Digital Epidemiology Reveals Global Childhood Disease Seasonality and the Effects of Immunization: Supporting Information

Kevin M Bakker[a,b], Micaela Martinez-Bakker[c], Barbara Helm[d] and Tyler J Stevenson[e]

[a]Department of Ecology and Evolutionary Biology and
[b]Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109
[c]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, 08544
[d]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Graham Kerr Building, Glasgow G12 8QQ, United Kingdom
[e]Institute for Biological and Environmental Sciences, University of Aberdeen, Aberdeen, United Kingdom, AB24 2TZ

May 16, 2016

Supporting Information

Methods and Materials

Acquisition and Analyses of Google Trends Data

Google Trends is a publically available data service provided by Google Inc [1] that allows internet users to view and download global information on internet search behaviour. Google Trends represent the relative number of searches for a specific key word, or combination of search terms. The numbers are standardized within each country such that the values range from 0 to 100. A search volume of 0 is assigned, by Google Trends, to weeks/months with a minimal amount of searches. Google Trends provides time-series of these abundance data, but gives no explanation of how the relative abundances were calculated. In addition to relative abundances, for each country, the downloaded Google Trends csv file provided a list of 'Top-searches for [the language-specific search term]'. Each top-search list included the context of the search term, each with an integer value of its relative abundance from $0-100$. The top searches are listed in descending order. Top searches were informative for determining how to interpret Google query data. Because searches for "shingles" on Google Trends could be referring to "the disease shingles" or "roofing shingles", the top search list could be used to distinguish among the search context when the single-word search term has an ambiguous context. The top-search list is therefore invaluable for ensuring data for specific search terms are being properly interpreted.

We used Google Trends data to evaluate childhood disease information seeking behaviour, and obtained country-specific data from the start of Google Trends, January 2004 to July 2015. We downloaded Google Trends data from 36 countries with high volume searches for chicken pox worldwide (Table S3). For each country the data were subset within the range that included consecutive weeks with > 0 search volume.

In order to relate Google Trends data to the dynamics of chicken pox (or other diseases of interest), care must be taken to select appropriate search terms. Chicken pox is the classical manifestation of disease, and therefore, language-specific queries of "chicken pox" are a straightforward choice for data-mining. In contrast, infections with generic symptoms, such as fever and diarrhea, could arise from many other diseases, making it difficult to identify appropriate queries. In either case, search terms vary subtly from country to country. For instance, in the US "chickenpox" is typically written as a single word, whereas in the U.K., people refer to "chicken pox" as two words; in Spanish, chicken pox is referred to as "varicela", with a single "l". We accounted for this variation among countries by careful choice of search terms, and downloaded the data for 36 countries using 21 language-specific queries of chicken pox. The csv file downloaded for each country included the top-searches for the country-specific search term. As an example, 48 top-searches were provided for Argentina; the top 5 searches–and their integer value, which we refer to as relative abundance–were:

Top searches for "varicela"

- la varicela, 100
- varicela sintomas, 35
- varicela vacuna, 30
- varicela contagio, 25
- sintomas de varicela, 20

We evaluated (and translated when needed) the top-search list provided for the US, Australia, Mexico, and Thailand, four of the countries for which we had data on reported cases, and these countries are highly variable in their varicella vaccination policy. Although the top-search lists had no metadata provided by Google, the clear difference between top searches among the four countries indicated that the top searches contained valuable epidemiological information. We therefore decided to systematically evaluate the top-search lists for epidemiological information. The top-search lists closely matched expectation based on each countries vaccination policy.

The 36 countries in our study differed in their VZV vaccination history and current policy (Table S3). The first VZV vaccine was licensed in the US in 1995 and was incorporated into the measles, mumps, rubella, and varicella vaccine (MMRV) in 2005 [2, 3]. These vaccinations were only implemented by a few countries, and at different times, which led us to expect country-specific differences in chicken pox query motivation. Therefore, for a subset of countries, we evaluated the context of Google Trends searches by categorizing searches based on whether they queried chicken pox as a disease, chicken pox vaccination, or other contexts (Fig S7, Table S2).

The significance of information seeking seasonality was tested using Morlet wavelet analyses for each country (Fig 1, S5) [4, 5]. Both wavelet analyses and General Additive Models (GAMs) are powerful methods for detecting periodicity in time series. The Google data for all countries other than Estonia, and the Czech Republic (which had monthly data) were examined for annual (52 week) periodicity using colored noise. First, we measured wavelet significance at the annual band. Second, of the 33 countries, 18 countries (Colombia, the UK, the US, Argentina, Brazil, Denmark, France, Hungary, India, Ireland, Italy, Mexico, The Netherlands, Poland, Romania, South Africa, Sweden, and Vietnam) had a significance band within the entire cone of influence of the annual band (time periods where you can test for significance). Seven countries (Australia, Finland, New Zealand, Philippines, Portugal, Spain, and Thailand) had significant power at the annual period for $> 50\%$ of their time series, with high-power periodicity (i.e., red banding) at the other non-significant time points. China had $\sim 40\%$ of its time series significant at the annual period, with high power annual periodicity at all other time points. In Germany the significance was lost about halfway through the time series, and the power of the annual period diminished. Chile and Japan did not have significant annual periodicity, although Japan did have high power at the annual period throughout its time series. The remaining four countries (Russia, Iran, Austria, and Venezuela) had time series that were too short to test for annual significance using wavelet analyses (i.e., they did not have 3+ years of data).

In order to characterize the seasonal shape of chicken pox information seeking, we used GAMs, which is a nonparametric extension of generalized linear models (GLMs) in which the linear predictor depends on smooth functions of predictor variables. We used the restricted maximum likelihood (REML) method with the linear predictor being the detrended Google data, while the predictor variables we tested included week number for seasonality and time for the overall trend. REML and maximum likelihood methods are less prone to local minima than the other criteria, and usually preferable. We used a GAM, rather than a generalized linear model or other model, because GAMs are flexible when fitting smooth curves to ecological data, and typically allow a better fit for time series than GLMs. A GAM was fit for all countries with significant annual wavelet periodicity, except Estonia and the Czech Republic, because their data were monthly and not weekly.

Validating Information Seeking in Chicken Pox

Once we characterized the seasonal variation in Google queries for chicken pox, we tested whether variation in information seeking behaviour paralleled variation in chicken pox incidence. The Google Trends data required validation against epidemiological data because variation in information seeking could be driven by cultural events rather than changes in disease incidence. For example, in the US, information seeking using the query “breast cancer” has a sharp seasonal peak in Google Trends each year in October, reflecting October as breast cancer awareness month, rather than a month with elevated incidence. For infectious diseases, the covariation between information seeking behaviour and clinical cases can be established using reported cases. We validated our Google Trends data, and evaluated search term context, using records of clinical cases from countries with active chicken pox surveillance (Fig 2).

We obtained data from five countries that report chicken pox: Australia, Thailand, and Estonia – which had monthly reported cases – and Mexico and the US, which reported cases weekly. The data from Australia were collected from the Australian National Disease Surveillance System [6], digitized on May 1st, 2015. Thailand chicken pox case data were downloaded from the Bureau of Epidemiology, Department of Disease Control, MoPH, Thailand [7] on April 12, 2015. The data from Mexico were digitized from the weekly disease surveillance reports of the Mexico General Directorate of Epidemiology, first published in [8] and provided to us by the authors. The US data, both historical (Fig S8) and modern, were obtained from the Project TYCHO database [9]. Data from Estonia were provided by the Estonian Health Board, Department of Communicable Disease Surveillance and Control [10]. Clinical data from these five countries span different time periods, but each overlapped with the Google Trends data for 4+ years. Clinical data spanned Jan 1995 – Feb 2011 in Mexico, Jan 2003–Dec 2014 in Thailand, Jan 2006–Feb 2015 in Australia, Jan 2006–Aug 2013 in the US, and Jan 1999–Dec 2014 in Estonia.

Forecasting

In order to determine if Google Trends data could be used to predict the magnitude and timing of chicken pox outbreaks, we built forecasting models. The models predicted the force of infection, which we defined as the monthly per capita rate at which children age 0–14 years are infected. We refer to this parameter as the force of infection, which is typically defined as the rate of infection per capita susceptible individual, because we are assuming all susceptible individuals are contained within the the 0–14 year age class, and therefore the number of 0–14 year olds is a surrogate for the number susceptible. The forecasting models containing Google data use Google data from the previous time intervals, $t-1$ and $t-2$, to predict the number of chicken pox cases at the current time interval, t. The models were fit to data from two countries that actively report chicken pox cases, one with active immunization (Australia) and one lacking immunization (Thailand). To determine whether Google Trends, T, was able to forecast the magnitude and timing of chicken pox outbreaks, we built and fitted multiple statistical models to forecast chicken pox case data. The correlation between chicken pox information seeking and chicken pox cases was weaker in Australia compared to Thailand ($R^2 = 0.26$ and 0.81, respectively). We therefore used the case data from Australia to test the power of the forecasting models, since Australia poses a more challenging forecasting problem.

The Google Trends data are weekly, whereas both Australia and Thailand reported chicken pox on a monthly basis. Thus, we forecast on a monthly basis and converted the weekly Google Trends data to monthly values. To do this, we repeated the weekly values at daily intervals (Google Trends data are relative search values and not absolute number of searches). We then assigned the daily values to their appropriate month of the year. For each month, we then found the mean of the daily values, which resulted in the values used for forecasting.

The null and four of our eight forecasting models included a cosine function to help predict chicken pox outbreaks. We discovered that the cosine function is required because it imposes cyclicity on the outbreaks, acting as a proxy for cyclical changes in (1) the number of susceptible individuals in the population and/or (2) the transmission rate. Although the Google Trends data are cyclical, since the forecasting model predicts one-month-ahead, without the cosine function, the Google Trends data alone would be limited in ability to forecast directionality (i.e., to determine if cases are increasing or decreasing). Including a cosine function with a period of 12-months allowed us to overcome this limitation. We tested eight different forecasting models, all slight variations of each other, and compared the model results to a null model that captured the annual seasonal patterns of chicken pox incidence (Table S1). It is unknown how the Google data scaled to chicken pox data. We therefore estimated scaling parameters in the various models (i.e. α, β_1, β_2, and β_3).

We evaluated the epidemiological information contained in Google Trends by comparing the Google Trends models with a seasonal null model that did not incorporate Google data (model B). The null model lacked information seeking in the force of infection λ_t. All models were fitted to the case data from a VZV-vaccinated population (Australia), which exhibited reduced seasonality. To estimate the number of symptomatic VZV infections each month, I_t, we used Google Trends data from the previous two months, T_{t-1} and T_{t-2}, where t is time in monthly time steps. The chicken pox process model with the best fit, tracked the force of infection, λ_t,

$$\lambda_t = \left[\beta_1 \cos\left(\frac{2\pi}{12}(t+\omega)\right) \mathrm{T}_{t-1} + \beta_2 |\mathrm{T}_{t-1} - \mathrm{T}_{t-2}| + \beta_3\right] \epsilon_t. \tag{A}$$

The model contained environmental stochasticity, ϵ_t, which was drawn from a gamma distribution with a mean of 1 and variance θ. We estimated the following parameters for the Google model: the mean and the phase of the seasonality (β_1 and ω), a parameter scaling the Google Trends data (β_2), the baseline force of infection (β_3), the process noise dispersion parameter (θ), and the reporting dispersion parameter (τ) of a normal distribution, with a mean of 1, from which case reports were drawn. The parameters were estimated using maximum likelihood by iterated particle filtering (MIF) in the R-package pomp [11, 12].

In order to estimate the number of symptomatic VZV infections per month, we multiplied the force of infection, λ, with an estimate of the population aged 0–14 years [13], C,

$$I_t = \lambda_t \mathrm{C}. \tag{1}$$

We modeled the observation process, which represents the number of cases reported. To account for stochasticity in the reporting of symptomatic VZV infections, case reports were drawn from a normal distribution with a mean report rate, $\rho = 1$, and dispersion parameter (τ) which was estimated from the data.

$$\text{chickenpox}_t \sim \mathcal{N}(\rho I_t, \tau I_t). \tag{2}$$

We evaluated the epidemiological information contained in Google Trends by comparing the Google Trends model with a seasonal null model where the force of infection did not incorporate Google Trends data. The null model force of infection was modeled as:

$$\lambda_t = \left[\beta_1 \cos\left(\frac{2\pi}{12}(t+\omega)\right) + \beta_3\right]\epsilon_t. \tag{B}$$

To explore other model possibilities, we tested seven other models that included Google data. All model parameters listed are in reference to the best fit Google Model (model A). Of the additional seven Google models, the first lacked the β_2 parameter;

$$\lambda_t = \left[\beta_1 \cos\left(\frac{2\pi}{12}(t+\omega)\right) \mathrm{T}_{t-1} + \beta_3\right]\epsilon_t. \tag{C}$$

The second model lacked the β_2 parameter, but included an additional Google Trends scaling parameter, α;

$$\lambda_t = \left[\beta_1 \cos\left(\frac{2\pi}{12}(t+\omega)\right) \mathrm{T}_{t-1}^{\alpha} + \beta_3\right]\epsilon_t. \tag{D}$$

The third model contained the α parameter,

$$\lambda_t = \left[\beta_1 \cos\left(\frac{2\pi}{12}(t+\omega)\right) \mathrm{T}_{t-1}^{\alpha} + \beta_2|\mathrm{T}_{t-1} - \mathrm{T}_{t-2}| + \beta_3\right]\epsilon_t. \tag{E}$$

The fourth model lacked the cosine function and β_2 parameter;

$$\lambda_t = [\beta_1(\mathrm{T}_{t-1}) + \beta_3]\,\epsilon_t. \tag{F}$$

The fifth model lacked the cosine function and β_2 parameter, but contained the α parameter;

$$\lambda_t = [\beta_1(\mathrm{T}_{t-1}^{\alpha}) + \beta_3]\,\epsilon_t. \tag{G}$$

The sixth model lacked the cosine function but included the α parameter;

$$\lambda_t = [\beta_1(\mathrm{T}_{t-1}^{\alpha}) + \beta_2|\mathrm{T}_{t-1} - \mathrm{T}_{t-2}| + \beta_3]\,\epsilon_t. \tag{H}$$

and the seventh model lacked the cosine function;

$$\lambda_t = [\beta_1(\mathrm{T}_{t-1}) + \beta_2|\mathrm{T}_{t-1} - \mathrm{T}_{t-2}| + \beta_3]\,\epsilon_t. \tag{I}$$

Results from all models are listed in Table S1. Models that included the cosine function, including the null (i.e. models A, B, C, D, and E), fit better than those that did not have the cosine function (i.e. models F, G, H, and I). The best fit Google forecasting model estimated six parameters and had an AIC of 1120.9, while the null model, which lacked Google Trends data, had an AIC of 1148.9. The best fit forecasting model without the cosine function estimated 4 parameters and had an AIC of 1179.3.

To further examine the difference between the Google model and the Null model, we ran 10000 simulations using the maximum-likelihood parameter set for each the Google model and the null model for Australia. We first examined each of the model fits (Fig S1) to the chicken pox case data. Since both models were seasonally forced, they were both able to capture the seasonal timing of outbreaks. However, the Google Trends model was able to predict the interannual variation in outbreak size, while the null model could not because the cosine function did not change interannually (Fig S1).

These results demonstrate that the Google Trends model was better able to capture the dynamics of chicken pox case data. The stochastic simulations showed more variation (larger standard deviation), and captured the data more often than the Null model. To visualize the relationship between the model simulations and the chicken pox case data, we plotted the mean predicted chicken pox cases (model results) against the actual cases for the Google Trends model and the Null model (Fig S1). Finally, to get a better understanding of why the Google Trends model fit the chicken pox case data better than the Null model, we explored the distribution densities of the troughs of each model against the data for each year (Fig S2). The Google Trends model achieved a better fit to chicken pox data (Fig S2). While the Google Trends model best captured the actual troughs in 2012, 2013, and 2014, its density distribution was always closer to the actual number of cases in the trough month relative to the Null model. The trough in 2006 was difficult to characterize because the model was estimating initial conditions, which could explain why neither the Google Trends model nor the Null model were able to accurately forecast the number of cases in May, 2006.

Information Seeking in other Childhood Diseases

To evaluate whether our findings based on chicken pox were representative of infectious childhood diseases in general, we examined information seeking behaviour for other childhood diseases. We obtained country-specific Google Trends data from the US and Australia for "hand foot and mouth", "croup", and "fifth disease" [1].

Google queries of croup, fifth disease, and HFMD in the US and Australia displayed variation in search volume (1) within and between years for each disease, (2) among diseases, and (3) across geographic locations for a given disease. HFMD displayed seasonal variation in the US and Australia (Fig S3). HFMD is caused by enteroviruses, which are notorious for their increased summer transmission in temperate regions [14]. The peak in HFMD information seeking generally occurred between June and August in the US and Australia, and was relatively synchronized between these countries. Seasonal variation in HFMD has been documented in clinical case data with peaks in the US occurring from spring to fall [15], which is in keeping with the seasonal variation observed in information seeking. In contrast, in Australia, the concurrent seasonal peak coincided with the southern hemisphere winter. This unexpected timing requires further investigation.

Croup information seeking also displayed seasonality, but unlike HFMD seasonal information seeking, it was asynchronous between the US and Australia. In the US, croup information seeking seasonally peaked between October and November, at the onset of the northern hemisphere winter; whereas in Australia, croup information seeking peaked from May–July, at the onset of the Australian winter. Croup is caused by Human parainfluenza viruses (HPIVs). HPIV-1 and HPIV-2, which cause croup in children, circulate in autumn, suggesting that the seasonality of croup information seeking in the US and Australia follow the seasonal circulation in HPIV-1 and 2.

Information seeking regarding fifth disease, which is caused by parvovirus B19, was highly seasonal in the US (Fig S3). The seasonal peak in fifth disease information seeking showed a distinct trough from August–October and peaked from April–May, roughly coinciding with the seasonal peak of clinically diagnosed fifth disease in late winter and early spring [16, 17]. The search volume of fifth disease was not sufficient outside North America for geographic comparison.

These preliminary examples further emphasize the untapped potential of analyzing information seeking behaviour of childhood infectious diseases. Digitally detecting pathogen-specific, large-scale spatio-temporal patterns can provide clues for identifying environmental and physiological drivers of the dynamics of these diseases.

Influenza Information Seeking Behaviour

Our data strongly suggest a signature of immunization on the seasonality of VZV. For the case of varicella, a readily apparent signature of immunization is perhaps to be expected because, when rolled-out into a population at high coverage, the VZV vaccine is highly effective [3, 18]. Therefore, we examined whether observable signatures of immunization can also be detected in diseases with lower vaccine efficacy. In order to determine whether vaccine effects could also be observed in Google data for other vaccine preventable disease, we obtained annual data on inactivated influenza vaccine efficacy and vaccine administration, weekly influenza and pneumonia mortality, and weekly information seeking regarding influenza and flu symptoms in the US state of Wisconsin (Fig S4). We measured effective immunization as the percent of the population expected to be immunized based on doses administered and vaccine efficacy, which varies substantially from year-to-year. Influenza mortality and information seeking displayed interannual variation not readily attributable to variation in effective immunization. We interpret this to be due to low effective immunization for influenza, which was $< 25\%$ in all years. Although flu and flu symptoms information seeking did not contain a signature of immunization, influenza and pneumonia mortality covaried with information seeking ($R^2 = 0.34$ and $R^2 = 0.50$ for flu and flu symptoms, respectively). This suggests that if seasonal flu immunization accounted for the interannual variation in influenza mortality, the effect of immunization would be reflected in flu information seeking.

For the state of Wisconsin, weekly influenza information seeking data were obtained from Google trends using the search terms "flu" and "flu symptoms". Wisconsin was chosen because published studies of inactivated influenza vaccine efficacy included patients from Wisconsin. The adjusted vaccine effectiveness estimates for influenza seasons were obtained from the CDC [19]. In years when the lower bound of the 95% CI of vaccine efficacy was negative, the efficacy was set to 0. Weekly influenza and pneumonia mortality was extracted from the Mortality Surveillance Data from the National Center for Health Statistics [20]. Influenza data are provided as a csv in the Supplemental Information.

Model	Model Structure	LogLik	# Params Est.	AIC	Δ AIC
A	Google Model	**-554.47**	6	**1120.9**	**0.0**
B	Null Model	−569.47	5	1148.9	28.0
C	-β_2	−558.32	5	1128.0	7.1
D	-β_2, $+\alpha$	−563.35	6	1138.7	17.8
E	$+\alpha$	−565.43	7	1144.9	24.0
F	- Cosine, -β_2	−585.63	4	1179.3	58.4
G	-Cosine, -β_2, $+\alpha$	−585.02	5	1180.0	59.1
H	-Cosine, $+\alpha$	−584.96	6	1181.9	61.0
I	-Cosine	−586.08	5	1182.2	61.3

Table S1: The equation letter, model structure, log–likelihood values, number of estimated parameters, AIC, and difference from top AIC values are shown above. Equation letter matches the equations from the supplement (main text references to the Google model, refer to the best-fit model, model A). The model structure refers to how the model varies from the top performing Google model. A cosine function, α parameter, or the β_2 parameter were either added (+) or removed (-) from the best fit Google Model. LogLik are the log–likelihood values for each models maximum likelihood parameter set. AIC refers to Akaike Information Criterion which penalizes models that use more parameters. The lowest AIC value represents the best fit model. Δ AIC was the difference in AIC from the best fit model.

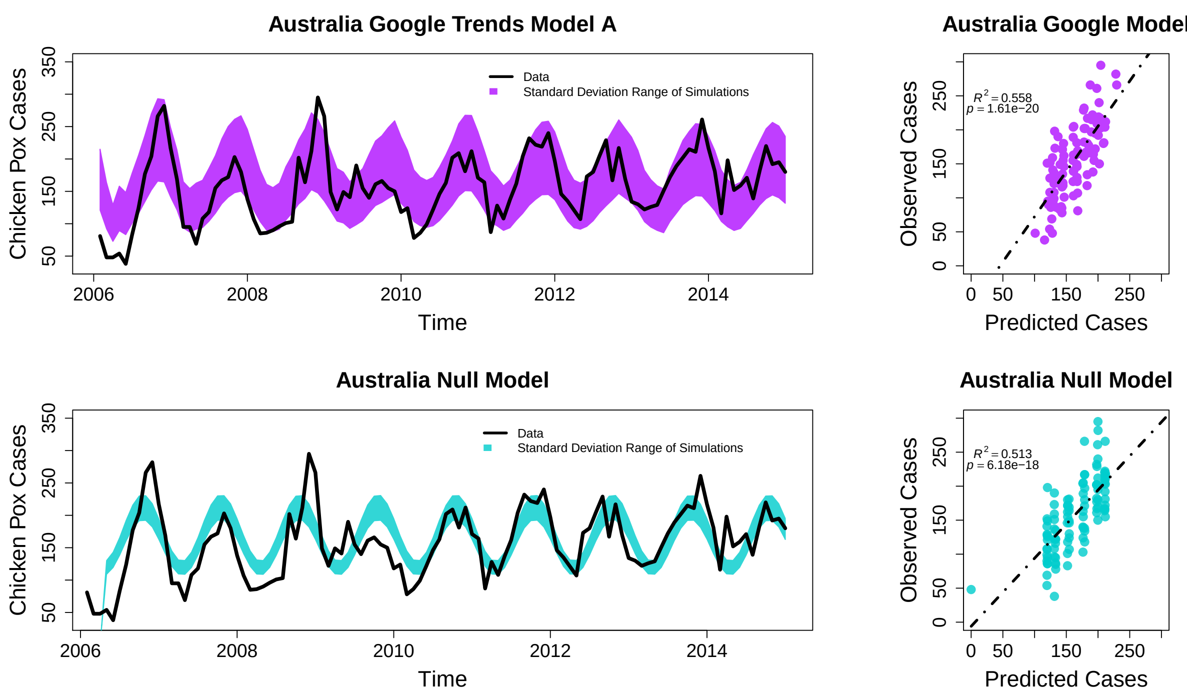

Figure S1: (top left panel)The best–fit Google Trends model (Model A), was simulated 10000 times and the range of standard deviation from the mean are plotted in purple against the actual Australian chicken pox case data (black). (top right panel) The relationship between the Google Trends model predicted chicken pox cases and the observed chicken pox cases. (bottom left panel) The Null model (Model B), was simulated 10000 times and the range of standard deviation from the mean are plotted in light blue against the actual chicken pox case data (black). (bottom right panel) The relationship between the Null model predicted chicken pox cases and the observed chicken pox cases.

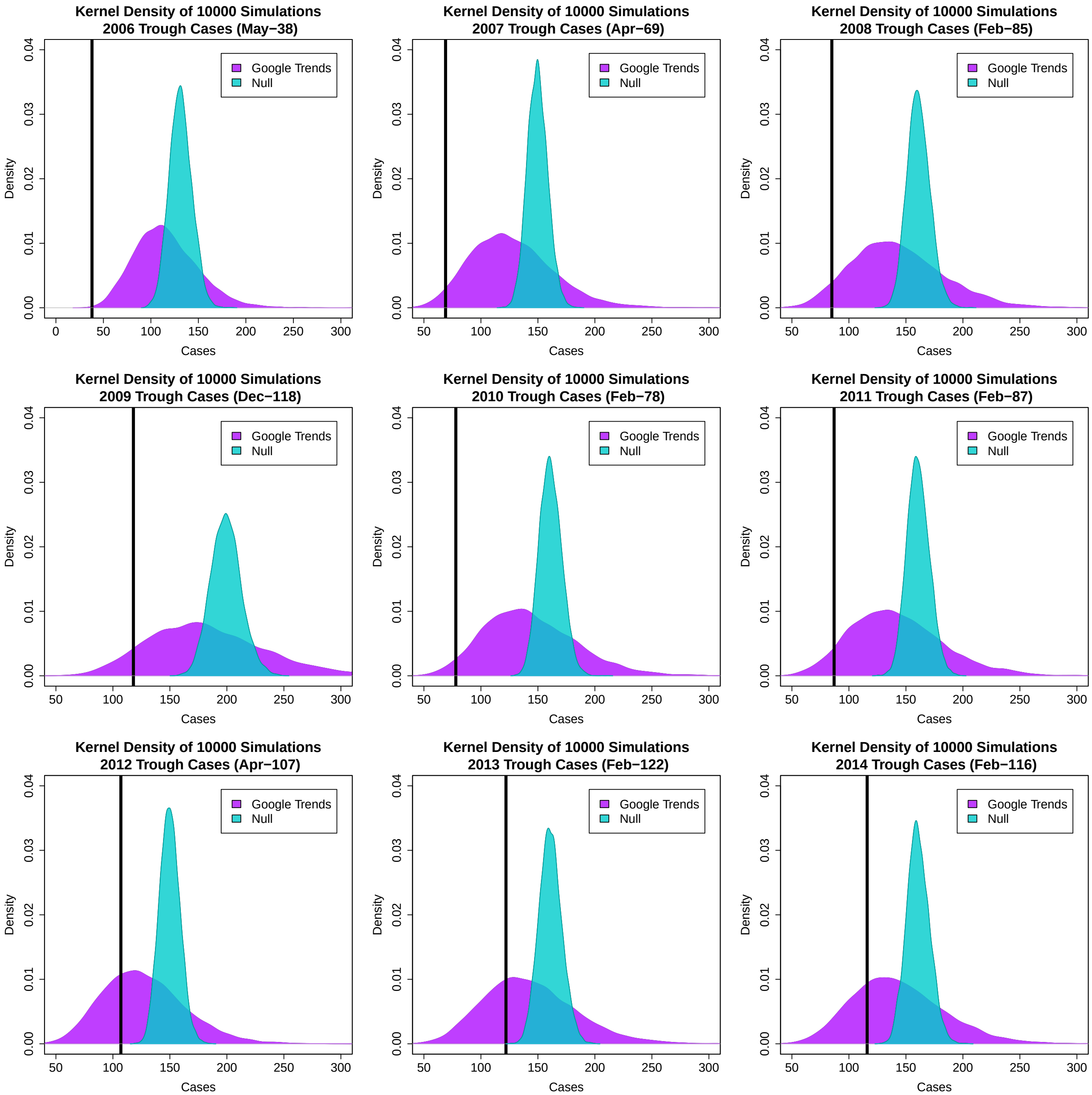

Figure S2: Density distributions of the 10000 simulations for each the Google Model (purple) and Null model (light blue) during the trough month in chicken pox cases for each year. The actual number of reported cases are in each panel title, and shown with a vertical black band.

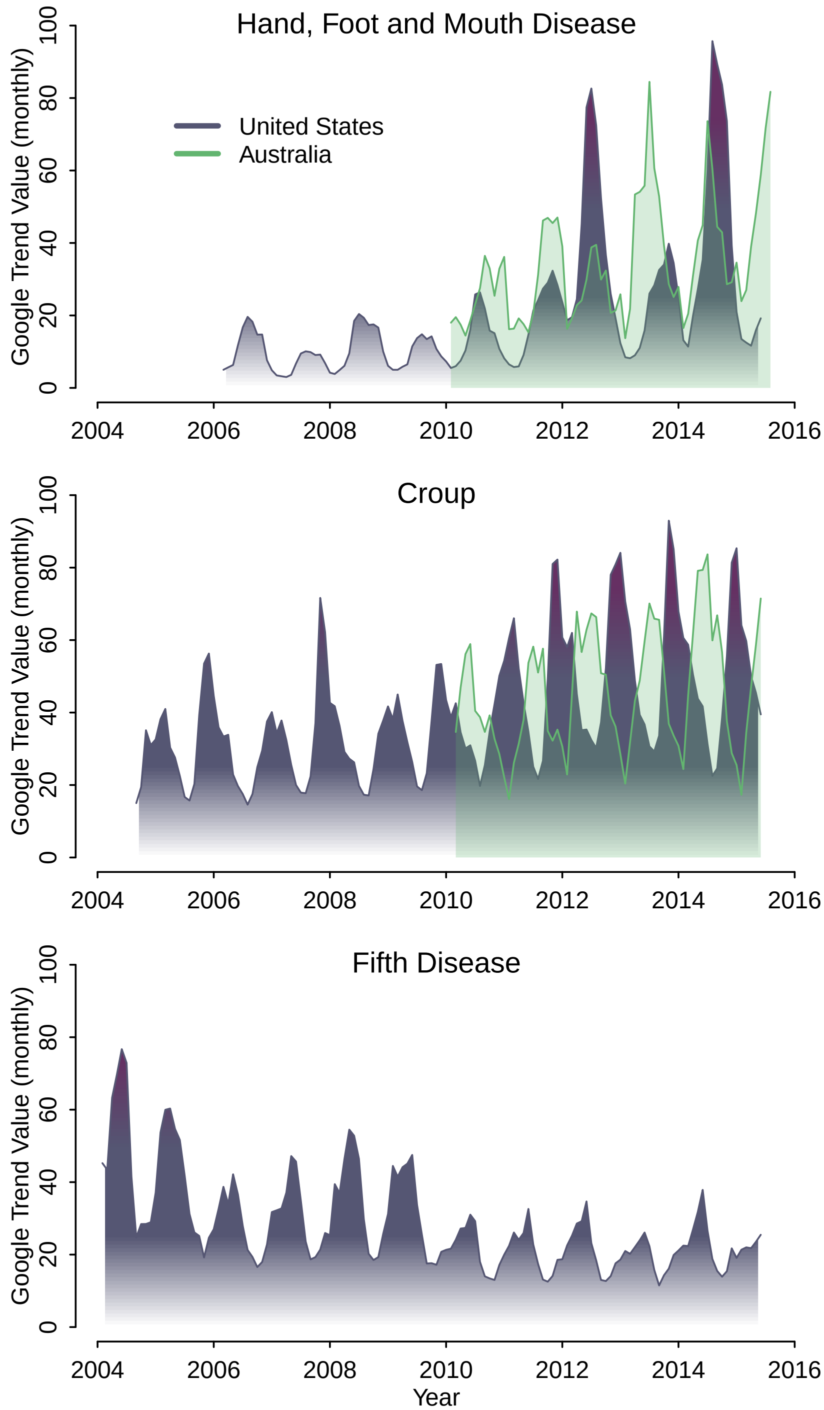

Figure S3: Seasonal variation in childhood disease information seeking. Time series of monthly information seeking for Google queries of "hand foot and mouth", "croup", and "fifth disease" in the US and Australia. Hand, foot and mouth disease (HFMD) queries in the US and Australia were relatively in phase with one another, whereas croup queries in the US and Australia were out of phase, with both occurring in the autumn of their respective hemisphere.

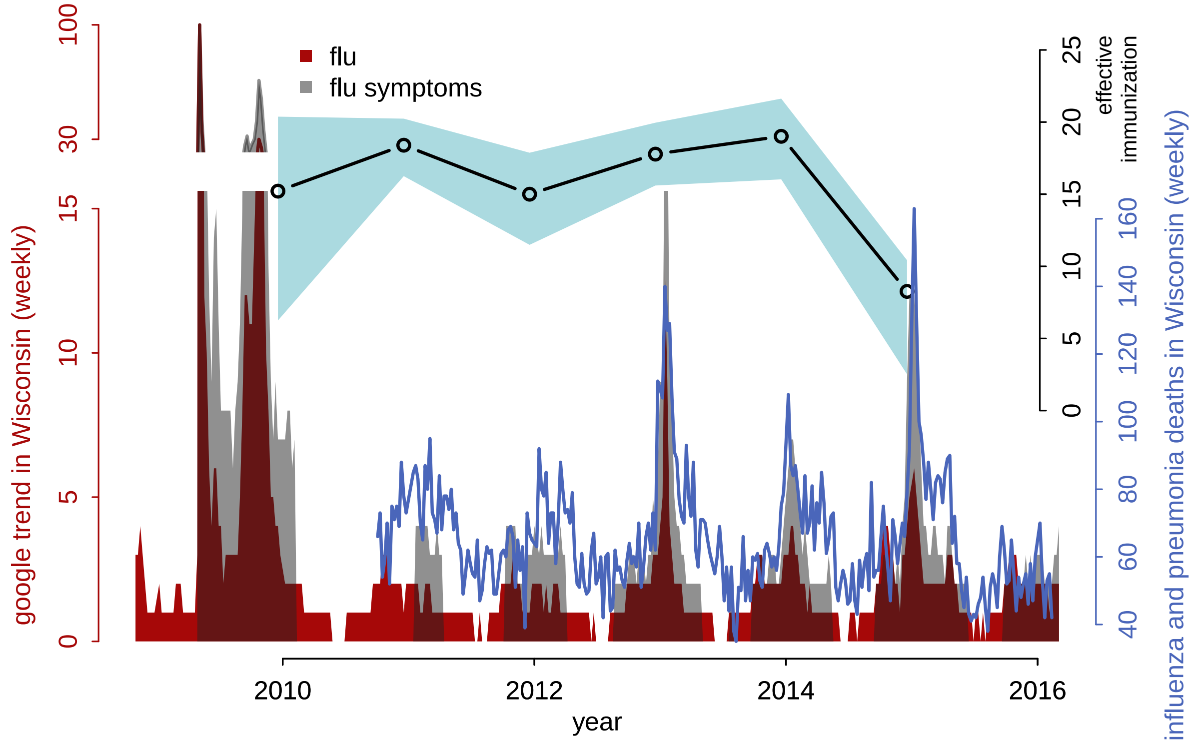

Figure S4: Influenza digital epidemiology and mortality in Wisconsin, USA. Weekly influenza and pneumonia mortality (blue line). Weekly Google trends based on the search terms "influenza" (red) and "influenza symptoms" (grey). Effective influenza immunization (black). The expected value and range of effective immunization, measured as the percent of the population immunized, was calculated by multiplying the percent of the population vaccinated by the annual vaccine efficacy. The expected values (black line with blue range) are based on the point estimates of vaccine efficacy, the range is based on the 95% CIs for reported efficacy.

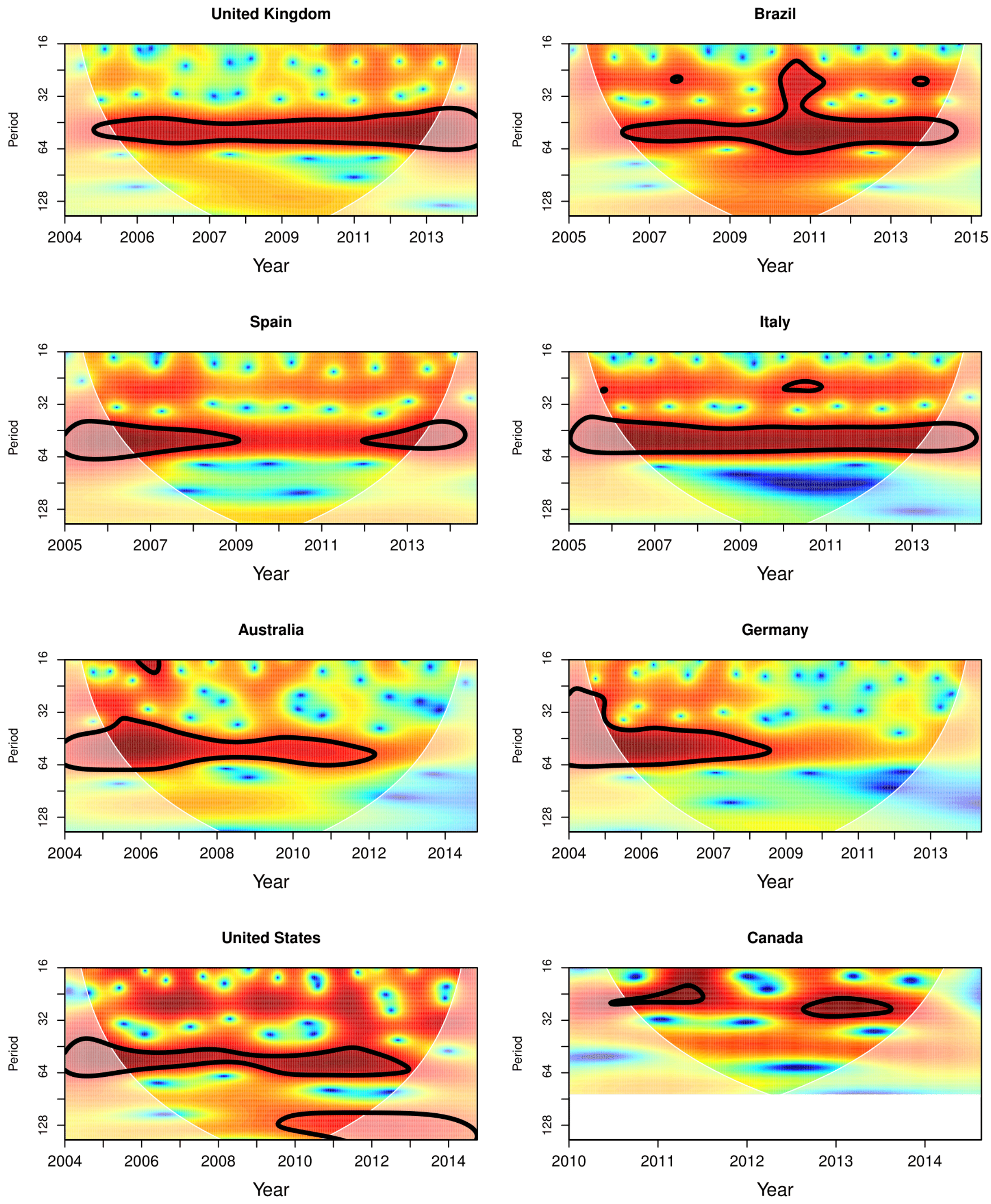

Figure S5: Wavelet analyses [4] for eight countries. Wavelets are used to identify periodic signals in non-stationary time series data. These signals can vary in amplitude, frequency, and phase over long temporal scales. Wavelets decompose time series data into signals of identifiable period and amplitude, both of which can change with time. For each figure, time is plotted on the x-axis, and the periodicity, in weeks, is plotted on the y-axis, with the colors representing the power of each frequency (blue=low, red=high). Areas circled in black have significant periodicity for that period. In the UK for example, the data are significant at 52 weeks throughout the entire time series (i.e. annual peaks). Meanwhile in both Australia and Germany, significant periodicity was lost during the time period analyzed, while in Spain, significant periodicity was lost between 2009-2012. Canada can only be tested for significance up to 96 weeks because of the short time series available.

Table S2: Chicken pox search term context. Search terms have been translated from Spanish to English for Mexico and from Thai to English for Thailand. Search terms that could not be properly translated from Thai are indicated by "could not translate" in the category. Note, there were unique searches in Spanish and Thai that resulted in the same English translation.

top chicken pox search terms	relative abundance	category	country	indicator
chicken pox	100	disease (common name or virus)	US	disease
chicken pox vaccine	80	vaccine or vaccination	US	vaccine
shingles chicken pox	75	similar disease	US	other
chicken pox symptoms	45	symptoms	US	disease
chicken pox pictures	30	disease (images)	US	disease
symptoms of chicken pox	30	symptoms	US	disease
what is chicken pox	25	disease (common name or virus)	US	disease
chicken pox adults	25	disease (based on stage of life)	US	disease
chicken pox virus	20	disease (common name or virus)	US	disease
pictures of chicken pox	20	disease (images)	US	disease
chicken pox in adults	20	disease (based on stage of life)	US	disease
chicken pox contagious	20	disease (other)	US	disease
chicken pox rash	20	symptoms	US	disease
chicken pox vaccine	20	vaccine or vaccination	US	vaccine
shingles vaccine	20	similar disease	US	other
varicella chicken pox	15	disease (common name or virus)	US	disease
varicella	15	disease (common name or virus)	US	disease
chicken pox and shingles	15	similar disease	US	other
chicken pox symptoms	15	symptoms	US	disease
signs of chicken pox	15	symptoms	US	disease
chicken pox pictures	10	disease (images)	US	disease
chicken pox in children	10	disease (based on stage of life)	US	disease
shingles from chicken pox	10	similar disease	US	other
never had chicken pox	10	uncategorized	US	other
is chicken pox contagious	10	disease (other)	US	disease
shingles contagious	10	similar disease	US	other
vaccine for chicken pox	10	vaccine or vaccination	US	vaccine
chicken pox treatment	10	care or treatment	US	disease
what is shingles	5	similar disease	US	other
chicken pox virus	5	disease (common name or virus)	US	disease
chicken pox images	5	disease (images)	US	disease
pregnancy and chicken pox	5	disease (based on stage of life)	US	disease
are chicken pox contagious	5	disease (other)	US	disease
chicken pox incubation	5	disease (other)	US	disease
incubation period chicken pox	5	disease (other)	US	disease
chicken pox history	5	uncategorized	US	other
what causes chicken pox	5	disease (other)	US	disease
is shingles contagious	5	similar disease	US	other
incubation for chicken pox	5	disease (other)	US	disease
causes of chicken pox	5	disease (other)	US	disease
cdc chicken pox	5	disease (other)	US	disease
exposure to chicken pox	5	disease (other)	US	disease
cause of chicken pox	5	disease (other)	US	disease
chicken pox transmission	5	disease (other)	US	disease
symptoms for chicken pox	5	symptoms	US	disease
stages of chicken pox	5	symptoms	US	disease
chicken pox vaccination	5	vaccine or vaccination	US	vaccine
varicella vaccine	5	vaccine or vaccination	US	vaccine
treatment for chicken pox	5	care or treatment	US	disease
treatment of chicken pox	5	care or treatment	US	disease
chicken pox	100	disease (common name or virus)	Thailand	disease
a chicken pox	85	disease (common name or virus)	Thailand	disease
chicken pox symptoms	50	symptoms	Thailand	disease
chicken pox treatment	50	care or treatment	Thailand	disease
chicken pox vaccine	40	vaccine or vaccination	Thailand	vaccine
vaccine	40	vaccine or vaccination	Thailand	vaccine
chicken pox medicine	35	care or treatment	Thailand	disease
chicken pox children	20	disease (based on stage of life)	Thailand	disease
chicken pox scars	20	disease (other)	Thailand	disease
prevent chicken pox	15	disease (other)	Thailand	disease
contact chicken pox	15	disease (other)	Thailand	disease
chicken pox blister	15	symptoms	Thailand	disease

Continued on next page

Table S2 – *Continued from previous page*

top chicken pox search terms	**relative abundance**	**category**	**country**	**indicator**
chicken pox blisters	15	symptoms	Thailand	disease
treatment of chicken pox	15	care or treatment	Thailand	disease
chicken pox solve	15	care or treatment	Thailand	disease
as chicken pox	10	disease (common name or virus)	Thailand	disease
chicken pox	10	disease (common name or virus)	Thailand	disease
chicken pox infection	10	disease (other)	Thailand	disease
chicken pox is caused by	10	disease (other)	Thailand	disease
chicken pox symptoms	10	symptoms	Thailand	disease
symptoms of chicken pox	10	symptoms	Thailand	disease
itchy chicken pox	10	symptoms	Thailand	disease
chicken pox symptoms	10	symptoms	Thailand	disease
symptoms of chicken pox	10	symptoms	Thailand	disease
price chicken pox vaccine	10	vaccine or vaccination	Thailand	vaccine
chicken pox vaccine	10	vaccine or vaccination	Thailand	vaccine
chicken pox treatment	10	care or treatment	Thailand	disease
treatment of chicken pox	10	care or treatment	Thailand	disease
treatment of chicken pox	10	care or treatment	Thailand	disease
chicken pox scars	10	disease (other)	Thailand	disease
green medicine (traditional medicine)	10	care or treatment	Thailand	disease
chicken pox green medicine (traditional medicine)	10	care or treatment	Thailand	disease
chicken pox cure	10	care or treatment	Thailand	disease
roy chicken pox	10	could not translate	Thailand	other
E-cooked	10	could not translate	Thailand	other
do not eat chicken pox	10	could not translate	Thailand	other
food chicken pox	10	could not translate	Thailand	other
the chicken pox	5	disease (common name or virus)	Thailand	disease
the chicken pox	5	disease (common name or virus)	Thailand	disease
chicken pox in children	5	disease (based on stage of life)	Thailand	disease
topical chicken pox	5	disease (other)	Thailand	disease
measels	5	similar disease	Thailand	other
chicken pox vaccine	5	vaccine or vaccination	Thailand	vaccine
chicken pox vaccine	5	vaccine or vaccination	Thailand	vaccine
to prevent chicken pox	5	disease (other)	Thailand	disease
scar treatment	5	care or treatment	Thailand	disease
chicken pox scar treatment	5	care or treatment	Thailand	disease
chicken pox scars	5	disease (other)	Thailand	disease
chicken pox wound healing	5	care or treatment	Thailand	disease
chicken pox hole	5	could not translate	Thailand	other
the chicken pox	100	disease (common name or virus)	Mexico	disease
chicken pox symptoms	25	symptoms	Mexico	disease
what is chicken pox	15	disease (common name or virus)	Mexico	disease
chicken pox infants	15	disease (based on stage of life)	Mexico	disease
chicken pox in infants	15	disease (based on stage of life)	Mexico	disease
symptoms of chicken pox	15	symptoms	Mexico	disease
measels	15	similar disease	Mexico	other
small pox	15	similar disease	Mexico	other
chicken pox adults	10	disease (based on stage of life)	Mexico	disease
chicken pox in adults	10	disease (based on stage of life)	Mexico	disease
pregnancy chicken pox	10	disease (based on stage of life)	Mexico	disease
treatment chicken pox	10	care or treatment	Mexico	disease
chicken pox vaccine	10	vaccine or vaccination	Mexico	vaccine
rubella	10	similar disease	Mexico	other
zoster chicken pox	5	disease (common name or virus)	Mexico	disease
chicken pox virus	5	disease (common name or virus)	Mexico	disease
chicken pox images	5	disease (images)	Mexico	disease
images of chicken pox	5	disease (images)	Mexico	disease
chicken pox pdf	5	disease (images)	Mexico	disease
chicken pox in pregnancy	5	disease (based on stage of life)	Mexico	disease
chicken pox babies	5	disease (based on stage of life)	Mexico	disease
chicken pox in babies	5	disease (based on stage of life)	Mexico	disease
symptoms chicken pox infants	5	disease (based on stage of life)	Mexico	disease
chicken pox spread	5	disease (other)	Mexico	disease
spread of chicken pox	5	disease (other)	Mexico	disease
hemorrhagic chicken pox	5	disease (other)	Mexico	disease
remedies for chicken pox	5	care or treatment	Mexico	disease
care chicken pox	5	care or treatment	Mexico	disease
aciclovir	5	care or treatment	Mexico	disease
treatment of chicken pox	5	care or treatment	Mexico	disease

Continued on next page

Table S2 – *Continued from previous page*

top chicken pox search terms	relative abundance	category	country	indicator
chicken pox home remedies	5	care or treatment	Mexico	disease
scars of chicken pox	5	disease (other)	Mexico	disease
treatment for chicken pox	5	care or treatment	Mexico	disease
marks of chicken pox	5	disease (other)	Mexico	disease
chicken pox postulates	5	symptoms	Mexico	disease
chicken pox care	5	care or treatment	Mexico	disease
vaccine for chicken pox	5	vaccine or vaccination	Mexico	vaccine
vaccine against chicken pox	5	vaccine or vaccination	Mexico	vaccine
the small pox	5	similar disease	Mexico	other
small pox and chicken pox	5	similar disease	Mexico	other
zoster herpes	5	similar disease	Mexico	other
chicken pox (misspelling)	0	disease (common name or virus)	Mexico	disease
chicken pox (misspelling)	0	disease (common name or virus)	Mexico	disease
photos of chicken pox	0	disease (images)	Mexico	disease
chicken pox and pregnancy	0	disease (based on stage of life)	Mexico	disease
chicken pox twice	0	disease (other)	Mexico	disease
scarlet fever	0	similar disease	Mexico	other
measel symptoms	0	similar disease	Mexico	other
small pox symptoms	0	similar disease	Mexico	other
symptoms chicken pox	100	symptoms	Australia	disease
rash	60	symptoms	Australia	disease
chicken pox rash	60	symptoms	Australia	disease
chicken pox	55	disease (common name or virus)	Australia	disease
chicken pox vaccine	50	vaccine or vaccination	Australia	vaccine
shingles chicken pox	45	similar disease	Australia	other
shingles	45	similar disease	Australia	other
measles	45	similar disease	Australia	other
chicken pox pictures	35	disease (images)	Australia	disease
adults chicken pox	30	disease (based on stage of life)	Australia	disease
chicken pox pregnancy	30	disease (based on stage of life)	Australia	disease
chicken pox children	25	disease (based on stage of life)	Australia	disease
chicken pox australia	25	disease (other)	Australia	disease
chicken pox contagious	25	disease (other)	Australia	disease
chicken pox baby	20	disease (based on stage of life)	Australia	disease
chicken pox spots	20	symptoms	Australia	disease
chicken pox vaccination	20	vaccine or vaccination	Australia	vaccine
chicken pox treatment	20	care or treatment	Australia	disease
chicken pox virus	15	disease (common name or virus)	Australia	disease
chicken pox babies	15	disease (based on stage of life)	Australia	disease
chicken pox pregnant	15	disease (based on stage of life)	Australia	disease
chicken pox immunisation	15	vaccine or vaccination	Australia	vaccine
varicella	10	disease (common name or virus)	Australia	disease
varicella chicken pox	10	disease (common name or virus)	Australia	disease
chicken pox images	10	disease (images)	Australia	disease
chicken pox in adults	10	disease (based on stage of life)	Australia	disease
chicken pox twice	10	disease (other)	Australia	disease
chicken pox incubation	10	disease (other)	Australia	disease
measles rash	10	similar disease	Australia	other
measles symptoms	10	similar disease	Australia	other
mumps	10	similar disease	Australia	other
chicken pox signs	10	symptoms	Australia	disease
chicken pox scars	10	disease (other)	Australia	disease
chicken pox photos	5	disease (images)	Australia	disease
chicken pox picture	5	disease (images)	Australia	disease
chicken pox toddler	5	disease (based on stage of life)	Australia	disease
chicken pox stages	5	disease (other)	Australia	disease
chicken pox herpes	5	disease (other)	Australia	disease
chicken pox mild	5	disease (other)	Australia	disease
small pox	5	similar disease	Australia	other
rubella	5	similar disease	Australia	other
german measles	5	similar disease	Australia	other
symptoms of chicken pox	5	symptoms	Australia	disease
shingles symptoms	5	similar disease	Australia	other
rashes	5	symptoms	Australia	disease
chicken pox rash	5	symptoms	Australia	disease
chicken pox vaccine	5	vaccine or vaccination	Australia	vaccine

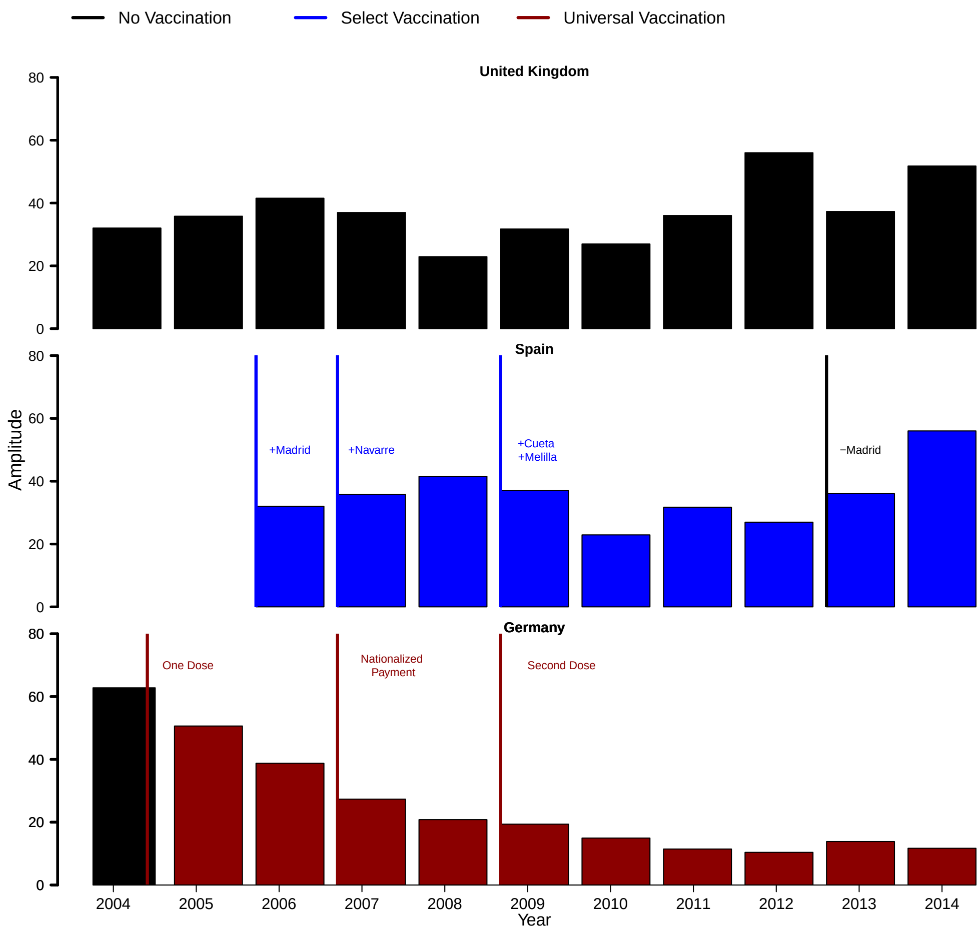

Figure S6: Annual amplitude values of Google searches for the United Kingdom, Spain, and Germany. Amplitudes were computed by first calculating the difference between the maximum each year and the mean each year. Second, we subtracted the minimum each year from the mean each year. Third, we found the difference between those two values and divided by two to get the final amplitude. Amplitude $= ((max(yr1) - mean(yr1)) - (min(yr1) - mean(yr1))/2$. In Spain, municipalities differed in their implementation of VZV vaccination. The Madrid metro region represents $\sim 14\%$ of the Spanish population (6.5m/46.7m), meaning that the vaccination policy in Madrid will have a large impact on overall chicken pox incidence and chicken pox Google Trends for the country. Spain initially had a significant seasonal period. However, after VZV vaccination was implemented in Madrid and additional cities, the significant seasonal periodicity was lost. Interestingly, the seasonality became significant again when Madrid withdrew VZV vaccination. This is similar to Germany (Fig 3 and S5), where the loss of significant wavelet periodicity followed the implementation of routine immunization after a few years. To examine this loss of seasonality in Google Trends in closer detail, we analyzed the annual amplitude for these two countries and the United Kingdom, which all differed in immunization mandates. The United Kingdom, which has no requirements, Spain, which implemented vaccination in certain municipalities for varying time periods, and Germany, which gradually increased its requirements over the course of a few years: first it required one shot, then made the payments nationalized, and finally required a second dose. In the UK, with no immunization requirements, the annual amplitude of Google searches for chicken pox remains relatively constant. In Spain, when all four municipalities were immunizing, the amplitude decreased from $\sim 40\%$ to $\sim 20\%$ in two years, before Madrid stopped vaccinating, after which the amplitude increased to over 50%. Meanwhile, in Germany pre-vaccine amplitudes in Google searches were $\sim 60\%$, before dropping to $\sim 40\%$ after the requirement of one dose, then dropping to $\sim 20\%$ after instituting nationalized payments, and finally dropping to $\sim 10-15\%$ after requiring a second dose. This additional analysis clearly elucidates the impact of immunization on search seasonality.

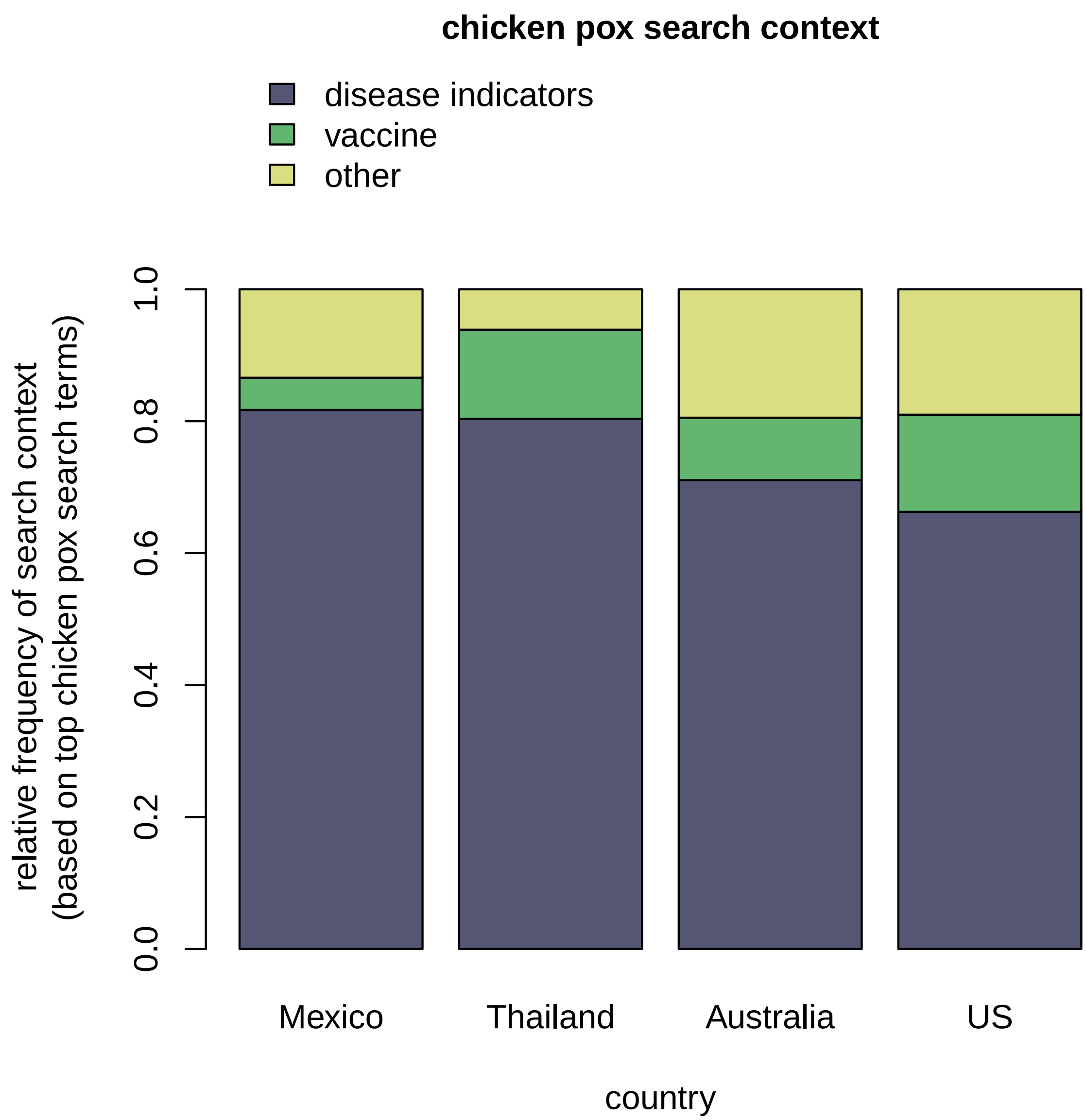

Figure S7: Relative frequency of language-specific chicken pox searches. Top searches were categorized into three broad categories: disease indicators, vaccine, or other. Disease indicators were searches considered to indicate chicken pox in the household/community. Vaccine indicators were searches regarding the VZV vaccine, and all other search contexts were placed in the "other" (also see Table S2).

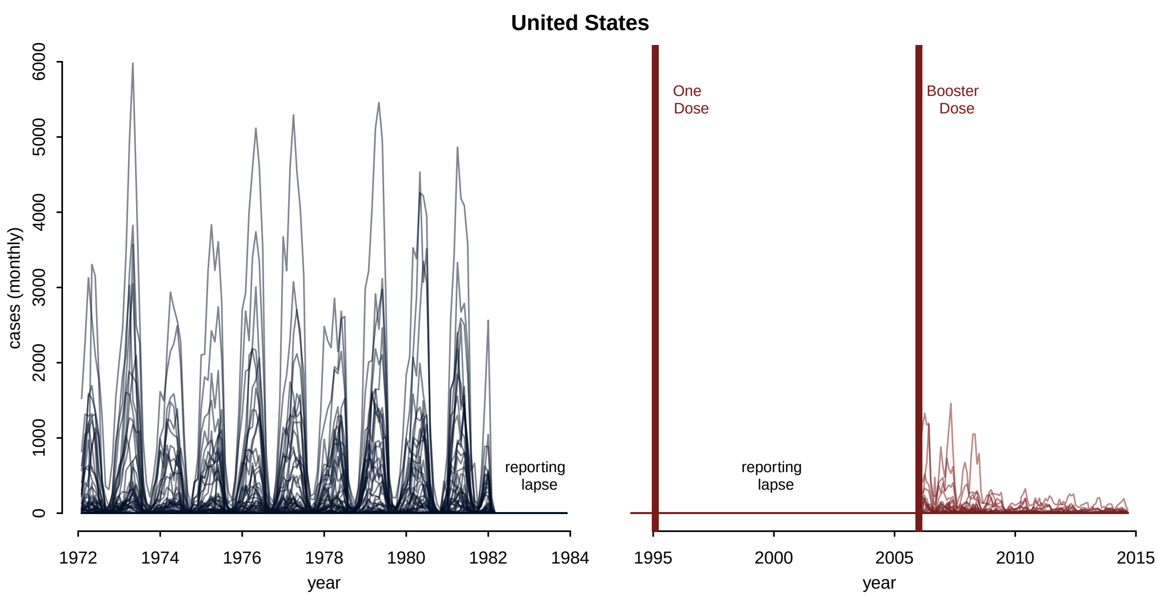

Figure S8: Weekly, state-level, data on chicken pox cases from [9] during the years 1972-2015 for all US states. Each state is plotted as an individual line. Black lines represent reported chicken pox cases during the pre-vaccine era, while red lines represent reported chicken pox cases during the vaccine era in the US

Table S3: Countries included in our chicken pox information seeking dataset from Google Trends. Countries in black lack nationwide immunization, countries in red began nationwide immunization in 2013 (Brazil) or 2014 (Japan), and countries in green have had nationwide immunization for multiple years. Countries with significant annual (seasonal) periodicities are in *italics*.

Countries			
Argentina	*Australia*	Austria	*Brazil*
Canada	Chile	*China*	*Colombia*
Czech Republic	*Denmark*	*Estonia*	*Finland*
France	Germany	*Hungary*	*India*
Iran	*Ireland*	*Italy*	*Japan*
Mexico	*Netherlands*	*New Zealand*	*Philippines*
Poland	*Portugal*	*Romania*	Russia
South Africa	*Spain*	*Sweden*	*Thailand*
United Kingdom	*United States*	Venezuela	*Vietnam*

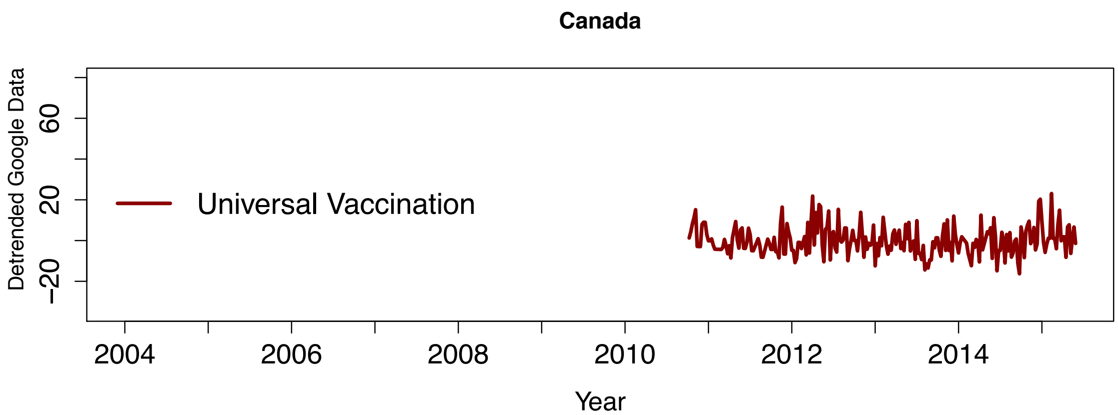

Figure S9: Canada Google Trends time series, data only available for the period with Universal Immunization (in red). This figure displays the lack of seasonality in the Google Trends data for Canada, a country with active immunization since 2000.

References

1. Google. (2015) Google Trends. https://www.google.com/trends/.

2. CDC. (2015) Chickenpox (Varicella) Vaccine Safety. http://www.cdc.gov/vaccinesafety/vaccines/varicella-vaccine.html. Accessed Sept 15, 2015.

3. Seward, J. F, Watson, B. M, Peterson, C. L, Mascola, L, Pelosi, J. W, Zhang, J. X, Maupin, T. J, Goldman, G. S, Tabony, L. J, Brodovicz, K. G, Jumaan, A. O, & Wharton, M. (2002) Varicella disease after introduction of varicella vaccine in the United States, 1995-2000. *Journal of the American Medical Association* **287**, 606–611.

4. Torrence, C & Compo, G. (1998) A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* **79**, 61–78.

5. Hastie, T & Tibshirani., R. (1990) *Generalized additive models.* (CRC Press) Vol. 43.

6. Australian-Government. (2015) National Notifiable DIseases Surveillance System. http://www9.health.gov.au/cda/source/rpt_1_sel_a.cfm. Accessed May 1, 2015.

7. Thai-Government. (2015) National Disease Surveillance. http://www.boe.moph.go.th/boedb/surdata/index.php. Accessed April 12, 2015.

8. Vergara-Castañeda, A, Escobar-Gutiérrez, A, Ruiz-Tovar, K, Sotelo, J, Ordoñez, G, Cruz-Rivera, M. Y, Fonseca-Coronado, S, Martinez-Guarneros, A, Carpio-Pedroza, J. C, & Vaughan, G. (2012) Epidemiology of varicella in mexico. *Journal of Clinical Virology* **55**, 51–57.

9. van Panhuis, W. G, Grefenstette, J, Jung, S. Y, Chok, N. S, Cross, A, Eng, H, Lee, B. Y, Zadorozhny, V, Brown, S, Cummings, D, & Burke, D. S. (2013) Contagious diseases in the United States from 1888 to the present. *New England Journal of Medicine* **369(22)**, 2152–2158.

10. Estonian-Government. (2015) Terviseamet Health Board. http://www.terviseamet.ee/. Accessed Sept 1, 2015.

11. King, A. A, Nguyen, D, & Ionides, E. L. (2015) Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software* **arXiv preprint arXiv:1509.00503**.

12. King, A. A, Ionides, E. L, Bretó, C. M, Ellner, S. P, Ferrari, M. J, Kendall, B. E, Lavine, M, Nguyen, D, Reuman, D. C, Wearing, H, & Wood, S. N. (2015) *pomp: Statistical Inference for Partially Observed Markov Processes,* (http://kingaa.github.io/pomp). R package, version 1.2.1.1.

13. UN. (2015) http://esa.un.org/unpd/wpp/ Accessed June 18, 2015.

14. Repass, G. L, Palmer, W. C, & Stancampiano, F. F. (2014) Hand, foot, and mouth disease: Identifying and managing an acute viral syndrome. *Clevel* **81(9)**, 537–543.

15. Khetsuriani, N, LaMonte-Fowlkes, A, Oberst, S, & Pallansch, M. A. (2006) Enterovirus surveillance United States, 1970D2005. *Morbidity and Mortality Weekly Report surveillance summary* **8**, 1–20.

16. Rodis, J. F, Jr, T. J. H, Quinn, D. L, Rosengren, S. S, & Tattersall, P. (1988) Human parvovirus infection in pregnancy. *Obstetrics & Gynecology* **72**, 733–738.

17. Anderson, L. (1987) Role of parvovirus b19 in human disease. *The Pediatric infectious disease journal* **6(8)**, 711–718.

18. Baxter, R, Tran, T. N, Ray, P, Lewis, E, Fireman, B, Black, S, Shinefield, H. R, Coplan, P. M, & Saddier, P. (2014) Impact of vaccination on the epidemiology of varicella: 1995-2009. *Pediatrics* **134**, 24–30.

19. CDC. (2016) Seasonal Influenza Vaccine Effectiveness, 2005-2015. http://www.cdc.gov/flu/professionals/vaccination/effectiveness-studies.htm Accessed March 4, 2016.

20. CDC. (2016) National Center for Health Statistics Mortality Surveillance Data. http://www.cdc.gov/flu/weekly/nchs.htm Accessed March 4, 2016.