Title: Measured voluntary avoidance behaviour during the 2009 A/H1N1 epidemic

Authors: Jude Bayham[1]*, Nicolai V. Kuminoff[2], Quentin Gunn[2], Eli P. Fenichel[1]

Affiliations:

[1] Yale School of Forestry and Environmental Studies, Yale University. 195 Prospect Street, New Haven, CT 06511.

[2] Department of Economics, Arizona State University, Tempe, AZ 85287.

*Correspondence to: jude.bayham@yale.edu

Abstract:

Managing infectious disease is among the foremost challenges for public health policy. Interpersonal contacts play a critical role in infectious disease transmission, and recent advances in epidemiological theory suggest a central role for adaptive human behaviour with respect to changing contact patterns. However, theoretical studies cannot answer the question: are individual responses to disease of sufficient magnitude to shape epidemiological dynamics and infectious disease risk? We provide empirical evidence that Americans voluntarily reduced their time spent in public places during the 2009 A/H1N1 swine flu, and that these behavioral shifts were of a magnitude capable of reducing the total number of cases. Moreover, we simulate ten years of epidemics (2003 – 2012) based on mixing patterns derived from individual time-use data to show that the mixing patterns in 2009 yield the lowest number of total infections relative to if the epidemic had occurred in any of the other ten years. The World Health Organization and other public health bodies have emphasized an important role for "distancing" or non-pharmaceutical interventions. Our empirical results suggest that neglect for voluntary avoidance behaviour in epidemic models may overestimate the public health benefits of public social distancing policies.

Introduction

Managing infectious disease is among the foremost challenges for public health policy. The World Health Organization and other public health bodies have emphasized an important role for "social distancing" or non-pharmaceutical interventions such as school and workplace closure [1]. Indeed, studies have shown that distancing policy can effectively mitigate disease spread by reducing contact between susceptible and infected individuals [2–4]. However, distancing policy can impose large economic and social costs [5–8]. In order to understand the public health benefits of social distancing policies, we must establish a behavioural baseline during an epidemic – the public health outcomes resulting from private actions of individuals to reduce their risk of infection [9].

Economic-epidemiology theory suggests that susceptible individuals may forgo beneficial contacts in order to reduce the probability of contracting a costly infectious disease [7,10–13]. Following standard epidemiological theory, this voluntary avoidance behaviour would in turn mitigate disease transmission and imply a dynamic feedback between humans and pathogens over the course of an epidemic [14]. We study whether individuals engaged in epidemiologically avoidance behaviour during the 2009 A/H1N1, and if the magnitude of individual behavioural shifts was of sufficient magnitude to alter epidemiological dynamics in the US. This enables us to quantify the approximately size of adaptive behavioural feedbacks on public health outcomes.

Several studies have used simulation to illustrate the potential public health benefits of avoidance behaviour [15,16]. Empirical efforts to quantify individuals' responses to infectious disease risk are often based on one-off surveys in the wake of the epidemic, but not coupled with epidemiological dynamics [17–20]; or infer potential avoidance behaviour ex post from observed

epidemic outcomes [21–23]. While these studies provide empirical insights into the role of avoidance behaviour during an epidemic, no study has quantified avoidance behaviour based on observable time-use data and coupled that behavioural shift with an epidemiological model to provide an empirical estimate of the public health consequences of avoidance behaviour.

We fill this important gap in the literature by estimating voluntary avoidance behaviour during the 2009 A/H1N1 (swine) flu epidemic using a detailed dataset with daily observations on how Americans spent their time between 2003 and 2012. We find that the average American reduced his time spent in potentially risky public locations by approximately 22 minutes with a 95% confidence interval of (6, 34) at the peak of the epidemic. Simulation results suggest that this behavioural shift may have reduced peak prevalence by 31% (7%, 48%) and the attack rate by 16% (3%, 25%).

We provide further evidence for epidemiologically meaningful avoidance behaviour by estimating empirical age by household size structured contact matrices by year over the ten year sample (2003-2012) and constructing counterfactual epidemics based on the outbreak period (April 20-December 20) and the pre-outbreak period (January 1 – April 19). We find that the attack rate of an epidemic based on the contact patterns observed during the 2009 outbreak period is substantially lower than that based on the average contact patterns observed during the outbreak period during 2003-2008 and 2010-2012. Moreover, contact patterns prior to the 2009 outbreak period are nearly identical to the average contact patterns prior to the outbreak period in 2003-2008 and 2010-2012. These results suggest that individual voluntary avoidance behaviour was of sufficient magnitude to meaningfully alter disease dynamics and impact transmission of the A/H1N1 influenza virus. Neglecting to account for voluntary avoidance behaviour may lead policy analyses to overstate the policy benefits of costly social distancing policies such as school

and work closure. Worse still, failing to understand behavioural responses to disease risk could potentially lead to policies that crowd out individuals own incentives to shift toward more protective behaviours.

Our methods to quantify feedbacks between avoidance behaviour and disease prevalence differ from those of related studies. Caley et al. [21], and He et al. [23] both fit epidemiological models to epidemic data from 1918 that attribute the model residual to avoidance behaviour. In contrast, survey-based studies [17–20] document individuals' reported avoidance behaviour, but are unable to characterize the human-pathogen feedbacks. Furthermore, our efforts to build behavioural feedbacks into models of 2009 A/H1N1 epidemic complements insights from models targeted at other aspects of the epidemic [24–29]. Our study bridges methods by utilizing surveyed behavioural data and reported epidemic data (as well as news and internet trends) in an epidemiological model.

Methods

Data. Data for this study come from multiple sources. Time use data for the general population were compiled from the American Time-Use Survey (ATUS) (2003-2012) [30]. The ATUS is subsampled from the U.S. Current Population Survey that contains detailed demographic and socioeconomic information about respondents older than 15 years old and their family members (including children under 15 years of age). Survey respondents report a 24-hour diary of activities, locations, and accompanying persons for every minute of the day. We supplement the ATUS data with time use data on children at school from the National Health and Activity Patterns Survey (NHAPS), a similar time-use survey, that includes children under 15 years old [31].

The weekly number of laboratory-confirmed cases were collected from [32], who obtain

data from the Centers for Disease Control and Prevention (CDC) Influenza Surveillance System. We used this measure of disease prevalence to capture the objective risk of spending time in public. Laboratory-confirmed cases peaked at 9,734 during the week of October 18-24, 2009 (Supplementary Figure S1). Data from Google Trends were used to measure the subjective risk of infection over the course of the epidemic. Google monitors search volume of specific key words and the occurrence of key words in media headlines from a variety of sources (Supplementary Figure S1). Extreme weather data were collected from The National Oceanic and Atmospheric Administration National Climatic Data Center Storm Events Database [33]. Extreme weather was used to control for additional time spent at home to avoid weather rather than flu risk.

Regression. We specify a series of fixed effects regression models to test the hypothesis that individuals engaged in avoidance behaviour in response to subjective (media attention) and objective (laboratory confirmed cases) measures of risk. The fixed effects regression model is

$$TIME_{tsm} = \beta_0 + \beta_1 CASES_t + \beta_2 MEDIA_{st} + \beta_3 WEATHER_{tsm} + \beta_4 X_{tsm} + \gamma_1 d_s + \gamma_2 d_m + a_{sm} + u_{st}$$

The variable $TIME$ denotes the number of minutes spent at home (ATUS) and subscript *t* indexes date, and the subscripts *s* and m index state and month. Time spent at home is considered safer than in public microenvironments during an epidemic and is indeed the motivation for social distancing policy such as school closure. Moreover, 26% of U.S. households consist of a single individual, which eliminates all household infection risk for this 26% of households [30]. We develop this argument further in the epidemic model section.

The variable $MEDIA$ represents the Google search index, $CASES$ represents the number of CDC laboratory confirmed weekly cases, $WEATHER$ represents instances of extreme

weather, X is a vector of demographic characteristics described in the electronic supplementary material, and d and a are month and state dummy variables that form the fixed-effects model.

We estimate three models using state fixed effects (model 1), month fixed effects (model 2), and state by-month fixed effects (model 3) to illustrate that our estimates are robust across model specifications (two additional models are presented in the SI material to further examine model specification uncertainty). All regressions are based on 27,091 observations of ATUS from 2008 to 2010 so pre- and post-epidemic periods are included in the model. Summary statistics are reported in Supplementary Table S1. All regression analyses were conducted in Stata 12.

Epidemic Simulations. We construct a SIR compartmental model to develop a first approximations to counterfactual epidemics of the 2009 A/H1N1 outbreak [34–36]. We specify the set of differential equations governing the transmission dynamics as

$$\dot{\boldsymbol{S}} = -\left(\boldsymbol{S} \circ \delta \boldsymbol{C}\left(\frac{\boldsymbol{I}}{\boldsymbol{N}}\right) + \boldsymbol{\sigma} \circ \left[\frac{\boldsymbol{S}}{\boldsymbol{H}} \circ \alpha\delta \boldsymbol{C}^{h}\left(\frac{\boldsymbol{I}}{\boldsymbol{N}}\right)\right]\right);$$

$$\dot{\boldsymbol{I}} = \left(\boldsymbol{S} \circ \delta \boldsymbol{C}\left(\frac{\boldsymbol{I}}{\boldsymbol{N}}\right) + \boldsymbol{\sigma} \circ \left[\frac{\boldsymbol{S}}{\boldsymbol{H}} \circ \alpha\delta \boldsymbol{C}^{h}\left(\frac{\boldsymbol{I}}{\boldsymbol{N}}\right)\right]\right) - v\boldsymbol{I}; \quad (1)$$

$$\dot{\boldsymbol{R}} = v\boldsymbol{I}$$

where $\circ$ and / denotes element by element multiplication and division. $\boldsymbol{S}, \boldsymbol{I},$ and $\boldsymbol{R}$ are $K \times 1$ vectors of susceptible, infectious, and recovered health classes where K is the number of subpopulations (e.g., age groups). $\boldsymbol{N}$ is an $K \times 1$ vector of subpopulations in each segment. $\boldsymbol{H}$ is a $K \times 1$ vector of the number of households in each subpopulation. $\boldsymbol{C}$ and $\boldsymbol{C}^{h}$ are $K \times K$ public and household probabilistic contact matrices that describe the interaction between and individual in subpopulation j (rows) and subpopulation k (columns). $\boldsymbol{\sigma}$ is a $K \times 1$ vector indicating the number of infected households where each element must be strictly between $\mathbf{0}$ and $\boldsymbol{I}$. δ is the disease-specific infectivity parameter, or conditional probability of transmission per minute of

contact between a susceptible and infected individual. α is a scalar that adjusts the relative infectiveness of a contact minute in the home relative to one in public. $1/v$ is the average infectious period constant across classes.

Epidemic dynamics depend on time spent in public and household environments [37,38]. The first term in equation (1) captures public transmission and is the contact time analog to common specifications. An individual makes potentially infectious contacts with household members if and only if there is at least one infected person in the household. The second terms capture within household transmission in the infectious home environment. This model makes a number of conservative assumptions that inflates the within household transmission. These assumptions will work to mask the epidemiological effects of individuals attempting to avoid infection by allocating more time to the household. We find significant effects of avoidance behaviour in spite of these assumptions. At any point in time, the expected number of susceptible individuals within the average household for each subpopulation is at most $\boldsymbol{S/H}$. This is an expectation across the entire population, but knowing that infected households must have at least 1 infectious individual and that household sizes are fixed implies the expected number of susceptible individuals in an infectious household must be less than $\boldsymbol{S/H}$, unless all households are infected. We approximate the number of infectious households in subpopulation j as I, which maximizes the potential for within household transmission. These assumptions are conservative and overestimate within household transmission because it allows the greatest number of households to be infected, imply a larger number of susceptibles in the infectious household environment than are truly at risk at home, and implicitly allows members of infectious households to "mix" freely among infectious households regardless of true home.

We assume δ is common to all population types. This assumption could be generalized

with estimates of δ that are location and attribute class specific that are independent of behaviour. To our knowledge such estimates do not exist because the multiplicative relationship between contact time and infectivity makes identification of location-specific δ difficult without imposing additional assumptions. Using age or location specific estimates of δ that did not control for contact time would confound our results.

While many detailed models of the H1N1 epidemic have been proposed (e.g., [24,39,40]), our work focuses on the behavioural responses and feedback to epidemiological dynamics. We use a relatively simple model in order to focus attention on the potential role of voluntary avoidance behaviour as a feedback mechanism. We model an epidemic over a short period of time such that births and deaths are negligible, a common assumption for influenza [41–43]. The model assumes that the entire population is susceptible prior to the introduction of the pathogen.

We simulate the epidemiological relevance of the avoidance behaviour estimated in the regression model using a homogeneous mixing model ($K = 1$). In the baseline case, public contact time is equal to the average of time spent in public in the ATUS ($C_0 = 316$ minutes or 5.26 hours per day) and remains constant throughout the epidemic (no avoidance). Alternatively, individuals respond to disease risk by shifting time in public to their household. Formally, $C(I) = C_0 - \beta_1 * \phi * \frac{I(t)}{1000}$, where β_1 is the minutes of avoidance behaviour per 1000 cases estimated from the regression model and $\phi = 8.33\%$ is the proportion of infected individuals confirmed through laboratory testing described in the Supplementary Material. Likewise, $C^h(I) = C_0^h + \beta_1 * \phi * \frac{I(t)}{1000}$ is the contact time the average individual experiences at home. We assume that the population has no memory and only responds to disease risk at time t, which yields conservative estimates of the epidemiological impact of avoidance behaviour. This

infection dependent contact rate is similar to the effective rate of transmission characterized by Funk et al [15] to model avoidance behaviour as a function of information about disease risk.

While the relationship between time spent interacting in public and transmission may be complex, and depend on many factors, δ can be interpreted as a first-order approximation of infectivity conditional on contact. We calibrate the conditional infectivity, δ, such that the maximum of the simulated prevalence path under avoidance behaviour equals the peak prevalence observed during the 2009 A/H1N1 outbreak (9,734 cases), which yields $\delta = 1.4\ 10^{-3}$. The simulation without avoidance then represents the possible epidemic outcomes had no individual engaged in avoidance behaviour. We set the household contact scalar to unity ($\alpha = 1$). This assumption is based on a systematic review of the empirical literature on household transmission that finds no consistent patterns [44]. Cauchemez et al. [45] and House et al. [38] find that larger households do not appear to have greater within home transmission, suggesting that a minute in proximity with an infected person within in a household is probably not qualitatively different from a minute spent with an infected person outside the household. We investigate the sensitivity of the results to this assumption in the supplementary materials.

We assume a population of 4.1×10^6 with 1.5×10^6 households, representing a U.S. city the size of the Phoenix Metropolitan Statistical Area (MSA), which was among the most affected areas during the 2009 A/H1N1 epidemic. We initialize the epidemic by introducing 20 infected individuals into the susceptible population. Increasing the number of initially infected individuals accelerates the time until large-scale outbreak but has no effect on the avoidance results.

Probabilistic Contact Matrix (PCM). We provide an alternative test for the impact of avoidance behaviour on epidemic dynamics by constructing counterfactual epidemics via

simulation based on age-household size ({0-4,5-12,13-17, 18-24, 25-49, 50-64, 65+}×{1,2,3,4,5}∈ P) where set P is of length K probabilistic contact matrices (PCMs) derived from the ATUS for each year between 2003 and 2012 [46]. The PCMs specify the amount of time (excluding time asleep) an individual of group $j \in P$ is exposed to populations $k \in P$.

These PCMs capture the fact that individuals can modify their schedule to avoid potentially infected individuals as an alternative form of avoidance behaviour. For example, young adults may go to health clubs and gyms in the evening to socialize, while other adults may go early in the morning to avoid congestion. Our PCM construction approach captures an individual's reallocation of time throughout the day across many activities including time at home. We break each individual year from 2003 to 2012 into two periods and construct a total of 20 PCMs. The first period is April 20 to December 20, the period of the actual A/H1N1 outbreak during 2009. The second period is January 1 to April 19, the period prior to the 2009 A/H1N1 outbreak. For brevity, we refer to April 20 through December 20 as the outbreak period and January 1 through April 19 as the pre-outbreak period. Counterfactual simulations are conducted for each empirical PCM calculated from the ATUS and NHAPS data according to the SIR model. We provide more detail on the PCM methods in the supplementary material.

Bootstrap Method. We employ bootstrap techniques to calculate confidence intervals around simulation results because of the nonlinearity in the SIR model and lack of closed form solution. The fixed-effects regression yields a parameter estimate of β_1 with distribution $N(\beta_1, \sigma_1^2)$. We simulate the SIR model 1000 times. In each simulation, the avoidance parameter is drawn from the distribution $N\left(\hat{\beta}_1, \hat{\sigma}_1^2\right)$ where the parameters are estimated in the regression. We report the 2.5 and 97.5 percentiles of the 1000 simulated results.

The construction of the PCM and subsequent epidemic simulations are deterministic. However, the ATUS is a stratified random sample of the U.S. population and thus sampling error exists. We employ a bootstrap approach to estimate standard errors for each element of the contact matrices as well as epidemic simulation outcomes (e.g., cumulative cases). We generate 1,000 resamples of the U.S. population over 2003-2012. We sample at the respondent-level and not the activity level so if an individual was selected, his or her entire 24-hour diary was used. We calculate the probabilistic contact matrices for each year (2003-2012) for each of the 1,000 bootstrap samples and simulate an epidemic based on each sample. We then construct 95% confidence intervals around the model outcomes using the 2.5 and 97.5 percentiles of the estimates from the 1,000 replications. Supplementary Figure S2 illustrates the ATUS sampling error propagating through the epidemic simulation.

Results

Additional Time at Home. We find that individuals increased their time spent at home in response to CDC confirmed cases by a statistically significant amount (Table 1, table with results for all control variables provided in Supplementary Table S2). We quantify sample uncertainty through 95% bootstrap confidence intervals and model uncertainty by estimating several model specifications for robustness. We use the point estimate from model three that controls for state and month fixed effects, which suggests that people spent 2.38 additional minutes at home for every 1,000 CDC confirmed cases with a 95% confidence interval of (0.278, 4.48). The avoidance response is statistically significant across model specifications (models 1-3 and the two additional models in the supplemental material).

Our estimates suggest that the average person in the population spent an additional 22.11 minutes at home, with 95% confidence interval of (5.76, 33.57), at the peak of the epidemic

when the CDC reported 9,734 new cases in a single week. This effect size is an average across the entire population, with some individuals likely spending substantially more time at home and others spending less. For comparison, the average individual spent 34.47 (10.56, 58.38) additional minutes at home during extreme weather events, e.g., snowstorms. Furthermore, we find no evidence that historically sensitive groups (persons 65+ years old and parents with children) engage in additional avoidance despite spending substantially more time at home regardless of the epidemic state (Supplementary Table S2 models (4) and (5)).

We illustrate the potential epidemiological significance of this avoidance response by comparing two simulated epidemics: 1) when individuals fail to respond to the epidemic, and maintain a constant level of contacts – the standard assumption in most epidemic studies; and 2) when individuals reduce their time spent in public by 2.38 (0.278, 4.48) minutes per thousand confirmed cases (model 3 in Table 1). As prevalence of the infection rises, individuals that engage in avoidance gradually shift time in public into their household relative to the no avoidance case. This is true despite the fact that the small fraction of infected households, in our model always less than 5% at a point in time, may not be safer than public. The substitution of relatively safe household time, for time in public, drives a wedge between the simulated epidemics with (dashed) and without (solid) avoidance behaviour (Figure 1). At the peak of the simulated epidemic on day 92, individuals spend 22.11 minutes less in public with a bootstrap confidence interval of (5.76, 33.57), which reduces the peak prevalence by 31.20% (7.46%, 48.27%) from 4.22% of the population without a behavioural avoidance model to 2.90% (2.18%, 3.90%) with the avoidance model. As the epidemic wanes, so too does the incentive to stay at home. The daily incidence during the last half of the infection is greater when individuals avoid infection early on because more of the population remains susceptible and avoidance behaviour

fades. Nevertheless, by the end of the epidemic avoidance behaviour reduces the attack rate by a proportional 13% (2.99%, 21.03%) from 50% of the population without avoidance behaviour to 42.22% (37.46%, 48.29%) with avoidance, which is comparable to the simulated attack rate of A/H1N1 reported in Towers and Chowell [28].

Adjusting Contact Patterns. Simulation results using the empirical PCMs suggest that individuals modified contacts during the A/H1N1 outbreak reducing transmission rates and the impact of the epidemic relative to the average across all years other than 2009 (Figure 2). During the outbreak period, peak prevalence falls from 2.83% (2.68%, 2.98%) of the population in the average simulation to 1.98% (1.49%, 2.52%) in 2009, a 30% decrease. This reduction in peak prevalence is comparable to the 28% decrease found by simulating the avoidance behaviour based on the regression results. The smaller epidemic in 2009 translates into a lower attack rate, 32.72% (28.91%, 36.64%), compared to the average case 38.60% (37.73%, 39.47%). Moreover, an epidemic based on the contact patterns in 2009 pre-outbreak period appears similar to the average across all years. The attack rate in 2009 is 39.96% (37.40%, 47.42%) while the attack rate in the average across all years scenario is 38.71% (37.56%, 39.84%). These results and sensitivity analyses are contained in Supplementary Table S3.

Figure 3 graphically presents a one-tailed test, and associated p-values, for the difference between simulated attack rate in the pre-outbreak period and outbreak period. 2009 is the only year with statistically significant positive difference at α=0.05, 5.88 percentage points with p-value of 0.02.

The simulation model based on the empirical PCMs disaggregates the population by age and household size. We find less behavioural heterogeneity across household size than across

age groups. The household size heterogeneity that is present indicates that single-person households and large households of five or more suffer lower attack rates than households of 2-4 individuals. This finding is consistent with Cauchemez et al. [45] and may indicate that members of larger households spend more time at home. We provide more detail in the supplementary material.

The regression model indicates that many factors influence how people spend their time. Because the empirical PCMs simply reflect probabilistic interactions between age groups, alone they do not reveal the mechanism responsible for the change in behaviour. However, the combined evidence from the simulations based on the empirical PCMs and the regression model that does identify avoidance behaviour as an epidemiologically significant factor suggests that people changed their behaviour during the A/H1N1 epidemic in a way that measurably affected epidemiological dynamics.

Discussion

We measure the extent to which Americans engaged in voluntary avoidance behaviour during the 2009 A/H1N1 epidemic and show that such behaviour is of epidemiologically meaningful magnitude. Our estimates derive from a national time-use survey conducted by the U.S. Census Bureau nearly every day since 2003. We show that individuals spent on average 2.38 (0.278, 4.48) additional minutes at home for every 1,000 CDC confirmed cases during the 2009 A/H1N1 epidemic. Moreover, simulations based on empirical contact matrices suggest that individuals adjusted behaviour in a manner that reduced contact time during the outbreak period in 2009 unlike the pre-outbreak period in 2009, or the outbreak period in any other year. These results

are further supported by recent anecdotal evidence of avoidance behaviour in American churches frequented by immigrants from west Africa during the ongoing Ebola crisis [47].

Social distancing policies are an important public health tool for controlling epidemics, particularly during the early stages. However, the social and economic costs of social distancing policies imply that public health officials must weigh the costs and benefits of such measures to determine when to employ the social distancing policy [13,40]. Most research on social distancing policy attributes all behavioural response to the policy [48,49]. Our results provide empirical evidence that individuals respond to disease risk with behavioural shifts that are likely sufficiently large to influence the course of an epidemic. Therefore, emergency response plans - based on epidemic forecasts that neglect self-directed behavioural response - may prescribe costly measures to reduce transmission rates that would occur because of voluntary avoidance behaviour. It is also possible that poorly planned social distancing policies could counteract innate responses and "crowd out" avoidance responses [7,50]. Furthermore, retrospective analysis of social distancing policies may appear beneficial when compared to a baseline forecasts that did not account for feedbacks and adaptive avoidance behaviour. Including accurate self-directed adaptive behavioural responses in baseline models is imperative for those models to accurately guide public health policy.

We have described the epidemiological implications of uniform avoidance behaviour in a population. However, our control variables in the fixed effects regressions make it possible to consider heterogeneity in avoidance behaviour across subpopulations (e.g., age classes). This heterogeneity may reflect variation in risk perception (the cost of avoidance) as well as the benefits of contact. The development of a heterogeneous mixing model with heterogeneous

avoidance behaviour is beyond the scope of this work, but could have important implications for targeted public health policies [48–50].

Feedbacks between human behaviours and biological processes are challenging to measure, but are receiving great attention in public health, ecology, earth systems, and sustainability. While many human-environmental feedbacks likely exist, the strength of these feedbacks is an empirical question. Our measure of the strength of feedback between adaptive human behaviour and epidemiological conditions suggests meaningful feedbacks that have implications for the study of infectious disease, the costs of infectious diseases, and public health policy.

Authors Contributions

JB conducted simulations and drafted the manuscript; NK and QG conducted the regression analysis and helped draft the manuscript; EF oversaw the study and helped draft the manuscript. All authors gave final approval for publication.

Acknowledgments This publication was made possible by grant number 1R01GM100471-01 from the National Institute of General Medical Sciences (NIGMS) at the National Institutes of Health and NSF. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS. This work was also funded by NSF grant 1414374 as part of the joint NSF-NIH-USDA Ecology and Evolution of Infectious Diseases program. We thank Simon Levin for helpful comments. We acknowledge the Yale University Omega High Performance Computing cluster.

References

1. Bell, D., Nicoll, A., Fukuda, K., Horby, P. & Monto, A. 2006 Nonpharmaceutical Interventions for Pandemic Influenza, National and Community Measures. *Emerg Infect Dis* **12**, 88–94. (doi:10.3201/eid1201.051371)

2. Ciofi degli Atti, M. L., Merler, S., Rizzo, C., Ajelli, M., Massari, M., Manfredi, P., Furlanello, C., Scalia Tomba, G. & Iannelli, M. 2008 Mitigation Measures for Pandemic Influenza in Italy: An Individual Based Model Considering Different Scenarios. *PLoS ONE* **3**, e1790. (doi:10.1371/journal.pone.0001790)

3. Ferguson, N. M., Cummings, D. A. T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S. & Burke, D. S. 2005 Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* **437**, 209–214. (doi:10.1038/nature04017)

4. Kelso, J. K., Milne, G. J. & Kelly, H. 2009 Simulation suggests that rapid activation of social distancing can arrest epidemic development due to a novel strain of influenza. *BMC Public Health* **9**, 117. (doi:10.1186/1471-2458-9-117)

5. Smith, R. D., Keogh-Brown, M. R., Barnett, T. & Tait, J. 2009 The economy-wide impact of pandemic influenza on the UK: a computable general equilibrium modelling experiment. *BMJ* **339**. (doi:10.1136/bmj.b4571)

6. Baum, N. M., Jacobson, P. D. & Goold, S. D. 2009 'Listen to the People': Public Deliberation About Social Distancing Measures in a Pandemic. *The American Journal of Bioethics* **9**, 4–14. (doi:10.1080/15265160903197531)

7. Maharaj, S. & Kleczkowski, A. 2012 Controlling epidemic spread by social distancing: Do it well or not at all. *BMC Public Health* **12**, 679. (doi:10.1186/1471-2458-12-679)

8. Fenichel, E. P., Kuminoff, N. V. & Chowell, G. 2013 Skip the Trip: Air Travelers' Behavioral Responses to Pandemic Influenza. *PLoS ONE* **8**, e58249. (doi:10.1371/journal.pone.0058249)

9. Ferguson, N. 2007 Capturing human behaviour. *Nature* **446**, 733–733. (doi:10.1038/446733a)

10. Fenichel, E. P. et al. 2011 Adaptive human behavior in epidemiological models. *PNAS* **108**, 6306–6311. (doi:10.1073/pnas.1011250108)

11. Chen, F., Jiang, M., Rabidoux, S. & Robinson, S. 2011 Public avoidance and epidemics: Insights from an economic model. *Journal of Theoretical Biology* **278**, 107–119. (doi:10.1016/j.jtbi.2011.03.007)

12. Morin, B. R., Fenichel, E. P. & Castillo-Chavez, C. 2013 Sir Dynamics with Economically Driven Contact Rates. *Natural Resource Modeling* **26**, 505–525. (doi:10.1111/nrm.12011)

13. Perrings, C. et al. 2014 Merging Economics and Epidemiology to Improve the Prediction and Management of Infectious Disease. *EcoHealth* , 1–12. (doi:10.1007/s10393-014-0963-6)

14. Funk, S., Salathé, M. & Jansen, V. A. A. 2010 Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface* **7**, 1247–1256. (doi:10.1098/rsif.2010.0142)

15. Funk, S., Gilad, E., Watkins, C. & Jansen, V. A. A. 2009 The spread of awareness and its impact on epidemic outbreaks. *PNAS* **106**, 6872–6877. (doi:10.1073/pnas.0810762106)

16. Perra, N., Balcan, D., Gonçalves, B. & Vespignani, A. 2011 Towards a Characterization of Behavior-Disease Models. *PLoS ONE* **6**, e23084. (doi:10.1371/journal.pone.0023084)

17. Lau, J. T., Griffiths, S., Choi, K. C. & Tsui, H. Y. 2010 Avoidance behaviors and negative psychological responses in the general population in the initial stage of the H1N1 pandemic in Hong Kong. *BMC Infectious Diseases* **10**, 139. (doi:10.1186/1471-2334-10-139)

18. Wong, L. P. & Sam, I.-C. 2011 Behavioral responses to the influenza A(H1N1) outbreak in Malaysia. *J Behav Med* **34**, 23–31. (doi:10.1007/s10865-010-9283-7)

19. SteelFisher, G. K., Blendon, R. J., Ward, J. R., Rapoport, R., Kahn, E. B. & Kohl, K. S. 2012 Public response to the 2009 influenza A H1N1 pandemic: a polling study in five countries. *The Lancet Infectious Diseases* **12**, 845–850. (doi:10.1016/S1473-3099(12)70206-2)

20. Greer, A. L. 2013 Can informal social distancing interventions minimize demand for antiviral treatment during a severe pandemic? *BMC Public Health* **13**, 669. (doi:10.1186/1471-2458-13-669)

21. Caley, P., Philips, D. J. & McCracken, K. 2008 Quantifying social distancing arising from pandemic influenza. *J. R. Soc. Interface* **5**, 631–639. (doi:10.1098/rsif.2007.1197)

22. Cauchemez, S., Valleron, A.-J., Boëlle, P.-Y., Flahault, A. & Ferguson, N. M. 2008 Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature* **452**, 750–754. (doi:10.1038/nature06732)

23. He, D., Dushoff, J., Day, T., Ma, J. & Earn, D. J. D. 2013 Inferring the causes of the three waves of the 1918 influenza pandemic in England and Wales. *Proc. R. Soc. B* **280**, 20131345. (doi:10.1098/rspb.2013.1345)

24. Fraser, C. et al. 2009 Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Science* **324**, 1557–1561. (doi:10.1126/science.1176062)

25. Chowell, G., Bertozzi, S. M. & Miller, M. A. 2009 Severe Respiratory Disease Concurrent with H1N1 Influenza Circulation. *The New England Journal of Medicine* , 674–679. (doi:DOI: 10.1056/NEJMoa0904023)

26. Nishiura, H., CastilloChavez, C., Safan, M. & Chowell, G. 2009 Transmission potential of the new influenza A(H1N1) virus and its age-specificity in Japan. *Euro surveillance : bulletin européen sur les maladies transmissibles = European communicable disease bulletin* **14**.

27. Bajardi, P., Poletto, C., Ramasco, J. J., Tizzoni, M., Colizza, V. & Vespignani, A. 2011 Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic. *PLoS ONE* **6**, e16591. (doi:10.1371/journal.pone.0016591)

28. Towers, S. & Chowell, G. 2012 Impact of weekday social contact patterns on the modeling of influenza transmission, and determination of the influenza latent period. *Journal of Theoretical Biology* **312**, 87–95. (doi:10.1016/j.jtbi.2012.07.023)

29. Springborn, M., Chowell, G., MacLaughlan, M. & Fenichel, E. 2015 Accounting for Behavioral Responses during a Flu Epidemic Using Home Television Viewing. *BMC Infectious Diseases*

30. United States Department of Labor. Bureau of Labor Statistics 2012 American Time Use Survey.

31. Klepeis, N. E., Nelson, W. C., Ott, W. R., Robinson, J. P., Tsang, A. M., Switzer, P., Behar, J. V., Hern, S. C. & Engelmann, W. H. 2001 The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *Journal of exposure analysis and environmental epidemiology* **11**, 231–252.

32. Brammer, L. et al. 2011 Surveillance for Influenza during the 2009 Influenza A (H1N1) Pandemic–United States, April 2009–March 2010. *Clin Infect Dis.* **52**, S27–S35. (doi:10.1093/cid/ciq009)

33. NCDC 2014 The National Oceanic and Atmospheric Administration National Climatic Data Center Storm Events Database.

34. Kermack, W. O. & McKendrick, A. G. 1929 Contributions to the Mathematical Theory of Epidemics, Part 1. *Proceedings of the Royal Society, London Series A* , 700–721.

35. Anderson, R. M. & May, R. M. 1985 Age-Related Changes in the Rate of Disease Transmission: Implications for the Design of Vaccination Programmes. *The Journal of Hygiene* **94**, 365–436.

36. Brauer, F. & Castillo-Chávez, C. 2013 *Mathematical Models for Communicable Diseases*. Philadelphia: Society for Industrial and Applied Mathematics.

37. Longini, I. M., Koopman, J. S., Monto, A. S. & Fox, J. P. 1982 Estimating Household and Community Transmission Parameters for Influenza. *Am. J. Epidemiol.* **115**, 736–751.

38. House, T. & Keeling, M. J. 2009 Household structure and infectious disease transmission. *Epidemiology & Infection* **137**, 654–661. (doi:10.1017/S0950268808001416)

39. Cauchemez, S. et al. 2011 Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *PNAS* **108**, 2825–2830. (doi:10.1073/pnas.1008895108)

40. Brown, S. T. et al. 2011 Would school closure for the 2009 H1N1 influenza epidemic have been worth the cost?: a computational simulation of Pennsylvania. *BMC Public Health* **11**, 353. (doi:10.1186/1471-2458-11-353)

41. Chowell, G., Viboud, C., Wang, X., Bertozzi, S. M. & Miller, M. A. 2009 Adaptive Vaccination Strategies to Mitigate Pandemic Influenza: Mexico as a Case Study. *PLoS ONE* **4**, e8164. (doi:10.1371/journal.pone.0008164)

42. Longini, I. M., Halloran, M. E., Nizam, A. & Yang, Y. 2004 Containing Pandemic Influenza with Antiviral Agents. *Am. J. Epidemiol.* **159**, 623–633. (doi:10.1093/aje/kwh092)

43. Galvani, A. P., Reluga, T. C. & Chapman, G. B. 2007 Long-standing influenza vaccination policy is in accord with individual self-interest but not with the utilitarian optimum. *PNAS* **104**, 5692–5697. (doi:10.1073/pnas.0606774104)

44. Lau, L. L. H., Nishiura, H., Kelly, H., Ip, D. K. M., Leung, G. M. & Cowling, B. J. 2012 Household transmission of 2009 pandemic influenza A(H1N1): a systematic review and meta-analysis. *Epidemiology* **23**, 531–542. (doi:10.1097/EDE.0b013e31825588b8)

45. Cauchemez, S., Donnelly, C. A., Reed, C., Ghani, A. C., Fraser, C., Kent, C. K., Finelli, L. & Ferguson, N. M. 2009 Household Transmission of 2009 Pandemic Influenza A (H1N1) Virus in the United States. *New England Journal of Medicine* **361**, 2619–2627. (doi:10.1056/NEJMoa0905498)

46. Zagheni, E., Billari, F. C., Manfredi, P., Melegaro, A., Mossong, J. & Edmunds, W. J. 2008 Using Time-Use Data to Parameterize Models for the Spread of Close-Contact Infectious Diseases. *Am. J. Epidemiol.* **168**, 1082–1090. (doi:10.1093/aje/kwn220)

47. Kelto, A. 2015 How Ebola Took A Toll On One American Church. *NPR.org*.

48. Bootsma, M. C. J. & Ferguson, N. M. 2007 The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *PNAS* **104**, 7588–7593. (doi:10.1073/pnas.0611071104)

49. House, T. et al. 2011 Modelling the impact of local reactive school closures on critical care provision during an influenza pandemic. *Proc. R. Soc. B* **278**, 2753–2760. (doi:10.1098/rspb.2010.2688)

50. Fenichel, E. P. 2013 Economic considerations for social distancing and behavioral based policies during an epidemic. *Journal of Health Economics* **32**, 440–451. (doi:10.1016/j.jhealeco.2013.01.002)

51. Roberts, M. G. & Heesterbeek, J. a. P. 2003 A new method for estimating the effort required to control an infectious disease. *Proc. R. Soc. Lond. B* **270**, 1359–1364. (doi:10.1098/rspb.2003.2339)

52. Shuai, Z., Heesterbeek, J. a. P. & Driessche, P. van den 2013 Extending the type reproduction number to infectious disease control targeting contacts between types. *J. Math. Biol.* **67**, 1067–1082. (doi:10.1007/s00285-012-0579-9)

53. Roberts, M. G. & Heesterbeek, J. a. P. 2013 Characterizing the next-generation matrix and basic reproduction number in ecological epidemiology. *J. Math. Biol.* **66**, 1045–1064. (doi:10.1007/s00285-012-0602-1)

Tables and Figures

Table 1. Regression results for time spent at home.

State fixed effects	x		x
Month fixed effects		x	x
Month*State fixed effects			x
	coefficient estimates		
	(1)	(2)	(3)
CDC reported cases	1.663*	2.378**	2.379**
	(0.944)	(1.057)	(1.072)
Google media index	-22.33	-15.02	-17.66
	(18.98)	(19.61)	(20.10)
Extreme Weather	30.88***	33.54***	34.47***
	(11.97)	(11.95)	(12.20)

*p<.1, **p<.05, ***p<.01

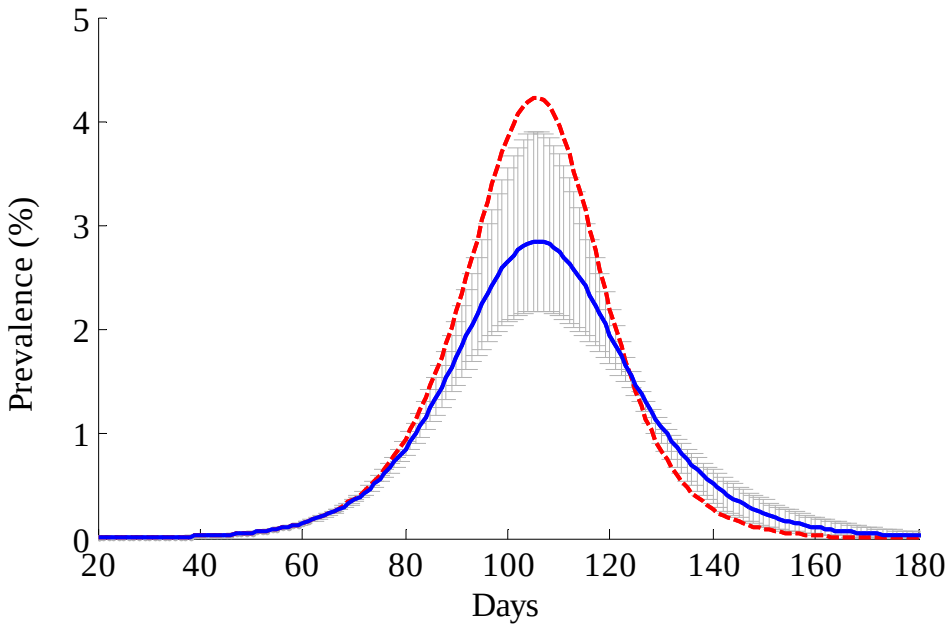


Figure 1. Simulated epidemic curves. The solid (blue solid) line indicates epidemic with avoidance behaviour and the dashed (red dashed) line without and the gray bars represent 95% bootstrapped confidence intervals. The simulations are based on the estimated 2.38 minute reduction in time spent in public per thousand cases. The susceptible population is 4.1×10^6, the

recovery rate is three days, and the infectivity parameter is chosen such that the basic reproduction number is 1.4.

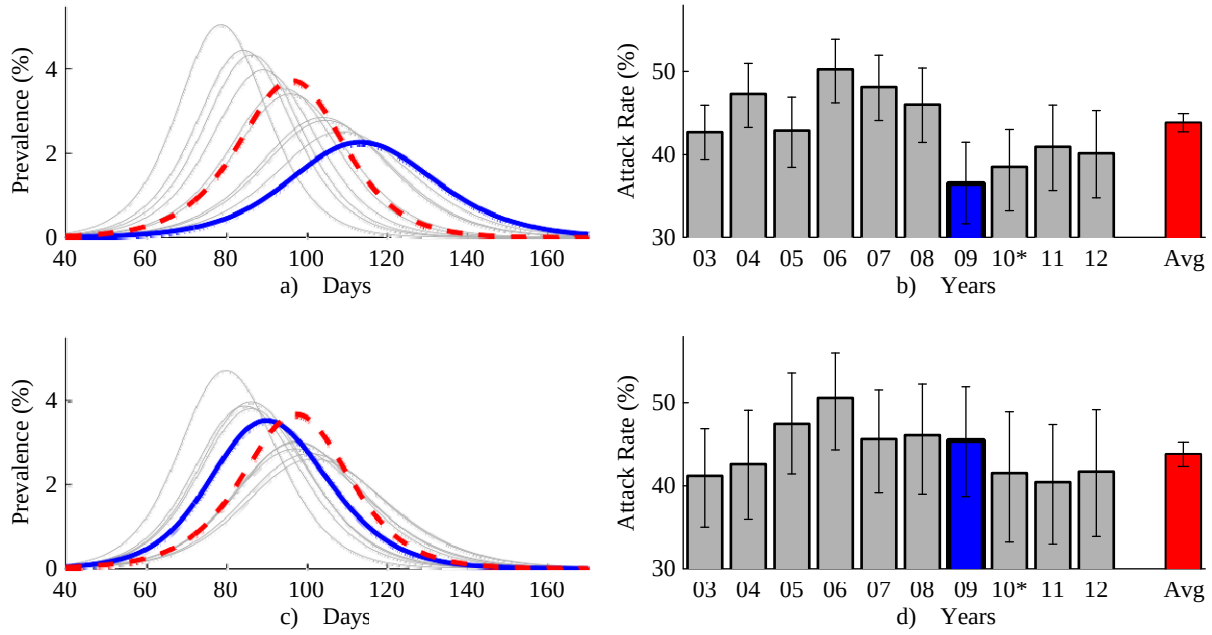


Figure 2. **Simulated epidemic curves and cumulative cases** based on contact matrices during the epidemic period April 20 to December 20 (panels a and b) and the pre-epidemic period January 1 to April 19 (panel c and d). Panels a and c illustrate the percent of the population infected by day. Solid (blue) lines are 2009, dashed (red) lines are non-2009 average, and thin (gray) lines are non-2009 by year. Panels b and d contain the cumulative number of infected and recovered individuals at the end of the epidemic where the bars indicate 95% bootstrap confidence intervals. The asterisk indicates that the pandemic was not declared over until June 23, 2010 even though very few cases were reported in 2010.

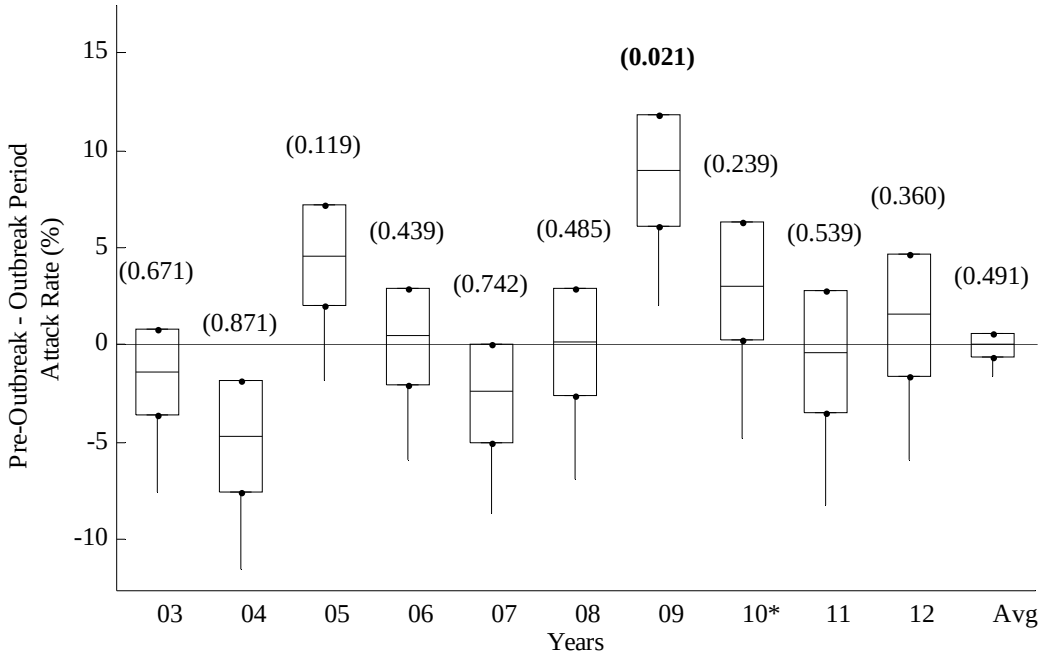


Figure 3. **Comparing the epidemic and pre-epidemic simulations.** The difference between simulated cumulative cases from the pre-epidemic period January 1 to April 19 (Fig 2 panel D) and the epidemic period April 20 to December 20 (Fig 2 panel B) with 95% bootstrapped lower confidence bound represented by bars and p-value of a one-sided hypothesis test with a null that the difference is less than or equal to zero. We reject the null hypothesis in 2009.

Supplementary Materials:

Additional Data Description.

Extreme weather is used to control for additional time spent at home to avoid weather rather than flu risk. The following conditions are used to classify extreme weather events: significant weather phenomena causing loss of life, injuries, significant property damage and/or disruption to commerce; rare and unusual weather phenomena garnering media attention; and other significant meteorological events.

We collect Google search data to represent a subjective measure of risk. The measure is subjective because it is based on news coverage and general interest in the population rather than true risk as measured by the disease prevalence. Google presents the information in the form of a normalised index from 0-100 where 100 represents the most search volume over a given time interval. We collect Google Trends for the search terms “swine flu” or “h1n1” in the U.S. only between the dates April 19, 2009 to April 04, 2010. Google search volume peaks (100) during the week April 19-26, 2009 (Supplementary Figure S1). We use Google Trends to measure media attention rather than Google Flu Trends, a proprietary prediction tool offered by Google, because our objective is to measure the population’s discussion about infection risk, which is not necessarily correlated with actual prevalence.

Regression Model. In addition to regression models 1-3 described in the main text, we include two additional models in the supplementary material. Model 4 includes a dummy for age 65 and older individuals and an interaction of that dummy variable with the number of laboratory-confirmed cases to test whether avoidance behaviour was stronger among more sensitive individuals. Model 5 includes a similar set of terms for parents spending time with children. The ATUS does not survey individuals less than 15 years; however, respondents specify which

family members were present during each activity and their demographic information. We use parents' time at home with children as a proxy for the behaviour of children less than 15 years old. Model 6 includes a third set of interactions between parents with children and Saturday to further investigate avoidance behavior on the weekend.

Regression Results.

We present the full set of parameter estimates for all variables used in the regression model in Supplementary Table S2. The constant coefficient represents the time spent at home by the baseline individual represented by the omitted categories or binary outcomes in the regression (e.g., less than college degree, other race, female, not extreme weather). The baseline respondent spends 335.1 (model 1) to 357.9 (model 2) minutes at home on Monday in January in Alabama (excluding sleep and other personal activities). The coefficients on laboratory confirmed cases ($CASES$) is statistically different from zero in models 1-6, which suggests that the result is robust to various model specifications.

In all models, individuals with a spouse or partner spend more time at home. People spend more time at home as their age increases, but at a decreasing rate. Similarly, people tend to spend more time at home during the weekend (40-60 minutes). We find that ill and disabled individuals do not spend any statistically significant additional time at home. We do find that students, employed individuals, and those with a college degree and beyond spend much less time at home. The estimated coefficient on the indicator for extreme weather conditions indicates that individuals spent between 30 and 34 additional minutes at home during an adverse weather event. We would expect this effect to provide an approximate upper bound on avoidance behaviour to H1N1 because it occurs in situations where individuals have a very strong incentive

to stay home. That is, a severe storm effectively forces people to stay indoors whereas they are not forced to stay indoors by health risk.

The $MEDIA$ coefficient indicates that individuals spent less time at home during the peak of the Google flu-related searches. This is not entirely surprising given the divergence between public attention to the epidemic and media attention suggested by figure 2. However, this effect is not statistically different from zero at the 10% level. In contrast, the statistically significant coefficients on $CASES$ suggest that people respond to the objective measure of risk.

We expect seniors to be more responsive to flu risk because they have a higher probability of becoming seriously ill or dying, and because they are less likely to be time constrained by work and caring for young children. The $ELDER$ variable in model 4 indicates that individuals 65 years and older spend 32 additional minutes at home with a 95% confidence interval of [12.20, 51.08] regardless of infection risk. The interaction of $ELDER$ with $CASES$ suggest that seniors spent an additional 14 [-27.65, 56.23] minutes at home in response to increased flu risk during the week with the most confirmed H1N1 cases. However, the interaction is not statistically significant, due in part to the smaller number of seniors in the ATUS sample, and perhaps because these individuals already spend more time at home creating a small margin on which to adjust behavior.

Model 5 adds an interaction between $CASES$ and whether the respondent spent time with their own children at some point during the day. Similar to seniors, young children are more susceptible to negative health consequences of contracting the flu. Therefore, we would expect their parents to have a greater incentive to take actions to reduce their probability of contracting the flu. This intuition is not mirrored by our results. The estimated coefficient is negative. This may indicate that parents with young children have less flexibility to adjust their schedules, that

they did not perceive the health risks to be heightened, or that having children leads people to behaviour fatalistically [1] with respect to flu.

Model 6 interacts parent's activities with children and $CASES$ along with a dummy variable Saturday. People spend an average of 54.57 minutes at home on Saturday. Parents with children spend an average of 16.58 additional minutes at home on Saturday whereas individuals without children spend 12.64 minutes less at home. Importantly, interacting parent's avoidance response with Saturday causes the negative coefficient to be statistically insignificant while the population average avoidance response is 3.345 per 1000 cases and is statistically significant. Therefore, we find no conclusive evidence of additional avoidance behavior by parents with children.

Our model accounts for approximately 20% of the variation in time spent at home ($R^2 = 0.20$ in models 3-6).The magnitude and statistical significance of estimates on employment, weather, and day of week demonstrate that our covariates capture certain critical factors that influence time-use patterns.

Epidemic Simulations Adjustment

Laboratory confirmed cases represent only a fraction of total cases suspected in the population [2]. Reed et al. [2] estimated that 43,677 laboratory confirmed cases through July 23, 2009 represented between 1.8 and 5.7 million cases in the population. Using the most conservative estimate, we assume that 2.4% of simulated cases are confirmed by laboratory testing. Furthermore, confirmed cases are measured at the national level so we extrapolate the simulated number of infected individuals (based on Phoenix MSA) to the national level by multiplying daily prevalence by 3.47 (9,734 national cases/2,800 Phoenix MSA cases during the peak of the epidemic in the third week of October [3]). The product of the proportion of laboratory-

confirmed cases and the proportion of national cases to those reported in Phoenix MSA yields $\phi = 8.33\%$.

Sensitivity of Household Contact Scalar (α)

We investigate the sensitivity of our results to the assumption of $\alpha = 1$ made on the basis of empirical studies of household transmission during the 2009 A/H1N1 outbreak. We re-simulate the model assuming α =3 and α =5. We keep the infectivity parameter constant in each simulation δ =1.4 10^3, which prevents the model from generating the number of cases observed in the data. Nevertheless, the simulations provide intuition about the impact of increasingly infectious household contacts. The peak prevalence and attack rates from the avoidance model and the empirical PCM model are reported in Supplementary Table S3.

As α increases, time spent engaging in household contacts is more infectious in households with at least one infected person. If α=3, household contacts are three times more infectious per minute than public contacts. The attack rate rises in the simulations with and without avoidance to over 80% of the population, but the percent change between the two simulations with and without avoidance falls by 7.9 percentage points. If α=5, the attack rates with and without avoidance reach nearly 100% of the population. Under such a severe epidemic, most households become infectious, which exacerbates the effect of increased infectivity of household contacts.

Probabilistic Contact Matrix (PCM).

We compare simulated epidemic outcomes in each year 2003-2012 using an age-household size model based on PCMs derived from ATUS data. Following Zagheni et al. [4], let $\ell \in L$ denote a public location reported in the ATUS. Public locations include: School; Restaurant or bar; Place of worship; Grocery store; Mall and other stores; Gym and Health club; Personal Services e.g.,

laundry, beauty salon; Public building, library and post office; Public bus; Subway and train; Boat and ferry; Airplane; Office building and bank, Factory/Manufacturing center, Hospital and health care center, Workplace not elsewhere classified; Public location not elsewhere classified. Let $m \in [0,1440]$ denote an index of time (minutes in a 24 hour period). Let $k \in P$ denote an age-household size group (P ={0-4, 5-12, 13-17, 18-24, 25-49, 50-64, 65+}× {1,2,3,4,5}) within which all members mix homogeneously. Then $g_k^{\ell m}$ is the number of individuals in age group k at location ℓ at minute m. The exposure of group k to group $j \neq k$ is proportional to the size of each population at location ℓ at minute m and is given by

$$z_{kj}^{\ell m} = g_k^{\ell m} \times \frac{g_j^{\ell m}}{\sum_l^K g_l^{\ell m}}$$

The total time of public exposure on an average day between groups k and j at location ℓ is $z_{kj}^{\ell} = \sum_{m=0}^{1440} z_{kj}^{\ell m}$ and the total time of exposure between groups k and j at all public locations is $z_{kj} = \sum_{\ell \in L} \sum_{m=0}^{1440} z_{kj}^{\ell m}$. The exposure matrix, $\boldsymbol{Z}$, has dimensions $(K, K, L, 1440)$ and is symmetric in (K, K)-space, by construction. The structure of the exposure matrix implies a symmetric contact matrices such that $z_{kj} = p_{kj} w_j = p_{jk} w_k$ where p_{kj} is the contact rate between population k and individual j, and w_j is the population of group j (i.e., a contact between groups k and j is necessarily a contact between groups j and k) [5]. However, the relevant measure is the contact between an individual in group j and the whole of population of group k. Therefore, we divide the columns of the matrix $\boldsymbol{Z}$ by population w_j, given by the sample weights in the ATUS, to form the PCM. A column in the PCM sums to the total number of minutes an average individual in group j spends in public. All contact matrices and the code written to generate them in Matlab are available upon request from the authors.

The ATUS surveys a subset of respondents to the Current Population Survey (CPS), which interviews only individuals older than 15 years. However, ATUS asks respondents to report information on all family members and whether any family members were present during an activity. We use the activity and location information on children with their parents (primary ATUS respondents) in public to construct a partial measurement of children's public contact patterns. Children younger than 13 years old are rarely in public microenvironments without an adult (see Supplementary Figure S2). School contacts are largely omitted from this calculation because parents generally leave their children at school for long periods during the weekday. The National Health and Activity Patterns Survey (NHAPS) was conducted from 1992-1994 and collected time use information on children by microenvironment (e.g., school, home) at minute resolution. We use the NHAPS to construct children's contact patterns at school. We then derive an index for children's time at school based on 15-17 year olds from the ATUS and use that index to rescale school contact patterns from the NHAPS. The base of the index is the average time spent at school from 2003-2012 omitting 2009. The index is constructed separately for the pre-outbreak and outbreak periods. Indices are created for each bootstrap sample and used to adjust the contacts at school in each bootstrap simulation.

Supplementary Figures S3 depict the household and public PCMs during the outbreak period averaged across 2003-2012 as heat maps with associated population distributions directly above. Panel a) illustrates the contact patterns by age and household size in the household environment. Segmenting the population by household size allows our population-based model to capture intra-household transmission but importantly, prevents inter-household transmission in the home environment. The household PCM clearly shows interaction between children and their parents or guardians in households of three or more, and strong affinity mixing between

adults in two-person households. The contact with children in single-person households is a result of children visiting parents in cases where the children do not live with the parent surveyed regularly. The population density just above the heat map shows that this population is small, so that this contact time has almost no impact on the simulation results.

Panel b) illustrates that contact time in public are heterogeneous across age groups. Adults primarily contact other adults in public, especially those in smaller households. Children, who are generally in households with three or more people, experience most of their contact time with other children at school. Together, the household and public PCMs capture a rich set of epidemiologically meaningful contact patterns.

Sampling Uncertainty

Like any sampling method, the ATUS samples include uncertainty, which we propagate through our simulations. In the main text, we illustrate sample uncertainty with 95% confidence bars around cumulative attack rate in Figure 2. In Supplementary Figure S4, we illustrate the sampling uncertainty associated with the 2009 ATUS survey data during the outbreak period (panel A) and pre-outbreak period (panel B) on each day of the epidemic (daily frequency). During the outbreak period (panel A), the peak prevalence of the epidemic curve based on the average PCM lies outside of the 95% confidence interval of the 2009 epidemic curve. This figure suggests that our result is not due to sampling uncertainty in the ATUS. In contrast, panel B illustrates that the 2009 and average epidemic curves are statistically indistinguishable during the pre-outbreak period.

Supplementary Figure S4 shows that the sampling errors are asymmetric near the peak prevalence. The pattern of uncertainty is consistent across both simulations because of the mechanics of the model and the source of the uncertainty. We hold the contact matrix constant

during each simulation. When a particular Monte Carlo sample from the ATUS yields a contact matrix with high contact minutes (i.e., respondents that happen to collocate more frequently), the simulated epidemic will spread through the population faster and the peak prevalence will occur earlier. This approach accounts for sampling uncertainty in the ATUS but not the stochastic variation in day-to-day contacts within the data. We find that stochastic variation in day-to-day contacts is very minimal and yields simulation results nearly identical to those based on the PCM held constant across the simulation. The Matlab code and bootstrap simulation data are available upon request.

References

1. Kremer, M. 1996 Integrating Behavioral Choice into Epidemiological Models of AIDS. *The Quarterly Journal of Economics* **111**, 549–573. (doi:10.2307/2946687)

2. Reed, C., Angulo, F. J., Swerdlow, D. L., Lipsitch, M., Meltzer, M. I., Jernigan, D. & Finelli, L. 2009 Estimates of the Prevalence of Pandemic (H1N1) 2009, United States, April–July 2009. *Emerging Infectious Diseases* **15**, 2004–2007. (doi:10.3201/eid1512.091413)

3. Arizona Department of Health Services In press. Arizona - Weekly Influenza Summary MMWR Week 31-32 (August 1, 2010 - August 14, 2010).

4. Zagheni, E., Billari, F. C., Manfredi, P., Melegaro, A., Mossong, J. & Edmunds, W. J. 2008 Using Time-Use Data to Parameterize Models for the Spread of Close-Contact Infectious Diseases. *Am. J. Epidemiol.* **168**, 1082–1090. (doi:10.1093/aje/kwn220)

5. Wallinga, J., Teunis, P. & Kretzschmar, M. 2006 Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents. *Am. J. Epidemiol.* **164**, 936–944. (doi:10.1093/aje/kwj317)

Supplementary Table S1. Summary Statistics

	Mean	St. Dev.	min	max
Time at Home (Minutes)	412.445	238.824	0	1400
CASES (1000)	0.729	1.660	0	9.735
Age (Years)	45.867	17.512	0	85
MEDIA (index 0-1)	0.027	0.087	0	1

Frequency Tables

Regression Variables		PCM Variables	
Male	43.1%	Age:	
Race:		0-4	7%
White	81.4%	5-12	11%
Black	13.3%	13-17	7%
Other Race	5.3%	18-24	9%
Hispanic	13.1%	25-49	35%
Partner Present	52.8%	50-64	18%
Education:		65+	13%
Less than College	38.4%	Household Size:	
College Degree	48.9%	One	11%
Advanced Degree	12.7%	Two	26%
Current Student	11.0%	Three	18%
Currently Employed	63.1%	Four	23%
Manual Labor	7.8%	Five or more	20%
Ill/Disabled	0.3%		
Extreme Weather	1.3%		
Survey Day:			
Holiday	1.6%		
Monday	9.3%		
Tuesday	9.2%		
Wednesday	9.4%		
Thursday	9.6%		
Friday	9.9%		
Saturday	26.0%		
Sunday	26.6%		

Supplementary Table S2. Fixed effect regression results for time spent at home.

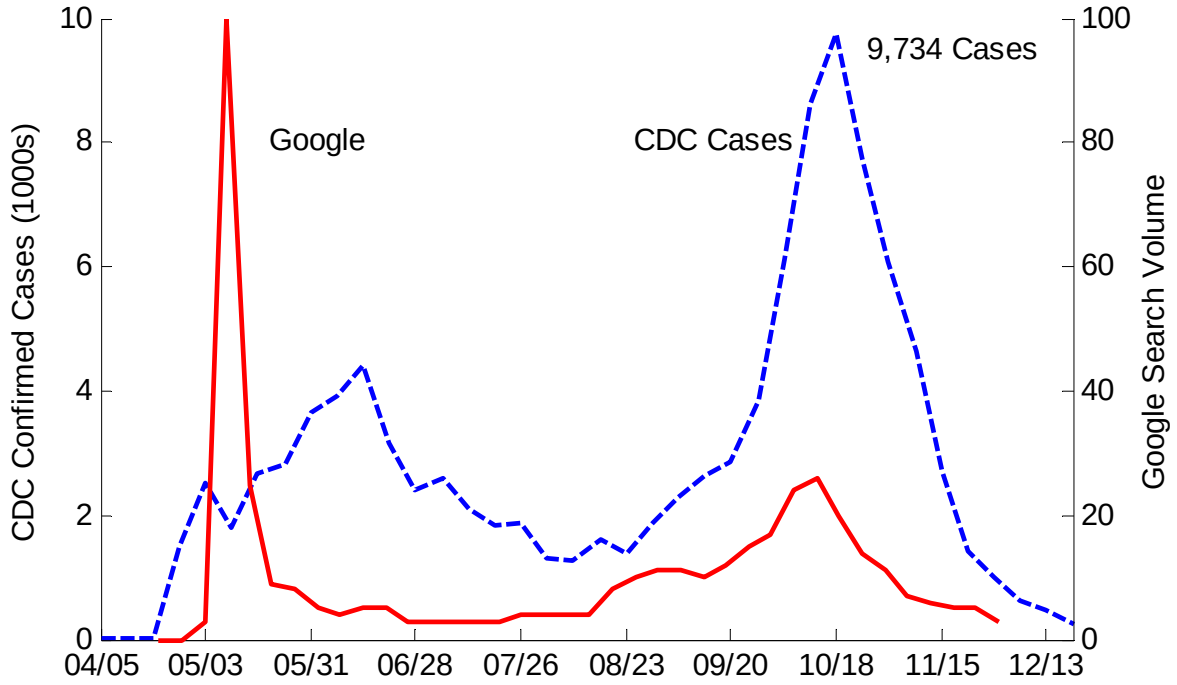
State fixed effects	x		x	x	x	x
Month fixed effects		x	x	x	x	x
Month*State fixed effects			x	x	x	x
	Coefficient Estimates of fixed effects models					
	(1)	(2)	(3)	(4)	(5)	(6)
CASES (1000)	1.663*	2.378**	2.379**	2.114*	3.357***	3.345***
	(0.944)	(1.057)	(1.072)	(1.129)	(1.219)	(1.218)
Elderly (65+ yrs)				31.64***		
				(7.924)		
CASES *Elderly				1.468		
				(2.198)		
Child (<18 yrs)					15.50***	16.58***
					(3.490)	(3.927)
Child*Saturday						-12.64**
						(6.369)
No Child*Saturday						16.58***
						(3.927)
CASES*Child					-3.263*	-2.084
					(1.690)	(1.867)
CASES*Child*Saturday						-3.809
						(3.084)
MEDIA	-22.33	-15.02	-17.66	-17.17	-17.06	-16.80
	(18.98)	(19.61)	(20.10)	(20.07)	(20.12)	(20.11)
Extreme Weather	30.88***	33.54***	34.47***	34.36***	34.46***	34.52***
	(11.97)	(11.95)	(12.20)	(12.20)	(12.21)	(12.21)
Holiday	5.931	5.240	6.910	7.029	5.237	5.757
	(11.93)	(11.96)	(11.93)	(11.93)	(11.96)	(11.95)
Tuesday	-6.286	-6.418	-5.172	-5.202	-5.240	-5.194
	(5.805)	(5.792)	(5.875)	(5.871)	(5.867)	(5.867)
Wednesday	-11.51**	-11.67**	-10.72*	-10.71*	-10.76*	-10.66*
	(5.738)	(5.730)	(5.818)	(5.817)	(5.809)	(5.809)
Thursday	-19.65***	-19.25***	-18.20***	-18.38***	-18.46***	-18.47***
	(5.696)	(5.680)	(5.786)	(5.784)	(5.779)	(5.778)
Friday	-16.08***	-16.08***	-14.60**	-14.66**	-15.16***	-15.20***
	(5.728)	(5.708)	(5.797)	(5.796)	(5.795)	(5.796)
Saturday	41.23***	41.49***	41.27***	41.25***	39.76***	54.57***
	(4.972)	(4.957)	(5.024)	(5.022)	(5.042)	(6.711)
Sunday	59.69***	60.07***	59.70***	59.66***	57.87***	57.59***
	(4.835)	(4.818)	(4.892)	(4.891)	(4.919)	(4.927)

Supplementary Table S2 cont. Fixed effect regression results for time spent at home.

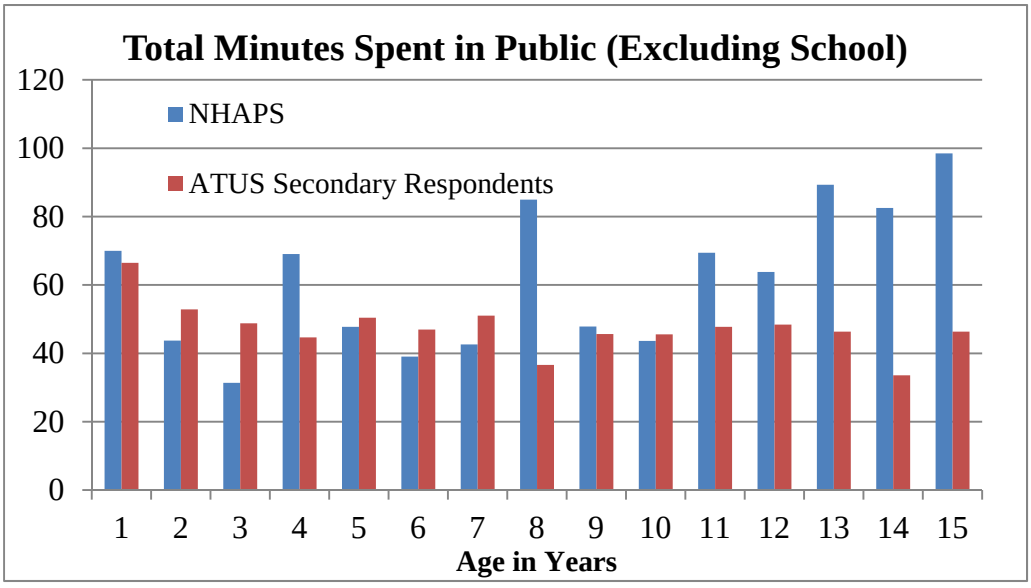
State fixed effects	x		x	x	x	x
Month fixed effects		x	x	x	x	x
Month*State fixed effects			x	x	x	x
	Coefficient Estimates of fixed effects models					
	(1)	(2)	(3)	(4)	(5)	(6)
College Degree	-5.780*	-5.493*	-5.707*	-6.141*	-5.773*	-5.797*
	(3.123)	(3.101)	(3.152)	(3.154)	(3.151)	(3.151)
Advanced Degree	-20.75***	-19.84***	-21.74***	-22.22***	-21.96***	-21.97***
	(4.563)	(4.527)	(4.612)	(4.612)	(4.612)	(4.612)
Hispanic	-13.37***	-14.79***	-13.83***	-14.12***	-14.66***	-14.71***
	(4.222)	(3.899)	(4.266)	(4.266)	(4.271)	(4.271)
White	-6.797	-4.978	-7.507	-7.540	-7.311	-7.349
	(6.251)	(6.080)	(6.278)	(6.277)	(6.276)	(6.276)
Black	-24.03***	-23.43***	-25.36***	-25.29***	-24.86***	-24.93***
	(7.280)	(7.049)	(7.308)	(7.306)	(7.305)	(7.306)
Ill/Disabled	-8.718	-6.528	-12.42	-10.82	-13.10	-13.17
	(23.10)	(23.02)	(22.81)	(22.89)	(22.89)	(22.88)
Age	4.229***	4.191***	4.234***	5.637***	4.109***	4.105***
	(0.484)	(0.483)	(0.488)	(0.582)	(0.490)	(0.490)
Age2	-0.019***	-0.018***	-0.019***	-0.038***	-0.016***	-0.016***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)
Male	-18.70***	-18.73***	-18.86***	-18.76***	-17.26***	-17.18***
	(2.787)	(2.783)	(2.809)	(2.808)	(2.833)	(2.834)
Employed	-144.5***	-144.6***	-144.5***	-143.8***	-143.7***	-143.7***
	(3.367)	(3.358)	(3.391)	(3.396)	(3.396)	(3.395)
Manual Labor	-0.496	-1.209	-1.331	-1.311	-1.326	-1.407
	(5.306)	(5.290)	(5.394)	(5.395)	(5.393)	(5.391)
Student	-34.36***	-34.70***	-35.69***	-33.42***	-33.79***	-33.85***
	(5.267)	(5.282)	(5.368)	(5.390)	(5.375)	(5.376)
Partner	42.99***	43.04***	43.20***	42.27***	40.33***	40.38***
	(2.942)	(2.935)	(2.961)	(2.968)	(3.049)	(3.050)
Constant	335.1***	355.6***	357.9***	333.7***	355.0***	354.9***
	(17.15)	(13.47)	(47.07)	(47.45)	(47.17)	(47.26)
R2	0.176	0.176	0.195	0.196	0.196	0.196
Observations	27,091	27,091	27,091	27,091	27,091	27,091

Supplementary Table S3. Sensitivity of results from homogeneous mixing model and heterogeneous age-household size model. All peak prevalence and attack rate results are in percent.

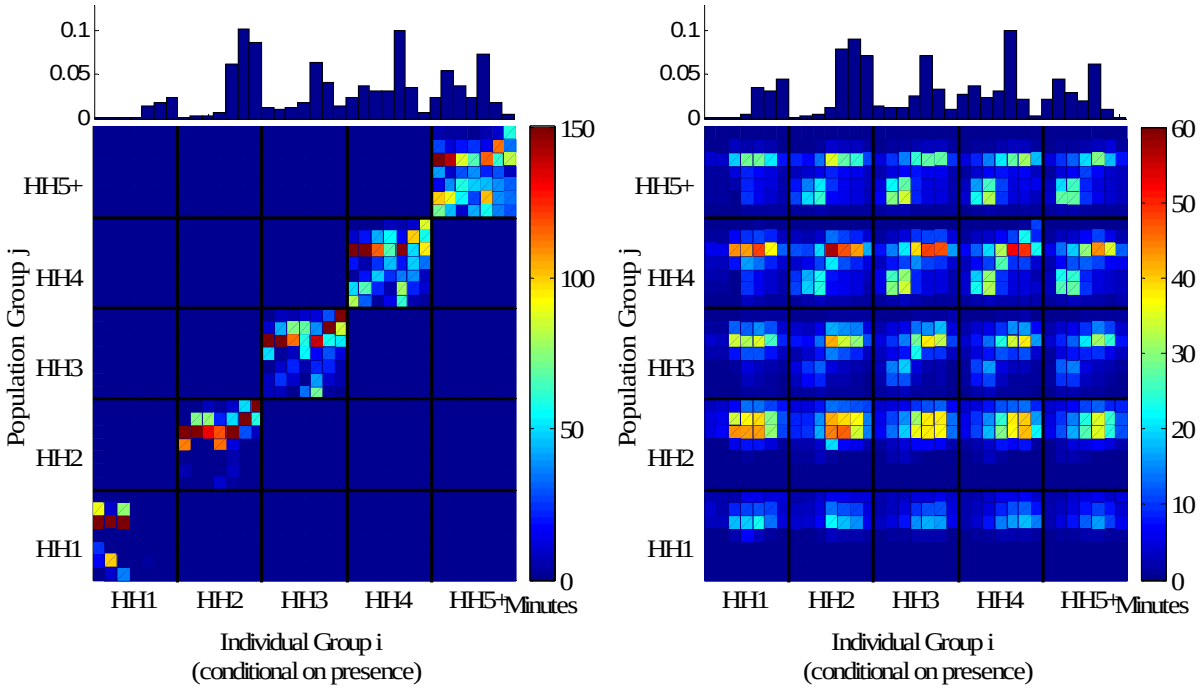
Result	$\alpha = 1$	95% CI	$\alpha = 3$	95% CI	$\alpha = 5$	95% CI
Homogeneous mixing model						
Peak Prev. w/ Av.	2.9	(2.18, 3.91)	14.93	(7.57, 19.82)	55.25	(46.51, 62.09)
Peak Prev. w/o Av.	4.22	-	20.34	-	44.16	-
% Change Peak Prev	31.2	(7.46, 48.27)	26.58	(2.56, 62.76)	-25.1	(-40.59, -5.30)
Atk. Rt. w/ Av.	42.22	(37.46, 48.29)	82.66	(65.49, 89.45)	99.98	(99.81, 100.00)
Atk. Rt. w/o Av.	50	-	89.5	-	99.59	-
% Change Atk. Rt.	15.55	(3.42, 25.07)	7.65	(0.06, 26.83)	-0.39	(-0.41, -0.22)
Age-household size model						
Pre-Outbreak Period						
Peak Prev. 2009	4.16	(2.82, 5.70)	17.94	(12.46, 22.49)	38.3	(35.26, 40.73)
Peak Prev. Avg	3.71	(3.41, 4.01)	16.61	(15.24, 17.92)	38.18	(37.06, 39.20)
Attack Rate 2009	45.45	(38.66, 51.93)	83.14	(75.02, 87.82)	94.58	(94.02, 95.13)
Attack Rate Avg	43.8	(42.32, 45.23)	82.27	(80.41, 83.93)	94.94	(94.76, 95.13)
Outbreak Period						
Peak Prev. 2009	2.52	(1.81, 3.32)	11.33	(6.83, 15.92)	35.72	(33.12, 38.03)
Peak Prev. Avg	3.76	(3.53, 4.00)	16.76	(15.87, 17.70)	38.37	(37.36, 39.21)
Attack Rate 2009	36.48	(31.62, 41.45)	71.86	(59.67, 81.07)	94.02	(93.43, 94.52)
Attack Rate Avg	43.81	(42.70, 44.91)	82.46	(81.21, 83.65)	95	(94.88, 95.13)
Difference 2009	8.97	(2.03, ~)	11.29	(1.65, ~)	0.57	(-0.06, ~)



Supplementary Figure S1. Weekly number of CDC reported cases of H1N1 (dashed blue) and Google search volume index (solid red) over the outbreak period April 20, 2009 – December 20, 2009. The horizontal axis displays dates in one week increments with the starting date in the format "mm/dd".

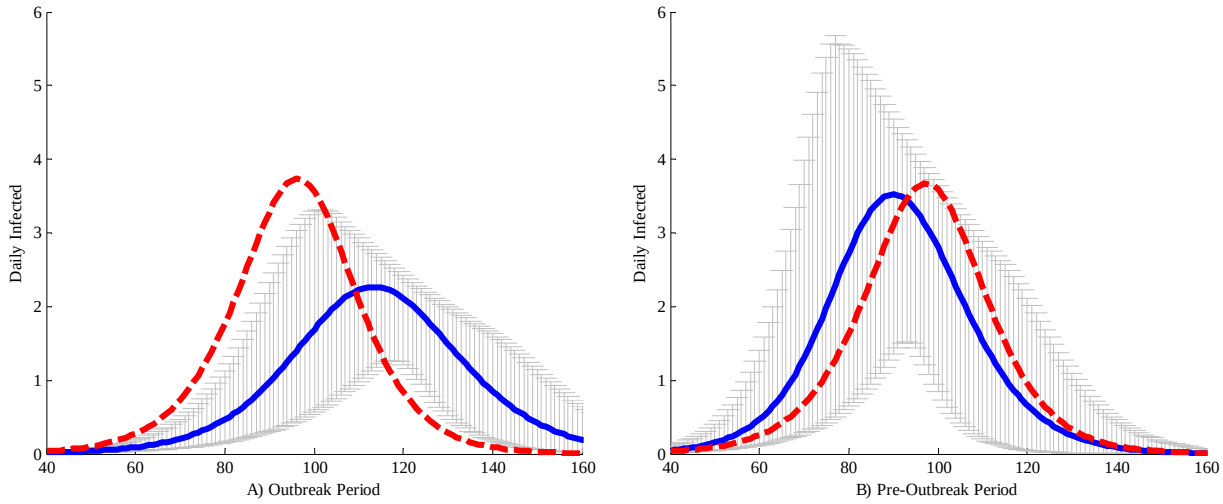


Supplementary Figure S2. Comparison of time spent in public using the NHAPS data from 1992-1994 and children reported with primary ATUS respondents (age 15 and older).



a) Household **b) Public**

Supplementary Figure S3. Household and public PCMs with empirical population distributions above. Dark lines denote the five household size groups. Each household size category consists of the seven age groups for a total of 35 demographically meaningful groups (P ={0-4, 5-12, 13-17, 18-24, 25-49, 50-64, 65+}×{1,2,3,4,5+}), note these bins are not of equal width. A cell represents the number of contact minutes an individual in group i interacts with the population of group j. The vertical sum of cells in a single column is equal to an individual's total contact minutes. The vertical axis of the population distribution (bar chart) is group i percent of the population. Note, these are the transpose of $\boldsymbol{C}$ described in the main text.



Supplementary Figure S4. Simulated epidemic curves based on contact matrices for 2009 (solid blue) and the average of all non-2009 years (dashed red) during the outbreak period (panel A) and the pre-outbreak period (panel B). The bars represent the bootstrap 95% confidence interval of the 2009 curve.