infected cases in India. Curves corresponding to the periods of different durations (14–16 days) intersect around 31 July 2009, indicating that the value of the average exponential slope is relatively robust with respect to the choice of the period about this date. b, Distribution of bootstrap estimates of the exponential slope for the period 31 July–15 August 2009. The average slope $\langle \lambda' \rangle$ obtained from 1000 bootstrap samples is 0.166 with a standard deviation of 0.024, which agrees with the approximate value of $\lambda = 0.15$ (corresponding to $R_0 = 1.45$) calculated in Figure 3.

Table 1. Regional variation of basic reproduction number for 2009 influenza A(H1N1)v in India

Region/city	Period[a]	$\langle\lambda\rangle$	R_0
Pune	30/07–14/08	0.25 ± 0.04	1.74 ± 0.14
Mumbai	05/08–20/08	0.22 ± 0.06	1.65 ± 0.18
Delhi	13/08–28/08	0.12 ± 0.02	1.36 ± 0.06
Southern region[b]	15/08–30/08	0.11 ± 0.02	1.34 ± 0.05

R_0 is estimated by the method of exponential curve fitting from 5-day moving averages of incidence data for different regions/cities. In each case, bootstrap estimates yielded similar values.
[a]For each region/city, the time interval over which R_0 is determined is chosen on the basis of exhibiting the highest rise in disease incidence.
[b]The Southern region comprises the cities of Bangalore, Chennai and Hyderabad.

Figure 4 *b* shows the distribution of the bootstrap estimates of the exponential slope for a particular period, 31 July–15 August 2009. The average slope $\langle\lambda'\rangle$ obtained from 1000 bootstrap samples for this period is 0.166 with a standard deviation of 0.024, which indicates that the spread of values around the average estimate of $\langle\lambda'\rangle = 0.15$ is not large. This confirms the reliability of the estimated value of the exponential slope, and hence of our calculation of the basic reproduction number.

In addition to estimating R_0 for the entire country, we have also separately evaluated the basic reproduction number for the different regions in which the epidemic occurred (Table 1). It may appear surprising that there was a large number of infections in Pune (1238 positive cases up to 30 September), despite it being less well-connected to the other major metropolitan cities of India, in comparison to urban centres that did not show a high incidence of the disease. For example, the Kolkata metropolitan area, which has a population around three times that of the Pune metropolitan area (http://www.world-gazetteer.com/; retrieved on 18 February 2010), had only 113 positive cases up to 30 September. This could possibly reflect the role of local climatic conditions: Pune, located at a relatively higher altitude, has a generally cooler climate than most Indian cities. In addition, the close proximity of Pune to Mumbai and the high volume of road traffic between these two cities could have helped in the transmission of the disease. Another feature pointing to the role of local climate is the fact that in Chennai, most infected cases were visitors from outside the city, while in Pune, majority of the cases were from the local population, even though the total number of infected cases listed for the two cities in our dataset is comparable (928 in Chennai and 1213 in Pune). This suggests the possibility that the incidence of the disease in Pune could have been aided by its cool climate, in contrast to the hotter climate of the coastal city of Chennai. The rapid spread of the disease in Pune may also have originated in transmission amongst the large crowds of people who had gathered in the H1N1 testing centres, given that the numbers appearing for testing here were much larger than elsewhere.

The calculation of R_0 for India assumes well-mixing of the population (i.e. homogeneity of the contact structure) among the major cities in India. Given the rapidity of travel between the different metropolitan areas via air and rail, this may not be an unreasonable assumption. However, some local variation in the development of the epidemic in different regions can indeed be seen (Figure 1). Around the end of July, almost all the cities under study showed a marked increase in the number of infected cases – indicating spread of the epidemic in the local population. This justifies our assumption of well-mixing in the urban population over the entire country for calculating the basic reproduction number.

To conclude, we stress the implications of our finding that the basic reproduction number for pandemic influenza A(H1N1)v in India lies towards the lower end of the values reported for other affected countries. This suggests that season-to-season and country-to-country variations need to be taken into account to formulate strategies for countering the spread of the disease. Evaluation of the reproductive rate, once control measures have been initiated, is vital in determining the future pattern of spread of the disease.

1. 2009 H1N1 flu: international situation update, 22 January 2010; http://www.cdc.gov/h1n1flu/updates/international
2. Jameel, S., The 2009 influenza pandemic. *Curr. Sci.*, 2010, **98**, 306–311.
3. World Health Organization, Pandemic (H1N1) 2009 – update 96, 16 April 2010; http://www.who.int/csr/don/2010_04_16/en/index.html
4. Ministry of Health and Family Welfare, Government of India, Situation update on H1N1, 11 April 2010; http://mohfw-h1n1.nic.in/documents/PDF/SituationalUpdatesArchives/april2010/Situational%20Updates%20on%2011.04.2010.pdf
5. 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.
6. Cruz-Pacheco, G. *et al.*, Modelling of the influenza A(H1N1)v outbreak in Mexico City, April–May 2009, with control sanitary measures. *Euro Surveill.*, 2009, **14**(26), pii: 19254.
7. Fraser, C. *et al.*, WHO Rapid Pandemic Assessment Collaboration, Pandemic potential of a strain of influenza A(H1N1): early findings. *Science*, 2009, **324**, 1557–1561.
8. Yang, Y. *et al.*, The transmissibility and control of pandemic influenza A(H1N1) virus. *Science*, 2009, **326**, 729–733.
9. McBryde, E. *et al.*, Early transmission characteristics of influenza A(H1N1)v in Australia: Victorian state, 16 May–3 June 2009. *Euro Surveill.*, 2009, **14**(42), pii: 19363.
10. Anderson, R. M. and May, R. M., *Infectious Diseases of Humans*, Oxford University Press, Oxford, 1992.