NBER WORKING PAPER SERIES

DOES DISEASE CAUSE VACCINATION? DISEASE OUTBREAKS AND VACCINATION RESPONSE

Emily Oster

Working Paper 22464
http://www.nber.org/papers/w22464

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2016

Min Kim, Motaz Al-Chanati and Kim Sarnoff provided outstanding research assistance. I thank Kate Baicker, Amitabh Chandra, Ben Olken, Jesse Shapiro and Andrei Shleifer for helpful comments. All mistakes remain my own. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Does Disease Cause Vaccination? Disease Outbreaks and Vaccination Response
Emily Oster
NBER Working Paper No. 22464
July 2016
JEL No. I12,I18

ABSTRACT

Childhood vaccinations are an important input to disease prevention, but vaccination rates have declined over the last decade due largely to parental fears about vaccine dangers. Education campaigns on the safety of vaccines seem to have little impact. Anecdotal evidence on disease outbreaks suggests that they may prompt vaccination behavior. I use newly compiled data on vaccinations and outbreaks to estimate whether vaccinations respond to disease outbreaks. I find that the pertussis vaccination rate increases among children at school entry following an outbreak in the year prior. A large outbreak in the county can decrease the share of unvaccinated children by 28% (1.2 percentage points). These responses do not reflect true changes in the future disease risk. I argue these facts may be explained by 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. I use survey data from health departments to show that states which directly coordinate outbreak responses have substantially larger vaccination increases in the wake of an outbreak, suggesting centralized management may better take advantage of this opportunity.

Emily Oster
Brown University
Department of Economics
64 Waterman Street
Providence, RI 02912
and NBER
emily_oster@brown.edu

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 many other developed countries) have declined. These declines are on average fairly small, but they are geographically dispersed, 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 in the US have roughly doubled since 2000, with a corresponding increase in deaths, primarily among infants.[1] A large measles outbreak in 2015 stemmed from unvaccinated children visiting Disneyland. Areas with low rates of vaccination have been shown to have an increased risk of outbreaks (Omer et al, 2006). 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]

Parental fears about vaccine safety are largely misplaced; vaccines do not cause autism, and adverse reactions are extremely rare (Mnookin, 2011). A natural policy response to low vaccination rates is, therefore, education for parents about vaccine safety (or, more generally, about the consequences of not vaccinating). Evidence does not suggest much impact of these educational approaches (Nyhan et al, 2014; Sadaf et a, 2013). An alternative response is to

[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.

change school vaccination rules to make it more difficult to enroll unvaccinated children in school. The latter approach appears to be more successful (Sadaf et al, 2013) but risks ghettoizing groups of non-vaccinated children in non-traditional schools, and therefore increasing disease outbreaks. Understanding whether failures of existing educational campaigns reflect the details of those campaigns or general lack of response to information is policy-relevant.

Response to information may also speak to theories of non-vaccination. It is possible to explain patterns of under-vaccination with fully informed, fully rational agents. Parents may (correctly) understand the disease risks to be small and they may weigh this against the (small but real) costs to vaccinations (child pain, discomfort) and decide not to vaccinate. Such parents would not be responsive to information as they are already fully informed. It is also possible that under-vaccination reflects some misinformation or uncertainty which could be altered with information.

This paper brings new data to bear on these issues. I combine (1) county-year data on vaccination among kindergarten children; (2) county-year data on outbreaks of disease and (3) Google search data on disease and vaccination related terms. I use these data to generate a series of facts about vaccination behavior in the cross section and over time within an area. I focus on the response of vaccinations to disease outbreaks. The central finding of the paper is that vaccination rates for pertussis show a sizable and significant response to pertussis outbreaks.

I demonstrate this finding in the first section of results. I show first that, in the cross section, vaccination rates vary with demographics - they are lower for areas with more education but higher for those with more income - and also correlate with Google searches for vaccination-related terms. Specifically, vaccination rates are higher in areas with a greater volume of search for information about the disease or vaccination, and lower in areas where searches for "autism and vaccines" are higher. This alone suggests a possible role for misinformation in driving low vaccination rates.

I then demonstrate that within a county over time, pertussis vaccination rates among kindergarten entrants are higher in the year after a pertussis outbreak.[3] The effect size suggest 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.

These results do not reflect changes in the true pertussis risk. Within a county, current outbreaks do not predict future outbreaks (if anything, the relationship is negative), implying that the impacts observed do not reflect responses to changes in actual disease risk. I also show the effects are concave: the per-case impact on vaccination is larger for the first cases of the disease relative to subsequent cases. Together, these are the primary facts which I will use to inform theory in Section 4.

This section also references a number of supportive analyses. I explore interactions and find that the strongest interactions are with initial vaccination level. Areas with very high vaccination rates do not respond much to outbreaks (in essence, there is not much change possible since their rates are already at or very close to 100%).[4] In addition I use evidence on other vaccinations and Google searches to provide further mechanism evidence. The impact of outbreaks is limited to pertussis vaccinations – it does not spill over to other vaccines – suggesting an instrumental response. I find that outbreaks increase searches for "pertussis" and "pertussis vaccination". This suggest information about outbreaks does spread among the population.

This evidence suggests people are responsive to local cases of disease, even holding constant the true risk of disease. The sizable impacts suggests that these marginal individuals may represent a substantial fraction of the unvaccinated population. In Section 4 I discuss what theory explains the behavior of these marginal individuals. 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

[3] 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.

[4] Outbreaks are also rare in places with high vaccination rates, although they do occur, for example among infants too young to be vaccinated.

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 and then they infer some positive probability. This model has the advantage that it can rationalize the response to disease outbreaks alongside the *lack* of response to information on vaccine safety seen in other data. Put simply, if someone perceives no disease risk we should not expect them to vaccinate even if they believe costs are low.

The key policy argument is that people may be responsive to information about disease even if it is not informative about the true risk. This suggests an important role for disseminating information about disease outbreaks. This dissemination is done at the state and local level, and Section 5 uses a survey of state and county health departments to explore the determinants of the magnitude of vaccination response.

The evidence from the survey reveals several things. First, it gives a sense of why the response to outbreaks is non-linear. Nearly all counties and states, regardless of population, classify a pertussis outbreak as 3 or fewer linked cases.[5] The protocols following this do not seem to differ with the size of the outbreak. Second, the survey makes clear that notifying affected (i.e. with a pertussis case) schools is a crucial part of the response in virtually all cases, which may help support the fact that we see a response among school-age children.

The survey identifies two important differences across areas. First, there is variation across states in whether the response to outbreaks is coordinated centrally by the state or region (4 states) or coordinated by the individual counties (8 states). In principle, either of these might be preferable. In practice, the evidence strongly suggest that the centrally-organized response dominates. The states with this dissemination structure have larger responses to outbreaks at all outbreak levels and, supportive of the mechanisms, also show larger responses in Google search volume. One interpretation of this is that a centralized structure generates more specialized experience: every state has some outbreaks in a given year, whereas many counties go years between outbreaks.

Second, I exploit variation across counties in states with county-coordinated responses. I

[5] An exception is some areas of California with a substantial pertussis burden where they wait for a larger number of cases - perhaps 10 - before responding.

look at two differences. First, whether the county systematically notifies the media about outbreaks. Second, at what case count the county defines an outbreak - some consider 1 or 2 cases an outbreak, others 3 or more. In both cases I adjust for the possibility of differences in response across states and across county population. I find that counties which have a systematic media notification plan have larger magnitude response, suggesting media notification may play a role. Further, counties which use a smaller number of cases to define an outbreak also see significantly larger vaccination responses. One interpretation of this is that if people are responsive to hearing about even one or two cases then there are wasted opportunities to influence behavior if we limit the definition of outbreaks to larger numbers of cases.

Overall the evidence in this paper suggests that disease outbreaks have an important impact on vaccination rates. A sizable fraction of vaccine-hesitant households seem to be responsive to these outbreaks. The constellation of facts suggests that some vaccine resistance may be driven by a perception that the risk of disease is zero. Evidence that this is not the case - provided by even a small number of cases of disease - can change behavior. The final analysis suggests that better centralized coordination of information dissemination about these outbreaks may encourage further response.

2 Data and Empirical Strategy

2.1 Data

This paper uses three primary data sets. The first is on vaccination rates, the second on disease outbreaks and the third is on Google searches.

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 was available for 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 vaccination. Summary statistics for the states used in the analysis appears in Panel A of Table 1. The years of coverage for states varies, as can be seen in this table. 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. The pertussis rate per 100,000 people ranges from 2 to just over 8 per year.

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 relates

[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.

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 and Varian (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. An issue with these data is that, although they do adjust for overall Google search volume, they do not adjust for domain-specific search volume. In particular, if there are some areas which simply use the internet more for health-related matters we are likely to observe them as high-search-volume areas for all the terms here. To address this, 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]

2.2 Empirical Strategy

The primary empirical findings in this paper focus on estimating vaccination behavior response to disease risk. For both the vaccination behavior and the Google searches, I use a panel data strategy in which I estimate the impact of disease outbreaks controlling for location and time fixed effects.

[7]A technical note: 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.

Vaccination Behavior

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

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

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 as the rate 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. The 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 by this since by this age they will already have several pertussis vaccinations. Any effects must therefore be driven by catch up vaccinations.

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.

However, 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. And since children cannot become unvaccinated, this generates a mechanical relationship.

To be more concrete, consider a county-year-cohort with low vaccination rates. The low vaccination rate prompts disease outbreaks, which then prompt an increase in vaccination rates. If we could observe vaccinations at age a and $a-1$ we could directly calculate the vaccination increase. However, we observe data only at a single age. If the vaccination rate is low at $a-1$ then it will be low (on average) in age a. This means if low vaccination rates prompt increased disease outbreaks we will underestimate the impact of outbreaks on vaccination.

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 rates. This is not very surprising. Changes in vaccination rates in a single cohort do not have a very large impact on the overall vaccination rate and it is this overall rate that drives disease. Put differently, 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. If this effect is truly 0, then β can be interpreted as a causal effect. If it is not zero, then β is an under-estimate of the effect.

Google Trends

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

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

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.

3 Results: Correlates of Vaccination and Response to 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 sub-section provides the primary evidence on the response of vaccination to pertussis outbreaks. Following this, I use evidence from vaccination for other disease and data from Google trends to providing supporting evidence suggesting that this impact reflects behavior change. Finally, in Section 3.4, I briefly discuss existing evidence on response to education campaigns as a counterpoint.

3.1 Correlates of Vaccination

Figure 1 shows evidence on changes in pertussis vaccination rates over time. The data for this figure are coefficients on year dummies from regressions which include county fixed effects. They are therefore identified off of changes within a county over time. Vaccination rates are trending downward over time. From the high point in the late-1990s, pertussis vaccination rates are down approximately 4 percentage points.

There is significant variation in pertussis vaccination rates across space. Evidence for this is shown in Figure 2. 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).

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).

The table shows that, in general, 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%.

3.2 Response to Outbreaks

Figure 3 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 3. 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.[8] The effects are larger for the smaller population areas, but they do not scale proportionally given the large differences in population across these groups. In Section 3.2.2 I return more specifically to the issue of functional form.

Table 3 shows the effect sizes in raw magnitudes and as a share of the unvaccinated population. It is also 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 kindergarteners. 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.

The primary results focus on the impacts of one-year-lagged outbreaks, but it is informative to look at the full time path. Figure 4 graphs coefficients on measures of large outbreaks, or the pertussis rate, over a number of years leading up to and following kindergarten entry. The effects are largest in the year prior to kindergarten entry although there are some smaller effects of lagged outbreaks. There is also some (insignificant) evidence of an impact of cases in the year of entry, which make sense given that entry is in September.

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

Future cases do not seem to matter, consistent with Column (5) of Table 3. The coefficients differ in this regression from Table 3 since the need to include future outbreaks limits the sample size.

The interpretation of the lags is somewhat complicated. Lagged outbreaks could matter because people remember them or, more likely, that they prompt vaccination behavior at the time. The fact that the effects are much smaller suggest that outbreaks in the immediate period before school entry matter more for vaccination than outbreaks when children are younger. This may reflect the fact that in the period before school entry parents must either vaccinate or obtain an exemption, thus making this a salient time for this choice. It would be interesting to observe impacts of outbreaks in infancy, when parents also make this choice, but looking this far back severely limits the data.

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 it is the other group with somewhat substantial coverage, likely 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) of Appendix Table A.2 shows the impact of measles outbreaks on measles vaccination for entering kindergarteners. 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.

3.2.1 Relationship between Current and Future Cases

One view of these relationships is that they reflect true changes in disease risk. If more cases of disease in the current year predict more cases of disease in future years then we should expect these changes in behavior even among individuals who are fully informed *ex ante.* 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; 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.

3.2.2 Functional Form of Relationship

Figures 5a and 5b provide more detail on the functional form of this relationship.

Figure 5a 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 5b 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.

3.2.3 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 6 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 6 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 do not 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.

3.3 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 a internet search response to the outbreak.

It is perhaps important to note that this paper is agnostic about whether the impacts here are driven by people hearing about the disease from the government, from the media, from their friends, from their doctor or in some other way.

3.3.1 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.

Table 5 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.

3.3.2 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 7a and 7b 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 7a estimates the impact of a linear control for number of cases. Figure 7b 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.4). 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 dangers.

3.4 Literature Evidence: Response to Education Campaigns

The evidence in this paper focuses on the response to disease. When thinking about the theory underlying this in the next section, it will be useful to combine this with existing evidence on the response to education about vaccine safety. I briefly summarize that evidence here. Broadly, education campaigns seem to be successful at changing perceptions about vaccine risks, but do not have impacts on vaccination or vaccination intentions.

Perhaps the most directly parallel paper is Nyhan et al (2014). This paper 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 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. In the end 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 2013) on the impact of educational interventions and conclude there is not much evidence to support the efficacy of particular educational interventions.

The evidence in the literature suggests that it is possible to alter beliefs about vaccine safety. However, such alterations do not seem to translate to intended or realized vaccination behavior. This stands in contrast to the changes in behavior observed above. In the next section I discuss what model of behavior may rationalize these facts.

4 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 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.

4.1 Setup

The model here considers a simple case in which vaccination decisions depend on the tradeoff 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 $D_{ij.}$ For simplicity given the focus of the work here, 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. I assume that 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, note that 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
$\frac{\partial S_{ij}}{do_j}\|\hat{p}_j > 0$	Vaccination increases with observed cases of disease, holding true disease rate constant
$\frac{\partial^2 S_{ij}}{\partial o_j^2}\|\hat{p}_j < 0$	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 particular model of limited information which I argue can fit all of these facts.

4.2 Fully Informed Model

Consider first a fully informed model of non-vaccination. The key aspect of the fully informed model is that the elements of the model will not change with observed cases (conditional on the true probability) or with education campaigns. We therefore have $C_{ij} = \hat{C}_{ij}$, and $p_{ij} = \hat{p}_j$. 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 $\hat{C}_{ij}$. That is, 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. In particular, it is straightforward to observe that $\frac{\partial S_{ij}}{do_j} = 0$ in the model, 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.

4.3 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.

Effectively, 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 parametrization 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.[9]

A key prediction - and policy implication - of this model is that making people aware of cases of disease may be a crucial input to increasing vaccination behavior. In the next section I use a final set of data - on the behavior of state and county health departments - to comment on policy.

[9]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.

5 Local Determinants of Vaccination Response

The model above suggests a need to understand the spread of information about disease outbreaks and, more importantly, whether there are policy behaviors which prompt larger responses. To do this, I conducted a survey of state and county health departments regarding their handling of disease outbreaks. I begin by summarizing some qualitative evidence from the survey. The second subsection looks at variation in response across states which either coordinate responses centrally or do not. The third subsection looks at variation in notification rules across counties.

A general caveat to this section is that there are a limited number of states and counties in the data, which makes it hard to draw strong causal conclusions. I view this analysis as suggestive.

5.1 Qualitative Conclusions

Local health departments - either the county or the state - are responsible for dealing with outbreaks of pertussis (and other notifiable diseases). There were several common themes among the notification structure. First, states and counties have rules about what constitutes a pertussis "outbreak" and therefore prompts a series of notification steps. In most cases the definition of an outbreak is between 1 and 3 cases and the cutoff is largely unrelated to the county population. This may provide some clue as to why the effects we observe do not scale much with population.

A second consistent theme is school notification. Virtually all states and counties have a policy of notifying schools or day care centers if a student at that location is diagnosed with pertussis.

States and counties differ in two ways, which I exploit below. First, for a subset of the states the response to disease outbreaks is centralized (either at the region or state level). In others, the response is coordinated by the county. The first subsection below exploits this difference. Second, within those states which coordinate response at the county level, there is variation across counties in (a) how many cases they consider necessary to define an outbreak (i.e. 1 versus 2 versus 3) and (b) the media notification process. This is explored in Section 5.3.

5.2 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 behest of the state or regional office, which coordinates the response. In the other eight states counties have autonomy with respect to their response to disease. 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 6 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 (which will typically cross county lines and involve the state anyway).

Columns (3) - (6) of Table 6 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 7. Motivated by that figure I simply 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.

5.3 Variation across Counties

In the states without central coordination, I focus on two differences across counties. First, whether the county said they have a systematic way to notifying the media about outbreaks. Second, the count of the number of cases required to define an outbreak. This figure is typically between 1 and 3, although in some cases it was higher (or based on historical figures). I define counties as having a low notification threshold if they define an outbreak as 1 or 2 cases.

We focus on the eight states with county-autonomy in notification. There are 828 counties in these states, of which 244 of them never have any outbreaks in the course of our data. Among the others, I was able to collect data from 137. 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 6 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 either (a) a dummy for whether the county typically notifies the media or (b) a dummy for whether the outbreak trigger case number is low. These regressions also include interactions between the county population and the outbreak variables and between state dummies and these variables. This ensures that any differences we pick up are not due to differences across states in notification policies.

The results are shown in Table 7. Column (1) shows the role of media notification and Column (2) shows the variation in response by the outbreak trigger level. The evidence on the media suggests that the impacts of large outbreaks are significantly higher in areas with a media contact plan.. The evidence in Column (2) shows response to both small and large outbreaks is larger in places where they consider a lower cutoff value to define an outbreak. Again, this is true even controlling for population interactions so this is not driven by differences in county size. The fact that these counties show larger response to large outbreaks - which would qualify as outbreaks in all counties - suggests that this variable may be picking up something other than just the trigger value. It may, for example, indicate counties which take pertussis cases more seriously in their policy approach.

Together, the results from this analysis and Section 5.2 indicate that the structure of how states respond to disease outbreaks matters in whether these outbreaks prompt vaccination behavior.

6 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 kindergarteners 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. Put simply, I observe large changes in vaccination behavior with no change in objective 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 which use lower disease thresholds for defining an outbreak 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

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

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.

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.

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.

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.

——, **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.

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.

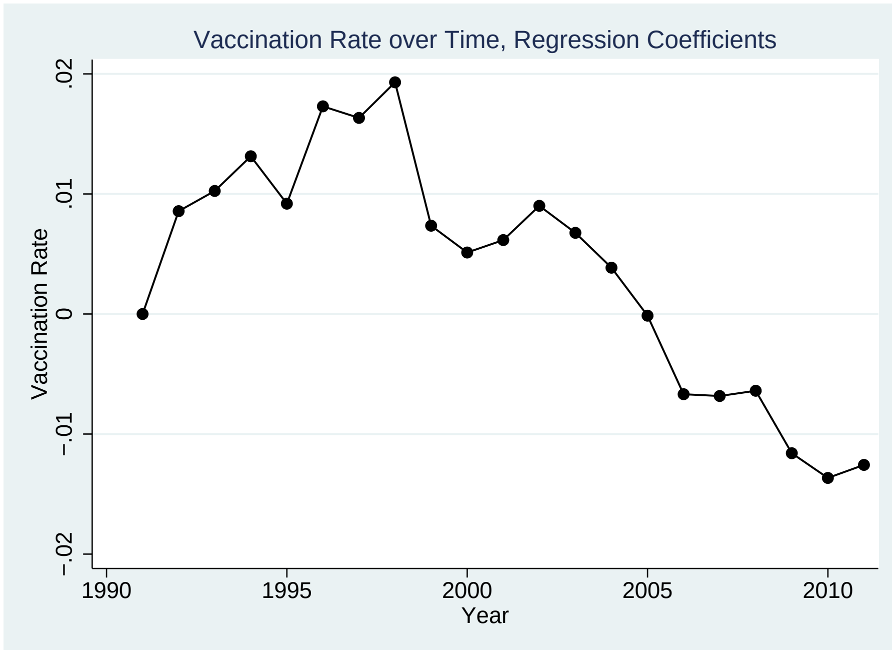
Salmon, Daniel A., Lawrence H. Moulton, Saad B. Omer, Patricia deHart, Shannon Stokley, and Neal A. Halsey, "Factors associated with refusal of childhood vaccines among parents of school-aged children: A case-control study," *Archives of Pediatrics and Adolescent Medicine*, 2005, *159* (5), 470–476.

Stephens-Davidowitz, Seth and Hal Varian, "A Hands-on Guide to Google Data," 2014.

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.

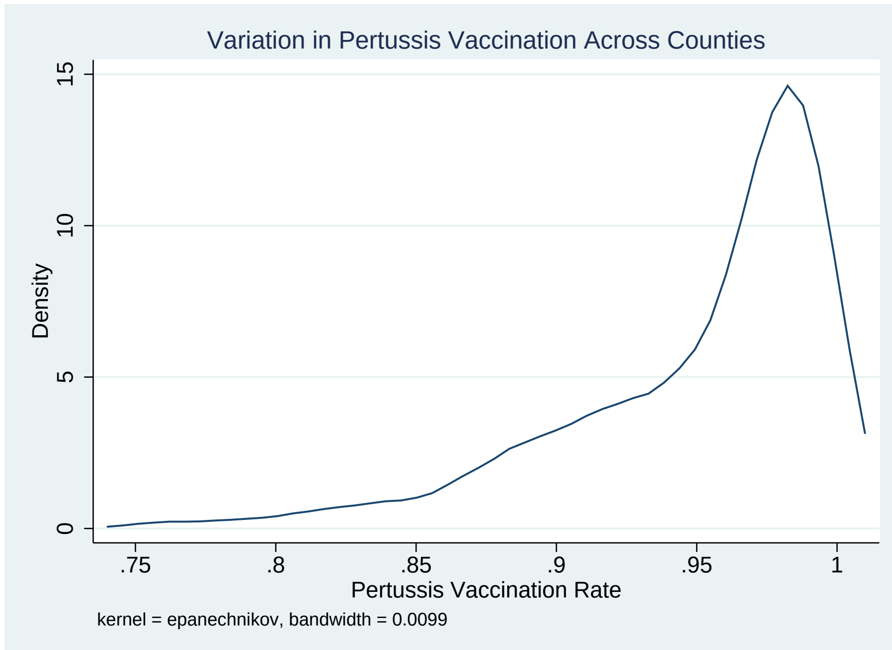
Yang, Tony Y. and Ross D. Silverman, "Legislative Prescriptions for Controlling Nonmedical Vaccine Exemptions," *The Journal of the American Medical Association*, 2015, *313* (3), 247–248.

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



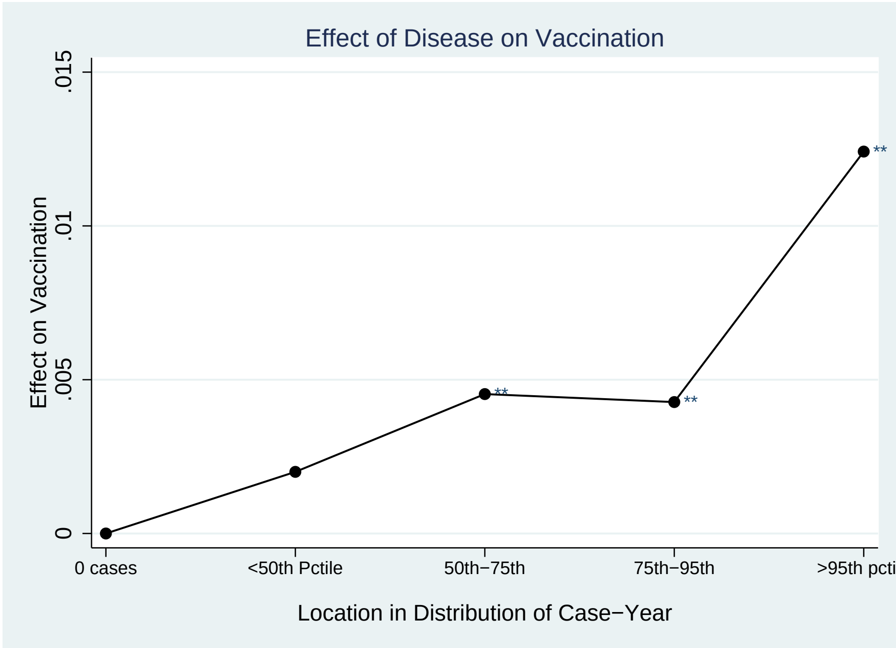
Notes: This graph shows estimates of changes in pertussis vaccination rates over time. The observations are coefficients on year dummies from regressions of vaccination rate on these dummies and county fixed effects. The regressions use an unbalanced panel, although do control for county fixed effects.

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



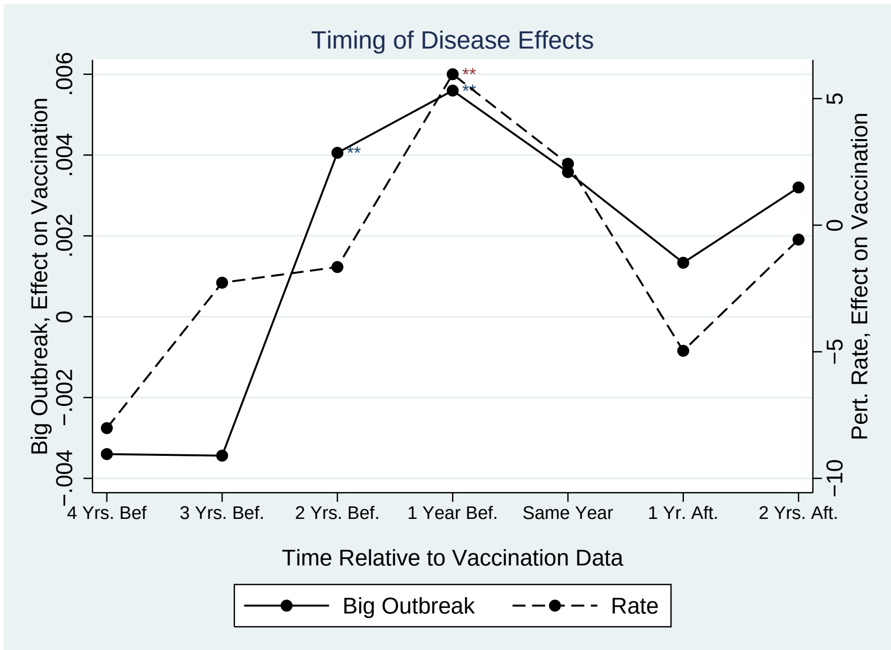
Notes: These figures show the density of vaccination rates across counties in the 2010/2011 period. I average vaccination rates for 2010 and 2011 and drop the bottom 1% of counties.

Figure 3: **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.

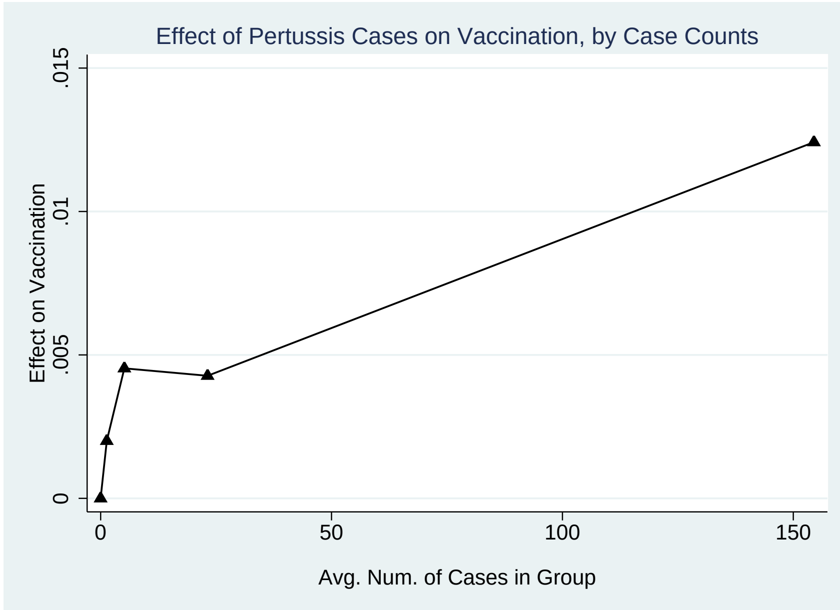
Figure 4: **Timing of Disease Impacts**



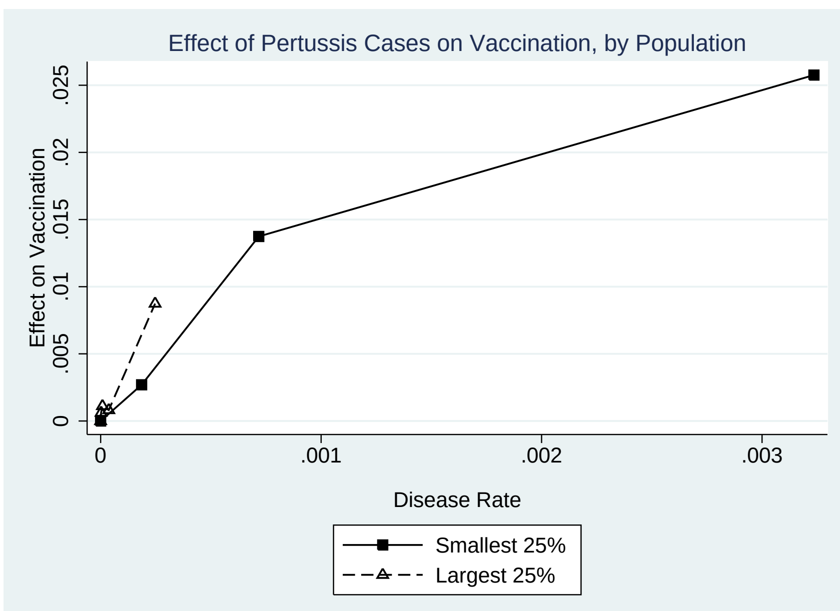
Notes: These figures show the impact of disease cases in the county on vaccination behavior in lags and leads. The outbreak is measured either as a dummy for a large outbreak or as the disease rate. The vaccination data is at the time of school entry and the outbreaks are measured in the year prior.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

Figure 5: **Functional Form Analysis**

(a) Vaccination Impact by Case Count

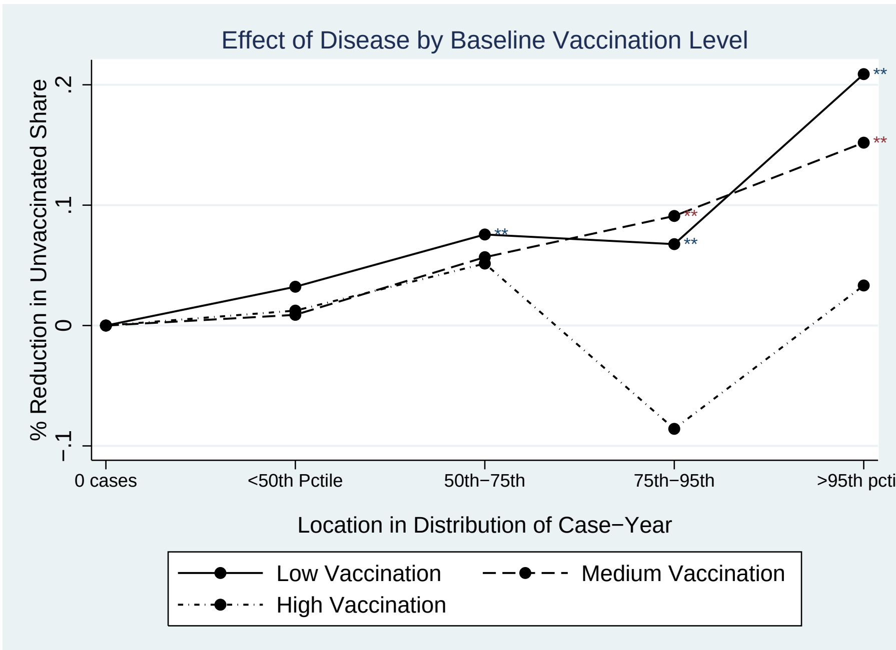


(b) Vaccination Impact by Population



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.

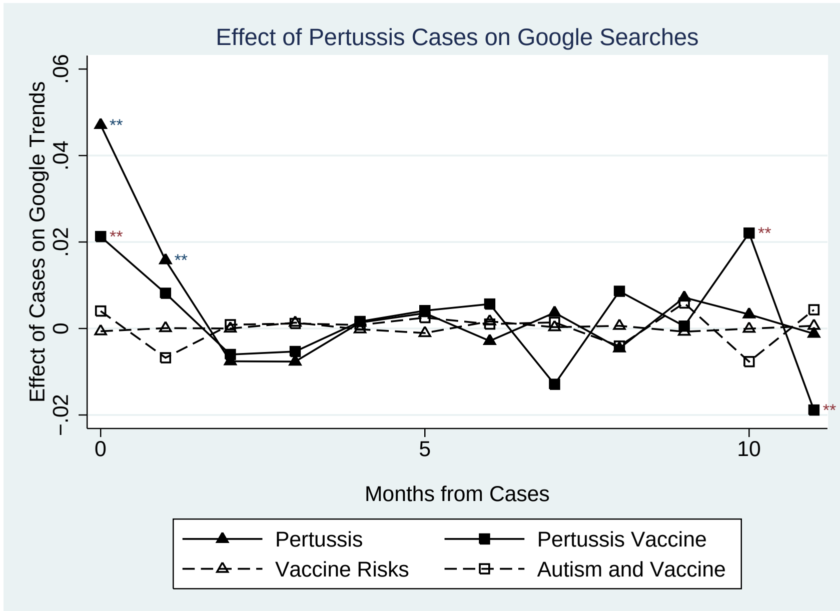
Figure 6: **Interactions with Vaccination Levels**



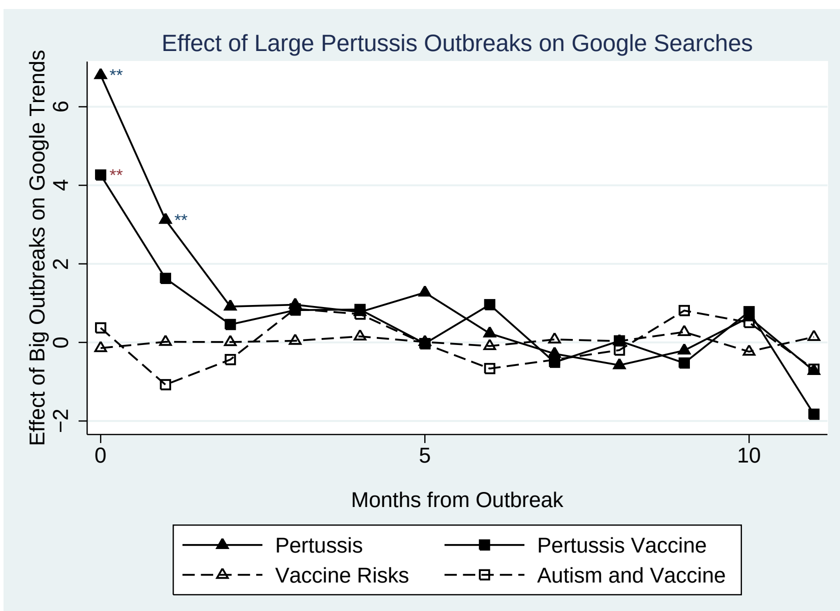
Notes: These figures show the impact of outbreaks on vaccination separated by areas with high and low vaccination rates. Counties are divided into three groups based on the minimum county-year vaccination rate. All points shown are regression coefficients from regressions with county and year fixed effects,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

Figure 7: **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.

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 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)
<50th Pctile of Cases	0.002 [4.5%]		0.001 [3.2%]	0.003* [5.8%]	0.002 [5.0%]
	(0.001)		(0.001)	(0.001)	(0.001)
50th to <75th Pctile	0.0045*** [10.3%]		0.0021 [4.8%]	0.0056*** [12.8%]	0.0055*** [12.5%]
	(0.001)		(0.001)	(0.002)	(0.002)
75th to <95th Pctile	0.0042***[9.7%]		0.002[4.1%]	0.0059***[13.5%]	0.0049***[11.0%]
	(0.002)		(0.002)	(0.002)	(0.002)
>=95th Pctile	0.012***[28.1%]		0.011***[24.4%]	0.010***[23.0%]	0.014***[31.8%]
	(0.003)		(0.003)	(0.003)	(0.004)
# 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 [-4.5%]
					(0.001)
50th to <75th Pctile, t+1					-0.002 [-5.5%]
					(0.002)
75th to <95th Pctile, t+1					-0.003 [-6.4%]
					(0.002)
>=95th Pctile, t+1					-0.010 [-24.4%]
					(0.009)
County FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
County-Specific Trends	NO	NO	NO	YES	NO
R-squared	0.58	0.58	0.58	0.70	0.58
Number of Observations	7472	7472	7472	7472	6422

Notes: This table shows the impact of pertussis outbreaks on vaccination rates. Everywhere except Column (2) outbreaks are defined in groups. The omitted category is 0 cases. The groups are then based on the distribution of positive county years. "Low pop" (Column (3) 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: **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 6: **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 7: **Impacts of Outbreaks by County Notification Details**

Outcome:	*Pertussis Vaccination*	
	(1)	(2)
<95th Pctile, $t-1$	-0.015**	-0.015*
	(0.008)	(0.008)
>=95th Pctile, $t-1$	0.005	-0.008
	(0.007)	(0.011)
<95th Pctile × Media Notified	0.004	
	(0.005)	
>=95th Pctile, × Media Notified	0.016**	
	(0.007)	
<95th Pctile × Low Outbreak Trigger		0.008*
		(0.005)
>=95th Pctile, × Low Outbreak Trigger		0.017***
		(0.007)
County FE	YES	YES
Year FE	YES	YES
Population X Outbreak Controls	YES	YES
State Dummy X Outbreak Controls	YES	YES
R-squared	0.65	0.65
Number of Observations	2569	2458

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 in Column (1) are those between the outbreak groups and whether the county reports systematically notifying the media of outbreaks. Those in Column (2) are between the outbreak dummies and a dummy for reporting that either 1 or 2 cases is considered an outbreak. The alternative is reporting 3 or more cases, or reporting that it is undefined (rare). Both 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 A: Figures and Tables

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 kindergarteners. 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 outbreaks on vaccination rates, using either county-level outbreak measures. The case groups are defined as follows. Pertussis, county level: 0 cases (omitted), 1 case, 2-10 cases, 11-100 cases, 101-1100 cases. Measles, county-level: 0 cases (omitted), 1-10 cases, 11-20 cases, 21-50 cases.. Columns (3), (4) and (7), (8) separate by whether there is any media coverage of the disease in that state-year. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table A.4: **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

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”
MMR	“mmr”	“mmr vaccine”
Measles	“measles”	“measles outbreak”
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 (e.g. “mmr vaccine” captures “vaccine mmr” and “vaccine schedule for mmr”).