

Supporting Online Material for

Demographic Variability, Vaccination, and the Spatiotemporal Dynamics of Rotavirus Epidemics

Virginia E. Pitzer,* Cécile Viboud, Lone Simonsen, Claudia Steiner, Catherine A. Panozzo, Wladimir J. Alonso, Mark A. Miller, Roger I. Glass, John W. Glasser, Umesh D. Parashar, Bryan T. Grenfell

*To whom correspondence should be addressed. E-mail: vep2@psu.edu

Published 17 July 2009 *Science* **325**, 290 (2009)

DOI: 10.1126/science.1172330

This PDF file includes:

Materials and Methods

Figs. S1 to S12

Tables S1 to S4

References

Supporting online material

Demographic variability, vaccination, and the spatiotemporal dynamics of rotavirus epidemics

Virginia E. Pitzer^{1,2}, Cécile Viboud², Lone Simonsen³, Claudia Steiner⁴, Catherine A. Panozzo⁵, Wladimir J. Alonso², Mark A. Miller², Roger I. Glass², John W. Glasser⁵, Umesh D. Parashar⁵, & Bryan T. Grenfell^{1,2}

¹ *Center for Infectious Disease Dynamics, Pennsylvania State University, State College, Pennsylvania, USA*

² *Fogarty International Center, National Institutes of Health, Bethesda, Maryland, USA*

³ *School of Public Health and Health Services, George Washington University, Washington DC, USA*

⁴ *Healthcare Cost and Utilization Project, Center for Delivery, Organization and Markets, Agency for Healthcare Research and Quality, US Department of Health and Human Services, Rockville, Maryland, USA*

⁵ *Epidemiology Branch, Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA*

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention (CDC).

Contents

- Comparison of spatiotemporal patterns in timing of rotavirus epidemics, climate indicators, and birth rates
- Geographic spread of infection: evidence for local persistence
- Transmission model details
- Relationship between birth rates and the timing of epidemics predicted by the transmission model
- Vaccination impact
- Sensitivity analyses
- Supporting figures and tables

Data availability: The key data necessary to evaluate the model are available from the authors upon request.

Comparison of spatiotemporal patterns in timing of rotavirus epidemics, climate indicators, and birth rates

Since climatic factors have sometimes been hypothesized to drive the traveling wave of rotavirus epidemics across the United States, we explored the statistical association between spatial and temporal variation in timing of rotavirus epidemics and environmental covariates. For comparative purposes, we conducted a similar analysis between timing of rotavirus epidemics and birth rates.

Definition of climate indicators

We compiled data on a variety of environmental factors, including solar radiation, precipitation, vapor pressure, and temperature, and derived summary environmental indicators for each winter season and state. To obtain a comparable rotavirus season in each state, we average climate indicators for December and January of each winter. Since rotavirus epidemic patterns appeared to change around the late 1990s, we compared climate indicators for two time periods: early years (1995-97) when the rotavirus traveling wave was most pronounced, and recent years when it was less pronounced (2000-2002). A description of the sources of the climate data is given below.

a) Solar radiation data

We compiled measurements of the amount of electromagnetic energy (solar radiation) incident on the surface of the earth across the U.S. Because measurement systems and processing of solar radiation data have changed in the last decades and may have affected

overall trends in observations (1), we considered two independent datasets, derived from ground observations and satellites. Since both datasets were highly correlated, we only present statistical results for ground observation exclusively.

Ground observations were obtained from the National Solar Radiation Database (NSRDB) of the National Renewable Energy Laboratory (http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/hourly/list_by_USAFN.html). We used data for the station closest to the more populated area in each state. The units are expressed in Wh/m²/hour (energy by square meter averaged by hours - including night hours) (Fig. S1).

Satellite-derived solar surface irradiance data were obtained from NASA GEWEX Surface Radiation Budget (SRB, http://eosweb.larc.nasa.gov/PRODOCS/sse/table_sse.html). Units in Wh/m²/day (daily total irradiance by square meter averaged by hours) (Fig. S2).

b) Data on precipitation, vapor pressure and temperature

Monthly climatic data were obtained from worldwide climate maps generated by the interpolation of climatic information from ground-based meteorological stations with a monthly temporal resolution and 0.5° (latitude) by 0.5° (longitude) spatial resolution (2). The climatic variables used were precipitation, monthly average of daily minimum and maximum temperatures, and vapor pressure.

These monthly climatic variables were extracted from the pixels with more than 10,000 inhabitants within each U.S. state and averaged for the two periods of interest (1995-97 and

2000-02). Average daily maximum temperatures (TMX) and minimum temperatures (TMN) are expressed in degrees Celsius, vapor pressure (VAP) in hecta-Pascals, and precipitation (PRE) in millimeters. Maps in Fig. S3-S6 show spatial and temporal patterns in precipitation, vapor pressure and temperature across the U.S.

c) Birth rate data

We used the on-line tool from the National Center for Health Statistics to retrieve annual birth rates for each state, 1992-2006

(<http://www.cdc.gov/nchs/datawh/vitalstats/VitalStatsbirths.htm>). Maps of average birth rates are presented in the main text (Fig. 1B).

Statistical approach

We conducted univariate and multivariate regression analyses to explore the relative contribution of climate and birth rates on the timing of rotavirus epidemics in U.S. states.

Timing of epidemics (T) was defined by the mean week of rotavirus activity for each season and epidemic year, as follows:

$$T_{s,y} = \frac{\sum_{w \in [1,52]} w * cases_{s,y,w}}{\sum_{w \in [1,52]} cases_{s,y,w}}$$

Here $T_{s,y}$ is the mean timing of rotavirus activity for state s and epidemic year y , w is an index for the week of year running from 1 to 52, where week 1 indicates the first week of July and week 52 is the last week of June, $cases_{s,y,w}$ is the number of cases of rotavirus hospitalizations reported in state s , epidemic year y and week w . Hence T is the mean week of the epidemic, where each week is weighted by the number of rotavirus cases.

If there is an association between rotavirus epidemic timing, climate indicators, and/or birth rates, this relationship is expected to explain spatial variations in epidemic timing each winter, as well as time trends in each state (e.g. California). To test the association, we first conducted an analysis of spatial variation in mean rotavirus timing stratified by two periods of interest, focusing on early years of the study when the geographic differences in timing of rotavirus epidemics was most pronounced (1993-1998), and in recent years when it was less pronounced (2001-2006). Second, we conducted an analysis of temporal trends in mean rotavirus timing in each state between 1993 and 2006. Given that each exploratory analysis involved multiple testing for 6 covariates (5 climate indicators + birth rate), we focused on factors with a P -value below $0.05/6=0.0083$, using Bonferroni correction.

Results

The results of the univariate and multivariate statistical analyses are presented in table S1. Birth rate was the covariate most strongly associated with rotavirus timing in all univariate analyses, and it was the only factor associated both with temporal and spatial patterns in rotavirus epidemic timing across states and epidemic years (all $P<0.008$). The association was quite strong and results were significant despite the small sample size in the analyses of early years (11 states due to data availability issues). Stepwise multivariate regression identified birth rate as the only predictor of rotavirus timing explaining spatial variations in early years as well as temporal changes, while the best model predicting spatial variations in recent years included birth rates and vapor pressure (Table S1). We note that in the latter analysis, birth rate had a stronger association with rotavirus timing than vapor pressure. Overall, birth rate alone explained 65-71% of the spatial variation in rotavirus epidemic

timing across U.S. states in the early and recent period, and 61% of temporal trends in rotavirus timing observed in 10 states with long-term records. Further, a separate analysis of monthly variations in climate indicators suggests that the timing of maximum and minimum indicators have not changed over the study period (Fig. S7).

In conclusion, this statistical analysis identifies birth rates as a strong predictor of spatial and temporal patterns in rotavirus epidemics across the U.S., and suggests that more homogeneous birth rates across the U.S. in recent years (mostly due to declining birth rates in southwestern states, Fig. 1B) have driven rotavirus epidemics to become more synchronized across the country.

Geographic spread of infection: evidence for local persistence

Another possible explanation for the apparent southwest-to-northeast spread of rotavirus is that antigenically novel strains are introduced each year from endemic regions with high year-round rotavirus activity, such as Central America, and spread geographically across the U.S. This hypothesis is reminiscent of the geographical spread of another acute viral infection, influenza, which has also been characterized in the U.S. (3). However, the pattern of spread exhibited by rotavirus differs considerably from that of influenza, for which epidemics tend to begin relatively synchronously in major population centers then spread hierarchically to more rural areas (3). Such patterns of spread inherently assume local extinction followed by annual reinvasion from outside of the U.S., which has been confirmed by phylogenetic analysis of influenza virus sequences (4).

In the case of rotavirus, a variety of epidemiologic and genetic evidence suggests that infection persists in the population during epidemic troughs. Studies focusing on rotavirus infections in adults have found relatively little seasonal variability in the occurrence of infection, as evidenced by both IgM titres and viral isolation (5, 6). Furthermore, a study noted that certain G1 electropherotypes persisted in the population for more than a year in Australia; also, there was no consistent direction of spread of individual strains from city to city, despite epidemics peaking 1-2 months earlier in Western Australia compared to eastern cities (7). In the U.S., the distribution of rotavirus strains differs considerably from state to state in a given season, even within neighboring states (8, 9), suggesting that geographic diffusion of novel strains plays a limited role in the overall dynamics.

Thus, hierarchical spread of infection seems an unlikely explanation for the observed spatiotemporal pattern of rotavirus epidemics. While there may be some local fadeout of rotavirus infection during epidemic troughs in less populous regions, rotavirus cases are reported throughout the year at state and national levels. We conclude that geographic spread of strains may occur, but cannot explain the overall patterning of rotavirus epidemics in the U.S.

Transmission model details

We used a compartmental model to describe the dynamics of rotavirus infection using a series of differential equations (Fig. S8). The model structure is similar to that developed

for other imperfectly immunizing infections, e.g. (10), but is specifically tailored to the epidemiology of rotavirus. Briefly, individuals are born into compartment M , during which they are protected from infection by the presence of maternal antibodies. These maternal antibodies wane at a rate ω_0 , leaving the individual susceptible to primary infection (S_0), which occurs at a rate λ . Individuals with primary rotavirus infection (I_1) recover at a rate γ_1 into the R_1 state, during which they are temporarily immune to re-infection. This immunity wanes at a rate ω_1 and individuals enter the S_1 state, from which they can be re-infected at a reduced rate $\sigma_1\lambda$. Secondary infections (I_2) are assumed to have infectiousness reduced by a factor ρ_2 and a faster recovery rate γ_2 . After recovering from secondary rotavirus infection, individuals are again assumed to be temporarily immune (R_2), and this immunity wanes at a rate ω_1 , upon which individuals enter the partially-immune S_2 class. Once in S_2 , individuals can be re-infected at a rate $\sigma_2\lambda$, but subsequent infections (I_A) are assumed to be asymptomatic (or only mildly symptomatic) and have infectiousness that is reduced by a factor ρ_A . Recovery from asymptomatic infection is again assumed to occur at an increased rate γ_2 and lead to a state of temporary immunity (R_A) that wanes at a rate ω_2 .

Parameter estimates are given in Table S2. We based our estimates of the relative rate of infection and the probability of developing severe diarrhea given infection on a study by Velazquez et al (11), but the results from other cohort studies conducted in Argentina, Guinea Bissau, India, the U.S., and the U.K. (12-16) have for the most part been in agreement. Other studies (17, 18) suggest that there is a short-lived period of homotypic immunity following infection, which may be linked to rotavirus-specific T-cell responses (19). However, the role that serum antibodies versus cell-mediated immune responses play in protection against rotavirus infection is still not entirely understood (20). Studies of

household transmission of rotavirus stress the important role of children as sources of infection (21). Furthermore, there is evidence to suggest that viral loads tend to be higher in children with more severe disease (22), which may indicate greater infectiousness.

We included age structure in the model so as to keep track of the age-specific incidence of disease in accordance with the level of detail in our data (see below). The population was divided into 21 age classes, with 1 month age classes from 0-11 months of age, 1 year age classes from 1-4 years of age, then 5-9 year, 10-19 year, 20-39 year, 40-59 year, and 60+ year age groups. The rate of aging into the next age class was equal to $1/(\text{width of the age class})$. Individuals were born into the first age class at a rate $B(t)$ equal to the annual mean crude birth rate for each state in our dataset from 1998 to 2004 (23). The rate of aging out of the last age class was chosen to be consistent with the median age of individuals in the U.S. over this time period.

The force of infection is given by $\lambda = \beta(I_1 + \rho_2 I_2 + \rho_A I_A)$, where $\beta(t)$ is the seasonally-varying transmission parameter. Seasonality in the rate of transmission is modeled using a simple sinusoidal function, such that $\beta(t) = \beta_0(1 + a \cos(2\pi(t - \varphi)))$, where β_0 is the baseline transmission rate, a is the amplitude of seasonality, and φ is the seasonal offset parameter. Note that the transmission rate depends on the density of infectious individuals, which is typical of directly transmitted infections (24) and previously-developed models for rotavirus (25-27). Thus, the number of potentially infectious contacts will scale with the population size. We initially assumed that mixing is homogeneous with respect to age, such that all individuals have an equal probability of contacting any other individual in the population and of being infected given contact with an infectious individual. We estimated

the parameters of $\beta(t)$ by fitting the model output D (the number of cases of severe rotavirus-related diarrhea) to age-specific rotavirus hospitalization data. Individuals with primary rotavirus infection (I_1) have a probability d_1 of developing severe diarrhea, while those with secondary rotavirus infection (I_2) develop severe diarrhea with probability $d_2 < d_1$, where d_1 and d_2 are equal to 11% and 2.9%, respectively, consistent with data from Velazquez, et al (11). We assume that only a proportion hD of diarrhea cases are hospitalized, where h is estimated to be consistent with the best-fitting model.

The model fit was assessed by calculating the log-likelihood of the data under the assumption that the number of hospitalizations in each age class i at time t ($H_{i,t}$) follows a Poisson distribution with mean $hD_{i,t}$ as follows:

$$LL_{i,t} = -hD_{i,t} + H_{i,t} \ln(hD_{i,t}) - \sum_{j=1}^{H_{i,t}} \ln j$$

$$LL = \sum_t \sum_{i=1}^{16} LL_{i,t}$$

The best-fitting model was the one for which the set of estimated parameters $\{\beta_0, a, \varphi, h\}$ yielded the maximum log-likelihood. We fit the model after an initial burn-in period of 25 years to allow the model to reach equilibrium.

Rotavirus hospitalization data

We fit our model to age-specific hospitalization data available from the State Inpatient Databases (SID) of the Healthcare Cost and Utilization Project (HCUP) (<http://www.hcup-us.ahrq.gov/databases.jsp>) maintained by the Agency for Healthcare Research and Quality (AHRQ). This database includes all hospital discharge records from community hospitals

in participating states. HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of patient-level health care data (28). We extracted all records of hospital discharges that listed the specific code for rotavirus [008.61] from the International Classification of Diseases 9th revision (ICD9-CM), in any position among the up to 15 discharge diagnoses. For each record, we noted the child's age, the state, year and month of admission.

We analyzed a subset of all available SIDs from January 1993 to December 2004. The 16 states with monthly hospital discharge data for all study years were: Arizona (AZ), California (CA), Colorado (CO), Connecticut (CT), Iowa (IA), Illinois (IL), Kansas (KS), Massachusetts (MA), Maryland (MD), New Jersey (NJ), New York (NY), Oregon (OR), Pennsylvania (PA), South Carolina (SC), Washington (WA), and Wisconsin (WI). Age was recorded in monthly age categories from 0 to 11 months of age, then in 1 year age categories from 1 to 4 years of age. It is known that the use of the rotavirus ICD-9-CM discharge code increased during the period from 1993 to 1999 (29). Thus, we only focused on the period from July 1998 to December 2004.

Mixing assumptions

We explored four possible assumptions about how individuals belonging to different age classes mix with one another in ways relevant to rotavirus transmission. The simplest assumption is that mixing is homogeneous, such that individuals of all age groups have equal probability of contacting a member of any other age group and of becoming infected given an infectious contact. Since the primary means of transmission of rotavirus is fecal-

oral, we also explored the possibility that individuals <1 year of age, 1-year-olds, and 2-year-olds have unique (and presumably elevated) rates of transmission to them, since infants tend to be more inclined to stick their hands and other objects in their mouth and exhibit other behaviors that could lead to an increased rate of infection. Another possibility is that mixing is assortative, such that individuals are more likely to contact other individuals belonging to the same age class as opposed to those of a different age; again, we allowed for unique rates of contact among those <1 year, 1 year, and 2 years of age, as well as a unique rate for within-age-class mixing among older individuals. Finally, since there has been some speculation that airborne transmission of rotavirus may occur (30), we explored the possibility that age-related mixing is proportional to self-reported conversational data (31).

We fit our model to the age-specific hospitalization data from California (the state with the greatest number of cases) under each of the four mixing assumptions and estimated the relevant parameters, including the four baseline parameters $\{\beta_0, a, \varphi, h\}$, plus whatever additional age-specific parameters were necessary. We compared the model fits using Akaike information criterion (AIC).

We found that the model which allows for unique rates of transmission to <3 year olds provides the best fit to the data (Table S3). When we assume mixing is homogeneous, the model tends to overestimate the proportion of cases in the <1 year age classes while underestimating the proportion of cases in 1 year olds (Fig. S9). However, if we assume that there are age-related differences in contact rates, such that children <3 years of age

have 1.5-2.5 times as many infectious contacts compared to other individuals, then the age distribution of cases predicted by the model more closely resembles the data (Fig. S9).

After determining the most appropriate mixing assumption, we then fit this model to remaining 15 states, independently estimating four baseline parameters plus three age-specific contact rates for <3 year olds for each state. State-specific estimates for R_0 for primary infections (calculated as the maximum eigenvalue of the R_{ij} matrix), the amplitude of seasonality, seasonal offset parameter, and proportion of cases hospitalized and reported are presented in Fig. S10. We also calculated R_0 for the average infectious individual (\bar{R}_0) from the equilibrium distribution of infection using the following formula:

$$\bar{R}_0 = \beta_0(\gamma_1 I_1 + \rho_2 \gamma_2 I_2 + \rho_A \gamma_A I_A) / (I_1 + I_2 + I_A).$$

We then estimated the mean and 95% confidence interval for each parameter for the best-fitting model from each state for each of the four census regions and the U.S. as a whole (Table S2).

The variation in state-specific estimates of fitted parameters observed in Fig. S10 is possibly related to other demographic parameters such as population density and/or the number of children attending day care centers. Estimates of R_0 tended to be higher in more densely populated states; however, this relationship was not significant ($P>0.05$). There was a significant positive correlation between the estimated proportion of cases hospitalized in a given state and the number of child care centers per 1,000 children ($P=0.014$). Day care attendance has been found to be a significant risk factor for rotavirus-

associated diarrhea (32). It is possible that there is a dose effect leading to more severe rotavirus cases in children in states with high rates of day care attendance.

Relationship between birth rates and the timing of epidemics as predicted by the transmission model

We also had data on the number of weekly rotavirus-positive specimens voluntarily reported by laboratories throughout the U.S. to the Centers for Disease Control and Prevention (CDC) through the National Respiratory and Enteric Virus Surveillance System (NREVSS) (<http://www.cdc.gov/surveillance/nrevss/>) for the period from July (calendar week 27) 1991 to June (calendar week 26) 2006. Reporting for some states was poor and/or intermittent over the study period. Thus, we only included 23 states with “good” surveillance (defined as at least 40 positive specimens reported each season) in our analysis, as follows: Alabama, Arkansas, California, Delaware, Georgia, Illinois, Indiana, Kentucky, Louisiana, Massachusetts, Missouri, Nebraska, New York, North Dakota, Ohio, Oklahoma, South Carolina, Tennessee, Texas, Virginia, Washington, West Virginia, and Wisconsin. In some cases, multiple positive tests for the same patient are reported. A comparison of the weekly laboratory-confirmed detections to the monthly hospitalization rates in seven states for which we had both types of data suggests that the rate at which rotavirus cases are laboratory-confirmed and reported through NREVSS varies considerably by state, and that at most 6-75% of hospitalizations with an ICD-9-CM discharge code for rotavirus receive laboratory confirmation. .

When fitting our model to the NREVSS data, we assumed that R_0 and the age-specific rates of transmission varied by region in accordance with the mean values for the best-fitting models to the age-specific hospitalization data from 16 states (Fig. S10A). Thus, only three parameters were estimated for each state—the amplitude of seasonality (a), seasonal offset parameter (φ), and proportion of cases reported (i.e. lab-confirmed) (h). We first fit the model to the data from each state assuming that the birth rate, $B(t)$, was constant over the interval from 1991 to 2006 and equal to the mean crude birth rate for the U.S. during this time period (constant in both time and space). We then fit the model allowing the birth rate to vary both temporally and geographically in accordance with the yearly crude birth rates for each state. The log-likelihoods of the best-fitting models for each state were compared to determine whether allowing for geographical and temporal variation in the birth rate improved the model fit.

The model with constant birth rate was not able to capture the observed variability in epidemic timing as well (Fig. S11A), and led to a worse fit (lower log-likelihood) for 18 of the 23 states, and for the 23 states combined ($P < 0.001$). The state-to-state variance in the seasonal offset parameter (φ) of the best-fitting model was approximately the same for both models; accounting for variability in the birth rate in the model led to a positive correlation between the seasonal offset parameter and the mean state-specific birth rate ($\rho = 0.599$, $P = 0.0025$) rather than a negative correlation when a constant birth rate was used ($\rho = -0.609$, $P = 0.002$) (Fig. S11B).

We attempted to correlate the residual variability in seasonality between states, including variation in the seasonal amplitude (a) and offset parameter (φ), to variability in other

demographic or environmental variables. Birth rates in the U.S. tend to vary seasonal with a small (2-7%) amplitude and peak in the late summer/early fall. However, the residual variability in seasonality does not appear to be related to state-to-state variation in birth rate seasonality. Furthermore, such variation in the birth rate is not comparable to similar amplitudes of seasonal forcing estimated for the transmission parameter (i.e. the observed variation in birth rate not sufficient to drive the strong epidemic patterns observed), as has been previously noted (33, 34). Population density may have some effect on the variation in seasonality parameters, but the correlations tended to be weak and non-significant. There were no significant correlations between any of the climatic variables we examined and the residual variability in seasonality. Thus, while factors such as increased virus survival in low relative humidity may help explain why transmissibility tends to peak in the winter (35), environmental factors do not explain the residual seasonality in our model. This small residual seasonality may not of itself be biologically meaningful; instead it may simply reflect the fact that we are fitting a necessarily approximate, nonlinear model to real data.

Relationship between birth rate and rotavirus dynamics in other countries

Qualitatively similar relationships between rotavirus seasonality and secular changes in birth rate have been observed in other countries, as well. For example, the incidence of rotavirus in the Basque region of Spain went from having a relatively high incidence of rotavirus cases year-round during the period from 1984-90 to strongly seasonal winter outbreaks in 1991-97; the authors hypothesized that this could be linked to a large decline in the birth rate during the preceding decade (36), and our model is able to reproduce this general pattern if we assume transmission rates are somewhat higher than in the U.S.

Rotavirus epidemics also occurred increasingly later in Japan throughout the 1990s (37), while birth rates declined considerably between 1975 and 1985 in this country (38). Thus, while the epidemiology of rotavirus may vary slightly between countries, demography may be a universally important driver of patterns in rotavirus incidence.

Vaccination impact

We modeled the effect of vaccination by assuming vaccinated individuals in the M and S_0 compartments who are successfully immunized bypass the I_1 compartment and enter the R_1 state. Studies have shown that vaccine efficacy is comparable in breastfed infants, compared to non-breastfed infants (39), so we assume vaccinated infants can be successfully immunized prior to the waning of maternal antibodies. We assume that 96% of vaccinated individuals are successfully immunized, consistent with the proportion of individuals who seroconverted following two doses of the Rotarix vaccine in phase II clinical trials (40). This yields a vaccine efficacy of 36.5% ($=0.96*(1-0.62)$) against infection and 80.3% ($=0.96*(1-0.62*0.029/0.11)$) against severe diarrhea, similar to that estimated for the Rotarix vaccine during two years follow-up in Latin American infants (41). Efficacy estimates have been slightly higher for the Rotateq vaccine (42), which is the only one currently licensed in the U.S., but the differences may be explained by the use of different classifications of disease severity and differences in the populations studies in vaccine trials (43). Since our model does not explicitly incorporate information on the different serotypes of rotavirus, it seems more appropriate to model the impact of a human-

derived monovalent vaccine such as Rotarix rather than a reassortant pentavalent vaccine such as Rotateq.

We validated our model against the weekly NREVSS data from 2006-2008; *note that we did not fit our model to the vaccine era data.* We examined the effect of vaccination administered at 3 months of age. Estimates of 1-dose vaccination coverage were derived from health insurance claims data gathered by Surveillance Data Inc. (SDI) (www.sdihealth.com). The number of Rotateq doses administered during the first 3 months of life was used as an estimate of the numerator, while denominators were estimated as the total number of physician visits in which any standard childhood vaccine was administered prior to 4 months of age. A similar method was shown to produce reliable estimates of vaccination coverage using health insurance data from Germany (44). Furthermore, estimates showed good agreement with coverage estimates from 8 immunization information system (IIS) sentinel sites and the Vaccine Safety Datalink (VSD) (45). Both the relative efficacy of 1 dose of the vaccine and the proportion of children who go on to receive the full dose are both somewhat uncertain; thus, we explored estimates of the relative effectiveness of the 1-dose coverage estimates (compared to the full 3-dose protective effect) between 50-100%. We found that our model predictions were in close agreement with the data given a relative effectiveness of 70-80%. This seems reasonable, given that current estimates of 1-dose efficacy have varied between 29-83% (46), and ~50% of individuals go on to receive ≥ 3 doses of the vaccine (45).

Since vaccination primarily protects against the development of symptoms given infection rather than infection itself, it is not obvious that vaccination should affect the timing of

epidemics. If the force of infection were mostly dependent on the large number of asymptomatic infections in the population, one would expect to see smaller epidemics occurring around the same time of year in response to vaccination. However, by preventing the more infectious primary infections that tend to drive the epidemics, vaccination has non-linear effects on both the size and timing of epidemics. Thus, the dynamics that drive the observed response to vaccination are the same that lead to the association between birth rates and the spatiotemporal pattern of epidemics.

Vaccination provides both direct protection for those who are successfully immunized with the vaccine and indirect protection (also known as herd immunity) for those who are not immunized, by decreasing the likelihood that they will come into contact with an infectious individual. To explore the relative contributions of direct and indirect protection in determining the long-term impact of rotavirus vaccination, we assumed that a proportion $c \in [0,1]$ of all individuals are fully immunized by the vaccine at a certain age. The recommended protocol is to vaccinate at 2 and 4 months of age for the two-dose Rotarix vaccine and at 2, 4, and 6 months of age for the three-dose Rotateq vaccine, although both vaccines may be administered up to 8 months of age; neither vaccine is not recommended for children over the age of 8 months, as the risk of intussusception associated with vaccination has not been assessed in older infants (47). High levels of maternal antibodies may interfere with the vaccine in individuals <2 months of age, despite evidence which suggests that breastfeeding does not interfere with vaccination under the current vaccine schedule (39). Furthermore, multiple doses of the vaccine may be required to achieve the full immunological effect (46). Thus, we examined three possible ages at immunization—

(1) a “best-case” scenario in which individuals are immunized at birth (i.e. upon first dose of the vaccine, prior to the waning of maternal immunity), (2) a “realistic” scenario in which a proportion c of susceptible (M and S_0) individuals are immunized at 4 months of age (i.e. following 2 doses of the vaccine according to the recommended schedule), and (2) a “pessimistic” scenario in which a proportion c are fully immunized at 8 months of age (i.e. following last dose of the vaccine at the oldest recommended age).

To estimate the degree of indirect protection afforded by vaccination, we compared the reduction in both the incidence in severe diarrhea and the overall prevalence of infection predicted by the model 10 years after the vaccine is introduced at coverage c to the direct effect of vaccination. The direct effect of vaccination is estimated as $1-0.803*c$ for the incidence of severe diarrhea and $1-0.365*c$ for the prevalence of infection. This calculation assumes all individuals who receive the vaccine are protected from birth.

Effect of Repeat Vaccination

We explored the potential effect of a booster vaccine that could mimic secondary infection in previously vaccinated (or infected) infants. Alternatively, if secondary infections are primarily caused by a different serotype, this could represent full protection against multiple serotypes. To model this effect, we assume that individuals in the R_1 and S_1 compartments who receive a second dose of the vaccine by-pass the I_2 state and enter the R_2 compartment. We assume that the booster dose is administered 2 months after the initial age at immunization, and that coverage is equal to that with the first dose. Again, we assume that 96% of individuals who receive the booster vaccine are effectively immunized. Thus, the vaccine efficacy is 59.9% against infection ($=0.96*0.96*(1-0.35)$) and 92.2%

against severe diarrhea ($=0.96*0.96*(1-0.35*0/0.11)$). The results suggest that such a vaccine could lead to the elimination of rotavirus from the population, but only at very high levels of coverage ($\geq 99\%$ when the first dose is administered at birth) (Fig. S12).

Sensitivity analyses

We examined the sensitivity of our results to the initial parameter estimates for four parameters about which there was some uncertainty, including the duration of infectiousness ($1/\gamma_1$ and $1/\gamma_2$), the relative infectiousness of secondary and subsequent infections (ρ_2 and ρ_A), the duration of complete immunity following infection ($1/\omega_1$ and $1/\omega_2$), and the possible waning of partial immunity in adults (in which we assume individuals move from the S_2 to the S_1 class at a rate ψ). For each of the four parameters, we re-fit our model (again assuming unique rates of transmission to <3 year olds) to the age-specific hospitalization data from the California HCUP SID under two different sets of parameter values; generally, one set of parameters was less than the initial estimate while the other was greater, except for the duration of infectiousness and the rate of waning partial immunity, for which we had reason to believe that our initial parameter values are on the lower end of the reasonable parameter range. We compared the model fits to our original model by comparing the negative log-likelihoods (with lower negative log-likelihoods indicating a better fit to the data). In all cases, our original model yielded the best fit (Table S4).

We then analyzed the maximum possible impact of vaccination for each of the parameter sets by assuming a coverage c of 100% at birth and examining the equilibrium prevalence of infection and incidence of severe diarrhea. Overall, the relative incidence of severe diarrhea given 100% vaccine coverage varied between 0 (when we assume the relative infectiousness of secondary and subsequent infections are 0.1 and 0.05, respectively) and 0.23 (when we assume secondary and subsequent infections are as infectious as primary infections) (Table S4). The relative prevalence of infection at equilibrium varied between 0 and 0.73, and was again most sensitive to our assumptions about the relative infectiousness of secondary and subsequent infections (Table S4). Note that the parameter values that predicted the highest relative rates of severe diarrhea (and thus the smallest impact of vaccination) also tended to be the ones that provided the worst fit to the age-specific hospitalization data. However, the models that included waning of partial immunity among adults, for which the fit to the data was only slightly worse, also predicted slightly elevated rates of infection and severe diarrhea at equilibrium.

Thus, the potential population-level impact of rotavirus vaccination we present in the main text is relatively robust to our parameterization of the model. However, predicting the true impact of vaccination may require a better understanding of the relative infectiousness and nature of immunity to rotavirus infection in adults.

There is some evidence to suggest that further refinement of our model may be necessary. For instance, an examination of laboratory-confirmed reports of rotavirus from Utah, which was excluded from our analysis since detections were only reported for the first 10 years of the study period, reveals that epidemics occurred later in this state beginning in 1996-1997

despite the fact that birth rates remained consistently high over the past two decades. Thus, it may be necessary to account for some spatial spread of infection from more populous regions or other factors to explain this pattern. Furthermore, our model suggests that a decline in birth rate should not only affect the timing of epidemics, but will also lead to a decrease in the mean incidence of infection over time. While there has been some year-to-year variability in the incidence of rotavirus, there is no evidence to suggest that rates of hospitalization and detection are declining along with birth rates in states such as California prior to the introduction of the vaccine. Nevertheless, the current formulation is able to capture many of the relevant features of the epidemiology of rotavirus in the U.S.A using a relatively simplified representation of the dynamics of infection.

Acknowledgments

We would like to thanks Steve Wilcox from the National Renewable Energy Laboratory for his assistance and helpful advices in finding and dealing with climatic datasets

Supplementary References

1. P. W. Stackhouse *et al.*, “Satellite based assessment of the NSRDB site irradiances and time series from NASA and suny/albany algorithms” (2008).
2. T. D. Mitchell, P. D. Jones, *International Journal of Climatology* **25**, 693 (May, 2005).
3. C. Viboud *et al.*, *Science* **312**, 447 (Apr 21, 2006).
4. M. I. Nelson, L. Simonsen, C. Viboud, M. A. Miller, E. C. Holmes, *PLoS Pathog* **3**, 1220 (Sep 14, 2007).
5. H. Nakajima *et al.*, *Lancet* **357**, 1950 (Jun 16, 2001).
6. M. J. Cox, G. F. Medley, *Epidemiol Infect* **131**, 719 (Aug, 2003).
7. R. F. Bishop, P. J. Masendycz, H. C. Bugg, J. B. Carlin, G. L. Barnes, *J Clin Microbiol* **39**, 1085 (Mar, 2001).
8. D. D. Griffin *et al.*, *J Clin Microbiol* **38**, 2784 (Jul, 2000).
9. D. O. Matson *et al.*, *J Infect Dis* **162**, 605 (Sep, 1990).
10. L. J. White *et al.*, *Math Biosci* **209**, 222 (Sep, 2007).
11. F. R. Velazquez *et al.*, *N Engl J Med* **335**, 1022 (Oct 3, 1996).
12. M. K. Bhan *et al.*, *J Infect Dis* **168**, 282 (Aug, 1993).
13. T. K. Fischer *et al.*, *J Infect Dis* **186**, 593 (Sep 1, 2002).
14. S. Grinstein, J. A. Gomez, J. A. Bercovich, E. L. Biscotti, *Am J Epidemiol* **130**, 300 (Aug, 1989).
15. L. H. Moulton, M. A. Staat, M. Santosham, R. L. Ward, *J Infect Dis* **178**, 1562 (Dec, 1998).
16. L. J. White, J. Buttery, B. Cooper, D. J. Nokes, G. F. Medley, *J R Soc Interface* (May 13, 2008).
17. D. I. Bernstein, D. S. Sander, V. E. Smith, G. M. Schiff, R. L. Ward, *J Infect Dis* **164**, 277 (Aug, 1991).
18. S. Chiba *et al.*, *Lancet* **2**, 417 (Aug 23, 1986).
19. M. Makela, J. Marttila, O. Simell, J. Ilonen, *Clin Exp Immunol* **137**, 173 (Jul, 2004).
20. B. Jiang, J. R. Gentsch, R. I. Glass, *Clin Infect Dis* **34**, 1351 (May 15, 2002).
21. J. S. Koopman, A. S. Monto, I. M. Longini, Jr., *Am J Epidemiol* **130**, 760 (Oct, 1989).
22. G. Kang *et al.*, *J Med Virol* **73**, 118 (May, 2004).
23. B. E. Hamilton, P. D. Sutton, S. J. Ventura, “Revised Birth and Fertility Rates for the 1990s and New Rates for Hispanic Populations, 2000 and 2001: United States” (National Center for Health Statistics, 2003).
24. M. C. M. De Jong, A. Bouma, O. Diekmann, H. Heesterbeek, *Trends in Ecology and Evolution* **17**, 64 (2002).
25. M. V. Jose, J. R. Bobadilla, *Vaccine* **12**, 109 (Feb, 1994).
26. E. Shim, Z. Feng, M. Martcheva, C. Castillo-Chavez, *J Math Biol* **53**, 719 (Oct, 2006).
27. L. J. White, M. J. Cox, G. F. Medley, *IMA J Math Appl Med Biol* **15**, 211 (Sep, 1998).
28. C. Steiner, A. Elixhauser, J. Schnaier, *Eff Clin Pract* **5**, 143 (May-Jun, 2002).
29. T. K. Fischer *et al.*, *J Infect Dis* **195**, 1117 (Apr 15, 2007).

30. P. H. Dennehy, *Pediatr Infect Dis J* **19**, S103 (Oct, 2000).
31. J. Wallinga, P. Teunis, M. Kretzschmar, *Am J Epidemiol* **164**, 936 (Nov 15, 2006).
32. R. R. Reves *et al.*, *Am J Epidemiol* **137**, 97 (Jan 1, 1993).
33. S. Altizer *et al.*, *Ecol Lett* **9**, 467 (Apr, 2006).
34. D. He, D. J. Earn, *Theor Popul Biol* **72**, 274 (Sep, 2007).
35. S. A. Ansari, V. S. Springthorpe, S. A. Sattar, *Rev Infect Dis* **13**, 448 (May-Jun, 1991).
36. G. Cilla, E. Perez-Trallero, M. C. Lopez-Lopategui, A. Gilsetas, M. Gomariz, *Epidemiol Infect* **125**, 677 (Dec, 2000).
37. H. Suzuki, T. Sakai, N. Tanabe, N. Okabe, *Pediatr Infect Dis J* **24**, 257 (Mar, 2005).
38. in *Statistical Handbook of Japan 2008*. (Statistics Bureau, Ministry of Internal Affairs and Communications, Japan), vol. 2008.
39. M. G. Goveia, M. J. Dinubile, M. J. Dallas, P. M. Heaton, B. J. Kuter, *Pediatr Infect Dis J* **27**, 656 (Jul, 2008).
40. T. Vesikari *et al.*, *Vaccine* **22**, 2836 (Jul 29, 2004).
41. A. C. Linhares *et al.*, *Lancet* **371**, 1181 (Apr 5, 2008).
42. T. Vesikari *et al.*, *N Engl J Med* **354**, 23 (Jan 5, 2006).
43. R. I. Glass, U. D. Parashar, *N Engl J Med* **354**, 75 (Jan 5, 2006).
44. H. Kalies, R. Redel, R. Varga, M. Tauscher, R. von Kries, *BMC Public Health* **8**, 82 (2008).
45. *MMWR Morb Mortal Wkly Rep* **57**, 398 (Apr 18, 2008).
46. D. Bernstein, *Pediatrics* **121**, 223; author reply 223 (Jan, 2008).
47. R. I. Glass *et al.*, *Lancet* **368**, 323 (Jul 22, 2006).
48. A. C. Linhares *et al.*, *Epidemiol Infect* **102**, 129 (Feb, 1989).
49. R. L. Ward *et al.*, *J Infect Dis* **161**, 440 (Mar, 1990).
50. J. Wilde, R. Yolken, R. Willoughby, J. Eiden, *Lancet* **337**, 323 (Feb 9, 1991).
51. E. J. Anderson, S. G. Weber, *Lancet Infect Dis* **4**, 91 (Feb, 2004).
52. R. L. Ward *et al.*, *J Infect Dis* **154**, 871 (Nov, 1986).

Figure S1: Map of average amount of solar radiation in December- January, based on ground observations. (A) 1995-97, (B) 2000-02, (C) difference between the two periods.

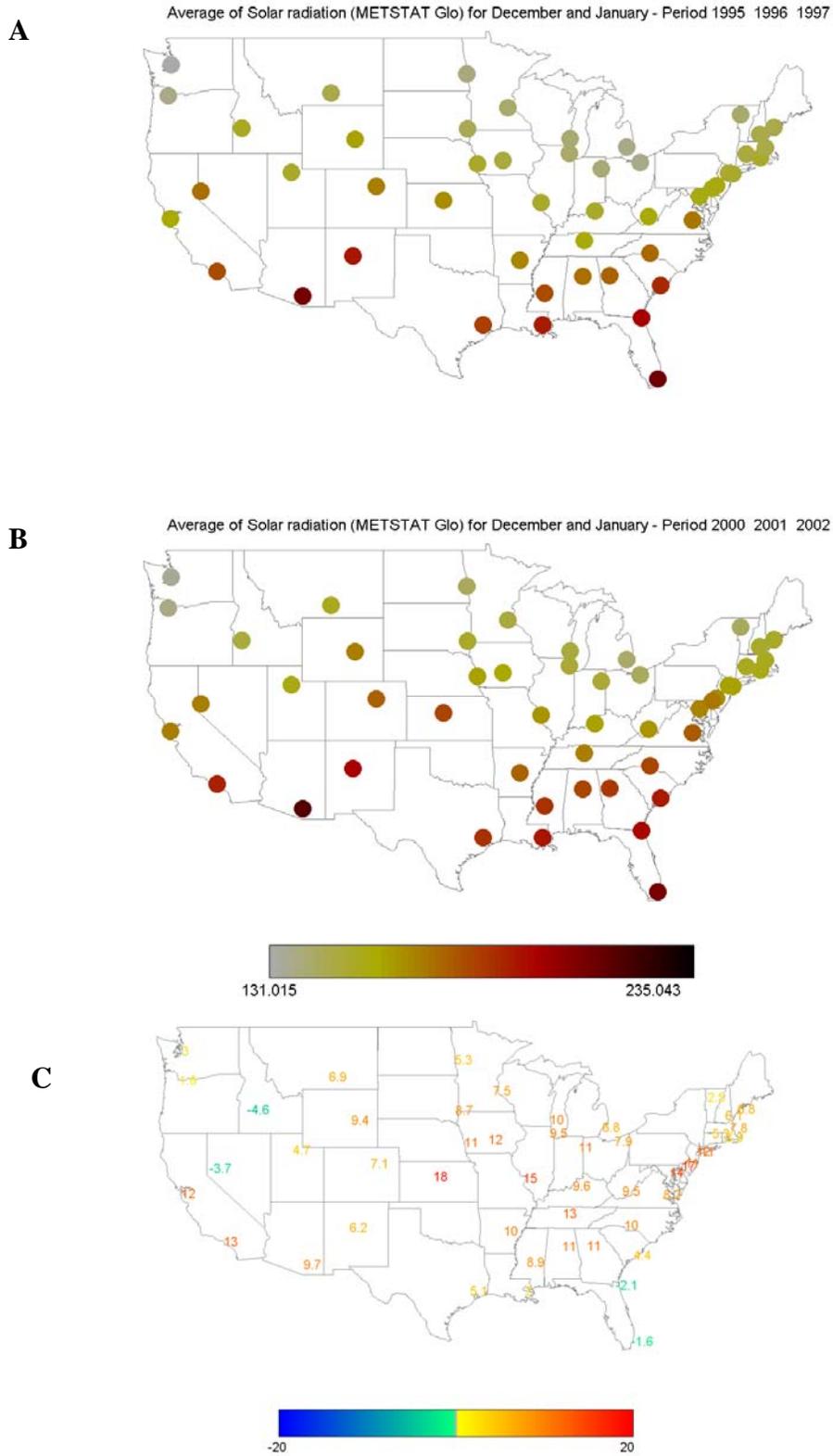


Figure S2: Map of average amount of solar radiation in December- January, based on satellite observations. (A) 1995-97, (B) 2000-02, (C) difference between the two periods.

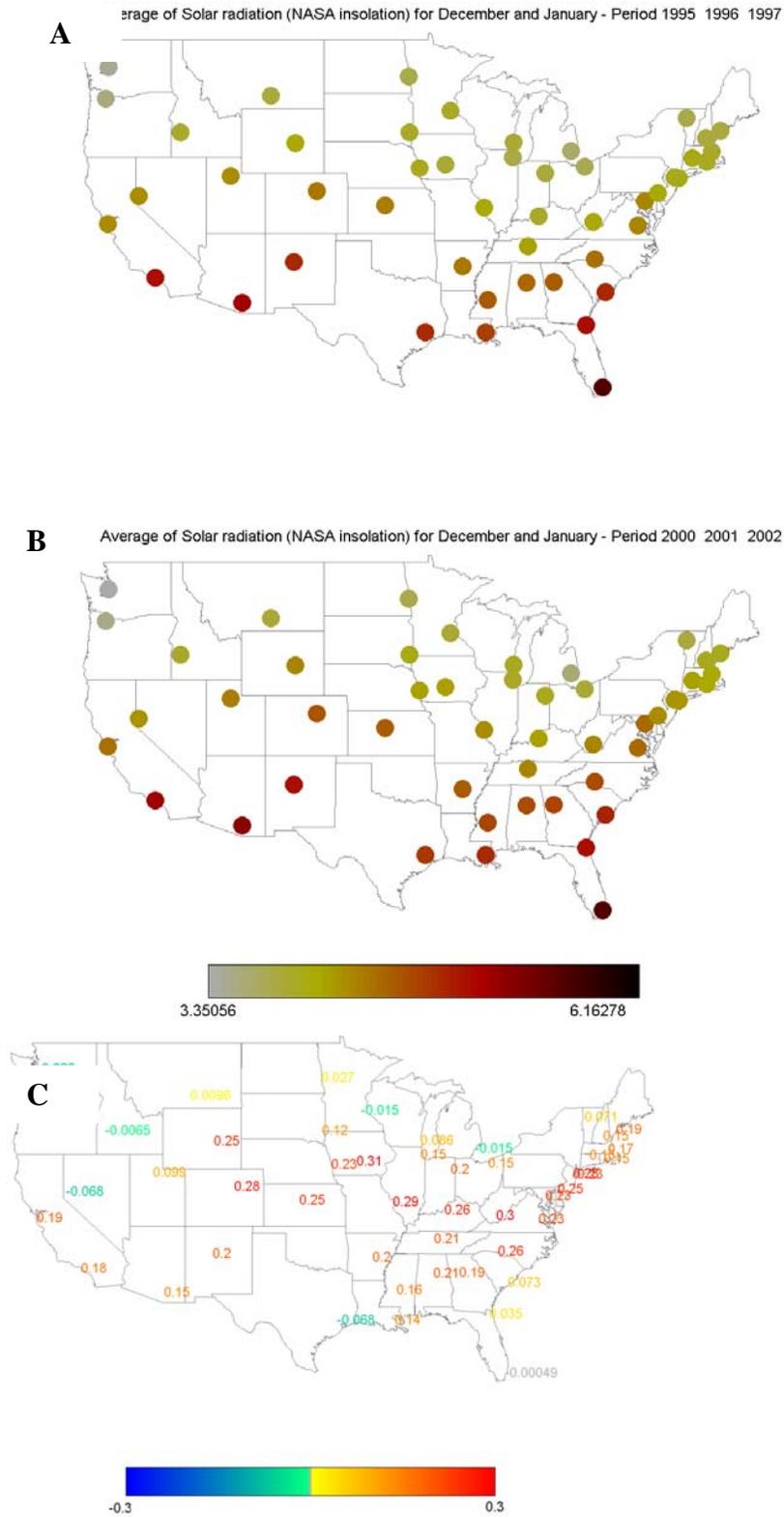


Figure S3: Map of average amount of precipitation in December-January. (A) 1995-97, (B) 2000-02, (C) difference between the two periods.

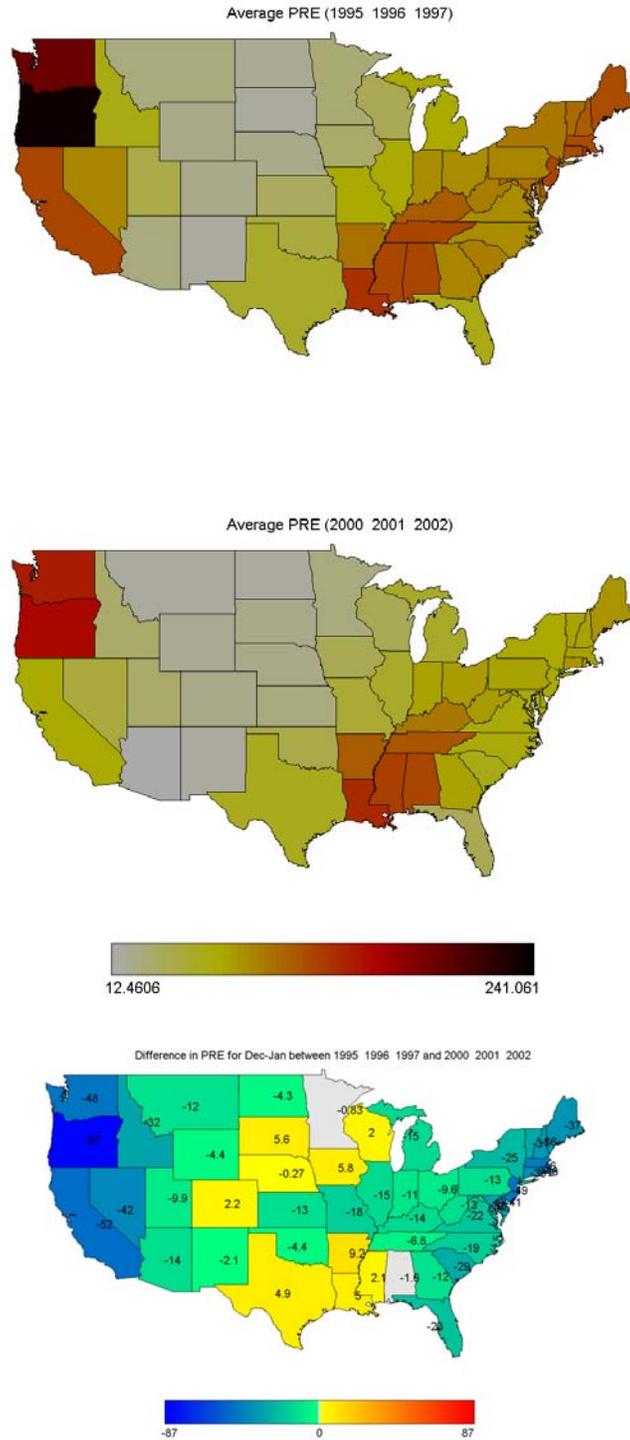


Figure S4: Map of average minimum temperature in December-January. (A) 1995-97, (B) 2000-02, (C) difference between the two periods.

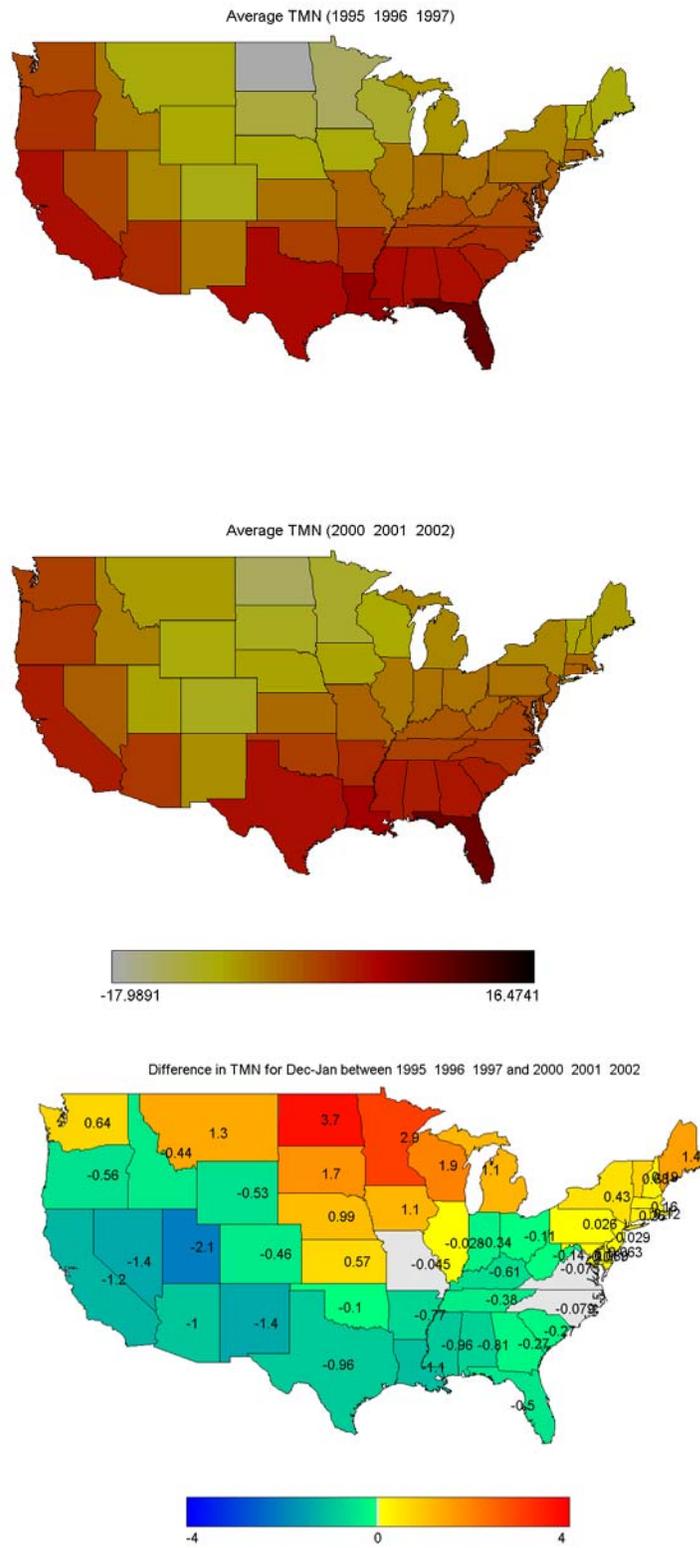


Figure S5: Map of average maximum temperature in December-January. (A) 1995-97, (B) 2000-02, (C) difference between the two periods.

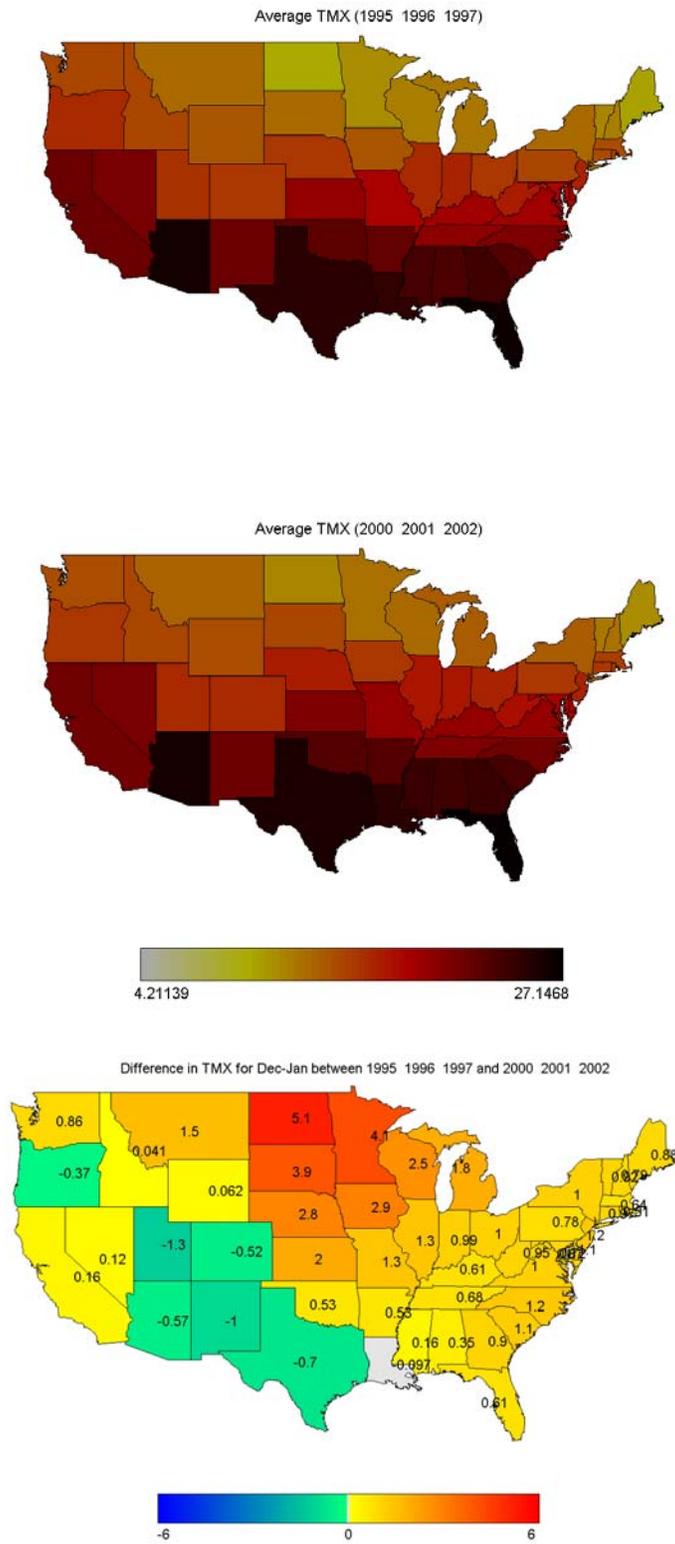


Figure S6: Map of average amount of vapor pressure in December-January. (A) 1995-97, (B) 2000-02, (C) difference between the two periods.

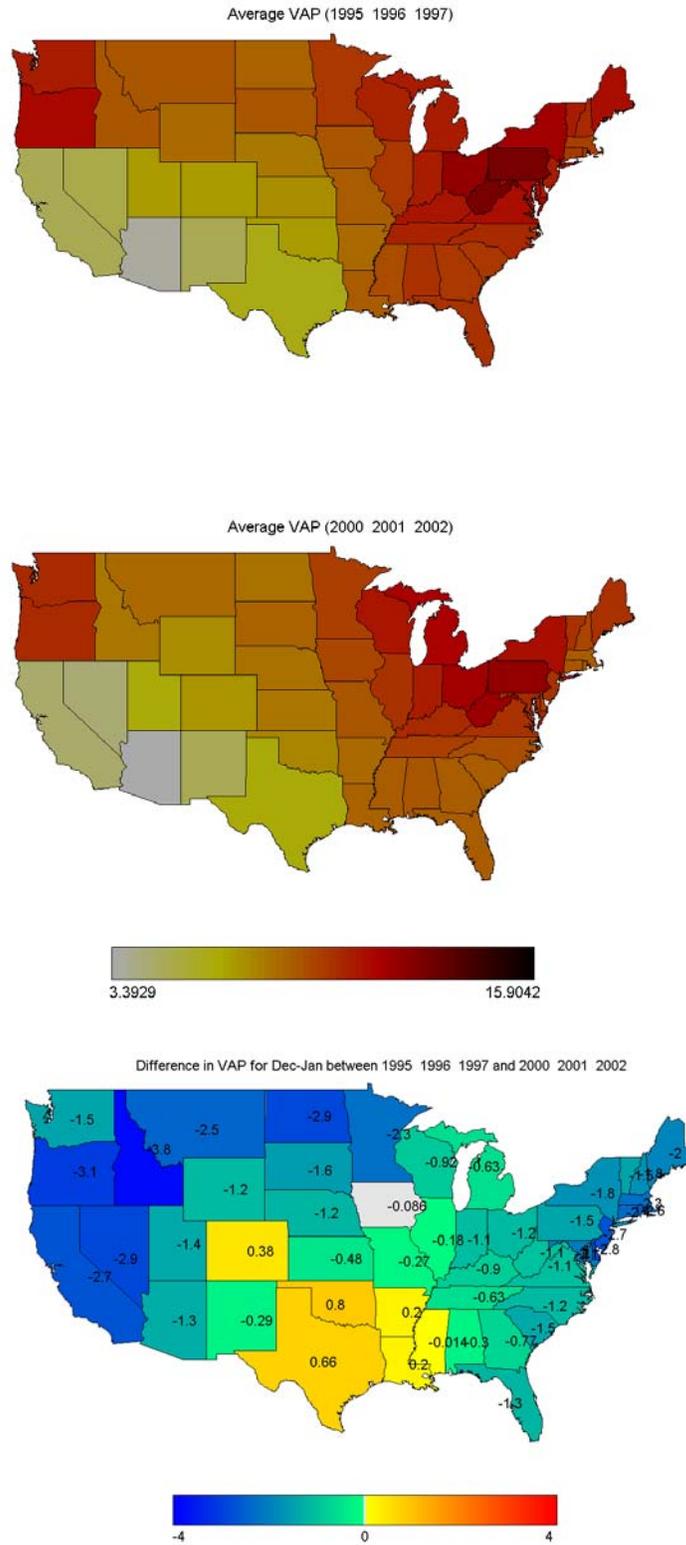


Figure S7: Maps of month of minimum climate indicators for 1995-97 (left panels) and 2000-02 (right panels). From top to bottom: precipitation (PRE), average daily maximum temperature (TMX), average daily minimum temperature (TMN), vapor pressure (VAP).

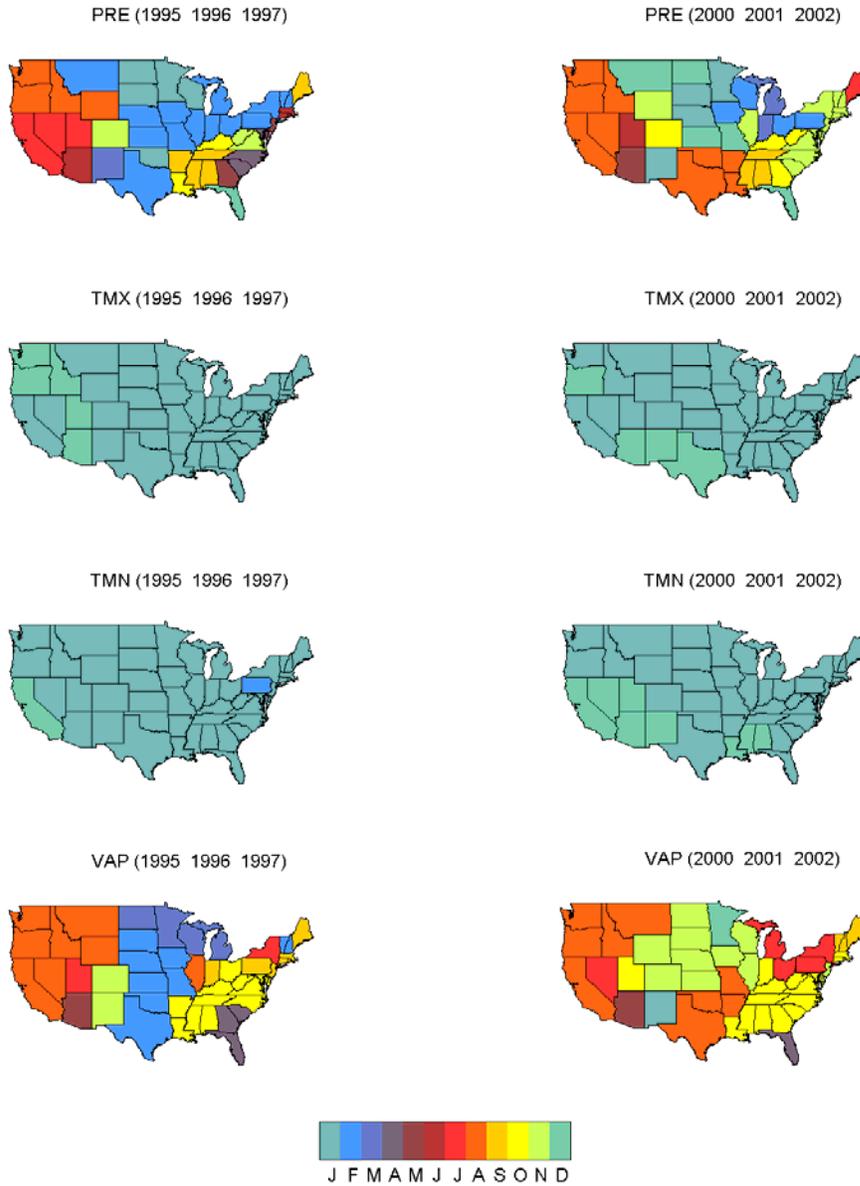


Figure S8: Detailed compartmental diagram illustrating transmission model. Parameter values are defined in Table S2.

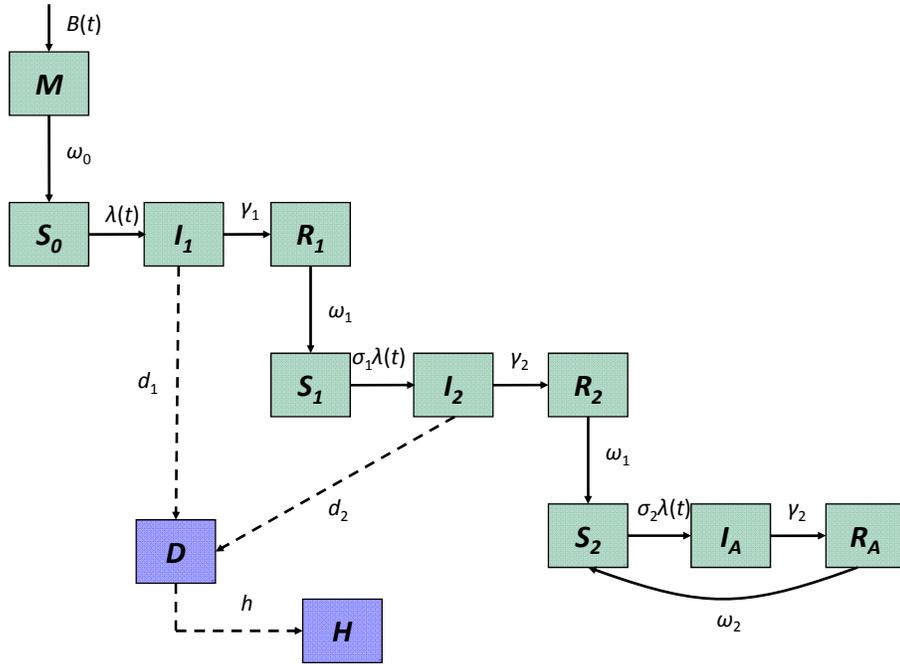


Figure S9: Age distribution of severe diarrhea cases predicted by the model assuming homogeneous mixing (blue) versus unique rates of transmission to <1, 1, and 2-year olds (green), compared to the age distribution of hospitalized cases from the California HCUP SID (red).

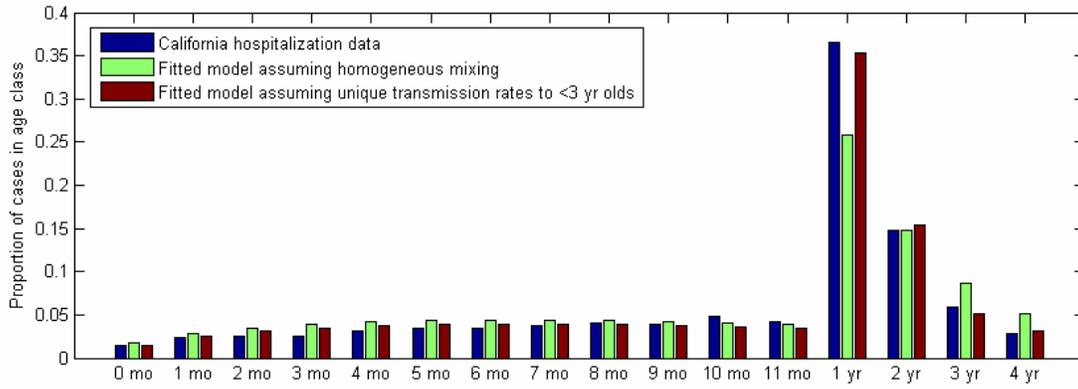


Figure S10: Estimated parameters for model fit to age-specific hospitalization data from HCUP SIDs for 16 states. (A) R_0 for primary infection, (B) amplitude of seasonality (a), (C) seasonal offset parameter (ϕ), and (D) proportion of severe diarrhea cases hospitalized and reported (h). Data is color-coded by U.S. census region (West=red, Midwest=blue, South=purple, Northeast=green).

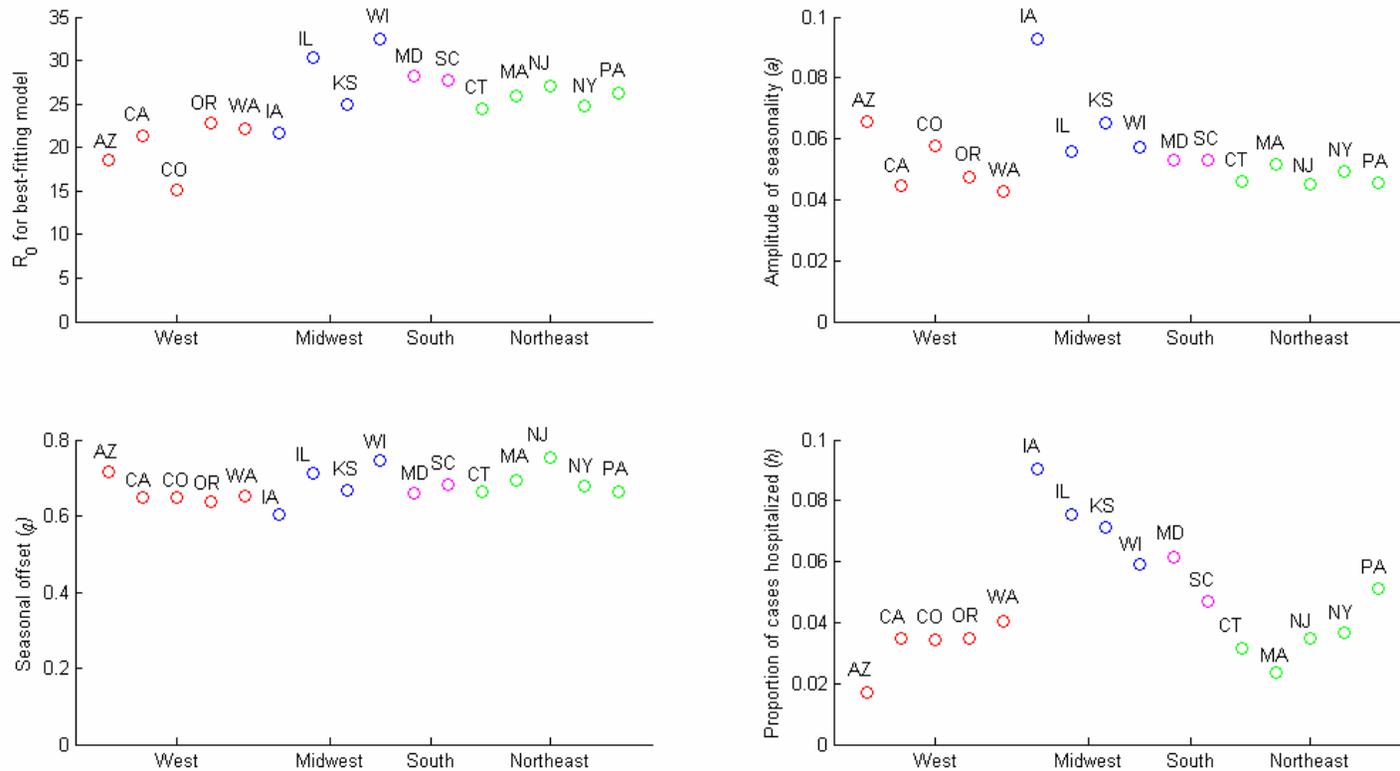


Figure S11: Comparison of model with constant birth rate to model incorporating state-specific estimates of annual birth rate. (A) Relationship between mean timing of rotavirus activity and birth rate in the preceding year. NREVSS data for 23 states from 1991 to 2006 is plotted in blue, while the red dots represent data for the fitted model with variable birth rate and the green dots represent data for the fitted model with constant birth rate. (B) Estimates of seasonal offset parameter for best-fitting model to data from 23 states, assuming variable birth rate (red) and constant birth rate (green).

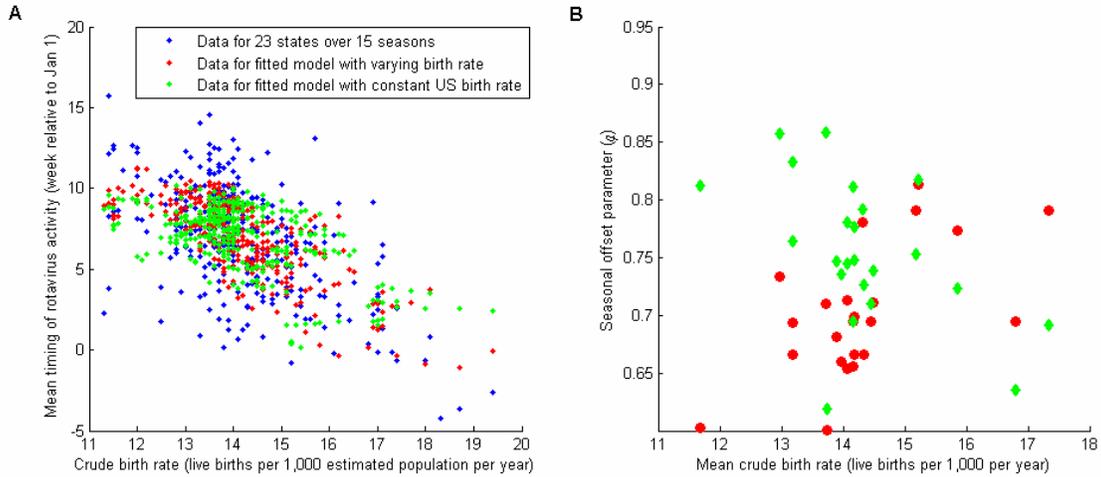


Figure S12: Effect of repeat vaccination at increasing levels of vaccine coverage on (A) the relative incidence of severe diarrhea and (B) the relative prevalence of infection after vaccination. Age at first immunization is at birth (blue), 4 months (green), or 8 months (red), while a booster dose is administered 2 months later. The dotted black line represents the direct effect of vaccination assuming full immunization at birth.

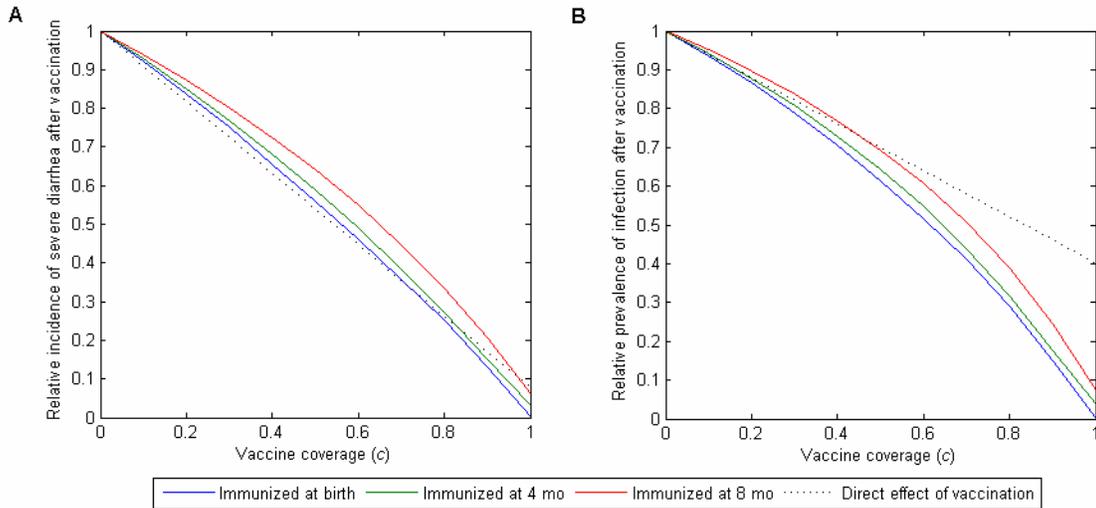


Table S1: Statistical association between spatial and temporal patterns in rotavirus epidemic timing and climate indicators and birth rates, U.S. states, 1993-2006. Results indicate regression coefficients and standard errors (se) of univariate and multivariate linear models, where the mean timing of rotavirus activity is the outcome and climate and birth rate indicators are explanatory covariates. Multivariate models are based on stepwise selection procedures.

Covariate	Univariate analysis : regression coefficient (se) §	Multivariate analysis: regression coefficient (se) §
Spatial patterns in early years: 1993-98 (n=11 states)		
Birth rate	-2.05 (0.43) ***	-2.05 (0.43) ***
Solar radiation	-0.07 (0.04)	
Precipitation	-0.006 (0.03)	
Min temperature	-0.23 (0.21)	
Max temperature	-0.28 (0.16)	
Vapor pressure	0.46 (0.31)	
Spatial patterns in recent years: 2001-06 (n=32 states)		
Birth rate	-1.24 (0.16) ****	-0.98 (0.19) ****
Solar radiation	-0.04 (0.02) *	
Precipitation	0.03 (0.01) *	
Min temperature	-0.02 (0.09)	
Max temperature	-0.12 (0.08)	
Vapor pressure	0.55 (0.11) ****	0.25 (0.10) *
Temporal trends, 1993-2006 (n=10 states)		
Birth rate	-1.93 (0.54) **	-1.93 (0.54) **
Solar radiation	-0.11 (0.30)	
Precipitation	-0.06 (0.03)	
Min temperature	-1.13 (0.80)	
Max temperature	-0.81 (0.69)	
Vapor pressure	-1.02 (0.68)	

§ Stars highlight regression coefficients that have significant P-values (* : <0.05; **:<0.01; ***<0.001,**** <0.0001)

Table S2: Model parameter values.

Parameter		Value	Source
Estimated from data			
ω_0	rate of waning maternal antibodies	0.333 mo ⁻¹	(48)
ω_1	rate of waning immunity following primary and secondary infection	0.111 mo ⁻¹	(18)
ω_2	rate of waning immunity following asymptomatic infection	0.083 mo ⁻¹	(49)
γ_1	rate of recovery from primary infection	4.3 mo ⁻¹	(50)
γ_2	rate of recovery from secondary and asymptomatic infection	8.6 mo ⁻¹	(49, 51, 52)
σ_1	relative susceptibility following first infection	0.62	(11)
σ_2	relative susceptibility following second infection	0.35	(11)
ρ_2	relative infectiousness of secondary infection	0.5	Assumption, (21)
ρ_A	relative infectiousness of asymptomatic infection	0.1	Assumption, (21)
d_1	proportion of first infections with severe diarrhea	0.11	(11)
d_2	proportion of second infections with severe diarrhea	0.029	(11)
Estimated from model fit to age-specific hospitalization data from HCUP for 16 states			
$R_0 (= \beta_0 \gamma)$	basic reproductive number for primary infection	24.6 (22.3, 26.9)	
	for average infectious individual	3.72 (3.43, 4.02)	
a	amplitude of seasonality	0.055 (0.048, 0.061)	
φ	seasonal offset parameter	0.68 (0.66, 0.70)	
h	proportion of severe diarrhea cases hospitalized	0.047 (0.036, 0.057)	

Table S3: Fit of model to age-specific hospitalization data from the California HCUP SID under various population mixing assumptions

Mixing assumption	Additional estimated parameters	R₀ (primary, average infection)	AIC
Homogeneous	--	29.9, 4.76	8766.3
Higher rates of transmission to 0, 1, and 2 year olds	$c_0 = 1.58$ $c_1 = 2.34$ $c_2 = 1.77$	21.3, 3.44	7728.3
Assortative with unique rates of transmission among 0, 1, and 2 year olds	$c_0 = 0.31$ $c_1 = 3.47$ $c_2 = 1.89$ $c_A = 1.06$	27.6, 4.60	8230.3
Mixing proportional to self-reported conversational data	--	12.5, 3.99	8750.9

Table S4: Sensitivity analysis of model parameters (fit to California HCUP SID age-specific hospitalization data)

Parameter	Value	Negative log-likelihood	Relative incidence of severe diarrhea given 100% vaccination coverage	Relative prevalence of infection given 100% vaccination coverage
Duration of infectiousness	$1/\gamma_1 = 1 \text{ wk}, 1/\gamma_2 = 0.5 \text{ wk}$	3857.1	0.107	0.209
	$1/\gamma_1 = 2 \text{ wk}, 1/\gamma_2 = 1 \text{ wk}$	4269.1	0.197	0.410
	$1/\gamma_1 = 4 \text{ wk}, 1/\gamma_2 = 2 \text{ wk}$	4366.1	0.219	0.459
Relative infectiousness of secondary and subsequent infections	$\rho_2 = 0.5, \rho_A = 0.1$	3857.1	0.107	0.209
	$\rho_2 = 1, \rho_A = 1$	4175.5	0.227	0.730
	$\rho_2 = 0.1, \rho_A = 0.05$	4308.4	0	0
Duration of complete immunity	$1/\omega_1 = 9 \text{ mo}, 1/\omega_2 = 12 \text{ mo}$	3857.1	0.107	0.209
	$1/\omega_1 = 1 \text{ mo}, 1/\omega_2 = 1 \text{ mo}$	4077.3	0.149	0.203
	$1/\omega_1 = 18 \text{ mo}, 1/\omega_2 = 24 \text{ mo}$	3966.3	0.100	0.246
Waning of partial immunity	$\psi = 0$ (no waning)	3857.1	0.107	0.209
	$\psi = 1/10 \text{ yr}^{-1}$	3896.0	0.191	0.494
	$\psi = 1/5 \text{ yr}^{-1}$	3912.4	0.209	0.570