Supporting information

Appendix A

Model Selection

Starting with a basic, endemic-only model (including a population offset and linear trend in time), potential extensions of the three core components were added in turn and measures of fit and predictive power were calculated. The addition which yielded the best improvement in the RPS of OSA predictions, subject to calibration (p not less than 0.1 for test of calibration based on RPS), was selected and then all remaining options tested again. This process was repeated until no further extension of the model made a significant (p <0.001) improvement to predictive power (as determined by a permutation test on the RPS). This stringent criterium was employed in order to prioritise simplicity over complexity. If at any point an individual model parameter lost significance, the element associated with this parameter was removed in subsequent models.

Empirical Coverage Probabilities

Again using a one-step-ahead approach, the 25th and 75th quantiles of the predicted distribution were calculated and a score of 0 or 1 assigned if the observed value fell inside or outside this quantile range respectively. This binary score was assigned for each block and each month in the test set, such that we could subsequently calculate a proportion of prediction intervals which did not capture the true count. Thus, the overall score, C, is given by

$$C = \frac{1}{n_i n_t} \sum_{i,t} \mathbb{1}[y_{it} \leq q_{i,t,0.25} | y_{it} \geq q_{i,t,0.75}] \tag{1}$$

where y_{it} is the observed count for block i at month t, n_i and n_t the total number of blocks and months respectively, and $q_{i,t,p}$ the p^{th} quantile of the predicted distribution. We also investigated such a score using 10th and 90th quantiles, to

ascertain whether these could be used as approximate lower and upper bounds for case counts.

Appendix B

Preliminary analyses

Dispersion

District-specific dispersion parameters were investigated, but ultimately not considered a viable option to be included in the model. Four districts in particular (Aurangabad, Banka, Jehanabad and Nawada) demonstrate extended periods of zero incidence with occasional sporadic cases or large spikes, which lead to very large dispersion estimates for these districts and therefore unrealistically high predictions. See Fig. S1 for an illustration of these patterns. Due to the neighbourhood effect, these high predictions in turn influence the predictions of any bordering blocks. Changes in detection effort could go some way to explaining these unusual patterns, however it is also likely that such patterns will become more common as elimination is approached. This suggests that an alternative modelling strategy will become necessary as cases become more sparse in space and time.

Distributed temporal lags

By sequentially adding further distributed lags to the best-fitting single-lagged model, neither a clear minimum nor an "elbow" in RPS was attained up to twelve months. The weights assigned to each lag did not show a rapid "drop-off" as a result of a high estimated decay parameter, and months substantially far back in time were still assigned non-negligible weight. PIT histograms of predictions from these lagged models are included in . We found that adding higher orders of distributed lags consistently improved both predictive power and fit. This appears to contradict analysis of individual block time series which suggested significant auto-correlation no more than four months back in time. In the current form of "hhh4addon", it is not possible to specify a different temporal lag length within the AR and NE components (for example, to incorporate neighbouring incidence from further back in time than within-block). Therefore, the contribution of distributed lags to both components had

to be considered and a balance had to be drawn. Comparing the PIT histograms of solely auto-regressive models, the very highest counts are vastly underestimated for all lag lengths. Since the highest values in each block often reflect sudden jumps they cannot be captured by auto-regression; more information - potentially from the surrounding area - is required to anticipate them. Models with no auto-regression but which incorporate neighbouring incidence are better able to reach the highest counts but in doing so over-estimate the moderate-to-high range. It was concluded that beyond four months of lags the improvement in prediction was small enough to discount, and much longer lags were difficult to justify epidemiologically. Therefore only four months of lags were considered for the final model.

Supplementary Figures

Fig. S1 Districts with unusual incidence patterns resulting in inflated dispersion estimates.

Fig. S2 Probability integral transform (PIT) histograms for models with increasing orders of geometric lags from 1 to 12 months (left to right, top to bottom) in the autoregressive component. The final model selection process considered up to four lags.

Table S1 Fit and prediction metrics for selected model at each stage. The reported AIC is for the fit to training data only, and RPS is of predictions made without updating this fit (i.e. fixed instead of rolling). C2575 and C1090 refer to the coverage of 50% and 80% quantile intervals, respectively, alongside the average interval width in cases. Model no. 42 is the final model.

Stage	Model No.	END	AR	NE	Dispersion	No. parameters	AIC	RPS	Calibration (p-value)	C1090	Avg. width
0	1	offset + 1 + t			1	3	65412	0.657	<0.0001	0.095	2.243
1	2	offset + 1 + t + seas(∼1, S=1)			1	5	65227	0.654	<0.0001	0.090	2.330
1	3	offset + 1			1	2	65811	0.698	<0.0001	0.044	4.158
1	4	offset + 1 + t + logpopdens			1	4	65708	0.662	<0.0001	0.094	2.180
1	5	offset + 1 + t	AR(1)		1	4	57100	0.495	0.109	0.060	2.386
1	**6**	**offset + 1 + t**	**AR(1) + seas(∼1, S=1)**		**1**	**6**	**57058**	**0.493**	**0.115**	**0.058**	**2.388**
1	7	offset + 1 + t	AR(1) + seas(∼1 + t, S=1)		1	7	57031	0.496	<0.0001	0.064	2.171
1	8	offset + 1 + t		NE(2)	1	5	56755	0.516	0.003	0.056	2.304
1	9	offset + 1 + t		NE(2) + logpopdens	1	5	56763	0.516	0.002	0.056	2.313
1	10	offset + 1 + t		NE(2) + seas(∼1, S = 1)	1	7	56685	0.515	0.001	0.054	2.308
1	11	offset + 1 + t		NE(2) + seas(∼1 + t, S = 1)	1	8	56680	0.516	0.203	0.057	2.201
1	12	offset + 1 + t			State	4	65310	0.659	<0.0001	0.098	2.145
2	13	offset + 1 + seas(∼1, S=1)	AR(1) + seas(∼1, S=1)		1	7	57024	0.502	<0.0001	0.048	2.627
2	14	offset + 1	AR(1) + seas(∼1, S=1)		1	5	57101	0.502	<0.0001	0.049	2.612
2	15	offset + 1 + t + logpopdens	AR(1) + seas(∼1, S=1)		1	6	57128	0.499	<0.0001	0.055	2.496
2	16	offset + 1 + t	AR(1) + seas(∼1 + t, S=1)		1	7	57031	0.496	<0.0001	0.064	2.171
2	17	offset + 1 + t	AR(1) + seas(∼1 + t, S=2)		1	9	56996	0.496	<0.0001	0.064	2.176
2	18	offset + 1 + t	AR(1) + seas(∼1, S=1)	NE(2)	1	8	53362	0.458	0.210	0.055	2.105
2	19	offset + 1 + t	AR(1) + seas(∼1, S=1)	NE(2) + seas(∼1, S = 1)	1	10	53300	0.457	0.294	0.053	2.101
2	20	offset + 1 + t	AR(1) + seas(∼1, S=1)	NE(2) + seas(∼1 + t, S = 1)	1	11	53301	0.458	0.125	0.053	2.122
2	21	offset + 1 + t	AR(1) + seas(∼1, S=1)	NE(2) + logpopdens	1	8	53398	0.458	0.144	0.054	2.111
2	22	offset + 1 + t	AR(1) + seas(∼1, S=1)		State	7	57059	0.493	0.123	0.058	2.389
2	**23**	**offset + 1 + t**	**AR(2) + seas(∼1, S=1)**		**1**	**6**	**53833**	**0.455**	**0.189**	**0.053**	**2.230**
2	24	offset + 1 + t	AR(3) + seas(∼1, S=1)		1	6	52279	0.439	0.005	0.061	2.017
2	25	offset + 1 + t	AR(4) + seas(∼1, S=1)		1	6	51342	0.428	<0.0001	0.064	1.877
3	26	offset + 1 + seas(∼1, S=1)	AR(2) + seas(∼1, S=1)		1	7	53806	0.457	<0.0001	0.043	2.395
3	27	offset + 1	AR(2) + seas(∼1, S=1)		1	5	53844	0.458	<0.0001	0.047	2.340
3	28	offset + 1 + t + logpopdens	AR(2) + seas(∼1, S=1)		1	6	53835	0.456	<0.0001	0.042	2.404
3	29	offset + 1 + t	AR(2) + seas(∼1 + t, S=1)		1	7	53815	0.455	0.002	0.056	2.087
3	30	offset + 1 + t	AR(2) + seas(∼1 + t, S=2)		1	9	53692	0.455	0.001	0.057	2.079

Stage	Model No.	END	AR	NE	Dispersion	No. parameters	AIC	RPS	Calibration (p-value)	C1090	Avg. width
3	31	offset + 1 + t	AR(3) + seas(∼1, S=1)		1	6	52279	0.439	0.005	0.061	2.017
3	32	offset + 1 + t	AR(2) + seas(∼1, S=1)	NE(1)	1	7	51749	0.437	0.181	0.054	1.974
3	**33**	**offset + 1 + t**	**AR(2) + seas(∼1, S=1)**	**NE(1) + seas(∼1, S = 1)**	**1**	**9**	**51675**	**0.437**	**0.122**	**0.055**	**1.966**
3	34	offset + 1 + t	AR(2) + seas(∼1, S=1)	NE(3) + seas(∼1 + t, S = 1)	1	11	51543	0.437	0.656	0.050	2.029
3	35	offset + 1 + t	AR(2) + seas(∼1, S=1)		State	7	53831	0.455	0.192	0.053	2.230
4	36	offset + 1 + seas(∼1, S=1)	AR(2) + seas(∼1, S=1)	NE(1) + seas(∼1, S = 1)	1	10	51701	0.437	0.085	0.056	1.961
4	37	offset + 1	AR(2) + seas(∼1, S=1)	NE(1) + seas(∼1, S = 1)	1	8	51673	0.437	0.194	0.055	1.969
4	38	offset + 1 + t + logpopdens	AR(2) + seas(∼1, S=1)	NE(1) + seas(∼1, S = 1)	1	9	51691	0.437	0.153	0.056	1.962
4	39	offset + 1 + t	AR(2) + t	NE(1) + seas(∼1, S = 1)	1	8	51670	0.439	0.001	0.059	1.865
4	40	offset + 1 + t	AR(2) + seas(∼1, S=2)	NE(1) + seas(∼1, S = 1)	1	11	51545	0.437	0.115	0.055	1.973
4	41	offset + 1 + t	AR(2) + seas(∼1 + t, S=2)	NE(1) + seas(∼1 + t, S = 1)	1	15	51446	0.441	0.563	0.054	1.959
4	**42**	**offset + 1**	**AR(4) + seas(∼1, S=1)**	**NE(1) + seas(∼1, S = 1)**	**1**	**8**	**50323**	**0.420**	**0.346**	**0.054**	**1.872**
4	43	offset + 1 + t	AR(2) + seas(∼1, S=1)	NE(1) + t	1	8	51749	0.437	0.545	0.053	2.003
4	44	offset + 1 + t	AR(2) + seas(∼1, S=1)	NE(1) + seas(∼logpopdens, S = 1)	1	9	51780	0.438	0.202	0.056	1.975
4	45	offset + 1 + t	AR(2) + seas(∼1, S=1)	NE(3)	1	8	51642	0.437	0.333	0.053	1.972
4	46	offset + 1 + t	AR(2) + seas(∼1, S=1)	NE(1) + seas(∼1, S = 1)	State	10	51676	0.437	0.118	0.055	1.964
4	47	offset + 1 + t	AR(2) + seas(∼1, S=1)	NE(1) + seas(∼logpopdens + t, S = 1)	State	11	51782	0.438	0.314	0.055	1.988
5	48	offset + 1 + seas(∼1, S=1)	AR(4) + seas(∼1, S=1)	NE(1) + seas(∼1, S = 1)	1	10	50342	0.420	0.297	0.055	1.867
5	49	offset + 1	AR(4) + seas(∼1 + t, S=1)	NE(1) + seas(∼1 + t, S = 1)	1	10	50296	0.424	0.614	0.052	1.864
5	50	offset + 1 + logpopdens	AR(4) + seas(∼1, S=1)	NE(1) + seas(∼1, S = 1)	1	9	50332	0.420	0.439	0.054	1.870
5	51	offset + 1	AR(4) + t	NE(1) + seas(∼1, S = 1)	1	7	50336	0.424	0.000	0.060	1.763
5	**52**	**offset + 1**	**AR(4) + seas(∼1, S=2)**	**NE(1) + seas(∼1)**	**1**	**10**	**50164**	**0.419**	**0.194**	**0.055**	**1.868**
5	53	offset + 1	AR(4) + seas(∼1 + t, S=1)	NE(1) + seas(∼1 + t, S = 2)	1	14	50097	0.423	0.782	0.052	1.851
5	54	offset + 1	AR(4) + seas(∼1, S=1)	NE(1) + seas(∼1 + t, S = 1)	1	9	50324	0.420	0.620	0.052	1.904
5	55	offset + 1	AR(4) + seas(∼1, S=1)	NE(1) + seas(∼logpopdens, S = 1)	1	8	50401	0.421	0.425	0.055	1.873
5	56	offset + 1	AR(4) + seas(∼1, S=1)	NE(1)	1	6	50416	0.420	0.251	0.054	1.877
5	57	offset + 1	AR(4) + seas(∼1, S=1)	NE(1) + seas(∼1, S = 1)	State	9	50325	0.420	0.342	0.054	1.873
5	58	offset + 1	AR(4) + seas(∼1, S=1)	NE(1) + seas(∼logpopdens + t, S = 1)	State	10	50405	0.421	0.537	0.055	1.876

Fig. S3 PIT histograms for the selected model at each stage. Model 42 is the final model. Model 52 offered minor improvement in RPS with additional complexity.

Fig. S4 Fitted seasonal waves in the auto-regressive (AR) and neighbourhood (NE) model components. Both reflect the first-quarter peak in reported cases but the magnitude of the waves differs, with the contribution of the AR component varying more than that of the NE.

Fig. S5 Blocks with average RPS greater than 2.5 over the test period (Jan 2017 - Dec 2018)