Does Disease Cause Vaccination? Disease Outbreaks and Vaccination Response*

Emily Oster

Brown University and NBER

October 7, 2016

Abstract

Childhood vaccination rates have recently declined in the US due to parental fear of vaccines. I test whether disease outbreaks increase vaccination using a new dataset of county-level disease and vaccination data. A large pertussis outbreak in a county decreases the share of unvaccinated children entering kindergarten by 30% (1.2 percentage points). These responses do not reflect changes in the future disease risk. I argue these facts may reflect a model in which perceived risk of disease is influenced by whether a household is aware of any cases of disease. This suggests better promotion of outbreaks could enhance the response.

1 Introduction

Childhood vaccinations are a crucial input to disease prevention. In the period from 1920 through 1940, prior to vaccination, the incidence of pertussis in the US was 150 cases and 6 deaths per 100,000 people (Kutty et al, 2013). By the early 1990s, case counts had dropped to just 1 per 100,000 with typically fewer than 10 deaths per year across the country (Davis et al, 1992). These long run trends in disease reflect trends in vaccination. The current vaccination rate for pertussis in the US is around 95%.

Within the past fifteen years, however, vaccination rates in the US (and in many other developed countries) have declined. These declines are on average fairly small, but they are geographically concentrated, leaving some areas with quite low vaccination rates (Omer et al, 2006). The decline in vaccination rates has contributed to incidence of disease. Pertussis rates

*Min Kim, Motaz Al-Chanati and Kim Sarnoff provided outstanding research assistance. I thank Kate Baicker, Amitabh Chandra, Matt Notowidigdo, Brendan Nyhan, Ben Olken, Jesse Shapiro, Andrei Shleifer and Heidi Williams for helpful comments. All mistakes remain my own.

in the US have roughly doubled since 2000, with a corresponding increase in deaths, primarily among infants.[1] Maintaining and increasing vaccination rates in the face of these issues is an important goal for both policy-makers and for many practitioners (Yang and Silverman, 2015; Orenstein and Seib, 2014).

These declines in vaccination rates seem to largely reflect parental choice (Healy and Paulson, 2015; Glassser et al, 2016; Omer et al, 2012). In surveys, parents express fears about vaccine safety and efficacy, and skepticism that their child is at risk for vaccine-preventable diseases (Omer et al, 2009; Salmon et al, 2005).[2]

The role for parental choice suggests that policies which ensure *access* to vaccines are unlikely to be important in the US. Instead, US policy has focused on either (a) educating parents about vaccines or (b) changing school vaccination policies to make it more difficult to enroll unvaccinated children in school. Both policies have limitations. Educational campaigns are appealing in that they preserve autonomy, but evidence suggests they are largely ineffective (Nyhan et al, 2014; Sadaf et al, 2013). Changes in school vaccination polices are more effective (Sadaf et al, 2013), but they may be seen as heavy-handed (Constable, Blank and Caplan, 2014).

A good example of this trade-off appears in California. Vaccination rates in California have declined substantially from the mid-2000s.[3] This decline has occurred in spite of a health department focus on developing materials to encourage parents and doctors to vaccinate.[4] In the wake of a large measles outbreak at Disneyland in 2014-2015, California passed a law completely eliminating personal belief exemptions to vaccination for school-age children (Mello et al, 2015). Early figures suggest this may already have had a large effect (LA Times Editorial Board, 2016). However, the backlash to this bill has been extreme, with lawsuit challenges and concerns about parents choosing to home-school their children to avoid vaccination (Mello et al, 2015; Siepel, 2016). At equal effectiveness, using information to convince parents they want to vaccinate their children should be preferred; at currently estimated effectiveness, legal remedies seem like the only effective option.

In a simple cost-benefit framework these policies act on different levers. Education campaigns seek to increase the perceived benefit of vaccines or lower their perceived cost, holding the true values of these constant. School vaccination policies increase the *actual* cost of not vaccinating. Recognizing this points to two explanations for the differing effectiveness.

---

[1]See evidence from the CDC: http://www.cdc.gov/pertussis/images/incidence-graph-age.jpg

[2]The concerns around vaccine risks include both debunked links with autism (Mnookin, 2011) as well as more general concerns about "toxins" and "metals" being present in vaccines.

[3]For reports on vaccination in California see http://www.cdph.ca.gov/programs/immunize/pages/immunizationlevels.aspx. Herd immunity for pertussis occurs around 92-94%.

[4]See, for example, the CA Health Department Immunization Branch page at https://www.cdph.ca.gov/PROGRAMS/IMMUNIZE/Pages/HealthProfessionals.aspx.

One is that non-vaccination represents a rational, fully informed parental decision. Given the small disease risk and the discomfort of vaccines, it is possible that a fully informed parent may choose not to vaccinate. Information will not affect this choice, although raising the actual cost of not vaccinating may do so. The other possibility is that parents *are* uninformed and some type of information provision is effective, but existing campaigns are not optimally designed.

Evidence that vaccination rates respond to data which is uninformative about the true risks or benefits of vaccines would be evidence for the latter theory. In this paper I provide evidence on this by estimating the response of vaccination rates to disease outbreaks. I find that outbreaks of pertussis increase pertussis vaccination rates, despite the fact that an outbreak in a single year is not predictive of outbreaks in future years. Cross-area evidence shows that the behavior of health departments may enhance or deflate this response. This is encouraging news for a role for education in changing vaccination rates.

The core contribution of this paper is to bring together new data on vaccination and disease. I combine (1) county-year data on vaccination among kindergarten children; (2) county-year data on outbreaks of disease; (3) Google search data on disease and vaccination related terms and (4) a survey of local health department. I use these data to generate a series of facts about the correlates of vaccination and the response of vaccination to information.

I show first that, in the cross section, vaccination rates vary with demographics and also correlate with Google searches for vaccination-related terms. I then demonstrate that within a county over time, pertussis vaccination rates among kindergarten entrants are higher in the year after a pertussis outbreak.[5] The effect size suggests that approximately 30% of vaccine-hesitant parents vaccinate after a large outbreak. These estimates survive inclusion of county-specific trends, and I show that future outbreaks do not impact current vaccination rates. This suggests the estimated impacts are causal.

I perform two analyses showing these results do not seem to reflect changes in the true pertussis risk. First, within a county, current outbreaks do not predict future outbreaks. Second, the effects are concave: the per-case impact on vaccination is larger for the first cases of the disease relative to subsequent cases. This suggests vaccination changes are not a function of true risk changes.

However, these behaviors do seem to be instrumental and to reflect information seeking. I show the impact of outbreaks is limited to pertussis vaccinations – it does not spill over to other vaccines – suggesting an instrumental response.Further, I find that outbreaks increase searches for "pertussis" and "pertussis vaccination". This suggest information about outbreaks

[5] I focus on pertussis since this is the disease which is the most common and, therefore, for which there is the most variation over time. In an appendix I show similar, but less precise, impacts for measles.

does spread among the population.

I use a survey of health departments to explore whether the response to disease varies systematically across areas. I find, first, that the response to outbreaks is much larger in states in which responses are coordinated by the state rather than individual counties. Second, I find that within states with county autonomy, counties with systematic media notification have larger responses.

The evidence presented shows that people are responsive to local cases of disease, even holding constant the true risk of disease. In Section 4 I discuss which theory might explain this behavior. I argue that the facts are inconsistent with a model with full information about disease risk. I further suggest that the patterns in the data - the concave response to cases and the fact that education about vaccine costs does not seem to have an impact - may be consistent with a particular model of limited information in which salience plays a role (as in, e.g., Bordalo, Gennaioli and Shleifer, 2012). Under this model, individuals perceive disease risk to be zero unless they are aware of cases of disease, in which case they infer some positive probability. This model has the advantage that it can rationalize the local response to disease outbreaks alongside the *lack* of response to information on vaccine safety seen in other data.

Under this model, information can play a crucial role in changing vaccination rates, but the key information is about disease risk not about vaccine risk. The evidence on area-level variation suggests that counties may take better or worse advantage of outbreaks to improve vaccination rates. To make this more vivid, the final section of the paper takes up the question of whether these responses are large enough to limit vaccination declines. I find that if the response is limited to only the age cohort entering kindergarten, it makes relatively little difference to vaccination rates or disease rates. If these responses occur across all children, the effect is more substantial. In general, this analysis requires a number of additional assumptions so should be taken with caution.

This paper contributes to the general literature on interventions to increase vaccination rates, well summarized by Sadaf et al (2013). Perhaps the most closely related paper on this topic, Wolf et al (2014), did not find an impact of a single large outbreak on vaccination of young children in Washington State. This differs markedly from my findings, which may be because I am able to look at a large number of outbreaks, or because my data focuses on school-age children and not younger children. More supportive of these findings, Cacciatore, Nowak and Evans (2016) find that the 2014-2015 Disneyland measles outbreak did increase confidence in vaccines, although they are not able to look at actual vaccination behavior.

The paper also contributes to the literature on economic epidemiology (e.g. Philipson, 1996; Geoffard and Philipson, 1996; Geoffard and Philipson, 1997; Philipson, 2000; Kremer, 1996; Adda, 2016). Philipson (1996) finds that states with more measles in the late 1980s also have higher vaccination rates. The direction of the effect is consistent with my findings, although the argument in Philipson (1996) is that the response is a rational one, given that in this time period measles was much more common.

The rest of the paper is organized as follows. Section 2 discusses background. Section 3 presents the data and empirical strategy. Section 4 presents the primary results on the changes in vaccination behavior in response to outbreaks. Section 5 presents a model which may explain the findings and Section 6 discusses the role of these responses in mitigating vaccine declines. Section 7 concludes.

2 Background

Literature on vaccination in the US and other developed countries has generally documented declines in vaccination in recent years. This change appears to be related to increasing parental decisions to avoid vaccinations, due to concerns about either vaccine efficacy or vaccine safety (Omer et al, 2006). In the late 1990s a now-discredited study was published in the Lancet suggesting a link between vaccines and autism. Although the paper was shown to be fraudulent, the effects on vaccination rates have been long-lasting (Mnookin, 2011). In the wake of declining vaccination rates there are serious concerns about outbreaks of vaccine-preventable diseases. Low levels of vaccination tend to be concentrated in particular areas, and these areas have been shown to have increased disease risk (Omer et al, 2008). This suggests vaccination rates in many areas are falling below what would be required to maintain herd immunity.

There are at least two common policy responses to declining vaccination rates. One is education. The other is changing rules for school vaccination exemptions.

Educational Interventions There is some existing evidence on the response of vaccination to education campaigns.

Nyhan et al (2014) uses a randomized trial design to test the impact of messaging about vaccine safety and vaccine-preventable diseases on vaccination intentions. The authors test four messages - a message refuting the link between MMR vaccine and autism, a generic text with disease information, images of children who are sick with vaccine-preventable disease and a dramatic narrative about an infant in danger from disease. Many of these had perverse effects; the narrative about an infant in danger, for example, increased fears about vaccine side effects. The education on the MMR/autism link did succeed in decreasing belief in this link, but did not change intention to vaccinate. None of these messages changed vaccination intentions.

Nyhan and Reifler (2015) evaluate an intervention designed to correct the misconception that the flu vaccine can give you the flu. They find that messaging can successfully decrease belief that this is the case, but their messaging also *decreases* the intent to vaccinate, more so among those who have an initially poor view of vaccination. Williams et al (2013) report on a small intervention among vaccine-hesitant parents of infants. Treatment group parents were shown a video designed to improve attitudes about vaccination and given some handouts on vaccines and on how to find better information on the internet. The authors, again, find that parental attitudes improve but they do not see any change in realized vaccination behavior.

Finally, Sadaf et al (2013) review the literature (as of 2012) on the impact of educational interventions and conclude there is not much evidence to support the efficacy of particular educational interventions.

Somewhat in contrast to this finding, and more closely related to this paper, Horne et al (2015) find that they *are* able to systematically alter vaccination attitudes with an intervention which focuses on the risk of communicable diseases. They suggest that this type of focus - rather than a focus on vaccine safety - may be more effective. This study is run on a convenience sample of Mechanical Turk participants and does not include anything about vaccination behavior or intentions, making it largely suggestive, but interesting in light of the hypothesis evaluated in this paper.

School Exemption Policies Schools in the US require students to be vaccinated. At school entry parents must either show proof that their child is vaccinated or request an exemption from the policy. All states allow exemptions on medical grounds but there is substantial variation across states in how difficult it is to obtain vaccine exemptions for non-medical reasons. Some states offer only religious exemptions, and others permit "personal belief" exemptions as well.

Omer et al (2006) show in the period from 2001 to 2004 states with relatively easily granted exemptions had a larger increase in the rate of non-medical exemptions than states which make the process more difficult. Between 1991 and 2004 the non-medical exemption rate doubled (from 1.25% to 2.5%) in states with easy exemptions and stayed constant in those with harder exemption policies. More recent evidence (Stadlin et al, 2012) suggests that even *medical* exemption rates may respond to state policy differences. In general, as noted by Sadaf et al (2013) in their review, these policies do seem to impact vaccination rates.

3 Data and Empirical Strategy

3.1 Data

This paper uses four primary data sets. The first is on vaccination rates, the second on disease outbreaks, the third is on Google searches and the fourth is from a survey of health departments.

The data on vaccination rates comes from individual states. The goal was to collect county-level vaccination data from as many states as possible. In some cases, states do not collect their own data on vaccinations, instead relying on the National Immunization Survey. For states which do have their own data collection, data came in one of two forms. In some cases data comes from annual school surveys, aggregated to the county level. In others, states used immunization registries. In the case of the latter, only a subset of the registries are mandatory. Optional registries tend to have quite poor coverage.

For data quality reasons, I use data from states that either have a mandatory registry or provided data from school reports.[6] I use the vaccination data at kindergarten entry because it provides consistent data for the largest number of locations, and focus on pertussis since this is the only illness with a significant number of outbreaks. In an appendix I will show results for older children and for measles vaccinations. Summary statistics for the states used in the analysis appear in Panel A of Table 1. The years of coverage for states varies. I have the longest time series for California, from 1991 to 2011. The shortest time series is for Missouri, with coverage only in 2011. Vaccination rates are generally high. In these data they are lowest in Michigan, which may reflect the fact that this is the only state where we use data from a registry (although the registry is mandatory). All results are robust to excluding Michigan.

The second data set covers disease outbreaks by county over time. These data are available from the CDC through the National Notifiable Diseases Surveillance System (NNDSS). The NNDSS is a nation-wide collaboration, run by the CDC, for public health departments at various levels (state, local) to share information about a set of notifiable diseases, of which pertussis is one. Reporting of these diseases is judicially mandated. The data provides counts, by county-year, of disease cases. It is likely that these figures are an understatement of total cases, especially for pertussis, but I expect them to be correlated with the true counts.

Merging these with county-level population data produces disease rates. Panel B of Table 1 summarizes the rates of pertussis by state. Pertussis cases per 100,000 people range from 2 to just over 8 per year.

[6] The school reports will include private schools but will not include students who are home schooled. Home schooling accounts for only about 3% of children so any bias from this is likely to be small.

The third data set is from Google trends. I focus on three categories of searches. The first are searches for terms related to pertussis ("pertussis" or "whooping cough"), the second are searches related to the vaccines ("pertussis vaccination") and the third are searches related to vaccine risks (either "vaccine injury" and related terms or "autism and vaccines"). The full list of search terms in each category appears in Table B.1 in the Appendix.

An issue in constructing the Google trends data is that the data are subject to privacy thresholds. It is not possible to generate data from rare search terms. To get around this, I use a technique from Stephens-Davidowitz (2014). In broad terms, this involves searching for the term of interest along with another common word (for example, "joke" or "sponge") and then searching for the common word alone and subtracting the two. Details of the implementation appear in Appendix B.

I use two sets of Google data. First, I use DMA-level measures which I average over the entire period from 2004 to 2015. These data indicate which areas have the overall highest interest in a particular term over this period. These data do not adjust for domain-specific search volume so I also collect data on searches for non-vaccination-related health terms (cancer, diabetes) and generate area-level residuals with respect to these terms.

Second, I use Google trends at the state-month level in the estimation of search response to disease outbreaks. These are merged with disease outbreak data at the state-month level from the NNDSS system.[7]

The fourth dataset used comes from a survey of state and county health departments. Focusing on the 12 states for which I have vaccination data, we first attempted to contact state health departments to learn about how outbreak response was coordinated. In four of the states these responses are centrally coordinated; in 8 they are not. Our first analysis will separate these groups.

Within the 8 states without central coordination we attempted to contact all county health departments to learn about their procedures including, crucially, whether they systematically notify the media about outbreaks. There are 828 counties in this set, of which we attempted to contact the 584 which ever had a case of pertussis in our data. We were able to contact 54% of counties.[8] 24% of these report a systematic procedure of notifying the media.

[7] In the period prior to 2011 there is an excess mass of reported cases in December, seemingly due to a policy of listing all cases with unknown timing as occurring in the last week of the year. I will therefore drop December from this monthly analysis.

[8] Reasons for lack of contact include an inability to find contact information for the health department, being unable to get a response after many contacts or, in two cases, a refusal to answer questions.

3.2 Empirical Strategy

This paper uses a panel data strategy in which I estimate the impact of disease outbreaks controlling for location and time fixed effects.

Vaccination Behavior

The vaccination data is merged with the disease data at the county-year level. I run regressions of the following form

$$Vacc_{a,c,t} = \alpha + \beta(Disease_{c,t-1}) + \gamma_c + \lambda_t + \epsilon_{a,c,t}$$

where $Vacc_{a,c,t}$ is a measure of the vaccination rate for children of age a in county c in year t and $Disease_{c,t-1}$ is a measure of disease in county c in year $t-1$. I explore various specifications for the disease risk, including rate, counts of cases and groups of cases. The vaccination rate is specified on the interval between 0 and 1. I will also report, throughout, the impact as a share of the unvaccinated population. This can be interpreted as the share of vaccine-hesitant people whose behavior is changed by the independent variable.

In all cases the regression includes county and year fixed effects. Effectively, this asks whether - within a county over time - years with disease outbreaks are followed by years with higher vaccination rates.

It is important to be clear on what is driving any impacts we observe. Vaccination is measured at kindergarten entry and the disease outbreak is in the previous year. Children with up to date vaccinations would not be affected since by this age they will already have several pertussis vaccinations. Any effects must therefore be driven by catch up vaccinations.

In the analysis of local variation in effects I run similar regressions by state groups, or within state with interactions between outbreaks and media notification plans at the county level.

The coefficient β can be interpreted as causal if disease outbreaks are assigned randomly within a county over time. However, there is a potential reverse causality issue in this regression. If changes in vaccination rates at the cohort level influence disease outbreaks, then the coefficient on β will be biased downward. To see why, consider the following relationship:

$$Disease_{c,t-1} = \delta + \Psi(Vacc_{a-1,c,t-1}) + \phi_c + \eta_t + \nu_{c,t}$$

This posits the possibility that a low vaccination rate for a particular cohort could influence the rate of disease. We expect $\Psi \leq 0$, implying that when vaccination rates are higher, disease rates are lower.

There is a mechanical relationship between $Vacc_{a,c,t}$ and $Vacc_{a-1,c,t-1}$ because the cohort of age a in year t was aged $a-1$ in $t-1$. Put differently, these are the same children. Since children cannot become unvaccinated, this generates a mechanical relationship. This will bias β towards zero.

It is possible to adjust for this directly by estimating Ψ in the data. Appendix Table A.1 does this analysis. There is no evidence that Ψ differs from zero. In other words, within a county over time there is no relationship between the vaccination rate of this cohort and the disease rate. This is not surprising. Much of the variation in disease rate across counties is accounted for by the county fixed effects - which effectively capture the overall level of vaccination in each county. This suggest the downward bias in β is not large, if it is present at all.

Google Trends

The analysis of Google trends will follow a similar structure. I will estimate:

$$Google_{s,t} = \alpha + \beta_{\mathbf{1}}(Disease_{s,t}) + \beta_2(Disease_{s,t-1}) + ... + \beta_{12}(Disease_{s,t-11}) + \gamma_s + \lambda_t + \epsilon_{c,t}$$

Where s indexes state. The data in this case is at the monthly level and the analysis is contemporary. Google searches in a month are related to outbreaks in that month and outbreaks in previous months in the last year. This structure will allow me to estimate the immediate impact of outbreaks and also how it evolves over time.

4 Results: Vaccination Behavior and Disease Outbreaks

The first subsection below discusses the variation in pertussis vaccination across space and over time, and describes the cross-sectional correlates of vaccination. The second subsection provides the primary evidence on the response of vaccination to pertussis outbreaks. Finally, in Section 4.3 I discuss variation across space.

4.1 Correlates of Vaccination

Pertussis vaccination has declined in these data over time and varies substantially across space. Using regressions with county fixed effects - which are identified based on changes within a location over time - pertussis vaccination rates are down 4 percentage points from a high in the late 1990s. Further, although many counties have vaccination rates at or close to

100, there is a long tail of low vaccination rates. Seven-point-eight percent of county-years have pertussis vaccination rates below 80%, and 12% have rates below 90%. These areas of very low vaccination are especially at risk for disease outbreaks (Phadke et al, 2016). Details of variation over time and space are in Appendix Figures A.1 and A.2.

Table 2 estimates correlations between vaccination rates and both some basic demographics (drawn from the American Community Survey) and the DMA-level Google search volume on average over the 2004 - 2015 period. The demographics are income, education and race and the Google searches include the four search term groups discussed above. All regressions include state fixed effects and I estimate the impacts for both the average of the 2004/2005 period and the average of the 2010/2011 period (the last in my data).

Counties with more educated people have lower vaccination rates. Counties with higher income, holding education constant, have higher vaccination rates. This is consistent with the general perception that at least some of the resistance to vaccination comes from highly educated parents; if the income control is excluded, education remains negative, although the effect is smaller.

The Google search data reveals that vaccination rates are higher in areas where people show more interest in searching for information about pertussis, or about pertussis vaccination. However, vaccination rates are lower in areas where there is a greater intensity of search for terms related to the link between vaccines and autism. This is consistent with survey evidence suggesting that concerns about the negative consequences of vaccination are among the factors that drive vaccine hesitancy.

The magnitudes here are moderate. They suggest that a 1 standard deviation increase in searches for autism-vaccine links, for example, decreases the unvaccinated population by 8.5%. Moving from zero population with a college degree to everyone would decrease the unvaccinated population by 2.9%.

4.2 Response to Outbreaks

Figure 1 shows the estimated impact of pertussis outbreaks (grouped by the number of cases) in the county-year. The message of the graph is simple: more cases of pertussis lead to higher vaccination rates in the subsequent year. The largest outbreaks in the data - greater than the 95th percentile of county-years - increase the vaccination rate by 1.2 percentage point. This is 28% of the unvaccinated population.

Table 3 shows regression evidence corresponding to this figure. The figures in square brackets indicate the impact as a share of the unvaccinated population. The first column corresponds exactly to Figure 1. This analysis looks only at the count of cases, ignoring the fact that this count has very different implications for the rate depending on the population. Column (2) explores an alternative functional form, including a linear term in the number of cases and a term measuring the rate. The rate seems to dominate this regression. Column (3) interacts the case groups with a dummy for being in the bottom half of counties in terms of population.[9] The effects are larger for the smaller population areas, but they do not scale proportionally given the large differences in population across these groups. Below I return more specifically to the issue of functional form.

It is useful to estimate the size of the impact in terms of number of children vaccinated. Focusing on Column (1) of Table 3: for the average county, this predicts that observing about 5 cases (the mean in the 50-75th percentile group) prompts 20 new vaccinations among entering kindergartners. This is a lower bound if other ages also respond.

Columns (4) and (5) of Table 3 show two standard robustness checks to address the possibility of preexisting trends driving the results. Column (4) shows the regressions with county-specific trends included. Column (5) shows the impact of *future* cases of pertussis (in the following year) on the current vaccination rate. The results are not sensitive to county-specific trends, and we see no evidence that future cases drive current vaccination.

Finally, Column (6) adds a control for the state exemption policy; these data are based on data from Omer et al (2006). This effect is estimated off of states which change their policies during the course of the data. Exemption regimes are ranked from 1 to 3 with higher numbers indicating a less strict exemption policy. Including the exemption policy control does not impact the estimated effect of outbreaks, suggesting state policy changes are not driving this effect. Moreover, we can see the impact of a large outbreak is similar to a 1 point increase in the exemption policy.

Relationship between Current and Future Cases Crucial to the interpretation of these facts is the question of whether they reflect true changes in disease risk. Table 4 tests for a relationship between current and future pertussis cases, controlling for county and year fixed effects. It does not appear that more cases in a year predict more cases in the following year or the year after. If anything, most of the coefficients are negative[10]; this is likely a result of the cyclical nature of pertussis outbreaks. There is one positive and significant coefficient: a large outbreak in the current year does predict an increase in cases in the following year. The magnitude is much smaller than the cases in the current year, however, and this does not extend two years in the future. It seems likely this simply reflects some continuation of a large epidemic across years.

[9] There are no very large outbreaks in the smaller counties so this interaction is dropped.

[10] Based on the result in Table A.1 this does not seem likely to result from vaccination responses.

Impact of Lagged Cases The primary results focus on the impact of outbreaks in the year prior to kindergarten entry. This is motivated by the fact that this is the time at which vaccine-hesitant parents must either vaccinate their children or obtain an exemption. However, it is useful to look at the full time path of effects, including both leads and lags. These have different interpretations. Future outbreaks should not affect current vaccination behavior (see Column (5) of Table 3). Lagged outbreaks, however, may impact vaccines if they prompt parents to vaccinate children at the time of the outbreak *or* parents remember them when they are making decisions about kindergarten vaccines. In practice, based on data from the National Immunization Survey, vaccination rates (as measured here, with an indicator for any vaccination) do not increase much between 19 and 35 months of age. Given this, finding a significant impact of lagged outbreaks likely points to the role of memory.

Figures 2a and 2b graph coefficients on measures of the pertussis rate, or large outbreaks, over a number of years leading up to and following kindergarten entry. The leads are all insignificant, mostly negative and small. When we look at large outbreaks there is some effect of outbreaks two years prior to entry and in the year of entry; this latter result make sense given that school entry is in September. This picture overall suggests that outbreaks do not have much of a long-term effect.[11]

Older Children, Measles This paper focuses on the case of pertussis for kindergarten students because this is the age group with the best data coverage and this is the disease with the most frequent outbreaks. However, to the extent possible we would like to confirm that these effects are not limited to this setting. Appendix Table A.2 shows two additional tests. First, Column (1) shows the same regressions - pertussis vaccination on pertussis outbreaks - but for 11-year-olds. There is less data for this age group but there is some coverage from junior high entry. There is a very large vaccination response in this group; in fact, it is larger than the 5-year-olds (this is partially but not completely driven by the change in sample).

Columns (2) shows the impact of measles outbreaks on measles vaccination for entering kindergartners. Measles outbreaks are rare even relative to the pertussis outbreaks. About half of county-years have at least one case of pertussis. In contrast, only 5% of county-years have at least one case of measles. The effects here are therefore identified off of a very small number of outbreaks. However, we do see evidence that large outbreaks - here defined as county-years in which there are 15 or more cases in a county - prompt increases in vaccination. Both of these provide some helpful confirmation that the results are not limited to a single specification.

[11] To preserve sample size as much as possible the regressions which produce this table are run separately. Given the evidence in Table 4 this should not bias the results. However, the impacts are extremely similar if we run the regression together, differing largely due to the loss of sample size on estimating lags.

Functional Form of Relationship Figures 3a and 3b provide more detail on the functional form of this relationship.

Figure 3a graphs the coefficients from Column (1) of Table 3 against the average count of cases by group. The relationship is concave. The impact is largest for the group with the largest number of cases, but it is not proportionally larger. That is, the increase in vaccinations is approximately 0.4 percentage points with an average of 5 cases in the county-year, and 1.2 percentage points for an average of 150 cases.

A second way to look at this is to look by population group. If the response were linear in the number of cases then we should see the impact of disease *rate* is the same for large and small counties. If the first cases matter more, then a given rate should have a larger impact in larger counties. Figure 3b shows the relationship between rate and impact on vaccinations for the smallest 25% of counties in terms of population and the largest 25%. The largest impacts are seen in the smaller counties, reflecting the fact that the same count of cases delivers a larger rate increase. However, to the left of the graph we see evidence of concavity: the impact of a similar change in disease rate is twice as big in larger counties.

Both figures suggest the impact of disease is non-linear: the first cases matter more than subsequent cases. Another way to see this point is based on the evidence in Column (3) of Table 3. This table shows the impacts separately for the top half and bottom half of the population distribution. The table shows that the largest outbreak group has a similar effect in the large counties as the second largest group does in the smaller counties. There is a six-fold difference in average cases in these two groups, but a twenty-fold difference in county sizes. This suggests, again, non-linearity.

Interactions with Baseline Covariates It is possible to estimate variation in the response by baseline characteristics. The most obvious covariate of interest is the initial level of vaccination: is this effect larger or smaller in areas which have on average lower vaccination rates? Figure 4 shows three lines, corresponding to the impact by case group on vaccination behavior for low, medium and high vaccination areas. These areas are defined as terciles based on the average of the two lowest vaccination county-years. The graph shows the impacts as a share of the unvaccinated population since this figure is comparable across areas.

Figure 4 demonstrates that the effects are similarly sized in the bottom two terciles, but it shows no effect in the highest vaccination group. This should not be surprising since in the highest vaccination groups there is simply not much room for vaccination rates to move. We would not expect vaccination rates of 100% in all locations since some share of people cannot vaccinate for health reasons.

Appendix Table A.3 shows this interaction and interactions with further demographic variables. In general, there is not much interesting variation. There is some weak evidence that the response is smaller in areas with more education.

4.2.1 Mechanism Evidence

The results above point to a vaccination response to observed cases of disease. This section provides two pieces of evidence which bolster the claim to causality by, first, making clear the results are specific to the particular vaccination in question and, second, by showing that there is an internet search response to the outbreak.

Cross-Disease Responses The analysis thus far has focused on the impact of pertussis outbreaks on pertussis vaccination. A related question is whether pertussis cases also impact vaccination for other diseases. A finding that the response is disease-specific would make the claim that this effect is causal and driven by the outbreaks more compelling. It is worth noting, however, that there are some (non-rational) theories which would predict a cross-vaccine response.

Appendix Table A.4 shows the impact of pertussis outbreaks on other vaccinations (all measles vaccines, MMR vaccines and an overall measure of whether a child is fully up to date on vaccinations). There is no evidence of cross-disease interactions. Outbreaks of pertussis do not seem to increase vaccination rates for any other diseases.

Response of Google Trends The results in this paper rely on the assumption that people learn about disease outbreaks and respond to them. Although we cannot explicitly observe what people know about these outbreaks, we can proxy for their information with Google searches.

Figure 5a and 5b show the impact of outbreaks in a state-month on Google searches for information about the disease, vaccination and searches for terms related to vaccine dangers. Figure 5a estimates the impact of a linear control for number of cases. Figure 5b estimates the impact of a dummy for a large outbreak in the state. These figures show the impact over time - the effect in the month of the outbreak, the next month, and so on.

The results in either case are the same. Outbreaks prompt a significant increase in searches for information on the disease. They also prompt a significant increase in searches for information on the vaccine. Both of these effects are short lived - they last a month or two before dissipating. The effects are reasonably large. In the case of searches for pertussis, the impact of a large outbreak is to increase searches by 0.4 of a standard deviation. For "pertussis vaccine" this figure is 0.18 of a standard deviation. These results show that information about outbreaks is reaching the population.

In contrast there is no evidence that outbreaks increase searches for vaccine injury terms or for terms that link vaccines and autism. This is despite the fact that in general these searches move together (see correlations in Appendix Table A.5). In other words, although it is generally the case that increases in searches for pertussis vaccination also increase searches for terms related to vaccine dangers, the increase in vaccine searches that are prompted by outbreaks do not seem to be accompanied by an increase in interest in vaccine risks.

4.3 Response Variation across Areas

I analyze two sources of variation across areas. First, I estimate the difference in response between states which centrally coordinate responses and those which do not. Second, within the latter set, I estimate differences between counties which systematically notify the media about outbreaks and those which do not.

4.3.1 Variation across States

The full sample in the paper includes 12 states: Alabama, Arizona, California, Kansas, Kentucky, Michigan, Missouri, New York, North Carolina, North Dakota, Oregon and Texas. In four of these states - Alabama, North Dakota, Oregon and Michigan - the response to disease outbreaks is coordinated directly by either the state or, in the case of Michigan, regional health offices. In these states the counties may play a role in disseminating information about disease outbreaks but they do so at the direction of the state or regional office. In the other eight states counties have autonomy with respect to their response to disease outbreaks. They may notify the state and ask for help, and the state provides some guidelines, but they ultimately decide their own policies.

In principle both structures have advantages. State-coordinated response could dominate since those dealing with outbreaks will have more experience (a state experiences many more outbreaks than individual counties) and the state may have a better ability to see the whole picture, including cross-county spread, etc. On the other hand, locally-sourced responses may allow for better targeting of the response to the individual county circumstances.

Columns (1) and (2) of Table 5 show the primary regressions in the paper (replicating Column (1) of Table 3) divided by states with coordinated responses (Column (1)) and those without (Column (2)). The responses are significantly larger in the centrally coordinated states. In the non-coordinated states there is little evidence of any response other than to very large outbreaks. In the centrally coordinated states the data suggests about 10% of unvaccinated children are vaccinated even in response to a very small outbreak.

Columns (3) - (6) of Table 5 show evidence on the response of Google searches for "pertussis" and "pertussis vaccination" (and related terms) in the two groups of states. These regressions are at the state-month level, as in Figure 5. I report the impact of cases in the same month. Consistent with the evidence on vaccination response, the Google search response on both terms is substantially higher for centrally-coordinated states.

This suggests that state-coordinated notification dominates county autonomy, at least in the sense of better promoting disease outbreaks to coordinate increases in vaccination.

4.3.2 Variation across Counties

In the states without central coordination, I focus on whether counties reported a systematic way of notifying the media about outbreaks. There are 828 counties in these states, of which 244 never have any outbreaks in the course of the data. Among the others, I was able to collect data from 317. In the other cases I was unable to contact the county, the county had no health department or the health department was unable to provide information on their approach to outbreaks.

The regressions in Table 5 show, on average, limited response in these states. To enhance power, I redefine the pertussis case groups as either "small" (less than the 95th percentile of county-years with positive outbreaks) or "large" (the largest outbreak group). I then estimate the standard regressions in the paper with these independent variables and these variables interacted with a dummy for whether the county typically notifies the media. This regression also includes interactions between the county population and the outbreak variables and between state dummies and these variables. The results are shown in Table 6: the impacts of large outbreaks are significantly higher in areas with a media contact plan.

4.4 Summary

The evidence in this section shows that vaccination responds to disease outbreaks, even as those outbreaks are not informative about future disease in the data. The evidence for a causal link is bolstered both by standard robustness checks and by evidence on mechanisms drawn from internet searches. Importantly, from a policy standpoint, there is variation across space. This variation suggests that with "good management" outbreaks can have a larger impact on vaccination rates.

The next two sections develop two further implications of these findings. First, I return to the underlying question of what model of behavior explains lack of vaccination in general and use the evidence above to inform this issue. Second, I turn to better understanding the size of these impacts and ask whether they are large enough to counteract vaccination declines.

5 Model of Behavior

In this section I outline an extremely simple theory of vaccination behavior, designed to rationalize the set of facts above, along with observations from the existing literature. I focus on understanding the behavior of the marginal individuals. I begin by setting up an extremely simple framework and situating the key findings in the language of that framework. I then argue that a model of full information cannot explain these results, and suggest that a model of limited information with an important salience component may be able to rationalize them.

5.1 Setup

The model considers a simple case in which vaccination decisions depend on the trade-off between (perceived) benefits and (perceived) costs.

Consider a disease j with an associated vaccine. Household i perceives a utility cost to vaccination for disease j, C_{ij} and a utility cost to developing the disease $D_{ij.}$ I normalize $D_{ij} = 1$ so C_{ij} will then be interpreted as the utility cost to vaccination relative to the utility cost of developing the disease. Household i also holds a perception about the excess risk of developing the disease in the absence of the vaccine: p_{ij}.

Both cost and probability may differ in perception from the true values. Denote the fully informed (household-specific) cost of vaccination as $\hat{C_{ij}}$. Denote the true excess risk of developing disease j if not vaccinated as $\hat{p_j}$ and note this may be larger or smaller than p_{ij} for any i.

In line with the experiments in the data and existing literature, I consider two stimuli: education about vaccine costs and data on observed cases of disease j. Denote an education stimulus as $e \in [0, 1]$ and observed disease cases as o_j. Education may impact perceptions of vaccine costs, and observed cases of the disease may impact perceived probability. In addition, I allow for the perceived probability to be related to the actual probability.

The perceived vaccine cost in the model is therefore denoted $C_{ij}(\hat{C_{ij}}, e)$, the excess probability of developing the disease absent the vaccine is $p_{ij}(\hat{p_j}, o_j)$ Note that the perceived probability may be a function of the true probability but, importantly, the true probability is not a function of the observed cases, consistent with the evidence in Table 4.

Under this model denote the vaccination status for household i for disease j as S_{ij}, where $S_{ij} = 1\left\{p_{ij}(\hat{p_j}, o_j) > C_{ij}(\hat{C_{ij}}, e)\right\}$.

The data presented above - both the new results and the evidence from the existing literature - suggest two key comparative statics: vaccination increases with outbreaks, vaccination does not increase with education about vaccine safety, even though that education does change the perception of safety.

The table below summarizes these facts in the language of the model.

Comparative Static	Description
$\frac{\partial S_{ij}}{do_j} \vert \hat{p}_j > 0$	Vaccination increases with observed cases of disease, holding true disease rate constant
$\frac{\partial^2 S_{ij}}{\partial o_j^2} \vert \hat{p}_j < 0$	The impact of observed cases on vaccination, holding true rate constant, is concave.
$\frac{\partial S_{ij}}{\partial e} = 0$, $\frac{\partial C_{ij}}{\partial e} < 0$	Vaccination does not respond to education, but beliefs about costs do.

I argue that the facts above - in the form of these comparative statics - put restrictions on the set of models which could explain the results. Below, I begin with a benchmark full-information model of under-vaccination and argue this is ruled out by the data. I then discuss a model of limited information.

5.2 Fully Informed Model

Consider first a fully informed model of non-vaccination. This is the type of model outlined in the baseline "economic epidemiology" literature on vaccinations (see, for example, Geoffard and Philipson, 1996 and Geoffard and Philipson, 1997). In these models, the choice to vaccinate or not is a response to actual risk of disease.

The key aspect of the fully informed model is that vaccination behavior is not affected by observed cases (conditional on the true probability) or by education. We therefore have $C_{ij} = \hat{C}_{ij}$, and $p_{ij} = \hat{p}_j$. Household vaccination behavior in this model can be expressed as:

$$S_{ij} = 1\left\{\hat{p}_j > \hat{C}_{ij}\right\}.$$

Note that this model has no trouble rationalizing the choice not to vaccinate since I allow for individual heterogeneity in $\hat{C}_{ij}$. Parents in this model may rationally decide that the small probability of infection does not outweigh the small costs (pain, tiny chance of adverse reaction) of the vaccine.

This model fails to replicate the first or second comparative static above. It is straightforward to observe that $\frac{\partial S_{ij}}{do_j} = 0$, given that $\hat{p}_j$ does not depend on o_j.

As an extension, it is worth noting that no model in which $p_{ij} = \hat{p}_j$ - even those with very flexible allowances for the C_{ij} function - will be able to rationalize the patterns in the data. Effectively, the central results in this paper mean that perceived risk of disease must depend on something other than the true disease risk.

5.3 Model of Limited Information and Salience

I turn now to suggesting a particular model of limited information. The first comparative static is sufficient to reject full information. The second and third suggest a particular functional form. More specifically, we must have, first, that the perceived risk is increasing in o_j but concave. Second, the model would ideally also deliver the result that changes in the perceived cost of vaccines (due to an education stimulus) *do not* change vaccination behavior. This last point is difficult to rationalize with a simple modification of the p_{ij} function since if the decision is responsive to movements in p_{ij} we would also expect it to be responsive to movements in C_{ij} on the margin.

It is necessary to introduce something into the model to generate the asymmetry in response. This section suggest a model with a salience component may fit the data well.

The key element of this model is the particularization of p_{ij}. In particular, I assume:

$$p_{ij} = \begin{cases} 0 & \text{if } o_j = 0 \\ p_{ij}(\hat{p}_j, o_j) & \text{if } o_j > 0 \end{cases}$$

That is, assume that if people do not observe any cases of the disease, they assume there is no chance of developing the disease. If they observe any cases of the disease then the perceived chance is a function of the true chance and the number of cases they observe. This is effectively a model of salience (Bordalo, Gennaioli and Shleifer, 2012), in which the disease risk only becomes salient to people once they are aware of at least one case.

Vaccination behavior can be expressed as follows:

$$S_{ij} = \begin{cases} 0 & \text{if } o_j = 0 \\ 1\left\{p_{ij}(\hat{p}_j, o_j) > C_{ij}(\hat{C}_{ij}, e)\right\} & \text{if } o_j > 0 \end{cases}$$

This model clearly delivers movement in vaccination behavior with observed cases of the disease. This occurs both because of the discontinuity associated with the belief structure and because $p_{ij}(\hat{p}_j, o_j)$ may vary with o_j. The model also accommodates a concave response to outbreaks, again for two reasons. First, the discontinuity generates concavity in response directly. Second, it would be possible to generate more discontinuity with the functional form of $p_{ij}(\hat{p}_j, o_j)$.

This model also accommodates the idea that S_{ij} may not respond much (or at all) to e

even if C_{ij} does respond. If $o_j = 0$ then movements in $C_{ij}(\hat{C_{ij}}, e)$ will not impact vaccination behavior. Intuitively, if people believe there is no benefit of vaccination - nothing to protect against - they will not vaccinate even if the cost is very small. Given that there are relatively few cases of these diseases in a given time period, it is likely that campaigns to educate people about vaccine safety will primarily be accessing people who have not recently seen cases of disease and, hence, movements in beliefs will have no impact.[12]

To the extent that this model captures important features of the decision not to vaccinate it suggests that there may be scope for education to influence vaccination rates but this education should focus on the risks of not vaccinating, not the safety of vaccines. This is consistent with the survey evidence in Cacciatore, Nowak and Evans (2016). Importantly, this model also suggests that it may be possible to affect beliefs even without changes in disease rates. Making even a small number of local disease cases salient to families may increase vaccination rates.

6 Are Declining Vaccination Rates Self-Limiting?

The results in Section 4 show vaccination rates increasing with disease outbreaks. A natural way to think about the magnitude of these effects is to ask whether they imply “cycles” in vaccination rates. As vaccination rates fall, outbreaks are more likely to occur, and this may lead to increased vaccination rates again. This will only occur, however, if these responses are large. In this section I take up this question.

6.1 Setup

Consider the simplest “economic epidemiology” model, explained by two equations which relate vaccination rates over time, V_t, to disease rates d_t. The first equation captures the epidemiology - current vaccination rates affects current disease - and the second capture the economics - recent disease outbreaks affect current vaccination.

$$d_t = f(V_t)$$
$$V_t = g(d_{t-1})$$

Ideally, in order to estimate this we would have full information on both of these functions. In reality, both are a challenge.

[12] It is important to recall that this is a model intended to explain non-vaccination behavior - that is, the behavior of a set of marginal vaccinators. There are of course a large set of people who vaccinate their children who may or may not believe there are disease risks.

The evidence in Section 4 provides an estimate of the response of vaccination among a single age group to disease outbreak. I do not have direct evidence on the response among any other age group.

Further, there is not an obvious source of information on the relationship between vaccination rates and disease outbreaks. Epidemiological models of pertussis focus more on natural cycles in the disease rates in a large population and less on the impact of location variation in vaccination on individual outbreaks. Existing papers on the link between local pertussis vaccination rates and pertussis outbreaks use data similar to the data used in this paper, but less complete.

To move forward on this problem, then, I introduce a number of assumptions. Future work would ideally more fully elucidate this link.

Vaccination and Disease I use cross sectional variation in the data to estimate the relationship between the level of vaccination and the probability of a disease outbreak of various sizes. Specifically, I estimate an ordered probit model where the outcome is outbreak size categories and the treatment variables are categories of vaccination rates. I estimate this using variation in the cross section since, as noted in Appendix Table A.1, within a county over time the year-to-year variations in vaccination rates do not seem to relate to pertussis levels. In the cross section there is a strong relationship between vaccination rates and outbreaks. This can be seen in the ordered probit results in Appendix Table A.6.

The output from the ordered probit model is a predicted probability of being in each outbreak group for each vaccination rate group.

Disease and Vaccination I show two simulations. First, I assume that the only response to disease occurs among the cohort of children entering kindergarten. I use the estimates from Table 3. Second, I assume this response occurs among all children.

Simulation Structure I focus on disease and vaccination rates among children 10 and under, who are the group primarily affected by pertussis. I start the simulation with a 99.5% vaccination rate. I then introduce a "shock": a permanent reduction of vaccination rate to 95% for each entering cohort. The idea is to capture, for example, the introduction of some information (like the purported vaccine-autism link) which limits vaccination.

I consider two possibilities for the evolution of vaccination rates. In the conservative specification I limit response to the cohort at age 5. That is, no additional children are vaccinated between ages 1 and 4 or between 6 and 10. At age 5 some children are vaccinated according to whether there was an outbreak in that year and its size; he response is to increase vaccination as a share of the unvaccinated group, with the size of the impacts based on the results in Section 4. Outbreaks are generated in simulation based on random data draws and the predicted probability of outbreaks of various sizes given the prevailing vaccination rate in the cohort.

In the less conservative specification I imagine that all age cohorts respond. In both cases I use, first, the overall response and, second, the response among the centrally coordinated states (see Table 5).

Throughout I run each simulation 10,000 times and average.

6.2 Results

The results of the simulations are in Figure 6. The first two sub-figures, Figure 6a and 6b show the path of vaccination rates; Figures 6c and 6d show corresponding graphs of the risk of a large outbreak (>=75% of case-years) in each scenario.

In the most conservative specification this response makes little difference to vaccination rates overall. It is straightforward to see why: even though the responses at age 5 are relatively large, there is only a chance for the response to operate on a single cohort. Figure 6c demonstrates that this response makes no difference to the risk of a large outbreak.

In the less conservative specification (Figures 6b and 6d), however, the response has a large effect on vaccination rates. Without response, the prevailing vaccination rates is 95%. With the large response, the rate is around 97%. Figure 6d shows that this makes a difference to the risk of large outbreaks. These outbreaks occur 13.5% of the time in steady state in the large response condition, and 16.6% of the time in the baseline condition.

This analysis suggests that the results here could matter for the path of vaccination in the face of a shock, although the result is heavily dependent on whether the vaccination response occurs only in one cohort versus all cohorts. It is worth noting again that there are many assumptions inherent in this analysis and even if we are confident in the estimated relationship between vaccination rate and disease outbreaks within the range of data here, we know little about the relationship between vaccination rates and outbreaks at lower levels of vaccination. From historical evidence we can learn about the likely rate of pertussis in the absence of any vaccination, but it may be harder to learn about the impact of (for example) a steady state rate of 70% vaccinated.

7 Conclusion

Anecdotal evidence suggests that vaccine-resistant parents can be swayed toward vaccination by disease outbreaks. This paper provides evidence suggesting those anecdotes are borne out in the data. Using a data set of county-year vaccination rates and outbreaks, I show that vaccination rates among entering kindergartners are increased by outbreaks of disease. For large outbreaks, these effects are sizable. In the second set of results I show that these outbreaks increase interest in the disease and vaccinations, as measured by Google search volume, but do not result in an increase in searches for vaccine dangers. The vaccination effects exist in spite of the fact that current outbreaks are not informative about future outbreaks.

It is difficult to fit these facts with a fully informed model in which households react to the disease risk. Instead, I suggest the data may be better fit by a model in which being made aware of even a single case of disease prompts changes in perceived disease risk through a salience mechanism. A version of this model in which individuals perceive the risk to be zero when they do not observe any cases can also fit the fact that we see limited evidence that vaccination responds to education campaigns.

The key policy issue motivating this paper is how to increase childhood vaccination rates. The evidence here suggests that disease outbreaks may be a powerful motivator and, in particular, that they may be a useful motivation even if they are not actually informative. I show that the structure of outbreak response across counties and states can importantly influence the size of this response. In particular, states which coordinate their response through state health departments are much more effective at promoting vaccination response than those which coordinate at the county level. Within the latter, I see some evidence that counties with a systematic media notification plan have larger responses. This may reflect better general management.

In general, these results suggest that disease outbreaks may provide an important opportunity to promote vaccination behavior. Effective management of this promotion may enhance these effects.

References

Adda, Jérôme, "Economic Activity and the Spread of Viral Diseases: Evidence from High Frequency Data," *The Quarterly Journal of Economics*, 2016, p. qjw005.

Board, LA Times Editorial, "California's new vaccine law is already a success," *Los Angeles Times*, Feb 2016.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "SALIENCE THEORY OF CHOICE UNDER RISK," *The Quarterly Journal of Economics*, 2012, *1243*, 1285.

Cacciatore, Michael A, Glen Nowak, and Nathaniel J Evans, "Exploring The Impact Of The US Measles Outbreak On Parental Awareness Of And Support For Vaccination," *Health Affairs*, 2016, *35* (2), 334–340.

Constable, Catherine, Nina R Blank, and Arthur L Caplan, "Rising rates of vaccine exemptions: problems with current policy and more promising remedies," *Vaccine*, 2014, *32* (16), 1793–1797.

Davis, S. F., P. M. Strebel, S. L. Cochi, E. R. Zell, and S. C. Hadler, "Pertussis surveillance–United States, 1989-1991," *MMWR CDC Surveill Summ*, Dec 1992, *41* (8), 11–19.

Geoffard, Pierre-Yves and Tomas Philipson, "Rational epidemics and their public control," *International economic review*, 1996, pp. 603–624.

___ **and** ___ , "Disease eradication: private versus public vaccination," *The American Economic Review*, 1997, *87* (1), 222–230.

Glasser, John W, Zhilan Feng, Saad B. Omer, Philip J. Smith, and Lance E. Rodewald, "The effect of heterogeneity in uptake of the measles, mumps, and rubella vaccine on the potential for outbreaks of measles: a modelling study," *The Lancet Infectious Diseases*, 2016, pp. –.

Healy, Jack and Michael Paulson, "Vaccine Critics Turn Defensive Over Measles," *The New York Times*, 2015.

Horne, Zachary, Derek Powell, John E Hummel, and Keith J Holyoak, "Countering antivaccination attitudes," *Proceedings of the National Academy of Sciences*, 2015, *112* (33), 10321–10324.

Kremer, Michael, "INTEGRATING BEHAVIORAL CHOICE INTO EPIDEMIOLOGICAL MODELS OF AIDS.," *Quarterly Journal of Economics*, 1996, *111* (2).

Kutty, Preeta, Jennifer Rota, William Bellini, Susan B. Redd, Albert Barskey, and Gregory Wallace, *Manual for the Surveillance of Vaccine-Preventable Diseases, Chapter 7: Measles* Centers for Disease Control and Prevention, 6 ed. 2013.

Mello, Michelle M, David M Studdert, and Wendy E Parmet, "Shifting vaccination politicsâthe end of personal-belief exemptions in California," *New England Journal of Medicine*, 2015, *373* (9), 785–787.

Mnookin, Seth, *The panic virus: a true story of medicine, science, and fear*, Simon and Schuster, 2011.

Nyhan, B. and J. Reifler, "Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information," *Vaccine*, Jan 2015, *33* (3), 459–464.

Nyhan, Brendan, Jason Reifler, Sean Richey, and Gary L Freed, "Effective messages in vaccine promotion: a randomized trial," *Pediatrics*, 2014, *133* (4), e835–e842.

Omer, Saad B., Daniel A. Salmon, Walter A. Orenstein, M. Patricia deHart, and Neal Halsey, "Vaccine Refusal, Mandatory Immunization, and the Risks of Vaccine-Preventable Diseases," *New England Journal of Medicine*, 2009, *360* (19), 1981–1988.

____, Jennifer L. Richards, Michelle Ward, and Robert A. Bednarczyk, "Vaccination Policies and Rates of Exemption from Immunization, 2005-2011," *New England Journal of Medicine*, 2012, *367* (12), 1170–1171.

Omer, Saad B, Kyle S Enger, Lawrence H Moulton, Neal A Halsey, Shannon Stokley, and Daniel A Salmon, "Geographic clustering of nonmedical exemptions to school immunization requirements and associations with geographic clustering of pertussis," *American Journal of Epidemiology*, 2008, *168* (12), 1389–1396.

Omer, Saad B., William K. Y. Pan, Neal A. Halsey, Shannon Stokley, Lawrence H. Moulton, Ann Marie Navar, and Daniel A. Pierce Mathew annd Salmon, "Nonmedical exemptions to school immunization requirements: Secular trends and association of state policies with pertussis incidence," *The Journal of the American Medical Association*, 2006, *296* (14), 1757–1763.

Orenstein, Walter and Katherine Seib, "Mounting a good offense against measles," *New England Journal of Medicine*, 2014, *371* (18), 1661–1663.

Phadke, Varun, Robert A. Bednarczyk, Daniel A. Salmon, and Saad B. Omer, "Association between vaccine refusal and vaccine-preventable diseases in the united states: A review of measles and pertussis," *The Journal of the American Medical Association*, 2016, *315* (11), 1149–1158.

Philipson, Tomas, "Private vaccination and public health: An empirical examination for US measles," *Journal of Human Resources*, 1996, pp. 611–630.

____, "Economic epidemiology and infectious diseases," *Handbook of health economics*, 2000, *1*, 1761–1799.

Sadaf, A., J. L. Richards, J. Glanz, D. A. Salmon, and S. B. Omer, "A systematic review of interventions for reducing parental vaccine refusal and vaccine hesitancy," *Vaccine*, Sep 2013, *31* (40), 4293–4304.

Salmon, Daniel A., Lawrence H. Moulton, Saad B. Omer, Patricia deHart, Shannon Stokley, and Neal A. Halsey, "Factors associated with refusal of childhood vaccines among parents of school-aged children: A case-control study," *Archives of Pediatrics and Adolescent Medicine*, 2005, *159* (5), 470–476.

Siepel, Tracy, "California's vaccine law: Opponents moving, home schooling to avoid controversial mandate," *The Mercury News*, June 2016.

Stadlin, Stephanie, Robert A Bednarczyk, and Saad B Omer, "Medical exemptions to school immunization requirements in the United Statesâassociation of state policies with medical exemption rates (2004–2011)," *Journal of Infectious Diseases*, 2012, *206* (7), 989–992.

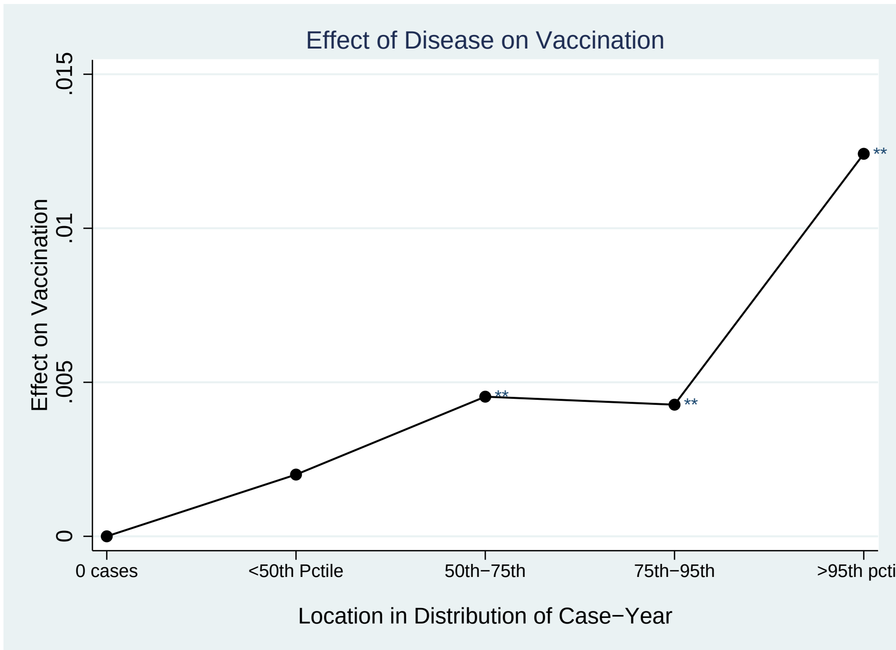
Stephens-Davidowitz, Seth, "The cost of racial animus on a black candidate: Evidence using Google search data," *Journal of Public Economics*, 2014, *118*, 26–40.

Williams, S. E., R. L. Rothman, P. A. Offit, W. Schaffner, M. Sullivan, and K. M. Edwards, "A randomized trial to increase acceptance of childhood vaccines by vaccine-hesitant parents: a pilot study," *Acad Pediatr*, 2013, *13* (5), 475–480.

Wolf, Elizabeth R., Douglas Opel, M. Patricia DeHart, Jodi Warren, and Ali Rowhani-Rahbar, "Impact of a pertussis epidemic on infant vaccination in Washington state," *Pediatrics*, 2014, *134* (3), 456–464.

Yang, Tony Y. and Ross D. Silverman, "Legislative Prescriptions for Controlling Nonmedical Vaccine Exemptions," *The Journal of the American Medical Association*, 2015, *313* (3), 247–248.

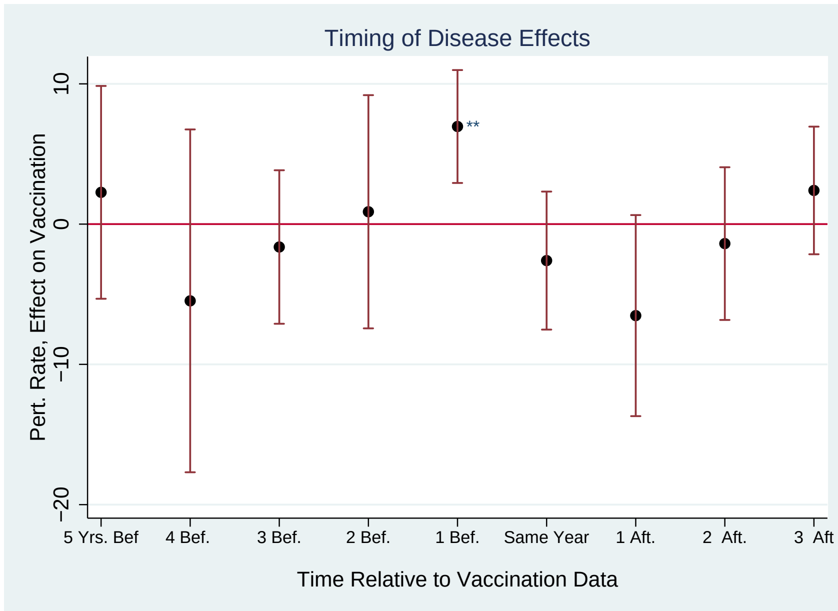
Figure 1: **Impact of Disease on Vaccination**



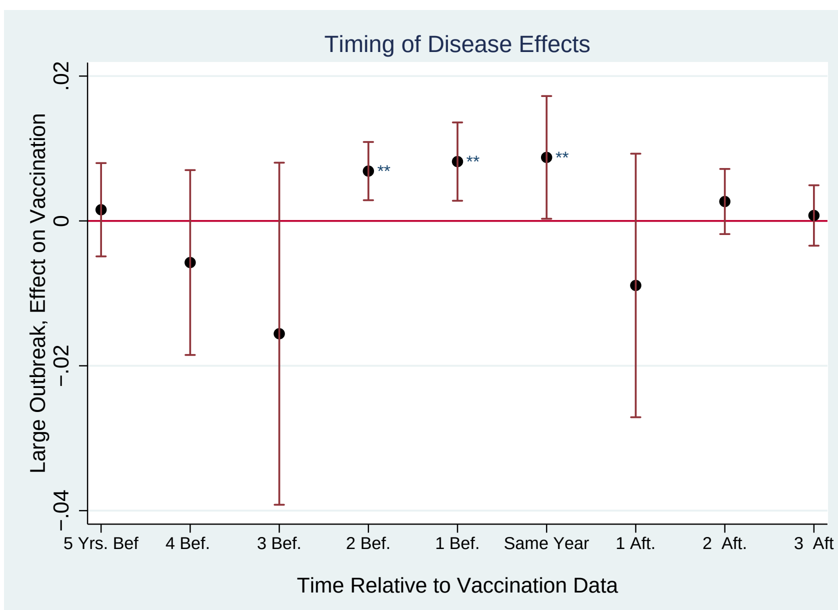
Notes: These figures show the impact of pertussis cases in the county on vaccination behavior. The vaccination data is at the time of school entry and the outbreaks are measured in the year prior. All points shown are regression coefficients from regressions with county and year fixed effects.

Figure 2: **Timing of Disease Impacts**

(a) Impact of Pertussis Rate on Vaccination



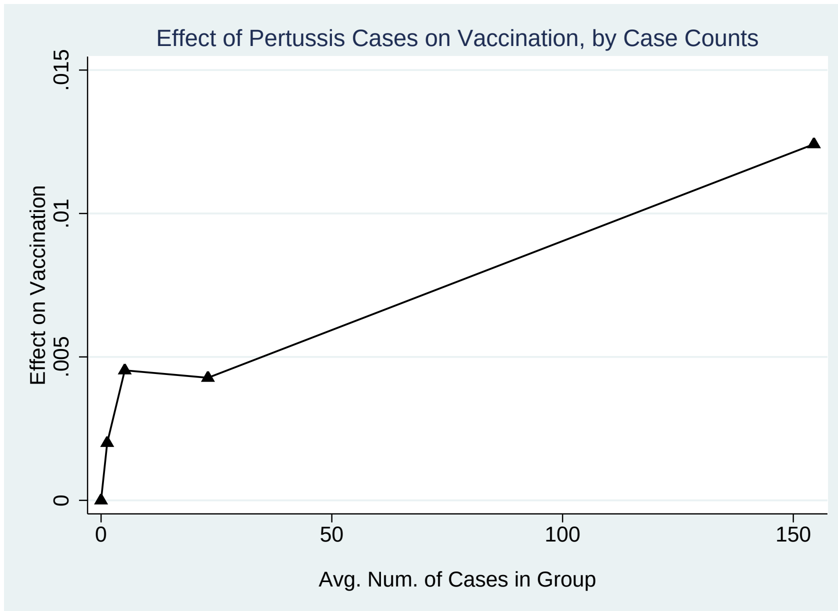
(b) Impact of Large Outbreaks on Vaccination



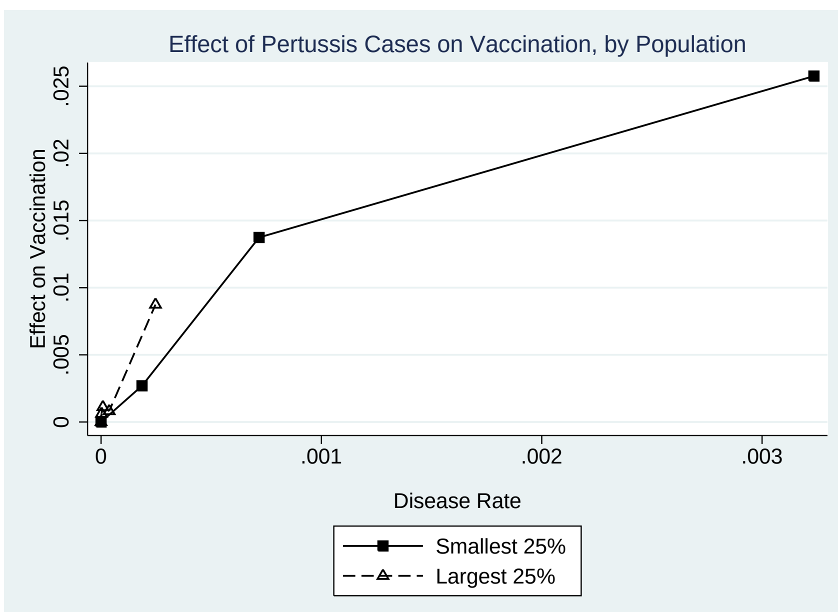
Notes: These figures show the impact of disease cases in the county on vaccination behavior in lags and leads. The outbreak is measured either as a dummy for a large outbreak or as the disease rate. All points shown are regression coefficients from regressions with county and year fixed effects. ** sig. at 5% level.

Figure 3: **Functional Form Analysis**

(a) Vaccination Impact by Case Count

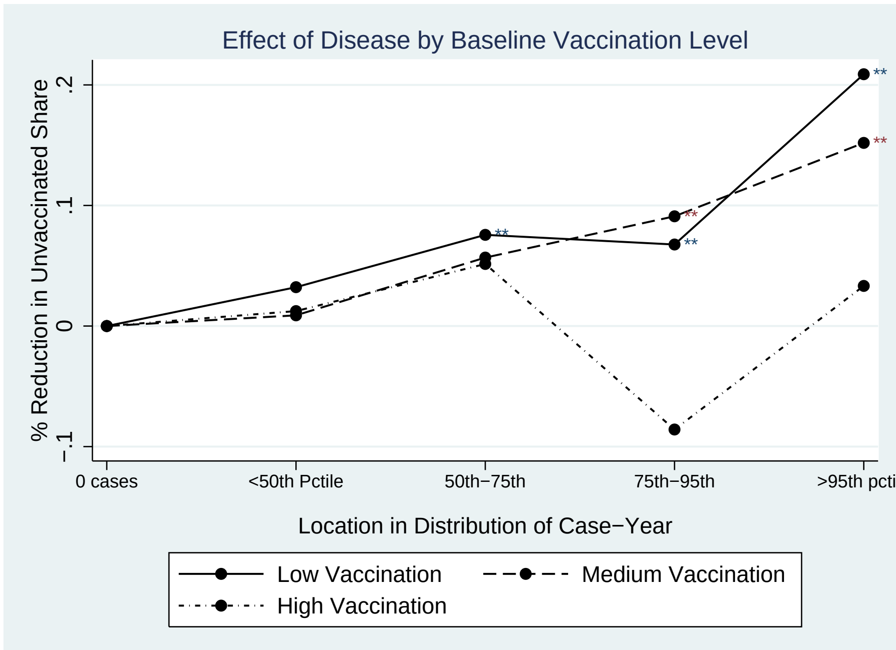


(b) Vaccination Impact by Population



Notes: These figures illustrate the functional form of the relationship between outbreaks and vaccination. In Sub-Figure b the two groups are the top 25% of counties in terms of population and the bottom 25% of counties. All points shown are regression coefficients from regressions with county and year fixed effects.

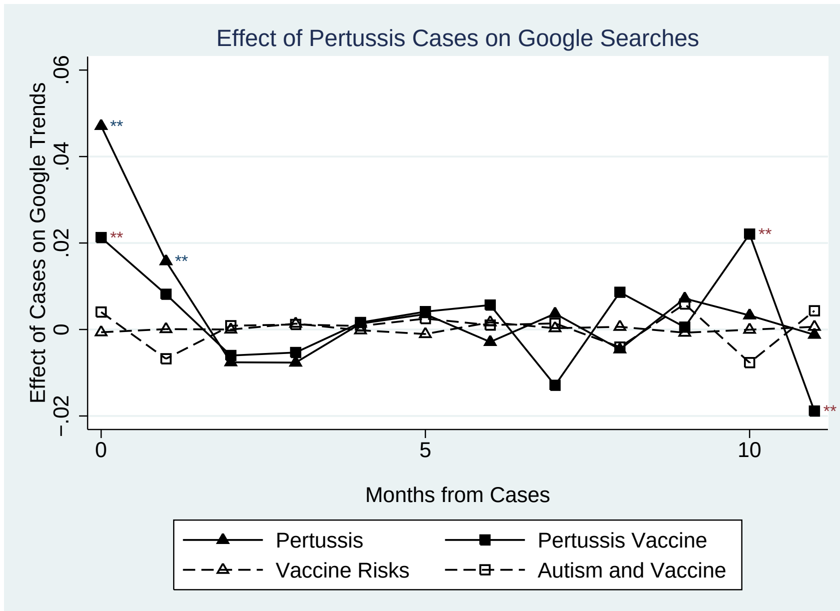
Figure 4: **Interactions with Vaccination Levels**



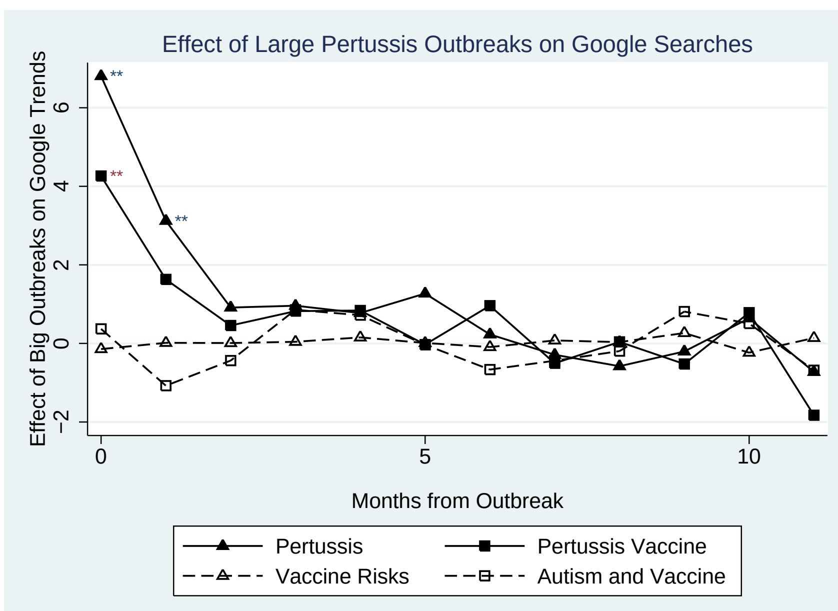
Notes: These figures show the impact of outbreaks on vaccination separated by areas with high and low vaccination rates. Counties are divided into three groups based on the minimum county-year vaccination rate. All points shown are regression coefficients from regressions with county and year fixed effects, scaled by the average under-vaccination rate in each group.

Figure 5: **Impact of Pertussis Cases on Google Searches**

(a) Count of Cases



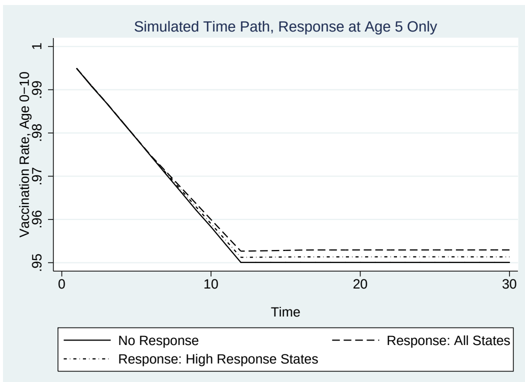
(b) Large Outbreak



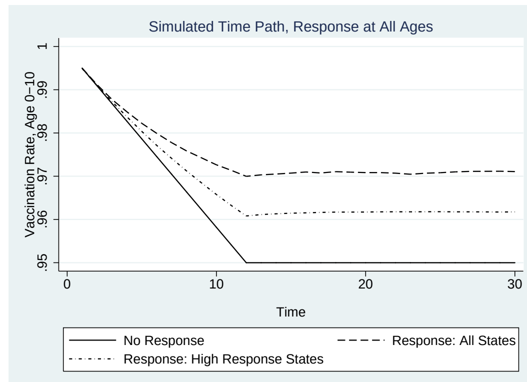
Notes: These figures show the impact of outbreaks on Google searches for four groups of terms: terms related to the disease, "pertussis vaccine" and related, "vaccine injury" and related and "vaccine and autism". The graph maps out the impact in the month in which the cases occur, the following month, the month after and so on up to a year after the outbreak. All coefficients are from regressions which include state and month fixed effects. Sub-figure (a) shows the impact of a linear control for number of cases. Sub-figure (b) shows the impact of a dummy for a large outbreak.

Figure 6: **Simulation Results**

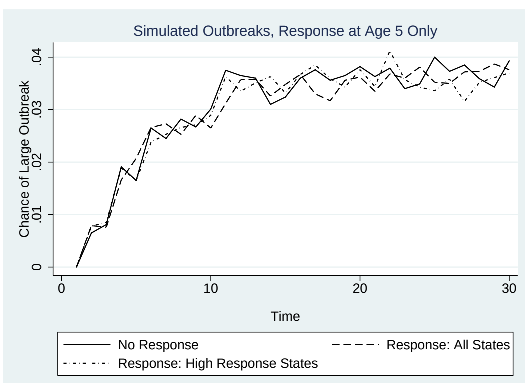
(a) Simulated Vaccination, Single Year Response



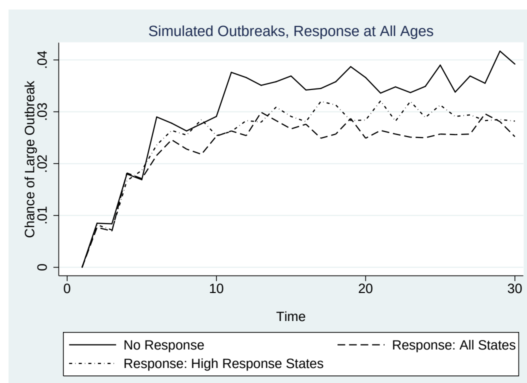
(b) Simulated Vaccination, All Ages Respond



(c) Simulated Outbreaks, Single Year Response



(d) Simulated Outbreaks, All Ages Respond



Notes: These figures show the results of the simulation. Subfigures (a) and (c) show results from the simulation assuming that only one age group (five-year-olds) responds. Sub-figures (b) and (d) show results assuming that all age groups respond. A large outbreak is an outbreak in >=95th percentile of county-years.

Table 1: **Summary Statistics**

State	Years of Coverage	Sample Size	Panel A: Vaccination Rates		Panel B: Disease Data	
			Mean	Std. Dev.	Mean	Std. Dev.
Alabama	2007-2011	304	0.978	0.042	2.673	5.177
Arizona	2009-2011	45	0.945	0.033	4.762	5.597
California	1991-2011	1218	0.934	0.041	4.670	11.540
Kansas	2009-2011	314	0.867	0.096	6.936	25.383
Kentucky	2004-2011	886	0.964	0.059	3.049	8.709
Michigan	2004-2011	662	0.906	0.034	8.765	31.470
Missouri	2011-2011	115	0.966	0.027	2.950	9.8656
New York	2002-2011	682	0.970	0.095	8.566	20.331
North Carolina	1999-2011	1293	0.991	0.030	2.060	10.954
North Dakota	2005-2011	352	0.935	0.094	6.470	20.466
Oregon	1992-2011	660	0.959	0.032	4.508	19.299
Texas	2007-2011	1260	0.972	0.048	5.014	17.980

Notes: This table shows pertussis vaccination and disease rates by state. Disease rates are quoted in rates per 100,000 people. As throughout the paper, vaccination rates are the share of children entering kindergarten with any pertussis vaccination.

Table 2: **Relationship Between Pertussis Vaccinations and Demographics**

Outcome:	**Vaccination Rate**	
	(1)	(2)
Period:	*2004-2005*	*2010-2011*
Share HS Degree	-0.091**	-0.061
	(.043)	(.047)
Share College Degree	-0.038*	-0.128*
	(.020)	(.067)
Share Black	-0.0197**	-0.020
	(.009)	(.019)
Median Family Income ('000s)	0.0005**	0.0013**
	(.0002)	(.0003)
Google: Pertussis	0.004***	-0.001
	(.002)	(.001)
Google: Pertussis Vacc.	0.001	0.003*
	(.001)	(.002)
Google: Vaccine Injury	-0.001	-0.002
	(.001)	(.002)
Google: Autism & Vaccine	-0.004**	-0.004*
	(.001)	(.001)
State FE	YES	YES
R-Squared	0.59	0.39
Number of Observations	479	892

Notes: This table illustrates the relationship between vaccination rates and county-level demographics. Vaccination rates are averaged for the 2004-2005 or 2010-2011 period. Demographic measures are from the 2010 census. Google searches are at the DMA-area level and are all standardized. All regressions include state fixed effects and are clustered at the DMA level. Figures in parentheses are standard errors. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 3: **Impacts of Pertussis Outbreaks on Vaccination**

Outcome:	*Pertussis Vaccination Rate, 5-year-olds*					
	(1)	(2)	(3)	(4)	(5)	(6)
<50th Pctile of Cases	0.002 [4.5%]		0.001 [3.2%]	0.003*	0.002	0.002
	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
50th to <75th Pctile	0.0045*** [10.3%]		0.0021 [4.8%]	0.0056***	0.0055***	0.0044***
	(0.001)		(0.001)	(0.002)	(0.002)	(0.001)
75th to <95th Pctile	0.0042***[9.7%]		0.002[4.1%]	0.0059***	0.0049***	0.0041**
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
>=95th Pctile	0.012***[28.1%]		0.011***[24.4%]	0.010***	0.014***	0.012***
	(0.003)		(0.003)	(0.003)	(0.004)	(0.003)
# of Cases		0.000 [0.00%]				
		(0.000)				
Rate		6.82*** [154%]				
		(2.13)				
<50th Pctile X Low Pop			0.0004 [0.9%]			
			(0.002)			
50th to <75th X Low Pop			0.0071** [16.0%]			
			(0.0034)			
75th to <95th X Low Pop			0.010**[24.0%]			
			(0.005)			
>=95th Pctile X Low Pop			N/A			
<50th Pctile of Cases, t+1					-0.002	
					(0.001)	
50th to <75th Pctile, t+1					-0.002	
					(0.002)	
75th to <95th Pctile, t+1					-0.003	
					(0.002)	
>=95th Pctile, t+1					-0.010	
					(0.009)	
Exemption Policy (1-3)						-0.010*
						(.006)
County FE	YES		YES	YES	YES	YES
Year FE	YES		YES	YES	YES	YES
County-Specific Trends	NO		NO	YES	NO	YES
R-squared	0.58		0.58	0.70	0.58	0.70
Number of Observations	7472		7472	7472	6422	7472

Notes: This table shows the impact of pertussis outbreaks on vaccination rates. Outbreaks are defined by groups. The omitted category is 0 cases. The groups are then based on the distribution of positive county years. "Low pop" (Column (2) interaction) is a dummy for being in the bottom half of the population distribution. The highest group interaction is omitted since no small counties have this outbreak rate. Figures in square brackets show the change as a share of the average unvaccinated population. Robust standard errors in parentheses, clustered at the county level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 4: **Impacts of Current Pertussis Cases on Future Cases**

Outcome:	*Pertussis Rate*		*Cases of Pertussis*	
	(1)	(2)	(3)	(4)
<50th Pctile of Cases, $t-1$	0.0000005		-1.18**	
	(0.000009)		(0.45)	
50th to <75th Pctile, $t-1$	-0.000008		-2.32***	
	(0.000008)		(0.75)	
75th to <95th Pctile, $t-1$	-0.0000006		0.357	
	(0.00001)		(1.63)	
>=95th Pctile, $t-1$	0.00002		22.04***	
	(0.00002)		(7.01)	
<50th Pctile of Cases, $t-2$	-0.000006		-0.20	
	(0.000006)		(0.34)	
50th to <75th Pctile, $t-2$	-0.00002*		-0.58	
	(0.00001)		(0.69)	
75th to <95th Pctile, $t-2$	-0.00004***		-1.87	
	(0.00001)		(12.6)	
>=95th Pctile, $t-2$	-0.00009***		-20.30***	
	(0.00001)		(7.34)	
Pertussis Rate, $t-1$		-0.020		864.6
		(0.038)		(1144.1)
Pertussis Rate, $t-2$		-0.140***		-8763.5**
		(0.047)		(4498.8)
County FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
R-squared	0.26	0.28	0.50	0.50
Number of Observations	7734	7734	7734	7734

Notes: This table shows the impact of past disease on current disease. Outbreaks are defined in groups. The omitted category is 0 cases. The groups are then based on the distribution of positive county years. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 5: **Impact of Outbreaks by Local Response Coordination**

Outcome	*Pertussis Vacc. Rate*		*Searches: Pertussis*		*Searches: Pertussis Vacc.*	
State Group:	State/Region Control	County Control	State/Region Control	County Control	State/Region Control	County Control
	(1)	(2)	(3)	(4)	(5)	(6)
<50th Pctile of Cases	0.0067** [11.3%]	0.0002 [0.6%]				
	(0.0030)	(0.0013)				
50th to <75th Pctile	0.012*** [19.7%]	0.0017 [4.6%]				
	(0.0030)	(0.0016)				
75th to <95th Pctile	0.015***[24.4%]	0.0004 [1.1%]				
	(0.0033)	(0.0021)				
>=95th Pctile	0.030***[49.4%]	0.007*[17.0%]				
	(0.008)	(0.004)				
# of Cases This Month			0.268***	0.050***	0.233***	0.006
			(0.013)	(0.003)	(0.041)	(0.004)
County FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
R-squared	0.47	0.61	0.81	0.73	0.57	0.65
Number of Observations	1976	5496	501	1007	333	867

Notes: This table shows the impact of outbreaks on vaccination rates and Google searches depending on the government coordination of outbreak response. Columns (1) and (2) look at outbreak impacts. The omitted category is 0 cases. The groups are then based on the distribution of positive county years. Figures in square brackets show the change as a share of the average unvaccinated population. Robust standard errors in parentheses, clustered at the county level. States with state or region coordinated responses are Alabama, North Dakota, Michigan and Oregon. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 6: **Impacts of Outbreaks by County Media Notification**

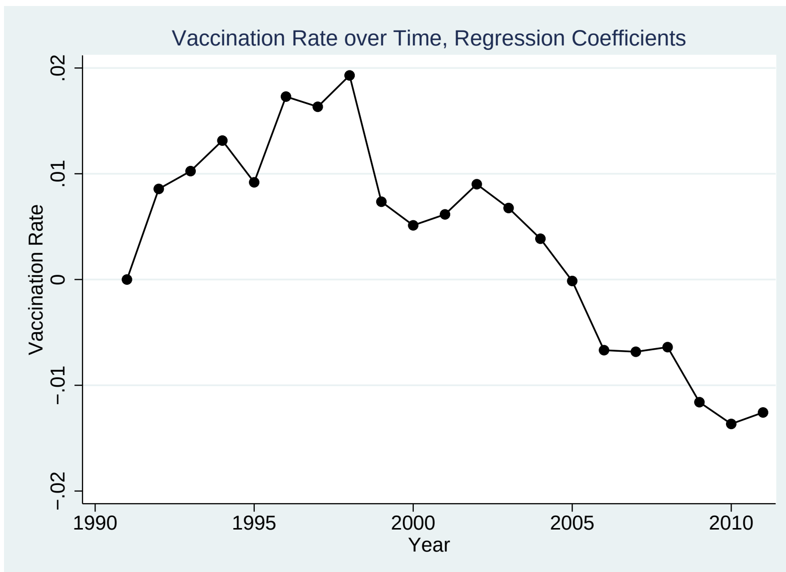
Outcome:	*Pertussis Vaccination*
	(1)
<95th Pctile, $t-1$	-0.011*
	(0.006)
>=95th Pctile, $t-1$	0.005
	(0.003)
<95th Pctile × Media Notified	0.001
	(0.004)
>=95th Pctile, × Media Notified	0.011**
	(0.005)
County FE	YES
Year FE	YES
Population X Outbreak Controls	YES
State Dummy X Outbreak Controls	YES
R-squared	0.64
Number of Observations	3740

Notes: This table shows how the impact of outbreaks varies with county notification details. The data excludes states with state coordination of response. The data includes all the counties with no outbreaks and the counties for which we have data on the details of outbreak response. Outbreaks are defined as either small or large to enhance power. The interactions of interest are those between the outbreak groups and whether the county reports systematically notifying the media of outbreaks. The regressions include interactions between county population and the outbreak groups and between state dummies and the outbreak groups. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Appendix: Online Publication Only

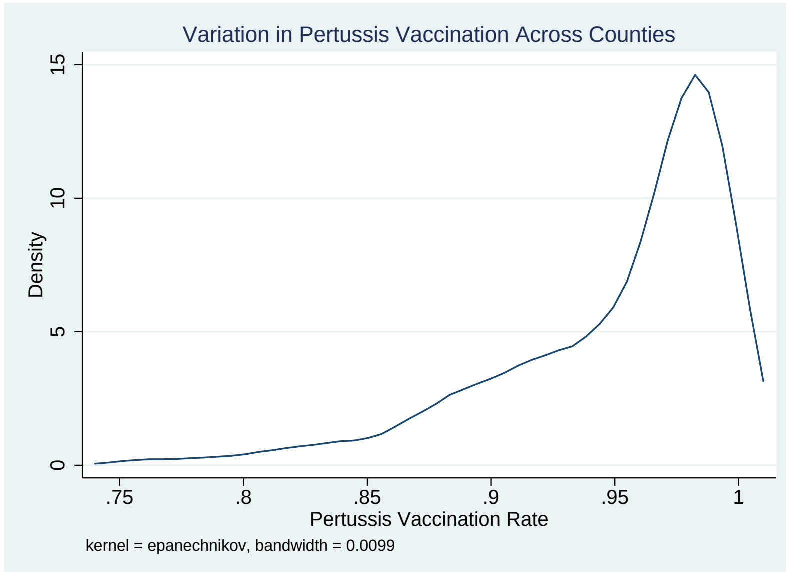
Appendix A: Figures and Tables

Figure A.1: **Pertussis Vaccination Rate Variation Over Time**



Notes: This graph shows estimates of changes in pertussis vaccination rates over time. The observations are coefficients on year dummies from regressions of vaccination rate on these dummies and county fixed effects. The regressions use an unbalanced panel, although do control for county fixed effects.

Figure A.2: **Pertussis Vaccination Rate Variation Across Space**



Notes: These figures show the density of vaccination rates across counties in the 2010/2011 period. I average vaccination rates for 2010 and 2011 and drop the bottom 1% of counties.

Table A.1: **Impact of Vaccination Rates on Outbreaks**

Outcome:	*Pertussis*	
	Pertussis Rate, t	Large Outbreak (0/1), t
Vaccine Rate, t	-0.000014	0.023
	(0.000033)	(0.017)
Population		0.00000060***
		(0.000000092)
County FE	YES	YES
Year FE	YES	YES
R-Squared	0.28	0.12
Number of Observations	7538	7614

Notes: This table shows the relationship between contemporaneous vaccine rates and disease rates. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table A.2: **Auxiliary Effects: Older Children, Measles Vaccination**

Outcome:	*Pertussis Vacc, 11-year-olds*	*Measles Vacc: 5-year-Olds*
	(1)	(2)
<50th Pctile of Cases	0.006	
	(0.009)	
50th to <75th Pctile	0.016*	
	(0.010)	
75th to <95th Pctile	0.027**	
	(0.011)	
>=95th Pctile	0.068***	
	(0.019)	
Measles Cases: 1-4		-0.0003
		(0.002)
Measles Cases : 5-14		0.002
		(0.003)
Measles Cases: >=15		0.012**
		(0.006)
County FE	YES	YES
Year FE	YES	YES
R-squared	0.71	0.76
Number of Observations	1249	7614

Notes: This table shows the impact of pertussis outbreaks on vaccination among 11-year-olds (Column (1)) and the impact of measles cases on vaccination rate of entering kindergartners. The pertussis case groups are defined as in the primary analysis in the paper. The omitted measles group is 0 cases. The maximum number of cases of measles in a county-year is 42. Robust standard errors in parentheses, clustered at the county level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table A.3: **Interactions between Response and Demographics**

Outcome:	*Pertussis Vaccination Rate*	
	(1)	(2)
Pertussis Rate, $t-1$	171.6	
	(108.2)	
% HS X Rate	-0.554	
	(0.572)	
Med. Income X Rate	-0.00006	
	(0.0004)	
Autism Search X Rate	0.613	
	(0.709)	
Vacc Level X Rate	-133.0	
	(91.42)	
Big Outbreak, $t-1$		0.258***
		(0.097)
% HS X Big		-0.0009**
		(0.0003)
Med. Income X Big		0.0000
		(0.0000)
Autism Search X Big		0.0004
		(0.0003)
Vacc Level X Big		-0.196**
		(0.093)
County FE	YES	YES
Year FE	YES	YES
R-squared	0.58	0.58
Number of Observations	7356	7356

Notes: This table shows the impact of pertussis rate on vaccination, interacted with various demographics. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table A.4: **Impacts of Pertussis Outbreaks on Other Vaccination Rates**

Outcome:	*All Measles Vacc.*	*MMR. Vacc.*	*Up-to-Date*
	(1)	(2)	(3)
<50th Pctile of Cases, $t-1$	0.0007	0.0007	-0.004
	(0.001)	(0.001)	(0.003)
50th to <75th Pctile, $t-1$	0.001	0.002	0.0004
	(0.002)	(0.002)	(0.003)
75th to <95th Pctile, $t-1$	-0.0008	-0.0006	0.001
	(0.002)	(0.003)	(0.003)
>=95th Pctile, $t-1$	-0.001	0.001	0.007
	(0.003)	(0.004)	(0.005)
County FE	YES	YES	YES
Year FE	YES	YES	YES
R-squared	0.76	0.73	0.55
Number of Observations	7614	6433	5339

Notes: This table shows the impact of pertussis outbreaks on vaccination for other diseases. Outbreaks are defined in groups. The omitted category is 0 cases. The groups are then based on the distribution of positive county years. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table A.5: **Correlations between Search Terms**

	"Pertussis"	"Pertvacc"	"Metalrisk"	"Autism"
"Pertussis"	1.000			
"Pertvacc"	0.2402	1.000		
"Metalrisk"	0.0753	0.0744	1.000	
"Autism"	0.0014	0.0343	0.1236	1.000

Table A.6: Regression of Outbreaks on Vaccination Levels, For Simulation

Outcome:	*Outbreak Category [1-5]*
	Method: Ordered Probit
96% to <98%	-0.136***
	(0.036)
98% to <99%	-0.287***
	(0.040)
>=99% to 100%	-0.629***
	(0.033)
Number of Observations	7472

Notes: This table shows the impact of vaccination levels in the cross section on disease outbreaks. The unit of observation is a county-year and we do not adjust for any fixed effects. The model is an ordered probit with the outcome categories as in the Tables in the paper: <50th pctile of cases by county-year, 50-75th pctile, 75-95th pctile and >=95th pctile. The omitted category for vaccination is a vaccination rate of less than 96%. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Appendix B: Google Trends Data Production

Google Trends reports data in two ways. First, they report changes in search interest over time within an area. This is reported relative to the time in that area with the highest search interest. Second, they report differences in search interest across space within a given time period. This is relative to the area with the highest search interest. Our estimation is identified off of changes within a location over time so we focus on the first type of data. These data are generated in the following way:

First define the search rate for a query ($[query]$) in a given area z at time y:

$$\theta_{y,z} = \frac{\text{Number of searches for } [query] \text{ at time } y \text{ in area } z}{\text{Total number of searches at time } y \text{ in area } z}$$

Then trend data (τ) for a given area z over a time period $Y = \{y_1, ..., y_n\}$ can be expressed as:

$$\tau_{y,z} = \frac{\theta_{y,z}}{\max\limits_{y \epsilon Y}(\theta_{y,z})} \times 100$$

Note that Google Trends only calculates these values on a random sample of searches, and so the values may change depending on the time the website was accessed.

The trends data ranges from 0 to 100. A score of 0 however does not usually indicate no searches for the query; instead, it usually indicates that the volume of searches for the query did not meet Google's privacy threshold. While I do not know the exact cut-off for the threshold, in general data is easier to produce for more common queries, larger time periods (i.e. months versus weeks) and larger areas. The data also improves over time.

Many of the terms I am searching do not meet the privacy restrictions. I take two steps to overcome this. First, Google Trends allows me to use an 'or' connector so I can combine many queries (up to 30 words) related to a common topic, which then reports the sum of their trend scores. This still does not fully solve the problem so the second step uses elements of the method described in Stephens-Davidowitz (2014).

The methodology is straightforward. I take a common word that is unrelated to our terms of interest (this common word should meet the privacy threshold by itself). I then search for two terms in the same query: the common word ($[word]$), and the common word or the term of interest ($[word + term]$). For example, if our term was "pertussis" and the common word was "joke", I would search for "joke" and "joke or pertussis" at the same time. The difference between the two trends gives the trends for "pertussis". Note that the scores are still given from 0-100, but the data-point with the 100 score is now given to the relative highest search rate across both terms.

There are trade-offs with how to select the common word. As the common word becomes more popular it is more likely to consistently pass the privacy threshold, even in smaller areas and shorter time periods. However, this also increases the probability of having a small (or zero) difference between $[word]$ and $[word + term]$. This is because Google Trends are reported on a relative scale, so the term of interest's score becomes smaller relative to the common word's score as the popularity of the common word increases.

With this as the general background, I follow the detailed steps below.

1. Scrape the data for $[word]$ and $[word + term]$ together at the area-month level
2. Collapse both queries to the year level (mean monthly trend)
3. Recode both queries in a given year as missing if:
    (a) One of them has a mean monthly trend equal to zero (i.e. every month in that year was below the privacy threshold)
    (b) $[word]$ trend is greater than $[word + term]$ trend (i.e. a negative difference, which can occur due to the random sampling and is more likely if the term of interest trend score is relatively small)
4. Eliminate any area with less than two non-missing year observations
5. Re-scale the trends relative to the highest area-year score within each area and across both queries
6. Take the difference between the scaled $[word + term]$ and $[word]$ - this generates the trends for $[term]$
7. Re-scale again relative to the highest area-year score within each area

To balance the popularity trade off, I repeat these steps using five common words with varying levels of popularity (sponge, joke, fax, chair, rainbow). I then repeat this process over three different days and average the results of the 15 scrapes. I do this at the state-month level for the time period 2004-2015 ($Y = \{2004, ..., 2015\}$). The search queries I scraped are listed in Table B.1.

Table B.1: **Google Trend Search Queries**

Topic	**Search Queries**	**Example Captured Searches**
Whooping Cough	“pertussis”, “whooping cough”	“what is pertussis”, “whooping cough symptoms”
Whooping Cough Vaccine	“pertussis vaccine”, “pertussis vaccines”, “whooping cough vaccine”, “whooping cough vaccines”, “pertussis vaccination”, “whooping cough vaccination”, “pertussis immunization”, “whooping cough immunization”, “dtap”, “tdap”, “dpt vaccine”, “dtp vaccine”	“pertussis vaccine age”, “whooping cough immunization side effects”
Vaccine Risk	“vaccine injury”, “vaccine danger”, “vaccine dangers”, “vaccines risk”, “vaccines risks”, “vaccine side effects”, “vaccine side effects”, “mercury vaccine”, “mercury vaccines”, “thimerosal vaccine”, “thimerosal vaccines”, “vaccine ingredient”, “vaccine ingredients”	“mmr vaccine side effects”, “vaccines mercury content”
Autism-Vaccine Link	“autism vaccine”, “autism vaccines”, “autism immunization”, “autism immunizations”, “autism mmr”, “autism measles”, “mercury autism”, “thimerosal autism”	“vaccines autism study”, “do vaccines cause autism”, “immunization linked to autism”

Notes: This table shows the collection of search queries run for each topic, where each query is written within quotation marks. Google Trends searches for the exact word written (e.g. “vaccine” is different to “vaccines”), but searches are not case sensitive. A comma in the table above indicates an ‘or’ connector (e.g. “Term 1”, “Term 2”’ means “Term 1” or “Term 2”, which gives us “Term 1” trends plus “Term 2” trends). Google Trends captures all searches that include all the words within a query, in any order, and with any words before, after, or between them.