Supplementary web appendix

Web appendix I – Million Death Study research ethics

The full protocol of the Million Death Study (MDS) is available at: http://www.cghr.org/project.htm. The MDS is conducted within the Registrar General of India's Sample Registration System (SRS) [1], a large, routine demographic survey and the primary system for the collection of Indian fertility and mortality data since 1971. SRS enrolment is on a voluntary basis, and its confidentiality and consent procedures are defined as part of the Registration of Births and Deaths Act, 1969. Oral consent was obtained in the first SRS sample frame. The new SRS sample obtains written consent at the baseline. Families are free to withdraw from the study, but the compliance is close to 100%. The study poses no or minimal risks to enrolled subjects. All personal identifiers present in the raw data are anonymized before analysis. The study has been approved by the review boards of the Post-Graduate Institute of Medical Education and Research, the Indian Council of Medical Research, and the Health Ministry's Screening Committee. Specific written consent procedures for additional biological measurements will be added, using international guidelines [2,3].

References

1. Office of the Registrar General (2001) Compendium of India's fertility and mortality indicators 1971–1999. New Delhi: Office of the Registrar General. 172 p.
2. Government of India (1998) Revised Guidelines for Exchange of Human Biological Material for Biomedical Research Purposes. Indian Journal of Pharmacology 30: 56-57.
3. Clayton EW, Steinberg KK, Khoury MJ, Thomson E, Andrews L, Kahn MJ, Kopelman LM, Weiss JO. (1995) Informed consent for genetic research on stored tissue samples. JAMA. 274: 1786-1792.

Web appendix II - Model structure

We adapted DynaMICE (Dynamic Measles Immunization Calculation Engine), a compartmental model of measles transmission and vaccination stratified by age, which was first developed to study measles in high burden countries [1]. We reproduce in this web appendix, from Verguet et al. Vaccine 2015 [1], the main model equations from DynaMICE.

The population in the model is categorized as: susceptible (S), infected (I), recovered (R), vaccinated susceptible (VS), vaccinated infected (VI), and vaccinated recovered (VR). Below are the differential equations used:

$$\frac{\partial S}{\partial t} + \frac{\partial S}{\partial a} = -\lambda S - \mu(a)S - \kappa(t,a)S \quad ,$$

$$\frac{\partial I}{\partial t} + \frac{\partial I}{\partial a} = \lambda S - \mu(a)I - \gamma I - \kappa(t,a)I \quad ,$$

$$\frac{\partial R}{\partial t} + \frac{\partial R}{\partial a} = \gamma I - \mu(a)R - \kappa(t,a)R \quad ,$$

$$\frac{\partial VS}{\partial t} + \frac{\partial VS}{\partial a} = -\lambda VS - \mu(a)VS + (1-\tau)\kappa(t,a)S - \tau\kappa_2(t,a)VS \ ,$$

$$\frac{\partial VI}{\partial t} + \frac{\partial VI}{\partial a} = \lambda VS - \mu(a)VI + \kappa(t,a)I - \gamma VI \ ,$$

$$\frac{\partial VR}{\partial t} + \frac{\partial VR}{\partial a} = \gamma VI - \mu(a)VR + \tau\kappa(t,a)S + \tau\kappa_2(t,a)VS \ ,$$

and the accompanying boundary conditions:

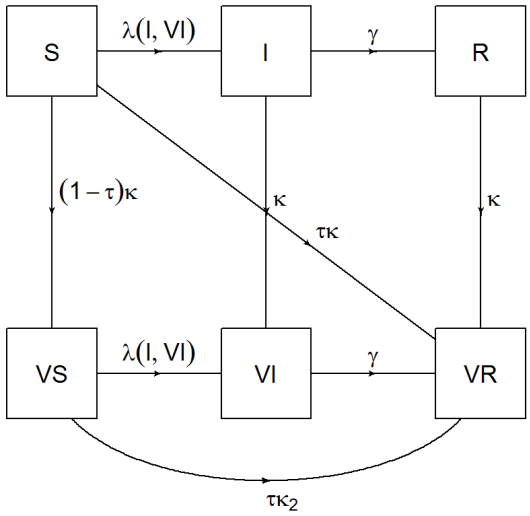
$$S(t,a) + I(t,a) + R(t,a) + VS(t,a) + VI(t,a) + VR(t,a) = N(t,a) \ ,$$

$$S(t,0) = \nu N(t,0) \ .$$

t is the time (in years), a is the age (in years), and N is the total population. $\kappa(t,a)$ is the coverage of immunization for individuals vaccinated for the first time, through either routine immunization (MCV1) (where $a < 1$) or through supplemental immunization activity (SIA) (where $a < 1$). $\kappa_2(t,a)$ is the coverage of immunization for individuals vaccinated for the second time, through SIA only (where $a < 1$). λ is the force of infection and depends on I and VI; ν is the birth rate in the population, $\mu(a)$ is the mortality rate at age a. γ is the infectiousness

period of measles; τ is the effectiveness of measles vaccine: $\tau = 0.85$ for individuals vaccinated once through MCV1, $\tau = 0.95$ for individuals vaccinated once through SIA, and $\tau = 0.98$ for individuals vaccinated twice.

Figure S1. Model of vaccine action.



Source: Reproduced from Verguet et al. Vaccine 2015 [1]. Licensed under Creative Commons Attribution (CC BY-NC-ND 4.0).

References

1. Verguet S, Johri M, Morris SK, Gauvreau CL, Jha P, Jit M. Controlling measles using supplemental immunization activities: a mathematical model to inform optimal policy. Vaccine 2015; 33:1291-1296.

Web appendix III – Parameter estimation

1. Calibration of the model

Figure S1 displays the variation in the oscillation period of estimated measles cases as the basic reproduction number and the amplitude of the forcing term vary in the model.

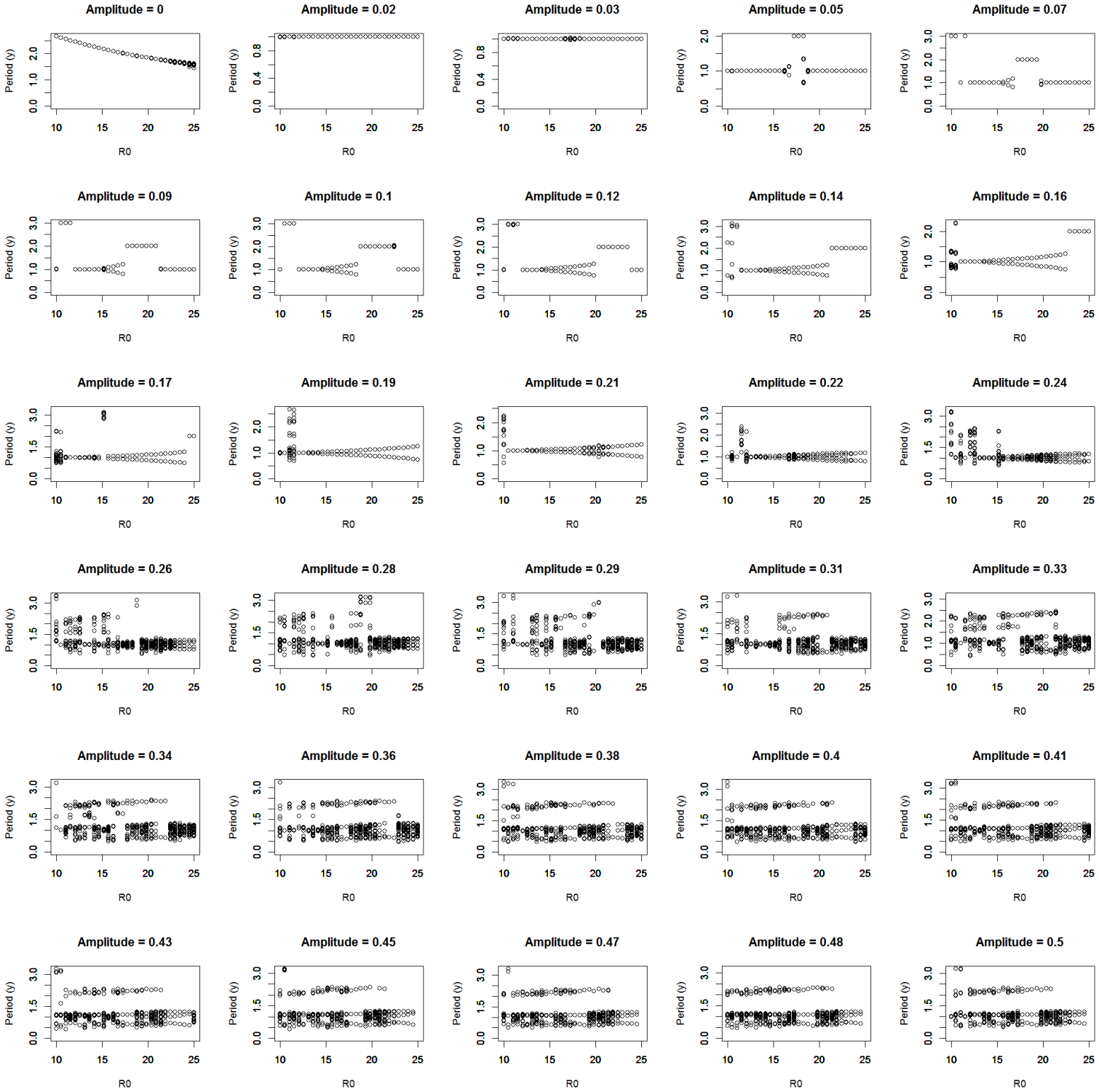
Table S1 below displays the estimates of measles deaths we used in our estimation of case fatality risk (CFR), taken from Morris et al. [1].

Table S1. Under-five measles deaths estimates, for India, Bihar, and Uttar Pradesh used in our estimation procedure.

	Estimated number of measles deaths (1-59 months)	**99% confidence interval**
India	92,000	63,200 to 137,200
Bihar	10,600	6,400 to 18,200
Uttar Pradesh	35,300	26,400 to 47,200

Source: Morris et al. [1].

Figure S1. Variation in the oscillation period of estimated measles cases as the basic reproduction number R_0 was varied between 10 and 25 and the amplitude of the forcing term a_0 (e.g. Amplitude) was varied between 0 and 0·5 in the model.



2. Coherence analysis of the measles deaths data and measles cases estimated by the model.

To compare the correlation or association in frequency f (or period) between two time series, we used the magnitude-squared coherence or coherence defined as:

$$C_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)}, \qquad (1)$$

where $G_{xy}(f)$ is the cross-spectral density between time series x and y, and $G_{xx}(f)$ and $G_{yy}(f)$ are the auto-spectral densities of x and y, respectively. When there is no relationship between x and y for a given f then $C_{xy}(f) = 0$. The significance of size α (e.g. $\alpha = 0.05$) is given as:

$$C^2 = 1 - \alpha^{1/(n-1)}, \qquad (2)$$

where C^2 is the coherence and n is the number of degrees of freedom dependent on the number of realizations for x and y [2]. (2) was derived assuming normally distributed data in close agreement with a Monte Carlo approach [3].

To be reliable and interpretable, coherence functions (1) must be smoothed across either multiple sub-averages or adjacent frequencies (or both). Yet, there are trade-offs: the more smoothed the more reliable the coherence function, but increasing smoothing reduces spectral resolution. A small number of frequencies were investigated (five, between 0·50 and 1·50 years) hence a minimal amount of smoothing (moving average of two weeks) was retained to maintain good spectral resolution.

Each combination (R_0; a_0) was simulated until equilibrium was reached. To compare the coherence of two time series, the two times series must overlap the same time period (3 years) and present the same time step (1 week). Thus, the model outputs were converted into weekly measles cases, from which $n = 1000$ time series of three-year length each were randomly extracted (to ensure that all dynamics were captured). Then, n coherence functions were estimated using the R package "spectrum" and the top 10% coherent simulated time series of measles cases were selected.

For five periods (0·50, 0·60, 0·75, 1·00, and 1·50 years) selected from the spectral analysis of the original measles mortality data time series (figure 1 in the main text), we plotted (figure S2) the coherence estimated for each (R_0; a_0) value against that of the equivalent estimated case fatality risk (CFR) among under-five children expressed in %. Each (R_0; a_0) yielded an estimated number of measles cases among under-fives which was subsequently compared to the total number of measles under-five deaths as procured by Morris and colleagues [1] for India as a whole and Indian states in 2005, to derive a CFR estimate.

Finally, to obtain the calibrated (R_0; a_0) and a subsequent CFR estimate, we used the pooled coherence methodology of Amjad et al. [4]. This method combines several independent coherence estimates into a single representative estimate. We used it to combine the results from the different periods yielded by the spectral analysis (figure 1 in the main text) of the original measles mortality time series. We recorded the range of CFRs produced from coherence estimates that were significant and the CFRs with the corresponding greatest coherence. We sampled (n = 10,000) the distribution of modeled cases, with replacement, using the probability of the corresponding coherence level (i.e. case estimates from a (R_0; a_0) combination that had a coherence value of 0·8 would have twice the probability of being sampled than with a coherence of 0·4). This allowed us to produce a distribution of cases. We

then combined this distribution of cases with the distribution of under-five deaths during the same time period (given by Morris et al. [1]; table S1) to finally produce a distribution of the CFR for each region (see figures S4, S7 and S10 below).

To use the pooled coherence methodology, we produced coherence estimates across the five periods/frequencies of interest and then pooled them by summing across the range of frequencies before estimating the magnitude of the coherence:

$$\hat{C}_{xy}(f) = \frac{\left|\sum_{i=1}^{k} \hat{G}_{x_i y_i}(f)_{df}\right|^2}{\left(\sum_{i=1}^{k} \hat{G}_{x_i x_i}(f)_{df}\right)\left(\sum_{i=1}^{k} \hat{G}_{y_i y_i}(f)_{df}\right)}, \qquad (3)$$

where we summed across *k* realizations and *df* was the number of degrees of freedom in the spectral estimate. Maitaining the assumption of independence the upper 95% confidence limit between the number of realizations was given by:

$$\hat{C}^2 = 1 - (0.05)^{\frac{1}{\sum df - 1}}, \qquad (4)$$

where $\sum df$ was the total number of degrees of freedom in the pooled coherence estimate. The highest pooled coherence was estimated at 0·67 and yielded both a CFR estimate of 0·66% (95% CI: 0·47-0·94) (figure S3) and estimated values for (R_0; a_0) (figure S4).

Similarly, detailed results are displayed for Bihar (figures S5, S6, and S7) where the highest pooled coherence was estimated at 0·87, and for Uttar Pradesh (figures S8, S9 and S10) where the highest pooled coherence was also estimated at 0·62.

One challenge in our analysis is whether a three-year time series of measles deaths (i.e. from the Million Death Study data) contains sufficient information to recover R_0 and a_0 of the underlying

infection process. To validate the ability of the pooled coherence procedure to estimate R_0 and a_0 parameters over a three-year window, we ran the model for a given (R_0; a_0) combination and took three hundred three-year time series from that simulation. We then looked at the maximum coherence estimates for each of the three hundred samples, calculated using the procedure described above, noting the (R_0; a_0) combination at the maximum. The distribution of (R_0; a_0) parameters produced could then be used to validate the procedure; ideally, we would find that the mean of this distribution is close to the true value of R_0 and a_0.

The 300 3-year time series were created by adding stochasticity to the deterministic model, in order to capture noise due to errors in the observation/measurement process. The detailed process was as follows:

(i) Take the time series of number of cases each week over the 3 years from the deterministic model.

(ii) Simulate a time series of number of deaths each week by sampling from the distribution Bin(n,p) where p=1% (the assumed case-fatality risk) and n=number of cases in that week.

(iii) Add further noise to this time series of deaths by adding to each week a random number sampled from a uniform distribution U(0,m), with the value of m estimated below.

To estimate the parameter m in the uniform distribution U(0,m) we estimated the magnitude of noise we were seeing in the Indian (Million Death Study, MDS) data we fit the original model to:

(i) Generate a time series of measles cases from our SIR model using values of R_0 and a_0 with the highest coherence to MDS data (as described in our manuscript).

(ii) Convert this to the number of deaths by applying a case-fatality risk of 1%.

(iii) Scale the number of deaths in this time series to match an "MDS data set" by multiplying it by the ratio of the deaths in the MDS sample with the number predicted for the whole of

India in Morris et al [1]. We used the Morris et al. paper because this study used exactly the same MDS time series to estimate the number of measles cases in India during that time period.

(iv) Calculate the residuals, i.e. the difference between our scaled time series of deaths with the actual number of deaths each week in the MDS time series.

Using this process, we calculated that the average residual was 3,531 with a range of 0-19,897. Hence we sampled noise from a uniform distribution U(0,5000) which is roughly equivalent. Just to test our method to the limits, we also generated a time series with noise sampled from U(0,20000) instead. The results are shown in Figure S11 below.

Note that the raw MDS data were filtered to detrend the data using a Baxter King filter (bkfilter in R). This process has the effect of removing much of the noise at frequencies with small periods (<1 month) which are highly unlikely to be relevant to measles transmission, but does not affect the periods we were interested in (> 6 months).

With the noise term sampled from U(0,5000), we obtain R_0 of 17.0 (95% interval 14.0-18.1) and a_0 of 0.26 (0.12-0.41). With the noise term sampled from U(0,20000), we obtain R_0 of 17.04 (13.4-19.5) and a_0 of 0.25 (0.04-0.46) (Figure S12). This compares with the original R_0 of 16.7 (95% Interval: 14.6-18.6) and a_0 of 0.27 (95% Interval 0.14-0.41). Hence we find that adding uniform noise increases the uncertainty intervals around the estimates of R_0 and a_0 but does not have much effect on the central estimate. We show in Figure S13, that with the addition of uniform noise, there is little effect on the underlying harmonics.

Clearly there are ways to add noise so that it becomes difficult to estimate the underlying periodicity of the data. We could add uniform noise of such great magnitude that the underlying periodicity (the signal) is swamped by the noise; it would then be difficult to pick up any meaningful periodicity at all. We could add noise with its own periodicity or a truncated pulse

of noise, such as seasonal reporting of cases/deaths, or data sampling of different efforts at varying periods. The spectral analysis of these time series would then pick up subharmonics in addition to the underlying periodicity. This is not surprising and would merely show that the method is working as intended, i.e. picking up periodic series without making any judgment about their origin.

As this paper was drawing from the MDS which has a high quality of reporting, we believe that MDS measles mortality data present less reporting biases and thus smaller noise, which enables the use of our method.

Figure S2. Estimated coherence plotted against estimated case fatality risk (CFR) for India, for each selected period (0·5, 0·6, 0·75, 1·0, 1·5 year). The green lines delimit the range at which the coherence values are 95% significant for the CFR (also indicated in the legend). The red line indicates the upper 95% significance limit for the coherence function.

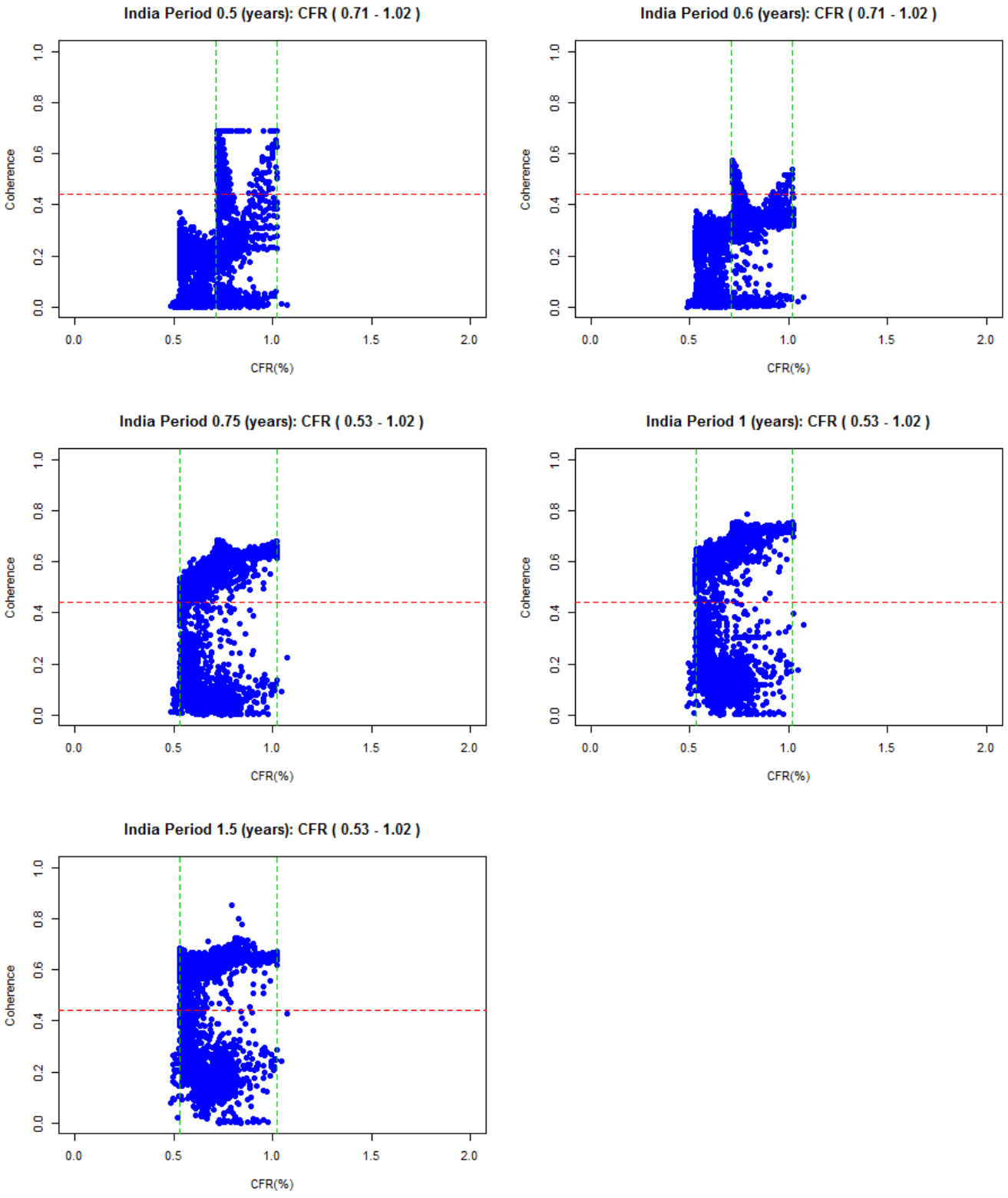


Figure S3. Pooled coherence across five periods (0·5, 0·6, 0·75, 1·0, 1·5 year) plotted against estimated case fatality risk (CFR) for India. The left and right green lines delimit the range at which the coherence values are 95% significant (0·47 and 0·95%) for the CFR. The middle green line indicates the CFR (0·66%) for which the highest pooled coherence was estimated. The red line indicates the upper 95% significance limit for the coherence estimate.

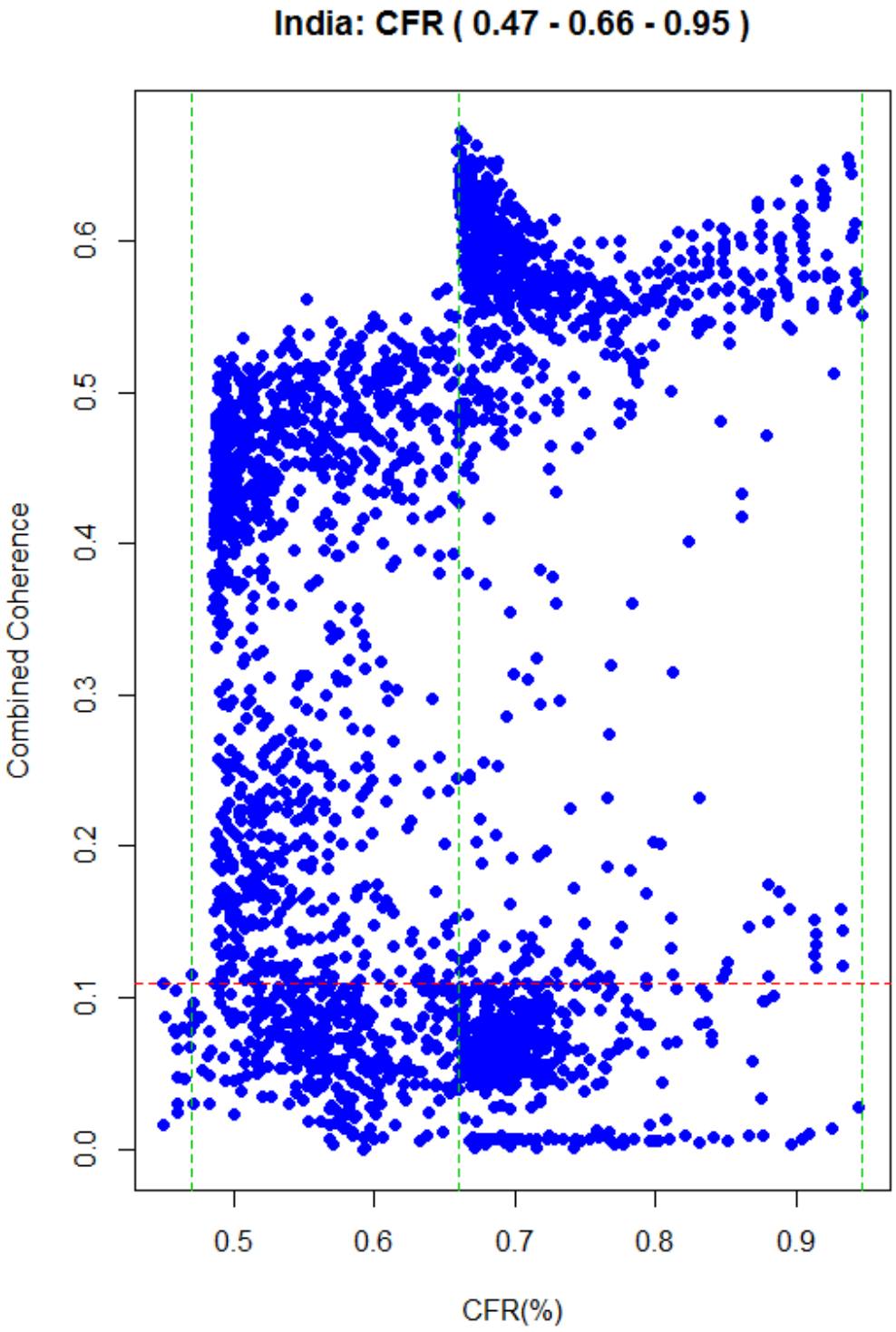
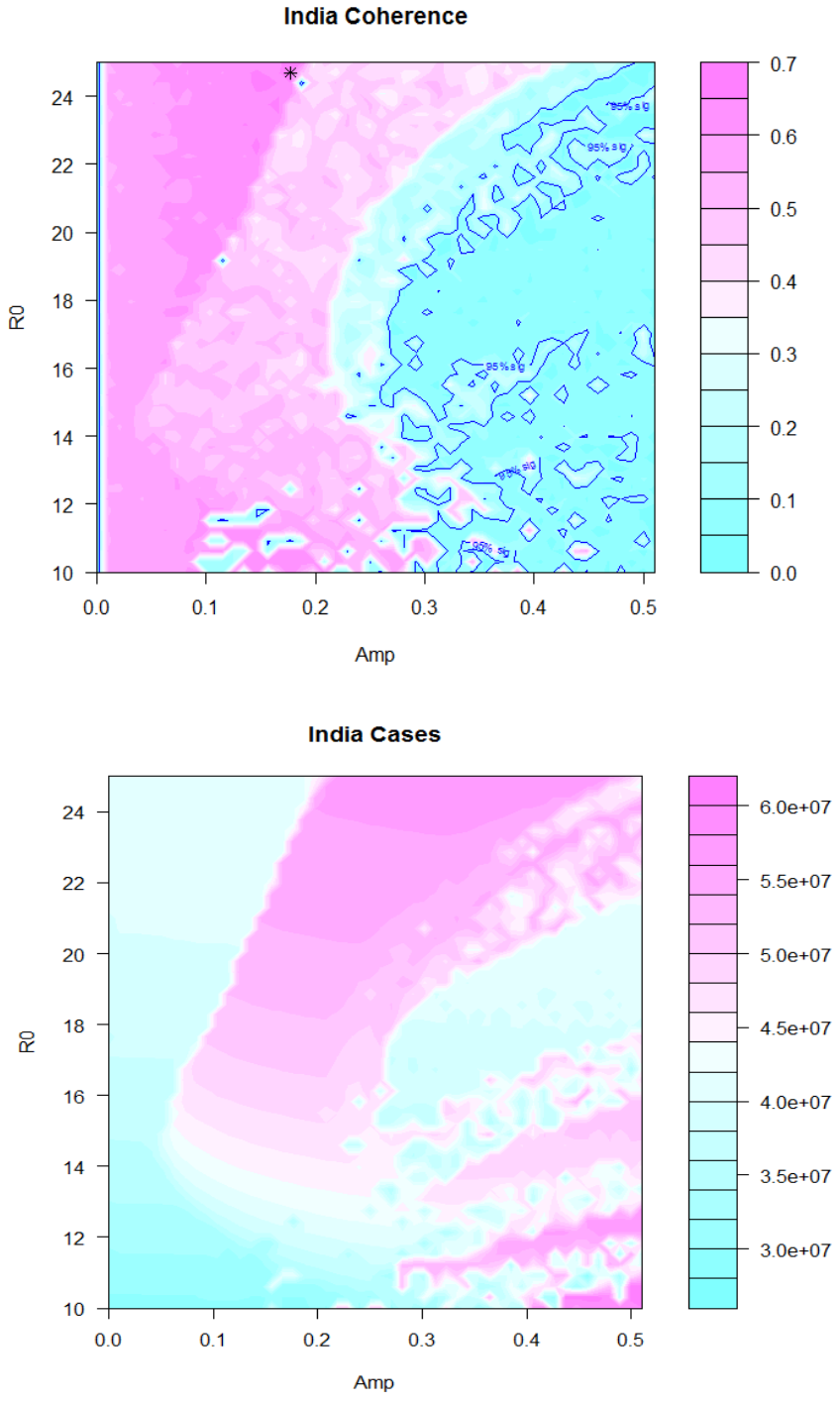
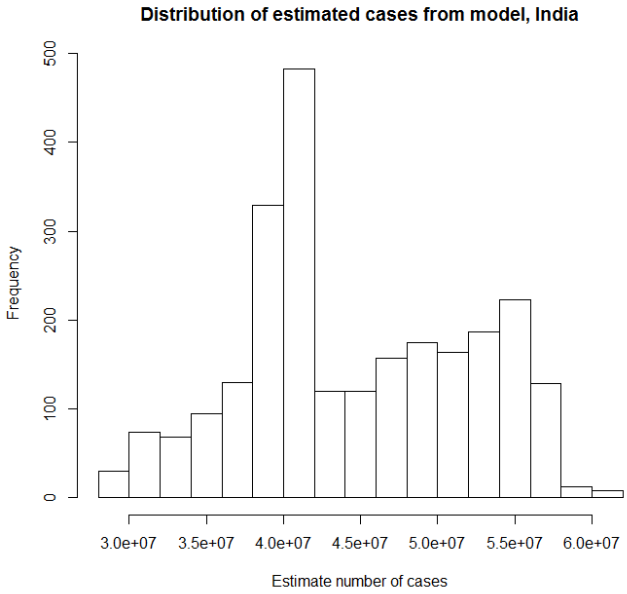
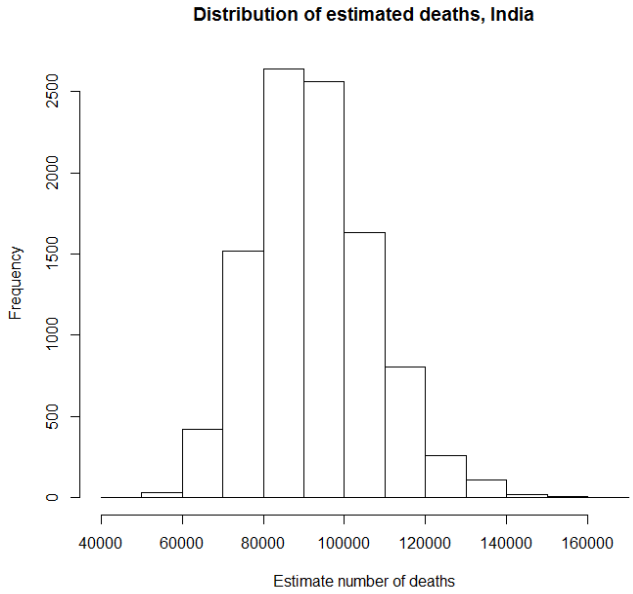


Figure S4. Pooled coherence estimate across five periods (0·5, 0·6, 0·75, 1·0, 1·5 year) and total cases over three-year period for India plotted against estimated basic reproduction number (R_0) and amplitude of the forcing term (a_0). The asterisk (*) denotes the location of the maximum combined coherence. Distributions of estimated measles cases and deaths [1] and resultant estimated case fatality risk (CFR) (0.63; CI: 0.40-1.0).




Distribution of estimated cases from model, India
Frequency
0
100
200
300
400
500
3.0e+07
3.5e+07
4.0e+07
4.5e+07
5.0e+07
5.5e+07
6.0e+07
Estimate number of cases


Distribution of estimated deaths, India
Frequency
0
500
1000
1500
2000
2500
40000
60000
80000
100000
120000
140000
160000
Estimate number of deaths

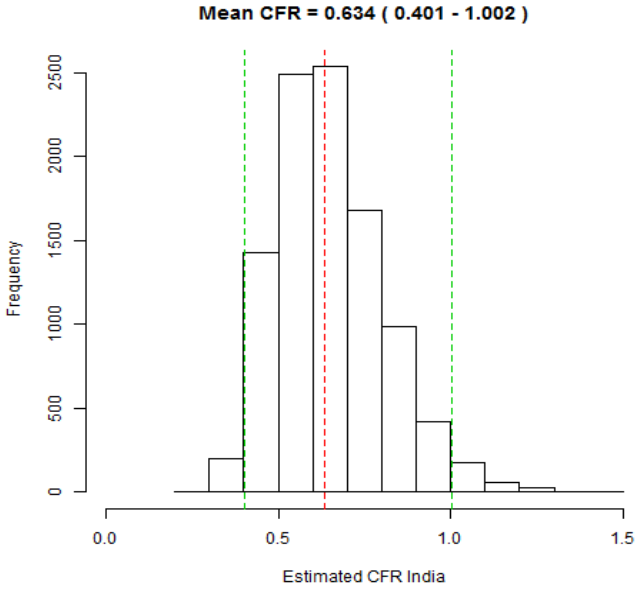

Mean CFR = 0.634 (0.401 - 1.002)
Frequency
0
500
1000
1500
2000
2500
0.0
0.5
1.0
1.5
Estimated CFR India

Figure S5. Estimated coherence plotted against estimated case fatality risk (CFR) for Bihar, for each selected period (0·5, 0·6, 0·75, 1·0, 1·5 year). The green lines delimit the range at which the coherence estimate is significant at the 95% level for the CFR (also indicated in the legend). The red line indicates the upper 95% significance limit for the coherence function.

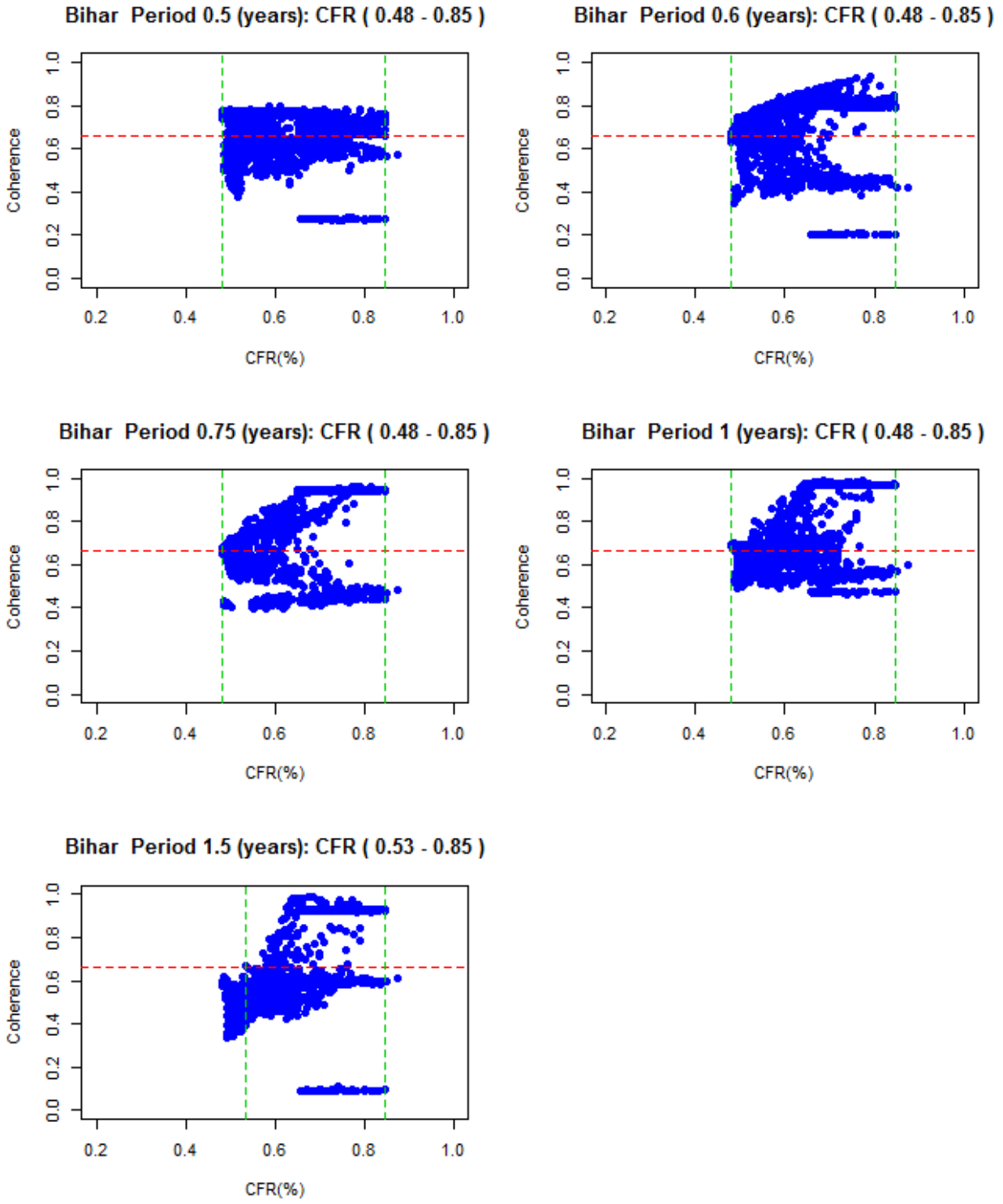


Figure S6. Pooled coherence across five periods (0·5, 0·6, 0·75, 1·0, 1·5 year) plotted against estimated case fatality risk (CFR) for Bihar. The left and right green lines delimit the range at which the coherence values are 95% significant (0·48 and 0·87%) for the CFR. The middle green line indicates the CFR (0·67%) for which the highest pooled coherence was estimated. The red line indicates the upper 95% significance limit for the coherence estimate.

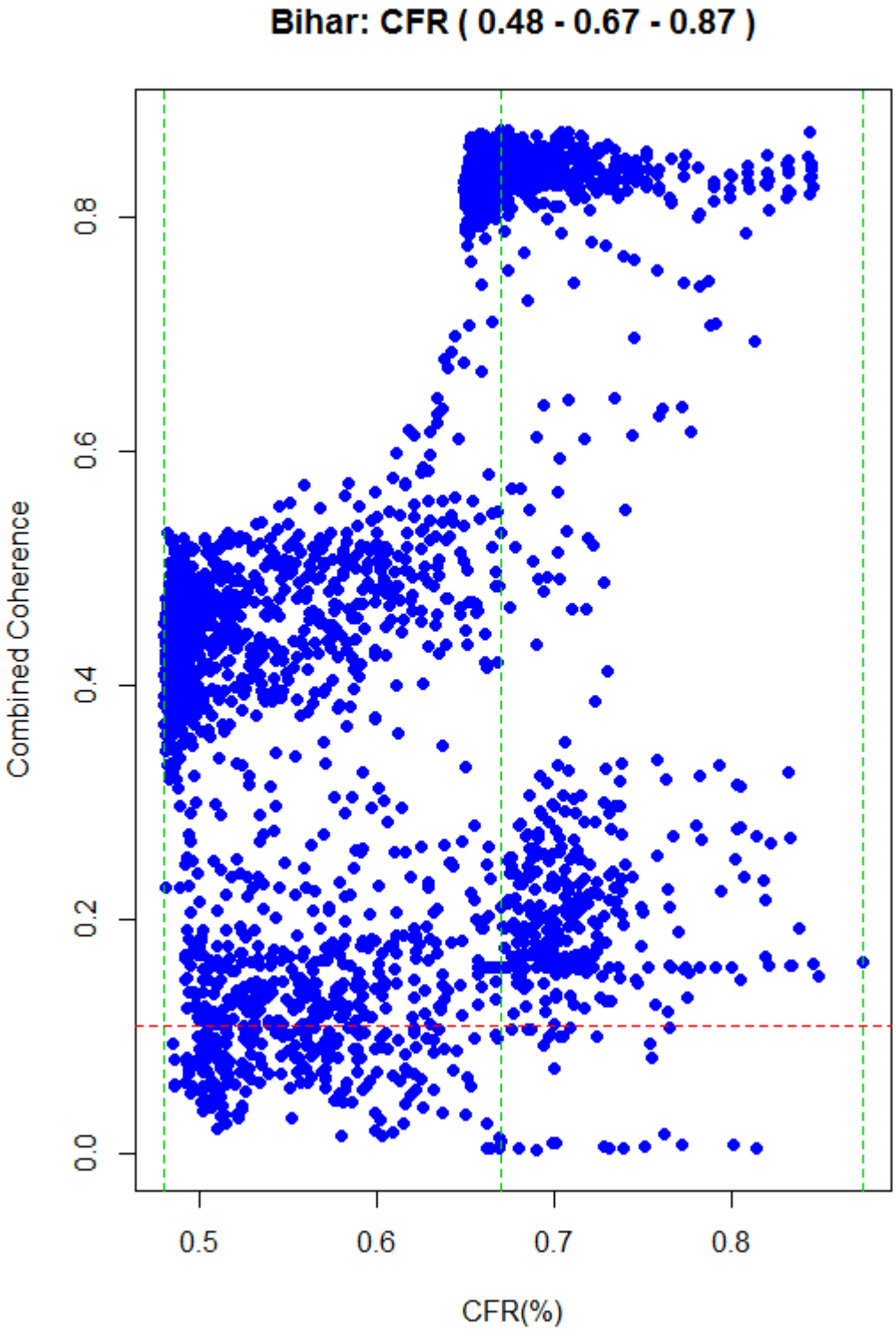
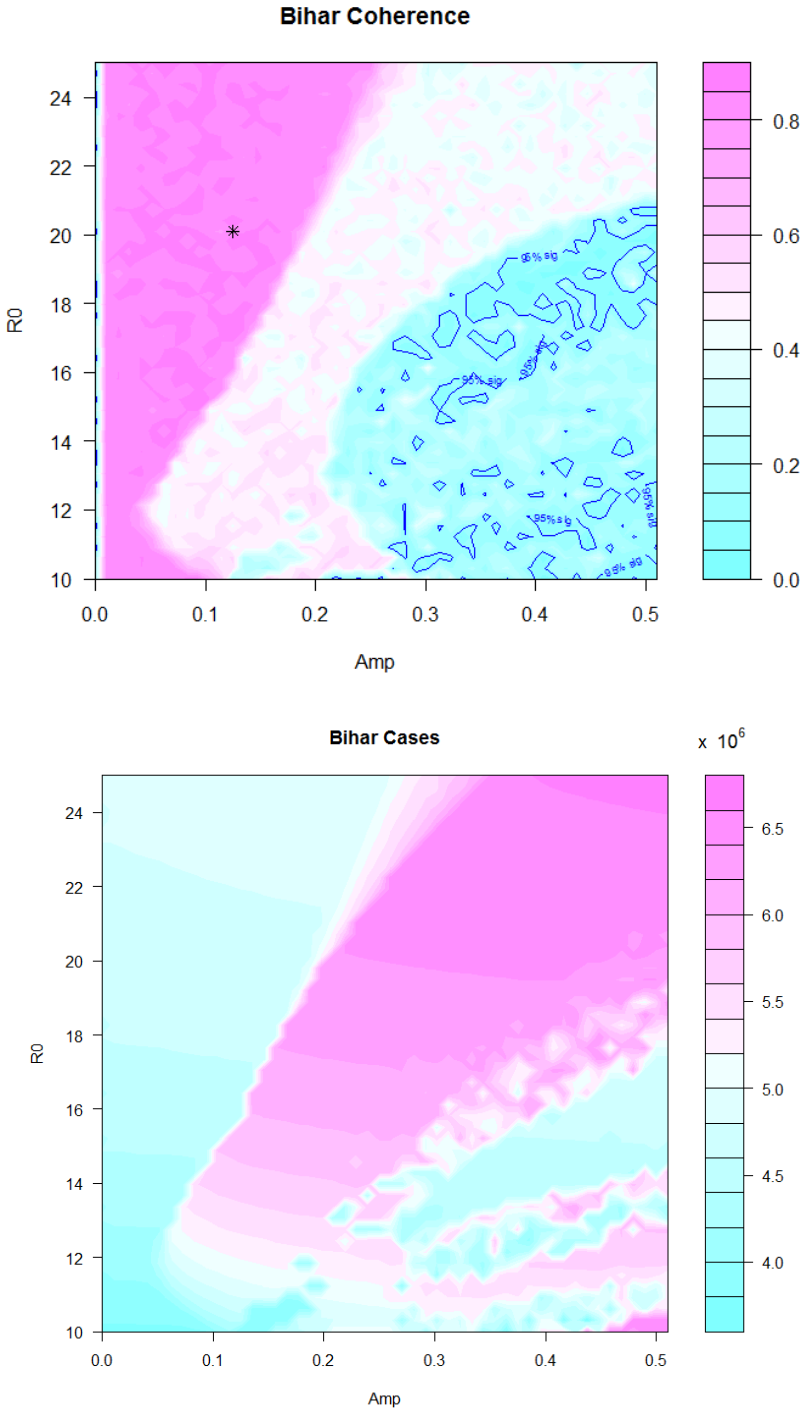
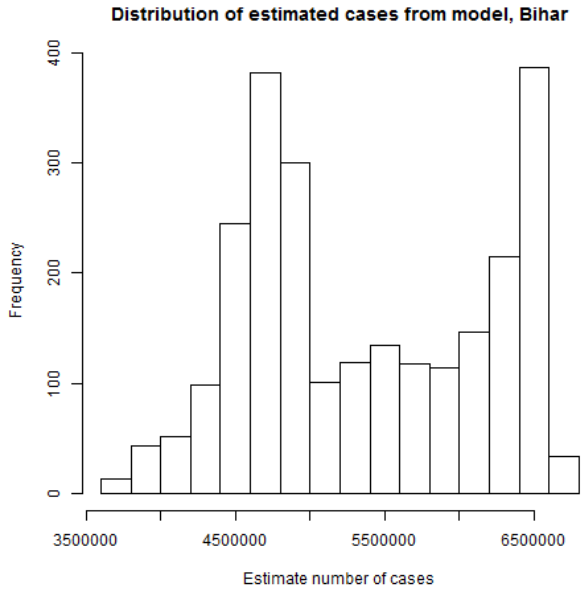
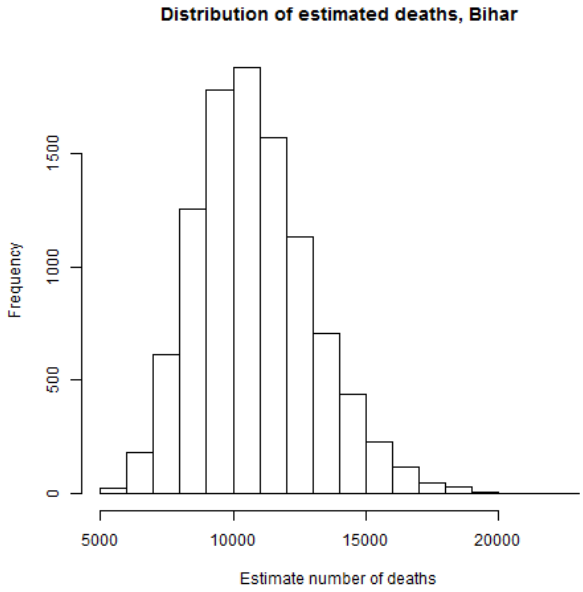


Figure S7. Pooled coherence estimate across five periods (0·5, 0·6, 0·75, 1·0, 1·5 year) and total cases over a three-year period for Bihar plotted against estimated basic reproduction number (R_0) and amplitude of the forcing term (a_0). The asterisk denotes location of the maximum pooled coherence. Distributions of estimated measles cases and deaths [1] and the resultant estimated case fatality risk (CFR) (0·70 CI: 0·46-1·06).




Distribution of estimated cases from model, Bihar
Frequency
0
100
200
300
400
3500000
4500000
5500000
6500000
Estimate number of cases


Distribution of estimated deaths, Bihar
Frequency
0
500
1000
1500
5000
10000
15000
20000
Estimate number of deaths

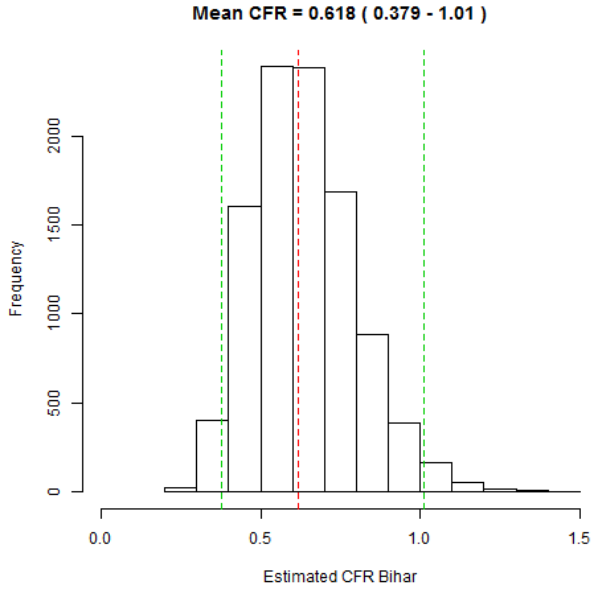

Mean CFR = 0.618 (0.379 - 1.01)
Frequency
0
500
1000
1500
2000
0.0
0.5
1.0
1.5
Estimated CFR Bihar

Figure S8. Estimated coherence plotted against estimated case fatality risk (CFR) for Uttar Pradesh, for each selected period (0·5, 0·6, 0·75, 1·0, 1·5 year). The green lines delimit the 95% confidence intervals for the CFR (also indicated in the legend). The red line indicates the upper 95% significance limit for the coherence function.

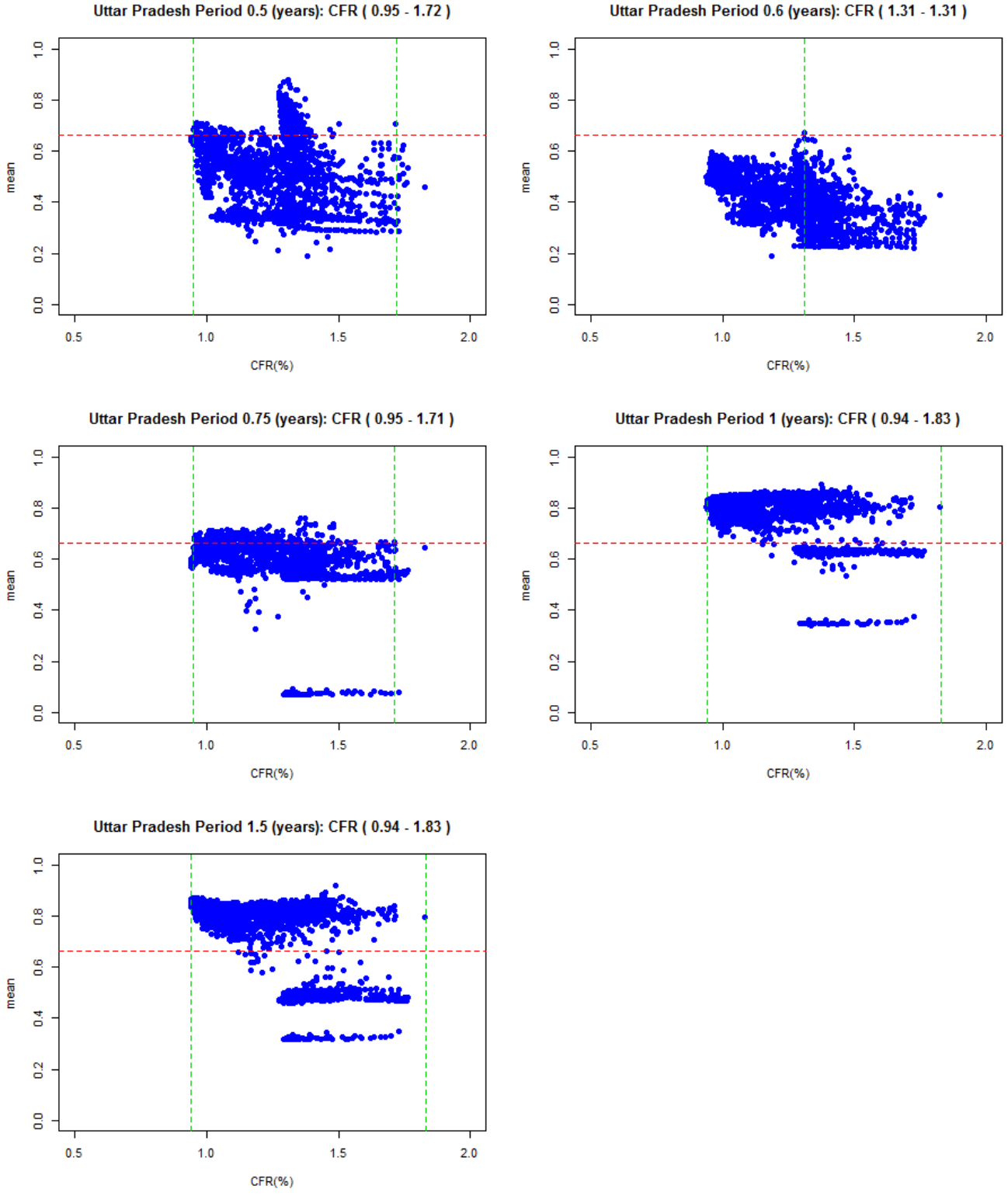


Figure S9. Pooled coherence across five periods (0·5, 0·6, 0·75, 1·0, 1·5 year) plotted against estimated case fatality risk (CFR) for Uttar Pradesh. The left and right green lines delimit the range at which the coherence values are 95% significant (0·94 and 1·76%) for the CFR. The middle green line indicates the CFR (1·14%) for which the highest pooled coherence was estimated. The red line indicates the upper 95% significance limit for the coherence estimate.

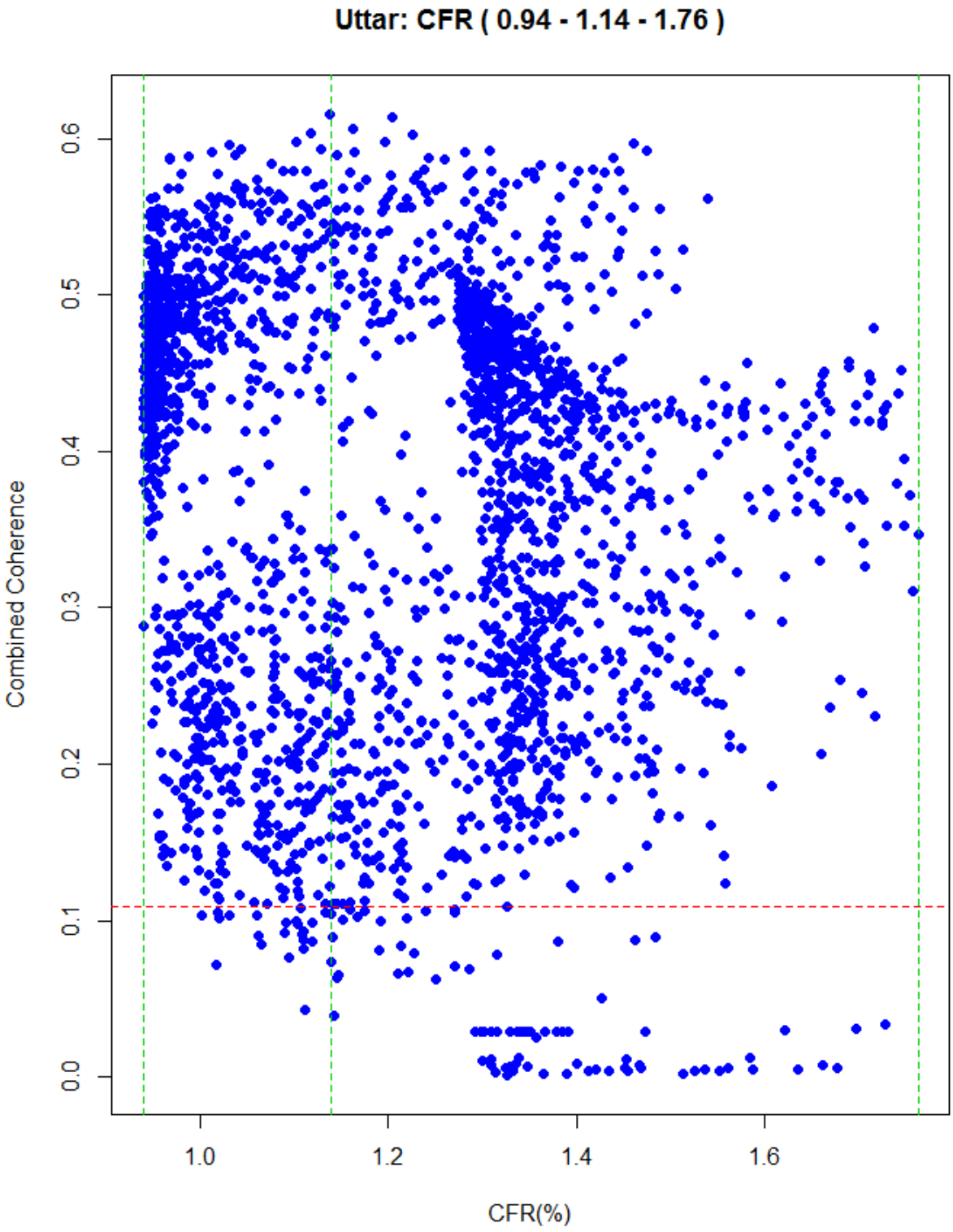
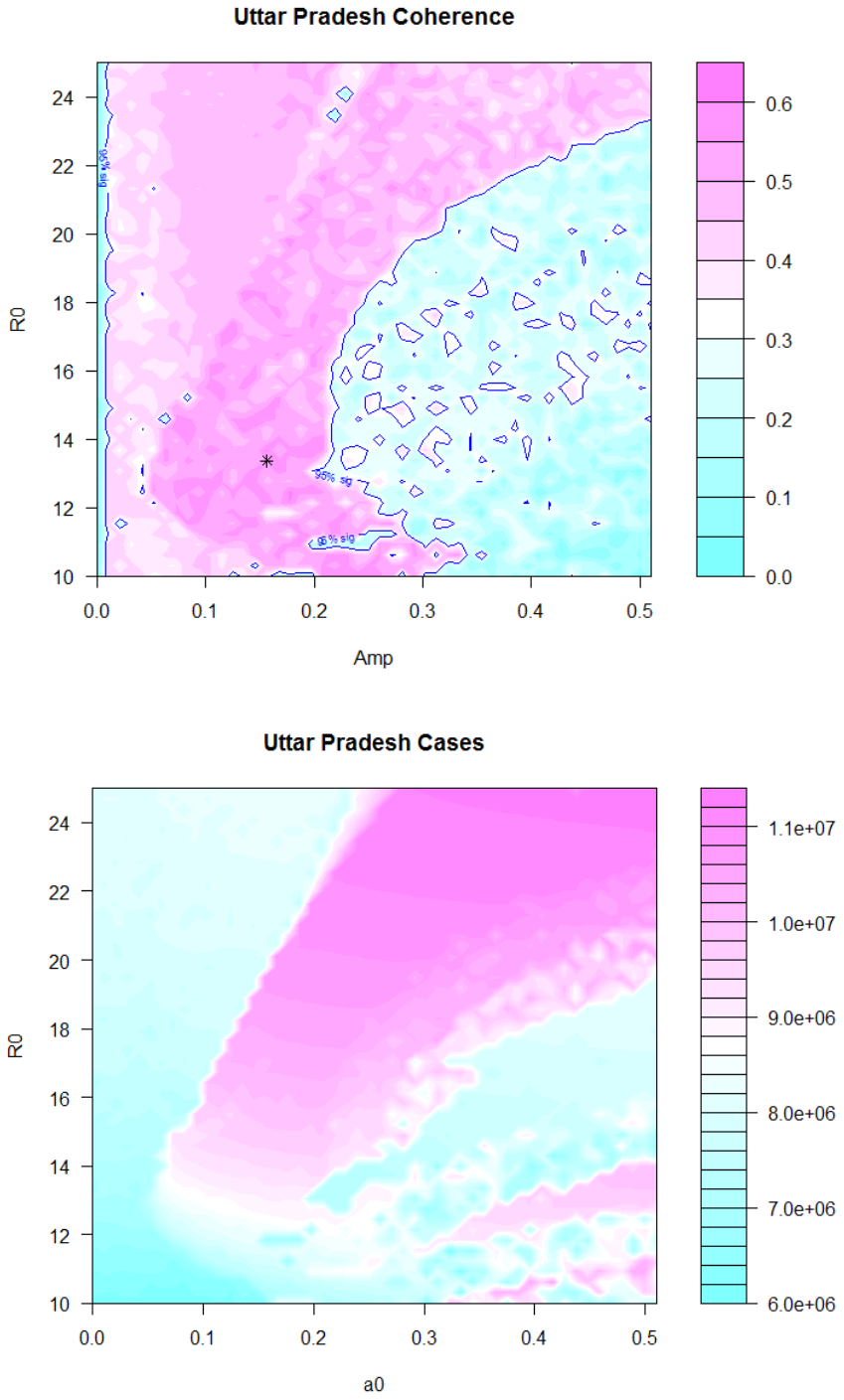
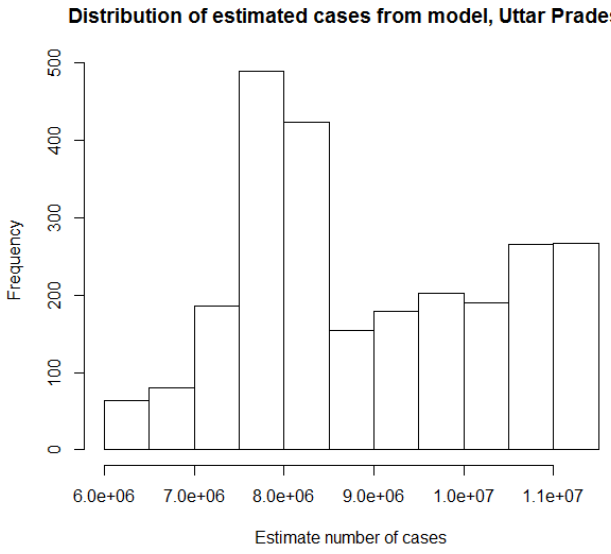
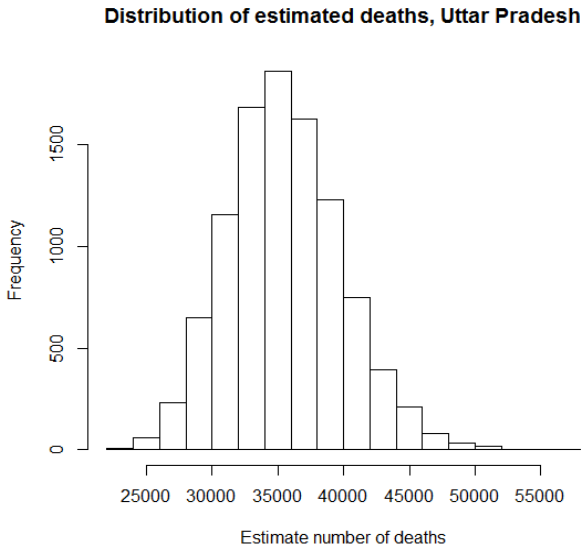


Figure S10. Pooled coherence estimate across five periods (0·5, 0·6, 0·75, 1·0, 1·5 year) and total cases over three-year period for UP plotted against estimated basic reproduction number (R_0) and amplitude of the forcing term (a_0). The asterisk (*) denotes the location of the maximum combined coherence. Distributions of estimated measles cases and deaths [1] and the resultant estimated case fatality risk (CFR) (1·19 CI: 0·80-1·76).




Distribution of estimated cases from model, Uttar Pradesh
Frequency
0
100
200
300
400
500
6.0e+06
7.0e+06
8.0e+06
9.0e+06
1.0e+07
1.1e+07
Estimate number of cases


Distribution of estimated deaths, Uttar Pradesh
Frequency
0
500
1000
1500
25000
30000
35000
40000
45000
50000
55000
Estimate number of deaths

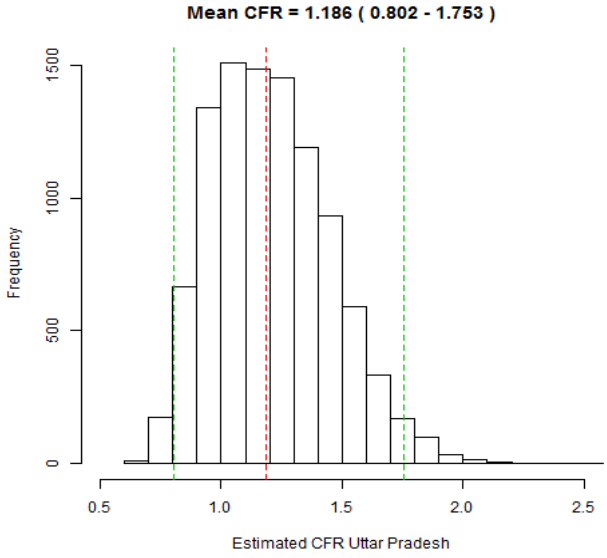

Mean CFR = 1.186 (0.802 - 1.753)
Frequency
0
500
1000
1500
0.5
1.0
1.5
2.0
2.5
Estimated CFR Uttar Pradesh

Figure S11. One of the 300 three-year sample time series of estimated deaths with added noise from either U(0,5000) (a) or U(0,20000) (b), respectively. Red line represents the data after Baxter-King filter.

(a) (b)

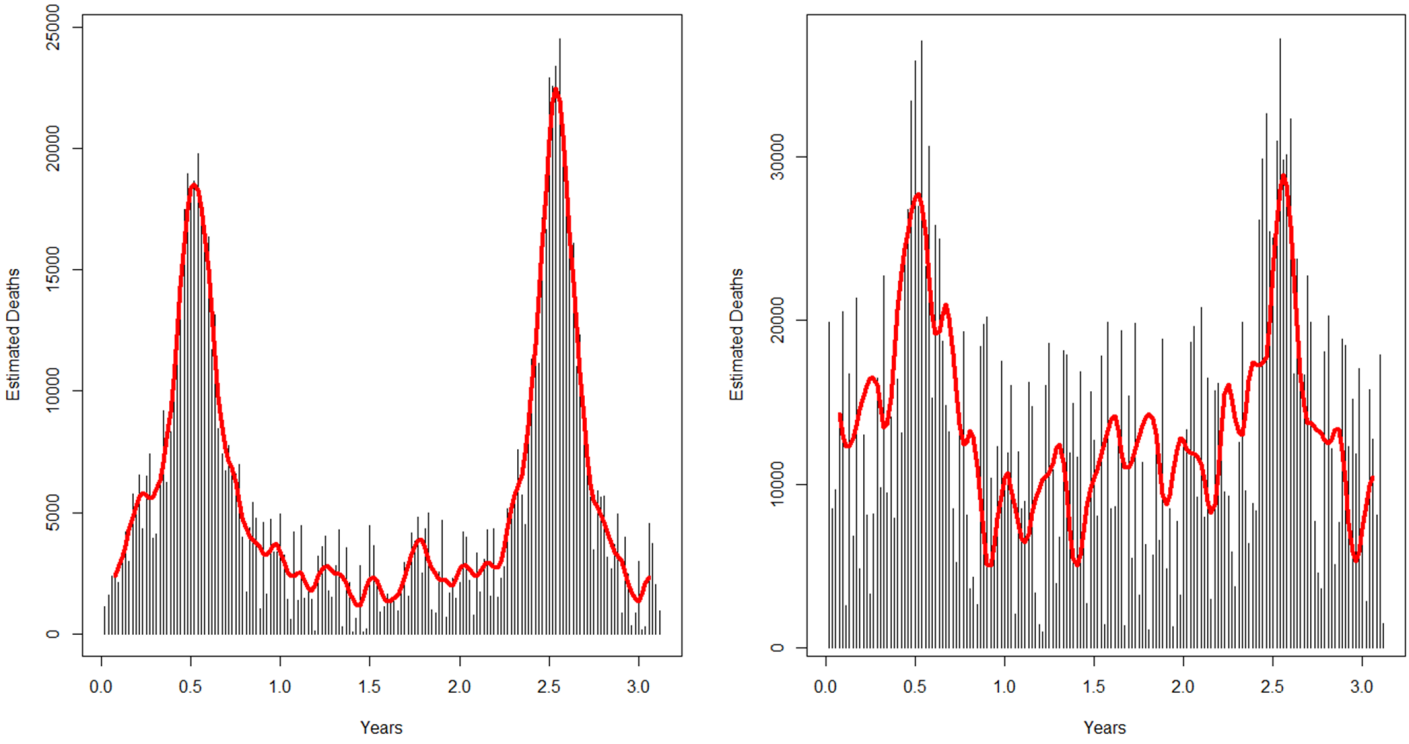
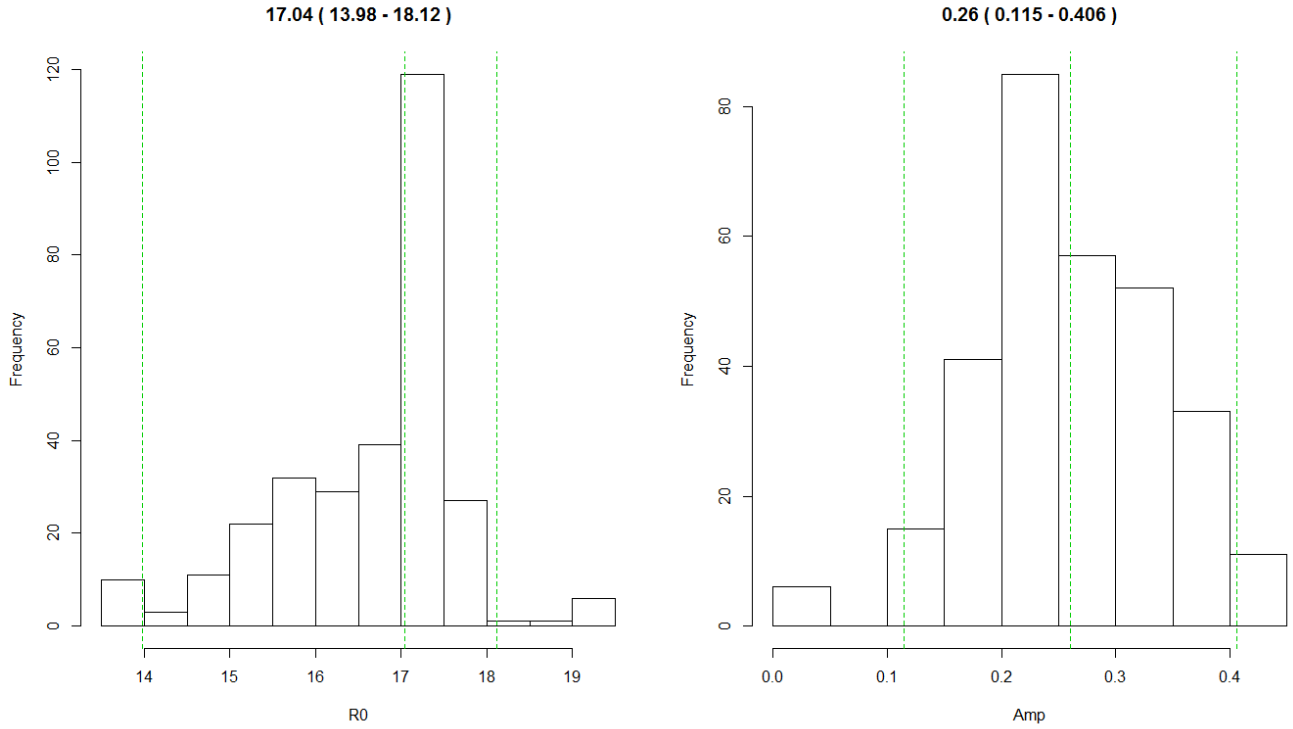


Figure S12. The distribution of estimates of R_0 and a_0 (Amp), in left and right columns respectively, with added noise from either U(0,5000) (a) or U(0,20000) (b), respectively.

(a)



(b)

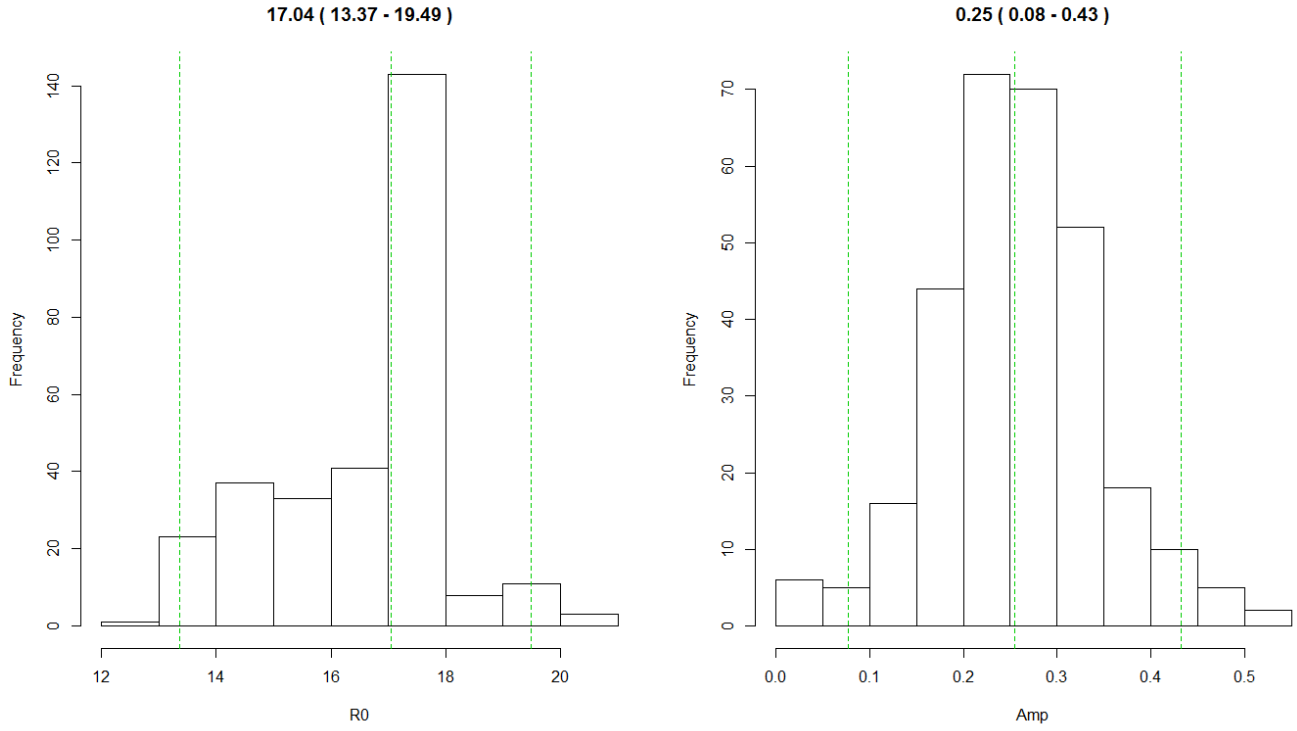
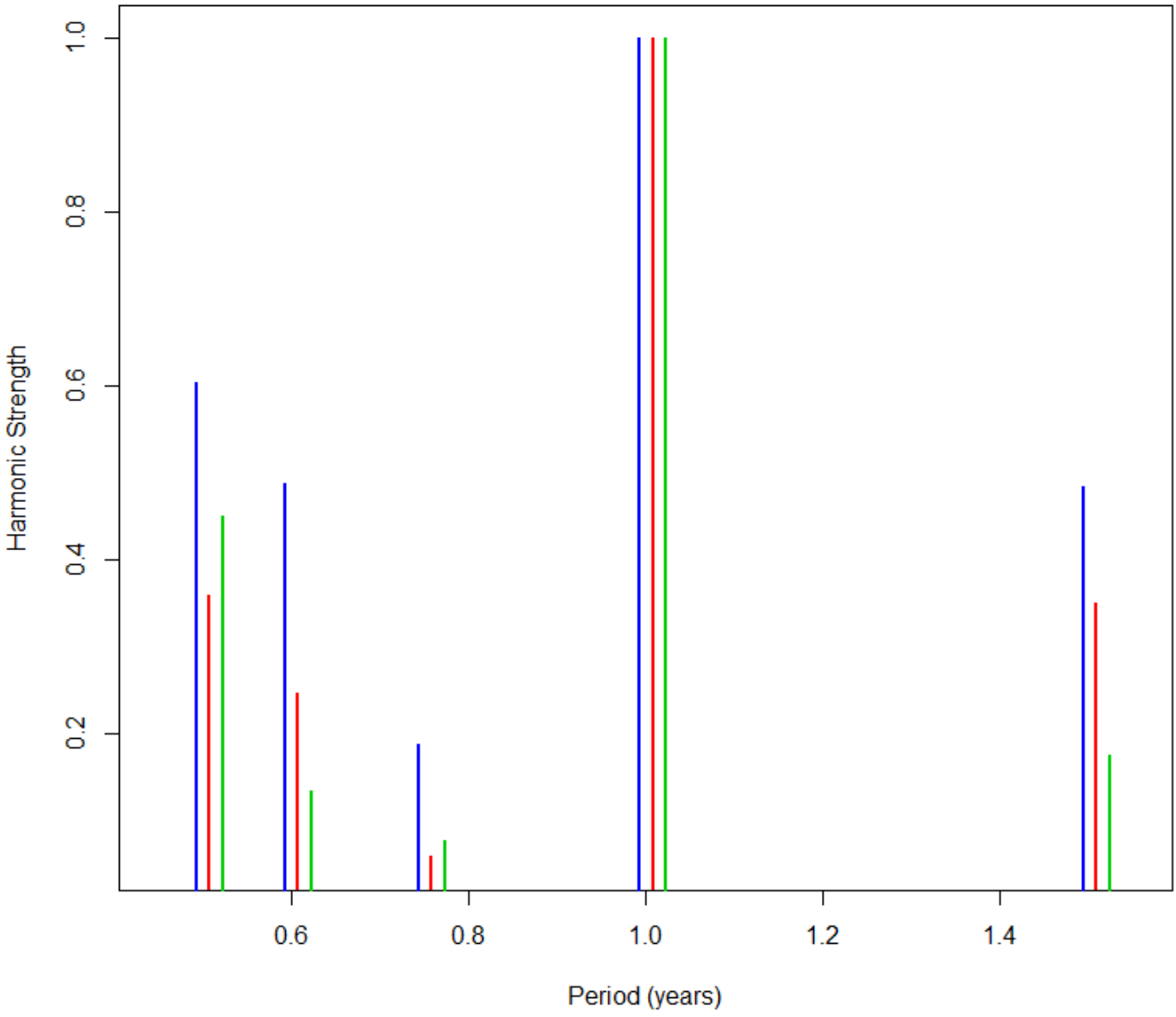


Figure S13. Frequency spectra showing longest five periods used in the analysis (from 6 months to 18 months), with normalized harmonic strength for comparison, of: the Million Death Study data (green), U(0,5000) (red) and U(0,20000) (blue).

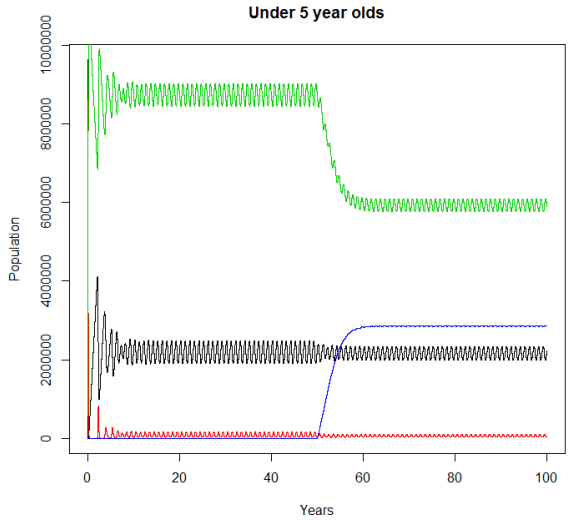


3. Dynamics of the model, pre- and post-vaccine equilibria

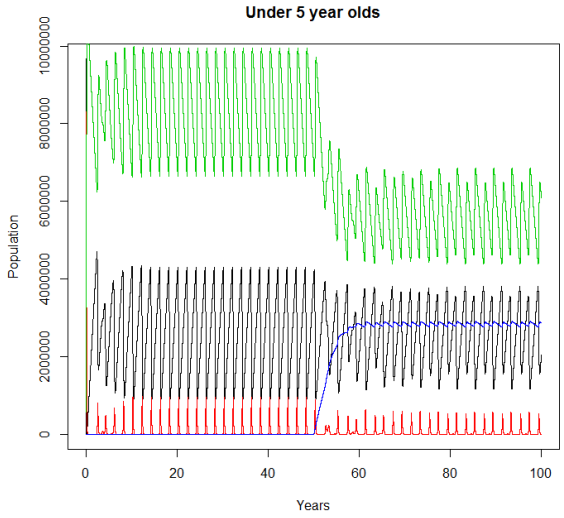
Figure S14 displays the dynamics of the model in India, Bihar, and Uttar Pradesh including the model outputs evolution to reach pre- and post-vaccine equilibria.

Figure S14. Time series of model outputs to reach the pre- and post-vaccine equilibrium, demonstrating the dynamics of the model in India, Bihar and Uttar Pradesh (UP) under two (R_0; a_0) combinations with high and low coherence (Coh) values respectively.
A, B and C plots are Bihar, India and UP respectively; 1 and 2 are high and low coherence respectively (Coh > 0.6, or Coh < 0.2). The black line represents the "Susceptibles"; the red line represents the "Infected"; the green line represents the "Recovered;" and the blue line represents the "Vaccinated". Vaccination was introduced at Year 50.

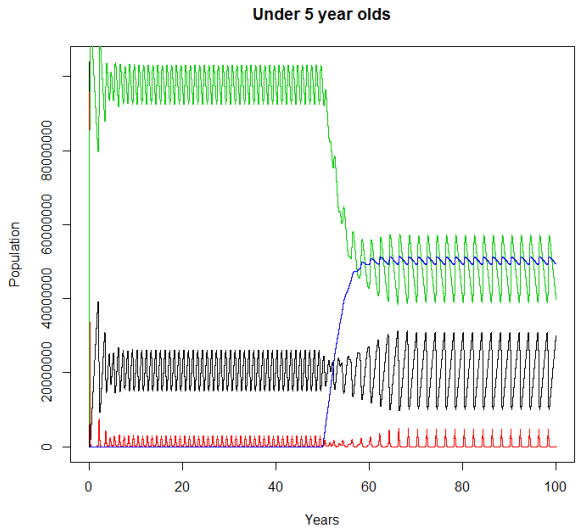
A1) Bihar R_0 =20; a_0=0.1; Coh=0.62



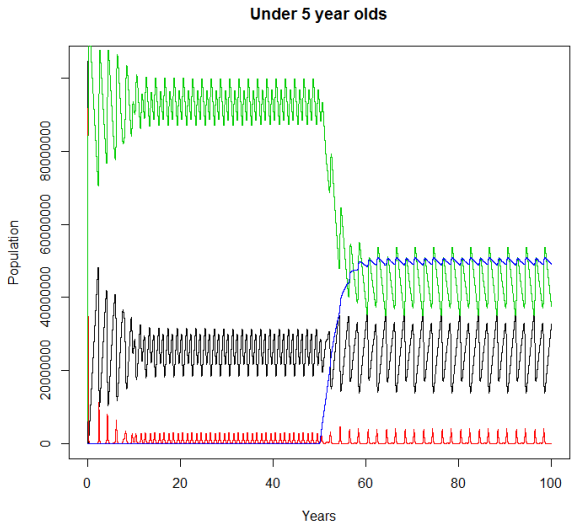
A2) Bihar R_0 =17; a_0=0.3; Coh=0.2



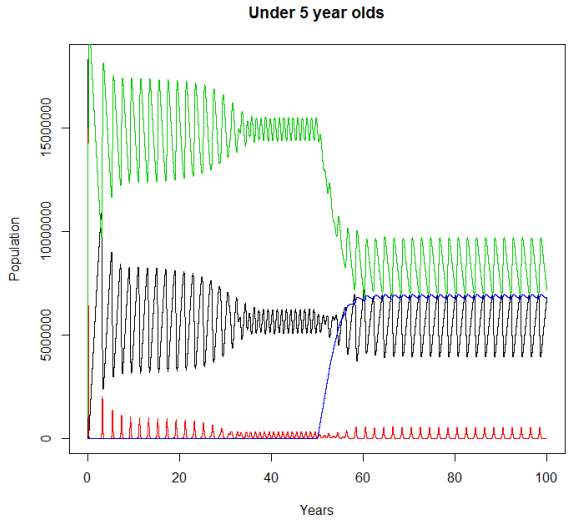
B1) India R_0 =24; a_0=0.15; Coh=0.64



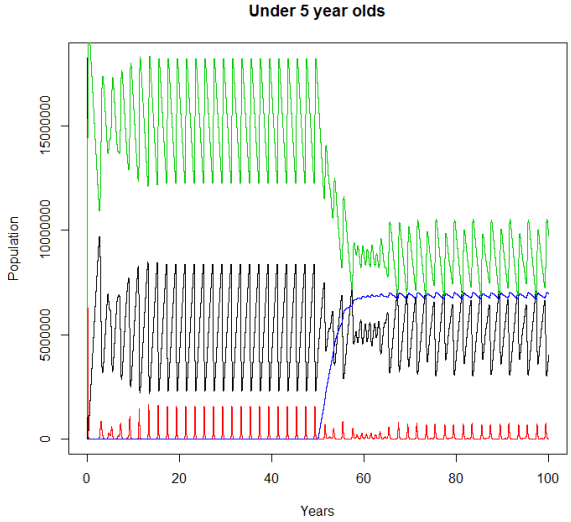
B2) India R_0 =17; a_0=0.25; Coh=0.15



C1) UP R_0 =13; a_0=0.15; Coh=0.6



C2) UPR_0 =16; a_0=0.35; Coh=0.2



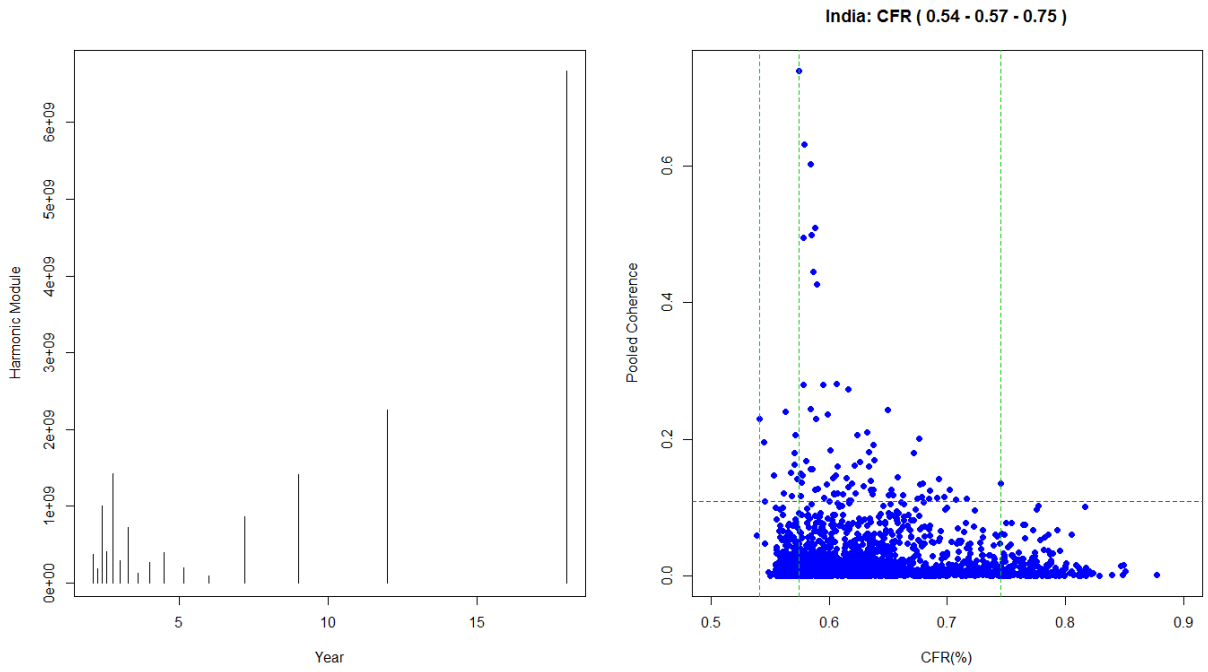
4. Use of World Health Organization measles reported cases

The World Health Organization (WHO) provides a measles notification time series with annual numbers of reported cases over 1980-2015 (36 years of data) for India as a whole. Conducting the same analysis (starting with the Fourier analysis on the WHO data), calculating the coherence between the estimated spectral density (Figure S15.A) and those produced by the model produced a CFR estimate of 0.57 (95% CI: 0.54-0.84) (Figure S15.B) for India, which is a lower prediction but overlaps with the CFR from the three-year Million Death Study data. For this analysis we used all periods, as even though the smallest harmonics of the Fourier analysis were seen for periods below 5 years, these were the periods of previously seen measles cycles. Though the data stretches over 36 years, there are far fewer data points in this time series (since only one data point per year is available) and this limits the possible periods identifiable in the Fourier analysis, the minimum period being 2 years (twice the sampling frequency) from an annual dataset, and the largest period being 18 years. With fewer data we can note there are fewer coherent matches and many are below being significant (Figure S15). Therefore for this

procedure to work or at least to identify periods shorter than biennial, we need time series data that is preferrably on a weekly or monthly time scale.

Figure S15. (A) Spectral density of WHO measles reported cases data for India (1980-2015), showing 16 periods from 2 years to 18 years. (B) Pooled coherence against calculated case fatality risk (CFR), demonstrating only a fraction of R_0 and a_0 combinations provided significant coherence estimates. The red line indicates the upper 95% significance limit for the coherence estimate.

(A) (B)



References

1. Morris SK, Awasthi S, Kumar R, Shet A, Khera A, et al. Measles mortality in high and low burden districts of India: estimates from a nationally representative study of over 12,000 deaths. Vaccine 2013; 31(41):4655-4661.
2. Goodman NR. On the joint estimation of the spectra, co-spectrum, and quadrature spectrum of a two-dimensional stationary Gaussian process. Scientific Paper No. 10 - Engineering Statistical Laboratory. New York: College of Engineering, New York University, 1957.
3. Thompson R. Coherence Significance Levels. Journal of the Atmospheric Sciences 1979; 36: 2020-2021.
4. Amjad AM, Halliday DM, Rosenberg JR, Conway BA. An extended difference of coherence test for comparing and combining several independent coherence estimates: theory and application to the study of motor units and physiological tremor. Journal of Neuroscience Methods 1997; 73(1):69-79.
5. World Health Organization. Measles surveillance data. Available from: http://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en/ (accessed December 29, 2015).