

Supplementary materials for

Respiratory virus transmission dynamics determines timing of asthma exacerbation peaks: evidence from a population-level model

RM Eggo, JG Scott, AP Galvani, L Ancel Meyers

correspondence to: r.eggo@lshtm.ac.uk

| | |
|--|----|
| Supplementary Materials and Methods | 1 |
| 1. Population size changes | 1 |
| 2. Air quality data | 3 |
| 3. Temperature data | 3 |
| 4. School calendar entered | 5 |
| 5. Hospitalization Data..... | 5 |
| 6. Extended information on the SIRS model | 7 |
| 7. Statistical Model | 8 |
| 8. MCMC Sampling..... | 9 |
| 9. MCMC Convergence..... | 11 |
| 10. Model Comparison..... | 11 |
| 11. Comparison without colds in the model | 12 |
| 12. With 0/1 holiday, linear model | 14 |
| 13. Best fitting model city-level results | 15 |
| 14. Daily residuals for each city | 18 |
| 15. Monthly residuals for each city..... | 21 |
| 16. Calculation of the average number of colds per person..... | 22 |
| 17. Change in school start date during study period..... | 23 |
| 18. Model fits by day of the week..... | 24 |
| 19. References..... | 27 |

Supplementary Materials and Methods

1. Population size changes

Population sizes for children aged 5–18, and adults 19–55 in each metropolitan area in 2000 and 2010 are taken from the 2000 and 2010 Census. The daily change in population is then calculated, and population estimates for each day in the study are interpolated. This maintains an approximately correct size and adult:child ratio throughout the study period. Since long-term immune protection does not occur in the SIRS model, there is no need for demographic turnover. Adjustments to population sizes are made in the susceptible class.

Figure S1 shows the population size changes for children 5–18 and adults 19–55 over the 7 years of the study period in the model for each metropolitan area. The metropolitan areas have a total population range of 385,000–5,282,000 (in 2002). 7 of the metropolitan areas grew in population over the course of the study period, and 1 decreased. Of note, the absolute total (including 0–5 year olds, and 55+) is larger than the value given here.

Figure S2 shows the proportion of 5–18 year olds in the modelled population, and changes through time. Some metropolitan areas have increasing proportion of children, others decreasing, with variation in proportion from 0.42 to 0.29. The proportion of children affects the dynamics of infections in the transmission model.

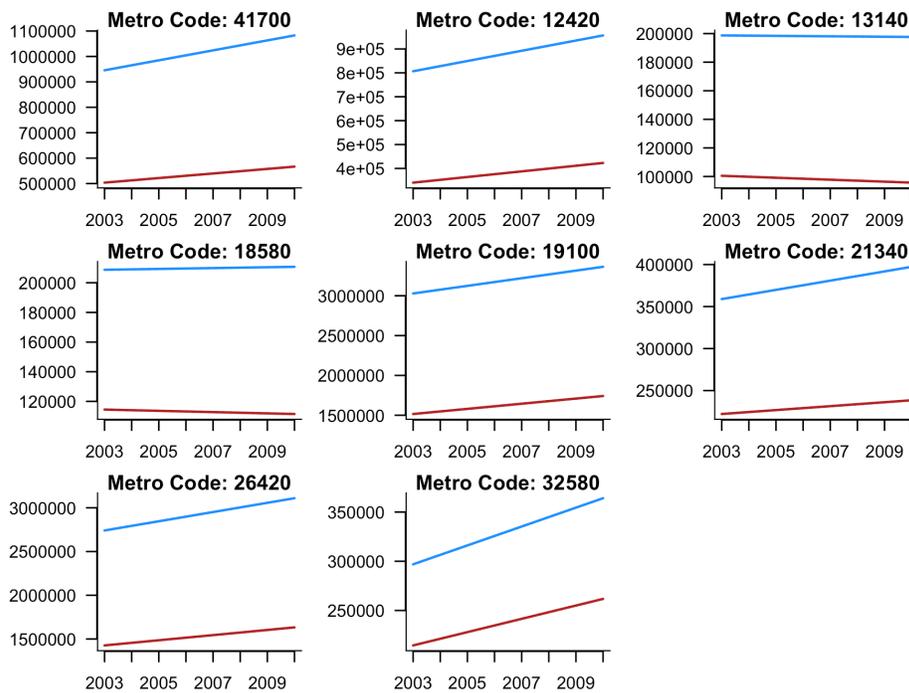


Fig. S1. Demographic change during study period. Number of adults (blue) aged 19-55, and children (red) aged 5-18 in each metropolitan area in the modelled population through the study period.

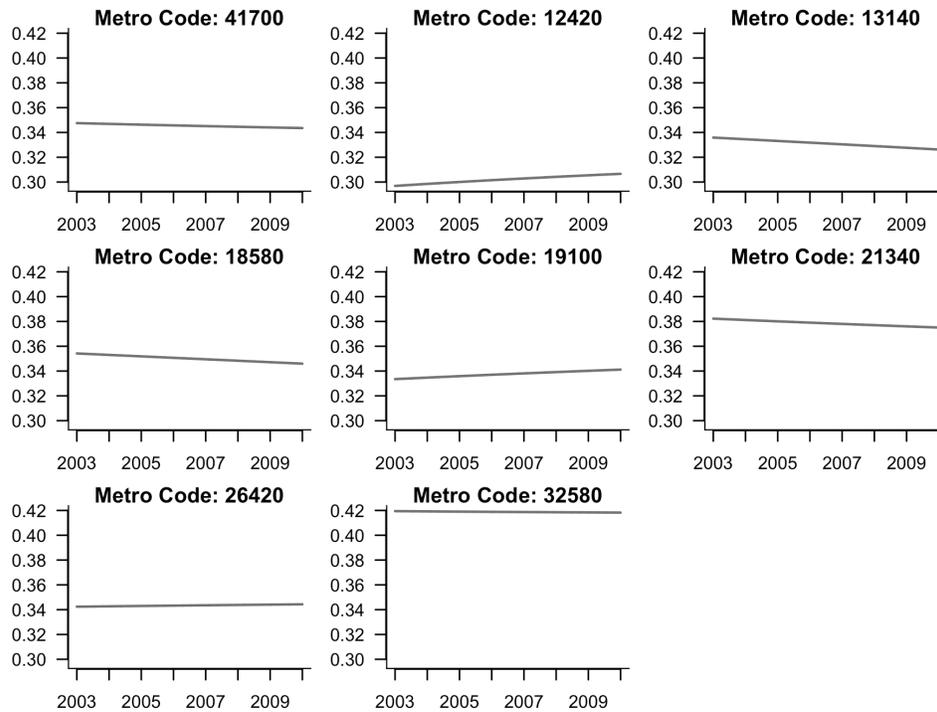


Fig. S2. Change in proportion of children.
Proportion of 5–18 year olds in the modelled population through time.

2. Air quality data

Particulate matter data are taken from the CDC Wonder database, which provides daily values for PM_{2.5} for each county from 2003–2011. We use the maximum value of PM_{2.5} ($\mu\text{g}/\text{m}^3$) from the counties that comprise each metropolitan area. The data were downloaded on September 24th 2013.

Ozone data are taken from the US Environmental Protection Agency (EPA) AIRS database, which provides a daily ozone measurement for each metropolitan area in Texas, for most days of the study period in Air Quality Index (AQI) values. There are 57 missing values in all 8 cities over the 7-year time period. Missing values are excluded. These data were downloaded on January 21st 2013.

3. Temperature data

Temperature data are taken from the CDC Wonder database, which provides air temperature values for each county from 1979–2011. The minimum temperature is the minimum daily value (in Celsius) from the counties that comprise the metropolitan area. The data were downloaded on September 24th 2013.

Minimum temperature is centered, so that the mean minimum temperature during the entire study period is 0 for each metropolitan area. Negative temperature values are therefore lower than the mean temperature, and thus a negative coefficient with a negative temperature increases the hospitalization rate. These data are shown in **Fig. S3**.

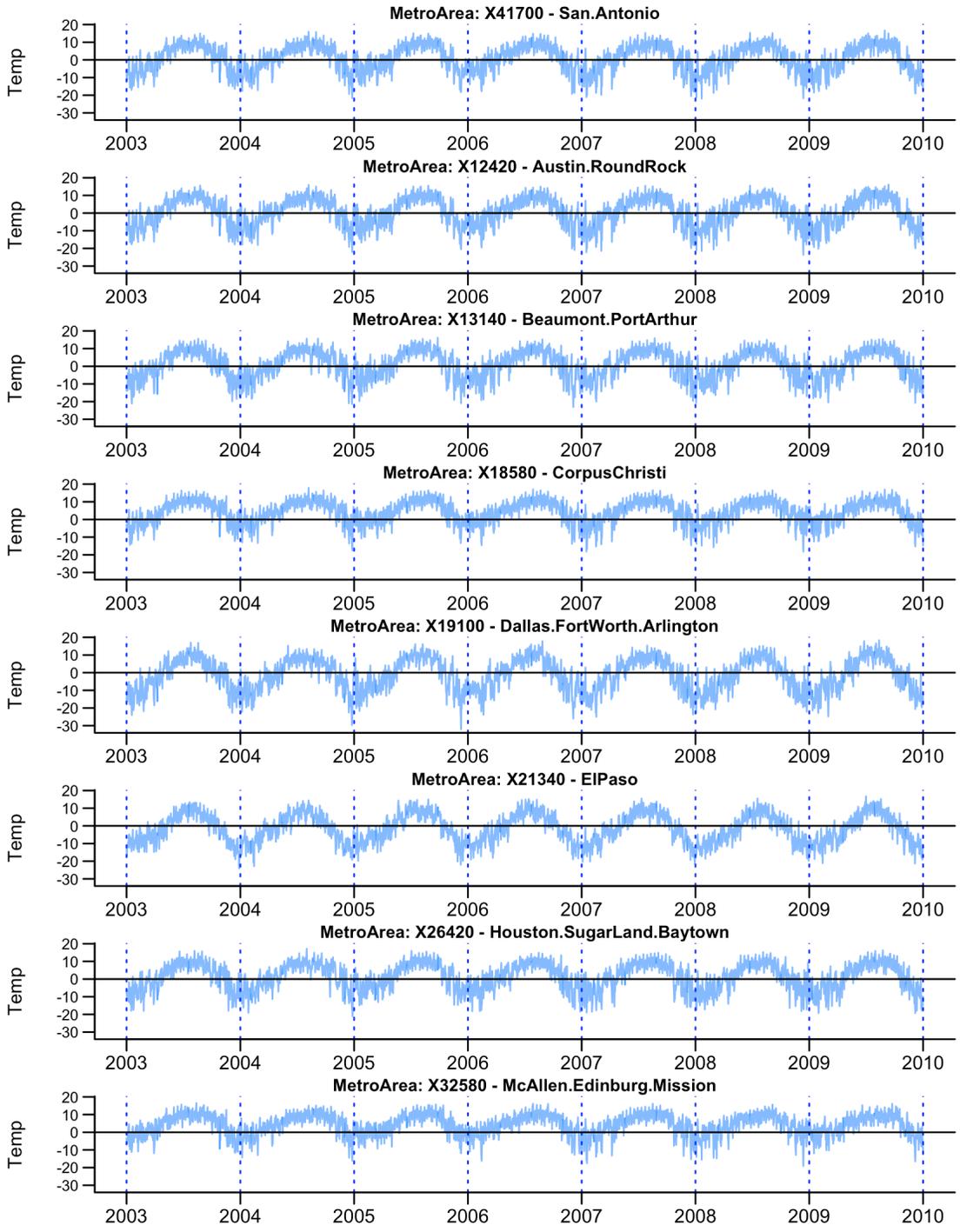


Fig. S3. Temperature during the study period.

Blue line marks the daily centred temperature on each day of the study period for each metropolitan area.

4. School calendar entered

There are over 1200 school districts in Texas, and separate fee-paying institutions may also set their own calendar. We used the school calendar for the largest independent School District (ISD) in each metropolitan area, excepting McAllen-Edinburg-Mission, where calendars for Edinburg ISD were not available. The second largest is a comparably sized district, Pharr-San Juan-Alamo ISD, and was used in its place. Table S1 shows for each metropolitan area, which school district calendar is used. There is one substitution, of one academic year in the calendar used for McAllen-Edinburg-Mission, where 2004-2005 is substituted with the calendar for McAllen ISD due to an error with the archived file received from Pharr-San Juan-Alamo ISD.

We gratefully acknowledge staff members at Austin ISD, Corpus Christi ISD, Dallas ISD, Houston ISD and Pharr-San Juan-Alamo ISD for providing historic school calendars.

| Metro Code | Metropolitan area name | ISD | # schools (2011) | Approx # children (2011) |
|------------|-----------------------------|----------------------|------------------|--------------------------|
| 12420 | Austin-Round Rock | Austin | 124 | 86697 |
| 13140 | Beaumont-Port Arthur | Beaumont | 40 | 19551 |
| 18580 | Corpus Christi | Corpus Christi | 62 | 38196 |
| 19100 | Dallas-Fort Worth-Arlington | Dallas | 236 | 157111 |
| 21340 | El Paso | El Paso | 106 | 63378 |
| 26420 | Houston-Sugar Land-Baytown | Houston | 309 | 202703 |
| 32580 | McAllen-Edinburg-Mission | Pharr-San Juan-Alamo | 44 | 31329 |
| 41700 | San Antonio | Northside | 110 | 92335 |

Table S1. School calendar information.

Summary information for locations and school calendars used in the study. The metropolitan reference code, name of the metropolitan area, Independent School District used for determining the school calendar in that metropolitan area, and the number of schools and approximate number of children in that ISD in 2011 are given.

5. Hospitalization Data

Hospitalization data for the 8 metropolitan regions by day of week are given in Fig S4 for children and Fig S5 for adults. These figures show the hospitalization rate due to asthma on each day of the study, stratified by day of the week, and the overlaid boxplots give the mean and interquartile range for the rate stratified by day of the week. It is important to note the large variation in observed hospitalization rates on every day of the week.

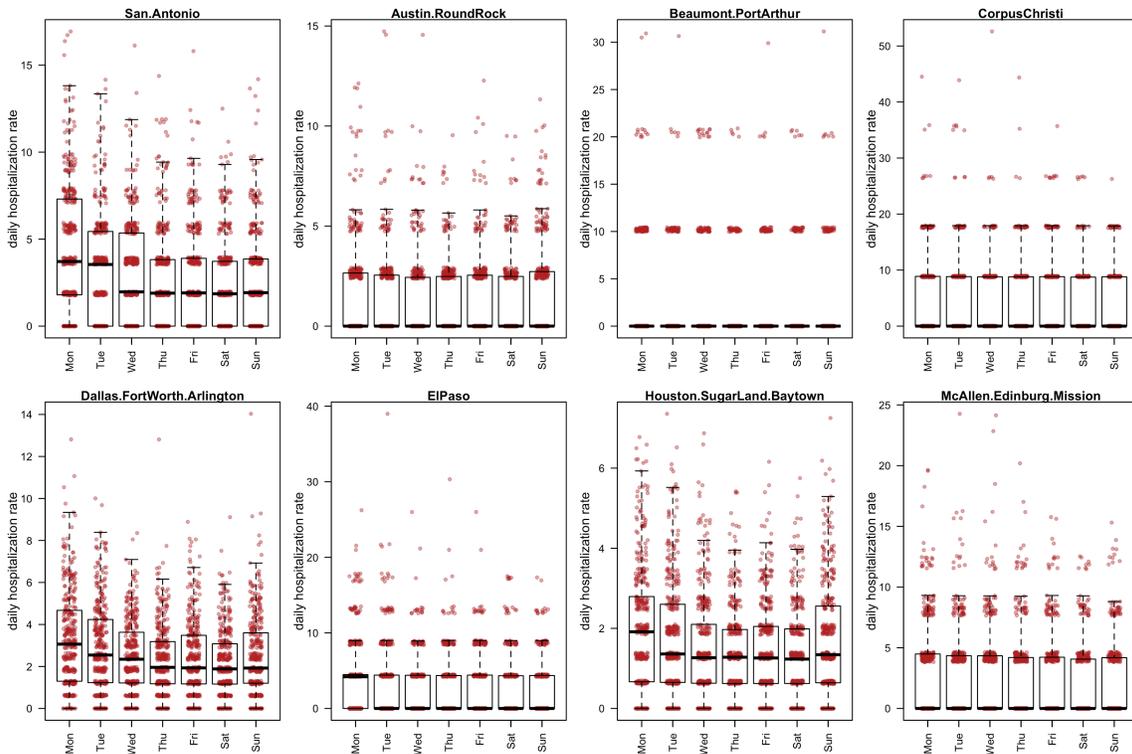


Fig S4. Daily hospitalization rate in children for each day of the study, stratified by day of the week.

Red points show the hospitalization rate for each day of the study, and boxplots show the mean and interquartile range of the data. In children there is a tendency for higher hospitalization rates at the start of the school week, but there is a large amount of variation in the data.

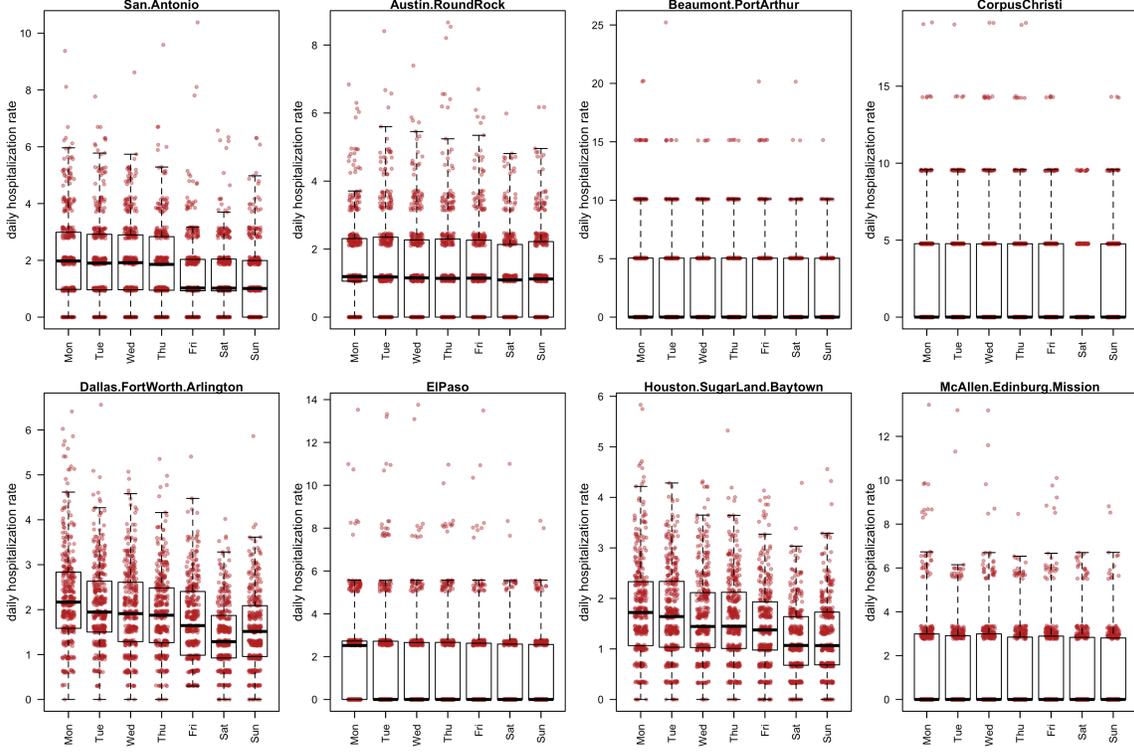


Fig S5. Daily hospitalization rate in adults for each day of the study, stratified by day of the week.

Red points show the hospitalization rate for each day of the study, and boxplots show the mean and interquartile range of the data. In adults there is a tendency for higher hospitalization rates at the start of the week, and lower on the weekends, but there is a large amount of variation in the data.

6. Extended information on the SIRS model

For each metropolitan area we use our age-stratified SIRS model to generate city-specific common cold prevalence:

$$\begin{aligned}\frac{dS_i}{dt} &= -\beta_{i,t}S_i + \omega R_i \\ \frac{dI_i}{dt} &= \beta_{i,t}S_i - \gamma I_i \\ \frac{dR_i}{dt} &= \gamma I_i - \omega R_i\end{aligned}$$

where, i = children or adults, and where,

$$\beta_{i,t} = \sum_{j=1}^2 \beta_0 \alpha_{ij} \sigma_{ij,t} \frac{I_{j,t}}{N_j}$$

and $\beta_{i,t}$ is the transmissibility of i , α_{ij} is the scaling of transmissibility between i and j , and, $I_{j,t}$ is the number of infected people in age group j at time t , N_j is the population of age group j , and $\sigma_{ij,t}$ is the effect of school vacation, as:

$$\sigma_{ij,t} = \begin{cases} 1, & \text{if } i \text{ or } j \text{ is an adult} \\ \text{or } \begin{cases} 1, & \text{if } t = \text{school in session} \\ \text{estimated, otherwise} \end{cases} \end{cases}$$

β_0 , α_{AA} , α_{AC} , $\sigma_{ij,t}$, γ and ω are estimated.

7. Statistical Model

We adopt a Bayesian approach and fit our model via Markov Chain Monte carlo (MCMC). To fix notation, we use i to index group (adults or children), j to index metropolitan area, and t to index days. Let y_{ijt} denote the observed number of hospitalizations in group i , area j , day t ; and let n_{ijt} denote the corresponding population size. Let $x_{ijt}(\theta_1)$ denote the common cold prevalence in group i , area j , day t , as a function of the six SIRS model parameter vector. (We denote this vector generically by θ_1 .) This function is not known in closed form, but can be evaluated for fixed θ_1 by numerically solving the SIRS system of differential equations described above. Finally, let z_{ijt} denote the vector of other covariates (given in Table 1 in the main text), but briefly: metropolitan area, day of week, time trend, influenza hospitalizations per million (smoothed), minimum temperature (centered), ozone (air quality index), and PM 2.5 ($\mu\text{g}/\text{m}^3$).

Our model assumes that the observed number of asthma-related hospitalizations follows a binomial sampling model whose parameter depends on the prevalence of colds in addition to the other covariates:

$$y_{ijt} \sim \text{Binomial}(n_{ijt}, p_{ijt})$$

$$p_{ijt} = \alpha + \beta x_{ijt}(\theta_1) + \gamma \cdot z_{ijt}$$

where α and β are scalar parameters, and where $\gamma \cdot z_{ijt}$ is the vector inner product (dot product) between the non-SIRS covariates and the parameter vector γ . We denote the non-SIRS model parameters (α, β, γ) generically by θ_2 .

Notice that the link function between parameters and the probability of a hospitalization is linear, which is nonstandard for the binomial model but justified by Gay *et al.* (1) the canonical link function would inappropriately imply that expected asthma hospitalizations scale highly nonlinearly with colds. Our binomial model is in fact nearly identical to Gay *et al.*'s Poisson model with a linear link: the Poisson distribution can be motivated as an approximation to the binomial distribution in the limit of a large sample size and a small probability (both of which hold for our data).

To fit the above model, we base our MCMC on the strategy described in Chapter 16, page 410 of (2) (3rd edition). These authors recommend that, when fitting Bayesian generalized linear models, it is computationally convenient to use a Laplace approximation to the likelihood function. The central limit theorem (or more specifically, the Bernstein-von Mises theorem) implies that such an approximation will be quite good when the sample size is large, which is certainly true in our case. Moreover, this approximation is especially helpful when running a partially collapsed MCMC like ours, in which some parameters are explicitly integrated out; see the discussion in the ‘‘MCMC Sampling’’ section below.

In our case, a Laplace approximation leads to a conditionally heteroscedastic Gaussian likelihood where the data are the observed hospitalization rates (which we express for convenience as average hospitalizations per 10,000 people), and the weights/inverse variances are proportional to the size of the metropolitan area. We adopt this approach in our MCMC, using the Laplace approximation to the binomial likelihood in the manner described in full detail by Gelman *et al.* (2).

8. MCMC Sampling

The parameters of our model fall into two blocks: (1) the six parameters of the SIRS model itself, which we have denoted as θ_1 ; and (2) the parameters associated with all other covariates in the regression model, such as day-of-week effects, influenza effects, and so forth, which we have denoted as θ_2 . A standard MCMC updating scheme would update each parameter in turn, given all the other parameters. We found that this approach mixed too slowly, given the number of parameters in our model. Therefore, we used a collapsed block-sampling scheme to explore the posterior distribution, which we now briefly describe.

The collapsed sampler involves four sub-steps for each step of the Markov chain:

- a) propose new values θ_1^* for the Block 1 parameters.
- b) Evaluate the marginal likelihood for the proposed value of θ_1^* , explicitly integrating out the parameters θ_2 in Block 2. Because we are using a Laplace approximation to the log likelihood in the manner described above, this integral is easily computed in closed form (see below).
- c) Use this marginal likelihood to calculate the Metropolis-Hastings acceptance probability, and accept or reject the proposed draw as appropriate to generate the updated value for θ_1 .
- d) Finally, sample new values for θ_2 , given the updated values for θ_1 .

Substep b, in which θ_2 is explicitly integrated out in order to assess the likelihood of the new point, is the key step that allows large gains in efficiency compared with the ordinary updating scheme. This is often referred to as a partially collapsed MCMC sampler, and inherits all the usual properties of MCMC. For a general description of the methodology and a discussion of the gains in convergence rate associated with collapsed samplers, see (3).

The ability to use the partially collapsed sampler depends crucially on the ability to carry out the integral in step (b) in closed form, thus providing strong computational motivation for the use of the Laplace approximation to the binomial likelihood. Specifically, we calculate the marginal likelihood from the Laplace approximation using the conditional maximum likelihood estimate and Fisher information matrix for θ_2 , given the proposed values of θ_1 . This result is reviewed in (4) and in most textbooks on Bayesian inference. As described above, in our case, the Laplace approximation is excellent, given that the total number of daily observations is in the tens of thousands, while the number of parameters is in the dozens or fewer. The Laplace method is far

cheaper computationally than computing the required integrals by a numerical method. Without the use of this step, MCMC runtimes become prohibitively long.

This procedure does not limit the predicted number of hospitalizations to be nonnegative. Doing so would lead to nuisance parameters. Consequently, the fitted model very occasionally generates predictions that are slightly negative in the summer period. We therefore modify the final prediction to be the maximum of the model-based forecast and zero.

We ran this scheme with single parameter updates for θ_l in step a, for 16,000 iterations, 4,000 of which are used as burn in. From this we generated the covariance matrix of θ_l . We use this to propose θ_l^* for all 6 parameters concurrently from multivariate normal distribution. This MCMC chain was run for 50,000 iterations, with 4,000 iterations burn in. The parameter estimates and covariance matrix for the best fitting model are given in Tables S2 and S3. Posterior means from the single component and block-updated chains were extremely similar, but mixing was greatly improved by block updating with multivariate jumps.

| | | Child Mean | Lower CI | Upper CI | Adult Mean | Lower CI | Upper CI |
|------------------------|--|---------------|-------------|-------------|---------------|-------------|-------------|
| Global Baseline | | -0.61 | -0.84 | -0.40 | 0.95 | 0.76 | 1.10 |
| Local Baseline | Austin-Round Rock | <i>ref</i> | <i>ref</i> | <i>ref</i> | <i>ref</i> | <i>ref</i> | <i>ref</i> |
| | Beaumont-Port Arthur | 0.45 | 0.21 | 0.69 | 0.99 | 0.86 | 1.12 |
| | Corpus Christi | 2.74 | 2.51 | 2.98 | 0.79 | 0.66 | 0.92 |
| | Dallas-Fort Worth-Arlington | 1.15 | 1.03 | 1.27 | 0.40 | 0.33 | 0.46 |
| | El Paso | 1.52 | 1.34 | 1.70 | -0.03 | -0.13 | 0.07 |
| | Houston-Sugar Land-Baytown | 0.21 | 0.09 | 0.34 | 0.10 | 0.04 | 0.17 |
| | McAllen-Edinburg-Mission | 1.27 | 1.09 | 1.45 | -0.14 | -0.25 | -0.03 |
| | San Antonio | 1.68 | 1.54 | 1.82 | 0.28 | 0.20 | 0.35 |
| SIRS Cases | 4 day | 0.020 | 0.018 | 0.022 | 0.020 | 0.013 | 0.032 |
| Time | day | -0.00001 | -0.00005 | 0.00003 | -0.00013 | -0.00015 | -0.00011 |
| | State level | | | | | | |
| Influenza | hospitalizations per million | 0.11 | 0.05 | 0.16 | 0.31 | 0.28 | 0.34 |
| Low Temperature | Average minimum temperature (centred) | -0.025 | -0.029 | -0.021 | -0.020 | -0.022 | -0.017 |
| Day of week | Monday | 0.91 | 0.79 | 1.02 | 0.46 | 0.39 | 0.52 |
| | Tuesday | 0.51 | 0.39 | 0.62 | 0.30 | 0.23 | 0.36 |
| | Wednesday | 0.23 | 0.11 | 0.35 | 0.20 | 0.13 | 0.27 |
| | Thursday | 0.07 | -0.05 | 0.19 | 0.14 | 0.07 | 0.20 |
| | Friday | <i>ref</i> | | | | | |
| | Saturday | -0.47 | -0.58 | -0.35 | -0.31 | -0.37 | -0.24 |
| | Sunday | 0.01 | -0.11 | 0.12 | -0.22 | -0.28 | -0.15 |

Table S2. Posterior mean and 95% credible intervals of the exacerbation model.
 Posterior mean and 95% credible intervals for the coefficients of the best fitting model. Italics indicate credible intervals that exclude zero. “ref” indicates that that variable was used as reference for the others in its class, and therefore its value is zero.

| | β_0 | α_{AC} | α_{AA} | σ | ω | γ |
|---------------|-----------|---------------|---------------|----------|----------|----------|
| β_0 | 0.0012 | -0.0010 | 0.0020 | -0.0007 | 0.0003 | 0.0027 |
| α_{AC} | -0.0010 | 0.0851 | -0.0003 | 0.0034 | 0.0004 | -0.0001 |
| α_{AA} | 0.0020 | -0.0003 | 0.0197 | -0.0038 | 0.0016 | 0.0073 |
| σ | -0.0007 | 0.0034 | -0.0038 | 0.0058 | -0.0007 | -0.0046 |
| ω | 0.0003 | 0.0004 | 0.0016 | -0.0007 | 0.0019 | 0.0015 |
| γ | 0.0027 | -0.0001 | 0.0073 | -0.0046 | 0.0015 | 0.0083 |

Table S3. Variance-covariance matrix of single component update chains.
 12,000 iterations after a burn in of 4,000.

9. MCMC Convergence

Chains were assessed for convergence visually, and by starting multiple chains from different areas of parameter space, to determine that they all reach the same stationary distribution. Posterior distributions appear unimodal and are approximately normally distributed.

10. Model Comparison

Table S4 shows Deviance Information criterion (DIC) scores and the components in each model. The table does not show models without varying metropolitan baseline, day of week, time trend, or influenza variable, although these models were tested. Their DIC values were much lower, and therefore they were excluded.

While the best model by DIC is model 138 at -196443.1, models 129, 130 and 137 all lie within 4 DIC points of this value. A difference of approximately 5 is required to distinguish models by this model, and so we choose the model with fewest parameters – model 129 – as the best fitting model. These models are very similar, with 129 and 130 differing from 137 and 138 by the choice of influenza variable; in the former, the state level influenza hospitalization rate in adults and children is used, and in the latter, the metropolitan-specific influenza hospitalization rate.

| DIC | Mo del Num | SIRS | Metro baseline | Day of week | time increment | Influenza | Low Temp | PM 2.5 | Lag days | Ozone | Lag days |
|------------|------------|------|----------------|-------------|----------------|-------------|----------|--------|----------|-------|----------|
| -196074.10 | 125 | Y | Y | Y | Y | state level | | | | | |
| -196086.19 | 126 | Y | Y | Y | Y | state level | | Y | 0 | | |
| -195685.43 | 127 | Y | Y | Y | Y | state level | | | | Y | 0 |

| | | | | | | | | | | | |
|------------|-----|---|---|---|---|-------------|---|---|---------|---|-------|
| -195690.20 | 128 | Y | Y | Y | Y | state level | Y | 0 | Y | 0 | |
| -196439.12 | 129 | Y | Y | Y | Y | state level | Y | | | | |
| -196442.27 | 130 | Y | Y | Y | Y | state level | Y | Y | 0 | | |
| -195997.62 | 131 | Y | Y | Y | Y | state level | Y | | Y | 0 | |
| -196004.21 | 132 | Y | Y | Y | Y | state level | Y | Y | 0 | Y | 0 |
| -196045.32 | 133 | Y | Y | Y | Y | metro level | | | | | |
| -196061.51 | 134 | Y | Y | Y | Y | metro level | | Y | 0 | | |
| -195676.77 | 135 | Y | Y | Y | Y | metro level | | | | Y | 0 |
| -195681.40 | 136 | Y | Y | Y | Y | metro level | | Y | 0 | Y | 0 |
| -196441.94 | 137 | Y | Y | Y | Y | metro level | Y | | | | |
| -196443.16 | 138 | Y | Y | Y | Y | metro level | Y | Y | 0 | | |
| -196006.61 | 139 | Y | Y | Y | Y | metro level | Y | | | Y | 0 |
| -196011.66 | 140 | Y | Y | Y | Y | metro level | Y | Y | 0 | Y | 0 |
| -195487.17 | 141 | Y | Y | Y | Y | | | | | | |
| -195525.10 | 142 | Y | Y | Y | Y | | | Y | 0 | | |
| -195188.71 | 143 | Y | Y | Y | Y | | | | | Y | 0 |
| -195196.89 | 144 | Y | Y | Y | Y | | | Y | 0 | Y | 0 |
| -196032.81 | 145 | Y | Y | Y | Y | | Y | | | | |
| -196035.24 | 146 | Y | Y | Y | Y | | Y | Y | 0 | | |
| -195623.87 | 147 | Y | Y | Y | Y | | Y | | | Y | 0 |
| -195625.58 | 148 | Y | Y | Y | Y | | Y | Y | 0 | Y | 0 |
| -196368.79 | 200 | Y | Y | Y | Y | state level | Y | Y | 0, 1 | | |
| -196291.07 | 201 | Y | Y | Y | Y | state level | Y | Y | 0, 1, 2 | | |
| -196365.83 | 202 | Y | Y | Y | Y | state level | Y | Y | 1 | | |
| -196288.98 | 203 | Y | Y | Y | Y | state level | Y | Y | 1,2 | | |
| -195944.45 | 204 | Y | Y | Y | Y | state level | Y | | | Y | 0, 1 |
| -195878.13 | 205 | Y | Y | Y | Y | state level | Y | | | Y | 0, 1, |
| -195991.17 | 206 | Y | Y | Y | Y | state level | Y | | | Y | 1 |
| -195924.35 | 207 | Y | Y | Y | Y | state level | Y | | | Y | 1,2 |
| -195916.81 | 208 | Y | Y | Y | Y | state level | Y | Y | 1 | Y | 1 |
| -195775.61 | 209 | Y | Y | Y | Y | state level | Y | Y | 1, 2 | Y | 1, 2 |

Table S4. Model comparison.

Table shows the DIC value for each model compared, where the other columns specify the model name and components of the model. A “Y” value indicates that the component is included, and other notes indicate information about the variable.

11. Comparison without colds in the model

To test how well we can fit the temporal pattern of asthma hospitalizations without using a dynamic transmission model, we compare the best fitting model described in the main paper with one without the SIRS component. The variables included are shown in **Table S5** below, where the difference is marked with a star. By likelihood ratio test, the SIRS model fits better ($p < .01$).

Fig. S6 shows the posterior mean model prediction in red and the data in grey. This is a summary figure for number of asthma hospitalizations per day in children in all metropolitan areas, although the model is fitted to individual metropolitan areas. Panel a is the best fitting model from the main paper with the model lacking the SIRS component in b. Fig. S6 clearly demonstrates that the temporal pattern of hospitalizations in children has a markedly different pattern by that generated without the SIRS transmission model.

Fig. S7 shows the same comparison for adults. Although the fit in the lower panel is not so obviously poor as for children, there are local patterns captured better by the model

with the dynamic SIRS model. By likelihood ratio test, the model with SIRS is preferred ($p < .01$).

| Data | Difference |
|----------------------|------------|
| Common Cold | * |
| Influenza prevalence | |
| Day of week | |
| Time trend | |
| Local Intercept | |
| Low Temperature | |

Table S5. Component comparison of model without SIRS common cold variable.

Table shows the components compared in this section, where the * indicates the difference between these 2 models.

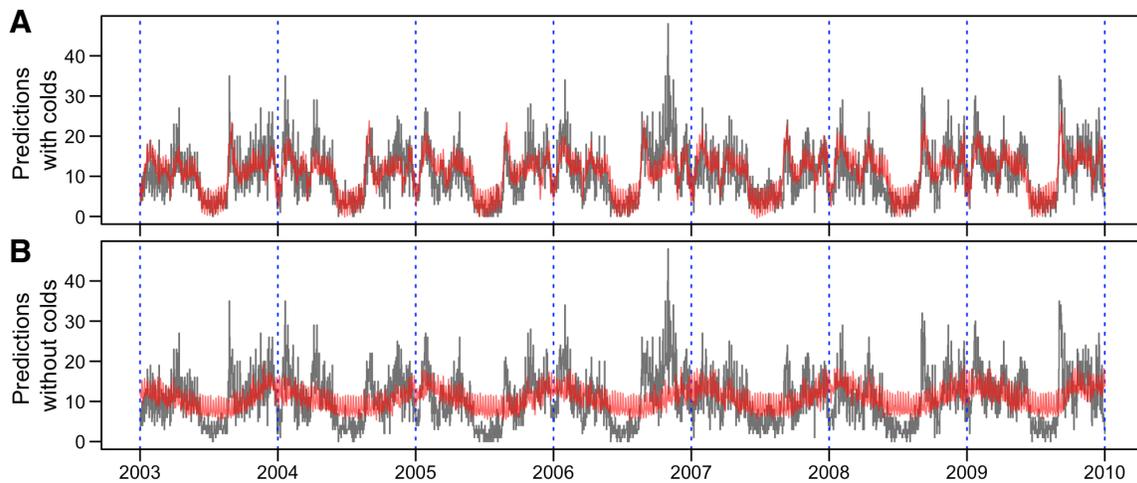


Fig. S6. Comparison of fit with and without common cold model (children).

The red line shows the posterior mean number of cases per day in children in all metropolitan areas. Grey line is the daily asthma hospitalizations in children aged 5-18 for the 8 metropolitan areas in the study. (a) best fitting model from the main text, with dynamic SIRS transmission model, (b) linear model with the SIRS model removed.

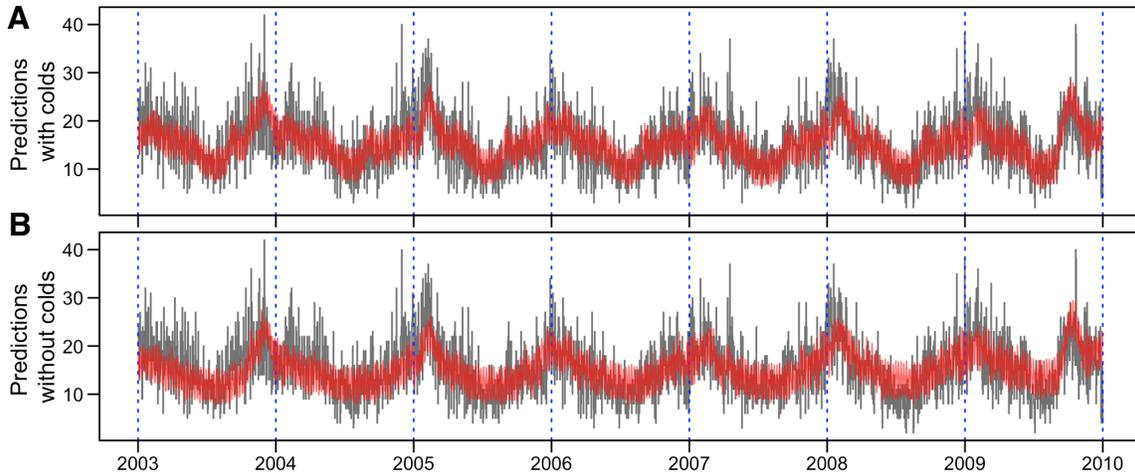


Fig. S7. Comparison of fit with and without common cold model (adults). The red line shows the posterior mean number of cases per day in adults in all metropolitan areas in the fitted model. Grey line is the daily asthma hospitalizations in adults aged 19-55 for the 8 metropolitan areas in the study. (a) best fitting model from the main text, with dynamic SIRS transmission model, (b) linear model with the SIRS model removed.

12. With 0/1 holiday, linear model

To test whether being at school is a driver for asthma exacerbation, rather than school being a place for the transmission of common cold viruses, we formulate a linear model where children have an indicator variable for being in school (the “school indicator model”), in place of the SIRS model. The components of the model are shown in **Table S6**. The rationale for this test is that children may be allergic to something in the school, such as chalk dust (5) the school buildings (6–8) which puts them at increased risk of exacerbation.

The results of fitting the school indicator model are shown in **Fig. S8a**, and compared with the best fitting model presented in the main text, and also shown in **Fig. S8b**. Compared with a model with no school variable (i.e. Fig. S6b), the school indicator model improves the appearance of fit. However, the complex pattern of summer troughs, September peaks, and school-vacation effects is clearly better captured by the complete model. The complete model is preferred by likelihood ratio test ($p < 0.01$, log likelihood value of indicator model: -194796, posterior mean log likelihood of complete model: -196449).

| Data | Children | Adults |
|-----------------------|----------|--------|
| Local Baseline | * | * |
| Day of week | * | * |
| Time trend | * | * |
| Influenza prevalence | * | * |
| Low Temperature | * | * |
| 0/1 for out/in school | * | |

Table S6. Component Comparison of indicator model.

The components of the child and adult model tested against the best fitting model from the main paper. * marks that that variable is present in the model.

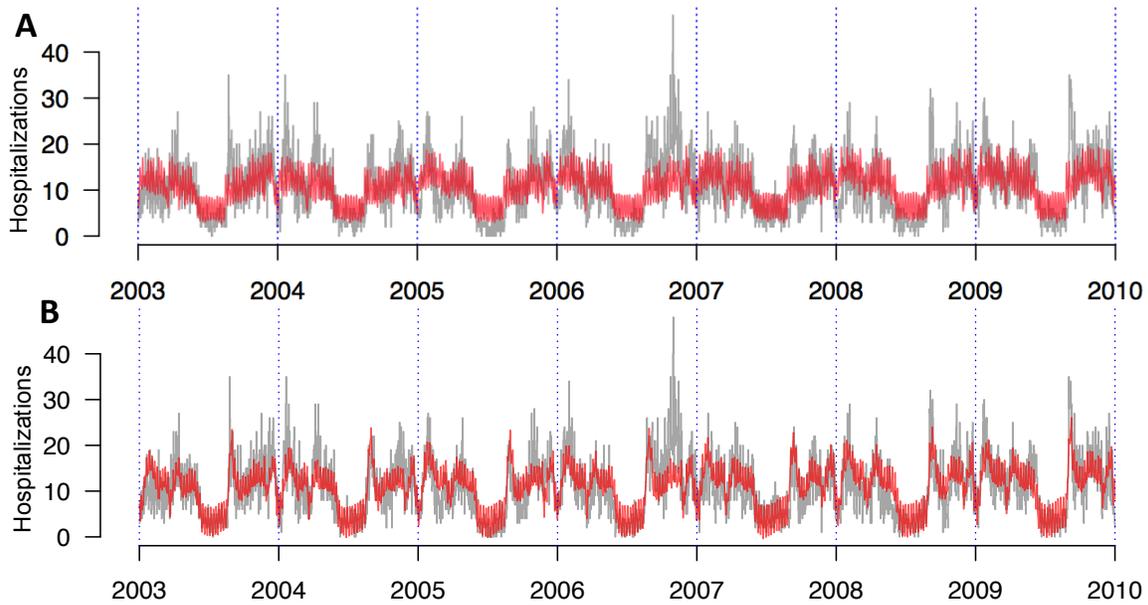


Fig. S8. Comparison of indicator model with full model (children).

Posterior mean model fit for children the model which has an indicator variable for school calendar (a), or the school calendar-driven SIRS model from the main text, (b). The model fit is the fitted number of cases in all 8 metropolitan areas, although the models are fitted to each metropolitan area separately. Red shows the posterior mean model fit, and grey the data.

13. Best fitting model city-level results

Figures in this section show the 7-day rolling mean number of asthma hospitalizations from 100 simulations in adults and children. We sampled 20 parameter sets from the joint posterior distribution, and calculated the predicted daily asthma hospitalization rate from the model. For each of those 20, we simulated 5 time series of incident cases using a non-homogenous Poisson simulation package, NHPPoisson v2.1, in R (9).

Fig. S9 and Fig. S10 show these 100 simulations for children and adults respectively, for each of the 8 metropolitan areas, with the true number of cases marked in black. These figures show good agreement between simulated number of cases and true data. The simulated cases in children have excellent concordance in the summer, return to school, and behavior around major holidays. The anomalously large peaks visible in children in Dallas-Fort Worth-Arlington (Metro Area 19100) and El Paso (Metro Area 21340) in 2006 are more obvious in the individual city plots. It is interesting that there is not a matching increase in asthma hospitalizations in adults in those time periods.

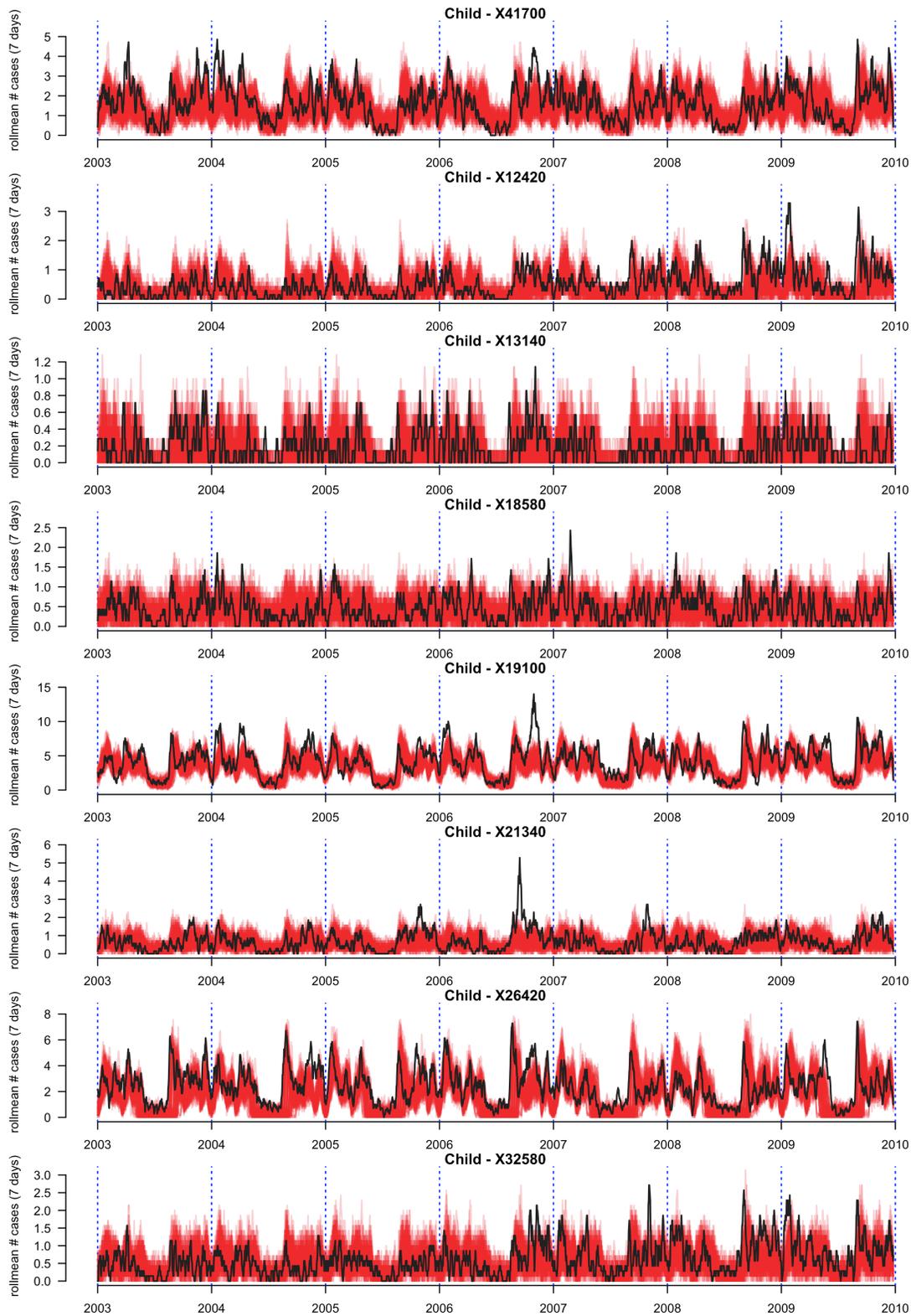


Fig. S9. Fit to each metropolitan area (children).

The black line shows the 7-day rolling mean number of hospitalizations in children in each metropolitan area. The red lines are 5 non-homogeneous Poisson simulations from each of 20 parameter sets sampled from the joint posterior distribution of the SIRS model. The red lines are semi-transparent, so darker colour represents areas where more simulated values lie.

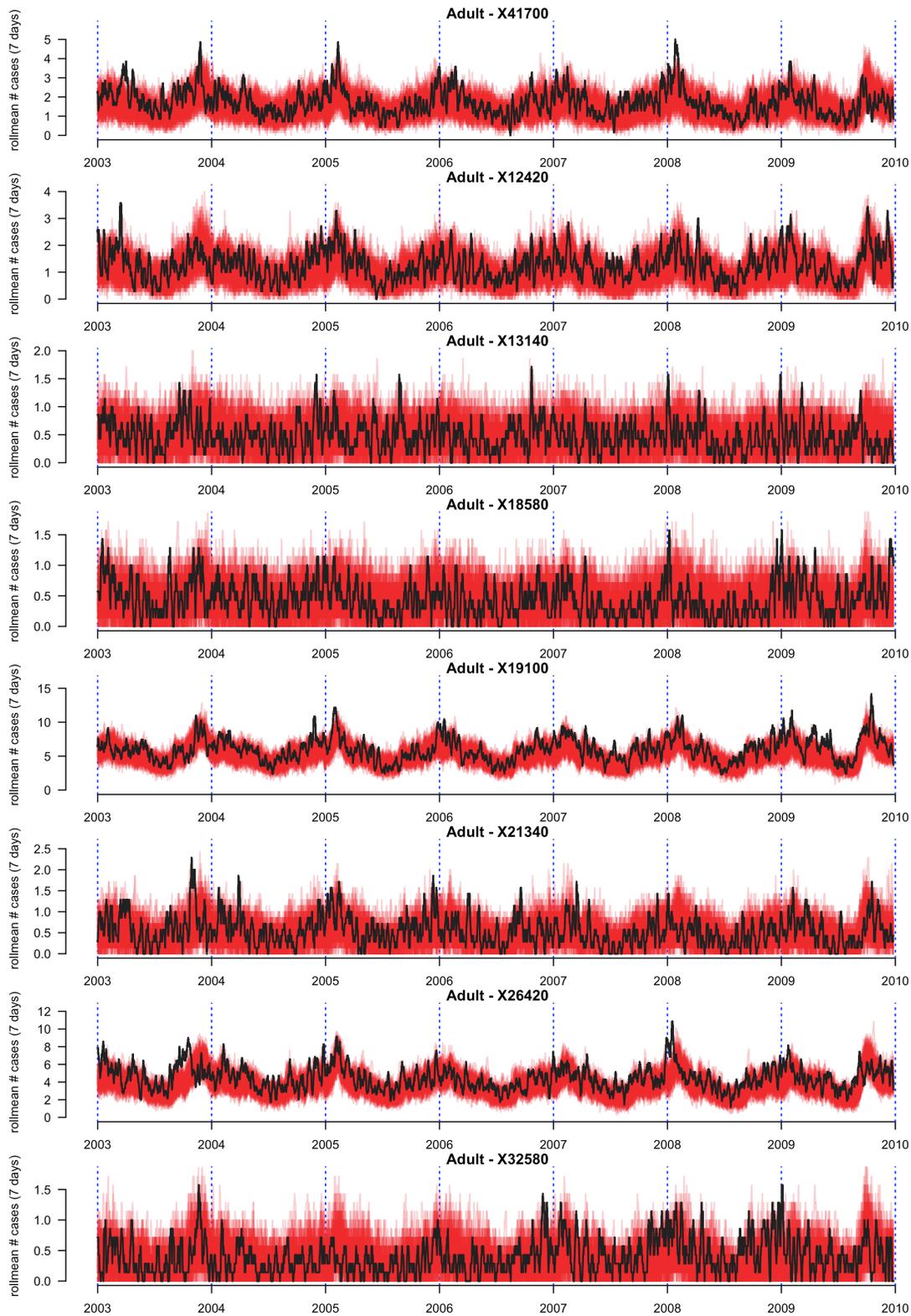


Fig. S10. Fit to each metropolitan areas (adults).

The black line shows the 7-day rolling mean number of hospitalizations in adults in each metropolitan area. The red lines are 5 non-homogeneous Poisson simulations from each of 20 parameter sets sampled from the joint posterior distribution of the SIRS model. The red lines are semi-transparent, so darker colour represents areas where more simulated values lie.

14. Daily residuals for each city

The daily residuals from the posterior mean SIRS parameters for children and adults are shown in Fig. S11 and Fig. S12 respectively. The figures show that the residuals are approximately normally distributed, and more so for larger metropolitan areas. Those with smaller asthma hospitalization rates appear with “banded” residuals because the low hospitalization rate results in the appearance of discrete hospitalization events.

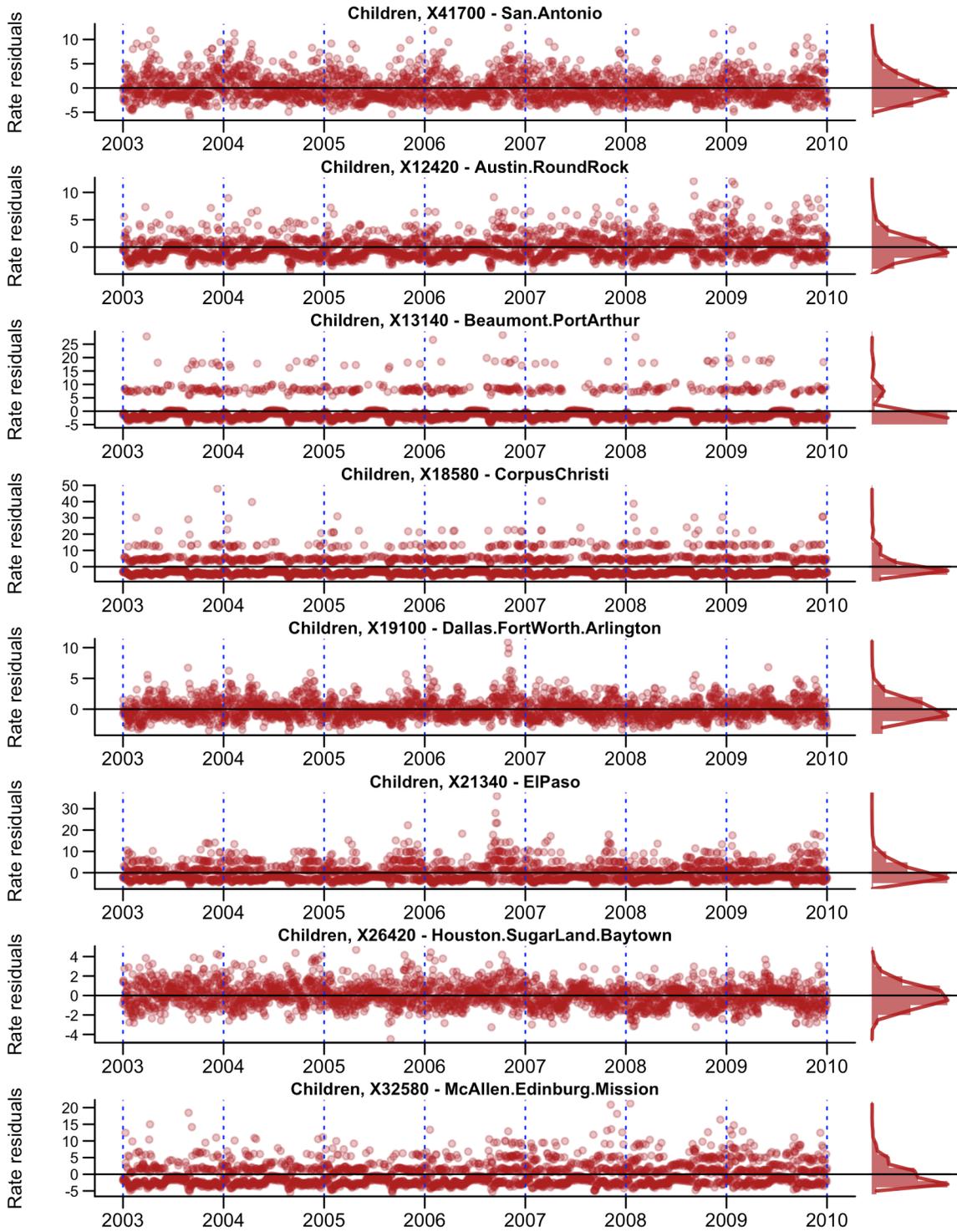


Fig. S11. Residuals of the model (children).

Daily residuals for best fitting model, at posterior mean of SIRS parameters. Values for children.

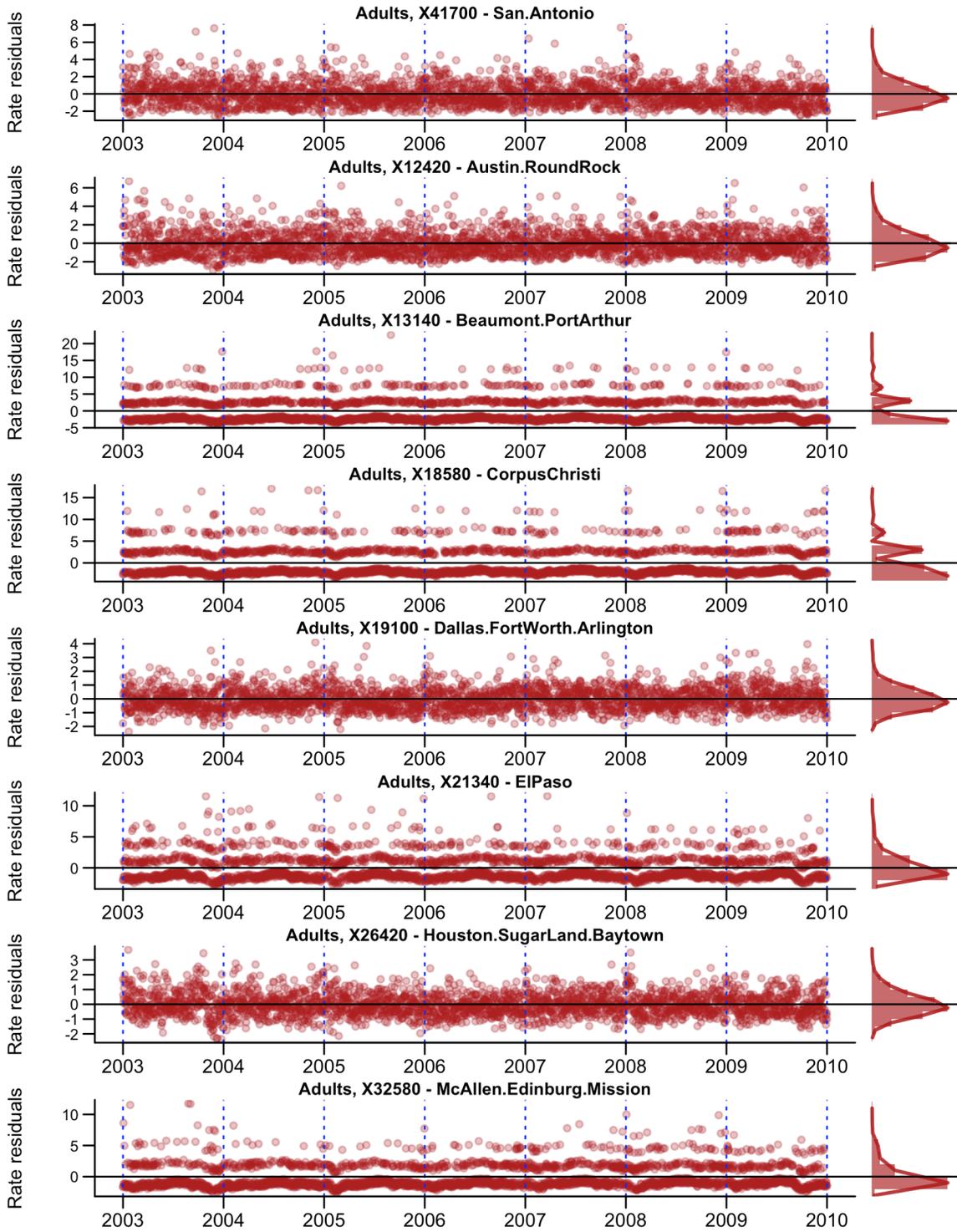


Fig. S12. Residuals of the model (adults).

Daily residuals for best fitting model, at posterior mean of SIRS parameters. Values for adults.

15. Monthly residuals for each city

To assess whether there is temporal variation in goodness of fit of the model, we check the monthly residuals. Fig. S13 and Fig. S14 show the monthly residuals for the posterior mean fitted model, for the best fitting model in the analysis for children and adults respectively. There is no striking temporal variation in residuals, e.g. all cities having low values in the summer, or winter.

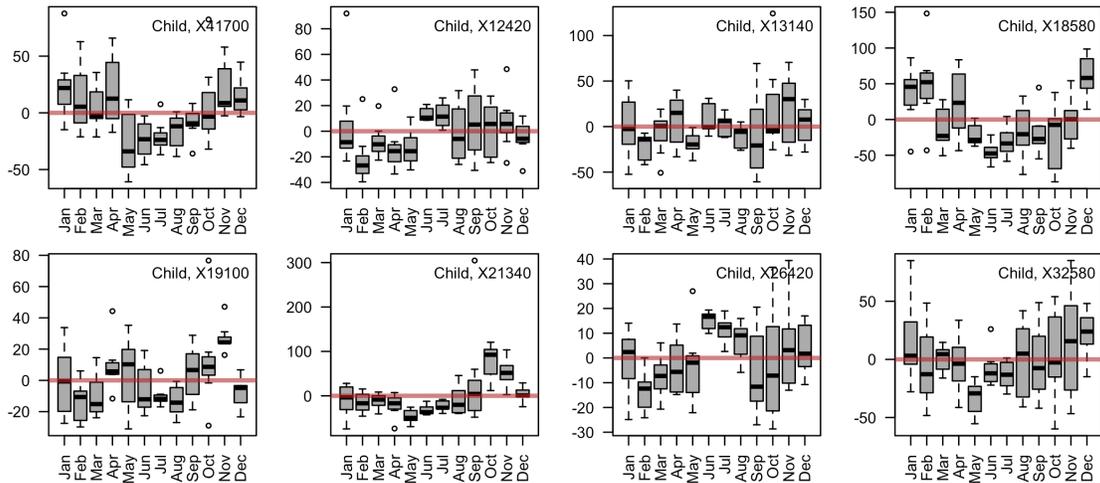


Fig. S13. Monthly residuals (children).

Residuals for each month of the study for the posterior mean model fit, of the best fitting model presented in the main paper. Plot shows results for children.

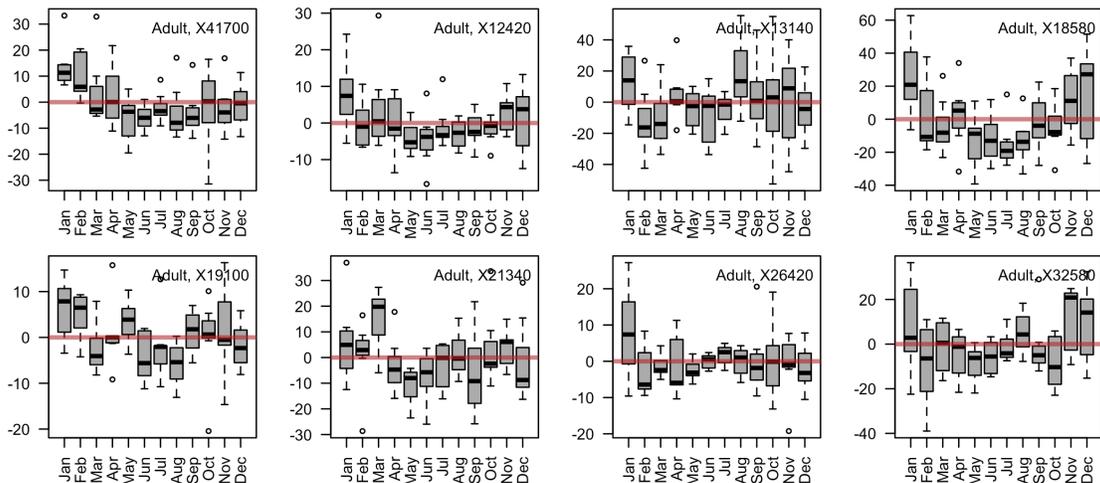


Fig. S14. Monthly residuals (adults).

Residuals for each month of the study for the posterior mean model fit, of the best fitting model presented in the main paper. Plot shows results for adults.

16. Calculation of the average number of colds per person

Average number of colds is calculated by dividing the cumulative number of infections in adults and children per year by the mid-year population of each metropolitan area. Due to uncertainty in the parameter estimates for the SIRS model, there is some uncertainty in the exact value. The values generated by the posterior mean of the SIRS estimates are shown in the main paper, and in Fig. S15, the average number of colds per person per year for each metropolitan area is shown for 20 parameter sets sampled from the joint posterior distribution. The red points are children, and the blue points are adults. As in the main paper, the red and blue windows represent the range of common cold infections per person from the UK NHS website on general information on the common cold (10). The change through time is the result of changing fraction of children in the population.

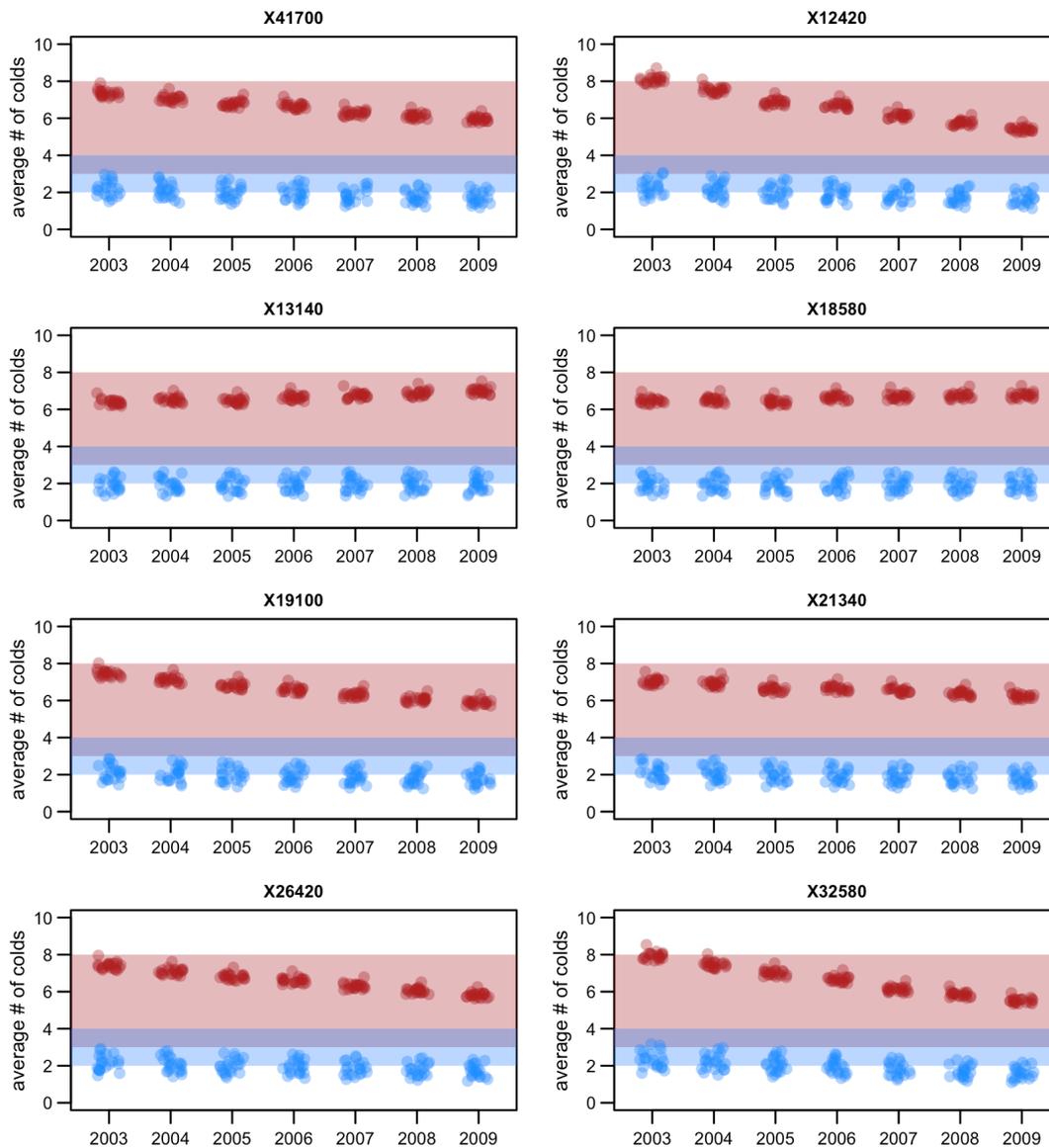


Fig. S15. Average number of colds per person per year in each metropolitan area.

Average number of colds per person per year for each metropolitan area is shown for 20 parameter sets sampled from the joint posterior distribution. Red is number of colds per year in children, and blue in adults. The red and blue shaded windows represent the range of common cold infections per person from the UK NHS website on general information on the common cold.

17. Change in school start date during study period

In 2007 a regulation went into effect preventing Texas schools opening before the 4th Monday in August (11). Fig. S16a shows the 7-day rolling sum of asthma hospitalizations in children in Texas aged 5–17 for years before 2007 (red) and after 2007 (blue). The September peak of asthma hospitalizations shifts after this regulation, indicating that return-to-school, and not just time of year is critical for generating the September peak. Our model easily accommodates this change in school start date, because there is a mechanistic connection between return to school and resurgence of asthma exacerbations. One simulation from the posterior mean SIRS parameters is shown in Fig. S16b.

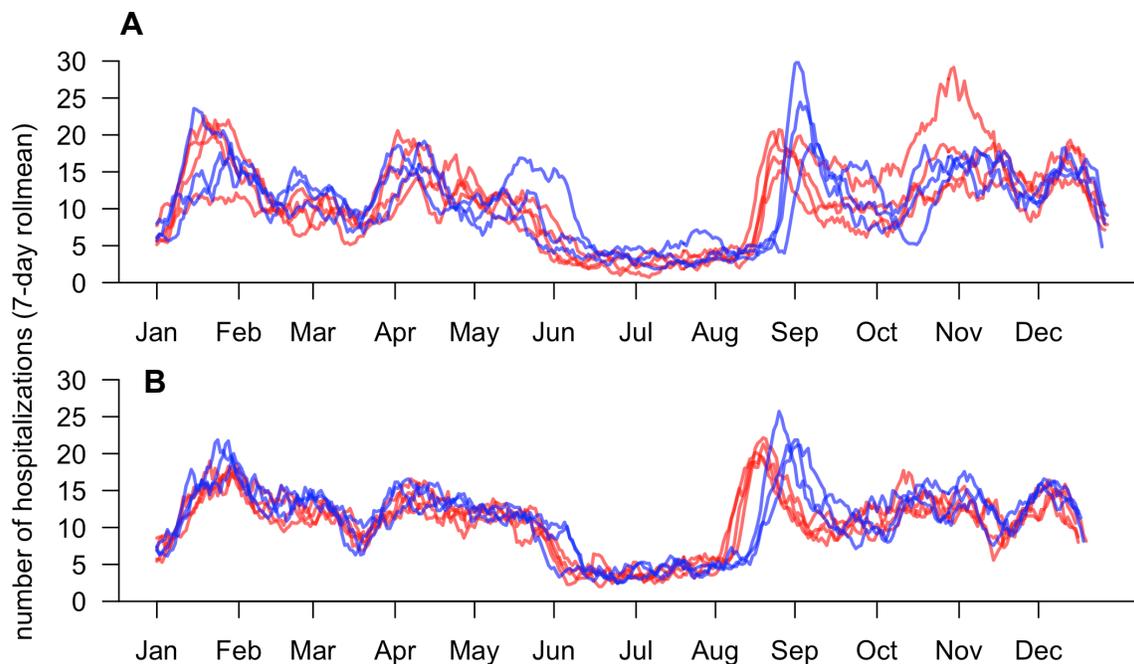


Fig. S16. Change in school start date in data and model.

7-day rolling sum of hospitalizations (principal code 493) for all areas of Texas. Red lines are years before the regulation went into effect, and blue lines are years after. (A) Observed number of hospitalizations. (B) One model realization showing predicted number of daily hospitalizations.

18. Model fits by day of the week

To visualize the combined impacts of different factors on asthma exacerbation rate, we show the fitted hospitalization rate and the contribution from each variable, for two example weeks: one randomly chosen week while school is in session (Figs S17a and S18a) and another randomly chosen week during a school closure period (Figs S18b & S18b). We only present these results for Dallas-Fort-Worth-Arlington, but all metropolitan areas show similar patterns. The results for children and adults are given in Figs 17 and 18, respectively.

While school is in session, the fitted hospitalization rate for children attributable to the common cold (red bar) is higher later in the week, because transmission between children is higher while in school (Fig S17a). To account for the several day window between infection and asthma exacerbation, we aggregated common cold prevalence over 4-day intervals. Therefore, the contribution from common cold prevalence rises gradually as prevalence increases during the school-week, and remains high on the weekend, despite the lower transmission among children on the weekend. The black points are the total fitted hospitalization rate, which is the sum of all bars on each day.

By comparison, when schools are closed in July, the overall hospitalization rate is much lower (Fig S17b), demonstrating the influence of school-based interactions on common cold prevalence. In addition, the day-of-the-week differences have a more pronounced effect on overall hospitalization rates, because of the lower impact of common cold prevalence relative to during school sessions.

In contrast, the adult hospitalization rate is more influenced by the baseline rate and the metropolitan baseline rate (combined into darkest bar) (Fig S18). Given the reduced impact of other covariates, the overall hospitalization rate is lower in October for adults than for children (Fig S18a), and the difference in hospitalization rates between the school-open week (Fig 18a) and the school-closed week (Fig 18b) is reduced in adults relative to the same difference between children.

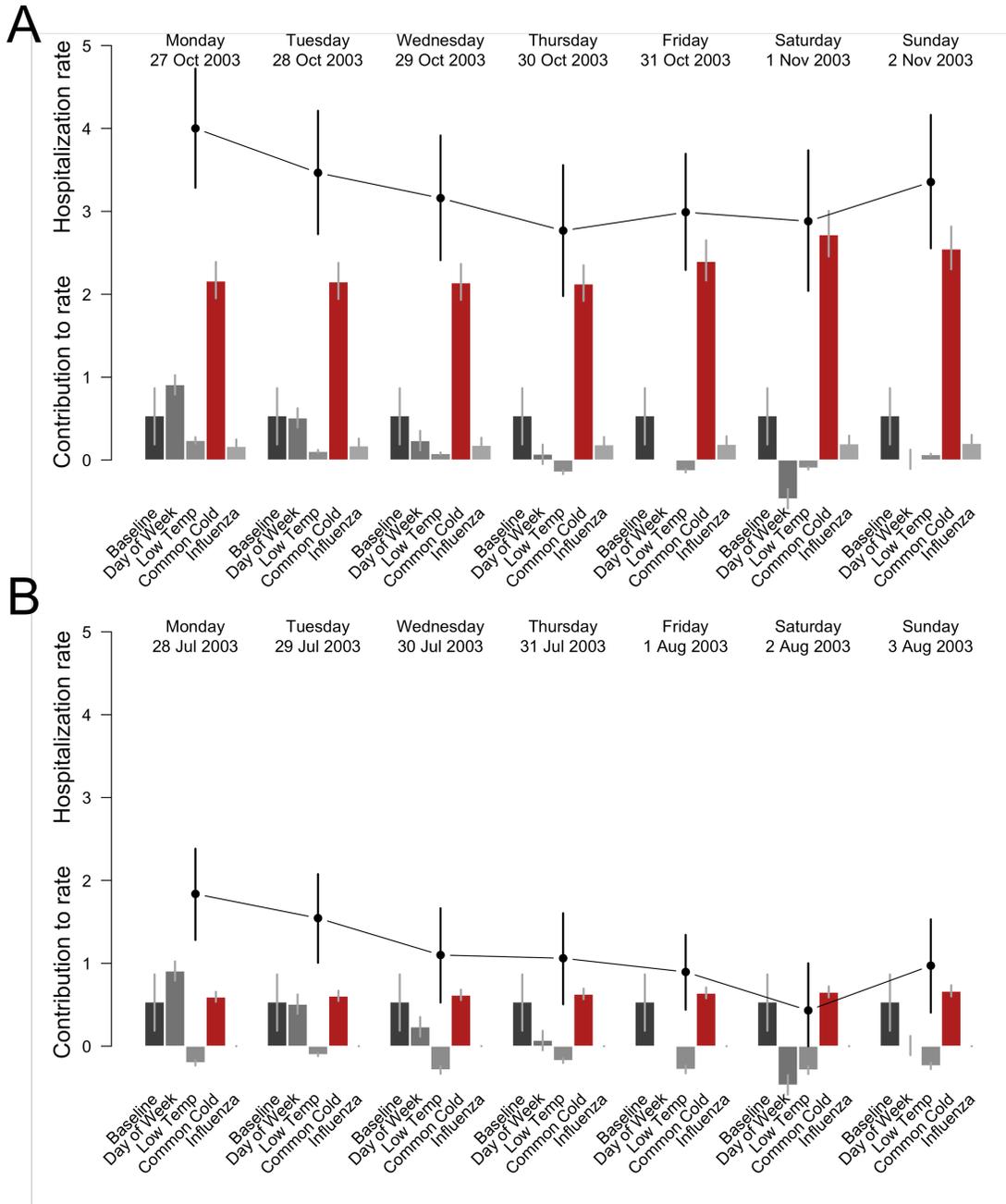


Figure S17. Fitted hospitalization rate and contribution to the fitted rate from each variable. Rate shown per 100,000 children, for two example weeks, in Dallas-Fort Worth-Arlington.

A) Week beginning 27th October 2003, while school is in session, B) Week beginning 28th July 2003, during summer school closure. Black dots indicate the fitted hospitalization rate (and 95% credible interval) for each day of the study. Barplots show the contribution of each variable in the model to the fitted rate of hospitalization.

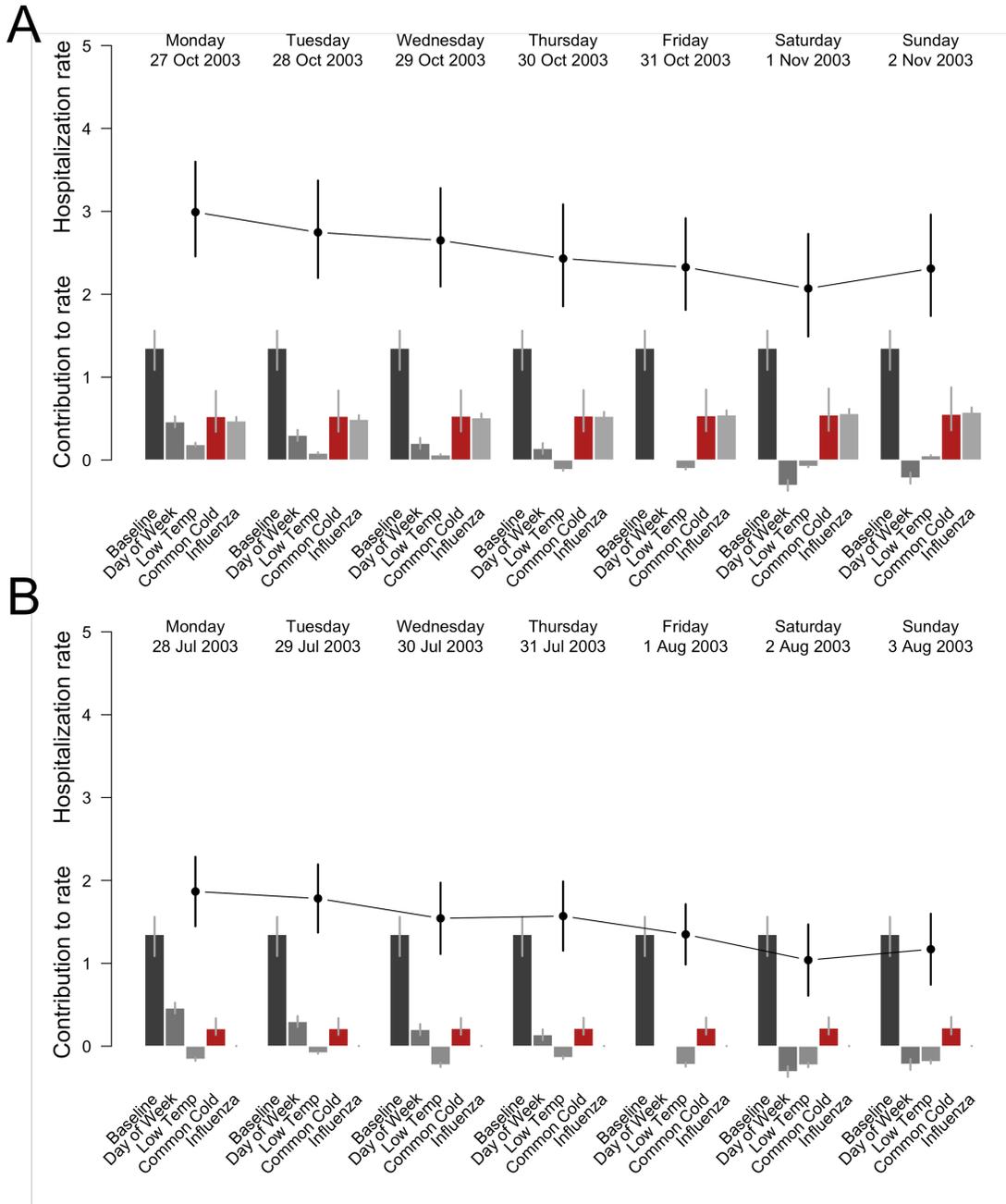


Figure S18. Fitted hospitalization rate and contribution to the fitted rate from each variable. Rate shown per 100,000 adults, for two example weeks, in Dallas-Fort Worth-Arlington.

A) Week beginning 27th October 2003, while school is in session B) Week beginning 28th July 2003, during summer school closure. Black dots indicate the fitted hospitalization rate (and 95% credible interval) for each day of the study. Barplots show the contribution of each variable in the model to the fitted rate of hospitalization.

19. References

1. Gay NJ, Andrews NJ, Trotter CL, Edmunds WJ. Estimating deaths due to influenza and respiratory syncytial virus. *JAMA* [Internet]. 2003 May [cited 2011 Mar 16];289(19):2499; author reply 2500–2. Available from: <http://jama.ama-assn.org>
2. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. 2nd Editio. Chapman & Hall/CRC; 2004. 668 p.
3. Dyk DA Van, Park T. Partially Collapsed Gibbs Samplers: Theory and Methods. *J Am Stat Assoc* [Internet]. 2008 [cited 2015 Feb 10];103(482). Available from: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.169.5077>
4. Kass R, Raftery A. Bayes Factors. *J Am Stat Assoc* [Internet]. 1995 [cited 2015 Feb 10];90:773–95. Available from: http://www.jstor.org/stable/2291091?seq=1&sid=21105307946651&uid=4&uid=3738032&uid=2#page_scan_tab_contents
5. Esquibel KP, Foster CR, Garnier VJ, Saunders ML. A program to help asthmatic students reach their potential. *Public Health Rep* [Internet]. 1984 Jan [cited 2015 Feb 15];99(6):606–9. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1424651&tool=pmcentrez&rendertype=abstract>
6. Einarsson R, Munir A, Dreborg S. Allergens in school dust: II. Major mite (I, I) allergens in dust from Swedish schools. *J Allergy Clin Immunol* [Internet]. Elsevier; 1995 May 5 [cited 2015 Feb 15];95(5):1049–53. Available from: <http://www.jacionline.org/article/S0091674995701087/fulltext>
7. Zock JP, Brunekreef B. House dust mite allergen levels in dust from schools with smooth and carpeted classroom floors. *Clin Exp Allergy* [Internet]. 1995 Jun [cited 2015 Feb 15];25(6):549–53. Available from: <http://doi.wiley.com/10.1111/j.1365-2222.1995.tb01093.x>
8. Dautel P, Whitehead L, Tortolero S, Abramson S, Sockrider M. Asthma Triggers in the Elementary School Environment: A Pilot Study. *J Asthma* [Internet]. 1999 [cited 2015 Feb 15];36(8):691–702. Available from: <http://informahealthcare.com/doi/pdf/10.3109/02770909909055421>
9. Cebrian AC. NHPoisson: Modelling and validation of non homogeneous Poisson processes [Internet]. 2014. p. <http://cran.r-project.org/web/packages/NHPoisson/NHPoisson.pdf>
10. NHS. NHS common cold [Internet]. 2013 [cited 2013 Aug 20]. Available from: <http://www.nhs.uk/Conditions/Cold-common/Pages/Introduction.aspx>
11. Association of School Boards T. School Start Date. 2012. p. <https://www.tasb.org/legislative/documents/lgibsta>.